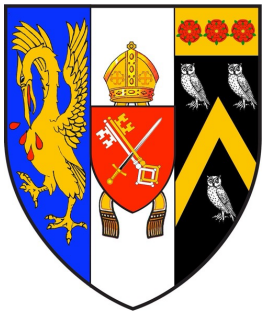
Characterising Haematopoietic Stem Cell Heterogeneity Using Single-cell Multi-omic Sequencing





Bowen Zhang

Corpus Christi College

Radcliffe Department of Medicine

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Medical Science

Hilary term 2025

Abstract

Haematopoietic stem cells (HSCs) sustain lifelong blood production through their unique ability to self-renew and differentiate into all blood lineages. Despite being classically defined as a homogeneous population, HSCs are now recognised as functionally diverse, with distinct subpopulations exhibiting stable lineage biases. Understanding the molecular basis of this heterogeneity is critical for advancing stem cell research and developing improved therapies for haematological disorders.

This thesis uses single-cell multi-omic sequencing to profile the transcriptomic, epigenomic and proteomic landscapes of mouse HSCs across developmental stages and during ageing. A key achievement is the development of an integrated analytical pipeline to infer HSC lineage potential from single-cell data. This enables the identification of molecularly distinct adult HSC clusters corresponding to *in vivo* lineage potentials. Using surface protein data, a fluorescence-activated cell sorting (FACS) panel was designed to prospectively isolate lineage-biased HSCs and allow ageing phenotyping without reliance on transgenic reporter mice.

This study also examines how ageing alters HSC composition and function. Lineage-biased HSCs expand with age and display increased quiescence, while multilineage HSCs show

inflammatory signatures, suggesting susceptibility to stress and functional decline. These changes could contribute to reduced lymphoid output and immune dysfunction in aged individuals.

Furthermore, a subset of young HSCs shares molecular features with foetal liver HSCs and exhibits activation of both lineage bias and multilineage programmes was identified, indicating a transitional population during the maturation of foetal to adult haematopoiesis.

While this work provides significant insights, limitations remain, including technical constraints of single-cell technologies and the need for validations in the human system. Future efforts should expand multi-omic sequencing to broader timepoints, species and associated haematopoietic progenitors. Overall, this thesis advances our understanding of HSC heterogeneity and provides a valuable framework for future investigations.

Acknowledgements

As I conclude this journey, I do not feel a sense of immense relief, nor was this experience particularly gruelling or arduous. The three years have passed smoothly and calmly, and I have genuinely enjoyed being fully dedicated to research. As I reach the end of my twenty-year student life, I find myself feeling somewhat reluctant to leave it behind. However, time waits for no one, and graduating from one of the world's best institutions is already a satisfying chapter of my life. This period will undoubtedly be worth reminiscing about time and time again.

My deepest gratitude goes to my wife, Yifan Bao, who came with me to Oxford and stood by my side through every step of this journey. You have been my unwavering pillar of strength, your calm resilience and devotion have continuously grounded and reassured me. I feel incredibly fortunate for all that you've given and even more so for the life we get to share together. Thank you, for everything.

I am deeply grateful to my parents for providing me with a strong financial foundation and the freedom to grow into who I am today. At every moment, I feel that I have stood on the shoulders of giants. I wish them good health and a joyful life.

I would like to thank my supervisors, Dr. Yiran Meng, for patiently teaching me laboratory

skills and helping me build a solid foundation. My principal supervisor, Prof. Claus Nerlov, has been central in shaping my ability to think critically and approach challenges as a scientist. I wish my junior colleagues, Simiao Zhao and Hazel Zhang, all the best in their academic journeys. I am also thankful for my colleagues and the facility staff at the Weatherall Institute, who have contributed to the creation of an outstanding academic environment.

I am extremely thankful to the mentors and teachers who have guided me throughout my academic Journey: Prof. Haonan Lu introduced me to the field of bioinformatics and data-driven science. Dr. Holly Simpson Ragdale offered me my first exposure to stem cell biology. Prof. David Selwood gave me the opportunity to participate in scientific research for the first time. I also owe much to my two former teachers, Mrs. Liwen Dong and Mr. Guangxin Liu, who set the example of a righteous person in my childhood.

Special thanks to my friends and the best travel mates at Oxford: Xingyi Yan (thank you for being my best man), Xinran Huang, Dr. Yu Hao and Boss Zehui Duan. Also, to Byron Lee, the motoring godfather, and to Oscar Lu from Shanghai. I deeply appreciate the memories we have created together.

My heartfelt gratitude extends to my high school friends: Jeffrey Sin (congratulations on your wedding), Thomas Feng, Freddie Jiang from Wujin, Changzhou, South Jiangsu,

Marco Yeung, Yijia Ma and Xinyang Pei (your braised beef brisket is unbeatable). You have been irreplaceable companions throughout my teenager years.

I am also must mention my lifelong friends: Haozhe Sun, Chuhan Wang, Yanmeng Shi (Doctor in progress), Zhuo Ai, Meixuan Wu, Yang Yu and Zhenxi Zhang, who have always supported me. May our friendship endure for years to come.

My college, Corpus Christi College, deserves a special recognition: formal hall dinners were always a delight, the summer garden was like a fairy tale, and the heated library benches were amazing. I extend this gratitude to the University of Oxford, and I hope to live long enough to witness your millennium celebration.

A special thanks to Four Seasons restaurant, your roast duck is the light of Britania.

I take pride in this achievement, brought about by a combination of my good fortune and perseverance. May I continue to keep a humbled heart, embrace life with passion, uphold integrity and carry forth with an unyielding spirit.

Lastly, I wish for the continued prosperity of my homeland and the enduring progress of human civilisation.

Declaration

This D.Phil. thesis is supervised by Dr. Yiran Meng, Prof. Paresh Vyas and Prof. Claus Nerlov at the Medical Research Council Weatherall Institute of Molecular Medicine (MRC WIMM), University of Oxford.

The single HSC clonal expansion and their bulk-RNA, ATAC sequencing were performed by Dr. Yiran Meng. The lineage potential signature gene and chromatin region lists were identified by Dr. Yiran Meng. The bone marrow harvest of various experiments was assisted by Dr. Yiran Meng, Dr. Xiying Ren, Yifan Bao, Xinran Huang and Zishan Zhang. The machine-learning-based label transfer was co-developed with Simiao Zhao. The *Gata1*-eGFP transgenic mice were bred and maintained by Dr. Alexandre Fagnan. All cell sorting works were assisted by the WIMM Flow Cytometry Facility. All other experiments and bioinformatic analyses were performed by the thesis author.

Contents

Abstract **1**

Acknowledgements **3**

Declaration **6**

Contents **7**

List of Figures **11**

List of Tables **13**

Abbreviations **14**

Chapter 1: Introduction **18**

1.1 The physiology of mammalian haematopoietic stem cells and their clinical significance 18

1.2 The haematopoietic hierarchy 19

1.3 HSC heterogeneity and lineage potential 22

1.4 Epigenetic regulation of HSC function 30

1.5 Ageing of the HSC compartment 31

1.6 Techniques for investigating HSC heterogeneity 37

1.7 Aims, research questions and potential impacts 51

Chapter 2: Methods 54

2.1 Animal 54

2.2 Bone marrow and foetal liver cell harvest 54

2.3 Haematopoietic cell c-Kit enrichment 55

2.4 c-Kit enriched cell preparation for FACS 55

2.5 Peripheral blood cell preparation for FACS 56

2.6 Single and bulk-HSC transplantation 56

2.7 Bulk RNA and ATACseq (Yiran Meng) 57

2.8 CITEseq antibody preparation 59

2.9 DOGMAseq single-cell multi-omic sequencing 60

2.10 Flow cytometry 62

2.11 Bioinformatic analysis 63

2.12 Fluidigm single-cell qPCR 69

Chapter 3: Single-cell multi-omic sequencing for mouse HSCs 71

3.1 Introduction and aims 71

3.2 Identification of HSC lineage potential signatures using bulk RNAseq data 72

3.3 Development of CITEseq antibody panel 73

3.4 Conjugation of CITEseq antibodies 76

3.5 Optimisation of DOGMAseq workflow for HSCs 78

3.6 Discussion 82

Chapter 4: Determining adult HSC lineage potentials *in silico* 86

4.1 Introduction and aims 86

4.2 HSC data quality control (QC) 88

4.3 Multi-omic unsupervised clustering of adult mouse HSCs 89

4.4 Annotation of HSC subpopulations through per-cell lineage potential scoring 90

4.5 Discussion 94

Chapter 5: Developing FACS-enrichment panel for HSC subpopulations 99

5.1 Introduction and aims 99

5.2 Enrichment of platelet-biased HSCs through *in silico* gating 101

5.3 FACS marker combination enriches *Vwf*+ HSCs 103

5.4 Differential expression of lineage potential genes in FACS-purified HSC subpopulations 105

5.5 Discussion 109

Chapter 6: Confirming lineage potential of FACS-isolated HSCs in transplantation 114

6.1 Introduction and aims 114

6.2 Bulk transplantation 115

6.3 Single-cell transplantation 120

6.4 Discussion 121

Chapter 7: Studying HSC subpopulation dynamics during ageing 126

7.1 Introduction and aims 126

7.2 Transfer of HSC subpopulation labels from 8-week to 24-month dataset 128

7.3 Investigation of ageing features 132

7.4 Discussion 137

Chapter 8: Evidence of postnatal HSC diversification 143

8.1 Introduction and aims 143

8.2 Identification of chromatin remodelling activities in 4-week HSCs 144

8.3 Discussion 148

Conclusion 152

Supplementary Figures 158

Supplementary Tables 161

Bibliography 165

List of Figures

Figure 1 Models for the haematopoiesis hierarchy.....22

Figure 2 HSC subpopulations with stepwise lineage-restricted reconstitution potential.....26

Figure 3 HSC ageing is associated with altered epigenetic landscape and clonal composition.35

Figure 4 DOGMAseq workflow for single-cell multi-omic sequencing.....50

Figure 5 Clonal expansion of single HSCs reveals candidate CITEseq targets.....75

Figure 6 Development of the CITEseq antibody panel.....77

Figure 7 Optimised DOGMAseq workflow enhances library quality.....82

Figure 8 8-week HSC dimensional reduction and unsupervised cluster annotation.....94

Figure 9 In silico gating simulation for enriching HSC clusters.....102

Figure 10 FACS marker combination enriches *Vwf*+ HSCs.....104

Figure 11 FACS-isolation using new markers enriches mouse HSC subpopulations.....108

Figure 12 Bulk transplantation of FACS-isolated HSC subpopulations.....119

Figure 13 Single-cell transplantation of FACS-isolated PLT-HSCs.....120

Figure 14 Transferring adult HSC cluster labels to the aged HSC dataset.....131

Figure 15 HSC ageing features in RNA modality.....133

Figure 16 HSC ageing features in ATAC modality.....136

Figure 17 Chromatin remodelling activities in 4-week HSCs.....147

Supplementary Figure 1 Dimensionality reduction components.....158

Supplementary Figure 2 FACS panel for transplantation reconstitution readouts.....159

Supplementary Figure 3 Mitochondrial genotyping.....160

Supplementary Figure 4 Fetal liver HSC WNN UMAP and lineage potential scoring.....160

List of Tables

Table 1 CITEseq antibody panel used in DOGMAseq experiments....78

Table 2 TaqMan probes used in Fluidigm experiments....106

Supplementary Table 1 Gating strategies in FACS analyses....161

Supplementary Table 2 Antibodies used in FACS analyses....162

Supplementary Table 3 Buffers....163

Supplementary Table 4 Primers....164

Abbreviations

ADT: Antibody derived tag

AML: Acute myeloid leukaemia

ATACseq: Assay for transposase-accessible chromatin using sequencing

BM: Bone marrow

CFU-S: Spleen colony forming unit

CH: Clonal haematopoiesis

CITEseq: Cellular indexing of transcriptomes and epitopes sequencing

CLP: Common lymphoid progenitor

CMP: Common myeloid progenitor

DE: Differential expression

FACS: Fluorescence-activated cell sorting

FCS: Foetal calf serum

FL: Foetal liver

G: Gravitational constant

GFP: Green fluorescent protein

GMP: Granulocyte-monocyte progenitor

GSEA: Gene set enrichment analysis

HSC: Haematopoietic stem cell

hHSC: Human haematopoietic stem cell

HSPC: Haematopoietic stem and progenitor cell

lncRNA: Long non-coding RNA

IMDM: Iscove's modified Dulbecco's medium

LMPP: Lymphoid-primed multipotent progenitor

LSC: Leukaemic stem cell

LSI: Latent semantic indexing

MFI: Mean fluorescence intensity

MKP: Megakaryocyte progenitor

MPP: Multipotent progenitor

mt-RNA: Mitochondrial-RNA

MULTI: Multilineage

NMP: Neutrophil macrophage progenitor

P: Platelet

PBMC: Peripheral blood mononuclear cell

PE: Platelet-erythrocyte

PEM: Platelet-erythrocyte-myeloid

PEMB: Platelet-erythrocyte-myeloid-B-cell

PB: Peripheral blood

PBS: Phosphate-buffered saline

PCA: Principal component analysis

RBC: Red blood cell

RFE: Recursive feature elimination

RT: Reverse transcription

TPM: Transcripts per million

TSS: Transcription start site

UMAP: Uniform manifold approximation and projection

WBC: White blood cell

WNN: Weighted nearest neighbour

WT: Wild-type

Chapter 1: Introduction

1.1 The physiology of mammalian haematopoietic stem cells and their clinical significance

Haematopoietic stem cells (HSCs) are the cornerstone of the mammalian haematopoietic system, responsible for replenishing mature blood cells required for maintaining homeostasis, immunity and oxygen transport throughout an organism's lifetime. Located predominantly in the bone marrow, HSCs possess the unique abilities of self-renewal and multipotent differentiation, enabling the generation of different blood cell lineages, including platelets, erythrocytes, myeloid cells and lymphocytes (Figure 1A).

The clinical significance of HSCs is exemplified by their central role in bone marrow transplantation, a life-saving therapy for patients with haematological malignancies, severe immune deficiencies and other blood disorders. Since its inception in the mid-20th century, bone marrow transplantation has been performed extensively, saving countless lives, highlighting the therapeutic potential of HSCs (Niederwieser et al. 2022). However, haematopoietic stem and progenitor cells (HSPCs) are also recognised for their potential to transform into leukaemic stem cells, a process contributed by genetic mutations, epigenetic alteration and inflammatory signalling. These transformed cells retain self-

renewal properties, allowing them to propagate malignancies that lead to the progression and relapse of haematological cancers (Colom Díaz, Mistry & Trowbridge 2023).

Emerging research has also focused on the mechanisms of HSC ageing, as aged HSCs show functional decline, causing deteriorated immune system and impaired engraftment potential (Muller-Sieburg et al. 2012; De Haan & Lazare 2018). Strategies aimed at rejuvenating aged HSCs, such as targeting signalling pathways like mTOR (Chen et al. 2009) and TGF-β (Xu et al. 2018; Valletta et al. 2020), and improving bone marrow niche conditions (Ramalingam et al. 2023), hold promise for restoring youthful HSC function.

Despite significant progress, understanding the molecular mechanisms that govern HSC function remains a critical research area with implications for regenerative medicine, ageing, and cancer biology.

1.2 The haematopoietic hierarchy

Haematopoiesis is a highly regulated hierarchical process, evolved to meet the physiological demands of the organism, and its study has advanced significantly over the last decades. Foundational discoveries began with the identification of clonogenic progenitor cells capable of forming spleen colonies (spleen colony-forming units, CFU-S), in irradiated mice (Till & McCulloch 1961). This was followed by the purification of long-term repopulating haematopoietic stem cells (HSCs) using surface markers and

transplantations, which demonstrated only a subset of bone marrow cells could self-renew and give rise to all blood lineages over serial transplantations (Spangrude, Heimfeld & Weissman 1988; Ikuta & Weissman 1992). These studies established the fundamental structure of the haematopoietic hierarchy, placing HSCs at the apex. HSCs give rise to multipotent progenitors (MPPs), which, although lacking durable self-renewal, retain capacity to differentiate into multiple blood lineages (Morrison et al. 1995). MPPs then progress into lineage-restricted unipotent progenitors, which are committed to specific branches of haematopoiesis. This multi-tiered structure of self-renewing, multipotent, unipotent and terminally differentiated cells forms the core framework of the haematopoietic system.

A major refinement of this model came with the discovery of common lymphoid progenitors (CLPs) and common myeloid progenitors (CMPs), representing a bifurcation in lymphoid and myeloid lineage commitment downstream of HSCs (Akashi et al. 2000). Further downstream, CMPs were shown to give rise to granulocyte-monocyte progenitors and megakaryocyte-erythroid progenitors, each restricted to specific myeloid outputs. Adding complexity to this framework, lymphoid-primed multipotent progenitors (LMPPs) were characterised as progenitors with lymphoid and granulocyte-monocyte potential, but lacking megakaryocyte and erythroid potential, indicating an earlier separation of these lineages than previously thought (Adolfsson et al. 2005). Additional granularity was

provided by the GATA1 model, which used *Gata1*-eGFP reporter mice to stratify myeloid progenitor populations into *Gata1*-expressing (GE+) and *Gata1*-negative (GE-) subsets. GE+ cells generated eosinophils, mast cells, megakaryocytes, and erythroid cells, while GE- cells gave rise to monocytes and neutrophils – evidence for an alternative segregation combination of myeloid lineages (Drissen et al. 2016). While these models differ in their precise definitions and branching logic, they all independently support the concept of stepwise lineage commitment occurring early in haematopoietic differentiation. Despite not necessarily agreeing on the order or nature of lineage restriction, these frameworks converge on the idea that fate decisions emerge from discrete progenitor populations with progressively limited potential, rather than from a single gradual continuum (Figure 1A) (Jacobsen & Nerlov 2019).

Upstream of these committed progenitors, studies have revealed that lineage bias arises even at the MPP level, previously thought to be a homogeneous transition state between HSCs and downstream progenitors. A key development was the definition of four functionally distinct MPP subsets – MPP1, MPP2, MPP3, and MPP4. MPP1, closely related to short-term HSCs, retains broad multilineage potential. In contrast, MPP2 shows a bias toward megakaryocyte production, MPP3 favours granulocyte-monocyte output, and MPP4 is skewed toward lymphoid differentiation (Figure 1B) (Pietras et al. 2015; Cheng, Zheng & Cheng 2020). These subsets emerge prior to full lineage commitment and

reflect an organised restriction of differentiation capacity. In this thesis, we adopt the MPP1–4 framework to characterise early stages of HSC differentiation, as it captures both the hierarchical structure and functional diversity of the haematopoietic progenitor landscape.

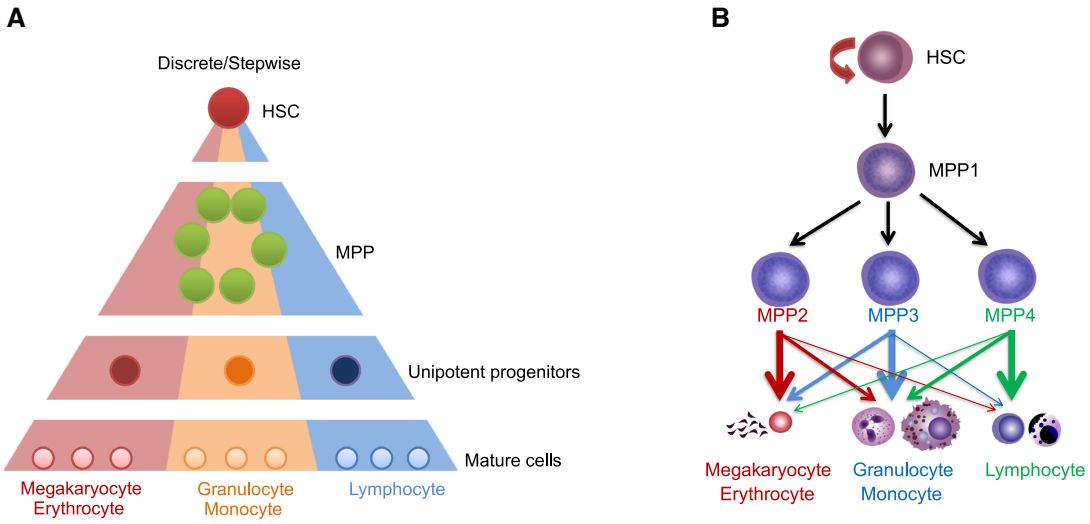


Figure 1 Models for the haematopoiesis hierarchy (Cheng et al. 2020)

A, The discrete differentiation model suggests that self-renewal HSCs differentiate into mature lineages through a stepwise process, following a tree-like hierarchy that progress through multipotent, bipotent and unipotent progenitors. **C,** In the refined model, the non-self-renewal multipotent progenitor population consists of MPP1(short-term HSCs), which give rise to MPP2, MPP3 and MPP4 that are biased to the megakaryocyte, myeloid and lymphoid lineages respectively.

1.3 HSC heterogeneity and lineage potential

1.3.1 Emergence and definition of lineage potential

Initially, HSCs were regarded as a uniform population of multipotent cells capable of giving rise to all blood lineages. Subsequent advances in clonal analysis, limiting dilution assays, and surface marker refinement revealed that even immunophenotypically defined HSCs

exhibit substantial functional variability. Using single-cell transplantation, studies demonstrated that individual mouse HSCs differ in lineage output and regenerative capacity, leading to the recognition that HSC compartment is functionally heterogeneous. Since then, numerous studies have expanded on this idea, showing that distinct HSC subsets possess stable lineage biases and differences in self-renewal, proliferation and engraftment potentials (Dykstra et al. 2007; Yamamoto et al. 2013).

Lineage potential has emerged as a key axis of HSC heterogeneity. Transplantation assays have revealed subsets of adult bone marrow HSCs that preferentially give rise to specific lineages: lymphoid, myeloid, erythroid or platelets, alongside a fraction with true multilineage capacity. These lineage preferences are stable across serial transplantations, suggesting HSC subpopulations are intrinsically programmed. This was further clarified by the use of fluorescent lineage reporters such as *Vwf*-eGFP and *Gata1*-eGFP, which enabled tracking of donor-derived platelet and erythroid output and revealed a progressive continuum of lineage potentials, from platelet-restricted HSCs to fully multipotent HSCs capable of regenerating all lineages including lymphocytes (Figure 2) (Carrelha et al. 2018). Although all lineage-restricted HSCs retain multipotency *in vitro*, their *in vivo* reconstitution patterns in transplantation setting suggest a functional phenotype shaped by both intrinsic and extrinsic factors.

In this thesis, we focus on the lineage potential aspect of HSC heterogeneity. We define

lineage potential based on the reconstitution patterns observed in *in vivo* transplantation assays, which reflect the ability of groups or individual HSCs to regenerate specific blood lineages under stress conditions. However, these outcomes may not mirror actual HSC fate decisions under homeostatic conditions, nor do they necessarily indicate intrinsic multipotency assessed by *in vitro* colony-forming capacity. Nonetheless, transplantation remains the gold standard for functionally evaluating HSC subpopulations, particularly in understanding their regenerative behaviour and lineage preference.

It is important to distinguish between the terms HSC 'lineage bias' and 'lineage restriction', as they are often used inconsistently across the field. These terms were originally developed in the context of progenitor populations, to describe the stepwise commitment of multipotent cells as they progressively lose the potential to give rise to multiple lineages. In the case of HSC biology, however, the use of 'bias' has been applied in a broader and sometimes ambiguous manner.

At the population level, 'lineage bias' often refers to the enrichment toward certain lineages, most commonly the myeloid compartment (Pang, Schrier & Weissman 2017). This observation may arise from changes in the ratio of functionally distinct HSC subpopulations, or from extrinsic factors such as physiological demands or microenvironmental influences. In contrast, at the single-cell level, 'lineage bias' is used to describe individual HSCs that generate higher contributions to certain lineages. A common

operational criterion is a three-fold higher output to one or few lineages relative to other lineages, assessed by the stable peripheral blood (PB) chimerism level in transplantation experiments (Carrelha et al. 2018; Meng & Nerlov 2024).

It is worth considering that the level of chimerism, and, consequently, the classification of lineage bias, may be influenced by the bone marrow condition. Factors such as the number of residual host cells, inflammatory status, and recipient age may all impact donor cell behaviour (Pei et al. 2017; Pei et al. 2020; Bogeska et al. 2022). However, lineage-biased HSCs have thus far been characterised mainly in transplantation settings with tightly controlled recipient conditions. Whether these HSCs would exhibit the same and stable lineage contributions under different physiological contexts or during homeostasis remains an open and important question.

On the other hand, studies using barcoding and single-cell transplantation have demonstrated that the capability to generate certain lineages, is remarkably stable across diverse conditions (Sanjuan-Pla et al. 2013; Carrelha et al. 2018; Jacobsen & Nerlov 2019; Belander Strålin et al. 2023). For example, single HSCs lacking lymphoid reconstitution under one condition consistently fail to do so under others, suggesting a hardwired restriction.

Therefore, in this thesis, we use 'lineage bias' to refer to the enrichment of certain lineages

in the PB, observed during ageing or in bulk transplantation assays. For single HSC analyses, we use 'lineage restriction' to define phenotypes based on the presence or absence of detectable output to a given lineage above a predefined minimal threshold, irrespective of the chimerism level. This approach allows us to classify single HSCs by their intrinsic lineage potential programming.

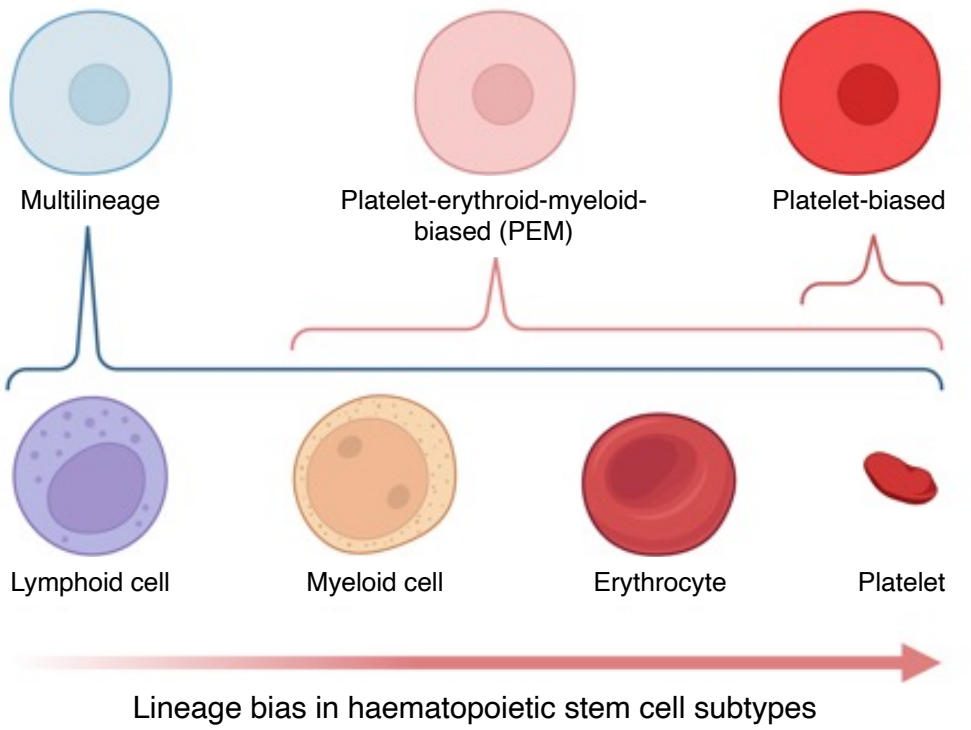


Figure 2 HSC subpopulations with stepwise lineage-restricted reconstitution potential

Mouse and human HSCs exhibit various potentials to long-term reconstitute blood cell lineages *in vivo*, ranging from HSCs restricted/biased to produce platelet, erythrocytes and myeloid cells, to multilineage HSCs that produce all lineages in a balanced output. All these HSC subtypes have been shown to retain multipotency with *in vitro* colony formation assays (Carrelha et al. 2018).

1.3.2 Dynamics, cross-species conservation and relevance of HSC lineage bias

Studies in foetal liver (FL) HSCs further illustrate how lineage potential evolves over developmental stages. FL HSCs are widely considered to be functionally homogeneous, with a consistent capacity to generate all blood lineages. However, (Benz et al. 2012) have observed instances of myeloid or lymphoid-skewed reconstitution from transplanted E14.5 FL HSCs. In contrast, other work using perinatal HSCs found that clones exhibiting lymphoid-biased reconstitution failed to engraft in secondary recipients, suggesting that such lymphoid-biased pattern may result from the longer half-life of lymphoid progeny rather than true long-term HSC function (Belander Strålin et al. 2023). The same study also observed no evidence of platelet restriction and only low degree of myeloid bias, reinforcing the notion that fully committed HSC subtypes are largely absent at this stage.

Further supporting FL HSC uniformity, studies using *Vwf* reporter mice showed that E14.5 FL HSCs are uniformly *Vwf*-positive (Sanjuan-Pla et al. 2013). This suggest that at this developmental stage, *Vwf* expression does not distinguish functionally polarised subsets, in contrast to adult bone marrow where *Vwf* enriches platelet-biased HSCs. Single-cell transplantation experiments also revealed that BM *Vwf*+ HSCs with multilineage potential can generate both *Vwf*+ and *Vwf*- HSCs in recipients, whereases *Vwf*- HSCs do not give rise to *Vwf*+ progeny. Notably, the *Vwf*+ HSC population in adult bone marrow

comprises a mixture of platelet-biased and multilineage HSCs, with the latter responsible for this bidirectional output. These findings imply that segregation into lineage-biased subtypes occurs downstream of multilineage *Vwf*+ HSCs and is not established during foetal liver development. Moreover, it raises the possibility that the adult bone marrow niche in transplantation analyses imposes lineage constraints that may not reflect the intrinsic developmental fate of FL HSCs, particularly given that the 16-week duration of transplantation assays exceeds the natural timeframe in which FL HSCs transition into fully mature adult HSCs. Whether fully lineage-restricted HSCs begin to emerge in the foetal stage remains an open question. Nonetheless, the prevailing consensus across studies is that FL HSCs are significantly more uniform in self-renewal and lineage output than their postnatal counterparts. The acquisition of stable lineage bias appears to accelerate rapidly after birth, coinciding with the transition of HSCs to the bone marrow niche and the onset of epigenetic diversification.

Furthermore, evidence suggests that during the first eight week of mouse life, the HSC compartment comprises a mixture of adult-like HSCs and HSCs with foetal-specific transcriptional programmes, which disappear after week eight (Christensen et al. 2004; Kim, Saunders & Morrison 2007). However, the molecular mechanisms governing this transition remain poorly defined, presenting a critical area for understanding how HSCs acquire functional diversity during maturation.

Importantly, HSC heterogeneity is conserved across species. Recent studies using high-throughput single-cell RNAseq and molecular barcoding in xenografted human HSCs (hHSCs) have identified lineage-biased and multilineage subpopulations, demonstrating that adult hHSCs share functional characteristics with murine counterparts (Aksöz et al. 2024).

The concept of lineage bias carries important implication for regenerative medicine and ageing. Multilineage HSCs are critical for robust lymphoid regeneration, while lineage-biased, particularly those skewed toward platelet and myeloid lineages are associated with enhanced quiescence and self-renewal (Sanjuan-Pla et al. 2013; Chen et al. 2017; Pei et al. 2017; Carrelha et al. 2018). Under normal conditions, this balance might support haematopoietic homeostasis and immune readiness. However, during ageing, the overall immunophenotypical HSC pool expands (Sudo et al. 2000), but this expansion is dominated by lineage-biased HSCs. In contrast, multilineage HSCs exhibit reduced functional potency, contributing to diminished lymphoid output (Gekas & Graf 2013; Yamamoto et al. 2013; Grover et al. 2016). This age-related skewing impairs immune function and contributes to increased susceptibility to infections and haematological disorders in the elderly (Müller-Sieburg et al. 2002; Cho, Sieburg & Muller-Sieburg 2008; Pang et al. 2011). Collectively, understanding the mechanisms driving HSC heterogeneity, and in particular, the determinants of lineage potential have wide-reaching implication for

stem cell therapy, immune rejuvenation and the treatment of age-associated haematological diseases.

1.4 Epigenetic regulation of HSC function

Epigenetic mechanisms, including DNA methylation, histone modifications and chromatin remodelling, play a crucial role in regulating HSC self-renewal and differentiation (Itokawa et al. 2022). These modifications influence gene expression without altering the DNA sequence, enabling HSCs to dynamically respond to physiological cues. A key method for investigating the epigenetic landscape of HSCs is the assay for transposase-accessible chromatin using sequencing (ATACseq), which identifies regions of open chromatin that correspond to active regulatory elements such as enhancers and promoters (Buenrostro et al. 2013; Itokawa et al. 2022). This technique utilises a hyperactive Tn5 transposase to cut and insert sequencing adaptors into accessible DNA regions, providing a detailed map of chromatin accessibility and enabling researchers to link epigenetic modifications to HSC functions. For instance, it was shown that the DNA methyltransferase DNMT3A is essential for HSC differentiation, with its loss leading to increased self-renewal but impaired differentiation, contributing to haematological malignancies (Challen et al. 2011; Challen et al. 2014).

As HSCs are largely quiescent, their functional potential is not only defined by active

transcriptional programmes but also embedded through epigenetic priming, which dictates their ability to differentiate when activated. A recent study has demonstrated that lineage potential in HSCs is also rooted in this epigenetic priming, with chromatin accessibility signatures at lineage-specific regulatory elements serving as predictive markers for differentiation outcomes. ATACseq has revealed that multilineage HSCs exhibit selective accessibility at lymphoid-specific elements, while platelet-restricted HSCs show enhanced accessibility at platelet-specific regions (Meng et al. 2023). These chromatin patterns, established through epigenetic priming, remain stable in quiescent HSCs and across serial transplantations, effectively encoding their differentiation potential. Additionally, primed lineage-biased HSCs exhibit accelerated lineage commitment in response to physiological signals, highlighting their role under emergent haematopoiesis (Carrelha et al. 2024).

1.5 Ageing of the HSC compartment

The functional decline of HSCs significantly contributes to the ageing of the haematopoietic system. Ageing is associated with an increase in the overall number of HSCs, but these cells exhibit reduced regenerative capacity and a pronounced lineage bias. This shift results in compromised immune function, impaired haematopoietic regeneration and increased susceptibility to haematopoietic malignancies (Yamamoto et al. 2018).

1.5.1 Epigenetic drift and clonal dynamics during HSC ageing

Epigenetic drift, defined as age-associated alterations in DNA methylation and chromatin accessibility, plays a crucial role in the functional decline of the HSCs compartment during ageing. This process is driven by factors such as replicative stress, genomic instability and increased inflammatory signalling (Shao et al. 2011; Kramer & Challen 2016; Wang et al. 2025). Aged HSCs exhibit impaired self-renewal, platelet-erythrocyte-myeloid-biased lineage output and reduced capacity to produce lymphocytes. These changes are linked to disruptions in the epigenetic landscape, including alterations in DNA methylation and histone modifications at regulatory regions associated with stem cell maintenance and lineage differentiation (Mortera-Blanco et al. 2013; Teschendorff, West & Beck 2013; Fujino et al. 2021).

A possible pathological outcome of age-related epigenetic drift is clonal haematopoiesis (CH), where mutant HSC clones with advantageous epigenetic alterations outcompete their counterparts, resulting in clonal dominance (Figure 3A). For example, *DNMT3A* mutations, which result in hypomethylation at transcription factor binding sites, such as those for *MYC* and *HIF1A*, promote gene expression linked to stemness and clonal expansion (Nam et al. 2022). Similarly, *TET2* mutations disrupt demethylation, skewing HSC differentiation towards the myeloid lineage while enhancing survival in an inflammatory environment (Izzo et al. 2020; Schiroli et al. 2024). These mutation-driven

clonal expansions can progress to haematological malignancies, such as acute myeloid leukaemia, myelodysplastic syndromes and myeloproliferative neoplasms (Genovese et al. 2014). Beyond malignancies, CH is also linked to non-haematological diseases, including cardiovascular conditions such as atherosclerosis, myocardial infarction and venous thrombosis. For instance, studies suggest that *TET2* mutations increase inflammatory responses in macrophages, exacerbating vascular damage and promoting the development of atherosclerotic lesions (Jaiswal et al. 2017; Marston et al. 2024).

Single-cell sequencing and functional studies have also revealed shifts of subpopulation composition within the HSC compartment. Ageing is associated with an expansion of lineage-biased HSC subpopulations and reduction of multilineage HSCs (Figure 3B) (Cho et al. 2008; Grover et al. 2016; Yamamoto et al. 2018). This shift not only reflects lineage bias but also highlights increased heterogeneity in aged HSC compartments. For example, a recent study demonstrated that HSC subsets separated by CD49b expression, undergo disproportional expansion during ageing. CD49b- HSCs, which are enriched for myeloid-biased potential, exhibit a greater expansion with age, while CD49b+ lymphoid-biased HSCs decline (Su et al. 2024). The increase in lineage bias with ageing is also observed in our reporter mouse model as we detected more expression of *Vwf*, a marker for platelet-erythrocyte-myeloid-bias in aged mouse HSCs. Importantly, it should be considered that multilineage HSCs produced greater numbers of mature blood cells across all lineages

compared to their biased counterparts (Grover et al. 2016; Pei et al. 2017). Therefore, a functional decline in multilineage HSCs with ageing may not only impair lymphoid output but also reduce overall erythropoiesis and myeloid function, contributing to broader haematopoietic insufficiency (Rossi et al. 2005; De Haan & Lazare 2018). Taken together, these observations support the trend toward lineage-biased HSC dominance in aged haematopoiesis.

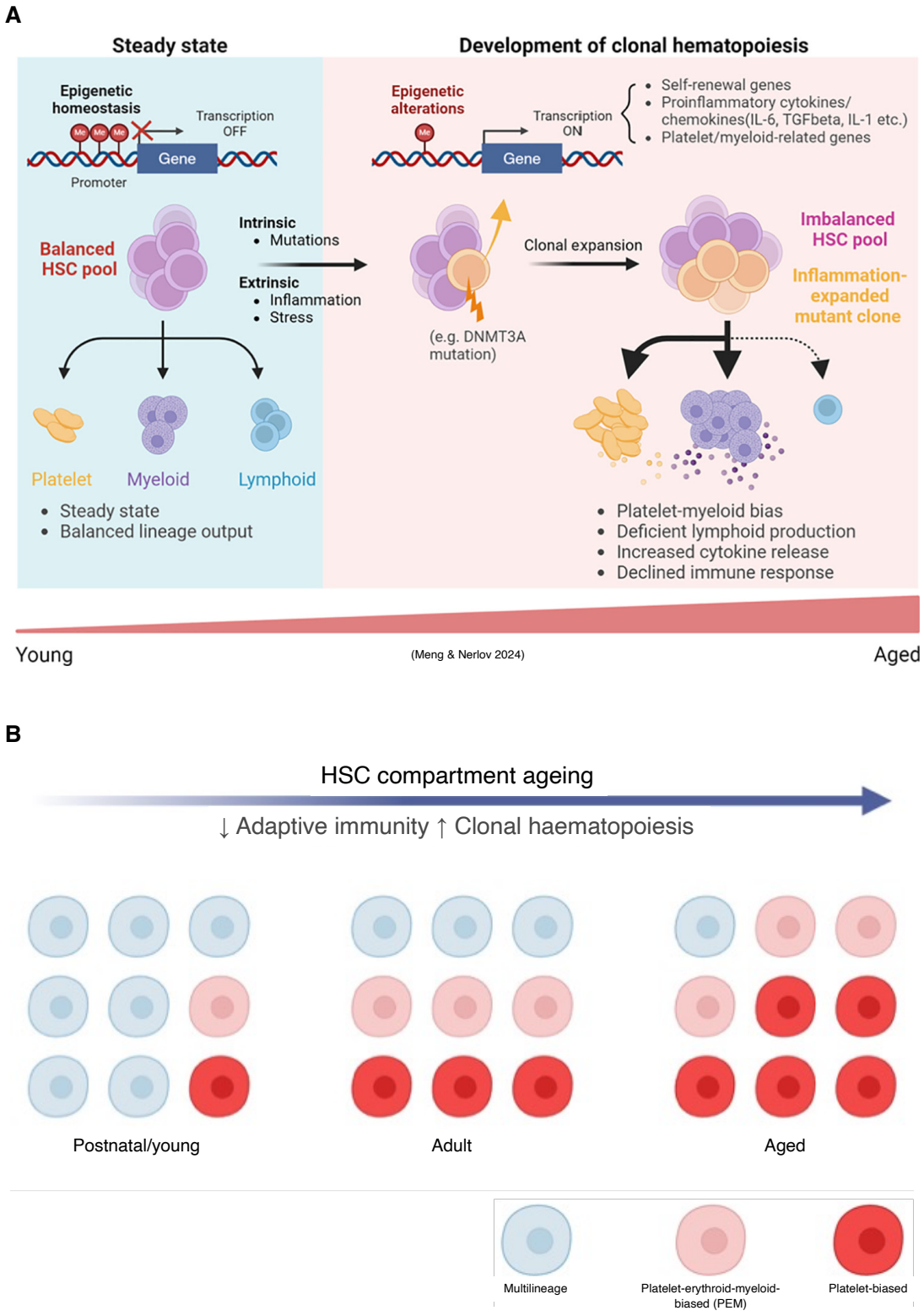


Figure 3 HSC ageing is associated with altered epigenetic landscape and clonal composition

A, Epigenetic drift contributes to age-related myeloid-bias and clonal haematopoiesis (Meng & Nerlov 2024)**.** In young individuals, haematopoiesis maintains a balanced lineage output under steady-state conditions. With ageing, increased genomic instability and inflammatory stress disrupt epigenetic regulation and promote the expression of genes associated with self-renewal and inflammation. These changes result in enhanced

myeloid output at the compartment level and favour the expansion of certain HSC clones. Mutations in epigenetic regulators such as DNMT3A and TET2 are commonly associated with this process, as they confer a competitive advantage to mutant HSCs, allowing them to expand disproportionately. **B,** While the entire HSC compartment increases in number with age, lineage-biased HSCs expand at a greater rate than multilineage HSCs, leading to skewed ratio. This shift contributes to a decline in lymphopoiesis and a subsequent reduction in adaptive immunity.

1.5.2 Reversibility and therapeutic implications

Addressing age-related haematopoietic decline requires a nuanced understanding of both universal ageing features and HSC subpopulation dynamics. Given that HSCs exist as functionally distinct subpopulations with embedded lineage potentials, therapeutic approaches must account for this heterogeneity. Strategies that fail to consider subpopulation dynamic risk inadvertently favouring certain clones, potentially exacerbating dysfunctions.

Pan-ageing factors such as inflammatory and oxidative stress, contribute to the progressive deterioration of HSC function. These factors lead to quiescence and impaired regenerative potential. Intervention targeting these systemic factors, such as anti-inflammatory treatments or approaches that mitigate oxidative damage, have shown promise in improving overall HSC function (Caiado, Pietras & Manz 2021). Modulating metabolic pathways including mTOR inhibition and autophagy enhancement (Leveque et al. 2013; Luo et al. 2014; Dellorusso et al. 2024) has been proposed as a strategy to rejuvenate aged HSCs by restoring their proliferative capacity.

Beyond systemic interventions, restoring youthful haematopoiesis requires strategies that specifically target HSC subpopulation imbalances. Some studies explored directly manipulating HSC subpopulations composition. For instance, antibody-mediated depletion of myeloid-biased HSCs in aged mice has demonstrated improvements in lymphoid regeneration, reduction in myeloid-driven inflammation and enhanced immune function (Ross et al. 2024). These findings suggest that selective modulation of HSC subpopulations could complement systemic interventions for a more robust haematopoietic rejuvenation strategy.

1.6 Techniques for investigating HSC heterogeneity

HSCs display remarkable heterogeneity, with subsets varying in molecular profiles, lineage potentials and regenerative capacities. Early investigation of HSC biology primarily relied on population-level analyses which provided foundational insights but averaged over diverse cell states. These approaches were instrumental in establishing the classical hierarchical model of haematopoiesis and identifying functional properties such as self-renewal and multilineage differentiation. However, they were limited in their ability to resolve intrinsic variability between individual HSCs.

The advent of single-cell technologies, particularly single-cell RNAseq, ATACseq and multi-omic platforms has transformed the field by enabling high-resolution profiling of

transcriptional and epigenetic states at the level of individual cells. These tools have revealed that phenotypically defined HSC populations are composed of distinct subpopulations with stable lineage restrictions and transcriptional priming, challenging the notion of a functionally homogeneous HSC pool. Together with genetically modified animal models and single-cell transplantation assays, these modern methodologies now offer a comprehensive toolkit for dissecting HSC heterogeneity.

1.6.1 Transplantation assays for mouse HSC

The foundational understanding of HSC function originated from early transplantation experiments, where bone marrow cells from donor mice were shown to rescue lethally irradiated recipients. These studies led to the development of spleen-forming unit assays, which measures the number of macroscopic colonies that form in the spleen of recipient mice within 8-12 days post-transplant, providing a readout of short-term progenitor activity and haematopoietic potential. (Till & McCulloch 1961). Building on these findings, the limiting dilution transplantation was developed to quantify HSC frequency and potency based on engraftment outcomes at defined cell dose, while the competitive repopulation assay became the benchmark for comparing repopulating ability of donor HSCs against a standard competitor population (Szilvassy et al. 1990).

Transplantation assays remain the gold standard for assessing HSC functionality, especially

their long-term reconstitution potential and lineage output. Long-term HSCs (LT-HSCs) are defined operationally by their capacity to sustain multilineage haematopoiesis over extended periods and across serial transplantations. This makes transplantation assays indispensable for distinguishing LT-HSCs from short-term progenitors (Benz et al. 2012; Yamamoto et al. 2018).

In a typical mouse HSC competitive transplantation experiment, donor HSCs (as few as one cell) are isolated and transplanted alongside competitor bone marrow cells into lethally irradiated recipient mice. The competitor bone marrow ensures a functional haematopoietic system post-irradiation. Following transplantation, peripheral blood (PB) is analysed at week 6, 10 and 16 with each time point reflecting distinct phases of donor HSC activation (Carrelha et al. 2018):

- Week 6: Platelets and erythrocytes are typically the first lineages to emerge, reflecting their rapid differentiation and shorter maturation time.
- Week 10: Myeloid cells become detectable with some early signs of lymphopoiesis, usually B cells.
- Week 16: Lymphocyte production, particularly CD4 and CD8 T cells, is observed. Lymphoid reconstitution requires longer maturation, making week 16 the appropriate time point for evaluating stable reconstitution and long-term engraftment.

At week 16, donor HSC reconstitution ability is assessed by measuring PB chimerism across platelet, erythrocyte, myeloid and lymphoid lineages (Supplementary Table 1). BM reconstitution is also analysed to confirm stable engraftment by the presence of donor-derived HSCs. In addition to PB lineage outputs, different HSC subpopulations exhibit distinct pattern of MPP production (Meng et al. 2023). For instance, lineage-biased HSCs generate a higher proportion of MPP2 and MPP3, which supports megakaryocyte and myeloid lineages differentiation; whereases multilineage HSCs produce a more even distribution of MPP subpopulations, sometimes favouring MPP4, which is associated with lymphopoiesis. These difference in MPP ratios highlight the functional heterogeneity of HSC subpopulations and their specific contributions to haematopoietic reconstitution.

Practically, lineage reconstitution was defined as > 0.1% donor-derived cells (identified by positive GFP or CD45.1 staining), based on prior analysis of negative control samples. These controls indicated that 0.1% represents the threshold above which true donor signal can be reliably distinguished from background noise on the flow cytometer (Carrelha et al. 2018), and thus serves as the operational detection limit for defining positive lineage contribution.

For secondary transplantation, stably reconstituted donor-derived HSCs are re-isolated from the bone marrow of primary recipients and transplanted into new lethally irradiated recipients. This secondary assay further tests the self-renewal capacity and lineage fidelity

of the transplanted HSCs. Single-cell experiments have demonstrated that platelet, erythrocyte and myeloid-restricted HSCs sustain their restricted differentiation potential across multiple rounds of transplantation, while multilineage HSCs maintain broader differentiation capacity (Carrelha et al. 2018). The fact that lineage-restricted HSCs maintain their reconstitution patterns even under the stress of serial transplantation suggests that heterogeneous HSC subsets are molecularly and functionally distinct, with stable lineage priming that persists even in extreme regenerative conditions. This highlights the long-term stability of HSC heterogeneity.

Together, these transplantation protocols provide a detailed picture of HSC function, cementing their role as an essential technique in HSC research.

1.6.2 Mouse models and the *Vwf* reporter system

Mouse models remain integral to understanding HSC biology *in vivo*. The *Vwf* reporter system (eGFP or tdTomato) has been a pivotal in defining platelet-erythrocyte-myeloid-biased (PEM) HSC subpopulations. In practice, approximately 60% of phenotypically defined 8-week mouse HSCs (Supplementary Table 1) express *Vwf*, a marker enriches for cells with biased PEM differentiation (Carrelha et al. 2018; Carrelha et al. 2024). The *Vwf* reporter enables fluorescence-activated cell sorting (FACS) isolation of these HSCs, facilitating downstream molecular and functional studies. By separating HSCs based on

Vwf expression, we have observed that *Vwf*-expressing HSCs increase in frequency with ageing, supporting the observed shift toward myeloid-biased haematopoiesis. Additionally, this model has revealed distinct kinetics among HSC subpopulation as *Vwf*+ HSCs exhibit accelerated thrombopoiesis, generating platelet more rapidly than their *Vwf*- counterparts (Carrelha et al. 2024). Overall, the *Vwf* enrichment efficiency has been studied across multiple ages in mouse models. It is also associated with megakaryocytic lineage differentiation, and has been widely used in studies to assess lineage bias. Thus, the *Vwf*-eGFP model serves as a useful benchmark for evaluating and developing new markers of lineage-biased HSCs.

In addition to marking HSC lineage potential, the *Vwf*-eGFP/tdTomato and *Gata1*-eGFP (Drissen et al. 2016) reporter lines are used to track donor-derived platelet and erythrocyte populations in transplantation assays. These cell types lack CD45.1 or CD45.2 antigen, commonly used for distinguishing donor and recipient white blood cells. By leveraging these reporters, researchers can robustly quantify donor platelet and erythrocyte output, determining the lineage potential of the transplanted HSCs.

1.6.3 Other platforms for functional HSC studies

In vitro colony formation assays demonstrate that lineage-restricted HSCs identified *in vivo* retain their intrinsic multipotency, a defining characteristic of HSC (Miller & Lai 2005).

For example, platelet-restricted HSCs, which only contribute to megakaryocytes and platelets *in vivo*, can form colonies comprising diverse lineages, including megakaryocyte, myeloid and lymphoid precursors, when stimulated with appropriate cytokines under controlled conditions (Carrelha et al. 2018). The assay provides functional confirmation that these subpopulations meet the classical criteria for HSCs.

Barcoding technologies have become powerful tools for studying clonal dynamics, lineage bias and functional heterogeneity in HSCs. A recent study used lentiviral barcoding to uniquely label purified human HSCs (hHSCs), which were then transplanted into immunodeficient mice. This approach enabled clonal tracking by sequencing the differentiated progeny, allowing the lineage contributions of individual hHSCs to be monitored over time. This system provided strong evidence that adult hHSCs, like their murine counterparts, comprise distinct subpopulations with stable lineage restrictions, including platelet-restricted and multilineage clones (Aksöz et al. 2024). Complementing this, the Polylox barcoding system provides an endogenous clonal tracing strategy for studying mouse HSC dynamics under physiological conditions. This method uses an artificial loxP cassette with high recombinatorial complexity, integrated into the genome and activated via Cre recombinase to generate unique barcodes *in vivo*. Unlike transplantations, Polylox barcoding enables fate mapping without altering HSC function, the study has revealed early lineage segregation patterns and supported models in which

myeloid and lymphoid reconstitution pattern varies between distinct HSC clones (Pei et al. 2017; Pei et al. 2020).

The emerging *ex vivo* culturing systems offer another valuable platform for studying HSC biology. Optimised culture conditions, including niche-mimicking matrices and cytokine cocktails (e.g., thrombopoietin, stem cell factor), allow for the long-term expansion and maintenance of HSCs (Wilkinson et al. 2019; Johnson et al. 2024; Meaker & Wilkinson 2024). These systems enable detailed studies of HSC responses to external stimuli and therapeutic interventions, and they are particularly advantageous for screening compounds and performing gene manipulations at a lower cost compared to animal models (Haney et al. biorxiv).

Together, these tools enable comparative studies of HSC biology across species while *ex vivo* cultures provide a controlled, cost-effective alternative approach for target screenings. Integrating these approaches complements our understanding of HSC heterogeneity, lineage bias and therapeutic potential.

1.6.4 Bulk RNA and ATAC sequencing

Bulk RNA sequencing (RNAseq) was one of the earliest omic approaches used to study HSCs, providing a comprehensive overview of the transcriptional profile in a cell population. While powerful for identifying global gene expression patterns and regulatory

pathways, bulk RNAseq cannot capture the heterogeneity within the HSC compartment, as it averages signals across diverse subsets of cells. Similarly, bulk ATAC sequencing (ATACseq) has been utilised to profile chromatin accessibility in HSC populations, revealing regulatory elements and transcription factor binding sites that control gene expression.

In one study, bulk sequencing techniques were applied to clonally amplified single-cell transplanted HSCs to investigate their transcriptional and epigenetic programming (Meng et al. 2023). While this method provided critical insights into HSC subpopulation-specific signature, it is still limited by the inherent lower resolution yielded by bulk approaches. Furthermore, the clonal expansion of single-cell transplanted HSCs is time and cost-intensive, and they do not allow for the direct integration of transcriptomes and epigenomes from the same cells, which is crucial for linking gene expression with chromatin accessibility. These constraints underscore the need for advanced single-cell and multi-modal sequencing technologies to unravel the complexities of HSC biology.

1.6.5 Single-cell sequencing: from plate-based to droplet technologies

Single-cell RNAseq has transformed haematology research by enabling the profiling of individual cells, allowing researchers to uncover cellular heterogeneity and map differentiation trajectories. Early single-cell approaches such as the plate-based Smart-seq

and Smart-seq2 platforms, provided full-length transcript coverage and high sensitivity, making it ideal for detailed transcriptomic studies (Picelli et al. 2014). In these plate-based methods, individual cells are sorted into separate PCR wells containing unique barcoding primers, where reverse transcription is performed to incorporate single-cell barcodes into the cDNA. While this approach offers high resolution, it is labour-intensive and have limited scalability, restricting their use in large-scale analyses. Similar workflow is also used in plate-based single-cell ATACseq, where chromatin accessibility is similarly profiled on a per-cell basis (Chen et al. 2018; Yan et al. 2020; Xu et al. 2021).

The advent of droplet-based platforms, such as the 10X Genomics Chromium system, introduced the use of barcoding beads to enable high-throughput single-cell sequencing. In these methods, cells are encapsulated in nanolitre-scale droplets along with gel beads coated with unique oligonucleotide barcodes. Each bead contains a large number of primers that include a cell-specific barcode, unique molecular identifiers (UMIs), and sequences for RNA capturing (Zilionis et al. 2016; Salomon et al. 2019). When a cell is lysed within its droplet, its RNA binds to the barcoding primers from the bead, and the resulting cDNA is labelled with the corresponding cell barcode and UMI (Figure 4B). This allows for the identification of RNA molecules from individual cells and ensures that quantification of gene expression is accurate, even when sequencing depth varies. After droplet barcoding, contents are pooled for sequencing library preparation. The droplet

barcoding system allows transcriptomes of thousands of cells to be analysed simultaneously, while maintaining the ability to trace each transcript back to its cell of origin.

1.6.6 Single-cell multi-omic sequencing

To comprehensively study single-cell biology, multi-omic sequencing technologies have been developed to integrate multiple layers of molecular information. The RNA+ATACseq Multiome platform allows the simultaneous profiling of transcriptome and chromatin accessibility at the single-cell level (Hao et al. 2021). This is achieved using barcoding beads that contain primers designed to capture both RNA and Tn5 transposed DNA fragments, allowing single-cell barcoding for both transcriptional and epigenetic data (Figure 4B). This approach provides insights into how regulatory elements influence gene expression.

Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITEseq) integrates RNAseq with surface protein quantification (Stoeckius et al. 2017). It uses antibody-derived tags (ADTs) to bind specific protein markers on the cell surface, enabling immunophenotyping akin to flow cytometry. These antibody-conjugated oligonucleotides contain a unique barcode for identifying the corresponding antigen, followed by either a poly-A tails (similar to 3' mRNA) or designated capture sequences, allowing them to be

recognised and barcoded by the same single-cell barcoding beads used in droplet-based platforms (Figure 4B). This capability allows for identification of protein markers alongside transcriptomic signatures, providing a direct link between cellular phenotypes and molecular states. In haematology studies, CITEseq is proven to be a highly effective tool for distinguishing functional cell types within heterogeneous single-cell clusters, facilitating the identification of phenotypic markers (Leblay et al. 2020; Man et al. 2021).

DOGMAseq is an advanced multi-modal sequencing technology designed to simultaneously profile transcriptome, chromatin accessibility and surface protein expression from individual cells (Mimitou et al. 2021). This integrative approach combines the ATAC+RNAseq Multiome and CITEseq ADTs into a unified workflow. DOGMAseq utilises a modified permeabilisation buffer which preserves surface-bound ADTs for protein quantification while still enabling Tn5 transposition (Figure 4A).

The main advantage of DOGMAseq is its ability to capture multiple molecular modalities from the same cell, allowing for a more comprehensive view of gene regulation, epigenetic priming and immunophenotype. This is particularly valuable in heterogeneous or developmentally dynamic populations such as HSCs. However, DOGMAseq presents some limitations, including lower cell recovery rate and higher cost (Xu et al. 2022). Additionally, the sensitivity for low-abundance transcripts and accessible regions may be lower due to compromises inherent in the droplet-biased platform. At the time of this

project, DOGMAseq was a relatively new technique that had not yet been established in the institute, requiring optimisation of reagents, instrument settings, and analytical pipelines prior to implementation.

By combining these technologies, we aim to gain a multi-dimensional view of single-cell biology. These integrated approaches provide opportunities to uncover the complex mechanisms driving functional heterogeneity and differentiation in HSCs, offering tools to study both normal haematopoiesis and its dysregulation in disease contexts.

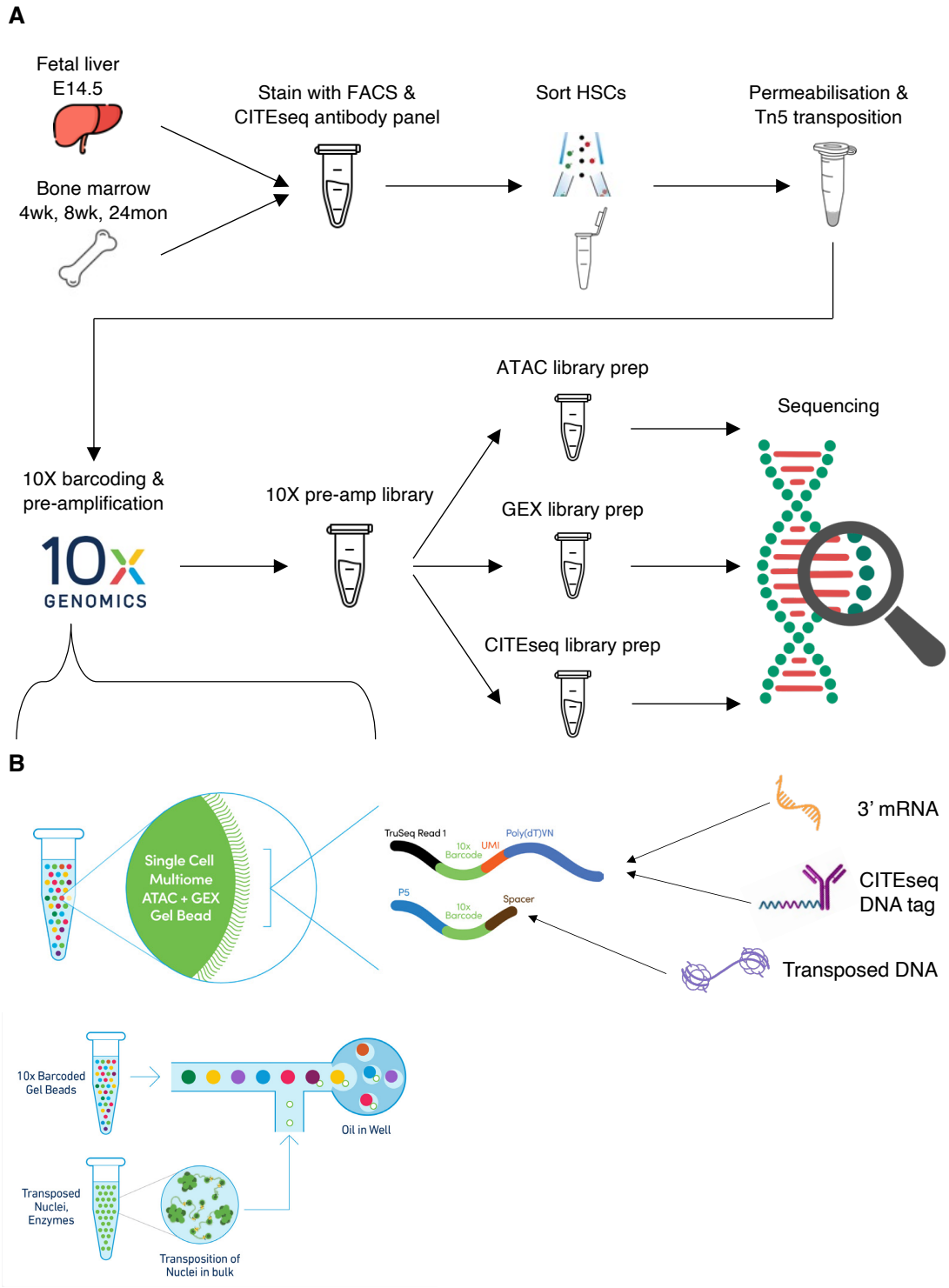


Figure 4 DOGMAseq workflow for single-cell multi-omic sequencing

A, In this thesis, bone marrow and foetal liver cells were harvested and stained using a FACS and CITEseq antibody panel. HSCs were FACS-sorted, permeabilised for Tn5 transposition, and loaded onto the 10X Multiome Gel Beads-in-emulsion (GEMs) barcoding reaction. The barcoded library was then split into three portions for separate library preparation of ATAC chromatin accessibility, RNA gene expression and CITEseq data. The resulting sequencing data were integrated into a tri-modality single-cell dataset using the shared single-cell barcodes. **B,** The 10X single-cell barcoding beads contain primers with poly-T capture

sequences for 3' mRNA and CITEseq antibody-derived tags; and primers to capture Tn5-transposed DNA fragments. Each successful GEM encapsulates one cell and one barcoding bead, which applies single-cell-specific barcodes to all three libraries

1.7 Aims, research questions and potential impacts

This thesis aims to establish a robust workflow for DOGMAseq in mouse HSC samples, enabling high-resolution multi-omic profiling of HSCs across key developmental and ageing stages. By integrating transcriptomic, epigenetic and proteomic data at the single-cell level, this study seeks to characterise the heterogeneity and lineage potentials of HSCs, and understand how these properties evolve with age.

The first objective is to develop and optimise a DOGMAseq workflow for mouse HSCs. This involves identifying candidate subpopulation surface markers for the CITEseq antibody panel, refining sample preparation and library construction steps of DOGMAseq (Chapter 3).

An important goal of this thesis is to identify and functionally validate distinct HSC subpopulations within the 8-week HSC dataset, as this age represent a snapshot of the standard adult HSC compartment. Using the multi-omic data, the study will classify HSC subpopulations based on molecular signatures (Chapter 4), followed by the development of FACS-based enrichment strategy to isolate these populations. This strategy will be validated functionally through series of molecular and *in vivo* analyses, ensuring that the identified subpopulations correspond to distinct haematopoietic potentials (Chapter 5-6).

This study will apply the established pipeline to further profile HSCs across key developmental and ageing stages, including foetal liver (FL) HSCs and young, adult, aged bone marrow (BM) HSCs (Chapter 7-8). By comparing these datasets, we aim to uncover how HSC transcriptional programmes, chromatin landscapes and protein expression profiles evolve over time. Of particular interest are the mechanisms governing ageing-related changes, and the transitions from FL to BM HSCs, which mark the emergence of lineage bias.

Another aim is to study how HSCs age at the subpopulation levels. This will involve identifying the same HSC subpopulations in the aged dataset and perform differential gene expression and chromatin accessibility analyses. By comparing lineage-biased and multilineage HSCs in adult and aged mice, this study will distinguish systemic ageing signatures and subpopulation-specific alterations. This study will assess whether ageing affects all HSC subpopulations uniformly or whether certain subpopulations are more resistant or susceptible to ageing-related factors (Chapter 7).

Beyond advancing fundamental stem cell biology, this research has significant implications for future studies and clinical applications. Establishing a reference multi-omic dataset for mouse HSCs across developmental and ageing stages will serve a valuable resource for the field. The development of FACS-based enrichment strategy will offer an alternative to transgenic reporter mouse lines, making it easier to isolate and study functionally distinct

HSC subsets. This approach may also aid in selecting the most potent HSC populations for transplantation therapies, improving clinical outcomes. Furthermore, by identifying key molecular regulators of HSC ageing, this study has the potential to reveal therapeutic targets for rejuvenating aged HSCs, providing new strategy for mitigating age-related haematopoietic decline.

By integrating high-resolution single-cell multi-omics with functional validation approaches, this thesis will provide a comprehensive characterisation of mouse HSC heterogeneity and ageing. This work will contribute to both basic and translational research, offering new perspectives on HSC regulation and potential therapeutic avenues for stem cell rejuvenation.

Chapter 2: Methods

2.1 Animal

All animal procedures, including breeding, maintenance, blood sampling and culling, were conducted in compliance with the UK Home Office regulations. Experimental mice were all from the C57BL/6 strain. For HSC clonal expansion, 8-week CD45.1 *Vwf*-tdTomato *Gata1*-eGFP mice were used as donors. For candidate FACS antibody validation, 8-week *Vwf*-eGFP (Carrelha et al. 2018) mice were used. For transplantations, CD45.1 *Gata1*-eGFP mice aged 8 to 12 weeks were used as HSC donors (Drissen et al. 2016). Wild-type (WT) CD45.2 recipient mice, aged 8 weeks, were obtained from Charles River UK (Model 632). These WT mice were also used for generating timed mating foetus and harvesting HSCs for single-cell sequencing and Fluidigm single-cell qPCR.

2.2 Bone marrow and foetal liver cell harvest

For bone marrow (BM), mice were firstly culled using schedule 1 method and their Sternum, Femur, Tibia and Pelvis were harvested. These bones were crushed with 10 ml FACS media (Supplementary Table 3) and filtered through 50 μM sterile filter (CellTrics).

For foetal liver (FL), timed mating was conducted, and foetuses were harvested on day E14.5. The FL cells were dissociated using 40 μm cell strainers (Corning, 431750) and

resuspended in 4 ml FACS media.

Sample white blood cell (WBC) concentrations were measured by a Horiba Pentra Haematology Analyser ES60.

2.3 Haematopoietic cell c-Kit enrichment

Whole BM and FL cells were enriched by c-Kit positive selection using MACS magnetic separation. The filtered BM/FL cells were pelleted by centrifugation at 500G for 5 min and resuspended at a concentration of 1 million (M) WBCs/μl of FACS media. The cells were then mixed with CD117 magnetic beads (Miltenyi, 130-097-146) at a ratio of 2.5 μl/100M WBCs and incubated on ice for 20 min. Subsequently, the cells were washed, resuspended in 4 ml of FACS media, and passed through a MACS LS column (Miltenyi, 130-042-401). The column was washed three times with 3 ml of FACS media, removed from the magnet, and the c-Kit enriched cells were collected by flushing 8 ml of FACS media through the column into a new tube.

2.4 c-Kit enriched cell preparation for FACS

Cells were resuspended in FACS media and blocked with 1X FcR blocking reagent (Miltenyi, 130-092-575) for 10 min, then stained with a 2X antibody cocktail (containing both FACS and CITEseq antibodies) (Supplementary Table 1) for 25 min. The stained cells were then washed, resuspended in FACS media with 0.525 μg/μl 7-AAD viability dye

(and 0.2 U/μl RNAse inhibitor for DOGMAseq), and filtered through a 50 μm sterile filter.

2.5 Peripheral blood cell preparation for FACS

Mouse peripheral blood (PB) was collected from tail vein into Heparin-coated tubes. To collect platelet and erythrocyte, blood was centrifuged 10 min at 100G, all the supernatant and 1 μl erythrocyte were transferred into a new tube containing 25 μl PB media (Supplementary Table 3). To collect WBCs, the remaining blood were gently mixed 100 μl PB media and 200 μl 2% Dextran (Sigma, 31392) and incubated at 37 °C for 40 min to pellet red blood cells (RBC). The supernatant was then transferred to a new tube, washed with 1 ml PB media and centrifuged at 1000G. Next, the pellet was resuspended with 200 μl 0.8% ammonium chloride (Stem cell technologies, 07850) for 4 min to lyse the remaining RBCs. Finally, the remaining WBCs were washed again with 1 ml PB media and resuspended in 25 μl PB media containing 1X FcR blocking reagents for 10 min before staining. Platelet, erythrocyte and WBC samples were stained with respective 2X antibody cocktail (Supplementary Table 1,2) for 25 min, washed by 1 ml PB media and resuspend in PB media for FACS analysis.

2.6 Single and bulk-HSC transplantation

Single or groups of 50 HSCs were sorted into 100 μl of injection media (Supplementary Table 3) in a U-bottom 96-well tissue culture plate. After sorting, an additional 100 μl of

injection media containing 200,000 fresh CD45.2 competitor BM cells were added to each well, bringing the total injection volume to 200 μl. The cells were then loaded into a 30G needle (Omnican50, B.Braun 9151125S) and injected intravenously into lethally irradiated (10.5 Gray) recipients via the tail vein.

The recipients were then given Baytril water for 4 weeks to prevent infection. PB was collected from the tail vein at week 6, 10, 16 for lineage reconstitution analysis (Platelet, Erythrocyte, Myeloid cell, T4 cell, T8 cell and B cell) (Supplementary Table 1). PB lineage reconstitution was defined as > 0.1% donor chimerism within a given lineage. BM was harvested as the endpoint analysis for reconstituted progenitors (MPP2, MPP3, MPP4) and HSCs (Supplementary Table 1).

A long-term engrafted lineage-restricted (P, PE or PEM-restriction) HSC was defined as an HSC that reconstitutes only certain PB lineages, not all, and maintains over 0.1% HSC BM chimerism at week 16.

2.7 Bulk RNA and ATACseq (Yiran Meng)

Single-cell transplanted HSCs were allowed to reconstitute in the recipient for 16 weeks and their lineage-potentials were determined by the week 16 PB readout (Figure 5A). Three donor contribution phenotypes were defined based on peripheral blood (PB) lineage reconstitution thresholds:

- Platelet-restricted (PLT): > 0.1% donor-derived platelets in PB, with no detectable reconstitution in other lineages.
- Platelet-erythrocyte-myeloid-restricted (PEM): > 0.1% donor-derived platelets, erythrocytes, and myeloid cells, but no lymphoid (B or T cell) reconstitution.
- Multilineage (MULTI): > 0.1% donor-derived contribution across all lineages, including platelets, erythrocytes, myeloid cells, B cells, CD4 and CD8 T cells.

For bulk RNAseq, 100 donor-derived HSCs from each recipient were sorted into 4 μl of RNAseq lysis buffer (Supplementary Table 3). Reverse transcription was performed using the SMARTScribe Reverse Transcriptase Kit (Takara, 639537) followed by DNA amplification PCR with SeqAmp Polymerase (Takara, 638504) according to the manufacturer's protocol. Samples were then purified using AMPure XP beads (Beckman Coulter, A63881), and the sequencing libraries were constructed using the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1024).

For bulk ATACseq, 500 donor-derived HSCs from each recipient were sorted into 25 μl of ATAC lysis buffer (Supplementary Table 3). To perform tagmentation, 1.25 μl of TDE1 transposase (Illumina, 20034197) was added to each sample, which was then incubated at 37 °C for 30 min with agitation at 300 rpm. Samples were immediately purified using the MiniElute Kit (Qiagen, 28004). Library construction PCR was performed with 10 μl of transposed DNA, 10 μl of water, 2.5 μl each of Ad1noMX and Ad2_x sample index

primers (Supplementary Table 4), and 25 μl of NEBNext High-Fidelity 2X PCR mix (NEB, M0541S) using the following settings (Corces et al. 2017):

(1) 72 °C 5 min

(2) 98 °C 30 sec

(3) 98 °C 10 sec

(4) 63 °C 30 sec

(5) 72 °C 1 min

Repeat (3) – (5), 17X

(6) 4 °C Hold

The PCR product was purified by AMPure XP beads at 1.8X.

2.8 CITEseq antibody preparation

For targets without off-the-shelf CITEseq antibodies, a biotin-streptavidin conjugation was performed using the Lightning-Link kit (Abcam, ab102921). Antibody-derived tag (ADT) oligos (Supplementary Table 4) were synthesised with a biotinylated 5' tail (IDT) and reconstituted to 100 μM in TE buffer. To link the unconjugated antibody with streptavidin, 30 μl of modifier solution was added to 300 μl of 0.5 μg/μl antibody. The antibody was then mixed with 100 μg of lyophilised streptavidin and incubated overnight in the dark at room temperature (RT). The reaction was quenched by adding 30 μl of

quenching solution followed by 40 μl of 0.1% Tween-20. To label each antibody molecule with 8 ADT oligos (as 4 docking sites per streptavidin molecule and 2 streptavidin per antibody), 150 μg of antibody was mixed with 8 nM biotinylated oligo and incubated overnight in the dark at RT.

Conjugated antibodies were purified using Amicon Ultra 50 kDa cut-off columns (Millipore, UFC505024) according to the manufacturer's protocol. The purified antibody was recovered with 150 μl PBS and stored at 4 °C. To verify the conjugation, 1 μg tagged antibodies were electrophoresed alongside with 1 μl stock oligo, 1 μg unconjugated antibody and 1 μg 4-week 4 °C stored + 37 °C 1-hour thermal incubated tagged antibody on a 4% agarose gel with SYBR Safe dye (Invitrogen, S33102) at 100V for 60 min.

2.9 DOGMAseq single-cell multi-omic sequencing

The workflow was adopted from (Mimitou et al. 2021) with modifications improving recovery rate for HSC samples. Cells were sorted into 200 μl of FACS media containing 0.2 U/μl RNase inhibitor (PN-EO0382, ThermoFisher) and centrifuged for 7 min at 500G. The supernatant was carefully removed until 5 μl remained, then 45 μl of DIG permeabilisation buffer (Supplementary Table 3) was added to the pellet, mixed by pipetting three times, and incubated on ice for 5 min. The permeabilised cells (nuclei) were washed with 100 μl of DIG wash buffer (Supplementary Table 3) and centrifuged for 7

min at 700G. The supernatant was carefully aspirated until 2 µl remained, and 3 µl of 1X nuclei buffer (10x Genomics, 1000285) was added to resuspend the pellet. The samples were then processed through the 10X Multiome pipeline (CG000338) with modifications to generate the CITEseq library: At step 4.1 pre-amplification, 1 µl of a 0.2 µM pre-amp primer (Supplementary Table 4) was added. To construct the CITEseq library, 30 µl of the pre-amplified library was mixed with 50 µl of 2X KAPA HiFi Mix (Roche, KK2601), 2.5 µl each of 10 µM SI-PCR and RPxx primers (Supplementary Table 4), and topped up to 100 µl with water for the sample index PCR:

(1) 95 °C 3 min

(2) 95 °C 20 sec

(3) 60 °C 30 sec

(4) 72 °C 20 sec

(5) 72 °C 5 min

Repeat (3) – (5), 19X

(6) 4 °C Hold

The PCR product was purified using SPRI beads (Beckman Coulter, B23318) at a ratio of 1.6X.

The library qualities were analysed by TapeStation D5000 screening tapes (Agilent, 5067-

5588) and their concentrations were determined using the NEBNext library quant kit (NEB, E7630) with correction to their average fragment size between 100 – 1000 bp. For sequencing, ADT libraries were mixed with gene expression libraries at a 1:9 ratio based on molecular concentration. The sequencing parameters have followed the 10X Multiome guideline: 50-8-24-49, 25,000 paired reads per cell for ATAC libraries and 28-10-10-90, 20,000 paired reads per cell for RNA+CITEseq libraries.

2.10 Flow cytometry

Flow cytometric analyses were performed using BD Fortessa and Fortessa X20 instruments. Single-colour positive and negative compensation beads (BD, 51-90-9001189 & 51-90-9001291) were stained with 1:400 antibodies for 5 min and used for creating the FACS panels. Fresh fluorescence minus one (FMO) controls were included in each individual experiment to guide gating and compensation. HSC sorting was carried out using BD Aria III, Aria Fusion, and S6 sorters. Bulk purity mode was used for sorting samples destined for bulk sequencing, DOGMAseq and bulk transplantation, while single-cell purity mode with indexing was employed for single-cell transplantation and single-cell qPCR experiments.

2.11 Bioinformatic analysis

2.11.1 Bulk RNAseq analysis (Yiran Meng)

FASTQ files were filtered by FastQC (Babraham Bioinformatics), followed by adaptor removal using Trimmomatic (Bolger, Lohse & Usadel 2014). Reads were aligned to the mm10 genome using STAR (Dobin et al. 2013) and unique reads were quantified using SubRead featureCounts (Liao, Smyth & Shi 2013). The output data were quality-checked again by MultiQC (Ewels et al. 2016). DESeq2 (Love, Huber & Anders 2014) was used to perform differential analysis ($p < 0.05$, $\log2FC > 0.5$) between paired HSC subpopulations: Platelet-restricted (P), Platelet-Erythrocytes-Myeloid-restricted (PEM) and Multilineage (MULTI). Differentially expressed genes were filtered for candidate CITEseq markers by the following criteria: TPM > 5 in at least one clone, maximum TPM variability < 200 between clones of the same lineage reconstitution pattern, localisation to the cell membrane (GO annotation), availability of a FACS antibody.

2.11.2 Bulk ATACseq analysis (Yiran Meng)

FASTQ files were filtered by FastQC, followed by adaptor removal using Trimmomatic. Reads were aligned to the mm10 genome using Bowtie2 (Love et al. 2014) and the output SAM files were converted to BAM files using SAMtools (Li et al. 2009), then deduplicated by Picard tool (Broad Institute). Read shift correction and mitochondrial gene removal was

performed by ATACseqQC (Ou et al. 2018). Peak calling was conducted using MACS2 (Gaspar) to generate a common list of open chromatin regions. This peak list was then used to count the number of unique reads in each sample using featureCounts (Liao, Smyth & Shi 2014). The differential chromatin features ($p < 0.05$, $log2FC > 0.5$) were identified using DESeq2.

2.11.3 Multiome library demultiplexing

Raw FASTQ data was demultiplexed using Cellranger-arc (10X) with the mm10 reference genome to generate RNA and ATACseq feature-by-barcode count matrices; the pipeline also performs joint-modality cell calling that identifies single-cell associated barcodes. The CITEseq data were demultiplexed using CITE-seq-Count (Grieshaber-Bouyer et al. 2021). Count data from different modalities were combined using Seurat V5 (Hao et al. 2024).

2.11.4 Quality control and normalisation

For each cell, the nucleosome signal, TSS enrichment score, mitochondrial RNA gene percentage, and ATACseq blacklist region (Amemiya, Kundaje & Boyle 2019) fragment percentage were calculated. Cells were filtered based on the following criteria: ATAC peak region fragments > 3,000, percentage of ATAC reads in peaks > 15%, ATAC blacklist reads < 5%, nucleosome signal < 4, TSS enrichment score > 2, and mitochondrial RNA gene percentage < 15%.

RNA expression data was log-normalised and scaled using Seurat. Antibody-derived tag (ADT) counts were normalised using centred log-ratio transformation (Lacar 2022). ATAC peak counts were normalised using term frequency-inverse document frequency (TF-IDF), followed by singular value decomposition (SVD).

Cell cycle phases were determined by the Seurat CellCycleScoring function using signature genes identified in (Tirosh et al. 2016).

2.11.5 Unsupervised clustering and dimensional reduction of multi-omic data

Dimensional reductions were performed on the normalised RNAseq and ATACseq data. For the RNAseq data, JackStraw scores were calculated for principal component analysis (PCA) results to select the best dimensions. For the ATACseq data, dimensions were calculated using latent semantic indexing (LSI), and the correlation between the total counts and LSI-reduced dimensions was computed.

To perform RNA+ATACseq joint analysis, the weighted nearest neighbour (WNN) (Hao et al. 2021) was computed using the appropriate dimensions from both modalities; unsupervised clusters and a uniform manifold approximation and projection (UMAP) were then generated.

2.11.6 *In silico* lineage potential scoring

Single HSCs were scored for platelet (PLT) and multilineage (MULTI) potential using the

bulk sequencing signatures. The RNA scores were calculated using AUCell (Aibar et al. 2017), which determines the enrichment level of a given gene set. ATAC scores were calculated using ChromVAR (Schep et al. 2017) that counts the number of fragments within a given chromatin peak list. Per-cell lineage potential scores were calculated by dividing the PLT score by MULTI score in each modality, a higher lineage potential score indicates more platelet-bias and vice versa.

2.11.7 *In silico* marker selection and gating simulation

Based on the RNA and ATACseq lineage potential scores, the WNN cluster with the highest average score was labelled as platelet-biased and the lowest-scored cluster was labelled as multilineage. Recursive feature elimination was used to rank the importance of CITEseq markers for separating the two extreme clusters. The top five ranked CITEseq markers were further tested *in vivo*.

2.11.8 Transfer of cluster labels

To understand ageing within specific HSC subpopulations, cluster labels from the 8-week HSC dataset was transferred the 24-month HSCs using two approaches.

First, an anchor-based reference mapping method in Seurat predicted the query dataset cell labels by incorporating both RNA and ATACseq information (Hao et al. 2021) . Briefly, after generating the WNN UMAP, a supervised PCA (sPCA) was performed on the 8-

week data to identify the RNA features best-aligning with the multi-omic WNN dimensional structures. Neighbouring cells were identified using sPCA dimension 1:50. A set of anchor cells was identified between 8-week and 24-month datasets using the Seurat FindTransferAnchor function with the reference.reduction = 'spca', reference.neighbors = 'spca.annoy,neighbors'. The transferred labels for 24-month HSCs were predicted using the anchor cells and the cells were also projected onto the 8-week WNN UMAP.

Alternatively, a machine-learning tool, Cellcano (Ma, Lu & Wu 2023) was applied with customised modifications. This tool originally takes labelled single-cell ATACseq data as the training set; it converts the ATACseq fragment counts to pseudo gene expression matrix using ArchR GeneScore feature (Granja et al. 2021) and selects the best 3,000 features during the training. To predict labels, Cellcano firstly identifies a set of anchor cells from the query data with the highest prediction confidence and uses these cells to semi-supervise a second round of prediction. In this thesis, Cellcano were modified to take both RNA and ATACseq data: the ArchR generated ATACseq gene score was merged with the normalised RNAseq data to form a multi-omic count matrix. This form of matrix was used to train the Cellcano model with the 8-week HSC data and predicate the labels for 24-month HSCs.

2.11.9 Single-cell data differential analysis

MAST (Finak et al. 2015) was used to identify the differentially expressed genes and Wilcox test was used to compare accessible chromatin sites. Differentially expressed genes or chromatin sites with log2FC > 0.2, adjusted $p < 0.05$ and minimal detection > 10% in a cell population were selected. Pathway enrichment analyses were performed using the clusterProfiler (Xu et al. 2024) gseGO function.

For comparison between datasets, including 8-week versus 24-month and foetal liver HSCs, their RNA expression data were scaled and centred before merging. For merging the ATAC data, a common peak list between the datasets was firstly created. Then, a new Seurat ATAC assay was created using the common peak count and fragment files from both datasets. The peak count was normalised as described above.

The motif accessibility was calculated using the RunChromVAR function of Signac (Stuart et al. 2021). The differentially accessible motifs between the 8-week and 24-month dataset were calculated by FindMarkers function with the mean.fxn parameter set to rowMeans.

2.11.10 Mitochondrial clone genotyping

To investigate potential clonal hierarchy in the single-cell datasets, we performed mitochondrial DNA mutation genotyping using mgatk (Lareau et al. 2021). The ATACseq BAM file generated by the Cellranger-arc pipeline was input into mgatk, which profiles

mitochondrial DNA mutations. The resulting data was subsequently imported into R using the ReadMGATK function from the Signac package. Variable mutations were identified and ranked based on the number of cells in which they were detectable.

2.11.11 Pseudotime trajectory analysis

Slingshot (Street et al. 2018) was used to perform pseudotime analysis with the 4-week HSC WNN UMAP. The 'Remodelling' cluster was specified as the initial cell cluster. The slingshot pseudotime and curves were generated using default parameters without pre-defining the number of lineages.

2.12 Fluidigm single-cell qPCR

Single-cell qPCR was performed using the Fluidigm 192.24 IFC platform (100-6170). First, single HSCs were FACS-sorted into 96-well PCR plates, each well containing 5 μl Fluidigm reverse transcription (RT) buffer (Supplementary Table 3). The plate was then incubated for the RT and targeted amplification by TaqMan PCR probes (Table 2):

(1) 50 °C 15 min

(2) 95 °C 2 min

(3) 95 °C 15 sec

(4) 60 °C 4 min

Repeat (3) – (4), 22X

(5) 4 °C Hold

After the incubation, 20 µl TE buffer was added to each well. Samples were then processed on a 192.24 IFC chip according to the manufacturer's protocol. Upon completion, a Ct-by-cell matrix was generated. The gene expression levels were normalised according to Fluidigm's official guide: an assumed baseline Ct value for negative results was set to Ct = 28; and any Ct > 28 were overwritten to 28. Then, 28 was subtracted from the Ct values and the delta Ct against the average Ct value of the two housekeeping genes was calculated (Table 2). A clustered heatmap was generated using the R package pheatmap.

To integrate data, delta Ct values were scaled and centred for each marker within each Fluidigm chip run before merging. After integrating data from multiple chips, a PCA analysis was performed. Cells were then projected onto a two-dimensional PC1-PC2 plot. To compare the lineage potentials of individual cells, the 95% confidence interval centroid coordinates of FACS-sorted PLT and MULTI cells were first calculated. A linear vector was then created, connecting the PLT (coordinate 0) and MULTI (coordinate 5) centroids. Each cell's perpendicular crossing coordinate along this vector was computed, reflecting its relative proximity to the two extreme reference.

Chapter 3: Single-cell multi-omic sequencing for mouse HSCs

3.1 Introduction and aims

Single-cell analysis is essential for understanding the regulatory mechanisms behind HSC physiology. It reveals the heterogeneity within the HSC compartment and helps identify molecular programmes that drive lineage priming and how these are altered by ageing and disease.

Recent advancements in single-cell multi-omic technologies, such as the Multiome platform, have enabled simultaneous profiling of RNA expression and chromatin accessibility at single-cell level. The epigenetic profiling is particularly valuable in the context of HSCs, as lineage potential is closely linked to epigenetic regulations (Meng et al. 2023). Using previously identified RNA and chromatin signatures, we aimed to identify lineage-biased HSC subpopulation and characterise their molecular features within the single-cell dataset.

This chapter focuses on the development and optimisation of the DOGMAseq workflow for mouse HSC samples. While DOGMAseq was originally validated on peripheral blood mononuclear cells, its application to HSCs required modifications. HSCs are rare cells,

often present in low numbers. In practice, each c-Kit enriched 4-week mouse bone marrow sample contains approximately 6 million white blood cells, of which only about 600 are HSCs. Considering sorting efficiencies and dropout rates, we estimated that around 400 HSCs can be obtained per mouse. Using two cell sorters, 12 mouse samples can be processed in 5 hours, yielding a maximum of 4,800 HSCs as the input for a single-cell sequencing experiment. As a result, the workflow needed to be adapted to accommodate input cell number under 5,000. Additionally, we aimed to reduce cell loss during the digitonin permeabilisation and washing steps, as pellets from a few thousand cells are invisible, making these steps critical for retaining as many cell as possible.

We also customised a CITEseq antibody panel to address the subtle differences between HSC subpopulations. To enhance the signal-to-noise ratio, we fine-selected candidate antibodies based on differentially expressed genes identified in our bulk RNAseq data. Furthermore, antibodies targeting high-abundance features were titrated with unconjugated counterparts to balance the total signal levels, enabling improved detection of low-abundance markers in the CITEseq library.

3.2 Identification of HSC lineage potential signatures using bulk RNAseq data

To identify potential FACS markers for differentiating HSC subpopulations, we focused

on genes that were differentially expressed across single-cell transplantation-expanded clones with varying lineage potentials. We compared the bulk RNAseq data from these HSC clones, as described in (Figure 5A) (Meng et al. 2023). From the 24,107 mapped genes, we applied a series of filters to identify candidate markers with the following characteristics:

1. High absolute expression levels (transcripts per million, TPM > 5) to ensure sufficient protein production.
2. Significant differential expression, defined as a maximum log2 fold change (log2FC > 1) across at least one of the pairwise comparisons between P vs. MULTI, P vs. PEM, and PEM vs. MULTI.
3. Low variability (< 200 TPM) among biological replicates within clones of the same lineage potential, ensuring that the marker reliably represents the phenotype.
4. Cell surface localisation based on Gene Ontology (GO) annotations, allowing detection by FACS or CITEseq using available antibodies.

Applying these criteria, we identified 25 genes with commercially available FACS antibodies (Figure 5B).

3.3 Development of CITEseq antibody panel

We then evaluated the quality of PE-conjugated FACS antibodies targeting the candidates on adult mouse HSCs. Positive signals were detected for 12 antibodies (Table 1), all of

which were sourced from Biolegend. In contrast, other antibodies obtained from Santa Cruz did not successfully stain any BM samples. We further examined their staining levels in relation to the *Vwf*-eGFP reporter and found that most markers displayed staining patterns consistent with their expression levels in the clone RNAseq data. For example, *CD357*, which was more highly expressed in multilineage compared to platelet-restricted HSC clones, exhibited a negative correlation with the *Vwf*-eGFP reporter signal (Figure 6A), confirming its association with multilineage potential.

To ensure balanced representation in different CITEseq markers in the sequencing library, we titrated antibodies with high binding quantities by using their unconjugated counterparts, effectively reducing overall signal strength. The mean fluorescence intensity (MFI) of the PE-conjugates was measured in adult mouse HSCs, with values ranging from 10^5 for highly abundant antigens like CD106, to as low as 10^3 for rare antigens such as CD68. By serial dilution with unconjugated antibodies, we determined the ratio for normalising the MFI of each candidate antibody to a range between 10^3 to 10^4 (Figure 6B). These optimised dilution ratios were subsequently used in the CITEseq antibody panel.

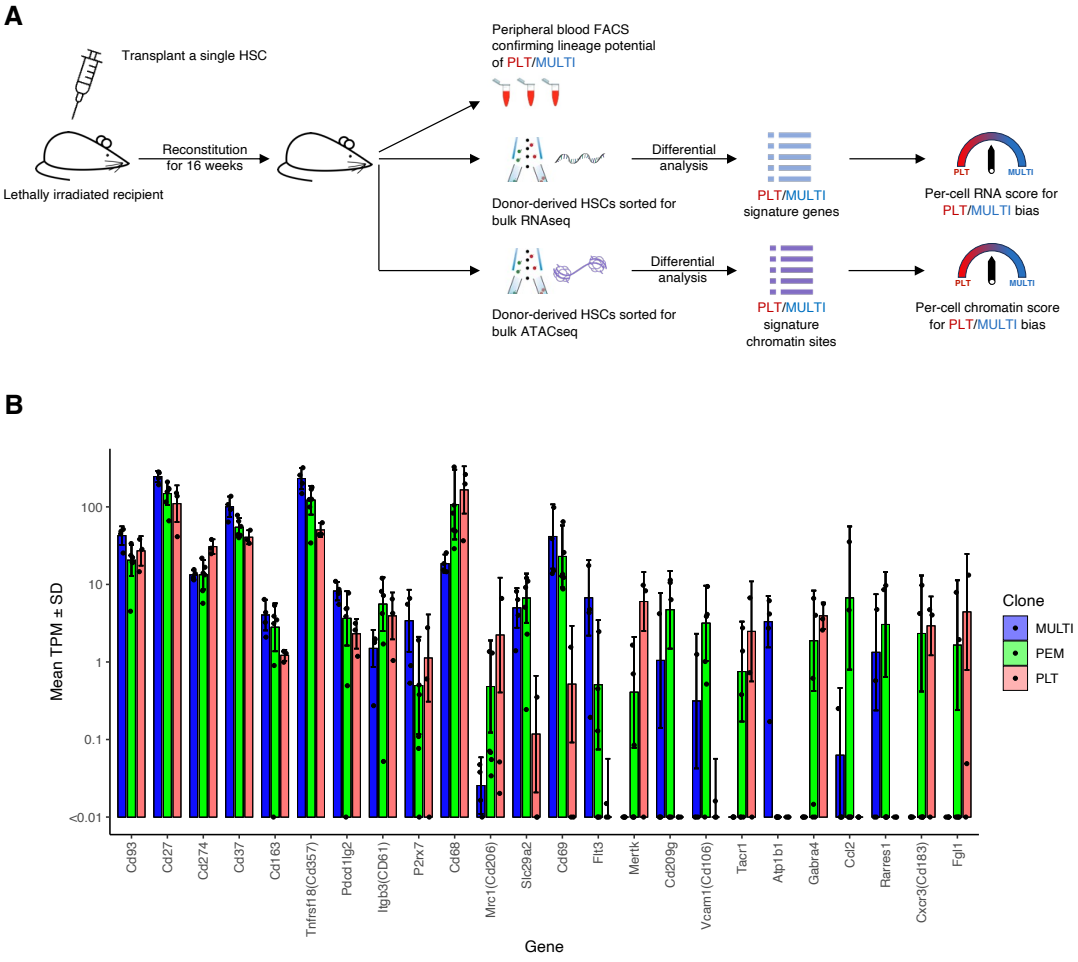


Figure 5 Clonal expansion of single HSCs reveals candidate CITEseq targets

A, Identification of lineage potential signature genes and chromatin sites. Single isolated HSCs were transplanted into lethally irradiated recipient mice. Clone lineage potentials (Chapter 2.6) were assessed based on peripheral blood reconstitution at week 16 post-transplantation. Donor-derived HSCs were collected for bulk RNA and ATAC sequencing. Differential gene expression and chromatin accessibility analyses were conducted among platelet-restricted (P, n = 3), platelet-erythrocyte-myeloid-restricted (PEM, n = 6), and multilineage (MULTI, n = 4) donor-derived HSCs. P and MULTI-HSCs signatures were subsequently utilised for per-cell lineage potential scoring in the DOGMAseq datasets, as shown in (Figure 8). **B,** Absolute expressions of differential CITEseq candidates. Candidate surface markers for isolating HSC subpopulations were identified by analysing differentially expressed genes among MULTI, PEM and P-HSC clones. Genes were filtered based on the following criteria: TPM > 5 in at least one clone, maximum TPM variability < 200 between clones of the same lineage reconstitution pattern, localisation to the cell membrane (GO annotation), availability of a FACS antibody, and ranking by the maximum fold change observed in paired DESeq2 analyses.

3.4 Conjugation of CITEseq antibodies

We performed in-house conjugation for candidate markers that did not have pre-conjugated CITEseq antibodies available from the Biolegend Totalseq A product line. Six anti-mouse antibodies were conjugated to their respective oligo tags using a biotin-streptavidin linkage method. Agarose gel electrophoresis confirmed successful conjugation, as the conjugated antibodies displayed significantly higher molecular weight bands compared to the free oligo tags, which were observed at 66 bp (Figure 6C).

To ensure the stability of the conjugated antibodies, we tested by storing the antibodies at 4 °C for 4 weeks and then incubating them at 37 °C for 1 hour (mimicking the ATAC transposition step). These conjugates showed no signs of dissociations that release 66 bp oligos (Figure 6C).

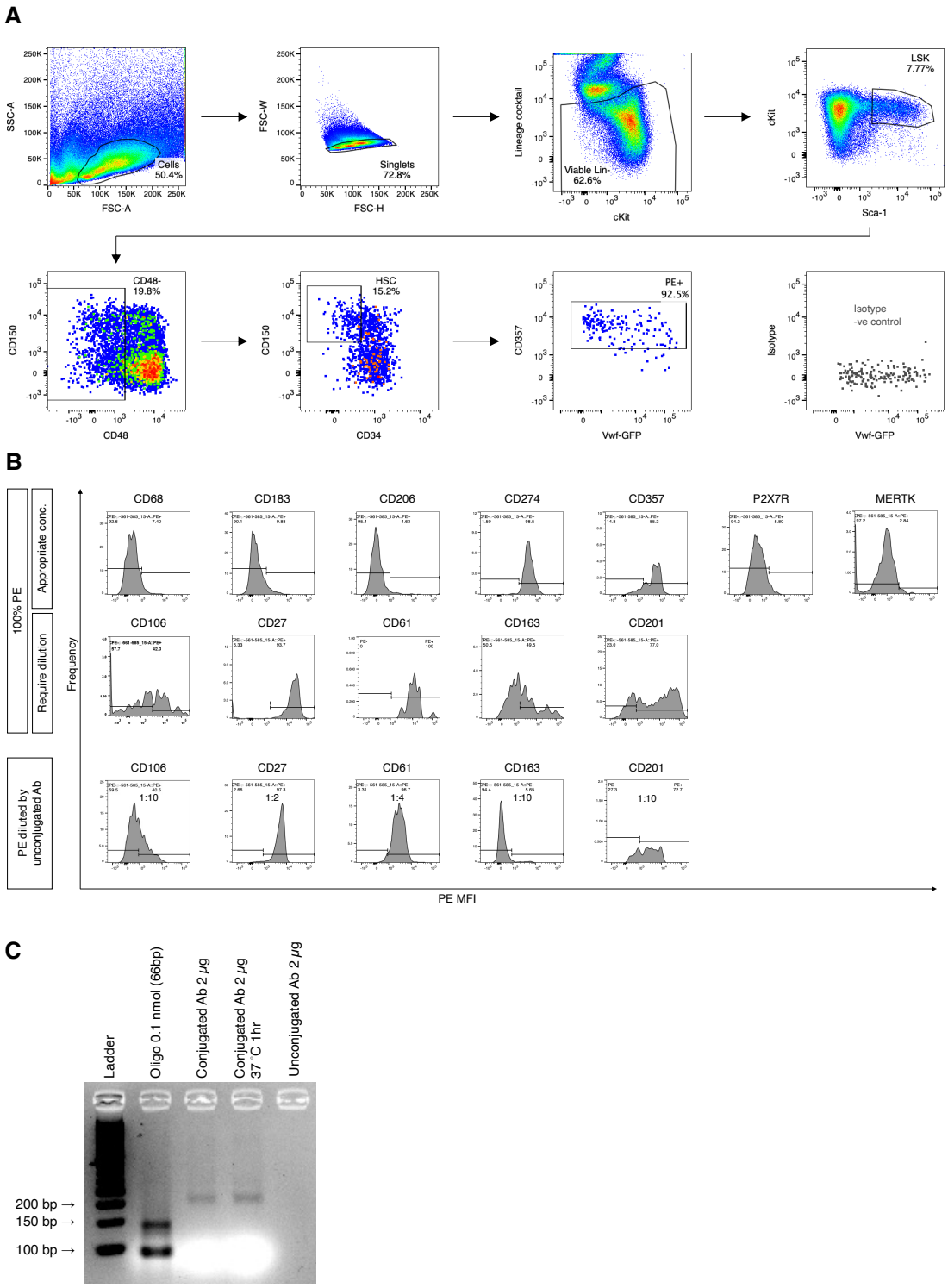


Figure 6 Development of the CITEseq antibody panel

A, FACS gating for long-term HSCs. PE-conjugated antibodies were tested on mouse long-term HSCs (Lineage- Sca-1+ c-Kit+ CD48- CD34- CD150+) carrying *Vwf*-eGFP reporters that indicate lineage bias. Positive staining was determined by comparison to an isotype negative control. This example shows the multilineage CITEseq candidate CD357 negatively correlates with *Vwf* expression level. **B,** Dilution of CITEseq antibodies. Candidate antibodies with higher absolute staining intensities (second row) were diluted

with their unconjugated forms to normalise signal levels across all candidates (third row). These dilution ratios were used in the final CITEseq antibody panel to enhance signal detection. **C,** In-house antibody-oligo conjugates were tested for purity. Biotin-streptavidin conjugations of DNA tags were performed for CITEseq candidates that have no off-the-shelf conjugated antibodies. Successfully conjugated samples exhibited a single, clean band at ~200 bp with no detectable free oligo contamination at 66 bp (and dimers at 132 bp). The conjugates remained stable without dissociation after long-term storage and incubation at 37 °C for 1 hour.

Target	ADT	Titration fold	Manufacturer	Catalogue	Pearson correlation
CD27	CAAGGTATGTCACTG	4	Biolegend	124235	-0.437
CD61	TTCTTTACCCGCCTG	4	Biolegend	104327	0.385
CD68	CTTTCTTTCACGGGA	1	Biolegend	137001	-0.040
CD106	CGTTCCTACCTACCT	10	Biolegend	105701	0.045
CD163	GACGCGCGTTGTCA	10	Biolegend	156702	0.069
CD183	CGTCCTAGGACATA	1	Biolegend	155902	-0.093
CD201	TATGATCTGCCCTTG	10	Biolegend	141509	0.717
CD206	GGTCGACTAGGTCG	1	Biolegend	141705	0.083
CD274	TCGATTCCACCAACT	1	Biolegend	153604	0.085
CD357	GGCACTCTGTAACAT	1	Biolegend	126319	-0.397
P2X7R	TGCTTCATTCATGTG	1	Biolegend	148711	0.003
MERTK	GACGAACGTCCGCG	1	Biolegend	151502	0.013

Table 1 CITEseq antibody panel used in DOGMAseq experiments

CITEseq antibodies were titrated by their unconjugated counterpart by molecular concentration. Person coefficients were calculated between the PE and *Vwf*-eGFP channels on 8-week mouse LT-HSCs as an indication of the candidate antigen's correlation with lineage bias. Underlined catalogue numbers denote unconjugated antibodies that were linked to oligo tags via in-house biotin-streptavidin reactions.

3.5 Optimisation of DOGMAseq workflow for HSCs

We encountered significant challenges in generating high-quality DOGMAseq libraries during our early attempts to apply the workflow to mouse HSCs. In the first trial, HSCs were sorted into standard 1.5 ml Eppendorf tubes, following the original protocol. Unfortunately, this approach resulted in minimal to no detectable fragment in the cDNA

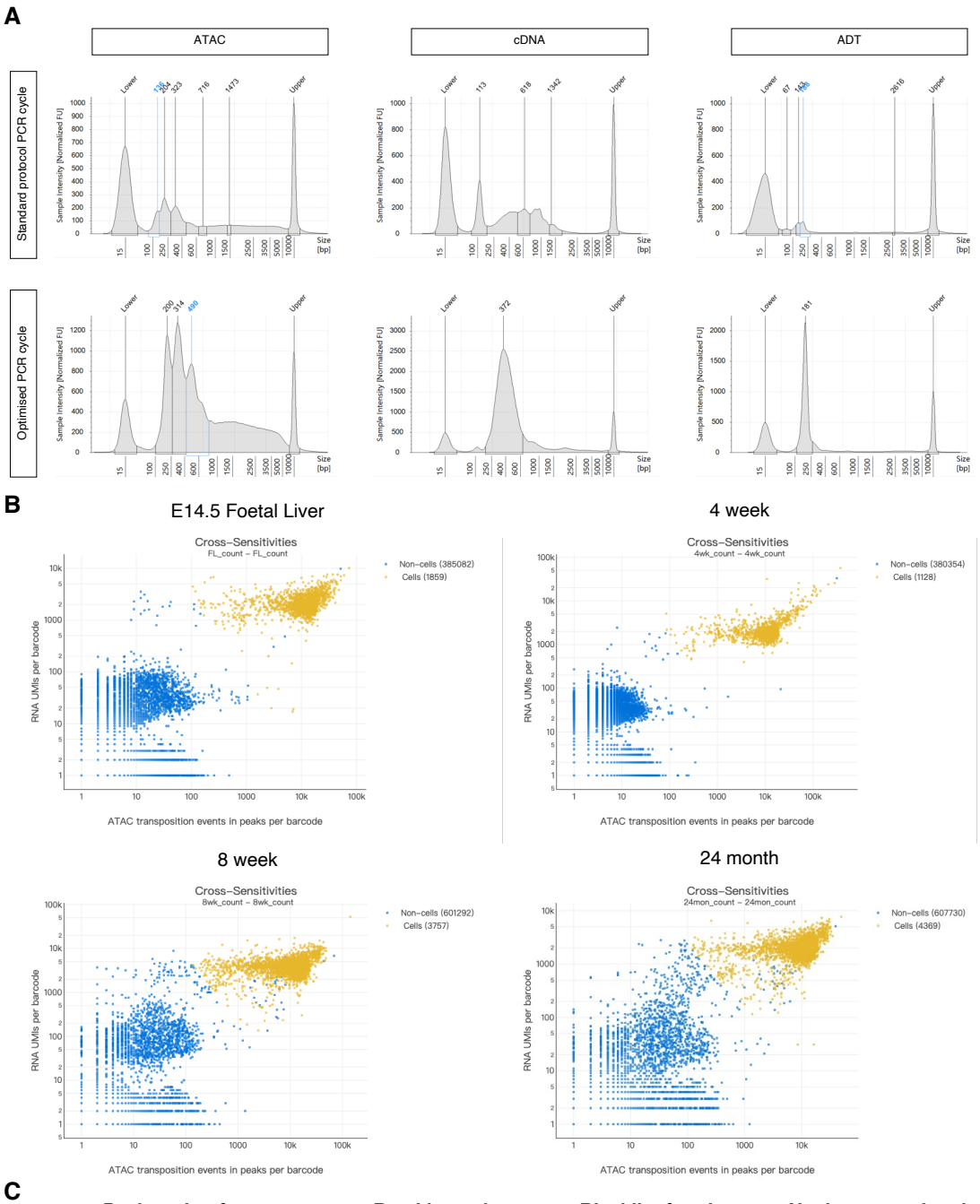
and ATACseq libraries, likely due to cell loss during the permeabilisation and washing steps.

In the second trial, we implemented several adjustments to improve cell recovery and library quality. HSCs were sorted directly into 200 µl PCR tubes to minimise handling and maximise retention. During the permeabilisation and washing steps, we reduced the number of washes from two to one while increasing the washing volume to 100 µl. Additionally, we extended the centrifugation time from 5 to 7 minutes to enhance pelleting of cells. These optimisations allowed us to detect cDNA, ATACseq fragments at the expected sizes during quality control, confirming improved cell retention (Figure 7A).

Despite improved recovery, the second trial still yielded low library concentrations across libraries, accompanied by primer dimer contamination. The reduced library concentration posed challenges for quality control and sequencing, potentially affecting data reliability. Additionally, no expected peaks were detected in the ADT library, likely due to the use of a smaller, titrated antibody panel and the need for additional PCR amplification.

To address these issues, we increased the PCR cycles for both ATACseq library sample index PCR and the cDNA amplification PCR by two cycles, which significantly improved library concentrations. For the ADT library, we increased the PCR cycles from 12 to 18 to compensate the small CITEseq panel. This adjustment successfully resolved the absence of ADT peaks in the second trial, producing a distinct peak at ~180 bp while minimising

primer dimer contamination (Figure 7A). These modifications ensured robust library preparation across RNA, ATAC and ADT libraries, making the DOGMAseq workflow more suitable for rare HSC samples with low input cell numbers. The optimised protocol resulted in a nuclei recovery rate of at least 25%, with cell-associated barcodes clearly separated from background barcodes in RNA UMI × ATACseq fragment count plots (Figure 7B-D), indicating effective RNA and ATACseq capture with minimal contamination. Furthermore, the RNA and ATACseq feature counts were comparable to publicly available PBMC Multiome and CITEseq datasets (Xu et al. 2022; 10X Genomics. 2011), confirming that the optimised workflow achieved high-quality data capture.



D

Sample	Foetal liver	4-week	8-week	24-month
Number of mice	9	12	10	2
HSC Sorted	9000	4000	26000	32000
HSC Recovered	1859	1128	3757	4369
Median RNA UMI	1282	1191	2014	1189
Median ATAC Fragment	9870	9084	10691	9728

Figure 7 Optimised DOGMAseq workflow enhances library quality

A, Adult mouse HSCs were processed using the protocol described in (Mimitou et al. 2021). library preparation using the standard PCR cycle number resulted in low library concentrations with significant primer dimer contamination below 150 bp in all three libraries. Optimisations, including increasing the ATACseq and cDNA amplification PCR cycles by two and extending ADT PCR to 18 cycles, improved library concentrations and minimised primer dimer contamination, making them suitable for next-generation sequencing. **B,** Foetal liver, 4-week, 8-week, and 24-month bone marrow HSC samples, processed using the optimised protocol, showed ideal quality and complexity. Quality control for cell-associated barcodes was performed based on RNA UMI counts, ATACseq peak fragment numbers, and doublet detection. **C,** Quality control metrics used for filtering high-quality cells in the Seurat workflow. Dashed lines indicate the thresholds applied for filtering. **D,** Summary of the final DOGMAseq libraries.

3.6 Discussion

This chapter describes the optimisation and application of the DOGMAseq workflow for profiling mouse HSCs at the single-cell level. A key component of this workflow is the inclusion of a customised CITEseq antibody panel, which enables simultaneous measurement of surface protein expression alongside RNA and ATACseq data, but its scope was limited by the availability of antibodies. Many top differentially expressed candidate genes lacked antibodies developed for flow cytometry, restricting the panel's ability to fully capture the molecular diversity of HSC subpopulations.

In-house conjugation of antibodies using biotin-streptavidin linkage allowed us to expand the panel, but alternative methods such as click chemistry may offer greater stability (Van Buggenum et al. 2016). Click chemistry forms stable covalent bonds, enabling the long-term storage of pre-mixed antibody cocktails. However, at the time of this project, the reagents required for click chemistry had a six-month lead time, which prevented their use.

Additionally, normalising antibody-derived tag (ADT) counts between separate DOGMAseq experiments remain a challenge due to the lack of housekeeping features in the CITEseq panel, making it difficult to compare the absolute abundance of surface markers across samples of different ages. One potential solution is to adopt a housekeeping feature for normalisation, similar to methods explored in flow cytometry. This strategy involves identifying a stable set of surface markers, serving as internal reference to adjust signal intensities across samples (Hahne et al. 2010). Applying similar approach to CITEseq could help correct for batch effects, allowing comparisons of biological variation between experiments. Additionally, including a negative control such as isotype antibodies could help monitoring background noise and improving interpretation of CITEseq data (Mulè, Martins & Tsang 2022).

The optimisation of the DOGMAseq workflow addressed several challenges posed by the rarity of HSCs. Initial trials, which followed the original DOGMAseq protocol designed for peripheral blood mononuclear (PBMC) cells, resulted in significant cell loss and poor library quality. Modifications, including sorting cells into smaller PCR tubes, reducing washing steps and increasing centrifugation times, substantially improved recovery and reduced contamination. While early libraries exhibited low concentrations, increasing library amplification PCR cycles resolved these issues, yielding feature counts comparable to PBMC datasets. Given the libraries contained a satisfactory level of complexity, we

suspected the concentration issue is caused by low PCR efficiency that may stem from the addition of customised ADT pre-amplification primers (although this has not been reported by other studies) and limitations of the local PCR instrument.

Despite these advances, the experimental design still has limitations. For example, the exclusion of cycling HSCs during sorting (via the CD34+ immunophenotype) may impact the comprehensiveness of the dataset. While CD34- cells are enriched for long-term HSCs (LT-HSCs), this marker's expression decreases at the compartment level with age as more HSCs entre quiescence. (Sudo et al. 2000; Wang et al. 2025). It is well established that aged individuals have a greater proportion of immunophenotypic LT-HSCs; however, this does not correspond to increased haematopoietic output or the number of functional HSCs, as aged HSCs exhibit reduced long-term engraftment potential and lineage output (Kovtonyuk et al. 2019). Conversely, younger individuals with fewer LT-HSCs typically demonstrate higher functional haematopoietic activity. Including both CD34+ and CD34- populations in future studies, coupled with CITEseq measurements of CD34 itself, could provide a more complete picture of the HSC compartment. However, CD34 is a lowly expressed marker and may not be reliably detected by CITEseq, presenting technical challenge for this approach.

In conclusion, the optimised DOGMAseq workflow represents a significant advancement in single-cell multi-omic profiling for rare cell populations like HSCs, enabling high-

resolution analysis of subpopulations. This study primarily focuses on capturing the 8-week LT-HSC population as a representative snapshot of the HSC compartment to address questions of lineage potential heterogeneity. However, future datasets could expand on this work by exploring functional differences across quiescence and cycling HSCs and other haematopoietic progenitor cell populations.

Chapter 4: Determining adult HSC lineage potentials *in silico*

4.1 Introduction and aims

This chapter focuses on the bioinformatic analysis of the 8-week mouse HSC single-cell multi-omic dataset. The primary objective was to develop a robust analytical framework for processing and integrating ATAC and RNAseq data. Ultimately, this framework aims to identify distinct HSC subpopulations and systematically quantify their lineage potentials.

Following stringent quality controls, we applied dimensionality reduction techniques to each data modality. For RNAseq data, principal component analysis (PCA) was used to transform the high-dimensional gene expression features into fewer dimensions by identifying linear combinations of genes that capture the largest variance in the dataset. This approach effectively condenses the transcriptomic information while retaining meaningful biological variation.

For ATACseq data dimensional analysis, latent semantic indexing (LSI), a method specifically suited for sparse, binary-like single-cell chromatin accessibility matrices (each unique fragment should be detected only once within an individual cell) was employed. Unlike PCA, which assumes continuous data, LSI is designed to handle the high sparsity

and the influence of sequencing depth of single-cell ATACseq data. By normalising fragment counts and correcting for depth biases, LSI focuses on capturing biologically relevant variability in chromatin accessibility. The first LSI dimension, often dominated by sequencing depth differences, was discarded, and subsequent dimensions were retained for downstream analysis to ensure the emphasis remained on biologically informative features.

Weighted nearest neighbour (WNN) analysis was then applied to integrate RNA and ATACseq data. WNN constructs modality-specific nearest neighbour graphs and combines them by assigning modality weights to each cell based on the quality of the signal contributed by RNA and ATACseq data. This approach balances the complementary strengths of both modalities, capturing the transcriptomic and epigenomic landscapes simultaneously. Using WNN, we performed unsupervised clustering and generated uniform manifold approximation and projection (UMAP) visualisations to identify distinct HSC subpopulations.

To further quantify lineage potential, we utilised the AUCell (Aibar et al. 2017) and ChromVAR tool (Schep et al. 2017) to calculate per-cell enrichment scores for previously identified lineage potential signatures (Figure 5A). AUCell assesses RNAseq data by quantifying the enrichment of gene sets within the highly expressed genes of each cell, providing a measure of lineage potential at the transcriptomic level. ChromVAR analyses ATACseq fragment data by evaluating the variability in chromatin accessibility associated

with regulatory regions linked to lineage potential signatures. Together, these tools enabled a systematic and integrated analysis of transcriptomic and epigenomic data, revealing the molecular characteristics of distinct HSC subpopulations.

4.2 HSC data quality control (QC)

Following sequencing, we utilised Cellranger-arc software for the initial quality control, focusing on the identification of cell-associated barcodes and the removal of doublets. The output included per-cell metadata, which consisted of RNA UMI counts and ATACseq fragment counts for each barcode. After importing the cell-associated barcode data into Seurat, we filtered out cells with extremely low or high RNA UMI and ATACseq fragment counts, as described in the method section (Figure 7C). This step further removed poorly captured cells, doublets, contaminated barcodes and those affected by over-tagmentation.

Additionally, high mitochondrial RNA (mt-RNA) content is often indicative of poor cell viability or pre-apoptotic states in single-cell sequencing (Osorio & Cai 2021). Therefore, we calculated the percentage of mt-RNA reads for each barcode and excluded cells with over 15% mt-RNA content to ensure high-quality gene expression data (Figure 7C).

We further assessed ATACseq quality by examining nucleosome signal (the ratio of mono-nucleosomal to nucleosome-free fragments), transcription start site (TSS) enrichment scores and the percentage of reads falling within called peaks (Figure 7C). Moreover, we

mapped chromatin fragments to the ENCODE blacklist regions, which are known to generate high background noise in next-generation sequencing (Amemiya et al. 2019). Barcodes with more than 5% of reads mapped to these blacklist regions were discarded to minimise noise and improve data reliability (Figure 7C).

4.3 Multi-omic unsupervised clustering of adult mouse HSCs

We implemented weighted nearest neighbour (WNN) methods to perform unsupervised cell clustering and UMAP projection using two modalities.

For RNA, the data was normalised and scaled as described in the method sections and PCA dimensions were computed. The JackStraw scores were calculated to evaluate the statistical significance of each principal component (PC), by comparing observed gene-PC associations to a null distribution from permuted data. PCs with low p-values indicate meaningful structure beyond random noise and were selected for downstream analysis.

For the 8-week adult dataset the significance score dropped significantly from PCA 7 (Supplementary Figure 1), thus, PCA 1:6 were used for dimensional reductions. For ATAC, we performed depth correlation to the LSI dimensions. The first dimension is discarded as it showed strong negative correlation between the depth and dimension components, suggesting it mainly differentiates the sequencing depth between cells (Supplementary Figure 1) (Yan et al. 2020). We then clustered cells using WNN with RNA PCA 1:6 and

ATAC LSI 2:15 at the default resolution, which produced a UMAP with 7 clusters (Figure 8A). Initially, we examined the expression of some known markers (Figure 8B), such as *Vwf*, *Pdgfc* for the platelet lineage, and *Ctla2a* associating with multilineage potential (Pei et al. 2020), which appeared to be locally enriched in the opposite cell populations on the UMAP. Besides, *Ezh2*, a histone methyltransferase associated with proliferating HSCs (Herviou et al. 2016) was highly expressed in a cluster of cells suggesting those are undergoing cell cycle.

4.4 Annotation of HSC subpopulations through per-cell lineage potential scoring

We systematically determined the lineage potentials at the single-cell level by applying gene expression and accessible chromatin signatures from the expanded clone experiment. These signatures, representing distinct lineage-reconstitution phenotypes enabled quantitative inference of lineage priming and heterogeneity across the dataset (Figure 5A) (Meng et al. 2023). For gene expression, we scored each cell using AUCell that calculates the enrichment levels of the P and MULTI signature genes. And for the epigenome, we counted the number of fragments that fall within the signature peaks using ChromVAR. To generate a single lineage potential score in each modality and to normalise the sequencing depth, we divided the P scores by the MULTI scores (Figure 8D).

Unsurprisingly, the cluster that enriched for *Vwf* expression showed the highest score for P signatures in both modalities. We therefore labelled this cluster as P-biased (PLT), the second highest as platelet-erythrocyte-myeloid-biased (PEM) and the lowest-scored two clusters as multilineage (MULTI)-1 and 2. Besides, we annotated the two cycling clusters based on S and G2M scores (Figure 8C); these cells also exhibited higher ATAC fragment counts (Figure 8E). Interestingly, the deviation of lineage potential scores is significantly larger in the epigenomes compared to transcriptomes. The PLT and PEM clusters also showed higher CD201 (EPCR) signal in CITEseq data (Figure 9A), a surface marker associated with superior self-renewal and long-term engraftment potential (Balazs et al. 2006), consistent with their expanded representation in the aged HSC pool and engraftment performance during serial transplantations.

We further performed differential gene expression analysis between the PLT and MULTI (1 and 2) clusters (Figure 8F). The PLT cluster was marked by genes such as *Vwf*, *Pdgfc*, *Plcl1*, and *Ncam2*, which are associated with megakaryocytic priming and self-renewal (Schnoeder et al. 2022; Singh et al. 2025). In contrast, the MULTI clusters showed higher expression of *Nrxn1*, *Arhgap15*, *Aff3*, *Ctla2a*, *Sox4*, and *Rgs1*, genes involved in stemness, cell lymphoid lineage and immune regulation (Han et al. 2005; Costa et al. 2011; Kinzfogl, Hangoc & Broxmeyer 2011; Sun et al. 2013; Tsukumo et al. 2022), supporting their multilineage potential.

We also annotated the differentially accessible chromatin regions to their nearest genes (Figure 8F). In the PLT cluster, we observed increased accessibility around *Peg3*, *Tnrc6a*, and *Ezh1*. These genes are involved in epigenetic and post-transcriptional regulation (Suzawa et al. 2017; Correra et al. 2018; Vo et al. 2018). In contrast, MULTI clusters showed higher accessibility near *Ifnar2*, *Arhgap30* and *Runx3*, with *Ifnar2* and *Runx3* being more directly implicated in immune lineage specification and differentiation (Passarelli et al. 2020; Hu et al. 2021; Fonseca et al. 2022). The remaining chromatin-accessible regions appear to be more associated with general signal integration and cellular plasticity rather than definitive lineage commitment. However, GSEA analysis did not yield statistically significant pathway enrichment in either transcriptomic or chromatin modalities, possibly due to relative shallow sequencing depth intrinsic to single-cell platforms.

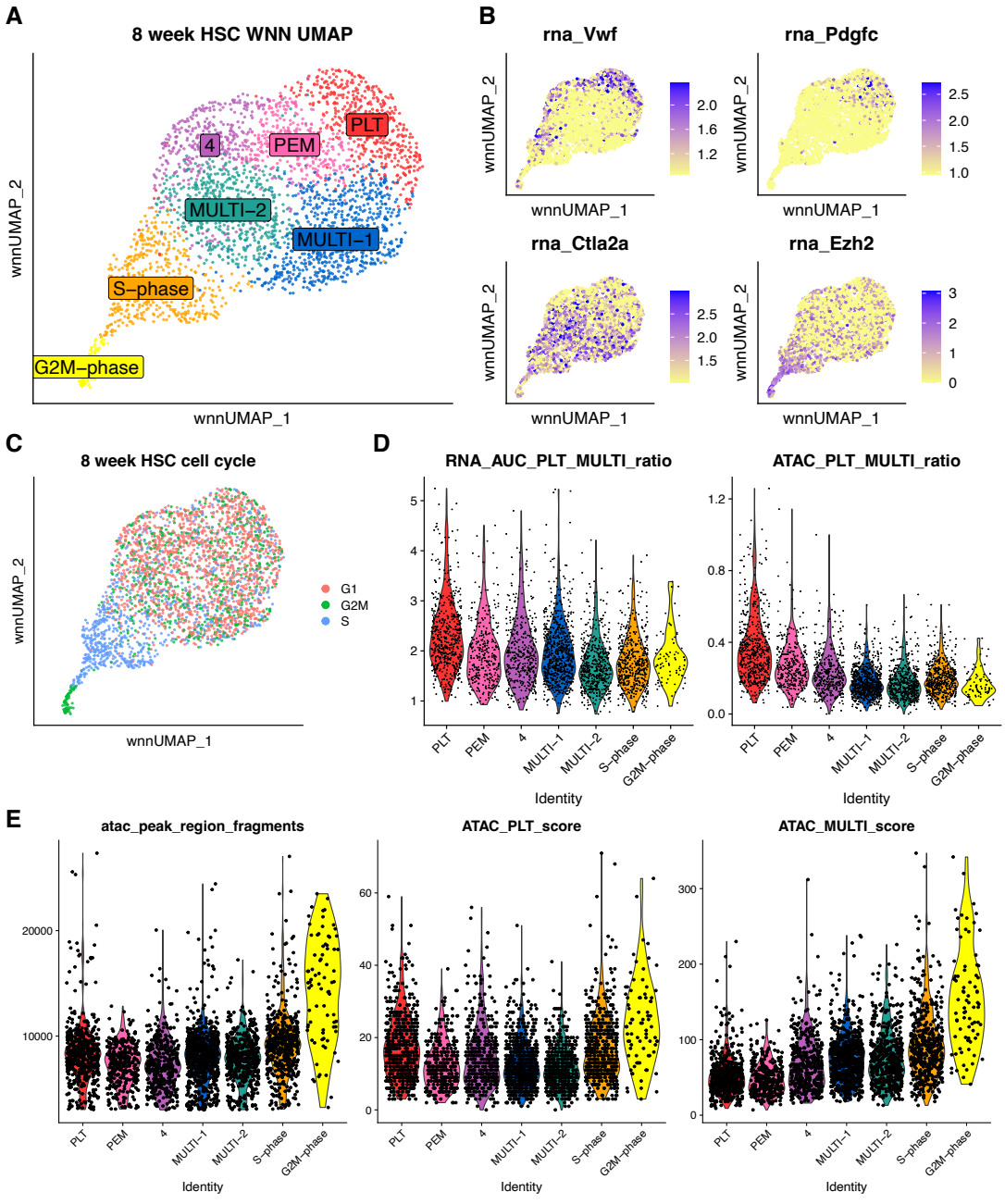

A
8 week HSC WNN UMAP
PLT
PEM
4
MULTI–2
MULTI–1
S–phase
G2M–phase
wnnUMAP_2
wnnUMAP_1
B
rna_Vwf
rna_Pdgfc
rna_Ctla2a
rna_Ezh2
C
8 week HSC cell cycle
G1
G2M
S
D
RNA_AUC_PLT_MULTI_ratio
ATAC_PLT_MULTI_ratio
Identity
E
atac_peak_region_fragments
ATAC_PLT_score
ATAC_MULTI_score
F
PLT vs. MULTI DE RNA features
PLT vs. MULTI DA ATAC features
−Log10 adjusted p
Log2 fold change
Vwf
Nrxn1
Plcl1
Pdgfc
Ncam2
Arhgap15
Meg3
Neo1
Aff3
Ctla2a
Ptpre
Rgs1
Sox4
Iqgap2
Pde3a
Peg3
Tnrc6a
Ryr2
Dok5
Sipa1l1
Chrm3
Ezh1
Gm26516
Arhgap30
Ppa2
Ccdc64
Sfrp2
Smtnl2
Smim3
Ifnar2
Runx3
Itpr1

Figure 8 8-week HSC dimensional reduction and unsupervised cluster annotation

A, Weighted nearest neighbour (WNN) UMAP reduction and unsupervised clustering of 8-week adult mouse HSCs utilising both the RNA and ATACseq data. **B,** Expression levels of platelet lineage (*Vwf*, *Pdgfc*), multilineage (*Ctla2a*) and cell cycling (*Ezh2*) marker genes were enriched in specific clusters on the UMAP. **C,** Cell cycle status was determined using the transcriptomic data, revealing distinct clusters corresponding to the S and G2/M phases. **D,** Per-cell scoring for lineage potential. HSC lineage potential scores were calculated using the PLT and MULTI signature RNA and chromatin regions identified in (Figure 5A). To quantify lineage bias, PLT scores were divided by MULTI scores, with higher ratios indicating stronger platelet bias. The resulting scores showed consistent heterogeneity across RNA and ATAC modalities, with differences appearing more pronounced at the chromatin level. **E,** Cycling (G2-M and S phase) 8-week HSCs exhibited higher detection of ATACseq fragments and signature region for subpopulation scoring. This highlights the necessity of excluding cycling cells for HSC subpopulation annotation. **F,** Differential expression (RNA) and accessibility (ATAC) analyses between the PLT and MULTI-1&2 clusters. Differential features were filtered for adjusted $p < 0.05$. ATAC chromatin regions were annotated by the nearest gene.

4.5 Discussion

In this chapter, we developed an analytical framework to characterise HSC subpopulations using single-cell multi-omic data, integrating transcriptomic and epigenomic dimensions through WNN analysis. This integration enabled the identification of seven distinct HSC clusters, providing a detailed map of HSC heterogeneity. Known markers associated with platelet-bias and multilineage potential, were distinctly enriched in specific clusters.

WNN integrates RNA and ATACseq datasets by constructing modality-specific nearest neighbour graphs and combining them with weights that balance the relative contributions of transcriptomic and epigenomic signals. This approach leverages the strengths of both modalities, allowing for a comprehensive analysis of HSC subpopulations while addressing potential confounders. Importantly, we used the RNAseq data to identify and exclude

cycling cells, as chromatin accessibility undergoes dynamic changes during cell division. Excluding these cells ensures that ATAC-based lineage potential scoring is not influenced by artefacts related to cell cycle-dependent chromatin remodelling.

Cell cycle states significantly impact both RNA expression and chromatin accessibility, adding complexity to data interpretation. While studies have reported chromatin accessibility generally decreases at transcription start sites (TSS) and enhancers as cells progress from G1 to G2-M (Estève et al. 2020), our dataset revealed an increased number of ATACseq fragments and higher accessibility in genes associated with lineage potential in cycling HSCs (Figure 8E). This suggests that, unlike other cell types where chromatin condenses during cell cycle progression, cycling HSCs may undergo chromatin remodelling that primes lineage-specific genes for differentiation. These findings highlight the cell and gene-specific nature of chromatin accessibility changes, making systematic normalisation challenging. Future studies could explore whether these patterns are unique to HSCs or apply more broadly to other self-renewing stem cell populations. Similarly, we chose not to use SCTransform (SCT) normalisation (Lause, Berens & Kobak 2021) for RNAseq data, as it is incompatible with the ATACseq modality used in WNN integration. While SCT effectively normalises RNAseq data and can regress out cell cycle effects, no equivalent method exists for chromatin accessibility, making it unsuitable for WNN reduction.

The lineage bias of each cluster was determined using AUCell for RNAseq data and ChromVAR for ATACseq data, scoring cells based on previously defined gene expression and chromatin accessibility signatures. This analysis predicted distinct subpopulations, including P-biased, PEM-biased and MULTI clusters. Interestingly, the deviation in lineage bias scores was more pronounced in the epigenomic data than in the transcriptomic data, highlighting the unique insights provided by chromatin accessibility.

While most clusters were well characterised, one cluster, termed Cluster 4, lacked clear definition. It did not exhibit significantly high or low scores for lineage potential signatures, nor did it display unique expression of markers linked to known functional states. This observation highlights a limitation of focusing solely on lineage potential as an axis for understanding HSC heterogeneity. Cluster 4 may represent a subpopulation with alternative functional characteristics, such as unique metabolic states, stress responses, or other specialised properties. Further exploration of these unidentified subpopulations may reveal additional dimensions of HSC biology that contribute to overall heterogeneity.

The gene expression profiles showed clearer distinctions between PLT and MULTI clusters, with interpretable lineage-specific signatures, whereas the chromatin accessibility differences were less directly aligned with lineage differentiation. This may reflect the inherent differences between modalities: gene expression provides a snapshot of active transcription, while chromatin accessibility reflects regulatory potential, which may or may

not be functionally realised. Only a subset of accessible regions, such as those near *Ifnar2* and *Runx3* in the MULTI cluster, corresponded to genes with known roles in lineage specification. The rest were more broadly associated with regulatory functions or cellular plasticity. It is also important to note that these annotations may not fully capture the true functional phenotype of these cells. Additionally, the relatively shallow sequencing depth typical of single-cell assays limits the resolution of both modalities, potentially obscuring weaker or transient regulatory programmes and contributing to the lack of significant pathway enrichment detected by GSEA. Nevertheless, differentially accessible chromatin sites may act as cis-regulatory elements influencing lineage decisions, and could still serve as useful biomarkers for predicting HSC subtypes. Future studies incorporating complementary approaches such as Hi-C or other chromatin conformation capture techniques will be valuable for linking these elements to their functional targets (Hughes & Davies 2024; Zhou et al. 2024).

In conclusion, the integration of RNA and ATACseq data through WNN provided a powerful tool for dissecting the complex regulatory landscape of HSCs. The identification of potential lineage-biased clusters and the observation of larger deviations in epigenomic lineage potential scores emphasise the importance of epigenome in understanding HSC functions. However, the presence of uncharacterised clusters such as Cluster 4 underscores the need to consider additional functional attributes beyond lineage potential when

studying HSC heterogeneity. This work provides a solid foundation for future studies investigating HSC subpopulations, functional states and their dynamics during ageing or in disease contexts.

Chapter 5: Developing FACS-enrichment panel for HSC subpopulations

5.1 Introduction and aims

Studying lineage-biased HSCs requires precise tools for phenotyping subpopulations with distinct functional features. Fluorescence-activated cell sorting (FACS) is a powerful method for isolating phenotypical cell populations. However, traditional approaches of designing FACS panels can be resource-intensive, relying on trial-and-error processes. These limitations are especially problematic when attempting to isolate lineage-biased subpopulations or monitor their dynamics under specific conditions, such as ageing or treatment interventions, without relying on genetically modified reporter models like *Vwf*-tdTomato/eGFP mice (Carrelha et al. 2018; Carrelha et al. 2024). Developing a robust FACS panel for this purpose would provide a practical alternative, enabling the identification and study of HSC subpopulations in broader genetic background or experimental setup.

To address these challenges, we used *in silico* simulations as a systematic and cost-effective approach for marker selection. By leveraging single-cell multi-omic datasets and computational tools, we aimed to identify the most informative markers for distinguishing

HSCs clusters discovered in the last chapter. This study applied recursive feature elimination (RFE), a machine learning algorithm designed to rank and select features based on their importance (Bahl et al. 2019). RFE operates by iteratively evaluating all candidate markers, removing the least relevant features in each step, and refining the set of markers until only the most discriminative ones remain. The method not only ensures that the selected markers offer maximal discrimination between populations but also minimises redundancy. Ultimately, we aimed to identify a minimal set of markers that could reliably separate platelet-biased and multilineage HSCs in real-world experiments, independent of reporter lines.

To validate the *in silico* findings, we first tested whether the selected FACS markers could enrich for *Vwf*-expressing cells in reporter mice. This provided an initial assessment of the markers' ability to isolate lineage-biased HSCs, as *Vwf* expression is a known indication of platelet, erythroid and myeloid lineage bias. This ensured that the selected markers were biologically relevant before applying them in more general contexts.

Beyond this initial validation, we experimentally tested the FACS panel's ability to enrich molecularly distinct HSC populations in wild-type mice. We employed Fluidigm single-cell qPCR to confirm the differential expression of lineage potential marker genes in the sorted populations. Fluidigm offers higher sensitivity for the targeted transcripts over sequencing, making it particularly suited for detecting low-abundance transcripts in HSCs. This

technique addresses a limitation of DOGMAseq, where the shallow 3' RNAseq depth often misses a significant number of differentially expressed genes. By confirming lineage-specific gene expression early in the workflow, Fluidigm provides molecular evidence for the identity of FACS-isolated populations before proceeding to transplantation assays, which require a 16-week timeline for functional validation. And we also expect this workflow to become an accurate molecular readout of HSC lineage potential in the future if it is successfully validated by transplantation assays in the next chapter.

5.2 Enrichment of platelet-biased HSCs through *in silico* gating

After assessing the lineage potentials of WNN clusters in adult mouse HSCs, we aimed to develop a FACS enrichment panel for the two extreme clusters: PLT and MULTI-HSCs (Figure 8A). Using *in silico* simulation, we systematically ranked the CITEseq features for distinguishing these two clusters. The top five most important CITEseq markers identified were CD274, CD61, CD357, P2X7R and CD27 (Figure 9A-B).

By filtering cells based on these five markers (CD61 > 9, CD274 > 35, CD27 > 16, CD357 > 95, P2X7R > 3), it increased the proportion of PLT-HSCs from 17% of the total dataset to 43%. Conversely, applying the opposite filtering criteria reduced the proportion of PLT-HSCs to just 2% (Figure 9C). With these results, we proceeded to validate this panel experimentally in the wet lab.

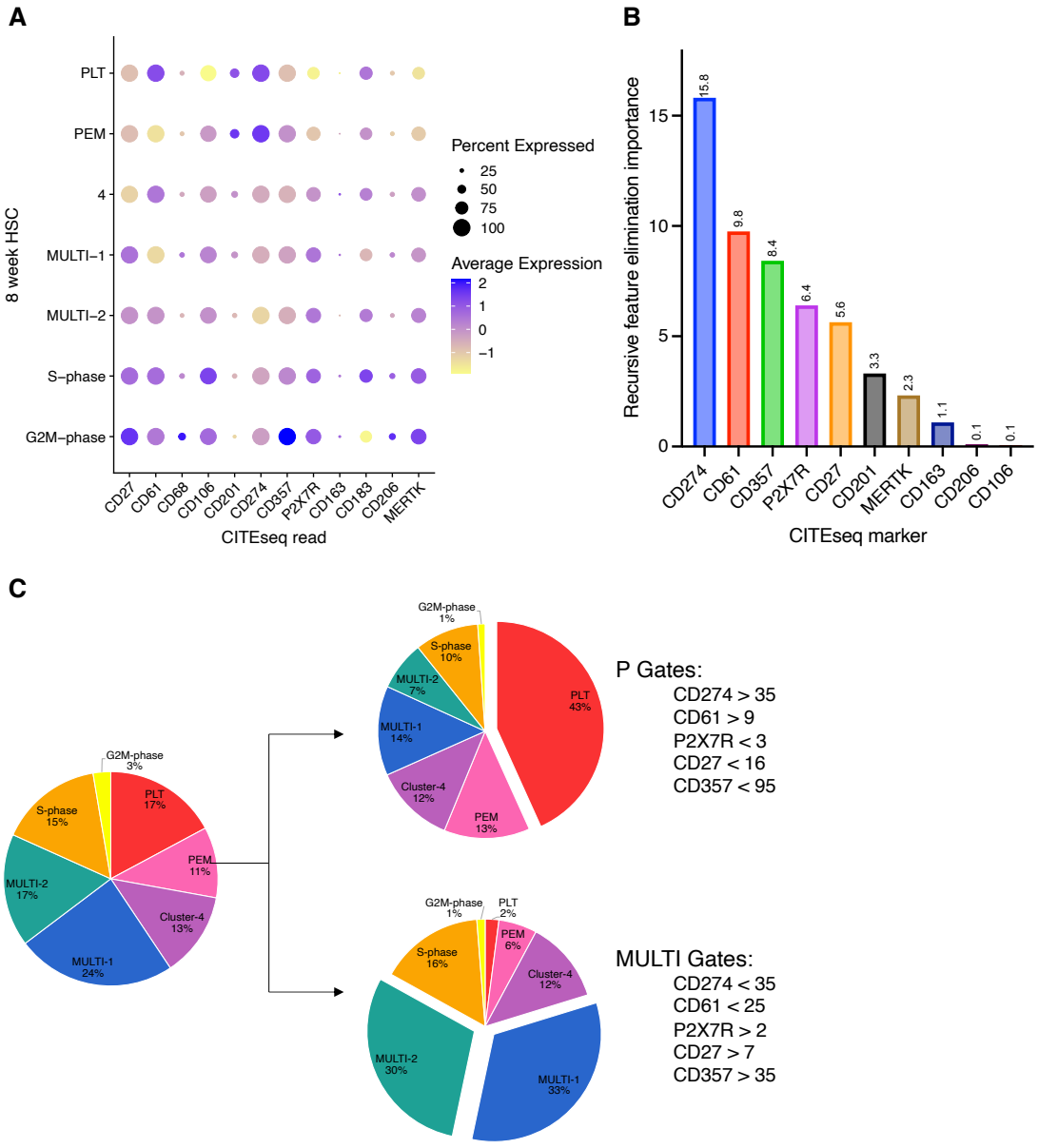


Figure 9 *In silico* gating simulation for enriching HSC clusters

A, Antibody-derived tag counts across the 8-week HSC clusters. **B,** CITEseq markers ranked by recursive feature elimination importance score for separating platelet-biased (PLT) and multilineage (MULTI-1&2) HSC clusters. **C,** *In silico* gating simulation. HSCs were filtered by the top five ranked CITEseq features, which effectively enriched for PLT and MULTI HSCs. The strategy was further validated with molecular and functional assays in (Figure 10-11).

5.3 FACS marker combination enriches *Vwf*+ HSCs

We next tested the FACS panel's ability to enrich for *Vwf*-expressing HSCs in GFP reporter mice (Figure 10). Without any gating, 52% of the general HSC population in adult female mice were *Vwf*+. By gating for HSCs that were P2X7R-, CD274 high, CD61 high, CD27 low and CD357 low, 94% of this population expressed *Vwf*, indicating a strong enrichment for lineage-biased HSCs. In contrast, gating for the opposite population resulted in only 5.9% *Vwf* positive cells.

Further testing of this FACS panel revealed that the marker P2X7R had poor staining quality, leading to suboptimal separation of cells in this channel and complicating the setting of positive and negative gates. It also significantly reduced the number of viable cells after filtering. By excluding P2X7R from the panel and using only the remaining four marker with more stringent gates, we achieve comparable levels of *Vwf*+ enrichment while preserving higher cell numbers (Figure 10A). As a result, we discarded P2X7R from the panel.

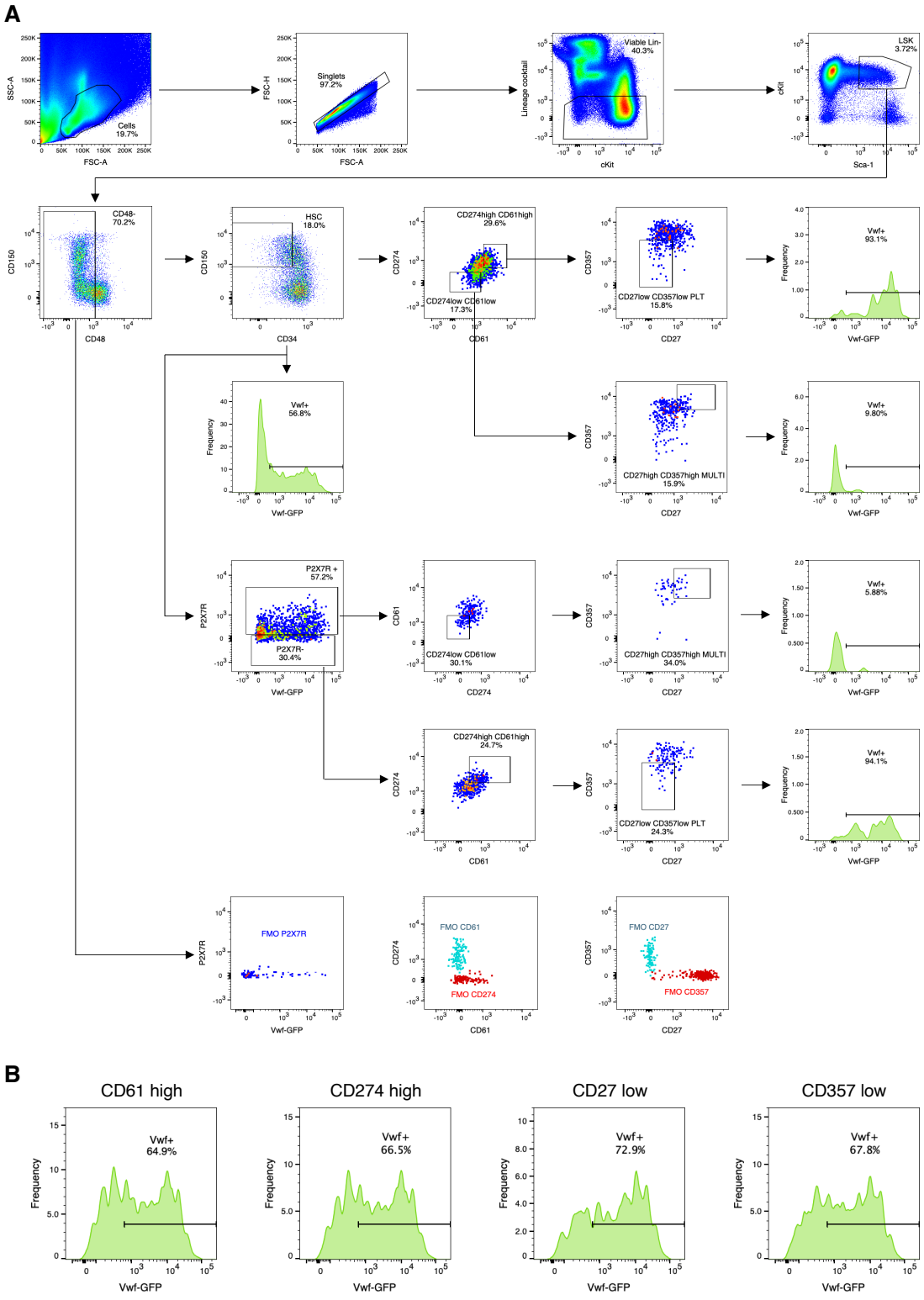


Figure 10 FACS marker combination enriches *Vwf*+ HSCs

The newly identified FACS markers were tested on *Vwf*-eGFP mouse HSCs. **A,** In unfractionated adult mouse HSCs, 57.6% were *Vwf*-eGFP+. Applying gating based on the five markers enriched the *Vwf*-eGFP expression to 94.1%, while selecting the opposite population reduced the expression to 5.9%. Excluding P2X7R from the gating strategy achieved similar levels of enrichment while preserving more cells. This

optimised four-marker strategy was used in subsequent experiments. Population gates were positioned to maximise the enrichment for *Vwf*+ HSCs. **B,** Single-marker gating was less efficient at enriching for *Vwf*+ HSCs

5.4 Differential expression of lineage potential genes in FACS-purified HSC subpopulations

To further assess the lineage potential of the FACS-purified HSC subpopulations, we conducted single-cell Fluidigm qPCR analyses. We sorted 64 PLT, 64 MULTI and 64 unfractionated HSCs from wild-type mice and probed for the top eight differentially expressed genes from the expanded clone analysis (Table 2). We also included *Vwf*, the four FACS markers and two housekeeping genes (*Hprt* and *Gapdh*) in the Fluidigm panel.

Target	Signature	Catalogue
Hprt	House keeping	Mm03024075_m1
Gapdh	House keeping	Mm99999915_g1
Vwf	Platelet-bias	Mm00550376_m1
Gpr4	Platelet-bias	Mm00558777_s1
Inhba	Platelet-bias	Mm00434339_m1
Kynu	Platelet-bias	Mm00551012_m1
Gabra4	Platelet-bias	Mm00802631_m1
Evc	Platelet-bias	Mm00469587_m1
Plek2	Platelet-bias	Mm00600165_m1
Mroh7	Platelet-bias	Mm01267126_m1
Pdgfc	Platelet-bias	Mm00480205_m1
Gapt	Multilineage	Mm01301496_m1
Cd27	Multilineage	Mm01185212_g1
Cpa3	Multilineage	Mm00483940_m1
Clec4d	Multilineage	Mm00501738_m1
Crip2	Multilineage	Mm00517877_m1
Flt3	Multilineage	Mm00439016_m1
C2	Multilineage	Mm00442726_m1
Sec16b	Multilineage	Mm00473719_g1

Ctla2a	Multilineage	Mm00484032_g1

Table 2 TaqMan probes used in Fluidigm experiments

The Fluidigm detected signals from five P-specific genes (*Vwf*, *Pdgfc*, *Plek2*, *Inhba* and *Gpr4*) and five MULTI-specific genes (*Flt3*, *Cd27*, *Ctla2a*, *Clec4d* and *Sec16b*). The other probes that did not yield signals in more than three cells, were excluded from the analysis. After normalising the data to the housekeeping genes, we observed highly enriched *Vwf* expression in the isolated PLT-HSCs, while *Ctla2a*, an established MULTI marker, was strongly expressed in isolated MULTI-HSCs (Figure 11A).

To systematically compare the sorted populations, we generated a hierarchically clustered heatmap, which revealed clear separation between the PLT and MULTI-HSCs (Figure 11B). Notably, nearly all PLT-HSCs were positive for *Pdgfc* and *Vwf*; while a subset of MULTI-HSCs expressed high levels of the progenitor marker *Flt3*, suggesting that some of the sorted MULTI population may consist of short-term HSCs, or progenitor-like cells. Additional evidence supporting this speculation will be presented in the next chapter.

We also tested the newly identified FACS marker and qPCR panel for phenotyping HSCs from young (8-week, n = 3) and aged (98-week, n = 2) mice in an independent experiment. All samples were stained using the same FACS antibody cocktail to ensure comparability across samples. As expected, aged HSCs showed a significant increase in cell falling within the PLT gate and a reduction in the MULTI gate (Figure 11C), consistent with an ageing-

associated shift towards platelet-bias.

We then compared the single-cell qPCR profile in young and aged HSCs against previously profiled PLT and MULTI-HSCs. Recognising that TaqMan probe signals can vary between runs, we applied a single-cell RNAseq data integration-like strategy: gene expression values were scaled and centred within each experiment prior to integration. We then projected all samples onto a PCA space. This revealed that aged HSCs clustered more closely with the PLT-HSC reference, while young HSCs aligned with the MULTI-HSCs (Figure 11D).

To quantify lineage potential on a linear scale, we defined a pseudo axis, setting the centroid of PLT-HSC cluster as coordinate 0 and the MULTI-HSC centroid as 5 (Figure 11E). Each cell was then positioned along this 1-dimensional trajectory. The aged HSCs appeared closer to the PLT end, and the young HSCs were located nearer to the MULTI end (Figure 11F).

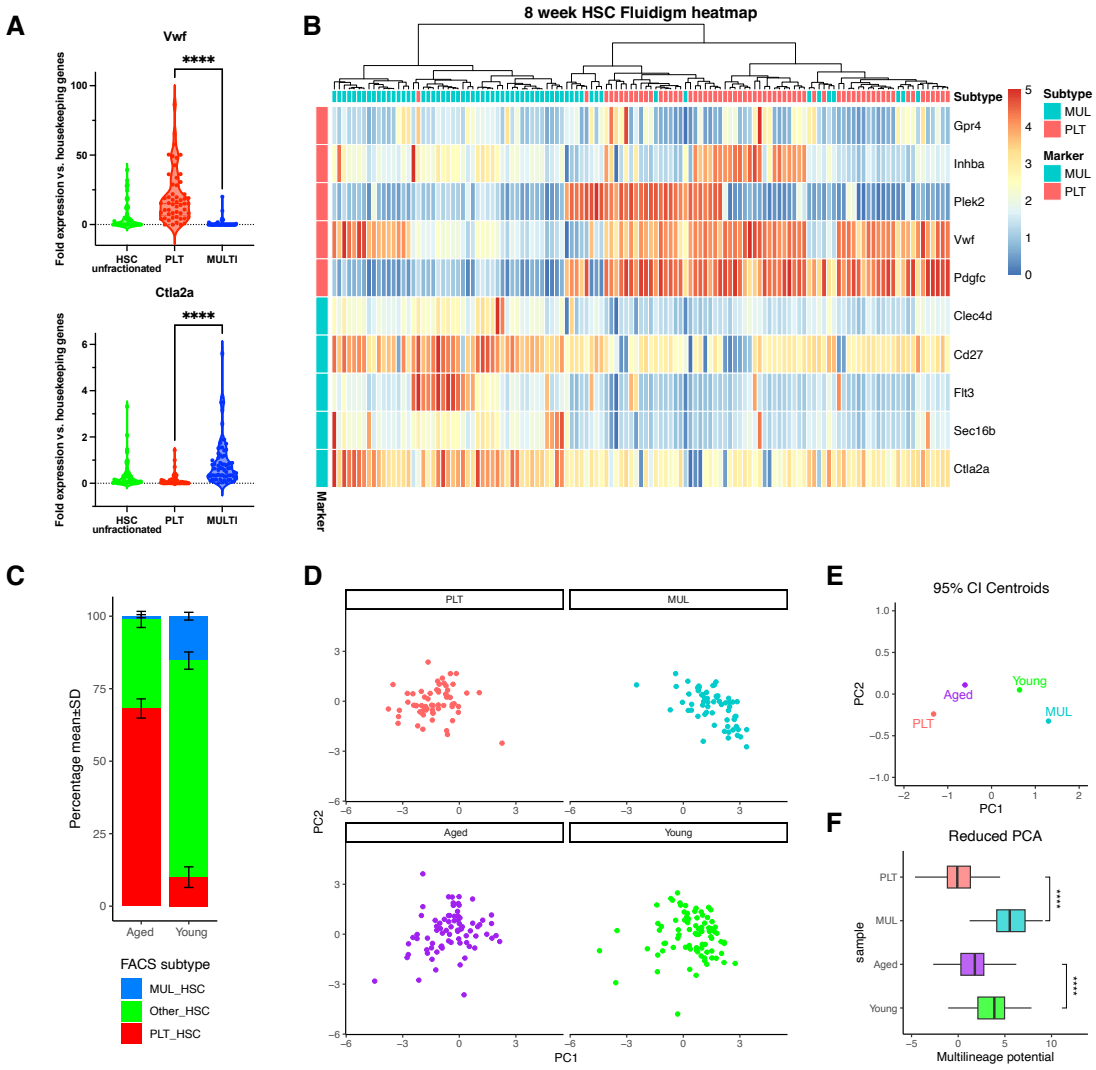


Figure 11 FACS-isolation using new markers enriches mouse HSC subpopulations

Using the four-marker strategy developed in (Figure 10A), adult 8-week unfractionated, PLT and MULTI-HSCs were FACS-isolated for single-cell qPCR of lineage potential marker expression (Table 2). **A,** Known platelet-bias marker *Vwf* and multilineage marker *Ctla2a* were enriched in the respective HSC subpopulations. **B,** Hierarchically clustered heatmap revealed distinct molecular patterns of FACS-sorted PLT and MULTI HSCs. **C-F,** FACS and qPCR phenotyping of young (8-week, n = 3) and aged (98-week, n = 2) mouse HSCs. **C,** FACS gating confirmed significant increased platelet-bias and decreased multilineage phenotype in aged HSC compartments. **D,** Principal component analysis (PCA) reduction of integrated qPCR data. **E,** 95% confident interval centroids of each sample. **F,** To compare sample similarity, a lineage vector connecting the PLT and MULTI centroids was created. Each cell's lineage potential was quantified by their perpendicular crossing coordinate on the PLT-MULTI axis. An unpaired T-test showed that aged HSCs were significantly more platelet-biased than young HSCs.

5.5 Discussion

In this chapter, we developed a FACS panel aimed for enriching lineage-biased HSCs using an *in silico* approach. This strategy used machine learning to identify a minimal set of markers that could distinguish PLT and MULTI-HSCs. The selected markers: CD274, CD61, CD357, P2X7R and CD27 were tested experimentally, with promising results demonstrating their ability to enrich for *Vwf*-expressing cells in reporter mice. These findings provided an initial validation of the panel's capacity to isolate lineage-biased HSCs, but several limitations and challenges emerged.

The marker P2X7R exhibited poor staining quality, complicating gating and significantly reducing the number of cells recovered. Consequently, P2X7R was excluded from the final panel, and the remaining markers provided comparable enrichment for *Vwf*-expressing cells while preserving higher cell numbers.

One critical assumption underpinning the *in silico* simulation is that the clusters with differential scores for lineage potential signatures represent the functional lineage-biased HSCs of interest. While this assumption is supported by the expression of known markers and the enrichment of *Vwf*-expressing cells, it remains to be fully validated functionally through downstream experiments such as transplantation. Furthermore, the cutoffs for *in silico* gating simulation are inherently arbitrary and difficult to directly translate into

measurable mean fluorescence intensity (MFI) values for FACS gating. This makes it challenging to estimate the purity of enrichment *in vivo*.

Another challenge is the complexity of the FACS panel itself (Supplementary Table 1). With 10 fluorophores, this panel is still relatively intricate for routine use in a laboratory setting. However, the panel still allows for accommodation of additional gene reporters, such as GFP or tdTomato (by moving CD61 PE to PerCPe710). An additional source of variability comes from the gating strategy: HSC subpopulations were defined based on relatively high or low expression of novel markers, rather than strict positive or negative gates. As a result, it is difficult to apply identical gating thresholds across experiments, animals, cytometers, and antibody batches. To mitigate this, we consistently included an 8-10 week control sample in each experiment and adjusted gates to maintain consistent percentages in the control. In the future, calibration tools such as rainbow beads, which contain particles with defined intensities across multiple channels, could help fine-tune and standardise gate placement across instruments and timepoints.

The Fluidigm single-cell qPCR results provided molecular evidence for the distinct lineage potentials of the sorted populations, despite some technical limitations. Some of the TaqMan probes failed to generate signals or show significant differences, possibly due to low expression levels or primer specificity. However, the remaining probes effectively highlighted the differential expression of molecular signatures in isolated PLT and MULTI-

HSCs. The systematic PCA and heatmap analyses further confirmed the distinct clustering of these populations, validating the FACS panel's ability to isolate biologically relevant subpopulations.

Although we mentioned that lineage potential differences are often more pronounced at the epigenomic level, we chose to use single-cell qPCR for its practical advantages. In particular, qPCR allows for index sorting, enabling us to directly link transcriptional data with FACS phenotypes. It also requires the lowest input cell numbers, making it well-suited for rare HSC populations. While single-cell ATACseq and DNA methylome (Nichols et al. 2025) protocols do exist, they are not compatible with index sorting and typically involve bulk tagmentation followed by re-sorting of nuclei, a process that leads to substantial cell loss and increases experimental complexity. Moreover, single-cell ATACseq generally yields lower per-cell sensitivity compared to qPCR, limiting its utility for detecting fine-grained signature activity. We therefore believe that single-cell qPCR remains a more practical and accessible method for functional stratification of rare HSC subsets. Looking ahead, the development of plate-based single-cell ATAC array technologies would be a valuable addition to the field.

The selected FACS markers are supported by parallel studies on HSC phenotype and function. CD61 and CD274 have both been implicated in HSC expansion and survival over ageing. CD61 is enriched in aged HSCs, marking a quiescent, self-renewing

subpopulation (Skinder et al. 2024), while CD274 (PD-L1) plays a role in HSC survival and myeloid differentiation pathways (Yamashita & Passegué 2019). The higher expression of CD61 and CD274 in PLT-HSCs supports the fact that lineage-biased HSCs expand with age. However, the cited study also showed that CD61 alone does not effectively enrich for lineage-biased HSCs in young adult mice, highlighting the necessity of using complementary markers for isolating functionally distinct subpopulations. Additionally, CD27 has been implicated in lymphoid proliferation, and its high expression correlates with shorter engraftment longevity (Wiesmann et al. 2000), while CD357 expression is found exclusively on lymphocytes among all mature blood cells. (Rodríguez-Bayona, Ramos-Amaya & Brieva 2011; Ronchetti et al. 2015). These might explain our finding that MULTI-HSCs contained more *Flt3*+ cells, suggesting that CD27 and CD357 high HSCs may be primed for lymphoid differentiation and include a subset of committed progenitors rather than true long-term multilineage HSCs. Importantly, when each of the four markers was tested individually for its ability to enrich *Vwf* + HSCs, none achieved greater than 15% enrichment (Figure 10B). This underscores the necessity of using a multi-marker approach.

To further validate the utility of the panel, we applied it alongside the qPCR gene expression panel to phenotype young and aged HSCs. The results showed an increase in platelet-biased and a reduction in multilineage HSCs with age, consistent with current understanding of HSC ageing dynamics. Importantly, we also developed an integration

framework for mapping qPCR data onto the reference PLT and MULTI-HSC transcriptional profiles. This may prove to be a powerful tool for phenotyping new samples in future studies, especially if the PLT and MULTI reference populations are further validated by functional assays.

In conclusion, this chapter demonstrates the potential of *in silico* simulations for designing a FACS panel to enrich lineage-biased HSCs. Validations confirmed that the panel effectively enriched *Vwf*-expressing HSCs, while also offering flexibility for incorporating additional reporter lines. Molecular analysis using Fluidigm further indicated the distinct lineage potentials of the sorted cells, providing strong evidence that the selected markers reliably distinguish HSC subpopulations. Despite the possibility of capturing some progenitor cells, the panel serves the purpose of isolating HSC subpopulations. These findings establish a solid foundation for further functional validation in transplantation assays described in the next chapter.

Chapter 6: Confirming lineage potential of FACS-isolated HSCs in transplantation

6.1 Introduction and aims

Transplantation assay is the gold standard for functionally validating the capacity of HSCs to reconstitute specific blood lineages. Previous studies have shown that approximately 50% of single, unfractionated transplanted HSCs are multilineage, with around 30% being lineage-restricted (Carrelha et al. 2018). When single *Vwf*+ HSCs were transplanted, the proportion of MULTI-HSCs, remained similar, but there was an increase in platelet (P), platelet-erythrocyte (PE) and platelet-erythrocyte-myeloid (PEM)-restricted HSCs, alongside a decrease in platelet-erythrocyte-erythrocyte-myeloid-B (PEMB)-HSCs (Figure 13B). These findings demonstrated that the *Vwf* reporter HSCs provide modest enrichment for HSCs with reduced lymphoid potential but still lack the precision necessary for studying platelet-biased HSCs in detail.

In the preceding chapters, we developed a refined FACS panel to isolate potentially lineage-biased HSCs, particularly focusing on the PLT and MULTI-HSCs identified *in silico*. These markers were selected through machine learning simulations, validated experimentally for enriching *Vwf*+ HSCs, and confirmed to represent distinct subsets via

molecular profiling using single-cell qPCR. However, while molecular evidence strongly supports the lineage potentials of the isolated cells, functional validation through transplantation remains essential.

The aims of this chapter are twofold. First, we sought to confirm the lineage potential of FACS-isolated HSC subpopulations through bulk transplantations, where HSCs from PLT, MULTI or unfractionated populations were transplanted into bone marrow-depleted recipients. This approach allowed us to assess early, intermediate and stable reconstitution in peripheral blood (PB) and bone marrow (BM). Second, we aimed to eliminate potential confounding factors, such as donor cell clonal competition, by performing single-cell transplantations. This strategy enables a direct comparison of the lineage potential in individual HSCs, providing a higher-resolution validation.

Through these experiments, we aimed to establish that our FACS panel offers improved enrichment of lineage-biased HSCs over the *Vwf* reporter, enabling more precise studies of HSC heterogeneity.

6.2 Bulk transplantation

We conducted two rounds of bulk transplantation using PLT, MULTI and unfractionated HSCs, sorted based on the four newly identified markers. For each recipient, 50 HSCs were transplanted alongside competitor bone marrow cells. Peripheral blood chimerism

was analysed at week 6, 10 and 16 to assess early, intermediate and stable reconstitution (Supplementary Figure 2A-B).

In the first cohort, the PLT recipients exhibited a strong bias toward platelet reconstitution, while the MULTI recipients showed more balanced lineage reconstitution in the week 6 PB readout. However, the overall percentage of donor-derived blood cell reconstitution declined over time, falling below 10% in most recipients by week 16 (Figure 12A). At this point, the PLT recipients still demonstrated higher donor-derived reconstitution in platelet, erythrocytes and myeloid cells compared to lymphoid cells, though lymphoid reconstitution has increased relative to earlier time points. This suggests that, while the sorted PLT-HSCs showed an overall platelet-bias, not all of them were fully lineage-restricted and might contain multilineage cells. According to our experience, multilineage HSCs tend to produce higher absolute number of blood cells compared to biased HSCs, which will help their domination in chimerism analyses. It also takes donor HSCs longer time to initiate lymphopoiesis than producing megakaryocytes, these might explain the slow rise of lymphoid cell percentage over time in this experiment.

The endpoint bone marrow chimerism analysis (Supplementary Figure 2C) revealed poor engraftment in a large portion of the recipients: 2/4 PLT, 3/4 MULTI and 1/3 unfractionated HSC recipients had < 1% donor-derived HSCs remained (Figure 12B), indicating suboptimal engraftment. Notably, the well-engrafted PLT recipients had the

highest chimerism in MPP2, supporting the idea of platelet-bias at the progenitor level. In contrast, MULTI recipients showed the highest reconstitution in MPP4, suggesting a potential lymphoid bias (Figure 12B) (Pietras et al. 2015). The fact that donor-derived MPPs in MULTI recipients displayed much higher reconstitution levels than HSCs indicates that the isolated MULTI cells may include short-term HSCs or progenitor-like cells primed for differentiation.

Due to the compromised results, we performed a second round of bulk transplantation (Figure 12C-D). Some PLT recipients again exhibited biased reconstitution toward platelet, erythrocyte and myeloid cells, with higher chimerism in MPP2 and MPP3 compared to MPP4. Meanwhile, MULTI cells again showed declining engraftment, with 3/4 recipients displaying < 10% donor-derived HSCs in bone marrow at the endpoint, along with a continued lymphoid bias in peripheral blood and at the MPP level.

The lineage reconstitution patterns in the unfractionated HSC group were inconsistent across biological replicates. We hypothesised that this variability could be due to unequal insufficient depletion of recipient bone marrow, resulting in variable selective pressure within the niche and exacerbated clonal competition among transplanted cells. Upon reviewing our local setup, we discovered that the irradiator configuration was flawed, resulting in a lower than required dose of irradiation. This finding prompted us to further question whether variability in residual host marrow contributed to the differences in

reconstitution kinetics and clonal dynamics observed. To eliminate these potential confounding factors, we moved on to single-cell transplantation assays using a corrected and optimised irradiation scheme.

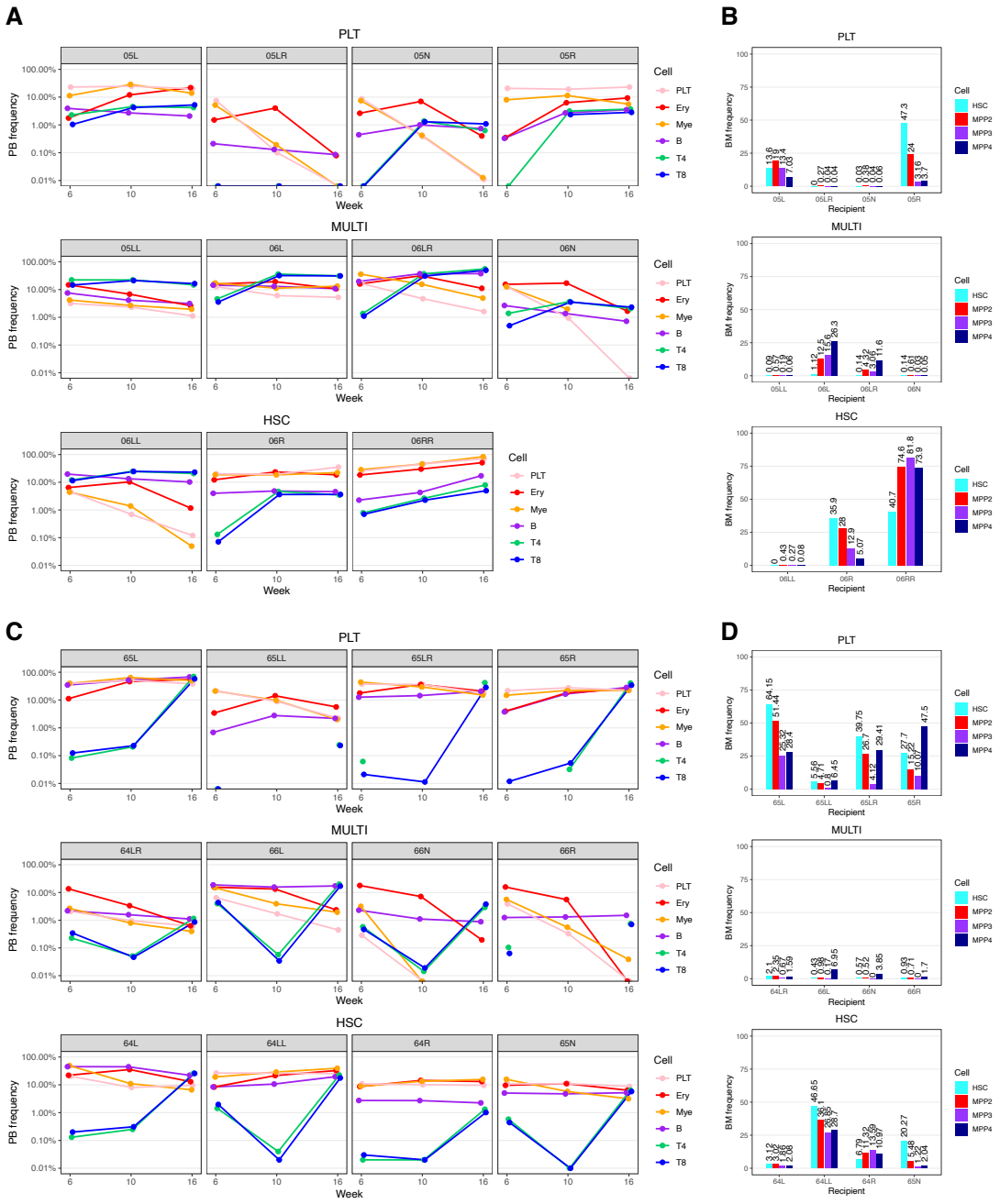


Figure 12 Bulk transplantation of FACS-isolated HSC subpopulations

A-B, Batch 1 of transplanting 50 HSCs into lethally irradiated recipient. **A,** Peripheral blood cell reconstitution pattern; **B,** Bone marrow HSC and multipotent progenitor reconstitution pattern. **C-D,** Batch 2 of transplanting 50 HSCs into lethally irradiated recipient. Initial platelet bias was observed in most of the PLT-HSC recipients, while reconstitution of lymphoid lineages eventually increased, possibly due to contaminating non-biased cells. At the 16-week bone marrow readout, there was an enrichment of megakaryocyte-biased MPP2 cells in PLT recipients. However, the control group receiving unfractionated HSCs showed inconsistent lineage output and variable overall engraftment. We suspected results from both batches were affected by suboptimal irradiator settings and potential clonal competition, leading us to adopt a single-cell transplantation approach for improved resolution.

6.3 Single-cell transplantation

We transplanted 60 single FACS-isolated PLT-HSCs (Supplementary Table 2 & Figure 10A) into lethally irradiated recipients. In total, 27 recipients demonstrated long-term engraftment and 24 (89%) showed various degrees of lineage restriction with 12 (44%) being P-restricted; only 3 donor cells (11%) reconstituted all lineages (Figure 13A). In contrast, previous studied that transplanted unfractionated HSCs reported around 50% of donor cells being multilineage (Figure 13B) (Carrelha et al. 2018). A Chi-square test showed $p < 0.005$ between the FACS-isolated PLT and general HSC lineage reconstitution pattern, indicating a significant enrichment of the lineage-biased HSCs.

We also observed that platelet-restricted donor HSCs typically results in lower PB chimerism on average compared with HSCs with a less pronounced lineage bias. This finding aligns with previous reports, which indicate that a higher degree of lineage bias is often associated with reduced reconstitution capacity (Carrelha et al. 2018).

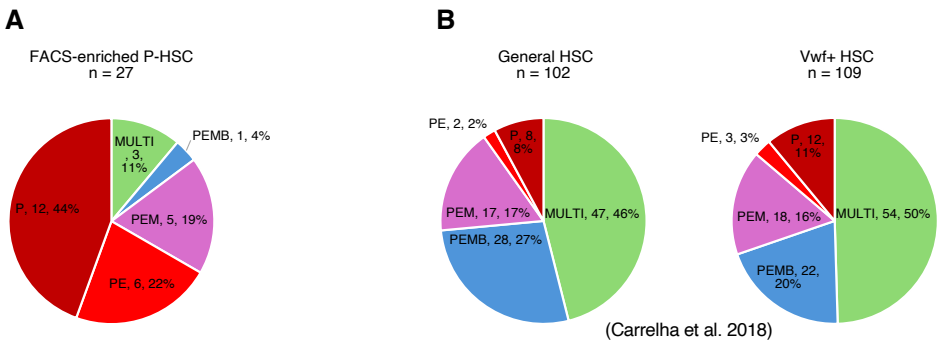


Figure 13 Single-cell transplantation of FACS-isolated PLT-HSCs

A, Single PLT-HSCs isolated by the four newly identified FACS markers (Supplementary Table 1) were transplanted into lethally irradiated recipients. Their lineage potentials were determined by the week 16

peripheral blood reconstitution readout as platelet-restricted (P), platelet-erythrocyte (PE), platelet-erythrocyte-myeloid-restricted (PEM), platelet-erythrocyte-myeloid-B-cell-restricted (PEMB) and multilineage (MULTI). **B,** Lineage potential distribution of unfractionated general HSCs and *Vwf*+ HSCs reported by previous studies (Carrelha et al. 2018).

6.4 Discussion

In this chapter, we explored the lineage reconstitution patterns of FACS-isolated PLT and MULTI-HSCs through bulk and single-cell transplantation experiments. Initial results from bulk transplantations revealed that groups of PLT-HSCs exhibited a pronounced bias toward platelet, erythrocyte and myeloid reconstituting kinetics, with gradual lymphoid production emerging over time. This delayed lymphopoiesis suggests that while the bulk 50 PLT-HSCs display a strong initial lineage bias, they might still include multilineage cells that dominate haematopoiesis over time due to their higher reconstitution capacity. Conversely, MULTI-HSCs showed a more balanced or lymphoid-biased initial reconstitution but demonstrated suboptimal long-term engraftment capacity, likely reflecting an enrichment of short-term HSCs or progenitor-like cells within this population. As CD61 and CD274 are both involved in HSC survival and self-renewal, selectively excluding these populations may inadvertently eliminate platelet lineage potential and self-renewal ability, leading to a bias toward lymphoid-primed short-term cells in the MULTI populations.

The poor overall engraftment and inconsistent lineage reconstitution, including in the

unfractionated HSC controls, prompted a careful review of the experimental conditions. We identified issues with the irradiation instrument, which required recalibration to ensure effective recipient bone marrow depletion. In addition, several factors may have contributed to variability in reconstitution kinetics and clonal competition across recipients. It is known that regenerative stress can trigger accelerated or alternative differentiation pathways (Carrelha et al. 2024), and bone marrow microenvironment conditionings can favour the expansion of different clones (Valletta et al. 2020). The extent of these factors likely varies depending on the actual biological dose received by each mouse. Despite identical genetic background and age, recipients in this cohort ranged from 15 to 20 grams in weight, potentially affecting radiation exposure. Body hydration status and subtle differences in positioning during irradiation might have further influenced the uniformity of marrow depletion.

To better assess depletion, we considered using CD45.1/CD45.2 competitor chimeras, where residual recipient-derived (CD45.2) cells could be distinguished from both donor (CD45.1) and competitor marrow. Other models that could be considered to avoid the need for irradiation and potentially reduce toxicity include genetic approaches such as Kit hypomorphic mice (Cosgun et al. 2014), or antibody-based conditioning targeting c-Kit (Czechowicz et al. 2007) or CD45 (Saha et al. 2024), which allow selective depletion of host HSPCs.

We also speculate that the stabilised lineage chimerism readout at 16 weeks might be over-represented by high-output donor clones. For example, among 50 PLT-gated HSCs, a single multilineage-competent cell could dominate haematopoiesis and contribute to over 80% of peripheral blood reconstitution, an outcome observed in previous single-cell transplantation studies. As a result, the 16-week readout may resemble multilineage output, despite the original input population containing a larger proportion of lineage-biased HSCs.

These concerns motivated our switch to single-cell transplantation assays, which avoids clonal competition among donor cells and provides a more precise assessment of individual HSC lineage potential. Our single-cell transplantation experiments demonstrated the superiority of the newly identified markers, with most sorted PLT-HSCs exhibiting lineage-restricted reconstitution patterns. Compared to both unfractionated and *Vwf*+ HSCs, the PLT-HSCs showed significant enrichment in lineage bias, indicating the robustness of this FACS panel. Notably, the observed difference in reconstitution abilities (engraftment percentage) suggests that lineage-biased subpopulations might not be responsible for the majority of steady-state haematopoiesis. Instead, they may play specialised roles under acute stress conditions, for example, undergo accelerated differentiation via alternative pathways (Carrelha et al. 2024). The control data of unfractionated HSC subpopulation ratio, while insightful, must be interpreted cautiously, as the data were derived from different donor mice (Carrelha et al. 2018). However, the

large ratio of P-restricted HSCs in our transplantation results provides strong evidence of substantial enrichment.

Looking ahead, future studies could investigate the response of these FACS markers to stress conditions, such as acute immune challenges, and explore their dynamics during ageing. This could reveal whether the expression levels of these markers change and provide insights into the adaptive roles of lineage-biased HSCs under different conditions. It also remains to be validated whether this enrichment strategy consistently captures the same lineage-biased phenotypes across different ages.

Another key consideration for future research is the development of a FACS-based strategy to selectively enrich multilineage long-term HSCs, as these cells are the most clinically relevant for transplantation therapies. Given that excluding CD61 and CD274 high cells may inadvertently remove subsets with higher self-renewal potential and platelet lineage reconstitution, a refined gating strategy may be necessary. One possible approach is to selectively sort HSCs based only on high CD27 and CD357 expression, ensuring robust lymphoid reconstitution while not eliminating platelet and myeloid potentials. Alternatively, a retrospective strategy could be employed: unfractionated HSCs could be transplanted while recording their expression levels for all four markers, allowing for post hoc identification of the marker combinations most predictive of robust multilineage reconstitution.

Overall, this chapter establishes that our newly identified FACS markers provide a more precise and effective method for isolating lineage-biased HSCs, offering significant advancements over previous approaches. These findings lay the groundwork for future research into the specialised roles of lineage-biased HSCs.

Chapter 7: Studying HSC subpopulation dynamics during ageing

7.1 Introduction and aims

Understanding and tracing HSC subpopulations is essential for characterising ageing within the compartment. While ageing is associated with well-documented conditions like declines in adaptive immunity and clonal haematopoiesis, it remains unclear whether these changes are primarily driven by overall HSC functional decline or shifts in specific subpopulations. This chapter addresses both aspects by examining ageing features at the compartment level and within distinct HSC subpopulations, aiming to clarify their contributions to ageing-associated functional changes.

To systematically study HSC ageing, we established and validated a robust label transfer pipeline that maps aged HSCs onto the adult HSC reference dataset, enabling consistent annotation of subpopulations across datasets. Two complementary approaches were employed for this task: the supervised principal component analysis (sPCA) method and the machine learning-based Cellcano framework (Figure 14A).

The sPCA approach identifies a projection of the transcriptome feature set that maximally captures the structure defined in the weighted nearest neighbour (WNN) graph (which

integrates both RNA and ATACseq modalities). By using the multi-omic WNN graph as a guide, sPCA ensures that the projected RNAseq data reflects label-specific molecular features encoded in the transcriptomic layer. This allows sPCA to align aged HSCs with lineage-specific clusters in the 8-week reference WNN UMAP, while also providing prediction confidence scores to evaluate the reliability of the label transfer. Importantly, the final features used for label transfer in sPCA are derived solely from the RNA modality, making it inherently transcriptomic-weighted.

In contrast, Cellcano adopts a machine learning approach that integrates RNA and ATACseq modalities equally, treating both datasets as complementary sources of information. For this study, Cellcano was adapted to use both ATAC-converted gene scores and normalised RNAseq count matrices as inputs, selecting the top 3,000 most informative features irrespective of their modality. Cellcano employs a two-stage classification process: an initial prediction using a multi-layer perceptron, followed by refinement via knowledge distillation using low-entropy pseudo-labelled data. This iterative refinement enables Cellcano to handle distributional shifts between reference and target datasets, ensuring robust predictions even for intermediate or less-defined phenotypic states.

Together, these methods establish a comprehensive strategy for annotating aged HSC lineage potential with high confidence. This pipeline offers a robust framework for tracing

molecular changes in ageing HSCs and supports the consistent analysis of datasets across experimental conditions.

Ultimately, this chapter aims to uncover the molecular characteristics of HSC ageing. By leveraging the subpopulation labels matched through our pipeline, we systematically investigated ageing-associated changes in transcriptomes and epigenomes at both the compartment level and within specific HSC subpopulations. Pathway enrichment and motif analyses were conducted to identify key regulatory mechanisms driving these changes. Particular attention was given to PLT and MULTI-HSCs, as these subpopulations exhibited asymmetrical expansions with ageing, suggesting they may be potential targets for interventions aimed at reversing the decline in haematopoiesis.

7.2 Transfer of HSC subpopulation labels from 8-week to 24-month dataset

To characterise HSC ageing, it is essential to deconvolute two key processes: the shift in clonal dynamics of HSC subpopulations, marked by an increase in platelet-biased clones and a decline in multilineage clones; and the universal ageing effects that impact all HSCs. To achieve this, we transferred HSC subpopulation labels identified in the 8-week dataset to the aged 24-month dataset.

Our first approach used the sPCA method to project the 24-month HSCs onto the 8-week

WNN UMAP. This method successfully mapped 24-month HSCs to the PLT, MULTI and cycling clusters with relatively high prediction confidence (Figure 14A). To evaluate the prediction quality, we generated an independent WNN UMAP for the 24-month HSCs and visualised the predicted label distribution. Cells with the same predicted labels were closely located on this unsupervised UMAP (Figure 14C), reflecting strong phenotypic similarities. Furthermore, the predicted PLT-HSCs exhibited the highest average chromatin and RNA scores for platelet-bias (Figure 14D). Examination of the CITEseq profiles of the 24-month HSCs corroborated these findings, as the predicted PLT-HSCs retained the characteristic high CD61, CD274 and low CD27, CD357 phenotypes (Figure 14 E).

We also employed Cellcano (Ma et al. 2023), a machine learning tool for the same task. Using ATACseq gene scores and RNAseq features from the 8-week dataset, we trained a classifier model that achieved 78% accuracy on the 8-week testing dataset. The Cellcano-predicted labels for 24-month HSCs aligned closely with the sPCA prediction, particularly for the PLT and MULTI labels (Figure 14A). However, a notable difference was observed in the classification of platelet-erythrocyte-myeloid (PEM) HSCs: while sPCA predicted these as PEM, Cellcano predominantly assigned them to either PLT or MULTI labels (Figure 14B). Further investigation revealed cells for which the two approaches disagreed generally exhibited lower prediction confidence in sPCA, suggesting they might represent

intermediate or less-committed phenotypic states transcriptomically.

Given the strong overall consistency between the two prediction approaches, we proceeded to investigate ageing features within specific HSC subpopulations. To ensure robust identification of ageing features, we performed differential analyses separately for each prediction result and took genes and chromatin regions that were consistently differentially regulated across both methods (Figure 15A).

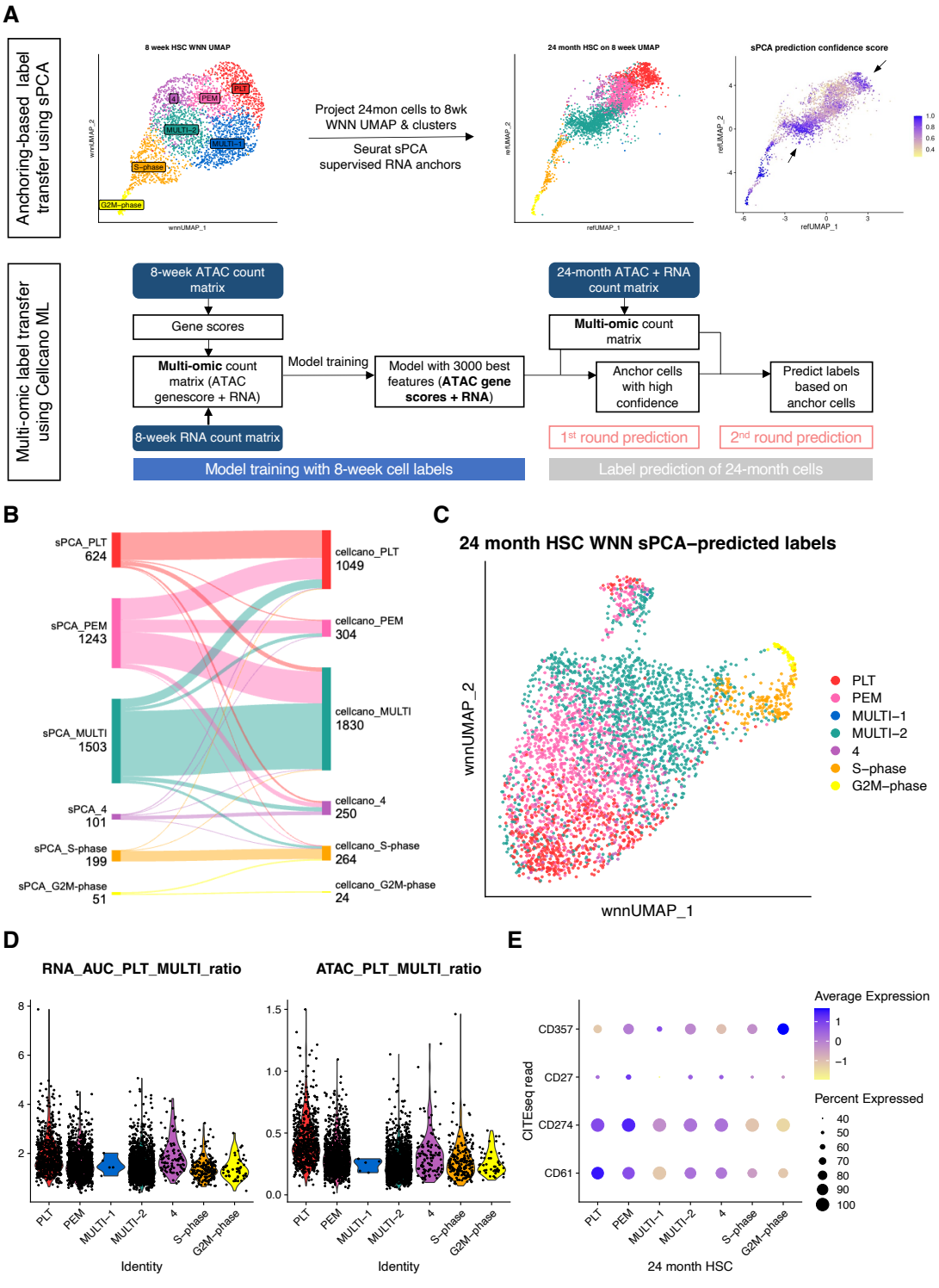


Figure 14 Transferring adult HSC cluster labels to the aged HSC dataset

A, Cluster labels from 8-week adult HSCs were transferred to aged 24-month HSCs using two approaches: supervised PCA (sPCA) projection and modified Cellcano machine learning model (Simiao Zhao). Both methods utilised multi-omic single-cell data for label prediction. The sPCA prediction was most confident in the PLT and MULTI labels. **B,** Comparison of predicted labels between the sPCA and Cellcano methods.

C-E, Predicted labels (sPCA) represent genuine HSC populations with distinct molecular patterns. **C,** An independent WNN UMAP was generated for the 24-month data showing localised distribution of transferred 8-week labels. Lineage potential scoring also confirmed the molecular consistency of these predicted labels. **D,** The sPCA-predicted 24-month PLT and MULTI cells exhibited the highest and lowest platelet-bias score for RNA and ATAC signatures. **E,** The CITEseq count patterns for the four validated markers were consistent in sPCA-predicted 24-month cell populations.

7.3 Investigation of ageing features

To investigate the molecular changes underlying HSC ageing, we compared the gene expression and chromatin accessibility profiles of 8-week adult and 24-month aged HSCs. At the compartment level, aged HSCs exhibited a general downregulation of genes related to mRNA processing, alongside a reduction in the size of cycle cell clusters (G2-M and S phase) from 18% at 8 weeks to 7% at 24 months. These findings suggest a general trend toward increased quiescence in the aged HSCs. In contrast, genes involve in immune response pathways were significantly upregulated, with notable increases in *Cd74*, *H2-Eb1*, *Ifitm3* and *Ifitm2* expressions (Figure 15B-C) (M.Hosseini et al. 2024).

Within specific HSC subpopulations, we observed distinct patterns between PLT and MULTI-HSCs. The significant reduction in mRNA processing pathways was found to be exclusive to PLT-HSCs, suggesting a decrease in cellular activity in this population during ageing. In contrast, MULTI-HSCs did not show differential expression in mRNA pathways, while displayed stronger enrichment for interferon-gamma response pathways, including genes such as *Ifitm2*, *Ifitm3* and *Cd74* (Tanese et al. 2015) (Figure 15D-E).

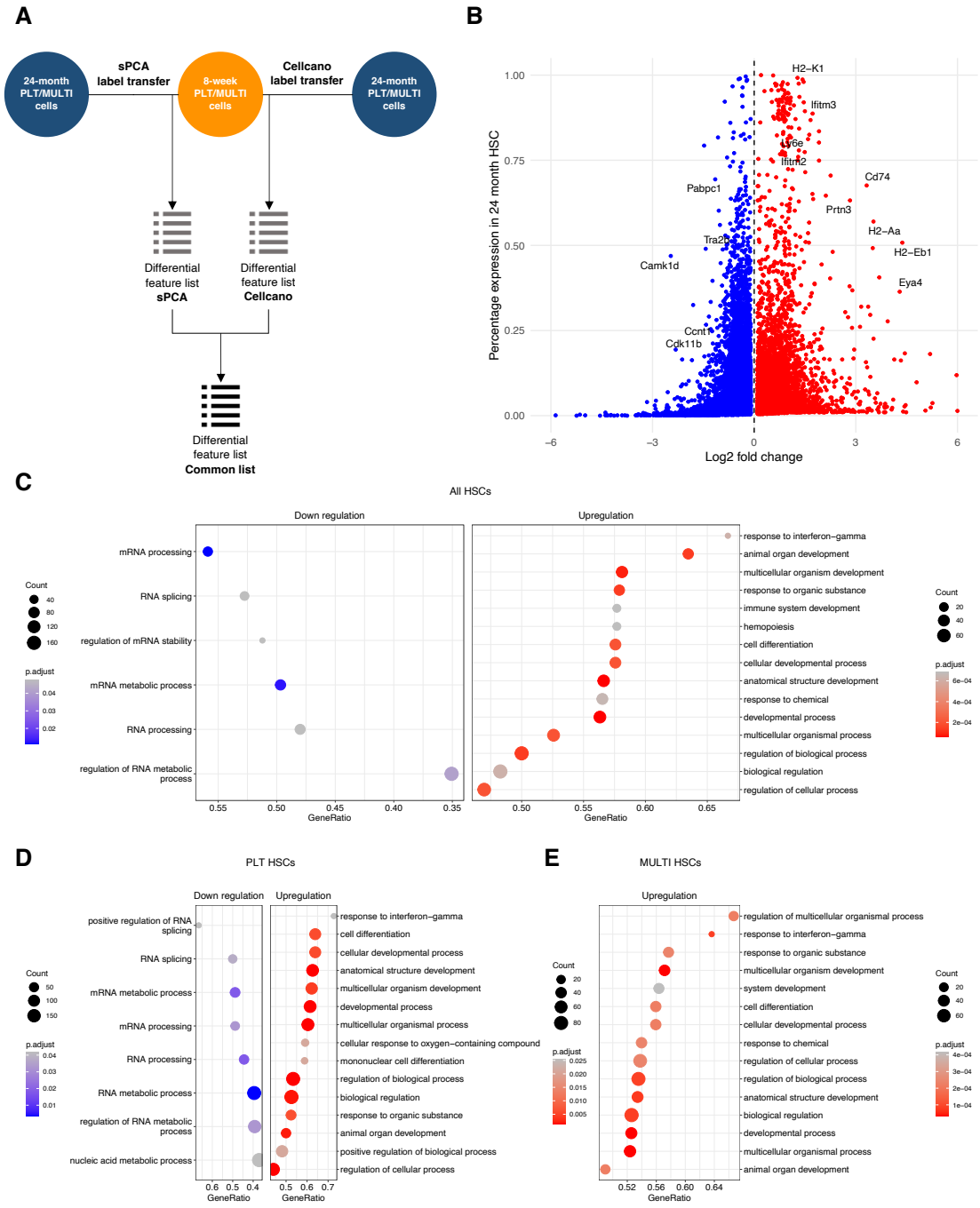


Figure 15 HSC ageing features in RNA modality

A, To investigate ageing features within specific subpopulations, differential RNA and ATACseq analyses were performed using the subpopulation labels predicted by sPCA and Cellcano (Figure 14). A common list of intersecting features was generated from these results. The fold change and percentage expression data of subsequent figures were calculated using the sPCA-predicted labels. **B,** Differential gene expression analysis of unfractionated HSCs identified ageing-associated changes. **C-E,** Pathway enrichment analysis of differentially expressed genes ($p < 0.05$) associated with ageing in **C,** All HSCs; **D,** PLT-HSCs and **E,** MULTI-HSCs.

Analysis of chromatin accessibility provided further insights into the mechanism of HSC ageing. Among the most upregulated accessible chromatin regions during ageing was *Hdac6*, a gene implicated in ageing processes (Chambers et al. 2007; Wang, Wang & Liu 2020). Consistently, RNA expression of *Hdac6* was also elevated in aged HSCs, reinforcing its significance (Figure 16A). On the other hand, chromatin accessibility for regions of the long non-coding RNA (lncRNA) *Gm26917* showed the most significant reductions during ageing (Figure 16B). *Gm26917* has been reported to regulate inflammatory response in macrophages through the NF-κB signalling pathway (Zhao et al. 2022), and reduced NF-κB signalling is known to impair HSC responses to stress and injury . (González-Murillo et al. 2015). At the subpopulation level, *Gm26917* showed the highest chromatin accessibility in the MULTI-2 cluster of 8-week HSCs and also exhibited the most pronounced reduction in this population (Figure 16C), with a Log2 fold change of 1.15 compared to 0.7 in PLT- HSCs and 0.9 at the compartment level.

Motif analysis provided further evidence of regulatory changes during HSC ageing. Across the HSC compartment, there was a significant reduction in the CTCF motif accessibility, a regulator known to restrain LT-HSCs from transitioning to short-term HSCs (Takayama et al. 2021). At the subpopulation level, inflammatory motifs accessibility, including IRF4, IRF8, IRF2 and IRF7, were found to be exclusively increased (plus IRF9 at a higher fold change compared to PLT-HSCs) in MULTI-HSCs during ageing (Figure 16D-E),

suggesting an accumulation of inflammatory activities within the MULTI subpopulation.

Finally, we conducted mitochondrial DNA mutation genotyping to investigate potential clonal haematopoiesis events. This analysis identified 20 mutations that were present in more than five cells (Supplementary Figure 3A). However, none of these mutations displayed a localised distribution on the WNN UMAP, suggesting that the cells carrying these mutations do not share a similar molecular identity, hence, are not clonally related.

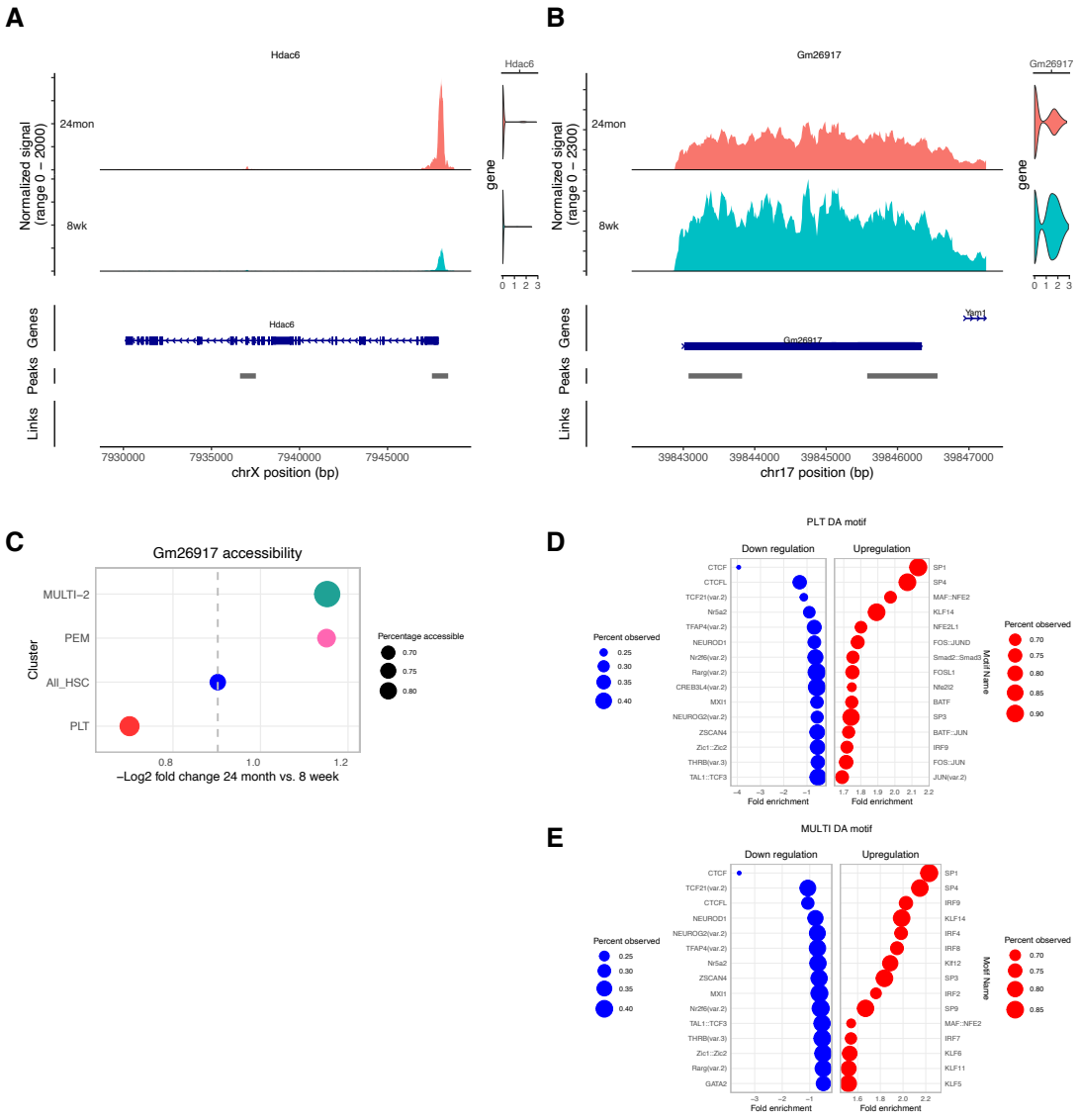


Figure 16 HSC ageing features in ATAC modality

A, *Hdac6* was identified as the most upregulated gene in chromatin accessibility in unfractionated HSCs, with elevated expression levels. **B,** *Gm26917* was the most downregulated gene in chromatin accessibility, showing reduced expression levels in unfractionated HSCs during ageing. **C,** *Gm26917* exhibited the greatest reduction in chromatin accessibility specifically in multilineage HSCs with ageing. **D-E,** Motif differential accessibility analysis in **D,** PLT-HSCs and **E,** MULTI-HSCs, revealing exclusive increase of interferon regulatory factor (IRF) motifs accessibility in MULTI-HSCs.

7.4 Discussion

This chapter examined the molecular characteristics of HSCs during ageing by transferring subpopulation labels from 8-week adult HSCs to 24-month aged HSCs and analysing their gene expression and chromatin accessibility profiles. Two complementary label transfer methods, sPCA and Cellcano, were employed, both of which successfully mapped aged HSCs onto adult subpopulation labels with strong concordance, particularly for PLT and MULT-HSCs. Discrepancies between the two methods were primarily observed in the classification of intermediate PEM cells, which were associated with lower prediction confidence.

Cellcano, leveraging equal integration of RNA and ATACseq modalities, assigned more sPCA PEM cells to either PLT or MULTI labels. This reassignment likely reflects the stronger lineage potential signals present in epigenomic data, enabling Cellcano to better distinguish subtle differences between intermediate populations. In contrast, sPCA, which uses RNA features guided by the WNN graph to align datasets, generated subpopulation ratios (P, PEM, MULTI) more consistent with single-cell transplantation experiments. This suggests that sPCA may better capture lineage potential distributions reflective of functional reconstitution, while Cellcano provides enhanced resolution for resolving less defined or ambiguous states.

Both methods mapped very few cells to cluster 4, indicating that this population largely disappears during ageing. Located between MULTI and lineage-biased clusters, cluster 4 likely represents a less specialised or primed subpopulation. Its relatively low expression of lymphoid (such as *Ctla2a*) and multilineage signatures suggests that its disappearance might not significantly contribute to the adaptive immunity decline associated with ageing. The functional role of cluster 4 remains unclear, but its extinction during ageing underscores the dynamic nature of the HSC compartment. Investigating whether these cells have unique non-lineage-specific roles could be an interesting avenue for future research, though it falls outside the scope of this thesis.

The molecular analysis of ageing HSCs revealed distinct ageing trajectories at the compartment and subpopulation levels. At the compartment level, we observed a general downregulation of mRNA processing genes, coupled with increased expression of genes involved in interferon-gamma response pathways. These findings, alongside the reduction in cycling cells suggest an overall shift toward increased quiescence in aged HSCs. Additionally, chromatin accessibility analysis identified a reduction in CTCF motif accessibility across aged HSCs, which may inhibit HSC activation and further enhance quiescence. The observed upregulation of inflammatory response pathways is consistent with previous studies linking ageing to chronic inflammatory priming and trained immunity (Beerman et al. 2010; De Haan & Lazare 2018a; Valletta et al. 2020). Given that the bone

marrow microenvironment plays a critical role in HSC maintenance (Matteini, Mulaw & Florian 2021), it would be valuable to extend these findings by investigating niche-associated changes that may contribute to the observed ageing phenotypes.

At the subpopulation level, our results indicate that ageing affects PLT and MULTI-HSCs differently. The increase in quiescence was more pronounced in PLT-HSCs, potentially contribute to self-renewal and relative expansion (Li & Bhatia 2011). Conversely, MULTI-HSCs exhibited stronger inflammatory priming, with significant enrichment of interferon-related genes and increased IRF motif accessibility. This differential ageing trajectory suggests that inflammation may apply stronger stress to MULTI-HSCs leading to their progressive loss or functional exhaustion with age (Bogeska et al. 2022). This hypothesis aligns with studies showing that inflammatory challenges can alter HSC function and reduce lymphoid potential. For example, chronic TLR stimulation through repeated low-dose LPS exposure was found to force HSCs into activation, resulting in a shift toward myeloid and a reduction of lymphoid progenitor generation (Esplin et al. 2011). Also, viral infections induce adaptive immune activation have been shown to reduce overall HSC self-renewal capacity while increasing myeloid bias in the long-term (Hirche et al. 2017). These findings support the idea that inflammatory stressors disproportionately impact MULTI-HSCs, which are more likely to respond to immune challenges. The loss of this this population with ageing may therefore result from prolongated inflammatory signalling.

Our analysis also identified molecular targets that may play roles in the ageing process. *Hdac6* emerged as one of the most upregulated genes in chromatin accessibility, which aligns with prior studies demonstrating that histone deacetylase inhibitors can promote HSC self-renewal (Wang et al. 2020). Specifically, HDAC inhibitors have been shown to maintain the primitive phenotype of HSCs in culture by preventing differentiation and preserving self-renewal markers such as Thy-1, c-Kit and Sca-1. These inhibitors also enhance the engraftment potential of human HSPC cells in xenotransplantation models, suggesting a role for histone acetylation in modulating HSC potency. The mechanism underlying this effect likely involves epigenetic reprogramming of genes essential for HSC maintenance.

Additionally, *Gm26917*, a long non-coding RNA that down-regulates NF-κB-mediated inflammatory responses, was the most downregulated accessible region during ageing, particularly in MULTI-HSCs. This reduction in accessibility could contribute to their increased inflammatory states (Bottero, Withoff & Verma 2006). Supporting this, our unpublished multi-omic analysis of Lin- c-Kit+ BM progenitor cells revealed that neutrophil-macrophage progenitors (NMPs) exhibited the highest chromatin accessibility for *Gm26917* compared to other progenitors, such as megakaryocyte progenitors (MkPs). This supports that *Gm26917* plays a role in regulating macrophage activity and its age-related reduction in MULTI-HSCs may contribute to their dysfunction. However,

Gm26917 lacks a conserved human homolog, limiting its clinical relevance. While its role in mouse HSC ageing is intriguing, future studies could focus on identifying human-equivalent regulatory elements that may serve similar functions.

Mitochondrial genotyping did not detect evidence of clonal haematopoiesis in the aged HSC dataset. However, the shallow sequencing depth of mitochondrial DNA fragments in the DOGMAseq workflow limits the sensitivity. Future studies could overcome this limitation by utilising dedicated clonal haematopoiesis mouse models. These models would allow for a more precise investigation of whether specific HSC subpopulations, as defined by the FACS or molecular markers identified in this study, are preferentially expanded during ageing.

Overall, this chapter confirmed increased quiescence and inflammatory priming as key ageing features at the HSC compartment level while revealing that ageing is not uniform across subpopulations. Using a robust label transfer pipeline, we successfully mapped aged HSCs onto the 8-week reference dataset, providing a framework for studying subpopulation dynamics across multi-omic datasets. Molecular analyses showed that MULTI-HSCs exhibited stronger inflammatory burden and signs of functional decline, whereas PLT-HSCs remained more quiescent. We also identified potential targets for modulating HSC ageing, such as Hdac6, which could influence self-renewal and differentiation potential. Given the distinct ageing trajectories observed among HSC

subpopulations, their dynamics should serve as an important readout in future studies aiming to rejuvenate aged HSC compartments. Expanding this work to investigate the bone marrow microenvironment and experimentally targeting Hdac6 or inflammatory pathways could provide impotent insights into reversing age-associated haematopoietic decline.

Chapter 8: Evidence of postnatal HSC diversification

8.1 Introduction and aims

HSCs undergo a critical transition during early postnatal development, moving from a homogeneous population with uniform multilineage potential in the foetal liver (FL) to a diverse compartment of lineage-biased subpopulations in the bone marrow (BM). However, the mechanisms driving this diversification remain poorly understood. Mouse model revealed that by eight weeks postnatally, foetal-like transcriptional programmes are fully replaced by adult HSC characteristics, marking the establishment of a functionally mature HSC compartment. Investigating this transition provide important understanding of how HSCs acquire lineage bias and functional specialisation, that are essential for adult haematopoiesis.

To study this transition, we used single-cell multi-omic data to examine HSC populations from the FL, 4 and 8-week BM. Previous studies have shown that approximately half of the 4-week BM HSCs still express the foetal marker *Sox17*, suggesting the presence of a subset with strong FL characteristics (Kim et al. 2007). We hypothesised that these cells represent a developmental bridge, reflecting a transitional state toward adult HSC. By

integrating sPCA prediction confidence scores with FL and adult BM signatures genes, we aimed to identify these intermediate HSCs that might maintain FL features while transitioning into adult states.

To trace the developmental trajectories of 4-week HSCs, we applied pseudotime analysis using Slingshot (Street et al. 2018), a tool suited for multi-lineage datasets. Slingshot offers flexibility in handling complex, unsupervised branching events without predefined lineage paths, making it ideal for the nature of HSC diversification. However, it is important to note that the WNN UMAP is constructed using both RNA and ATACseq data, whereas Slingshot computes pseudotime solely based on RNAseq data.

Through these analyses, we aimed to investigate whether a transitional programme exists at four weeks and to determine if it is possible to characterise the group of cells that bridges foetal to adult HSC states. If successful, this could provide insights into how HSC diversification is regulated, shedding light on the molecular and chromatin remodelling events underlying this critical developmental window.

8.2 Identification of chromatin remodelling activities in 4-week HSCs

To investigate foetal to adult HSC transition, we generated single-cell multi-omic libraries for E14.5 FL HSCs and 4-week BM HSCs and performed unsupervised WNN clustering

on the datasets.

In the FL dataset, we observed no significant deviations in either P or MULTI signatures across chromatin accessibility and RNA expression modalities (Supplementary Figure 4), suggesting a unitary lineage potential in FL HSCs. In contrast, the 4-week dataset revealed distinct PLT and MULTI clusters indicating the emergence of lineage-biased subpopulations (Figure 17A). Notably, we identified a unique cluster in the 4-week dataset, later named 'Remodelling (RML)' HSCs, which exhibited high expression of both P and MULTI signatures, a feature not observed at any other developmental stage we tested (Figure 17B). Unlike other clusters, RML cells displayed chromatin accessibility at both PLT and MULTI-associated regulatory regions within the same cells, accompanied by an overall increase in ATACseq fragment count per cell (Figure 17D). This suggests a dynamic chromatin landscape, potentially reflecting a state of increased plasticity. Additionally, these cells expressed high levels of foetal-specific genes such as *Lin28b* (Lee et al. 2013).

To further investigate the identity of RML cells, we computed differentially expressed gene signatures between FL and 8-week adult HSC datasets and calculated their AUCell enrichment scores for 4-week HSCs. Surprisingly, RML cells scored the highest for both FL and adult signatures. Mapping 4-week HSCs onto the FL and adult datasets using sPCA revealed that RML cells had stronger alignment to the FL dataset (Figure 17C), suggesting that they may resemble a foetal HSC state while beginning to express transcriptional

programmes characteristic of adult HSCs. Motif analysis identified ZBTB33 as the most significantly increased accessible motif in the RML cluster and the adjacent 'Transition' cluster (Figure 17A), suggesting increased activity of this zinc finger in these populations.

We also performed pseudotime analysis using Slingshot, which inferred a trajectory initiating from the RML cluster, passing through a central 'Transition' cluster, and branching into paths toward PLT, MULTI, and cycling clusters (Figure 17F).

To validate these findings and ensure they were not artifacts or outliers, particularly considering that the 4-week dataset was generated from BM pooled from 12 mice, we conducted additional analyses. Mitochondrial mutation genotyping, revealed no mutations exclusive to the RML cluster despite the shallow sequencing depth (Supplementary Figure 3B), supporting the integrity of the data. Additional cell cycle phase analysis showed that RML cells were not confined to any specific phase of the cell cycle, indicating that their unique features were not influenced by proliferative states or cell cycle phases (Figure 17E). These validations confirm that the RML cluster represents a genuine and biologically relevant subpopulation rather than an experimental artifact or outlier.

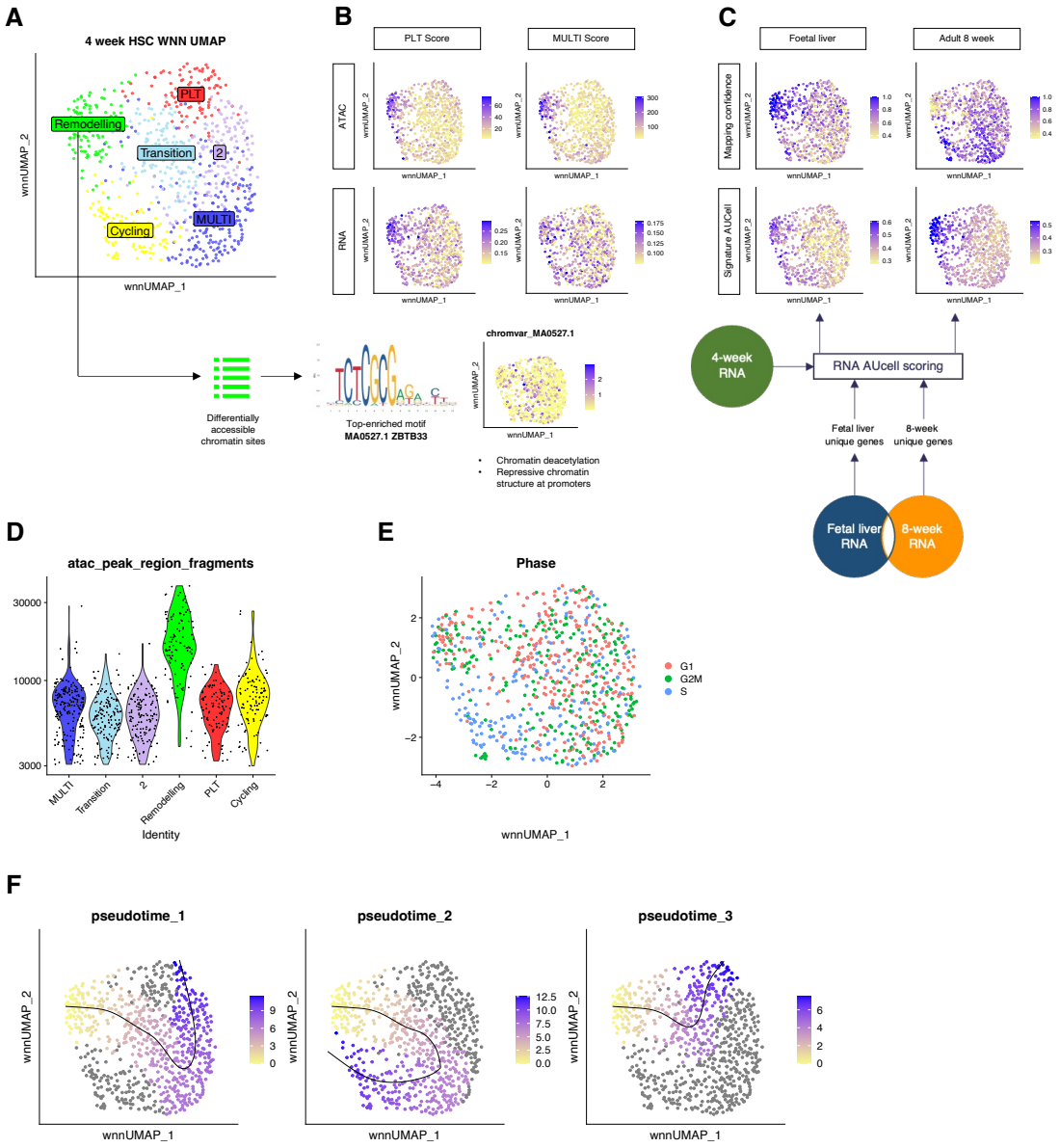


Figure 17 Chromatin remodelling activities in 4-week HSCs

A, WNN UMAP reduction of 4-week HSCs. The PLT and MULTI clusters were annotated by their lineage potential scoring described in (Figure 5A). Differential analysis revealed ZBTB33 as the most increased accessible motif in the 'Remodelling (RML)' and 'Transition' cluster. **B,** Lineage potential scoring also revealed the RML cluster co-expressing both platelet-bias and multilineage signatures **C,** RML HSCs exhibited higher sPCA mapping confidence to the foetal liver dataset while co-expressing both foetal and adult HSC markers, suggesting a unique cell population that retains foetal liver characteristics while initiating transcriptional programmes associated with the heterogeneous subpopulations of adult HSCs. **D,** RML cluster is associated with higher detectable ATACseq fragment count. **E,** The clustering of RML cells is not associated with cell cycle states. **F,** Pseudotime analysis showing RML HSCs diversify into distinct clusters.

8.3 Discussion

This chapter provides insights into the transitional states of HSCs during early postnatal development, focusing on the discovery of the 'Remodelling' (RML) cluster in 4-week HSCs. These cells displayed unique molecular patterns associated with both foetal and adult-specific signatures, reflecting a potential intermediate state in the diversification of young HSCs. However, several points warrant further discussion and investigation.

Although *Sox17* is a well-established foetal HSC marker and expected to be expressed in a subset of 4-week HSCs, our sequencing data did not detect any *Sox17* expression. This could be attributed to technical limitations, as *Sox17* expression may fall below the detection threshold of 3' RNA-sequencing. Future studies could address this limitation by employing the *Sox17*-eGFP reporter mouse line to specifically isolate and sequence *Sox17*+ HSCs at this developmental stage, providing direct molecular characterisation of this transitional subpopulation.

A key observation was the extensive chromatin accessibility of P and MULTI regions within the RML HSCs. This accessibility pattern was not observed in either foetal or adult HSCs and appeared independent of cell cycle phase. However, our sequencing data did not identify a unique transcriptional programme initiating this phenomenon. This could be due to the inherent limitations of 3' RNAseq, which may fail to capture low-abundance

transcripts or regulatory RNAs. Alternatively, the initiation of this chromatin remodelling events may occur at an earlier timepoint. To test this hypothesis, future studies could profile HSCs at two weeks postnatally to capture earlier stages of chromatin and transcriptional transitions.

Although the pseudotime and lineage analyses suggested the RML cluster might be upstream of the clusters with polarised lineage potentials, we note several caveats in its interpretation. The method assumes a hierarchical or progressive relationship between clusters within the dataset, which may not reflect true biological lineage progression. In reality, the populations observed may arise in parallel prior to sample capture, originate from different precursors, or descend from unprofiled cells in the same animal. Therefore, while the inferred trajectory may help visualising transcriptional transitions, it should be interpreted with caution and not taken as definitive evidence of lineage progression or developmental ordering.

Importantly, mature HSC subpopulations do not switch lineage potential, as demonstrated by secondary transplantations, particularly not from a restricted state back to a multilineage. This is further supported by our 24-month HSC dataset, which revealed a loss of intermediate subpopulations and an increasing molecular polarisation of HSC subpopulations. Identifying the initiator of this chromatin remodelling event may provide an opportunity to reprogramme existing HSCs, rather than relying on expanding the

remaining HSCs, which may have already accumulated stress and mutations over time.

It is also important to note that our comparisons involved foetal liver HSCs but did not consider foetal BM HSCs, which may represent an alternative source of HSCs. Since most evidence suggests that foetal BM HSCs contribute minimally to the adult HSC compartment due to their limited self-renewal capacity (Hall et al. 2022), it is less likely that the RML cluster are derived from foetal BM HSCs.

To further validate the findings, single-cell transplantation experiments with RML cells would be a possible next step. We hypothesise that these cells can give rise to both lineage-biased and MULTI-HSCs, and this can be tested by monitoring donor-derived HSCs using the FACS markers and Fluidigm signatures identified in earlier chapters. Such transplantation experiments would provide direct functional evidence for the transitional role of RML cells and offer insight into their contribution to lineage potential diversification.

Motif analysis identified ZBTB33 as a potential regulatory factor in RML cells. However, this zinc finger protein is involved in a wide range of cellular activities, including transcriptional repression, chromatin organisation and methylation-mediated gene silencing (Yoon et al. 2003; Blattler et al. 2013). Its precise role in HSC chromatin remodelling and diversification remains unclear. Given its systemic effects on multiple

organs, *in vivo* knockout or inhibition of ZBTB33 in a mouse model would likely result in broad physiological consequences (Kulikov et al. 2016). A more feasible approach to initially investigate its role in HSCs would be to inhibit or knock down ZBTB33 in an *ex vivo* culturing system.

Overall, we hypothesise that lineage potential heterogeneity in HSCs emerges to meet the increasingly variable physiological demands encountered after birth. As HSCs transition from the foetal liver to the bone marrow niche, or in response to cues during birth and early development, they may undergo a global chromatin opening phase that permits accessibility to a broad set of lineage-specific regulatory elements. This opening state may subsequently resolve through selective chromatin closure, conferring lineage priming that results in the heterogeneous reconstitution pattern found in adult HSCs. Whether the decision of an HSC to acquire a specific lineage-biased phenotype is predetermined earlier in development or arises stochastically at the single-cell level remains unknown.

Conclusion

This thesis has explored the heterogeneity of haematopoietic stem cells (HSCs) using single-cell multi-omic sequencing, revealing distinct lineage-biased subpopulations and their functional implications in haematopoiesis. By integrating transcriptomic, epigenomic and proteomic data, this study provided a comprehensive molecular characterisation of HSC subpopulations, identifying novel regulatory mechanisms underlying lineage bias and ageing. These findings contribute to the broader understanding of how stem cell function is regulated and how perturbations in these mechanisms can lead to age-related dysfunction or disease.

One of the primary contributions of this work is the development of a refined fluorescence-activated cell sorting (FACS) panel, which effectively enriches lineage-biased HSC subpopulations. This panel was constructed based on *in silico* marker selection, followed by rigorous experimental validations, ensuring the accurate isolation of HSC subpopulations with distinct lineage potentials. The implementation of this FACS panel in transplantation assays provided functional confirmation of the predicted lineage potentials, reinforcing the notion that HSC subpopulations are molecularly and functionally distinct. This methodological advance paves the way for future studies seeking to investigate

lineage-biased HSCs with greater precision.

This thesis has also contributed significantly to the development of methodologies that enhance our ability to study HSC biology at high resolution. The optimisation of multi-omic sequencing for mouse HSC samples has enabled the simultaneous profiling of gene expression, chromatin accessibility and proteomic data, offering an unprecedented view of stem cell biology. The application of unsupervised clustering methods has facilitated the identification of distinct subpopulations, while the development of a robust lineage potential scoring approach has provided a quantitative framework for predicting functional heterogeneity. Additionally, *in silico* gating simulations have proven to be an effective strategy for refining FACS marker selection, streamlining the isolation of lineage-biased HSCs. A key advance is the establishment of a robust label transfer pipeline for mapping HSC subpopulations across multi-omic datasets. This pipeline enables the systematic comparison of datasets from different developmental and ageing stages, ensuring that future experiments can be integrated into a comprehensive multi-omic reference atlas of mouse HSCs. These advancements lay a strong foundation for future studies aiming to further dissect HSC biology, providing powerful analytical tools that will facilitate deeper insights into haematopoiesis and stem cell ageing.

In addition to lineage bias, this thesis examined the effects of ageing on HSCs, uncovering distinct age-related changes across different subpopulations. Ageing is marked by a shift in

clonal dynamics, with a notable expansion of platelet-biased HSCs and a decline in multilineage HSCs. Single-cell multi-omic analysis revealed that aged platelet-biased HSCs not only increase in number but also acquire a more quiescent state, potentially contributing to their enhanced persistence over time. In contrast, multilineage HSCs exhibit heightened inflammatory signatures, accumulating transcriptomic and epigenomic markers of stress and activation. This inflammatory burden may accelerate their depletion, further reducing the lymphopoietic potential of aged individuals. These findings highlight the divergent ageing trajectories of HSC subpopulations and provides potential targets for interventions aimed at restoring the functional balance of HSCs in aged individuals.

Beyond these primary findings, this thesis also uncovered additional molecular insights with potential therapeutic implications. The identification of Hdac6 as a targetable epigenetic regulator suggests that modulating histone deacetylation could influence HSC functional viability. Furthermore, the discovery of chromatin remodelling events in young HSCs highlights a dynamic epigenetic landscape that may play a crucial role in the transition from foetal to adult haematopoiesis. These findings open new avenues for exploring epigenetic interventions aimed at restoring HSC functionality, either by reversing age-related changes or by enhancing the regenerative capacity of HSCs in therapeutic contexts.

Despite these advances, several limitations remain. One challenge in single-cell sequencing

is its inherent technical noise and low transcript capture efficiency, which may obscure subtle transcriptional differences between populations. Additionally, while *in silico* gating provides an efficient strategy for marker selection, its functional relevance must be continuously validated through experimental approaches such as transplantation and lineage-tracing studies. Another limitation is the reliance on mouse models, which, while highly informative, may not fully recapitulate human haematopoiesis. Future studies should aim to validate these findings in human HSCs, using patient-derived samples and xenotransplantation models.

To further enhance our understanding of HSC heterogeneity and ageing, future research should leverage newer technologies such as spatial transcriptomics and advanced single-cell epigenomics. Spatial transcriptomics, for instance, would allow researchers to map HSC subpopulations within their native bone marrow microenvironment, providing insights into how niche interactions influence lineage commitment and ageing. Similarly, advanced single-cell epigenomic approaches with higher resolution, such as CUT&Tag (Kaya-Okur et al. 2019) could provide more detailed characterisations of chromatin dynamics in HSCs. Integrating these methods with functional assays, such as *in vivo* lineage tracing and competitive transplantation, will further elucidate the mechanisms underlying lineage bias and age-related haematopoietic decline.

Beyond fundamental research, the findings of this thesis have clinical implications. The

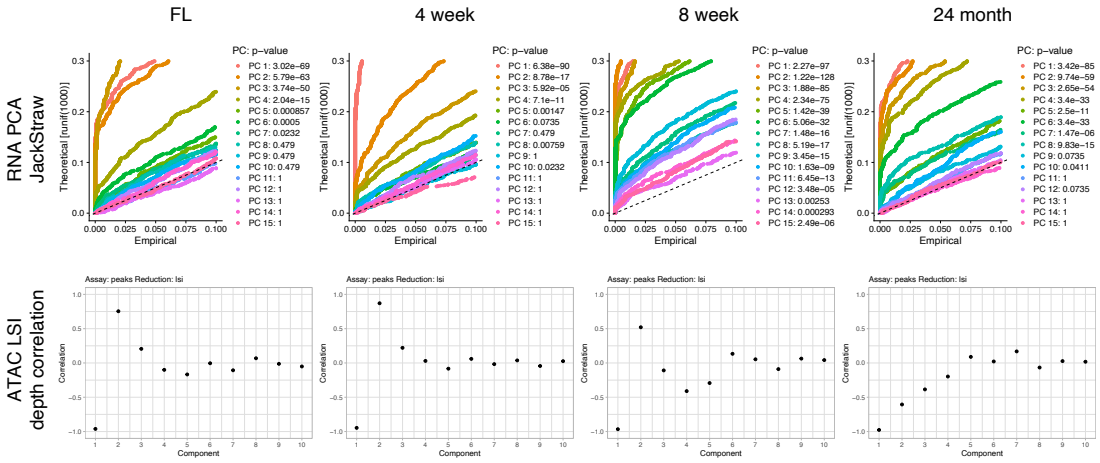
ability to isolate functionally distinct HSC subpopulations with the refined FACS panel opens new possibilities for improving stem cell transplantation therapies. By selecting the most potent HSC populations for transplantation, clinicians may be able to enhance engraftment efficiency and long-term haematopoietic reconstitution while reduce the risk of donor-engrafted clonal haematopoiesis (Gibson et al. 2022) in patients undergoing bone marrow transplants. Additionally, understanding the molecular basis of HSC ageing provides a foundation for developing targeted therapies to counteract age-related haematopoietic decline.

Furthermore, the characterisation of lineage-biased HSCs has implications for cancer biology. Given that lineage-biased HSCs exhibit superior self-renewal capacities, it is plausible that certain HSCs may be more susceptible to leukemic transformation. Investigating the oncogenic potential of different HSC subpopulations could provide insights into the cellular origins of haematological malignancies. Future research should explore whether specific epigenetic alterations in lineage-biased HSCs predispose them to leukemic transformation and whether targeting these alterations could provide novel therapeutic strategies for blood cancers.

In conclusion, this thesis has provided a comprehensive characterisation of HSC heterogeneity, lineage bias and ageing using state-of-the-art single-cell multi-omic technologies. By integrating computational analysis with experimental validation, this work

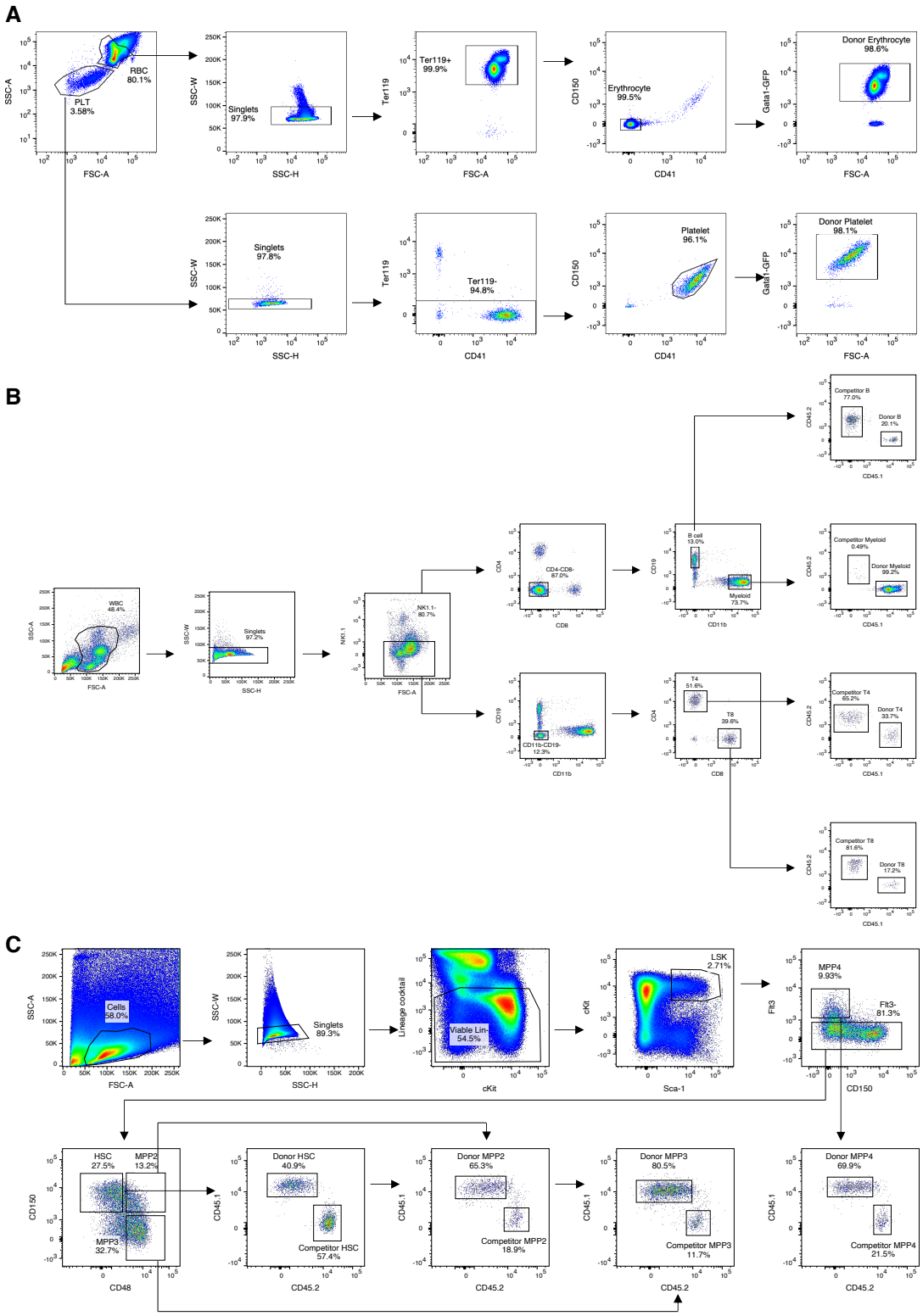
advances our understanding of the molecular underpinnings of HSC function and ageing. The findings have broad implications for both basic stem cell biology and translational medicine, offering new perspectives on stem cell regulation and potential therapeutic strategies for haematopoietic ageing and disease. Future research should continue to build upon these discoveries, leveraging emerging technologies and interdisciplinary approaches to further unravel the complexity of HSC biology and its implications for health and disease.

Supplementary Figures



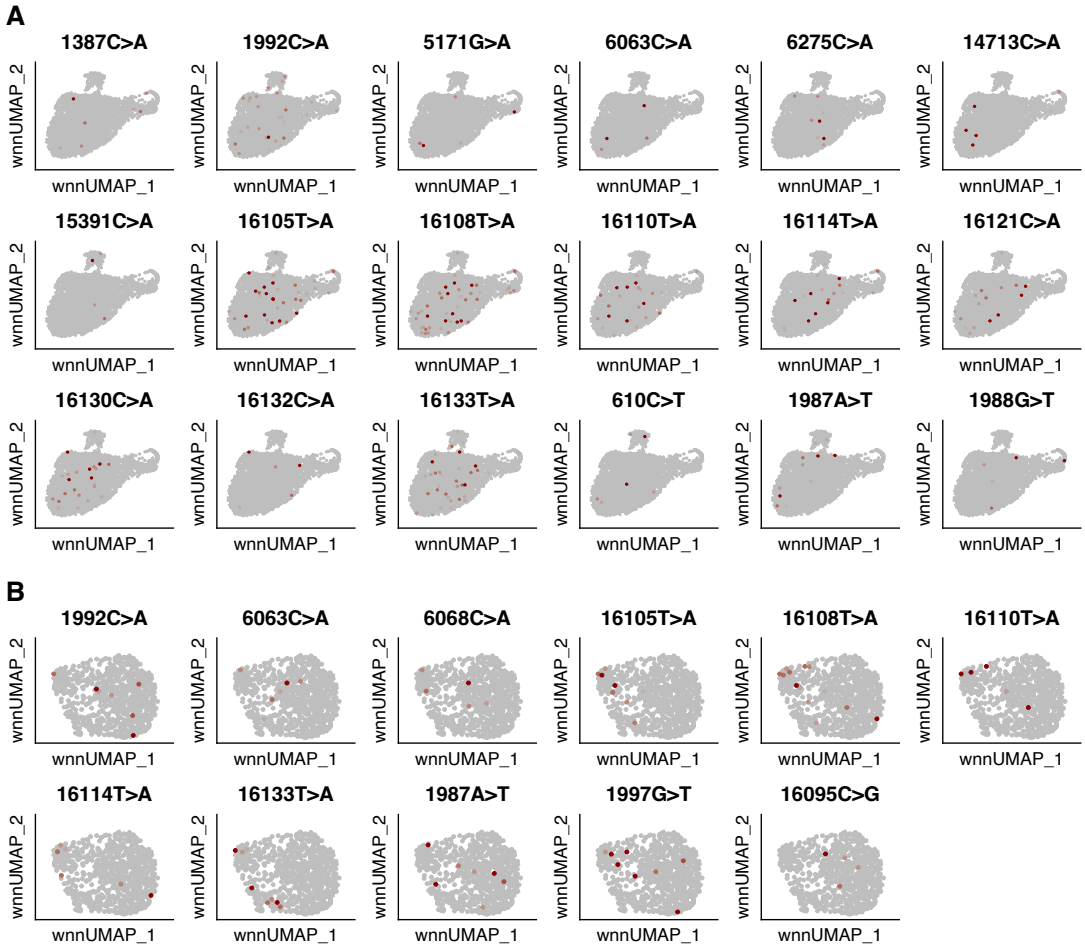
Supplementary Figure 1 Dimensionality reduction components

Dimensionality reduction components used for WNN clustering were selected based on Jack Straw scores for the RNA modality and depth correlation of LSI components for the ATAC modality.



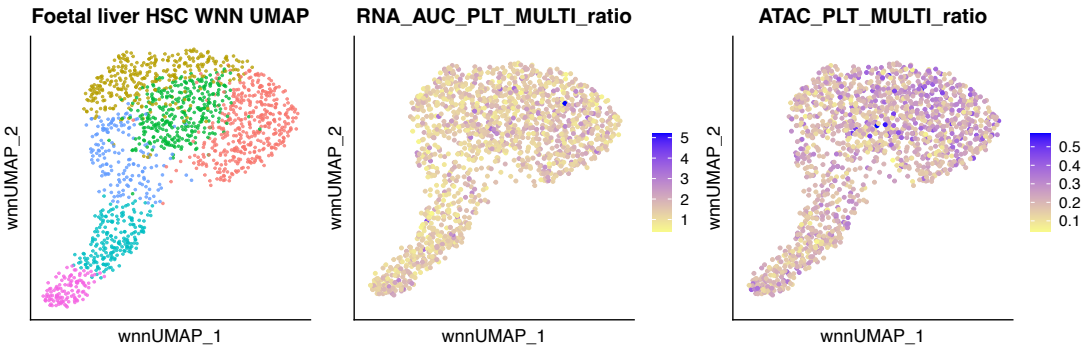
Supplementary Figure 2 FACS panel for transplantation reconstitution readouts

CD45.1 *Gata1*-eGFP HSCs alongside 200,000 CD45.2 WT bone marrow cells were transplanted into lethally irradiated CD45.2 WT recipients. **A-B,** Peripheral blood chimerism readouts were taken at week 6, 10, 16 with week 16 regarded as steady-state reconstitution. **A,** Platelet and erythrocyte; **B,** Myeloid, B, T4 and T8 cells. **C,** Bone marrow endpoint readouts for HSC and multipotent progenitor (MPP 2-4) chimerism analyses.



Supplementary Figure 3 Mitochondrial genotyping

Mitochondrial DNA mutation genotyping was used to identify **A,** Potential clonal haematopoiesis in the 24-month and **B,** Outlier mice in the 4-week HSC datasets. No significant clonal relationship has been identified despite the sparse and low-abundant results.



Supplementary Figure 4 Fetal liver HSC WNN UMAP and lineage potential scoring

Lineage potential scores of foetal liver HSCs were not enriched in specific subsets, supporting the molecular and functional homogeneity of foetal liver HSCs

Supplementary Tables

Population	Sample	Marker
LSK	BM	Lin- Sca1+ c-Kit+
HSC	BM&FL	LSK CD48- CD34- CD150+
PLT-HSC	BM	HSC CD274hi CD61hi CD27low CD357low
MULTI-HSC	BM	HSC CD274low CD61low CD27hi CD357hi
MPP2	BM	LSK Flt3- CD48+ CD150+
MPP3	BM	LSK Flt3- CD48+ CD150-
MPP4	BM	LSK Flt3+ CD150-
Platelet	PB	Ter119- CD41+ CD150+
Erythrocyte	PB	Ter119+ CD41- CD150-
Myeloid cell	PB	NK1.1- CD4- CD8- Mac1+ CD19-
T4 cell	PB	NK1.1- CD4+ CD8- Mac1- CD19-
T8 cell	PB	NK1.1- CD4- CD8+ Mac1- CD19-
B cell	PB	NK1.1- CD4- CD8- Mac1- CD19+
Donor-derived HSC	BM	LSK CD48- CD150+ CD45.1+ CD45.2-
Donor-derived WBC	PB&BM	CD45.1+ CD45.2-
Donor-derived platelet & erythrocyte	PB	GFP+

Supplementary Table 1 Gating strategies in FACS analyses

Target	Panel	Fluorophore	Titration	Manufacturer	Catalogue
CD4	Lin cocktail	PECy5	400	eBioscience	15-0042-82
B220	Lin cocktail	PECy5	400	Biolegend	103210
Ter119	Lin cocktail	PECy5	600	eBioscience	15-5921-83
Gr1	Lin cocktail	PECy5	800	eBioscience	15-5931-02
CD5	Lin cocktail	PECy5	800	Biolegend	100610
Mac1	Lin cocktail	PECy5	800	eBioscience	15-0112-82
CD8a	Lin cocktail	PECy5	1200	eBioscience	15-0081-83
Live/dead	Viability	7AAD (PECy5)	0.525 μg/μl	Biolegend	420403
Sca1	HSC	PECy7	400	eBioscience	25-5981-87
c-Kit	HSC	APCe780	1600	eBioscience	47-1171-82
CD150	HSC	BV785	200	Biolegend	115937
CD48	HSC	APC	600	eBioscience	17-0481-82
Flt3 (CD135)	HSC	BV421	25	Biolegend	135313
CD34	HSC	AF700	10	eBioscience	56-0341-82
Rat isotype	HSC	PE	400	eBioscience	12-4724-82
CD45.1	Host/Donor	BV650	25	Biolegend	110736
CD45.2	Host/Donor	AF700	100	eBioscience	56-0454-82
Sca1	HSC subtype	BV421	200	Biolegend	108128
CD48	HSC subtype	BV510	600	Biolegend	103443
CD274	HSC subtype	BV605	100	Biolegend	153606
CD61	HSC subtype	PE	400	Biolegend	104307
CD357	HSC subtype	PECy7	100	Biolegend	126318
CD27	HSC subtype	APC	200	Biolegend	124212
Mac1	WBC	PECy7	1600	eBioscience	25-0112-81
CD19	WBC	eF450	200	eBioscience	48-0193-80
CD4	WBC	APC	1200	eBioscience	17-0042-82
CD8a	WBC	APCe780	400	eBioscience	47-0081-82
NK1.1	WBC	PECy5	400	Biolegend	108716
CD4	PLT	BV421	100	Biolegend	133911
Ter119	PLT	APC	400	eBioscience	12-0452-82

Supplementary Table 2 Antibodies used in FACS analyses

Buffer	Reagent	Concentration
FACS media	PBS	1X
	FCS	5%
	EDTA	2 mM
PB media	PBS	1X
	FCS	1%
	EDTA	2 mM
DIG permeabilisation	Tris-HCl pH 7.5	20 mM
	NaCl	150 mM
	MgCl2	3 mM
	Digitonin	0.01%
	RNase inhibitor (Sigma, PN-3335399001)	2 U/µl
DIG wash	Tris-HCl pH 7.5	20 mM
	NaCl	150 mM
	MgCl2	3 mM
	RNase inhibitor (Sigma, PN-3335399001)	1 U/µl
Transplantation injection media	IMDM	1X
	Pen/Strep	1X
	L-Glutamine	1X
	2-Mercaptoethanol	0.1 mM
	BIT9500	20%
Fluidigm RT	Cell Direct 2X Reaction Mix (ThermoFisher, 11754)	1X
	SuperScript® III RT/Platinum® Taq Mix (ThermoFisher, 11754)	12%
	SUPERase·In™ RNase Inhibitor (Ambion, AM2694)	0.2 U/µl
	TE buffer	12%
	Taqman Assay	0.2X
Bulk RNA lysis	Triton X-100	0.18%
	RNAse inhibitor	1 U/µl
	dNTPs	2.5 mM
	Oligo(dT)	2.5 µM
ATAC lysis	TD buffer (Illumina, 20034197)	1X
	Digitonin	0.01%
	Tween-20	0.1%
	PBS	33%
	Water	10%

Supplementary Table 3 Buffers

Primer	Sequence 5' to 3'
Pre-amp	CCTTGGCACCCGAGAATTCC
SI-PCR	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TC
RPxx (sample index)	CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTC CTTGGCACCC GAGAATTCCA
Ad1noMX	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGA TGTG
Ad2_x	CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTCTCGTGGGCTC GGAGATGT
ADT	/5Biosg/CCTTGGCACCGAGAATCCAXXXXXXXXXXXXXXXGAAAAAA AAAAAAAAAAAAAAAAAAAAA*A*A

Supplementary Table 4 Primers

ADT oligos were biotinylated at 5' for streptavidin conjugation and the * in 3' denotes phosphorothioate bond that prevents nuclease degradation.

Bibliography

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.

Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., Oord, J. Van Den, Atak, Z.K., Wouters, J. & Aerts, S., 2017, 'SCENIC: Single-cell regulatory network inference and clustering', *Nature Methods*, 14(11), 1083–1086.

Akashi, K., Traver, D., Miyamoto, T. & Weissman, I.L., 2000, 'A clonogenic common myeloid progenitor that gives rise to all myeloid lineages', *Nature 2000 404:6774*, 404(6774), 193–197.

Aksöz, M., Gafencu, G.-A., Stoilova, B., Buono, M., Zhang, Y., Turkalj, S., Meng, Y., Asger Jakobsen, N., Metzner, M., Clark, S.-A., Beveridge, R., Thongjuea, S., Vyas, P. & Nerlov, C., 2024, *Hematopoietic stem cell heterogeneity and age-associated*

platelet bias are evolutionarily conserved, vol. 9.

Amemiya, H.M., Kundaje, A. & Boyle, A.P., 2019, 'The ENCODE Blacklist: Identification of Problematic Regions of the Genome', *Scientific Reports 2019 9:1*, 9(1), 1–5.

Bahl, A., Hellack, B., Balas, M., Dinischiotu, A., Wiemann, M., Brinkmann, J., Luch, A., Renard, B.Y. & Haase, A., 2019, 'Recursive feature elimination in random forest classification supports nanomaterial grouping', *NanoImpact*, 15, 100179.

Balazs, A.B., Fabian, A.J., Esmon, C.T. & Mulligan, R.C., 2006, 'Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow', *Blood*, 107(6), 2317–2321.

Beerman, I., Maloney, W.J., Weissmann, I.L. & Rossi, D.J., 2010, 'Stem cells and the aging hematopoietic system', *Current opinion in immunology*, 22(4), 500.

Belander Strålin, K., Carrelha, J., Winroth, A., Ziegenhain, C., Hagemann-Jensen, M., Kettyle, L.M., Hillen, A., Högstrand, K., Markljung, E., Grasso, F., Seki, M., Mazzi, S., Meng, Y., Wu, B., Chari, E., Lehander, M., Sandberg, R., Woll, P.S. & Jacobsen, S.E.W., 2023, 'Platelet and myeloid lineage biases of transplanted single perinatal mouse hematopoietic stem cells', *Cell Research 2023 33:11*, 33(11), 883–886.

Benz, C., Copley, M.R., Kent, D.G., Wohrer, S., Cortes, A., Aghaeepour, N., Ma, E., Mader, H., Rowe, K., Day, C., Treloar, D., Brinkman, R.R. & Eaves, C.J., 2012, 'Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs', *Cell stem cell*, 10(3), 273–283.

Blattler, A., Yao, L., Wang, Y., Ye, Z., Jin, V.X. & Farnham, P.J., 2013, 'ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes', *Epigenetics and Chromatin*, 6(1), 1–18.

Bogeska, R., Mikecin, A.M., Kaschutnig, P., Fawaz, M., Büchler-Schäff, M., Le, D., Ganuza, M., Vollmer, A., Paffenholz, S. V., Asada, N., Rodriguez-Correa, E., Frauhammer, F., Buettner, F., Ball, M., Knoch, J., Stäble, S., Walter, D., Petri, A., Carreño-Gonzalez, M.J., Wagner, V., Brors, B., Haas, S., Lipka, D.B., Essers, M.A.G., Weru, V., Holland-Letz, T., Mallm, J.P., Rippe, K., Krämer, S., Schlesner, M., McKinney Freeman, S., Florian, M.C., King, K.Y., Frenette, P.S., Rieger, M.A. & Milsom, M.D., 2022, 'Inflammatory exposure drives long-lived impairment of hematopoietic stem cell self-renewal activity and accelerated aging', *Cell Stem Cell*, 29(8), 1273-1284.e8.

Bolger, A.M., Lohse, M. & Usadel, B., 2014, 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), 2114.

Bottero, V., Withoff, S. & Verma, I.M., 2006, 'NF-κB and the regulation of hematopoiesis', *Cell Death & Differentiation 2006 13:5*, 13(5), 785–797.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J., 2013, 'Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics', *Nature methods*, 10(12), 1213.

Buggenum, J.A.G.L. Van, Gerlach, J.P., Eising, S., Schoonen, L., Eijl, R.A.P.M. Van, Tanis, S.E.J., Hogeweg, M., Hubner, N.C., Hest, J.C. Van, Bonger, K.M. & Mulder, K.W., 2016, 'A covalent and cleavable antibody-DNA conjugation strategy for sensitive protein detection via immuno-PCR', *Scientific Reports 2016 6:1*, 6(1), 1–12.

Caiado, F., Pietras, E.M. & Manz, M.G., 2021, 'Inflammation as a regulator of hematopoietic stem cell function in disease, aging, and clonal selection', *Journal of Experimental Medicine*, 218(7).

Carrelha, J., Mazzi, S., Winroth, A., Hagemann-Jensen, M., Ziegenhain, C., Högstrand, K., Seki, M., Brennan, M.S., Lehander, M., Wu, B., Meng, Y., Markljung, E., Norfo, R., Ishida, H., Belander Strålin, K., Grasso, F., Simoglou Karali, C., Aliouat, A., Hillen, A., Chari, E., Siletti, K., Thongjuea, S., Mead, A.J., Linnarsson, S., Nerlov, C., Sandberg, R., Yoshizato, T., Woll, P.S. & Jacobsen,

S.E.W., 2024, 'Alternative platelet differentiation pathways initiated by nonhierarchically related hematopoietic stem cells', *Nature Immunology 2024 25:6*, 25(6), 1007–1019.

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.

Challen, G.A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J.S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., Liang, S., Lu, Y., Darlington, G.J., Meissner, A., Issa, J.P.J., Godley, L.A., Li, W. & Goodell, M.A., 2011, 'Dnmt3a is essential for hematopoietic stem cell differentiation', *Nature genetics*, 44(1), 23–31.

Challen, G.A., Sun, D., Mayle, A., Jeong, M., Luo, M., Rodriguez, B., Mallaney, C., Celik, H., Yang, L., Xia, Z., Cullen, S., Berg, J., Zheng, Y., Darlington, G.J., Li, W. & Goodell, M.A., 2014, 'Dnmt3a and Dnmt3b have Overlapping and Distinct Functions in Hematopoietic Stem Cells', *Cell stem cell*, 15(3), 350.

Chambers, S.M., Shaw, C.A., Gatza, C., Fisk, C.J., Donehower, L.A. & Goodell, M.A., 2007, 'Aging Hematopoietic Stem Cells Decline in Function and Exhibit

Epigenetic Dysregulation', *PLOS Biology*, 5(8), e201.

Chen, C., Liu, Yu, Liu, Yang & Zheng, P., 2009, 'mTOR Regulation and Therapeutic Rejuvenation of Aging Hematopoietic Stem Cells', *Science signaling*, 2(98), ra75.

Chen, X., Deng, H., Churchill, M.J., Luchsinger, L.L., Du, X., Chu, T.H., Friedman, R.A., Middelhoff, M., Ding, H., Tailor, Y.H., Wang, A.L.E., Liu, H., Niu, Z., Wang, H., Jiang, Z., Renders, S., Ho, S.H., Shah, S. V., Tishchenko, P., Chang, W., Swayne, T.C., Munteanu, L., Califano, A., Takahashi, R., Nagar, K.K., Renz, B.W., Worthley, D.L., Westphalen, C.B., Hayakawa, Y., Asfaha, S., Borot, F., Lin, C.S., Snoeck, H.W., Mukherjee, S. & Wang, T.C., 2017, 'Bone marrow myeloid cells regulate myeloid-biased hematopoietic stem cells via a histamine-dependent feedback loop', *Cell stem cell*, 21(6), 747.

Chen, X., Miragaia, R.J., Natarajan, K.N. & Teichmann, S.A., 2018, 'A rapid and robust method for single cell chromatin accessibility profiling', *Nature Communications 2018 9:1*, 9(1), 1–9.

Cheng, H., Zheng, Z. & Cheng, T., 2020, *New paradigms on hematopoietic stem cell differentiation*, *Protein and Cell*, 11(1), 34–44.

Cho, R.H., Sieburg, H.B. & Muller-Sieburg, C.E., 2008, 'A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the

stem cell compartment but not individual stem cells', *Blood*, 111(12), 5553.

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.

Colom Díaz, P.A., Mistry, J.J. & Trowbridge, J.J., 2023, 'Hematopoietic stem cell aging and leukemia transformation', *Blood*, 142(6), 533–542.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., Kathiria, A., Cho, S.W., Mumbach, M.R., Carter, A.C., Kasowski, M., Orloff, L.A., Risca, V.I., Kundaje, A., Khavari, P.A., Montine, T.J., Greenleaf, W.J. & Chang, H.Y., 2017, 'An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues', *Nature Methods 2017 14:10*, 14(10), 959–962.

Correra, R.M., Ollitrault, D., Valente, M., Mazzola, A., Adalsteinsson, B.T., Ferguson-Smith, A.C., Marazzi, G. & Sassoon, D.A., 2018, 'The imprinted gene Pw1/Peg3 regulates skeletal muscle growth, satellite cell metabolic state, and self-renewal', *Scientific Reports 2018 8:1*, 8(1), 1–14.

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.

Costa, C., Germena, G., Martin-Conte, E.L., Molineris, I., Bosco, E., Marengo, S., Azzolino, O., Altruda, F., Ranieri, V.M. & Hirsch, E., 2011, 'The RacGAP ArhGAP15 is a master negative regulator of neutrophil functions', *Blood*, 118(4), 1099–1108.

Czechowicz, A., Kraft, D., Weissman, I.L. & Bhattacharya, D., 2007, 'Efficient transplantation via antibody-based clearance of hematopoietic stem cell niches', *Science*, 318(5854), 1296–1299.

Dellorusso, P. V, Proven, M.A., Calero-Nieto, F.J., Morrison, S., Gö, B., Wang, X., Mitchell, C.A., Hartmann, F., Amouzgar, M., Favaro, P., Devilbiss, A., Swann, J.W., Ho, T.T., Zhao, Z., Bendall, S.C. & Passegué, E., 2024, 'Autophagy counters inflammation-driven glycolytic impairment in aging hematopoietic stem cells', *Stem Cell*, 31, 1020-1037.e9.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T.R., 2013, 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics (Oxford, England)*, 29(1), 15–21.

Drissen, R., Buza-Vidas, N., Woll, P., Thongjuea, S., Gambardella, A., Giustacchini,

A., Mancini, E., Zriwil, A., Lutteropp, M., Grover, A., Mead, A., Sitnicka, E., Jacobsen, S.E.W. & Nerlov, C., 2016, 'Distinct myeloid progenitor-differentiation pathways identified through single-cell RNA sequencing', *Nature immunology*, 17(6), 666–676.

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.

Esplin, B.L., Shimazu, T., Welner, R.S., Garrett, K.P., Nie, L., Zhang, Q., Humphrey, M.B., Yang, Q., Borghesi, L.A. & Kincade, P.W., 2011, 'Chronic Exposure to a TLR Ligand Injures Hematopoietic Stem Cells', *Journal of immunology (Baltimore, Md. : 1950)*, 186(9), 5367.

Estève, P.O., Vishnu, U.S., Chin, H.G. & Pradhan, S., 2020, 'Visualization and Sequencing of Accessible Chromatin Reveals Cell Cycle and Post-HDAC inhibitor Treatment Dynamics', *Journal of Molecular Biology*, 432(19), 5304–5321.

Ewels, P., Magnusson, M., Lundin, S. & Käller, M., 2016, 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics*, 32(19), 3047–3048.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K.,

Miller, H.W., McElrath, M.J., Prlic, M., Linsley, P.S. & Gottardo, R., 2015, 'MAST: A flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data', *Genome Biology*, 16(1).

Fonseca, R., Burn, T.N., Gandolfo, L.C., Devi, S., Park, S.L., Obers, A., Evrard, M., Christo, S.N., Buquicchio, F.A., Lareau, C.A., McDonald, K.M., Sandford, S.K., Zamudio, N.M., Zanluqui, N.G., Zaid, A., Speed, T.P., Satpathy, A.T., Mueller, S.N., Carbone, F.R. & Mackay, L.K., 2022, 'Runx3 drives a CD8+ T cell tissue residency program that is absent in CD4+ T cells', *Nature Immunology*, 23(8), 1236–1245.

Fujino, T., Goyama, S., Sugiura, Y., Inoue, D., Asada, S., Yamasaki, S., Matsumoto, A., Yamaguchi, K., Isobe, Y., Tsuchiya, A., Shikata, S., Sato, N., Morinaga, H., Fukuyama, T., Tanaka, Y., Fukushima, T., Takeda, R., Yamamoto, K., Honda, H., Nishimura, E.K., Furukawa, Y., Shibata, T., Abdel-Wahab, O., Suematsu, M. & Kitamura, T., 2021, 'Mutant ASXL1 induces age-related expansion of phenotypic hematopoietic stem cells through activation of Akt/mTOR pathway', *Nature Communications 2021 12:1*, 12(1), 1–20.

Gaspar, J.M., no date, 'Improved peak-calling with MACS2'.

Gekas, C. & Graf, T., 2013, 'CD41 expression marks myeloid-biased adult hematopoietic stem cells and increases with age', *Blood*, 121(22), 4463–4472.

Genovese, G., Kähler, A.K., Handsaker, R.E., Lindberg, J., Rose, S.A., Bakhoum, S.F., Chambert, K., Mick, E., Neale, B.M., Fromer, M., Purcell, S.M., Svantesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S.B., Moran, J.L., Lander, E.S., Sullivan, P.F., Sklar, P., Grönberg, H., Hultman, C.M. & McCarroll, S.A., 2014, 'Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence', *New England Journal of Medicine*, 371(26), 2477–2487.

Gibson, C.J., Kim, H.T., Zhao, L., Murdock, H.M., Hambley, B., Ogata, A., Madero-Marroquin, R., Wang, S., Green, L., Fleharty, M., Dougan, T., Cheng, C.A., Blumenstiel, B., Cibulskis, C., Tsuji, J., Duran, M., Gocke, C.D., Antin, J.H., Nikiforow, S., DeZern, A.E., Chen, Y. Bin, Ho, V.T., Jones, R.J., Lennon, N.J., Walt, D.R., Ritz, J., Soiffer, R.J., Gondek, L.P. & Lindsley, R.C., 2022, 'Donor Clonal Hematopoiesis and Recipient Outcomes After Transplantation', *Journal of Clinical Oncology*, 40(2), 189–201.

González-Murillo, Á., Fernández, L., Baena, S., Melen, G.J., Sánchez, R., Sánchez-Valdepeñas, C., Segovia, J.C., Liou, H.C., Schmid, R., Madero, L., Fresno, M. & Ramírez, M., 2015, 'The NFKB Inducing Kinase Modulates Hematopoiesis

during Stress', *Stem Cells*, 33(9), 2825–2837.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y. & Greenleaf, W.J., 2021, 'ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis', *Nature Genetics*, 53(3), 403–411.

Grieshaber-Bouyer, R., Radtke, F.A., Cunin, P., Stifano, G., Levescot, A., Vijaykumar, B., Nelson-Maney, N., Blaustein, R.B., Monach, P.A., Nigrovic, P.A., Aguilar, O., Allan, R., Astarita, J., Austen, K.F., Barrett, N., Baysoy, A., Benoist, C., Brown, B.D., Buechler, M., Buenrostro, J., Casanova, M.A., Chowdhary, K., Colonna, M., Crowl, T., Deng, T., Desland, F., Dhainaut, M., Ding, J., Dominguez, C., Dwyer, D., Frascoli, M., Gal-Oz, S., Goldrath, A., Johanson, T., Jordan, S., Kang, J., Kapoor, V., Kenigsberg, E., Kim, J., Kim, K. wook, Kiner, E., Kronenberg, M., Lanier, L., Laplace, C., Lareau, C., Leader, A., Lee, J., Magen, A., Maier, B., Maslova, A., Mathis, D., McFarland, A., Merad, M., Meunier, E., Monach, P.A., Mostafavi, S., Muller, S., Muus, C., Ner-Gaon, H., Nguyen, Q., Novakovsky, G., Nutt, S., Omilusik, K., Ortiz-Lopez, A., Paynich, M., Peng, V., Potempa, M., Pradhan, R., Quon, S., Ramirez, R., Ramanan, D., Randolph, G., Regev, A., Rose, S.A., Seddu, K., Shay, T., Shemesh, A., Shyer, J., Smilie, C., Spidale, N., Subramanian, A., Sylvia, K., Tellier, J., Turley, S., Vijaykumar, B., Wagers, A., Wang, C., Wang, P.L., Wroblewska, A., Yang, L., Yim, A. & Yoshida,

H., 2021, 'The neutrotime transcriptional signature defines a single continuum of neutrophils across biological compartments', *Nature Communications*, 12(1).

Grover, A., Sanjuan-Pla, A., Thongjuea, S., Carrelha, J., Giustacchini, A., Gambardella, A., Macaulay, I., Mancini, E., Luis, T.C., Mead, A., Jacobsen, S.E.W. & Nerlov, C., 2016, 'Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells', *Nature communications*, 7.

Haan, G. De & Lazare, S.S., 2018a, *Aging of hematopoietic stem cells*.

Haan, G. De & Lazare, S.S., 2018b, 'Aging of hematopoietic stem cells', *Blood*, 131(5), 479–487.

Hahne, F., Khodabakhshi, A.H., Bashashati, A., Wong, C.J., Gascoyne, R.D., Weng, A.P., Seyfert-Margolis, V., Bourcier, K., Asare, A., Lumley, T., Gentleman, R. & Brinkman, R.R., 2010, 'Per-channel basis normalization methods for flow cytometry data', *Cytometry Part A*, 77A(2), 121–131.

Hall, T.D., Kim, H., Dabbah, M., Myers, J.A., Crawford, J.C., Morales-Hernandez, A., Caprio, C.E., Sriram, P., Kooienga, E., Derecka, M., Obeng, E.A., Thomas, P.G. & McKinney-Freeman, S., 2022, 'Murine fetal bone marrow does not support functional hematopoietic stem and progenitor cells until birth', *Nature*

Communications 2022 13:1, 13(1), 1–18.

Han, S.B., Moratz, C., Huang, N.N., Kelsall, B., Cho, H., Shi, C.S., Schwartz, O. & Kehrl, J.H., 2005, 'Rgs1 and Gnai2 Regulate the Entrance of B Lymphocytes into Lymph Nodes and B Cell Motility within Lymph Node Follicles', *Immunity*, 22(3), 343–354.

Haney, M.S., Shankar, A., Hsu, I., Miyauchi, M., Pálovics, R., Khoo, H.M., Igarashi, K.J., Bhadury, J., Munson, C., Mack, P.K., Tan, T.-K., Wyss-Coray, T., Nakauchi, H. & Wilkinson, A.C., no date, 'Large-scale in vivo CRISPR screens identify SAGA complex members as a key regulators of HSC lineage commitment and aging'.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E.P., Jain, J., Srivastava, A., Stuart, T., Fleming, L.M., Yeung, B., Rogers, A.J., McElrath, J.M., Blish, C.A., Gottardo, R., Smibert, P. & Satija, R., 2021, 'Integrated analysis of multimodal single-cell data', *Cell*, 184(13), 3573-3587.e29.

Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C. & Satija, R., 2024,

'Dictionary learning for integrative, multimodal and scalable single-cell analysis', *Nature Biotechnology*, 42(2), 293–304.

Herviou, L., Cavalli, G., Cartron, G., Klein, B. & Moreaux, J., 2016, 'EZH2 in normal hematopoiesis and hematological malignancies', *Oncotarget*, 7(3), 2284.

Hirche, C., Frenz, T., Haas, S.F., Döring, M., Borst, K., Tegtmeyer, P.K., Brizic, I., Jordan, S., Keyser, K., Chhatbar, C., Pronk, E., Lin, S., Messerle, M., Jonjic, S., Falk, C.S., Trumpp, A., Essers, M.A.G. & Kalinke, U., 2017, 'Systemic Virus Infections Differentially Modulate Cell Cycle State and Functionality of Long-Term Hematopoietic Stem Cells In Vivo', *Cell Reports*, 19(11), 2345–2356.

Hu, S., Zhang, W., Ye, J., Zhang, Y., Zhang, D., Peng, J., Yu, D., Xu, J. & Wei, Y., 2021, 'DNA methylation of ARHGAP30 is negatively associated with ARHGAP30 expression in lung adenocarcinoma, which reduces tumor immunity and is detrimental to patient survival', *Aging (Albany NY)*, 13(24), 25799.

Hughes, J.R. & Davies, J.O.J., 2024, 'Single-cell technologies meet Hi-C', *Nature Genetics*, 56(8), 1542–1543.

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), 1502–1506.

Itokawa, N., Oshima, M., Koide, S., Takayama, N., Kuribayashi, W., Nakajima-Takagi, Y., Aoyama, K., Yamazaki, S., Yamaguchi, K., Furukawa, Y., Eto, K. & Iwama, A., 2022, 'Epigenetic traits inscribed in chromatin accessibility in aged hematopoietic stem cells', *Nature Communications 2022 13:1*, 13(1), 1–15.

Izzo, F., Lee, S.C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R.R., Deochand, S.D., Ang, C., Jones, P.W., Nam, A.S., Kim, K.T., Kothen-Hill, S., Schulman, R.C., Ki, M., Lhoumaud, P., Skok, J.A., Viny, A.D., Levine, R.L., Kenigsberg, E., Abdel-Wahab, O. & Landau, D.A., 2020, 'DNA methylation disruption reshapes the hematopoietic differentiation landscape', *Nature Genetics 2020 52:4*, 52(4), 378–387.

Jacobsen, S.E.W. & Nerlov, C., 2019, *Haematopoiesis in the era of advanced single-cell technologies*, *Nature Cell Biology*, 21(1), 2–8.

Jaiswal, S., Natarajan, P., Silver, A.J., Gibson, C.J., Bick, A.G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., Baber, U., Mehran, R., Fuster, V., Danesh, J., Frossard, P., Saleheen, D., Melander, O., Sukhova, G.K., Neuberg, D., Libby, P., Kathiresan, S. & Ebert, B.L., 2017, 'Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease', *New England Journal of Medicine*,

377(2), 111–121.

Johnson, C.S., Williams, M., Sham, K., Belluschi, S., Ma, W., Wang, X., Lau, W.W.Y., Kaufmann, K.B., Krivdova, G., Calderbank, E.F., Mende, N., McLeod, J., Mantica, G., Li, J., Grey-Wilson, C., Drakopoulos, M., Basheer, S., Sinha, S., Diamanti, E., Basford, C., Wilson, N.K., Howe, S.J., Dick, J.E., Göttgens, B., Green, A.R., Francis, N. & Laurenti, E., 2024, 'Adaptation to ex vivo culture reduces human hematopoietic stem cell activity independently of the cell cycle', *Blood*, 144(7), 729–741.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K. & Henikoff, S., 2019, 'CUT&Tag for efficient epigenomic profiling of small samples and single cells', *Nature Communications 2019 10:1*, 10(1), 1–10.

Kim, I., Saunders, T.L. & Morrison, S.J., 2007, 'Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells', *Cell*, 130(3), 470–483.

Kinzfogl, J., Hangoc, G. & Broxmeyer, H.E., 2011, 'Neurexophilin 1 suppresses the proliferation of hematopoietic progenitor cells', *Blood*, 118(3), 565–575.

Kovtonyuk, L. V., Ashcroft, P., Spaltro, G., Tata, N.R., Skoda, R.C., Bonhoeffer, S. & Manz, M.G., 2019, 'Hematopoietic Stem Cells Increase Quiescence during

Aging', *Blood*, 134(Supplement_1), 2484.

Kramer, A. & Challen, G.A., 2016, 'The Epigenetic Basis of Hematopoietic Stem Cell Aging', *Seminars in hematology*, 54(1), 19.

Kulikov, A. V., Korostina, V.S., Kulikova, E.A., Fursenko, D. V., Akulov, A.E., Moshkin, M.P. & Prokhortchouk, E.B., 2016, 'Knockout Zbtb33 gene results in an increased locomotion, exploration and pre-pulse inhibition in mice', *Behavioural Brain Research*, 297, 76–83.

Lacar, B., 2022, 'Generation of Centered Log-Ratio Normalized Antibody-Derived Tag Counts from Large Single-Cell Sequencing Datasets', *Methods in Molecular Biology*, 2386, 203–217.

Lareau, C.A., Ludwig, L.S., Muus, C., Gohil, S.H., Zhao, T., Chiang, Z., Pelka, K., Verboon, J.M., Luo, W., Christian, E., Rosebrock, D., Getz, G., Boland, G.M., Chen, F., Buenrostro, J.D., Hacohen, N., Wu, C.J., Aryee, M.J., Regev, A. & Sankaran, V.G., 2021, 'Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling', *Nature Biotechnology*, 39, 451–461.

Lause, J., Berens, P. & Kobak, D., 2021, 'Analytic Pearson residuals for normalization of single-cell RNA-seq UMI data', *Genome Biology*, 22(1), 1–20.

Leblay, N., Maity, R., Barakat, E., McCulloch, S., Duggan, P., Jimenez-Zepeda, V.,

Bahlis, N.J. & Neri, P., 2020, 'Cite-Seq Profiling of T Cells in Multiple Myeloma Patients Undergoing BCMA Targeting CAR-T or Bites Immunotherapy', *Blood*, 136(Supplement 1), 11–12.

Lee, Y.T., Vasconcellos, J.F. De, Yuan, J., Byrnes, C., Noh, S.J., Meier, E.R., Kim, K.S., Rabel, A., Kaushal, M., Muljo, S.A. & Miller, J.L., 2013, 'LIN28B-mediated expression of fetal hemoglobin and production of fetal-like erythrocytes from adult human erythroblasts ex vivo', *Blood*, 122(6), 1034.

Leveque, L., Vu, T., Kuns, R.D., Teal, B.E., Lor, M., Lineburg, K.E., Bruedigam, C., Bagger, F.O., Boyle, G., Hill, G.R., MacDonald, K.P.A. & Lane, S.W., 2013, 'Autophagy Is Required For Long-Term Hematopoietic Stem Cell (HSC) Function and G-CSF-Induced HSC Mobilization', *Blood*, 122(21), 892.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R., 2009, 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics (Oxford, England)*, 25(16), 2078–2079.

Li, L. & Bhatia, R., 2011, 'Stem cell quiescence', *Clinical Cancer Research*, 17(15), 4936–4941.

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

Liao, Y., Smyth, G.K. & Shi, W., 2014, 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics (Oxford, England)*, 30(7), 923–930.

Love, M.I., Huber, W. & Anders, S., 2014, 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biology*, 15(12), 1–21.

Luo, Y., Li, L., Zou, P., Wang, J., Shao, L., Zhou, D. & Liu, L., 2014, 'Rapamycin enhances long-term hematopoietic reconstitution of ex vivo expanded mouse hematopoietic stem cells by inhibiting senescence', *Transplantation*, 97(1), 20–29.

Ma, W., Lu, J. & Wu, H., 2023, 'Cellcano: supervised cell type identification for single cell ATAC-seq data', *Nature Communications*, 14(1).

Man, T.-K., Najaf Panah, M.J., Elswood, J.L., Sumazin, P. & Redell, M.S., 2021, 'Cite-Seq Reveals Distinct Patterns and Potential Mechanisms of Relapse in Pediatric AML', *Blood*, 138(Supplement 1), 3458–3458.

Marston, N.A., Pirruccello, J.P., Melloni, G.E.M., Kamanu, F., Bonaca, M.P., Giugliano, R.P., Scirica, B.M., Wiviott, S.D., Bhatt, D.L., Steg, P.G., Raz, I., Braunwald, E., Libby, P., Ellinor, P.T., Bick, A.G., Sabatine, M.S. & Ruff, C.T., 2024, 'Clonal hematopoiesis, cardiovascular events and treatment benefit in 63,700 individuals from five TIMI randomized trials', *Nature Medicine 2024 30:9*,

30(9), 2641–2647.

Matteini, F., Mulaw, M.A. & Florian, M.C., 2021, 'Aging of the Hematopoietic Stem Cell Niche: New Tools to Answer an Old Question', *Frontiers in Immunology*, 12, 738204.

Meaker, G.A. & Wilkinson, A.C., 2024, 'Ex vivo hematopoietic stem cell expansion technologies: recent progress, applications, and open questions', *Experimental Hematology*, 130, 104136.

Meng, Y., Carrelha, J., Drissen, R., Ren, X., Zhang, B., Gambardella, A., Valletta, S., Thongjuea, S., Jacobsen, S.E. & Nerlov, C., 2023, 'Epigenetic programming defines haematopoietic stem cell fate restriction', *Nature Cell Biology 2023 25:6*, 25(6), 812–822.

Meng, Y. & Nerlov, C., 2024, *Epigenetic regulation of hematopoietic stem cell fate*, *Trends in Cell Biology*.

M.Hosseini, M., Enright, T., Duvall, A.S., Chitsazan, A., Lin, H.-Y., Ors, A., Davis, B., Nikolova, O., Bresciani, E., Diemer, J.L., Craft, K., Menezes, C., Merguerian, M., Chong, S., Calvo, K.R., Glushakow-Smith, S., Gritsman, K., Godley, L., Horwitz, M.S., Keel, S., Castilla, L.H., Demir, E., Mohammed, H., Liu, P.P. & Agarwal, A., 2024, 'CD74 Signaling Regulates Inflammatory Stress to Drive

Defective Hematopoiesis and Fitness Advantage of Progenitors in RUNX1-Familial Platelet Disorder', *Blood*, 144(Supplement 1), 198–198.

Miller, C.L. & Lai, B., 2005, 'Human and mouse hematopoietic colony-forming cell assays.', *Methods in molecular biology (Clifton, N.J.)*, 290, 71–89.

Mimitou, E.P., Lareau, C.A., Chen, K.Y., Zorzetto-Fernandes, A.L., Hao, Y., Takeshima, Y., Luo, W., Huang, T.S., Yeung, B.Z., Papalexi, E., Thakore, P.I., Kibayashi, T., Wing, J.B., Hata, M., Satija, R., Nazor, K.L., Sakaguchi, S., Ludwig, L.S., Sankaran, V.G., Regev, A. & Smibert, P., 2021, 'Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells', *Nature Biotechnology 2021 39:10*, 39(10), 1246–1258.

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*, 92(22), 10302–10306.

Mortera-Blanco, T., Dimitriou, M., Woll, P., Karimi, M., Papaemmanuil, E., Conte, S., Chowdhuri, O., Jansson, M., Campbell, P.J., Jacobsen, S.E.W. & Hellström-Lindberg, E., 2013, 'Hierarchical Analysis Of Recurrent Point Mutations In SF3B1 and TET2 In RARS Stem Cells', *Blood*, 122(21), 2749.

Mulè, M.P., Martins, A.J. & Tsang, J.S., 2022, 'Normalizing and denoising protein

expression data from droplet-based single cell profiling', *Nature Communications 2022 13:1*, 13(1), 1–12.

Müller-Sieburg, C.E., Cho, R.H., Thoman, M., Adkins, B. & Sieburg, H.B., 2002, 'Deterministic regulation of hematopoietic stem cell self-renewal and differentiation', *Blood*, 100(4), 1302–1309.

Muller-Sieburg, C.E., Sieburg, H.B., Bernitz, J.M. & Cattarossi, G., 2012, 'Stem cell heterogeneity: implications for aging and regenerative medicine', *Blood*, 119(17), 3900.

Nam, A.S., Dusaj, N., Izzo, F., Murali, R., Myers, R.M., Mouhieddine, T.H., Sotelo, J., Benbarche, S., Waarts, M., Gaiti, F., Tahri, S., Levine, R., Abdel-Wahab, O., Godley, L.A., Chaligne, R., Ghobrial, I. & Landau, D.A., 2022, 'Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation', *Nature Genetics 2022 54:10*, 54(10), 1514–1526.

Nichols, R. V., Rylaarsdam, L.E., O'Connell, B.L., Shipony, Z., Iremadze, N., Acharya, S.N. & Adey, A.C., 2025, 'Atlas-scale single-cell DNA methylation profiling with sciMETv3', *Cell Genomics*, 5(1).

Niederwieser, D., Baldomero, H., Bazuaye, N., Bupp, C., Chaudhri, N., Corbacioglu,

S., Elhaddad, A., Frutos, C., Galeano, S., Hamad, N., Hamidieh, A.A., Hashmi, S., Ho, A., Horowitz, M.M., Iida, M., Jaimovich, G., Karduss, A., Kodera, Y., Kröger, N., Latour, R.P. de, Lee, J.W., Martínez-Rolón, J., Pasquini, M.C., Passweg, J., Paulson, K., Seber, A., Snowden, J.A., Srivastava, A., Szer, J., Weisdorf, D., Worel, N., Koh, M.B.C., Aljurf, M., Greinix, H., Atsuta, Y. & Saber, W., 2022, 'One and a half million hematopoietic stem cell transplants: continuous and differential improvement in worldwide access with the use of non-identical family donors', *Haematologica*, 107(5), 1045–1053.

Osorio, D. & Cai, J.J., 2021, 'Systematic determination of the mitochondrial proportion in human and mice tissues for single-cell RNA-sequencing data quality control', *Bioinformatics*, 37(7), 963.

Ou, J., Liu, H., Yu, J., Kelliher, M.A., Castilla, L.H., Lawson, N.D. & Zhu, L.J., 2018, 'ATACseqQC: A Bioconductor package for post-alignment quality assessment of ATAC-seq data', *BMC Genomics*, 19(1), 1–13.

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 of the United States of America*, 108(50), 20012–20017.

Pang, W.W., Schrier, S.L. & Weissman, I.L., 2017, 'Age-associated changes in human hematopoietic stem cells', *Seminars in Hematology*, 54(1), 39–42.

Passarelli, C., Civino, A., Rossi, M.N., Cifaldi, L., Lanari, V., Moneta, G.M., Caiello, I., Bracaglia, C., Montinaro, R., Novelli, A., Benedetti, F. De & Prencipe, G., 2020, 'IFNAR2 Deficiency Causing Dysregulation of NK Cell Functions and Presenting With Hemophagocytic Lymphohistiocytosis', *Frontiers in Genetics*, 11, 555244.

Pei, W., Feyerabend, T.B., Rössler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., Chen, W., Sauer, S., Wolf, S., Höfer, T. & Rodewald, H.R., 2017, 'Polylox barcoding reveals haematopoietic stem cell fates realized in vivo', *Nature*, 548(7668), 456–460.

Pei, W., Shang, F., Wang, X., Fanti, A.K., Greco, A., Busch, K., Klapproth, K., Zhang, Q., Quedenau, C., Sauer, S., Feyerabend, T.B., Höfer, T. & Rodewald, H.R., 2020, 'Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding', *Cell Stem Cell*, 27(3), 383-395.e8.

Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S. & Sandberg, R., 2014, 'Full-length RNA-seq from single cells using Smart-seq2', *Nature Protocols 2013 9:1*, 9(1), 171–181.

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.

Ramalingam, P., Gutkin, M.C., Poulos, M.G., Tillery, T., Doughty, C., Winiarski, A., Freire, A.G., Rafii, S., Redmond, D. & Butler, J.M., 2023, 'Restoring bone marrow niche function rejuvenates aged hematopoietic stem cells by reactivating the DNA Damage Response', *Nature Communications 2023 14:1*, 14(1), 1–20.

Rodríguez-Bayona, B., Ramos-Amaya, A. & Brieva, J.A., 2011, 'Differential expression of SLAMS and other modulatory molecules by human plasma cells during normal maturation', *Immunology letters*, 134(2), 122–128.

Ronchetti, S., Ricci, E., Petrillo, M.G., Cari, L., Migliorati, G., Nocentini, G. & Riccardi, C., 2015, 'Glucocorticoid-Induced Tumour Necrosis Factor Receptor-Related Protein: A Key Marker of Functional Regulatory T Cells', *Journal of Immunology Research*, 2015, 171520.

Ross, J.B., Myers, L.M., Noh, J.J., Collins, M.M., Carmody, A.B., Messer, R.J., Dhuey, E., Hasenkrug, K.J. & Weissman, I.L., 2024, 'Depleting myeloid-biased haematopoietic stem cells rejuvenates aged immunity', *Nature*, 628(8006), 162–

170.

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 of the United States of America*, 102(26), 9194–9199.

Saha, A., Palchaudhuri, R., Lanieri, L., Hyzy, S., Riddle, M.J., Panthera, J., Eide, C.R., Tolar, J., Panoskaltsis-Mortari, A., Gorfinkel, L., Tkachev, V., Gerdemann, U., Alvarez-Calderon, F., Palato, E.R., MacMillan, M.L., Wagner, J.E., Kean, L.S., Osborn, M.J., Kiem, H.P., Scadden, D.T., Olson, L.M. & Blazar, B.R., 2024, 'Alloengraftment without significant toxicity or GVHD in CD45 antibody-drug conjugate–conditioned Fanconi anemia mice', *Blood*, 143(21), 2201–2216.

Salomon, R., Kaczorowski, D., Valdes-Mora, F., Nordon, R.E., Neild, A., Farbehi, N., Bartonicek, N. & Gallego-Ortega, D., 2019, 'Droplet-based single cell RNAseq tools: a practical guide', *Lab on a Chip*, 19(10), 1706–1727.

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., Carrelha, J., Tarrant, P., Atkinson, D., Clark, S.A., Nerlov, C. &

Jacobsen, S.E.W., 2013, 'Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy', *Nature 2013 502:7470*, 502(7470), 232–236.

Schep, A.N., Wu, B., Buenrostro, J.D. & Greenleaf, W.J., 2017, 'ChromVAR: Inferring transcription-factor-associated accessibility from single-cell epigenomic data', *Nature Methods*, 14(10), 975–978.

Schiroli, G., Kartha, V., Duarte, F.M., Kristiansen, T.A., Mayerhofer, C., Shrestha, R., Earl, A., Hu, Y., Tay, T., Rhee, C., Buenrostro, J.D. & Scadden, D.T., 2024, 'Cell of origin epigenetic priming determines susceptibility to Tet2 mutation', *Nature Communications 2024 15:1*, 15(1), 1–20.

Schnoeder, T.M., Schwarzer, A., Jayavelu, A.K., Hsu, C.J., Kirkpatrick, J., Döhner, K., Perner, F., Eifert, T., Huber, N., Arreba-Tutusaus, P., Dolnik, A., Assi, S.A., Nafria, M., Jiang, L., Dai, Y.T., Chen, Z., Chen, S.J., Kellaway, S.G., Ptasinska, A., Ng, E.S., Stanley, E.G., Elefanty, A.G., Buschbeck, M., Bierhoff, H., Brodt, S., Matziolis, G., Fischer, K.D., Hochhaus, A., Chen, C.W., Heidenreich, O., Mann, M., Lane, S.W., Bullinger, L., Ori, A., Eyss, B. von, Bonifer, C. & Heidel, F.H., 2022, 'PLCG1 is required for AML1-ETO leukemia stem cell self-renewal', *Blood*, 139(7), 1080–1097.

Shao, L., Li, H., Pazhanisamy, S.K., Meng, A., Wang, Y. & Zhou, D., 2011, 'Reactive

oxygen species and hematopoietic stem cell senescence', *International Journal of Hematology*, 94(1), 24.

Singh, I., Fernandez-Perez, D., Sanchez, P.S. & Rodriguez-Fraticelli, A.E., 2025, 'Pre-existing stem cell heterogeneity dictates clonal responses to the acquisition of leukemic driver mutations', *Cell Stem Cell*, 32(4), 564-580.e6.

Skinder, N., Fernández, I.S., Dethmers-Ausema, A., Weersing, E. & Haan, G. de, 2024, 'CD61 identifies a superior population of aged murine HSCs and is required to preserve quiescence and self-renewal', *Blood Advances*, 8(1), 99–111.

Spangrude, G.J., Heimfeld, S. & Weissman, I.L., 1988, 'Purification and Characterization of Mouse Hematopoietic Stem Cells', *Science*, 241(4861), 58–62.

Stoeckius, M., Hafemeister, C., Stephenson, W., Houck-Loomis, B., Chattopadhyay, P.K., Swerdlow, H., Satija, R. & Smibert, P., 2017, 'Simultaneous epitope and transcriptome measurement in single cells', *Nature methods*, 14(9), 865–868.

Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N., Purdom, E. & Dudoit, S., 2018, 'Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics', *BMC Genomics*, 19(1).

Stuart, T., Srivastava, A., Madad, S., Lareau, C.A. & Satija, R., 2021, 'Single-cell chromatin state analysis with Signac', *Nature Methods 2021 18:11*, 18(11), 1333–

1341.

Su, T.Y., Hauenstein, J., Somuncular, E., Dumral, Ö., Leonard, E., Gustafsson, C., Tzortzis, E., Forlani, A., Johansson, A.S., Qian, H., Månsson, R. & Luc, S., 2024, 'Aging is associated with functional and molecular changes in distinct hematopoietic stem cell subsets', *Nature Communications 2024 15:1*, 15(1), 1–16.

Sudo, K., Ema, H., Morita, Y. & Nakauchi, H., 2000, 'Age-Associated Characteristics of Murine Hematopoietic Stem Cells', *Journal of Experimental Medicine*, 192(9), 1273–1280.

Sun, B., Mallampati, S., Gong, Y., Wang, D., Lefebvre, V. & Sun, X., 2013, 'Sox4 Is Required for the Survival of Pro-B Cells', *The Journal of Immunology*, 190(5), 2080–2089.

Suzawa, M., Noguchi, K., Nishi, K., Kozuka-Hata, H., Oyama, M. & Ui-Tei, K., 2017, 'Comprehensive Identification of Nuclear and Cytoplasmic TNRC6A-Associating Proteins', *Journal of Molecular Biology*, 429(21), 3319–3333.

Szilvassy, S.J., Humphries, R.K., Lansdorp, P.M., Eaves, A.C. & Eaves, C.J., 1990, 'Quantitative assay for totipotent reconstituting hematopoietic stem cells by a competitive repopulation strategy.', *Proceedings of the National Academy of Sciences*, 87(22), 8736–8740.

Takayama, N., Murison, A., Takayanagi, S. ichiro, Arlidge, C., Zhou, S., Garcia-Prat, L., Chan-Seng-Yue, M., Zandi, S., Gan, O.I., Boutzen, H., Kaufmann, K.B., Trotman-Grant, A., Schoof, E., Kron, K., Díaz, N., Lee, J.J.Y., Medina, T., Carvalho, D.D. De, Taylor, M.D., Vaquerizas, J.M., Xie, S.Z., Dick, J.E. & Lupien, M., 2021, 'The Transition from Quiescent to Activated States in Human Hematopoietic Stem Cells Is Governed by Dynamic 3D Genome Reorganization', *Cell Stem Cell*, 28(3), 488-501.e10.

Tanese, K., Hashimoto, Y., Berkova, Z., Wang, Y., Samaniego, F., Lee, J.E., Ekmekcioglu, S. & Grimm, E.A., 2015, 'Cell Surface CD74–MIF Interactions Drive Melanoma Survival in Response to Interferon-γ', *Journal of Investigative Dermatology*, 135(11), 2775–2784.

Teschendorff, A.E., West, J. & Beck, S., 2013, 'Age-associated epigenetic drift: implications, and a case of epigenetic thrift?', *Human Molecular Genetics*, 22(R1), R7.

TILL, J.E. & McCULLOCH, E.A., 1961, 'A direct measurement of the radiation sensitivity of normal mouse bone marrow cells', *Radiation research*, 14, 213–222.

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., Kazer, S.W., Gaillard, A., Kolb, K.E., Villani, A.C., Johannessen, C.M., Andreev, A.Y., Allen, E.M. Van, Bertagnolli, M., Sorger, P.K., Sullivan, R.J., Flaherty, K.T., Frederick, D.T., Jané-Valbuena, J., Yoon, C.H., Rozenblatt-Rosen, O., Shalek, A.K., Regev, A. & Garraway, L.A., 2016, 'Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq', *Science*, 352(6282), 189–196.

Tsukumo, S.I., Subramani, P.G., Seija, N., Tabata, M., Maekawa, Y., Mori, Y., Ishifune, C., Itoh, Y., Ota, M., Fujio, K., Noia, J.M. Di & Yasutomo, K., 2022, 'AFF3, a susceptibility factor for autoimmune diseases, is a molecular facilitator of immunoglobulin class switch recombination', *Science Advances*, 8(34), 8.

Valletta, S., Thomas, A., Meng, Y., Ren, X., Drissen, R., Sengül, H., Genua, C. Di & Nerlov, C., 2020, 'Micro-environmental sensing by bone marrow stroma identifies IL-6 and TGFβ1 as regulators of hematopoietic ageing', *Nature Communications*, 11(1).

Vo, L.T., Kinney, M.A., Liu, X., Zhang, Y., Barragan, J., Sousa, P.M., Jha, D.K., Han, A., Cesana, M., Shao, Z., North, T.E., Orkin, S.H., Doulatov, S., Xu, J. & Daley, G.Q., 2018, 'Regulation of embryonic haematopoietic multipotency by EZH1',

Nature 2017 553:7689, 553(7689), 506–510.

Wang, P., Wang, Z. & Liu, J., 2020, 'Role of HDACs in normal and malignant hematopoiesis', *Molecular Cancer 2020 19:1*, 19(1), 1–21.

Wang, Y., Zhang, W., Zhang, C., Van, H.Q.T., Seino, T. & Zhang, Y., 2025, 'Reducing functionally defective old HSCs alleviates aging-related phenotypes in old recipient mice', *Cell Research 2024 35:1*, 35(1), 45–58.

Wiesmann, A., Phillips, R.L., Mojica, M., Pierce, L.J., Searles, A.E., Spangrude, G.J. & Lemischka, I., 2000, 'Expression of CD27 on Murine Hematopoietic Stem and Progenitor Cells', *Immunity*, 12(2), 193–199.

Wilkinson, A.C., Ishida, R., Kikuchi, M., Sudo, K., Morita, M., Crisostomo, R.V., Yamamoto, R., Loh, K.M., Nakamura, Y., Watanabe, M., Nakauchi, H. & Yamazaki, S., 2019, 'Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation', *Nature 2019 571:7763*, 571(7763), 117–121.

Xu, S., Hu, E., Cai, Y., Xie, Z., Luo, X., Zhan, L., Tang, W., Wang, Q., Liu, B., Wang, R., Xie, W., Wu, T., Xie, L. & Yu, G., 2024, 'Using clusterProfiler to characterize multiomics data', *Nature Protocols*.

Xu, W., Wen, Y., Liang, Y., Xu, Q., Wang, X., Jin, W. & Chen, X., 2021, 'A plate-

based single-cell ATAC-seq workflow for fast and robust profiling of chromatin accessibility', *Nature Protocols 2021 16:8*, 16(8), 4084–4107.

Xu, X., Zheng, L., Yuan, Q., Zhen, G., Crane, J.L., Zhou, X. & Cao, X., 2018, 'Transforming growth factor-β in stem cells and tissue homeostasis', *Bone Research 2018 6:1*, 6(1), 1–31.

Xu, Z., Heidrich-O'Hare, E., Chen, W. & Duerr, R.H., 2022, 'Comprehensive benchmarking of CITE-seq versus DOGMA-seq single cell multimodal omics', *Genome Biology*, 23(1), 1–17.

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).

Yamamoto, R., Wilkinson, A.C., Ooehara, J., Lan, X., Lai, C.Y., Nakauchi, Y., Pritchard, J.K. & Nakauchi, H., 2018, 'Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment', *Cell stem cell*, 22(4), 600-607.e4.

Yamashita, M. & Passegué, E., 2019, 'TNF-α Coordinates Hematopoietic Stem Cell Survival and Myeloid Regeneration', *Cell Stem Cell*, 25(3), 357-372.e7.

Yan, F., Powell, D.R., Curtis, D.J. & Wong, N.C., 2020, 'From reads to insight: a hitchhiker's guide to ATAC-seq data analysis', *Genome Biology 2020 21:1*, 21(1), 1–16.

Yoon, H.G., Chan, D.W., Reynolds, A.B., Qin, J. & Wong, J., 2003, 'N-CoR Mediates DNA Methylation-Dependent Repression through a Methyl CpG Binding Protein Kaiso', *Molecular Cell*, 12(3), 723–734.

Zhao, Q., Sheng, M.F., Wang, Y.Y., Wang, X.Y., Liu, W.Y., Zhang, Y.Y., Ke, T.Y., Chen, S., Pang, G.Z., Yong, L., Ding, Z., Shen, Y.J., Shen, Y.X. & Shao, W., 2022, 'LncRNA Gm26917 regulates inflammatory response in macrophages by enhancing Annexin A1 ubiquitination in LPS-induced acute liver injury', *Frontiers in Pharmacology*, 13.

Zhou, T., Zhang, R., Jia, D., Doty, R.T., Munday, A.D., Gao, D., Xin, L., Abkowitz, J.L., Duan, Z. & Ma, J., 2024, 'GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells', *Nature Genetics*, 56(8), 1701–1711.

Zilionis, R., Nainys, J., Veres, A., Savova, V., Zemmour, D., Klein, A.M. & Mazutis, L., 2016, 'Single-cell barcoding and sequencing using droplet microfluidics', *Nature Protocols 2016 12:1*, 12(1), 44–73.

No date, *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*.

No date, *Picard Tools - By Broad Institute*.

No date, *PBMC from a Healthy Donor - No Cell Sorting (3k) - 10x Genomics*.