# IDENTIFICATION OF NOVEL DRIVERS OF HAEMATOPOIETIC STEM CELL FATE USING LOW CELL NUMBER PROTEOMICS AND SINGLE CELL PROFILING 

## DANIEL BODE



## Clare College

University of Cambridge
September 2021

This Thesis is submitted for the Degree of Doctor of Philosophy

## Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the School of Clinical Medicine and Department of Haematology Degree Committee.

# IDENTIFICATION OF <br> NOVEL <br> DRIVERS <br> OF <br> HAEMATOPOIETIC STEM CELL FATE USING LOW CELL NUMBER PROTEOMICS AND SINGLE CELL PROFILING 

Daniel Bode


#### Abstract

The abrogation and de-regulation of cellular decisions in adult haematopoietic stem cells (HSCs) have been widely recognised as key contributing factors in ageing and disease. In particular, acquired mutations altering HSC self-renewal contribute to the formation of pre-leukaemic disorders, such as myeloproliferative neoplasms (MPNs), and ultimately to their transformation to acute myeloid leukaemia. Therefore, we urgently require a complete characterisation of the underlying molecular pathways of HSC self-renewal in order to provide crucial insights into disease progression and to help inform novel strategies for ex vivo expansion for gene therapy applications.


To date, advances in functional and molecular single cell technologies have provided unprecedented resolution of heterogeneous HSC populations and their transcriptional landscapes. In contrast, the scarcity of functional HSCs and technical limitations of unbiased proteomic screening technology have prevented comprehensive characterisation of protein networks governing HSC fate. To overcome these technical limitations, we developed an optimised mass spectrometry workflow to interrogate as few as 10,000 HSCs and multiplex up to 16 cell fractions. Using this approach, we were able to quantify in excess of 4,000 proteins, while reducing the required cell input 30-fold. To identify key molecular drivers of in vivo HSC self-renewal, we probed HSC populations with increasing self-renewal potency, including TET2-deficient HSCs exhibiting a self-renewal advantage in MPN mouse models. Here, we observed reshaping of extracellular matrix protein networks, indicating a potential physical role of the neighbouring bone marrow niche cells for regulating HSC fate. Next, we integrated proteomic and transcriptomic to characterise the molecular pathways
underlying HSC self-renewal across numerous -omics technologies. Here, we identified a wide range of intrinsic regulatory pathways and described molecular mechanisms regulating intracellular calcium levels in HSCs.

In the final results chapter, we explored the ex vivo expansion of HSCs, since this is of paramount importance for the delivery of gene and cellular therapies against a plethora of monogenic haematological diseases. Recently pioneered HSC expansion protocols greatly enhanced the yield of phenotypic HSCs, but such cultures exhibit significant clone-to-clone variability in long-term self-renewal potency and differentiation. By linking the transplantation outcomes with transcriptional profiles of individual single cell-derived clones, we derived a novel gene signature for ex vivo expanded HSCs with long-term self-renewal potency and characterised key molecular pathways governing HSC fate throughout expansion. Furthermore, we identified a reporter strategy for prospectively isolating expanded HSCs.

Together, these findings provided an insight into the molecular machinery underlying HSC self-renewal within the native bone marrow niche and during ex vivo expansion. Comprehensive multi-omic profiling also revealed the intricate relationship of gene expression profiles with the proteomic phenotype within the HSC compartment. Finally, we propose an optimised workflow for performing comprehensive proteomics on any rare cell populations which will be of use to researchers investigating a widerange of cellular biology questions.

## Preface

## Chapter 1

All work was conducted under the supervision of Dr David Kent. The protocol for rare cell proteomic mass spectrometry screening was co-developed with Dr Theodoros Roumeliotis and Prof Jyoti Choudhary (Institute of Cancer Research). Single cell RNA sequencing data of Hoxb8-FL cells was provided by Dr Iwo Kucinski and Prof Berthold Göttgens. Raw sequencing data was processed also performed by Dr Iwo Kucinski and the Göttgens group.

## Chapter 2

All work was conducted under the supervision of Dr David Kent. Cell isolation and mass spectrometry sample processing was conducted by me. The subsequently offline fractionation and operation of the Orbitrap Lumos mass spectrometer were performed by Dr Theodoros Roumeliotis. Prior to bulk sequencing, RNA sequencing was performed by me and the subsequently library preparation and sequencing were conducted by the Cambridge Stem Cell Institute (CSCI) Genomics Core Facility (Maike Paramor and Vicky Murray). Single cell RNA sequencing libraries were prepared by Dr Alyssa Cull and Dr Sally James (University of York). The raw of single cell RNA sequencing files were aligned by Dr John Davey (University of York). Cell sorting was conducted by Dr Rainer Schulte, Dr Chiara Cossetti and Gabriela Grondys-Kortoba at the Cambridge Institute for Medical Research (CIMR) Flow Cytometry Core Facility or Dr Karen Hogg and Dr Graeme Park at the Technology Facility, University of York. Extracellular matrix protein coating and subsequent cultures were conduction by Dr Juan Antonio-Rubio (University of York). Microscope operation for calcium imaging and subsequent data analysis was conducted by Dr Andrew James (University of York).

## Chapter 3

All work was conducted under the supervision of Dr David Kent. All transplantation assays were conducted by Dr James Che, Dr David Kent, Dr Alyssa Cull and Grace Boyd. Peripheral blood analysis was also conducted by Dr James Che and Dr Alyssa Cull. Single cell 28-day expansion cultures and bulk inhibitor cultures were conducted
and analysed by Dr James. RNA extractions and pre-processing for bulk RNA sequencing was performed in tandem by Dr James Che and me. The subsequent library preparations and sequencing were performed by Maike Paramor and Vicky Murray at the Cambridge Stem Cell Institute (CSCI) Genomics Core Facility. Single cell sequencing libraries (Smart-seq2) were constructed by Dr James Che and sequenced by the CSCI Genomics Core Facility. The Smart-seq2 data was aligned by Dr Iwo Kucinski. Single cell RNA sequencing (10X) to profile a bulk 28-day expansion culture was performed and processed by Dr Iwo Kucinski, Prof Bertie Göttgens, and Dr Adam Wilkinson. Cell sorting was conducted by Dr Rainer Schulte, Dr Chiara Cossetti and Gabriela Grondys-Kortoba at the CIMR Flow Cytometry Core Facility.

## Acknowledgements

I am very grateful for the immense support from friends, my family, colleagues, and the funding bodies. It has been a joy to learn, grow and laugh together!

First, I would like to express my gratitude to Dr David Kent. His dedication to sharing his knowledge, guiding, and teaching are unparalleled. His input has not only helped me mature as a scientist, but also played an enormous role in my professional development away from the bench. I am also particularly grateful for his trust to let me explore - His supervision has truly shaped my future career.

I am also immensely grateful to Jyoti Choudhary for the many fruitful discussions. The conversations about science, technology and biotech have been, and will continue to be, great beacons in my professional development. I would like to extend my gratitude to Prof Berthold Göttgens for taking on the primary supervisory responsibilities upon Dave's move to York. I am particularly grateful to the Wellcome Trust, who provided me the opportunity to embark on an exciting PhD program. Similarly, I am very grateful to the Cambridge Stem Cell Institute and particularly Dr Brian Hendrich for putting together an incredibly thorough PhD program. The rotations and assignments were of paramount importance to develop my ability to present, write and communicate.

None of this work would have been possible without my amazing colleagues! It has been an absolute pleasure to work with all members of the Kent lab, especially Alyssa Cull, Miriam Belmonte, Craig McDonald, Grace Boyd, James Che and Juan AntonioRubio. I am particularly grateful to James Che and Theo Roumeliotis - It has been a truly collaborative effort and a lot of fun building stories together. I am also grateful to Iwo Kucinski, Andy James and all members of the amazing core facilities, without whom this work would not have been possible.Finally, I would like to thank my family and friends. I am particularly grateful to my parents, Olga and Hans-Jürgen for their countless advice, support, and life lessons! Also, a big shoutout to the handball team and all my friends, particularly Felix, Tim, Chis, Jamie, David, Cloé, Klara, Beza, Sebastian, Florian, Phillipp - so many memories and I hope a great start to life-long friendships!

## Abbreviations

| 5-hmC | 5-Hydroxymethylcytosine |
| :--- | :--- |
| 5-mC | 5-Methylcytosine |
| AGM | Aorta-Gonad-Mesonephros |
| AML | Acute Myeloid Leukaemia |
| Ang-1 | Angiopoietin-1 |
| BM | Bone Marrow |
| BSA | Bovine Serum Albumin |
| BSA | Bovine Serum Albumin |
| CAR | Cxcl12-Abundant Reticular |
| CFU-C | Colony-Forming Unit Cell |
| CFU-S | Colony-Forming Unit Spleen |
| ChIP-seq | Chromatin Imuunopercipitation-Sequencing |
| CITE-seq | Cellular Indexing Transcriptome And Epitome By |
| CLP | Sequencing |
| CMML | Common Lymphoid Progenitor |
| CMP | Chronic Myelomonocytic Leukaemia |
| CRM | Common Myeloid Progenitor |
| CV | Cis-Regulatory Motif |
| DEG | Coefficient Of Variability |
| DGE | Differentially Expressed Gene |
| DGE | Differential Gene Expression |
| DoT | Differential Gene Expression |
| dpc | Direction Of Transition |
| DPE | Days Post Coitum |
| E | Differential Protein Expression |
| ECM | Embryonic Day |
| ELSK | Extracellular Matrix |
| EoBP | Epcr+Lin-Sca1+Ckit+ |
| EPCR | Eosinophil-Basophil Progenitor |
| ESI | Endothelial Protein C Receptor |
| ESLAM | Electrospray lonisation |
| ET | Epcr-Slam |
| FACS | Essential Thrombocythaemia |
| FAIMS | Fluorescence-Activated Cell Sorting |
| FBLN1 | High Field Assymetric lon Mobility Spectrometry |
| FCS | Fibulin |
| FDR | Foetal Calf Serum |
| FELSK | False Discover Rate |
| FGG | Fibrisk |
|  |  |


| $F^{\text {hi }} \mathrm{E}^{\text {hi }}$ | Fgd5Hiepcrhi |
| :---: | :---: |
| FL | Foetal Liver |
| $\mathrm{F}^{10} \mathrm{E}^{10}$ | Fgd5Loepcrio |
| FLT3L | Flt3 Ligand |
| G-CSF | Granulocyte Colony-Stimulating Factor |
| GFP | Green Fluorescent Protein |
| GM | Granulocyte-Macrophage |
| GMP | Granulocyte And Macrophage Progenitor |
| GO | Gene Ontology |
| GRN | Gene Regulatory Network |
| HA | Hyaluronic Acid |
| HPLC | High-Performance Liquid Chromatography |
| HSC | Haematopoietic Stem Cell |
| HSCT | Haematopoietic Stem Cell Transplantation |
| HSPC | Haematopoietic Stem/Progenitor Cell |
| IAA | lodoacetamide |
| IL-11 | Interleukin-11 |
| IL-6 | Interleukin-6 |
| ImmGen | Immunological Genome Project |
| IT | Intermediate |
| KEGG | Encyclopaedia Of Genes And Genomes |
| LCM | Laser Capture Microdissection |
| Lin | Lineage |
| LK | Lin-Ckit+ |
| LMPP | Lymphoid-Primed Multipotent Progenitor |
| LPS | Lipopolysaccharide |
| LSK | Ckit+ Sca-1+ Lin- |
| LT | Long-Term |
| m/z | Mass To Charge |
| MALDI | Matrix-Assisted Laser Desorption Ionisation |
| MEP | Megakaryocyte And Erythoid Progenitor |
| MF | Myelofibrosis |
| MkP | Megakaryocyte Progenitor |
| MolO | Molecular Overlap |
| MPN | Myeloproliferative Neoplasm |
| MPP | Multipotent Progenitor |
| MS | Mass Spectrometry |
| MSC | Mesenchymal Stromal Cell |
| MSI | Mass Spectrometric Imaging |
| nanoPOTS | Nanodroplet Processing In One Pot For Trace Samples |
| NegELSK | Elsks From Non-Repopulating Clones |
| NegNonELSK | Nonelsks From Non-Repopulating Clones |
| NoMO | No Molecular Overlap |


| OPN | Osteopontin |
| :---: | :---: |
| PBS | Phosphate Buffered Saline |
| PC | Principal Component |
| PCA | Principal Component Analysis |
| PCA | Principal Component Analysis |
| PCR | Polymerase Chain Reaction |
| PLG | Plasminogen |
| PosELSK | Elsks From Repopulating Clones |
| PosNonELSK | Nonelsks From Repopulating Clones |
| Progs | Progenitor Cell Populations |
| PSM | Peptide Spectral Matches |
| PTM | Post-Translational Modification |
| PV | Polycythaemia Vera |
| PVA | Polyvinyl Alcohol |
| qPCR | Quantiative Polymerase Chain Reaction |
| RepopSig | Repopulation Signature |
| RP | Reversed-Phase |
| SCF | Stem Cell Factor |
| scMS | Single Cell Proteomics |
| SCoPE-MS | Single Cell Proteomics By Mass Spectrometry |
| scRNA-seq | Single Cell RNA Sequencing |
| SDS | Sodium Dodecyl Sulphate |
| SIMS | Secondary-Ion Mass Spectrometry |
| SP | Side Population |
| SPF | Specified Pathogen Free |
| SR | Self-Renewal |
| SR1 | Stemregenin1 |
| ST | Short-Term |
| TCA | Trichloroacetic Acid |
| TCA | Trichloroacetic Acid |
| TCEP | Tris-2-Carboxymethyl Phosphine |
| TF | Transcription Factor |
| TGM2 | Transglutaminase 2 |
| TLR | Toll Like Receptor |
| TLR4 | Toll-Like Receptor 4 |
| TMT | Tandem Mass Tag |
| TPO | Thrombopoietin |
| VEGF | Vascular Endothelial Growth Factor |
| VWF | Von Willebrand Factor |
| WBC | White Blood Cell |
| WT | Wild-Type |

## Table of Contents

Chapter 1. Introduction ..... 1
1.1. Haematopoietic stem cells and their progeny ..... 1
1.1.1. Haematopoietic hierarchy ..... 1
1.1.2. The developmental origins of HSCs ..... 5
1.1.3. Isolation of haematopoietic stem cells ..... 6
1.2. Understanding cell fate in heterogeneous HSC populations ..... 8
1.3. The HSC niche ..... 12
1.3.1. Bone marrow structure and resident cell types ..... 12
1.3.2. Localisation of HSCs in the bone marrow niche ..... 14
1.3.3. The role of structural extracellular matrix proteins in the HSC niche ..... 15
1.4. HSC self-renewal regulation ..... 16
1.4.1. Extrinsic factors mediating HSC self-renewal ..... 16
1.4.2. Intrinsic molecular pathways governing HSC self-renewal ..... 19
1.5. Ex vivo expansion and maintenance of HSCs ..... 22
1.5.1. Feeder cell co-culture systems ..... 23
1.5.2. Expression of transgenes ..... 24
1.5.3. Replicating the 3D HSC niche in vitro ..... 25
1.5.4. Cytokines promoting ex vivo HSC expansion ..... 25
1.5.5. Culturing in presence of small molecules ..... 26
1.5.6. Best performing murine culturing system to date ..... 27
1.6. The role of HSCs in haematological malignancies ..... 28
1.6.1. Common driver mutation contributing to MPN disease ..... 28
1.6.2. TET2 deficiency and the acquired stem cell advantage ..... 30
1.6.3. Clinical significance of TET2 deficiency in MPNs ..... 31
1.7. Proteomic screening of rare cell populations and its importance in HSC biology ..... 33
1.7.1. The advent of proteomic mass spectrometry ..... 33
1.7.2. Proteomics on rare cell populations ..... 35
1.7.3. The HSC Proteome ..... 40
1.8. Thesis Aims ..... 42
Chapter 2. Material and Methods ..... 44
2.1. Mice ..... 44
2.2. Isolation of haematopoietic stem cells ..... 44
2.2.1. Bone marrow harvesting ..... 44
2.2.2. Red cell lysis ..... 45
2.2.3. Lineage depletion ..... 45
2.2.4. Isolation strategy of ESLAM-Sca1+ HSCs ..... 45
2.2.5. Isolation strategy of CD150 ${ }^{+}$and CD150 cell fractions ..... 47
2.2.6. Isolation strategy for ELSK cells post ex vivo expansion ..... 49
2.3. Ex vivo HSC expansion cultures ..... 50
2.3.1. Stemspan 10-day HSC maintenance cultures ..... 50
2.3.2. F12-based 28 -day expansion cultures ..... 50
2.3.3. F12-based 28 -day cultures, supplemented with signalling pathway modulators ..... 51
2.3.4. F12-based 10-day expansion cultures ..... 51
2.3.5. Hibernation cultures ..... 51
2.3.6. Functionalised Stembond cultures ..... 51
2.4. Flow cytometry analysis of ex vivo expansion cultures ..... 52
2.5. Gene expression profiling by RT-qPCR ..... 55
2.6. Calcium imaging ..... 56
2.7. Transplantation assays ..... 56
2.7.1. Peripheral blood analysis ..... 57
2.8. RNA sequencing ..... 58
2.8.1. Bulk RNA-seq library preparation and sequencing ..... 58
2.8.2. Bulk RNA-seq raw data processing ..... 59
2.8.3. Single cell RNA-seq library preparation and sequencing ..... 59
2.8.4. Single cell RNA-seq raw data processing ..... 60
2.9. Proteomic Analysis ..... 60
2.9.1. LC-MS/MS Sample Preparation ..... 60
2.9.2. Peptide fractionation ..... 61
2.9.3. LC-MS/MS Analysis ..... 61
2.9.4. Protein identification and quantification ..... 62
2.10. RNA-seq data analysis ..... 62
2.10.1. Published datasets ..... 63
2.10.2. Bulk RNA-seq quality control and processing ..... 63
2.10.3. Bulk RNA-seq batch correction ..... 63
2.10.4. scRNA-seq quality control and processing ..... 63
2.10.5. Differential gene expression analysis ..... 64
2.10.6. Gene signature scoring ..... 65
2.10.7. Principal Component analysis ..... 65
2.10.8. Correlation scoring using SingleR ..... 66
2.10.9. GO term and KEGG pathway analysis ..... 66
2.10.10. Reactome pathway analysis. ..... 66
2.10.11. Gene Set Enrichment Analysis ..... 66
2.10.12. scRNA-seq landscape projections ..... 67
2.10.13. Direction of Transition (DoT) Scoring ..... 67
2.10.14. Logistic and linear regression models ..... 68
2.10.15. Pairwise gene correlations ..... 68
2.11. Proteomic data analysis ..... 68
2.11.1. Protein interaction network analysis and data visualisation ..... 68
2.11.2. Differential protein expression ..... 69
2.11.3. GO term, KEGG and Reactome pathway analysis ..... 69
2.11.4. Principal component analysis and hierarchical clustering ..... 69
2.11.5. Proteome-Transcriptome correlation ..... 69
2.11.6. Other statistical analyses ..... 70
Results ..... 71
Chapter 3. Adapting multiplex LC-MS/MS to facilitate proteomic screening of rare cell populations ..... 71
3.1. TMT labelling provides sufficient accuracy to enable down-scaling of cell numbers ..... 73
3.2. Current TMT-based LC-MS/MS protocols do not support screening of small sample sizes ..... 74
3.3. TCA protein precipitation reduced protein loss throughout the sample handling process ..... 75
3.4. Introduction of a BSA spike-in improves protein quantification and reduces stochastic variability ..... 77
3.5. Introduction of carrier populations to enhance protein detection ..... 79
3.6. Offline fractionation maximised protein quantification and reproducibility ..... 83
3.7. Peptide fractionation improved identification of DNA-binding proteins ..... 87
3.8. Proteomics identifies ECM proteins and metabolic regulators, previously not detected by scRNA-seq ..... 89
3.9. The proteome provides a broader dynamic range than the transcriptome ..... 91
3.10. The LC-MS/MS protocol, optimised for small cell number experiments, reliably identifies key marker genes of Hoxb8-FL cells ..... 93
3.11. Discussion ..... 95
Chapter 4. Simultaneous proteomic and transcriptomic profiling to characterise molecular networks governing HSC self-renewal ..... 98
4.1. Experimental setup for multi-omic screening to decipher HSC self-renewal. ..... 100
4.2. Small cell number LC-MS/MS provides reproducible proteomic profiles of HSC populations ..... 101
4.3. The CD150-LT-HSC compartment holds most molecular differences between WT and TET2 ${ }^{-/}$animals ..... 104
4.4. Poor correlation between the proteome and transcriptome of LT-HSCs ..... 106
2.5. Proteomics identifies new pathways enriched in TET2-/ HSCs ..... 109
4.6. Intracellular calcium regulators, inflammatory response, extracellular matrix and cytoskeletal proteins are deregulated in TET2 $^{-/}$HSCs ..... 111
4.7. Single cell RNA-seq revealed enrichment of cell division and SUMOylation in TET2-/ HSCs ..... 114
4.8. TET2 knockout induces remodelling of HSC extracellular matrix proteins ..... 118
4.9. Bulk and scRNA-seq profiles of ECM-associated genes ..... 122
4.10. Ex vivo cultures of LT-HSCs on singular ECM proteins did not significantly alter HSC expansion ..... 125
4.11. Intracellular calcium binding proteins are deregulated in TET2 $^{-1}$ HSCs ..... 127
4.12. HSC and progenitors express distinct sets of calcium regulators ..... 130
4.13. Haematopoietic transcription factors are deregulated in TET2 ${ }^{-/}$HSCs ..... 131
4.14. Shear stress and ECM proteins are enriched in CD150 ${ }^{+}$TET2 ${ }^{-/}$cells but not their CD150 counterparts ..... 133
4.15. Niche interactors are enriched in CD150+ WT cells but not their CD150 counterparts ..... 137
4.16. Fluid shear stress and ECM regulation are ubiquitously enriched in cell populations harbouring increased self-renewal potency ..... 140
4.17. Discussion ..... 143
Chapter 5. Identifying drivers of HSC self-renewal and repopulation potency in ex vivo expansion ..... 147
5.1. EPCR and Fgd5 mark functional HSCs in ex vivo expansion cultures ..... 148
5.2. Deciphering clonal heterogeneity of long-term HSC expansion cultures ..... 152
5.3. Repopulating clones have distinct gene expression profiles ..... 157
5.4. Non-repopulating clones are largely comprised of differentiated progenitors ..... 159
5.5. Ex vivo expanded ELSKs retain the core HSC molecular profiles ..... 161
5.6. Non-repopulating clones lose expression of HSC-specific genes ..... 166
5.7. Erythrocytes and Megakaryocytes are enriched in non-repopulating clones ..... 168
5.8. Deriving a gene signature for repopulating HSCs ..... 172
5.9. Curation of the repopulation signature using linear models ..... 176
5.10. Extension of the repopulation gene signature using highly correlated genes ..... 179
5.11. The repopulation signature can stratify clones by HSC content ..... 183
5.12. Inhibition of Rho GTPases collapsed HSC clones, but over-stimulation did not enhance HSC proliferation ..... 184
5.13. Prospective identification of successful ex vivo expansion cultures using RepopSig ..... 186
5.14. RepopSig marks functional HSCs across multiple cellular states ..... 190
5.15. Discussion ..... 196
Chapter 6. Discussion ..... 200
6.1. Summary of major findings ..... 200
6.2. The advent of rare cell and single cell proteomics ..... 201
6.3. The HSC proteome ..... 202
6.4. Applications of molecularly defined ex vivo expanded HSCs ..... 203
6.5. Therapeutic utility of ex vivo expanded HSCs ..... 204
6.6. Future Studies and concluding remarks ..... 205
References ..... 208

## Table of Figures

Figure 1.1. The evolution of the haematopoietic lineage tree over the past three decades ..... 4
Figure 1.2. Ontogeny of murine HSCs. ..... 6
Figure 1.3. Lineage bias of different HSC subtypes ..... 10
Figure 1.4. HSC bone marrow niche ..... 13
Figure 1.5. The exponentially increased complexity of the genotype-phenotype paradigm. ..... 36
Figure 2.1. Representative gating strategy for ESLAM-Sca1+ HSCs ..... 46
Figure 2.2. Representative gating strategy for CD150 ${ }^{+}$and CD150 ${ }^{-}$HSCs ..... 48
Figure 2.3. Representative gating strategy for ex vivo cultured ELSK HSCs ..... 49
Figure 2.4. Representative gating strategy for flow cytometry analysis of ELSK cells cultured on plastic ..... 54
Figure 2.5. Representative gating strategy for flow cytometry analysis of ELSK cells cultured on functionalised Stembond hydrogels ..... 55
Figure 2.6. Representative gating strategy for peripheral blood flow cytometry analysis ..... 58
Figure 3.1. Detected protein abundance is proportional to cell number ..... 74
Figure 3.2. The standard TMT multiplex protocol does not suffice for the application to low cell number experiments ..... 75
Figure 3.3. Bulk label-free LC-MS/MS shows efficacy of TCA protein precipitation ..... 77
Figure 3.4. Introduction of a BSA spike-in reduces sampling stochasticity ..... 78
Figure 3.5. Carrier populations within a 10-plex experiment increase the number of identified proteins ..... 81
Figure 3.6. The carrier approach permits reduction of the required cell numbers per 10-plex experiment ..... 82
Figure 3.7. Offline fractionation prior to LC-MS/MS boosts protein identification and reduces data variability ..... 84
Figure 3.8. Optimised sample handling reduces sampling errors across biological samples ..... 86
Figure 3.9. Workflow of the final small cell number LC-MS/MS protocol. Diagram of the final small cell number proteomics protocol ..... 87
Figure 3.10. Protein Interaction network of DNA-binding proteins reveals consistent protein identification across samples ..... 88
Figure 3.11. The proteome of Hoxb8-FL cells correlates with the transcriptome, but permits identification of unique proteins, not detected by scRNA-seq. ..... 89
Figure 3.12. Proteins overrepresented in the proteome are enriched for extracellular matrix proteins and metabolic regulators ..... 90
Figure 3.13. Dynamic range of LC-MS/MS and RNA-seq data. ..... 92
Figure 3.14. Hoxb8-FL cells are primed for myeloid and lymphoid differentiation ..... 93
Figure 3.15. The small cell number LC-MS/MS protocol reliably identifies key marker genes of Hoxb8-FL cells.95
Figure 4.1. Haematopoietic cell types and their self-renewal potency ..... 99
Figure 4.2. Sample acquisition strategy for LC-MS/MS and bulk RNA-seq ..... 101
Figure 4.3. Small cell number proteomics captures most of the previously described HSC proteome ..... 103
Figure 4.4. LC-MS/MS and RNA-seq profiles are reproducible across biological replicates ..... 104
Figure 4.5. Proteomic and transcriptomic profiling reveals distinct molecular states between WT and TET2-/- HSCs ..... 105
Figure 4.6. TET2 knockout predominantly alters protein and mRNA profiles of finite HSCs ..... 106
Figure 4.7. LC-MS/MS and RNA-seq capture unique molecular profiles of WT and TET2-/- HSCs . ..... 108
Figure 4.8. Transcriptomic profiling, but not LC-MS/MS, preferentially identifies signalling pathways inTET2-/- HSCs110
Figure 4.9. WT and TET2-/- HSCs differ in extracellular matrix proteins and intracellular calcium regulators ..... 112
Figure 4.10. Protein network analysis reveals enrichment of cytoskeletal, inflammatory and cell cycle proteins in TET2-/- HSCs ..... 113
Figure 4.11. Single cell transcriptomics reveals distinct molecular profiles of WT and TET2-/- HSCs ..... 115
Figure 4.12. Knockout of TET2 induces mitosis and SUMOylation in HSCs ..... 117
Figure 4.13. Differentially expressed ECM proteins between WT and TET2-/- HSCs ..... 119
Figure 4.14. TET2 knockout HSCs are enriched for collagens and their respective integrin binding partners ..... 120
Figure 4.15. ECM protein abundance changes are proportional to the self-renewal potency ..... 122
Figure 4.16. ECM genes are differentially expressed between WT and TET2-/- HSCs ..... 124
Figure 4.17. In vitro expansion of HSCs in presence of ECM proteins ..... 126
Figure 4.18. Intracellular Ca2+ regulators are deregulated in TET2-/- HSCs ..... 128
Figure 4.19. Single cell transcriptomes do not corroborate LC-MS/MS-derived calciome deregulation in TET2-/- HSCs ..... 129
Figure 4.20. TET2-deficient HSCs are desensitised towards extracellular calcium ..... 130
Figure 4.21. The calciome of HSCs differs from progenitor cell populations ..... 131
Figure 4.22. Key haematopoietic transcription factors are downregulated in TET2-/- HSCs ..... 132
Figure 4.23. Comparison of proteome and transcriptome for CD150+ and CD150- TET2-/- cells ..... 134
Figure 4.24. GO term enrichment of genes enriched in TET2-/- CD150+ HSCs ..... 135
Figure 4.25. KEGG pathway enrichment of genes enriched in TET2-/- CD150+ HSCs ..... 137
Figure 4.26. Comparison of proteome and transcriptome for CD150+ and CD150- WT cells ..... 138
Figure 4.27. GO term enrichment of genes enriched in WT CD150+ HSCs. ..... 139
Figure 4.28. KEGG pathway enrichment of genes enriched in WT CD150+ HSCs ..... 140
Figure 4.29. Overlap between WT-derived and TET2-/- -derived CD150+ HSC functions revealed enrichment of shear stress and ECM function ..... 141
Figure 5.1. EPCR and Fgd5-ZsGreen allow identification of cycling HSCs ..... 150
Figure 5.2. EPCR and Fgd5-ZsGreen mark ex vivo expanded functional HSCs. ..... 152
Figure 5.3. Approach to inspect clonal Heterogeneity of clonally expanded HSCs ..... 154
Figure 5.4. Proportion of phenotypic HSCs in heterogeneous clonal cultures ..... 155
Figure 5.5. FELSK correlates with the repopulation outcome of ex vivo expanded HSCs ..... 156
Figure 5.6. Batch correction and pre-processing of bulk RNA-seq data ..... 158
Figure 5.7. Repopulating clones are enriched for LT-HSC gene signatures ..... 160
Figure 5.8. Differential gene expression between ELSKs and nonELSKs. ..... 161
Figure 5.9. Go term enrichment of ELSK and nonELSK fractions ..... 163
Figure 5.10. Pathway analysis of ELSK and nonELSK fractions ..... 165
Figure 5.11. Gene expression profiling of PosELSKs and NegELSKs ..... 167
Figure 5.12. Repopulating and non-repopulating clones are enriched for distinct NonELSK cell types ..... 169
Figure 5.13. Clonally expanded HSCs resemble freshly isolated HSCs ..... 171
Figure 5.14. Repopulating and non-repopulating cell fraction cluster separately ..... 173
Figure 5.15. Deriving a molecular signature for repopulation potency using PCA. ..... 174
Figure 5.16. Transplantation metadata is correlated with principal components ..... 175
Figure 5.17. PCA loading plot reveals genes correlated with the transplantation outcome ..... 176
Figure 5.18. Regression analysis reveals a core repopulation gene signature ..... 177
Figure 5.19. Manual curation of regression models reveals additional repopulation-assoiated genes ..... 180
Figure 5.20. Curated repopulation signature for HSCs ..... 181
Figure 5.21. Repopulation signature provides an improved separation of repopulating samples ..... 183
Figure 5.22. Signalling pathways enriched in single cell-derived clones with repopulation capacity. ..... 185

Figure 5.24. Repopulation signature genes are uniquely expressed in clones with functional HSCs 188 Figure 5.25. Repopulation Signature genes selectively mark clones with repopulation potential...... 189
Figure 5.26. Repopulation signature selectively marks freshly isolated LT-HSC ............................. 191
Figure 5.27. Repopulation signature identifies foetal liver and hibernating HSCs. ........................... 192
Figure 5.28. Repopulation signature reduces HSC identification bias towards freshly isolated cell
populations......................................................................................................................... 193
Figure 5.29. Single cell gene expression profiling of 28 -day HSC expansion cultures reveals that the repopulation Signature marks the HSC cluster............................................................................ 194
Figure 5.30. Repopulation signature genes are enriched in the HSC cluster.................................. 195

## List of Tables

Table 1. ESLAM-Sca1+ panel antibodies ..... 46
Table 2. CD150+ and CD150- HSC panel antibodies ..... 47
Table 3. ELSK panel antibodies ..... 49
Table 4. ELSK panel antibodies for flow cytometry ..... 53
Table 5. Antibody panel for peripheral blood analysis ..... 57
Table 6. Genes associated with fluid shear stress and enriched in WT and TET2-deficient CD150+ HSCs, as identified by the proteome and transcriptome ..... 142

## Chapter 1. Introduction

### 1.1. Haematopoietic stem cells and their progeny

### 1.1.1. Haematopoietic hierarchy

The blood system is one of the best studied adult tissues and has widely been used for understanding the function of adult mammalian stem cells ${ }^{1}$. Many blood cell types are short lived and thus the body requires daily regeneration of approximately one trillion cells ${ }^{2}$. These cells are classified into two major classes, the lymphoid and myeloid cells (or "lineages"). The lymphoid lineage includes T-, B-, natural killer and plasmacytoid dendritic cells, while granulocytes, macrophages, megakaryocytes, dendritic cells and erythrocytes are classified a myeloid cells ${ }^{3}$. Haematopoietic stem cells (HSCs) are a rare population of cells residing in the adult bone marrow (BM) and are responsible for lifelong maintenance of the entire haematopoietic system. Sitting at the apex of this hierarchically organised tissue, HSCs differentiate into highly proliferative precursors, which subsequently undergo further step-wise differentiation and maturation to ultimately produce terminally differentiated progeny ${ }^{2,4}$. To maintain the stem cell compartment and produce differentiated progeny, HSCs must on average create one daughter HSC and one non-HSC although any single HSC may divide symmetrically or asymmetrically with respect to stem cell potential, yielding 0 , 1, or 2 daughter $\mathrm{HSCs}^{5,6}$.

Independent work by Ford et al. and Jacobson et al. provided the first formal evidence of a common origin for the myeloid and lymphoid lineage ${ }^{7,8}$. Subsequently, the first successful bone marrow transplantation in patients confirmed the presence of selfrenewing and multipotent HSCs that could sustain lifelong haematopoiesis in a recipient patient receiving these donor cells ${ }^{9,10}$. Shortly after, Till and McCulloch developed the spleen colony assay, which showed that small BM samples were able to form haematopoietic tissue in the form of serially transplantable and multi-lineage colonies in irradiated recipient mice. They used this quantitative assay to formally demonstrate that the multiple mature cell types derived from a single cell were capable
of self-renewal by transplanting into secondary recipient mice ${ }^{11,12}$, paving the way for modern assessments of stem cell function in vivo. Together, these studies confirmed the presence of a rare HSC ( $\sim 1$ in 10,000 BM cells), capable of reconstituting the entire spectrum of haematopoietic cell types ${ }^{13}$, thus raising wide interest in their clinical application. The next decades resulted in a plethora of studies attempting to isolate pure HSC populations and, in 1996, Osawa et al. performed single cell transplantations and provided the first formal evidence that a single cell could carry out long-term multilineage reconstitution in recipient animals on its own, while the vast majority of other progenitor cells could not sustain haematopoiesis over a longer period of time ${ }^{14}$. This work collectively led to descriptions of long-term (LT) and shortterm (ST) HSCs, where ST-HSCs were multi-lineage but transient in a transplantation setting.

Since the original descriptions of this haematopoietic hierarchy, new evidence has led to several conceptual alterations of the traditional stick and ball diagrams that imply a step-wise progression through defined cell compartments. Originally, a simple bifurcation model was proposed (Figure 1) where LT-HSCs sequentially progress through to ST-HSCs and subsequently multipotent progenitors (MPPs) ${ }^{15}$. MPPs then differentiate into either a common lymphoid progenitor (CLP), which in turn gives rise to more restricted lymphoid progenitors, or a common myeloid progenitor (CMP), capable of forming cells of the myeloid lineage ${ }^{16,17}$. Most recently however, differences in lineage biases amongst the LT-HSC population were discovered ${ }^{18,19}$, highlighting the possibility that lineage decisions might be occurring at the very top of the hierarchy (Figure 1) and supposed cellular intermediates may be bypassed. These different HSC subtypes were shown to differentially give rise to CMPs or a lymphoid-primed multipotent progenitor (LMPP) via ST-HSCs ${ }^{15}$. Even more dramatically, megakaryocyte-biased HSCs were recently identified, where megakaryocyte progenitors could be derived directly from the LT-HSC pool, completely by-passing the classical differentiation cascade ${ }^{20}$. However, others have argued that this observation could in fact result from the contamination of the HSC compartment by long-lived megakaryocyte progenitors with finite self-renewal capabilities ${ }^{21}$. Most recently, the advent of single-cell transcriptomics allowed effective capture of the molecular states of single cells at different stages throughout the haematopoietic hierarchy ${ }^{22}$. As a result, several groups observed a range of molecularly heterogeneous haematopoietic
stem/progenitor cells (HSPCs) rather than distinct populations of hierarchically arranged naïve haematopoietic cell (extensively reviewed in ${ }^{23}$ ). Lineage commitment and formation of mature cell types subsequently results from a continuous acquisition of lineage-specific traits. Despite significantly expanding our horizon of understanding HSC development and the wide-range of potential cellular states, the technical limitations of single cell RNA sequencing (scRNA-seq) still leave a number of questions about haematopoietic development open for further investigation. For instance, ribosome depletion, polyA tail pulldowns and polymerase chain reaction (PCR)-induced amplification during sample preparation can introduce sampling and amplification biases ${ }^{24}$. Similarly, gene length biases have also been observed ${ }^{25}$. In addition, technical limitations paired with the rare nature of HSCs have limited our ability to interrogate the proteome and the associated post-translational modifications (PTMs).


Figure 1.1. The evolution of the haematopoietic lineage tree over the past three decades. (A) First bifurcation model of haematopoiesis. The model depicts stepwise progression along either myeloid or lymphoid trajectories, with mutually exclusive commitment to a single lineage. (B) Adapted model, accounting for observed heterogeneity and a fluid continuum within the stem and multipotent progenitor compartment. Furthermore, several downstream progenitors, such as the lymphoid-primed multipotent progenitor (LMPP) population retained some multi potent potential. (C) Most recent single cell transcriptomic profiling of the haematopoietic hierarchy revealed a continuum of lineage commitment, with single cells able to "travel" along fluid differentiation trajectories. LT-HSC, long-term HSCs; STHSC, short-term HSC; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; LMPP, lymphoid-primed multipotent progenitor; EoBP, eosinophil-basophil progenitor; GMP, granulocyte and macrophage progenitor; MEP, megakaryocyte and erythroid progenitor. The figure was adapted from Laurenti and Göttgens ${ }^{26}$.

### 1.1.2. The developmental origins of HSCs

Early mammalian haematopoietic development initiates in the yolk sac at E7.5, before haematopoietic cells migrate into the aorta-gonad-mesonephros (AGM) region by 10 days post coitum $(\mathrm{dpc})^{27,28}$. Here, the first HSCs with both long-term reconstitution capacity and the ability to form spleen colony forming units (CFU-S) are formed ${ }^{27}$. Transplantation of early haematopoietic cells and embryonic cells, derived from preE10 mouse embryos conversely showed a striking inability to reconstitute the haematopoietic system long-term ${ }^{29}$. At the E10.5 stage of mouse development, early HSCs reside in the aortic endothelium of the foetal liver (FL) and the AGM, where they undergo extensive expansion via symmetric self-renewal to yield definitive HSCs ${ }^{30}$. Following the symmetric expansion in the FL, HSCs migrate into the thymus, spleen and BM after E12 ${ }^{31}$. In the BM, between 3 and 4 weeks of age, HSCs adopt an adult state where they shift from a proliferating self-renewal state to a quiescent LT-HSC population ${ }^{32}$. The population then shifts into homeostatic maintenance where, on average, adult LT-HSCs undergo asymmetric division to produce one HSC and one multi-potent progenitor, the latter of which subsequently enters the haematopoietic differentiation cascade to form large numbers of mature blood cells ${ }^{1}$ (Figure 2). This underscores the importance of cell fate decisions in adult organisms for maintenance of tissues homeostasis.


Figure 1.2. Ontogeny of murine HSCs. Figure was adapted from Mikkola and Orkin ${ }^{33}$. Timeline of HSC development in the mouse. Functional and developmental aspects (bars) are indicated for each developmental location: grey, mesoderm development; red, HSC differentiation; yellow, HSC genesis; blue, functional HSCs are present.

The differences in self-renewal capacity between FL and adult HSCs imply differences in the underlying molecular regulatory networks. To identify the molecular drivers of self-renewal in HSCs, the transcriptomes of developing HSCs have been compared to that of adult $\mathrm{HSCs}^{34}$ and the unbiased gene network analysis revealed key differences in gene expression between FL and BM-derived adult HSCs. Such comparisons require reliable isolation of FL HSCs, which were also shown to be enriched (>1 in 5) within the ESLAM immunophenotypic fraction ${ }^{35}$. Hmga2 and Sox17 have been implicated in FL HSC specification, with Sox17 being a key determining factor for FL HSC identify and Hmga2 a key regulator of self-renewal ${ }^{36,37}$.

### 1.1.3. Isolation of haematopoietic stem cells

In order to understand the biology of HSCs and to exploit their regenerative potential for therapeutics, isolating pure populations is of great importance. Early efforts to track the progeny of HSCs involved the use of retroviral gene transfer methods where
patterns of integration sites could be tracked throughout serial transplantation studies to chart the progeny of single stem cells ${ }^{38-40}$. The most rapid advances came with the advent of multi-parameter flow cytometry where the expression (or not) of specific cell surface markers could be used in tandem to greatly accelerate the prospective isolation of functional HSCs. Antibodies raised against a set of mature cell markers of the myeloid and lymphoid lineages (Lin), differentiation-specific Thy-1 and the surface protein Sca-1 represented a major breakthrough for stem cell isolation with the Thy$1^{10} \mathrm{Lin}-\mathrm{Sca}-1^{+}$population greatly enriching for HSCs from other BM cells, despite the overall purity of the HSCs remaining at $\sim 10 \%{ }^{41-44}$. The Stem Cell Factor (SCF)receptor c-Kit was then found to be constitutively expressed on haematopoietic progenitors and the combination of c-Kit with Thy- $1^{\text {lo }} \mathrm{Lin}-\mathrm{Sca}-1^{+}$(subsequently termed KSL) further increased HSC purity ${ }^{45,46}$. Of note however, the evidence was based on primary transplantation experiments only and therefore likely represented an overestimate of functional HSC frequency. A landmark paper by Osawa and colleagues revealed enrichment of HSCs based on CD34 expression ${ }^{14}$ and it was demonstrated that the CD34 ${ }^{10 /-}$ fraction of murine BM cells contained high numbers of LT-HSCs, which were able to sustain multilineage reconstitution ability in single cell serial transplantation assays. ST-HSCs were shown to be present in the complementary CD34 ${ }^{+}$fraction since they could not maintain reconstitution long-term or in serial transplantations.

Together with increasing functional evidence for different HSC populations, further efforts to optimise purification strategies were focused on isolating LT-HSCs with the greatest serial reconstitution ability. KSL-isolated HSCs and MPPs differentially express member of the SLAM family of surface proteins, most notably CD48, CD150 and CD244. Indeed, CD150+CD48- -based isolation showed enrichment for LTHSCs ${ }^{47}$. Similarly, Endothelial Protein C Receptor (EPCR) enriches for LT-HSCs ${ }^{48,49}$. The combination of KSL-SLAM or EPCR-SLAM (EPCR ${ }^{+}$CD150 ${ }^{+}$CD45 ${ }^{+}$CD48; abbreviated ESLAM) and Sca- $1^{\text {bright }}$ yielded a purity of LT-HSCs above $50 \%$, based on the ability to sustain the haematopoietic system in single cell serial transplants ${ }^{35,50,51}$. Other distinguishing cell surface markers have been described, including the functional dyes Rhodamine123 and Hooechst 33342 staining, which reveals the drug-effluxing side population (SP) of cells, CD49b, ESAM, and Robo152,53.

While these advances were significant for understanding murine HSC biology, the lack of similarly robust assays for detecting LT-HSCs in human samples has hindered progress. The advent of xenograft models were a major advance, leading to improved discrimination in the detection of human HSCs, but these remain troubled by the inability to determine relative lineage contributions and the lack of a completely humanised micro-environmental niche ${ }^{54}$.

Notwithstanding these limitations, the majority of human HSCs share the expression of CD34. Additional isolation of human cord blood cells based on CD49f+CD38-CD45RA- increased the frequency of HSCs to $\sim 10 \%{ }^{55}$. Unfortunately, HSC purification from other sources, such as human adult BM and peripheral blood have significantly lower efficiency. As a result, detailed and accurate studies of HSC heterogeneity are restricted to the mouse HSC model where robust functional assays at the single cell level exist and purities exceed $50 \%$, making it possible to characterise the molecular drivers of cell fate.

### 1.2. Understanding cell fate in heterogeneous HSC populations

As described above, advances in describing molecular profiles of single HSCs and the diversity of functional output observed in a multitude of studies confirmed the presence of functional heterogeneity within the HSC compartment. Early attempts to define HSCs revealed differences in proliferation efficacy and phenotype of daughter cells within CFU-S ${ }^{12}$ and was the first indication of heterogeneity amongst the HSC population. Despite this, researchers broadly considered cells that met the criteria for long-term reconstitution were more or less the same as one another ${ }^{56}$. However, technological advances over the past decade, particularly in single cell biology, allowed for the formal description, and in some cases prospective isolation, of different HSC subtypes, both in vivo and in ex vivo colony forming assays ${ }^{57}$. Following evidence on the population level, it became apparent that cellular decisions are made by single, individual $\mathrm{HSCs}^{58}$. Aside from self-renewal capacity, other factors, such as life span, lineage bias, cell cycle status, and repopulation pattern, all contribute to distinct HSC states ${ }^{56,59-61}$.

Muller-Sieburg and colleagues were first to observe that a considerable proportion of clones derived from serial transplants of HSCs in limiting dilution assays, showed a lineage bias toward either the lymphoid or myeloid lineage ${ }^{58}$. Intriguingly, these clones also maintained the same self-renewal capacity and lineage bias throughout a series of serial transplants. Thus, the authors concluded that lineage biases and self-renewal capacity are inherited by the progeny and intrinsically regulated ${ }^{58,62}$. Furthermore, this study provided the first classification of HSCs into three subtypes, namely the lymphoid-biased, myeloid-biased and balanced HSC population. An independent study by Dykstra et al. provided formal support for this classification by performing a large number of single cell transplantations of freshly isolated and cultured clones of HSCs ${ }^{19}$. Here, the authors developed a slightly more intricate classification of HSC subtypes where single CD45 mid_in-Rho-SP-isolated cells ( 352 cells), of which at least 1 in 4 met long term reconstitution criteria ${ }^{63,64}$, were transplanted into irradiated mice, followed by a secondary and tertiary transplantation. White blood cell (WBC) counts were monitored throughout, and the contribution of the transplantation was determined. The authors observed four distinct patterns, namely $\alpha, \beta, \gamma$ and $\delta$, based on their lineage contribution and self-renewal durability (Figure 3). Here, $\alpha$-HSCs primarily give rise to myeloid cells, while $\beta$-HSCs appeared to produce a balanced lymphoid and myeloid output. Both HSC subtypes also sustained significant (> 1\%) contribution to the WBC counts beyond 16 weeks in secondary and tertiary transplants, indicative of long-term sustained self-renewal durability. In contrast, the contribution of $\gamma$ - and $\delta$-HSCs was prematurely exhausted, suggesting finite amounts of HSC self-renewal despite meeting the traditional criteria for an LT-HSC. Furthermore, $\gamma$ - and $\delta$-HSCs were generally lymphoid-biased, whereby $\delta$-HSCs contribute exclusively to the lymphoid lineage which may reflect the longer lifespan of some lymphoid cell types as opposed to active production of haematopoietic cells by the transplanted HSCs ${ }^{19,65}$. Interestingly, the proportion of HSC subtypes shifts throughout life. During HSC expansion in the foetal liver of the mouse embryo, $\beta$ HSCs form the dominant subtype ${ }^{50}$. However, after birth and throughout lifetime, $\alpha$-HSCs become increasingly dominant, with aged mice showing an enrichment for lymphoiddeficient HSCs ${ }^{66,67}$.


Figure 1.3. Lineage bias of different HSC subtypes. Several groups provided independent evidence, which led to the classification of HSCs into three classes. The progeny of $\alpha$-HSCs is lymphoid-deficient, $\beta$-HSCs produce a balanced population and $\gamma$ and $\delta$ HSCs are both myeloid deficient. $\alpha$ and $\beta$ HSCs exclusively possess long-term self-renewal (SR) capacity and can be distinguished from $\gamma$ and $\delta H S C$ s by the presence of the CD150 surface protein and their ability to repopulate secondary recipients in transplantation assays. This enables separation of both HSC subtype groups by flow cytometry ${ }^{68,69}$.

Subsequent studies characterising HSC heterogeneity largely accorded with the Muller-Sieburg et al. and Dykstra et al. studies. Broadly speaking, two major classes emerged; LT-HSCs, capable of extensive self-renewal and contribution to repopulation for over 12 months and ST-HSC, which lack the capacity to sustain contribution and self-renewal ${ }^{56,70,71}$. Notably, this "ST-HSC" is distinct from the original ST-HSCs described in the literature and some groups opt to classify all cells meeting the traditional HSC criteria as LT-HSCs with differences in self-renewal durability. To address the nomenclature issue, Benveniste and colleagues proposed that an intermediate (IT-) HSC class should be introduced for the lymphoid-biased subsets of HSCs since these showed extensive - but not sustained - reconstitution of the haematopoietic system ${ }^{53}$. Due to this evolving nomenclature, some terms, such as STHSC, have become conflated in the literature, causing some difficulties in interpreting experiments between laboratories.

The difference in self-renewal potential and lineage bias raised the interest in isolating HSC subtypes to subsequently characterise the underlying molecular networks driving these phenotypic differences. Notably, CD150 was identified as a key surface marker,
capable of distinguishing between LT-HSCs ( $\alpha$ and $\beta$ ) and ST-HSCs ( $\gamma$ and $\delta$ ) (Figure $3)^{35,59,69,72}$. Besides CD150, other phenotypic factors were identified as distinguishing factors between HSCs, exhibiting different lineage biases. Hoechst 33343 staining of the side population (SP) ${ }^{73}$ showed that the upper HSC population showed a lymphoid bias, while HSCs in the lower region of the gate were biased for the myeloid lineage. The same study also revealed a differential response to the growth factor TGF- $\beta$, which inhibits lymphoid-biased HSCs but activates myeloid-biased HSCs ${ }^{74}$. Moreover, the SLAM family member CD229 is exclusively expressed in lymphoid-biased HSCs ${ }^{75}$. However, these approaches (excluding CD150) fail to isolate balanced HSCs ${ }^{57}$.

Despite the above efforts, isolating individual HSC subtypes in sufficient purity to determine specific molecular signatures distinguishing such subtypes remains a challenge since no single strategy isolates the different HSC subtypes exclusively ${ }^{65}$. Separation by differential CD150 expression raises the exciting prospect of defining the molecular mechanisms underlying self-renewal durability with a more than 10-fold enrichment of $\alpha-/ \beta-H S C s$ in those expressing high CD150. Such information could also greatly improve efforts to expand HSC cultures ex vivo for therapeutic use and for gene therapy applications ${ }^{76}$.

More recently, Wilson and colleagues provided one of the first unbiased approaches to dissect HSC heterogeneity on the transcriptional level to identify a molecular signature associated with long-term self-renewing HSCs ${ }^{51}$. Using single cellquantitative polymerase chain reaction (qPCR), RNA-sequencing, and transplantation data for HSCs derived from different fluorescence-activated cell sorting (FACS) purification strategies, the authors described a molecular signature, enriched amongst functional LT-HSCs ${ }^{51}$. First, the authors identified five commonly used flow cytometry sorting strategies to isolate single LT-HSC and determined the expression of 48 selected genes, previously implicated in HSC biology and cell fate regulation. Under the assumption that true HSCs will share a common genetic signature amongst the pool of contaminating cells and by accounting for the functional outputs, the authors identified a population of single cells with overlapping molecular profiles (MolO cells). The remaining cohort, which also contains phenotypic HSCs, was termed the NoMO (No molecular overlap) population. As all single cells were index sorted, it became apparent that the MolO and NoMO clusters could be prospectively isolated
by Sca-1, which is predominantly expressed in MolO cells. It was shown that cells with superior repopulation and long-term self-renewal potential were associated with higher Sca-1 expression and were indeed situated within the MoIO signature ${ }^{51}$. To extend the identification of MoIO and NoMO-specific gene signatures, scRNA-seq profiles were acquired for 92 cells, which were isolated by Lincc-Kit+Sca-1+CD34-FIt3-CD48-CD150 ${ }^{+}$. Differentially expressed genes were identified based on a MolO score, which described each cell's fit to the MolO gene expression profile and these data represent an excellent starting point for identifying candidate regulators of durable self-renewal.

### 1.3. The HSC niche

To exert their role in maintaining homeostasis of the haematopoietic system, HSCs reside in a specialised microenvironment and sustain responsiveness to extrinsic signals that module HSC fate. The concept of a stem cell niche was first developed by Ray Schofield ${ }^{77}$ who argued that the functional heterogeneity of stem cells in CFU-S assays was, in part, caused by the association of stem cells with other cell types, which were able to modulate stem cell behaviour ${ }^{77}$. Following inception of the stem cell niche concept, a plethora of studies have set out to identify the localisation of HSCs in the BM and the impact of the niche on HSC fate ${ }^{78}$. Unsurprisingly, characterisation of the HSC niche was largely conducted using mouse models, however several human BM studies have also provided insight into the human HSC niche ${ }^{79}$. Of note, human haematopoiesis predominantly occurs in the axial skeleton, while all murine bones are sites of active haematopoiesis ${ }^{78}$.

### 1.3.1. Bone marrow structure and resident cell types

HSCs predominantly reside in the BM, a heterogeneous and complex tissue comprised of haematopoietic and non-haematopoietic cell types ${ }^{78}$. Intriguingly, only 1 out of 25,000-50,000 BM cells are HSCs, thus further highlighting the importance of the interactions between HSCs and their surrounding niche ${ }^{49}$.

The marrow of long bones is highly vascularised and contains an extensive nervous network ${ }^{80}$ (Figure 4A). The endosteum coats the inner layer of the bone tissue and
interfaces with the inner BM (Figure 4A). Interestingly, the periosteum, which lines the outer layer of bone tissue, has been shown to mediate the penetration of the vascular and neuronal networks into the inner cavity of the bone ${ }^{81}$ (Figure 4A).


Figure 1.4. HSC bone marrow niche. (A) Murine long bone cross section to depict the anatomy of the bone marrow. The distribution of blood vessels and the sympathetic nervous system is shown. (B) HSCs are localised in close proximity of the sinusoids and also the endosteum, in proximity to osteoclasts. Cell types capable of modulating HSC self-renewal and stem cell maintenance are depicted. Sympathetic nerves, Mesenchymal stromal cells (MSC), Cxcl12-abundant reticular (CAR) cells and osteoclasts have been particularly implicated in HSC niche localisation and functional modulation. Figures were adapted from Pinho and Frenette ${ }^{78}$ and Morrison and Scadden ${ }^{79}$.

At the interface of bone tissue and the endosteum, osteoblasts and osteoclasts facilitate the formation of the bone tissue and its remodelling ${ }^{82}$. Here, the endosteal niche is closely associated with the trabecular bone surface and the site of increased arterial branching ${ }^{83}$. The resulting network of arterioles stretches throughout the BM cavity and forms sinusoids to mediate homing and exchange of oxygen, molecules, and cells ${ }^{84}$. To facilitate such exchange, sinusoids are comprised of a thin endothelial wall and contain a large lumen ${ }^{79,84}$. In turn, the arterial branches ultimately enter a large central artery and vein, which runs throughout the length of the cortical bone ${ }^{78,84}$. In addition, blood vessels and in particular sinusoids are lined by mesenchymal stromal cells (MSCs) ${ }^{85}$ (Figure 4B).

As briefly mentioned above, the BM lumen is also penetrated by a nervous network, in particular sympathetic nerves. The association of nerve fibres with stromal cells at
arterial walls was first characterised by Yamazaki and Allen ${ }^{86}$ (Figure 4B). Subsequently, Mendez-Ferrer et al. showed that adrenergic nerve fibres associated with MSCs at arterioles to mediate HSC fate decisions ${ }^{87}$. Pioneering work by Katayama et al. confirmed a key role of the sympathetic nervous system in mediating HSC localisation ${ }^{88}$. This latter study utilised a galactrosyltransferase-deficient mouse model, which abrogated normal nerve conduction. Interestingly, the resulting inhibition of sympathetic nerve signals inhibited HSC mobilisation, while an adrenergic agonist induced the inverse ${ }^{88}$. Collectively, these studies outlined that sympathetic nerves form an essential part of the BM microenvironment and directly impact HSC function.

### 1.3.2. Localisation of HSCs in the bone marrow niche

Historically, the HSC niche was thought to be located in close proximity to the endosteal surface of the bone, which is lined by osteoblasts ${ }^{89}$. It was reported that osteoblasts are crucial for the recruitment of HSCs and their expansion ${ }^{90}$. Pioneering work by Arai et al. identified quiescent HSC population, directly adhering to osteoblasts ${ }^{91}$. Here, quiescent HSCs ubiquitously expressed Tie2 to recognise Angiopoietin-1 (Ang-1) ligand to maintain long-term repopulation potency ${ }^{91}$. The interaction between HSCs and its niche are mediated through the formation of an N cadherin/ $\beta$-catenin adhesion complex, which is regulated by c-Myc ${ }^{92}$. Interestingly, conditional knockout of c-Myc resulted in increased N -cadherin expression and thus retention of HSCs in the niche. Moreover, the c-Myc depletion induced a shift in cell fate, with mutant HSCs failing to differentiate. Instead, LT-HSC accumulated in the BM. Thus, Wilson et al. showed that normal c-Myc expression was essential for modulating the interaction of HSCs with the microenvironment and consequently the balance between self-renewal and differentiation ${ }^{92}$. More recently, the characterisation of CD150 and CD48 as distinguishing cell surface markers for HSCs resulted in the identification of HSCs at sinusoidal blood vessels ${ }^{47}$.

While some CD150 ${ }^{+}$CD48-CD41-Lineage ${ }^{-}$HSCs were identified at the endosteum, the majority of HSC were localised at the sinusoidal endothelium (Figure 4B) ${ }^{47}$. Such affinity to the sinusoid was proposed to be mediated via the CXCL12-mediated signalling ${ }^{93}$. A conditional deletion of the complementary receptor CXCR4 resulted in
a decreased number of viable HSCs and increased exit from a quiescent state. The authors subsequently identified CXCL12-abundant reticular (CAR) cells at the sinusoidal endothelium, where they secrete the majority of CXCL12 to promote retention and quiescence of HSCs in the BM niche ${ }^{93}$.

### 1.3.3. The role of structural extracellular matrix proteins in the HSC niche

In addition to the BM architecture and the surrounding cell types, the immediate HSC microenvironment is composed of the extracellular matrix (ECM). A broad range of HSC cell surface receptors enable interaction with structural components of the niche ${ }^{94}$. Interestingly, structural ECM proteins not only ensure structural integrity of the BM but are also able to modulate HSC function ${ }^{95}$. Well-characterised and notable examples include fibronectin, osteopontin (OPN), hyaluronic acid (HA), laminins and several types of collagens, all of which are highly abundant in the $\mathrm{BM}^{96-98}$.

Fibronectin has been widely recognised as a key protein for HSC localisation and regulation ${ }^{97}$. Using immunofluorescent imaging, Nilsson et al. demonstrated concentration of fibronectin in the endosteal niche of murine long bones ${ }^{97}$. In turn, HSCs bind to fibronectin via the $\alpha 5 \beta 1$ and $\alpha v \beta 3$ integrins ${ }^{99}$. In line with such observations, Dao et al. utilised ex vivo culturing strategies in the presence of fibronectin and demonstrated that the interaction of human HSCs with fibronectin induces proliferation and long-term self-renewal capacity ${ }^{100}$.

Several collagens, namely type I, type III and type IV, have been implicated in HSC biology ${ }^{94}$. As a major structural protein of the bone, type I collagen is amongst the most abundant ECM proteins ${ }^{97}$. Oswald and colleagues constructed threedimensional type I collagen matrices for ex vivo expansion of human CD34 ${ }^{+}$cells ${ }^{101}$. Interestingly, the type I collagen scaffold induced transcriptional changes in HSCs and increased the frequency of colony-forming unit cells (CFU-Cs) ${ }^{102}$. Type IV collagen has been shown to strongly adhere to several haematopoietic cell types ${ }^{101}$.

OPN has been identified in the endosteal niche, a major site of HSC localisation. Grassinger et al. demonstrated that HSCs directly interact with OPN via $\alpha 4 \beta 1$ and
$\alpha 9 \beta 1$ integrins ${ }^{103}$. In turn, such interaction mediates the localisation of HSCs and acts as a negative regulator of HSC proliferation ${ }^{96,104}$.

Several studies have identified laminin as a key structural component of the BM niche. In brief, Gu et al. and Sagar et al. demonstrated a key role of laminin in both human and murine HSC function ${ }^{105,106}$. Human CD34+ and CD34 ${ }^{+}$CD38- stem and progenitor cell populations expressed significant affinity towards laminins via integrin $\alpha 6^{105}$. Similarly, murine HSCs were shown to directly interact with laminin in an ex vivo culturing system, which increased the cells' self-renewal potency ${ }^{106}$. A more recent study by Susek et al. confirmed these observations and showed that deficiency of laminin $\alpha 4$ inhibited HSPC circulation and cycling ${ }^{107}$.

In contrast to several other ECM proteins, HA has been shown to not only be expressed by niche cell types but also by HSCs ${ }^{108}$. In turn, HA mediates engraftment of HSCs in the endosteal niche through the interaction with the cell surface receptor CD44 ${ }^{108,109}$. Interestingly, Nilsson et al. also demonstrated that the in vivo interaction of HSCs with HA prevents proliferation and maintains the undifferentiated, naïve state of HSCs ${ }^{108}$.

### 1.4. HSC self-renewal regulation

Early studies of HSC differentiation led to the hypothesis that HSC fate was stochastic and occurred somewhat randomly ${ }^{110,111}$. However it became apparent that HSC fate choices are tightly regulated by both intrinsic molecular pathways and environmental signals, provided by the stem cell niche ${ }^{90,112}$. Key intrinsic signalling pathways and potent extrinsic modulators of HSC self-renewal will be discussed below.

### 1.4.1. Extrinsic factors mediating HSC self-renewal

As outlined above, a multitude of extrinsic factors can regulate HSC function and in particular their differentiation and self-renewal capacity. One of the most prominent niche factors that regulate HSC fate, SCF was shown to be expressed by perivascular cells in the BM niche ${ }^{113}$. Here, SCF plays a crucial role in maintaining HSC populations
by effectively influencing cell fate ${ }^{114}$. SCF has been shown to be able to inhibit the activation of apoptotic mechanisms in HSCs ${ }^{115,116}$. Interestingly, it has been shown that varying levels of SCF influence the molecular signature of HSC, whereby high levels of SCF were shown to promote asymmetric self-renewal while low SCF concentrations in ex vivo culture conditions drove HSCs towards symmetric differentiation ${ }^{49}$. The significance of SCF-mediated stimulation of HSC self-renewal is further demonstrated by a comparison of adult and actively proliferating FL HSCs. In brief, Bowie et al. observed an increased sensitivity of FL-HSCs towards SCF in both, ex vivo and in vivo settings ${ }^{117}$. The SCF ligand is bound by KIT, a receptor tyrosine kinase ${ }^{118}$. Loss-of-function mutations in the KIT receptor caused the depletion of the HSC population, a similar phenotypic response to the depletion of SCF ${ }^{119}$. In turn, depletion of SPRED1, a negative regulator of KIT, resulted in a self-renewal advantage of HSCs ${ }^{120}$. Collectively, these studies attribute a critical function to the SCF-KIT signalling pathway in regulating HSC self-renewal.

Thrombopoietin (TPO), a versatile haematopoietic cytokine, has been shown to regulate megakaryopoiesis as well as HSC self-renewal ${ }^{121}$. The secretion of TPO has largely been attributed to the liver, however modest TPO expression has also been observed in the $\mathrm{BM}^{122,123}$. Upon secretion, the TPO ligand is recognised by the complementary receptor Mpl , which is highly expressed in purified HSC populations ${ }^{124}$. Several independent studies observed severe depletion of the HSC pool in mice, carrying either TPO or MPL knockouts ${ }^{125-128}$. Collectively, these studies demonstrated the importance of TPO-induced MPL signalling for positively regulating HSC self-renewal.

Interestingly, Buza-Vidas also observed a key role for an MPL adaptor protein, LNK (SH2B3), in the regulation of HSC self-renewal ${ }^{124}$. The association of LNK with MPL dampens the downstream signalling cascade ${ }^{129}$. Here, Buza-Vidas et al. demonstrated that Lnk deficiency significantly increased HSC proportions and thus their self-renewal potency ${ }^{124}$. Ema et al. reported a particularly striking phenotype, whereby the loss of Lnk induced a 3,000-fold expansion of HSCs in vivo ${ }^{130}$. Most intriguingly, loss of Lnk function additionally induced a clonal advantage when compared to wild-type HSCs in competitive transplantation assays ${ }^{130,131}$. These
observations further highlight the importance of Mpl and the downstream signalling cascade for regulating HSC self-renewal ${ }^{132}$.

Another very prominent regulator of HSC fate, Ang-1 was shown to be expressed by osteoblasts in the BM niche to maintain the quiescent state and long-term repopulation ability of HSCs ${ }^{133,134}$. Arai et al. showed that HSCs expressing the Ang-1 receptor Tie2 localise to osteoblasts and maintain a quiescent state ${ }^{133}$.

As outlined above, the chemokine CXCL12 induced HSC exit from quiescence in the endosteal niche. Several niche cell types express or modulate CXCL12 to regulate HSC localisation ${ }^{78}$. Conditional deletion of CXCL12 in perivascular cells resulted in a substantial loss of $\mathrm{HSCs}^{135}$. Similarly, CXCL12 deletion in nestin-negative mesenchymal progenitors or endothelial cell decimated the HSC compartment and induced HSC exhaustion ${ }^{136,137}$. Strikingly, CXCL12 deletion in osteoblasts or haematopoietic cells did not affect HSC maintenance ${ }^{136,137}$. Collectively, these observations suggest that the CXCL12-mediated regulation of HSC self-renewal is niche-dependent.

TGF- $\beta$ has been implicated in the regulation of tissue homeostasis in an array of organs ${ }^{138}$. In the context of haematopoiesis, TGF- $\beta$ has been shown to maintain HSC quiescence and prevent stem cell exhaustion ${ }^{139}$. In vitro assays of human and murine HSCs showed that TGF- $\beta$ inhibited differentiation and proliferation ${ }^{140-142}$. Yamazaki et al. showed that signalling pathways, downstream of the TGF- $\beta$ receptor, were activated in quiescent HSCs ${ }^{143}$. A subsequent study by Capron et al. inspected HSC fitness in TGF- $\beta 1$ knockout mice ${ }^{144}$. The authors observed significant HSC defects, with TGF- $\beta 1$-deficent stem cells losing short-term and long-term reconstitution potency in transplantation assays ${ }^{137}$.

Several cytokines, implicated in the regulation of the inflammatory response, have also been identified as potent modulators of HSC self-renewal and differentiation ${ }^{145}$. The granulocyte colony-stimulating factor (G-CSF) protein can disrupt the CXCL12 gradient to induce HSC differentiation and mobilisation ${ }^{146}$. Subsequently, pioneering
work by Katayama et al. identified that adrenergic signalling via the sympathetic nervous system modulated G-CSF and CXCL12 levels to regulate HSC mobilisation ${ }^{88}$.

Two key inflammatory cytokines, namely interleukin 11 (IL-11) and interleukin 6 (IL-6), have also been shown to impact HSC fate ${ }^{112,145,147}$. Several independent studies showed that IL-11 induced HSC proliferation and the overexpression of IL-11 also increased HSC self-renewal potency ${ }^{148-150}$. Similarly, IL-6 has been implicated in HSC function, with deletion of IL-6 abrogating the self-renewal potency of HSCs in vivo ${ }^{151}$. Both cytokines are potent ligands for the GP130 receptor, which is ubiquitously expressed by $\mathrm{HSCs}^{78}$. In line with the phenotypes of IL-6 deficiency and IL-11 overexpression, deletion of GP130 significantly reduced the frequency of HSCs in vivo ${ }^{152,153}$.

Finally, Takizawa et al. subjected mice to lipopolysaccharide (LPS) treatment to induce innate immune responses ${ }^{154}$. The authors showed that dormant HSCs recognised LPS stimulation via toll-like receptor 4 (TLR4) and, in turn, underwent extensive myeloid differentiation and proliferation ${ }^{154}$.

### 1.4.2. Intrinsic molecular pathways governing HSC self-renewal

While HSCs can respond to a plethora of external stimuli, they also maintain an intrinsic capacity for self-renewal regulation. For instance, it was shown that the WNT and NOTCH signalling pathways are widely implicated in the regulation of HSC fate choices ${ }^{155}$. Several Notch ligands were shown to promote self-renewal and expansion of HSCs. These effects were observed in vivo but also in ex vivo HSC expansion studies, when supplemented with additional factors ${ }^{156-158}$. The Wnt pathway was also implicated in the HSC expansion through activation of the canonical pathway ${ }^{155}$. Wnt5a, which inhibits the canonical Wnt pathway, was shown to maintain HSCs in a quiescent state and also enhance the repopulation activity ${ }^{159}$. However, a more recent study reported that the Wnt pathway is largely expendable for HSC function ${ }^{160}$.

Beside the large number of studies describing cell fate-modulating cytokines and ligands of selected signalling pathways, transcriptomic analyses of HSCs revealed
intricate gene regulatory networks (GRNs) which might govern cell fate. Under the assumption that molecular pathways are conserved amongst all long-term selfrenewing HSCs at any developmental stage, the first GRN for LT-HSC was built ${ }^{161}$. The transcription factors (TFs) Gata2 and Scl have previously been implicated in haematopoiesis ${ }^{162-164}$. Pimanda and colleagues subsequently showed that Gata2, Scl and the transcription factor Fli1 form a cross-regulating triad in LT-HSCs. All three TFs bind to the cis-regulatory region of the other two components to reinforce and maintain constitutive activation ${ }^{161}$, which is essential for normal haematopoiesis ${ }^{165}$. All three TF are activated by BMP4, which also plays a crucial role in haematopoietic development ${ }^{166-168}$. Using this triad as the focal point, the overlaying and extended GRN can be constructed using the bottom-up approach ${ }^{169}$. The approach is dependent on the identification of tissue-specific cis-regulatory motifs (CRMs) of Scl, Fli1 or Gata2. Here, the authors used the Scl enhancer as a template to determine the CRM binding pattern for the TFs in question. Subsequently, a genome-wide screen for sites matching the Scl enhancer yielded multiple target genes, which are also potentially bound by Gata2 and Fli1. Next, using chromatin immunoprecipitation sequencing (ChIP-seq) data for the TFs in question, genes associated with the central Scl-Gata2Fli1 triad were inferred ${ }^{169}$. Recent work by Hamey et al. extended the approach by inferring regulatory networks throughout haematopoietic differentiation, based on gene-gene correlations derived from single-cell RNA sequencing (scRNA-seq) data ${ }^{170}$. By projecting pseudotime trajectories, the authors were able to describe the transcriptional changes that LT-HSC undergo to form the LMPP or MEP lineage.

Other genes have also been implicated as regulators of HSCs (extensively reviewed in ${ }^{171}$ ). Cell cycle regulators, such as Cdk4/6, were shown to induce activation of adult LT-HSC, while genetic knockouts or small molecule inhibition of CDK4/6 sustained their quiescent state ${ }^{172,173}$. Furthermore, a myriad of TFs was associated with regulating HSC fate choices ${ }^{49}$. The Polycomb chromatin remodellers BMI1 and EZH2, are potent activators of HSC self-renewal and proliferation, but also prevent exhaustion of the HSC pool ${ }^{174,175}$. Notably, increased supply of SCF increased expression of the transcriptional regulators Bmi1 and Ezh2 ${ }^{49}$.

The above approaches provide an intriguing insight into key regulatory pathways in LT-HSCs and their fate choices. However, little knowledge of molecular pathways on
the protein level and regulation of chromatin states exists. Cabezas-Wallscheid and colleagues provided the first comprehensive network, integrating DNA methylation, transcriptome and proteome data from HSCs and several multipotent progenitor populations (MPPs) ${ }^{176}$. The authors reported a large overlap of the proteome and transcriptome and identified a protein network, exclusive for HSCs. Receptor proteins and multiple TFs were particularly enriched in the HSC compartment.

A comparison of symmetrically expanding FL HSCs with adult BM HSCs could provide an intriguing model for identifying key pathways involved in self-renewal regulation. Both divisional kinetics and transcriptional profiles of FL HSCs and BM HSCs have previously been directly compared ${ }^{177,178}$. Using limiting dilution transplantation assays, Bowie and colleagues first observed a greater HSC expansion rate of FL HSCs in a transplantation setting ${ }^{178}$. Subsequently, the authors also demonstrated that the notable difference in expansion kinetics between foetal HSCs and adult HSCs can most likely be explained by a higher rate of symmetric cell division of FL HSCs. Bowie et al. also demonstrated that HSCs from secondary transplantation of the progeny of FL HSCs displayed similar expansion kinetics to adult HSCs. Thus, suggesting the presence of a distinct molecular switch that drives fast-cycling, symmetrically selfrenewing HSCs into a quiescent state ${ }^{178}$. Interestingly, the authors were able to identify differential gene expression patterns using targeted qPCR - with self-renewal regulators, such as CyclinD2, Ikaros, MEF and Rae-28 being overexpressed in FL HSCs ${ }^{178}$. Subsequent work by Copley et al. revealed that postnatal transition, during which HSC undergo a reduction in self-renewal activity, was accompanied by reduced Lin28 expression and increased expression of let-7 microRNAs. In line with this observation, the authors showed that overexpression of Lin28 and the downstream let-7 target Hmga2 increased the rate of self-renewal in adult HSCs ${ }^{36}$. Furthermore, others have also shown that foetal and neonatal HSCs rely on Sox17 for maintenance of high self-renewal activity and their identity. In turn, adult HSCs do not express Sox17 and a conditional knockout of Sox17 severely disrupted foetal HSC function, while adult HSCs were not affected ${ }^{179}$. Moreover, unbiased bulk transcriptomic analysis ( $\sim 3000$ cells) of E14.5 FL and adult HSCs, conducted by Manesia et al. revealed metabolic differences, with FL HSCs primarily relying on oxidative phosphorylation to sustain their proliferative capacity ${ }^{177}$.

Similarly, a loss-of-function mutation in epigenetic regulators, such as TET2 and DNMT3A, have been associated with sustained self-renewal of HSCs bearing such mutations ${ }^{180,181}$. Such a Tet2 defect has consequently been implicated in the acquisition of haematopoietic malignancies and myeloproliferative disorders 182,183. Similarly, IDH1/2, DNMT3A and ASXL1 have been shown to induce myeloproliferative neoplasm (MPN) disorder and particularly ASXL1 mutations have been associated with poor prognosis ${ }^{184-186}$. Ultimately, the acquisition of mutations causing sustained self-renewal, such as DNMT3A and TET2, can lead to clonal haematopoiesis ${ }^{187}$.

As outlined above, both extrinsic and intrinsic pathways regulating cell fate in the haematopoietic system have been described. However, these studies are largely based on populations of HSCs and do not distinguish between HSC subtypes. Therefore, to understand the underlying molecular mechanisms of self-renewal and lineage bias, it is of paramount importance to distinguish between HSC subtypes. In this context, comparing HSCs isolated at different developmental stages and/or different genetic models of HSC expansion, such as loss-of-function TET2 mutations, could prove a promising avenue for the identification of novel self-renewal regulators.

### 1.5. Ex vivo expansion and maintenance of HSCs

A vast array of inherited and acquired haematological disorders induce the deregulation of stem, progenitor, or mature cell types. Being situated at the apex of the haematopoietic hierarchy and able to produce all cell types makes HSCs a promising target for therapeutic application. Unsurprisingly, the therapeutic prospect of HSCs for the treatment of a diverse set of disease has been widely recognised ${ }^{112}$. The first therapeutic application of HSCs was conducted over 60 years ago in form of HSC transplantation (HSCT) to treat haematological malignancies ${ }^{188}$. Since that first remarkable HSCT procedure in 1957, significant technological strides have established HSCT as a key component of the armamentarium for various haematological disorders ${ }^{188}$. The most common HSCT procedure, allogeneic transplantation, involves transplantation of HSCs, collected from healthy donors ${ }^{189}$. Alternatively, autologous HSCT procedures require the collection of the patent's

HSCs, which are re-introduced following aggressive radio- or chemotherapy to prevent BM aplasia ${ }^{189}$.

However, current HSCT procedures require a dose of $3-4 \times 10^{6} \mathrm{CD} 34^{+}$cells per kg of body weight ${ }^{190}$. In combination with the challenge of identifying matching healthy donors, these factors lead to a significant shortage of HSCs for clinical purposes. More recently, the advent of advanced gene therapy technologies raised the prospect of treating severe monogenic disorders ${ }^{191,192}$, potentially putting HSCs in even greater demand. Here, patient's HSCs are collected and subjected to ex vivo gene editing. Edited HSCs then undergo an expansion phase, prior to the re-introduction of the corrected HSCs ${ }^{193}$. While in vivo HSC gene therapies are currently being developed, these remain limited by inefficient delivery system and substantial immune responses against current delivery vectors ${ }^{192,194,195}$. Collectively, HSCT treatments for haematological malignancies and application in gene therapies have raised the need for efficient ex vivo culturing systems of HSCs.

Over the past decades, significant efforts have been expended to develop reliable ex vivo culturing systems for the expansion of human and murine HSCs, with a particular focus on the symmetric self-renewal of cultured HSCs. A multitude of such approaches are described below.

### 1.5.1. Feeder cell co-culture systems

Several cell types have been shown to sustain HSC self-renewal in vivo ${ }^{112}$. In the absence of a comprehensive characterisation of the components within the HSC microenvironment, several strategies leveraged the supportive nature of such cell types in feeder cell co-cultures. For instance, stromal cells, such as AFT024, have been shown to provide a cellular environment that supports HSC expansion in vitro ${ }^{196,197}$. Furthermore, MSCs have also been implicated in the effective maintenance and expansion of HSCs ${ }^{198}$. Alternatively, based on the idea that the embryo supported more HSC symmetric self-renewal than the adult, Oostendorp et al. derived a number of embryonic tissue-based cell lines which included the embryonic liver cell line EL081D2 for effective maintenance of HSCs in culture ${ }^{199}$. Most intriguingly, Oostendorp
were able to perform these cultures without the requirement of cell-to-cell contact between HSCs and the feeder cell line, thus implying that conditioning of culture media might provided a pool of secreted molecules, sufficient for HSC maintenance ${ }^{199}$. While such system showed effective HSC expansion, even for as long as 5 weeks, the introduction of secondary cell lines introduced molecular complexity. Clinical application requires the use of defined and reproducible systems, which is made considerably more difficult using a feeder cell approach.

### 1.5.2. Expression of transgenes

In contrast to co-culture systems, overexpression of key HSC maintenance factors provided a well characterised culturing strategy. The first transgene system for HSCs was developed by Helgason et al. ${ }^{200}$. Here, the authors engineered a mouse model with constitutively overexpressed HOXB4 and identified enhanced proliferation of several haematopoietic lineages ${ }^{200}$. In a subsequent study, Antonchuk et al. transduced cells with HOXB4 or green fluorescent protein (GFP) and maintained these, and an non-transduced control, for 14 days in liquid cultures ${ }^{201}$. The authors demonstrated a 40 -fold increase in HSC yield against the starting cell input and a striking 1000-fold increase when compared to the non-transduced control culture ${ }^{201}$. Similarly, overexpression of $\beta$-Catenin a downstream effector of Wnt signalling, and overexpression of HES-1, a Notch signalling target gene, were shown to increase symmetric ex vivo HSC expansion ${ }^{202,203}$.

However, stable or transient overexpression of self-renewal drivers is subject to a key limitation - how to stop symmetric self-renewal. The abnormal expression of a selfrenewal driver that may be a potential oncogene constitutes a great risk factor for clinical application. A notable example includes the NUP98-HOXA9 fusion protein ${ }^{204}$ with which Chung et al. transduced human CD34+ cells to mediate the expression of the NUP98-HOXA9 fusion gene, which in turn increased ex vivo proliferation of $\mathrm{HSCs}^{204}$. Despite such increase in self-renewal of HSCs during ex vivo expansion, the NUP98-HOXA9 fusion protein has been shown to drive the formation of acute myeloid leukaemia (AML) ${ }^{205}$.

### 1.5.3. Replicating the 3D HSC niche in vitro

As previously mentioned in section 1.3.2, a plethora of extracellular matrix proteins have been implicated in engraftment, proliferation, and maintenance of HSCs. These observations raised the question whether the three-dimensional structure of the ECM was required for effective ex vivo expansion of HSCs. Concurrent efforts yielded multiple intriguing approaches.

For instance, Bourgine et al. constructed an osteoblast niche-like bioreactor, embedded in a perfusion chamber, to mimic the human bone marrow niche ${ }^{206}$. For this purpose, human MSC were seeded within the bioreactor to construct the niche ECM compartment, in which CD34 ${ }^{+}$cells were subsequently engrafted ${ }^{206}$.


#### Abstract

Albeit without a bioreactor system, Tiwari et al. utilised a similar approach to Bourgine and colleagues ${ }^{207}$. Here, the authors generated three-dimensional scaffolds by first seeding stromal cells, then decellularising the scaffolds and finally seeding purified human CD34+ HSCs. This approach yielded an approximately 2-fold increase in HSC expansion when compared to control samples ${ }^{207}$.


Most recently, Bai et al. reported a novel 3D culturing system based on a zwitterionic hydroge ${ }^{208}$. In brief, such hydrogels provide a binding scaffold for selective engraftment of HSPCs, while not directly binding to any secreted molecules or cell debris. Human CD34 ${ }^{+}$HSPCs retained multi-lineage potency and successfully repopulated the host BM in transplantation assays. Most intriguingly, the authors achieved a 73 -fold expansion of human HSCs ${ }^{208}$. However, the reproducibility of this approach remains to be determined.

### 1.5.4. Cytokines promoting ex vivo HSC expansion

Several cytokines have been described as potent drivers of HSC self-renewal and expansion. As described previously, SCF, TPO, IL-11 and IL-6 have been identified as potent HSC activators in vivo ${ }^{112,145,209-211 . ~ T h i s ~ r a i s e d ~ t h e ~ i n t e r e s t ~ t o ~ t e s t ~ w h e t h e r ~}$ these cytokines can also facilitate the expansion of human and murine HSCs ex vivo.

A multitude of studies showed that all four cytokines were able to support HSCs ex vivo with murine HSC expansion being particularly supported by SCF, TPO and IL-11211-217. In one example, Sitnicka et al. highlighted that the ex vivo proliferative expansion of murine HSCs was dependent on simultaneous supplementation of SCF and TPO ${ }^{215}$. In another study, the addition of IL-11, a potent activator of the gp130 signalling pathway, was required for the retention of multilineage reconstitution capacity of cultured $\mathrm{HSCs}^{213}$. In line with the murine culturing strategies, human HSC cultures also required supplementation with SCF and TPO ${ }^{212}$. However, in contrast to murine expansion cultures, IL-6 provided potent stimulation of gp130213. Furthermore, human HSC cultures benefited from the addition of FIt3 ligand (FLT3L), which has been shown to be redundant in murine HSC cultures ${ }^{155,212,218}$.

### 1.5.5. Culturing in presence of small molecules

As described above, the introduction of cytokines to ex vivo cultures provided a potent avenue for modulating HSC function. However, expansion was modest and, in most cases, only HSC maintenance was reported. These modest effects prompted a number of groups to explore concurrent addition of a selected set of small molecules to further increase the yield of functional HSCs, particularly for human cells.

A landmark paper by Boitano et al. utilised a previously developed small molecule library, containing 100,000 putative kinase modulators ${ }^{219}$, to identify molecules driving proliferation of human HSCs ${ }^{220}$. The authors identified StemRegenin 1 (SR1) as a potent agonist of HSC expansion, inducing a nearly 3-fold increase in total CD34+ cell numbers and an up to 17-fold increase in the number of engrafting HSCs that were cultured for 7 days $^{220}$. Of note, the SR1-mediated increase in expansion required the presence of the previously defined cytokine cocktail, containing SCF, IL-6, TPO and FLT3L ${ }^{220,221 .}$

Following this success, Fares et al. identified the small molecule UM171 as a stimulant of HSC expansion ${ }^{222}$. The authors demonstrated that the addition of UM171 to liquid expansion cultures, containing SCF, TPO and FLT3L, induced 13-fold expansion of human HSCs ${ }^{222}$. Most intriguingly, the authors directly compared SR1 and UM171 and
observed that UM171 targets more primitive cell types in contrast to SR1. Subsequently, Chagraoui et al. suggested that UM171 likely modulates the regulatory pathways of the inflammatory response ${ }^{223}$. Most intriguingly, a recent phase I clinical safety study for the use of UM171-expanded cord blood in HSC transplantation procedures showed that UM171-expanded HSCs are safe to use and that culture in the presence of UM171 did not impair engraftment ${ }^{224}$.

### 1.5.6. Best performing murine culturing system to date

Despite the abundance of in vitro culture strategies, very few have been widely adopted by the broader haematopoietic research community. A major limitation of several systems is their reliance on feeder cells or undefined serum supplements. The absence of completely defined medium hinders clinical applications. Serum-free conditions have previously been proposed, with Miller and Eaves demonstrating 10day expansion of murine HSCs only supplemented with SCF, TPO and IL-11211. Nevertheless, functional heterogeneity of cells, partially caused by contaminating cells, causes large inconsistencies between individual cultures. Furthermore, the culturing system proposed by Miller and Eaves only permits maintenance for up to 10 days, thus limiting the scope of expansion and utility in long-term assays.

Pioneering work by Wilkinson et al. enabled expansion of single HSCs for as long as 28 days under defined conditions ${ }^{225,226}$. Strikingly, the authors were able to achieve between 236- and 899-fold expansion, as determined by limiting dilution assays and serial transplantation. The culturing strategy consisted of defined, serum-free base media, supplemented with SCF and TPO. To replace the carrier function of serum albumin with a defined and reproducible substitute, Wilkinson et al. screen 11 compounds. Here, polyvinyl alcohol (PVA) was shown to effectively replace serum albumin and increase HSC expansion ex vivo. Most importantly, the authors also immobilised HSCs by coating tissue culture plastic with fibronectin. The affinity of HSCs to fibronectin, facilitated via $\alpha 5 \beta 1$ and $\alpha v \beta 3$ integrins, has been widely recognised ${ }^{227,228}$. It has also been reported that fibronectin-induced signalling cascades contribute to HSC self-renewal ${ }^{100}$. Thus, fibronectin-mediated HSC
immobilisation permits routine application of complete media changes to remove differentiation driving molecules, secreted by cultured cells.

Despite the significant strides in the achievable fold expansion of functional HSCs and longevity of the cultures, several limitations remain. Individual clones and particularly single cell-derived clones remained subject to significant heterogeneity, with some cultures experiencing rapid differentiation, following by loss of long-term repopulation capacity ${ }^{225}$. Hence, it will be of paramount importance to decipher the molecular drivers of such heterogeneity to improve the consistency of HSC expansion. In addition, few strategies have been derived for the identification of functional HSCs, thus making the assessment of the output of ex vivo culture systems time-consuming and costly.

### 1.6. The role of HSCs in haematological malignancies

Haematopoietic malignancies, such as AML, have been shown to originate from HSCs ${ }^{229,230}$. In brief, the onset of AML occurs upon the acquisition of a proliferative mutation in a single HSC with durable self-renewal ${ }^{229}$. Myeloproliferative neoplasms (MPNs) share common somatic mutations with AML, are characterised by proliferative expansion of the myeloid compartment and can transform to AML ${ }^{231,232}$. Hence, MPNs provide a powerful model for uncovering mechanisms of early-stage haematological malignancies.

### 1.6.1. Common driver mutation contributing to MPN disease

MPNs are a group of disorders caused by proliferative expansion of the myeloid lineage ${ }^{231}$. In a subset of patients, such overproduction of myeloid cells also results in the transformation to AML ${ }^{232}$. A set of acquired genetic mutations in the HSC compartment induce constitutive activation of cell proliferation signals ${ }^{231}$. These are accompanied by an array of co-mutations, which contribute to the onset of MPN formation ${ }^{231}$.

Mutations of JAK2, MPL and CALR have been identified as the most common drivers of MPNs ${ }^{231}$. Interestingly, mutational signature screening of 151 MPN patients revealed that CALR and JAK2 mutations were mutually exclusive ${ }^{233}$. In brief, a gain-of-function point mutation in the wild-type JAK2 locus leads to an amino acid substitution at position 617 (JAK2 ${ }^{\mathrm{V617F}}$ ), leading to a substitution of valine with phenylalanine ${ }^{234}$. In turn, JAK2 ${ }^{\text {V617F }}$ induces constitutive activation of the JAK2 kinase function and thus downstream phosphorylation of STAT signalling factors ${ }^{235}$. Similarly, activating mutations in MPL, a cytokine receptor upstream of the JAK/STAT signalling pathway, induce receptor dimerisation to initiate the JAK/STAT phosphorylation cascade ${ }^{236,237}$. Multiple frameshift mutations in the calcium regulatory factor CALR have been reported ${ }^{238,239}$. While the disease mechanism remains not fully described, several reports suggested that mutant CALR binds to MPL to induce constitutive JAK/STAT signalling ${ }^{237}$.

While such driver mutations are crucial to induce a hyperproliferative disease phenotype, several studies highlighted the importance of secondary co-mutations in disease onset ${ }^{240-242}$. Xu et al. screened blood samples of a large group of donors without MPN diagnosis and phenotype ${ }^{243}$. The authors showed that 37 out of 3935 donors carried a JAK2-V617F mutation, but only one donor exhibited an MPN-like phenotype ${ }^{243}$. These results provided intriguing evidence that singular driver mutations were not sufficient to induce the onset of MPNs.

Multiple epigenetic regulators were identified as commonly co-mutated, with TET2 exhibiting the highest mutation frequency in JAK2 ${ }^{\text {V617F }}$ positive patients ${ }^{244}$. Intriguingly, such loss-of-function mutations of TET2 induced an HSC self-renewal advantage ${ }^{245-}$ ${ }^{248}$. Hence, increased HSC self-renewal, coupled with the JAK2 ${ }^{\text {V617F-mediated }}$ hyperproliferative phenotype drive MPN disease onset and progression. A subsequent study interrogated the importance of the mutation order in patients, harbouring TET2 loss-of-function and JAK2 ${ }^{\text {V617F }}$ mutations ${ }^{249}$. The authors showed that patients who acquired the JAK2 ${ }^{\text {V617F }}$ mutation first had a worse prognosis, while an initial loss of TET2 reduced the JAK2 ${ }^{\text {V617F }}$-mediated hyperproliferative drive in HSPCs ${ }^{249}$. Instead, HSPCs from patients who acquired a TET2 loss-of-function mutation first were predominantly carrying only a TET2 mutation. Together, these studies highlight the
importance of co-mutations for MPN disease progression and establish the role of HSCs in disease onset.

### 1.6.2. TET2 deficiency and the acquired stem cell advantage

MPN mutations, such as the loss-of-function TET2, alter HSC fate and thus contribute to the disease phenotype. Consequently, such observations raise the prospect of utilising TET2 mutant HSC to uncover key molecular pathways governing HSC selfrenewal. During homeostasis, the TET2 enzyme catalyses the conversion of 5methylctosine (5-mC) into 5-hydroxymethylcytosine (5-hmC) ${ }^{250}$. Such oxidation of the repressive DNA methylation initiates the demethylation cascade to revert gene silencing ${ }^{251}$. Interestingly, this process has been widely implicated in HSC biology and TET2 has been identified as the most abundantly expressed TET family member in adult HSCs ${ }^{252,253}$.

In MPNs and other malignancies, a broad range of mutations across multiple exons have been reported ${ }^{254,255}$. These include insertions, deletions, and frameshift mutations, all resulting in the loss of native TET2 function ${ }^{256}$. In turn, TET2 deficiency induced myeloid cell expansion ${ }^{257}$. Most intriguingly, the HSPC compartment of MPN patients was predominantly enriched for singular TET2 los-of-function mutations and TET2 mutations increased HSC fitness ${ }^{241}$.

The prevalence of TET2 mutations in patient samples raised the interest in the underlying molecular mechanisms. For this reason, multiple groups developed mouse models of loss-of-function TET2 mutations by introducing I) conditional knockout of exon 3, II) conditional knockout of exons 8-10, III) conditional knockout of exon 11, IV) insertion of $\beta$-galactosidase-neomycin at intron 9 to form a non-functional fusion protein and V ) insertion of a nlacZ/nGFP reporter into exon $3^{245-248}$. All four models showed significant abrogation of wild-type TET2 expression ${ }^{245-248}$. These studies showed that TET2 deficiency ( TET2 $^{-/}$) led to significant expansion of the stem cell compartment. In line with such observations, competitive transplantation assays revealed that TET2 ${ }^{-1}$ HSCs gained a significant self-renewal advantage ${ }^{245-248}$. Downstream progenitor populations remained largely unchanged, with the exception
of CMPs which were significantly expanded in TET2-/ mice. In addition to the enhanced self-renewal potency in vivo, serial replating CFU assays showed consistent self-renewal and significantly reduced differentiation of $T E T 2^{-/} \mathrm{HSCs}^{248}$.

Collectively, these studies demonstrated that TET2-mediated demethylation is of paramount importance for normal haematopoiesis, whereby abrogation of homeostatic TET2 function leads to a significant HSC defect. Aside from the clinical importance of uncovering the molecular drivers of TET2 deficiency, the model system also provides a valuable resource for uncovering the cellular regulation of HSC selfrenewal. Three of the four available TET2-/ mouse models, developed by MoranCrusio et al., Li et al. and Quivoron et al., developed chronic myelomonocytic leukaemia (CMML) ${ }^{246-248}$. In contrast, the TET2 ${ }^{-1}$ model, harbouring the loss of exons $8-10$ and developed by Ko et al., was subject to increased CD115+F4/80+ monocyte/macrophage cells but did not develop significant disease onset ${ }^{245}$. Instead, the phenotypic consequences of TET deficiency were limited to the previously described increased self-renewal potency, while other HSC functions remained intact. Hence, wild-type and TET2-/ HSCs are only phenotypically distinguishable by their rate of self-renewal and competitive advantage, making it a powerful model for uncovering the molecular machinery governing HSC self-renewal.

### 1.6.3. Clinical significance of TET2 deficiency in MPNs

Uncovering molecular pathways governing the acquired stem cell advantage of TET2 deficient HSCs also provides an intriguing insight into its disease pathology. As previously described, TET2 loss-of-function mutations have been widely implicated in MPNs, AML and CMML, among other diseases ${ }^{254,258}$. While the JAK2 ${ }^{\text {V617F }}$ point mutation has been identified as a key driver mutation of MPNs, several studies outlined that singular JAK2 ${ }^{\text {V617F }}$ mutant HSCs are subject to exhaustion. Only in the presence of a concurrent loss-of-function TET2 mutation were stem cells able to retain long-term potency to sustain the JAK2 ${ }^{\mathrm{V617F}}$-mediated proliferative advantage ${ }^{259,260}$. Consequently, double mutant models exhibited more severe MPN phenotypes ${ }^{259,260}$.

MPNs have been classified into three major disease subtypes, namely essential thrombocythaemia (ET), polycythaemia vera (PV) and myelofibrosis (MF) $)^{261}$. In brief, ET patients have elevated platelet levels and are at risk of developing thrombosis or transform to AML or myelofibrosis ${ }^{232}$. The prevalence of ET remains low, with 0.38 to 1.7 out of 100,000 individuals diagnosed with ET per year. Furthermore, patients are diagnosed with ET at a median age of $65-70^{262}$. TET2 frameshift, nonsense, or missense mutations, leading to the catalytic inactivation of TET2, were detected in $5 \%$ of patients ${ }^{254}$.

In contrast, the PV phenotype includes the excessive synthesis of erythrocytes and elevated haemoglobin levels in combination with reduced levels of erythropoietin ${ }^{263}$. The incidence rate of PV amongst several European registries lies between 0.4 and 2.8 per 100,000 individuals and patients were diagnosed between 65 and 74 years of age (median value) ${ }^{262}$. PV patients are at risk of developing thrombosis and have a median life expectancy below 2 years post disease onset if untreated ${ }^{264}$. In addition, 2-7\% of PV patients transform to AML and 10-20\% transform to MF. Approximately $16 \%$ of PV patients have been shown to carry a loss-of-function TET2 mutation ${ }^{254}$.

MF, the third major and arguably most severe MPN subtype, is comprised of multiple phenotypic symptoms. MF patients develop BM fibrosis or reticulin deposits in the bone marrow ${ }^{231}$. Furthermore, patients suffering from MF have the highest rate of AML transformation of all MPN subtypes ${ }^{265-269}$. The incidence rate of MF across multiple European cohorts ranged from 0.1 to 1 per 100,000 people ${ }^{262}$. TET2-deficiencies have been identified in $17 \%$ of patients with $\mathrm{MF}^{254}$. In addition, patients that underwent transformation from either PV or ET to MF were shown to harbour loss-of-function TET2 mutations in $14 \%$ of cases $^{254}$. Finally, $17 \%$ of patients that underwent blast phase, or AML, transformation from PV, ET or MF carried TET2 mutations ${ }^{254}$.

In addition, TET2 deficiency has been implicated in the malignant propagation of AML and CMML. In brief, Tefferi et al. screened a small cohort of patients with non-MPN myeloid malignancies for TET2 mutations ${ }^{258}$. As expected, the authors reported a broad range of frameshift and nonsense mutations across several exons ${ }^{258}$. Interestingly, 5 out of 15 CMML patents were shown to carry mutant TET2, but their disease severity did not differ from CMML patients without TET2 mutations ${ }^{258}$. A later
study of 175 CMML patients identified that $46 \%$ of patients carried TET2 mutations and that the presence of mutant TET2 provided a favourable prognosis ${ }^{270}$. In contrast, loss-of-function TET2 mutations in primary AML patients have been associated with an unfavourable disease prognosis and have been identified in 64 out of 486 patients ${ }^{271}$.

Collectively, these data highlight the clinical importance of TET2 deficiency in various haematological diseases and malignancies. While the stem cell advantage, observed in TET2-/ mouse models, provide an intriguing system for uncovering the molecular networks governing self-renewal, it can also provide an insight into disease initiation and progression.

### 1.7. Proteomic screening of rare cell populations and its importance in HSC biology

### 1.7.1. The advent of proteomic mass spectrometry

Originally, biological discovery was predominantly conducted in the form of forward genetic approaches, driven by the observation of phenotypic phenomena with the subsequent aim to identify the underlying genes and their protein products ${ }^{272}$. In this approach, gene sequence and function were often inferred from isolated proteins associated with the phenotypic phenomenon. For this purpose, isolated proteins were sequenced using Edman degradation, where the amino group of each amino acid is sequentially cleaved and subsequently identified by high-performance liquid chromatography (HPLC) $)^{273-275}$. Gene sequences could then be inferred from the amino acid sequence, and the corresponding gene identified by targeted $P C R^{276}$. The introduction of expression arrays and other genomic approaches prompted the assembly of gene sequence databases to ultimately enable discovery-driven reverse genetics. In proteomics, this reduced the downstream processing following Edman degradation and enabled the identification of several hundred proteins, using protein maps and two-dimensional gel electrophoresis ${ }^{277}$. The human genome project not only pioneered today's genomic and transcriptomic discovery-based science, but also revolutionised the field of proteomics. Access to complete genomic sequences rapidly
fuelled curation of comprehensive genetic databases, removing the need for de novo protein sequencing. This shifted the focus to developing techniques to overcome the remaining limiting factors: I) speed of protein sequencing, II) sensitivity within a complex protein mixture and III) reliability of mapping sequences to the genome. With Edman degradation being low-throughput and suffering from poor sensitivity, radical innovation was required.

The need was quickly met by the introduction of mass spectrometry (MS), which had previously been widely used for determining the mass and structure of small molecules ${ }^{278}$. In principle, the molecules of interest are ionised and passed into the vacuum chamber of the mass analyser. Here, the ions are accelerated to facilitate separation based on their mass to charge ( $\mathrm{m} / \mathrm{z}$ ) ratio ${ }^{272}$. One of the first major breakthroughs by Fenn and colleagues was electrospray ionisation (ESI), which allowed reliable ionisation of large biomolecules, such as proteins, without dissociating their structures ${ }^{279}$. While intact proteins do not have distinctive masses, unique peptide fragments from such proteins can be used for protein identification. This led to the introduction of proteolytic digestion ${ }^{280}$. Here, proteolytic enzymes cleave proteins at defined sequence-dependent sites, yielding small peptides. The resulting unique peptide masses could then be matched to theoretic peptide masses, derived from newly assembled sequence databases. This principle was supported by the introduction of tandem mass spectrometry (MS/MS). This allowed injection of peptide mixtures, which are first separated in MS1 to yield precursor ions. Based on their m/z values, specific peptide ions can be selected and directed into a collision chamber, where these ions are fragmented along the peptide backbone ${ }^{272,281}$. Fragmented ions are then passed into the second mass analyser (MS2). With the $\mathrm{m} / \mathrm{z}$ value for each amino acid being unique, the acquired MS/MS spectrum can be used to derive the amino acid sequence of each sampled peptide. Pioneering work by Matthias Mann and colleagues further revolutionised MS/MS analysis by introducing nanoelectrospray to reduce the droplet size produced by ESI, which in turn minimised sample loss due to ionisation inefficiency, and reduced the required sample input 282,283 . Collectively, this analytical process largely depends on reliable mapping of MS/MS spectra to theoretical spectra predicted based on genomic sequence databases. Thus, it is not surprising that the rapid technological advances were also
accompanied by the development of search algorithms, matching MS/MS spectra to databases for confident protein identification ${ }^{284}$.

For decades, protein biologists relied on 2D gel electrophoresis for protein isolation. However, its low throughput combined with the difficulty of recovering proteins remained a major hurdle on the road towards ubiquitous screening of complex proteomes. This was first overcome by introducing in-gel digestion ${ }^{285}$. Shevchenko et al. showed that proteins could be proteolytically cleaved using trypsin and the resulting peptides extracted efficiently ${ }^{286}$. Most importantly, the subsequent protein identification was significantly more sensitive than previous Edman degradation or MS techniques. Several years later, increasingly complex protein lysates became accessible for rapid MS/MS analysis through the introduction of online liquid chromatography, in particular HPLC. In principle, this allows for pre-fractionation of complex peptide mixtures, which are sequentially passed directly to the mass spectrometer ${ }^{287,288}$. This facilitates greater peptide separation from the same sample, thus maximising the number of peptides detected and fragmented in MS2 ${ }^{281}$. The combination of ever-increasing sequencing speed of modern mass spectrometers, automation of precursor ion selection in MS1, and the technological advances described above facilitated rapid, high-throughput protein identification, ultimately termed shotgun proteomics ${ }^{289}$.

### 1.7.2. Proteomics on rare cell populations

To date, approximately 20,000 protein-coding genes have been described, of which at least half are being transcribed by a cell at any given time ${ }^{290}$. However, the greatest complexity within the proteome lies beyond the number of expressed genes and their copy numbers (Figure 5). PTMs increase the number of functionally different proteins many-fold by affecting cellular localisation, protein half-lives, folding, activation and beyond ${ }^{291-295}$. This phenomenon drove rapid technical development to facilitate the large-scale unbiased inspection of most, or even all, proteins of cells in a given state. Within a decade, groups have achieved an ever-increasing sequencing depth from $\sim 5,000$ identified proteins to now 15,000 proteins ${ }^{296-298}$. However, these approaches require substantial starting protein material, which in turn coincides with large cell
numbers for each condition or experiment. Typical experiments are performed with upwards of $10^{6}$ cells per condition. Hence, proteomic screens for rare cell populations remained scarce. One such notable example are HSCs. As outlined above, little is known about the HSC proteome, and most studies have been limited to inspecting singular proteins, using immunofluorescence and, more recently, mass cytometry ${ }^{299,300}$. However, these studies rely on prospective selection of putative targets.

## Proteome Complexity



Figure 1.5. The exponentially increased complexity of the genotype-phenotype paradigm. Alternative splicing during transcription of protein-coding genes increases the scope of functionally distinct proteins. Following the translation of the respective mRNA molecules, protein folding further increases the molecular variation. In addition, differences in the rate of transcription of a given mRNA molecule and the corresponding rate of translation introduces non-linear relationships between transcripts and their respective protein. Furthermore, subcellular localisation alters the scope of binding partners for a given protein, often substantially altering its molecular function. Finally, post-translational modifications, such as phosphorylation, ubiquitinylation and others, also impact the functional output of a given protein. Collectively, the proteome holds the greatest functional complexity.

The proteomic community has long been striving towards establishing reliable protocols to perform deep, unbiased proteomic screens with little material, even down to the single cell level ${ }^{301}$. Originally termed microproteomics, significant efforts were concentrated on improving sample processing prior to proteomic analysis, development of improved mass analysers, and downstream protein identification ${ }^{302}$. A prominent example of early microproteomics included the development of mass spectrometric imaging (MSI) ${ }^{303}$. In brief, this pioneering technique utilised matrixassisted laser desorption ionisation (MALDI) for screening the proteome across preserved tissue cross sections ${ }^{303-305}$. The MSI approach, developed by Stoeckli and colleagues, was shown to sequentially ionise spots with an average diameter of 25
$\mu m^{303}$. Subsequent MS analysis and data processing enabled identification of up to 200 proteins per spot, with up to 30,000 spot arrays analysed for a given cross section. Thus, MSI enabled spatially resolved proteomic screening of complex tissues at high resolution. More recent advances in secondary-ion mass spectrometry (SIMS) and MALDI-based approaches have further increased resolution, enabling screening for a multitude of molecules, such as metabolites, peptides and lipid, at single-cell resolution ${ }^{306}$. However, throughput and the depth of capture remain key challenges in these approaches.

While spatially resolved proteomic screening provides powerful insight into complex tissues, such techniques do not permit prospective isolation of target cell types. To address the unmet need Poulhac et al. proposed the use of laser capture microdissection (LCM) for isolating populations of target cells ${ }^{306}$. Here, the authors isolated a target cell by LCM from cryopreserved tissue sections. Regions of interest in tissue sections were extracted and pooled prior to downstream proteins extraction. Following extraction, protein identification and quantification was performed using nanoscale liquid chromatography paired with tandem mass spectrometry (LCMS/MS). Such LCM-based microproteomic techniques enabled identification of fewer than 200 proteins from as few as 1,000 cells. While such technique enhances the scope of experimental tools, the sequencing depth of <200 proteins limits resolution of biological differences and the scope of addressable biological questions. Similarly, a tissue extraction-based approach limits application to homogenous tissues, as LCMbased extraction does not permit isolation of rare target cells residing in heterogeneous tissues with a broad range of cell types sharing the same niche.

Recent technological advances in the speed and resolution of mass analysers, have sparked efforts to address the unmet need of screening rare cell populations of any origin. Early work by Russell et al. to develop the TMTcalibrator (Tandem Mass Tag calibrator) protocol for diagnostic screening of low abundant biomarkers, revealed that significant sample reduction is achievable through isobaric labelling approaches ${ }^{307}$. Building on the principles of reducing protein loss during sample preparation, Zhu et al. developed the nanodroplet processing in one pot for trace samples (nanoPOTS) framework ${ }^{308}$. In brief, photolithography was utilised to manufacture glass chips, comprised of an array of nanowells for sample processing. As few as 10-100 cells
were deposited in each nanowell using a robotic setup for handling of picoliter volumes. Using nanoPOTS, the authors were able to detect as many as 1,500 proteins from 10 cells, increasing to $\sim 3,000$ proteins when $\sim 140$ cells were pooled per sample ${ }^{308}$. These studies demonstrated the technological feasibility of routine rare cell proteomics, sparking great interest in achieving single-cell resolution. In addition, these studies also highlighted two parallel developmental pathways, one utilising labelling-based multiplexing and the second focusing on label-free technologies.

Prominent examples of novel isobaric labelling-mediated multiplexing protocols enabling single-cell proteomics include pioneering work by Budnik et al. and Specht et al. ${ }^{309,310}$. The first single cell proteomics by mass spectrometry (SCoPE-MS) protocol was proposed by Budnik et al. ${ }^{311}$. The authors manually isolated single cells and performed Covaris sonication to avoid sequential sample handling and the associated loss of material. To enable pooled LC-MS/MS analysis of single cell, isobaric TMT labels were utilised. TMT labelling has previously been established to enable multiplexed proteomic screening, which in turn increased sample throughput, reduced cost, and increased protein ID recovery ${ }^{312,313}$. Most importantly, TMT labelling also provided sensitive quantification of each mass tag, allowing accurate computation of relative difference in protein abundance between samples. In SCoPE-MS, 8 single cells were multiplexed alongside two carrier populations containing 100 cells each. Using this approach, the authors demonstrated identification of several hundred proteins. However, low sample throughput and a 100-fold dynamic range between single cells and the carrier population provide significant limitations to SCoPE-MS. The importance of controlling the dynamic range between single cells and the carrier population were also highlighted by Cheung et al. ${ }^{314}$. The authors demonstrated that an increasing dynamic range introduces stochasticity in single cell-derived ion quantification, thus masking technical biases as biological variability. To address these issues, the authors developed a second iteration, SCoPE2 ${ }^{310}$. In brief, Specht et al. introduced a 5-cell reference population to reduce the dynamic range to within a 5 -fold abundance differential. In addition, the authors implemented cell lysis in a 384-well format and developed an optimised analysis pipeline. Using Bayesian inference, the authors utilised chromatographic retention time to inform the confidence of peptide spectrum matches, thus increasing the rate of protein identification ${ }^{315,316}$. Overall, SCoPE2 increased the number of identified proteins (< 1\% false discover rate) to $\sim 900$
proteins per cell. The authors demonstrated the power of SCoPE2 by screening $\sim 1,500$ monocytes and macrophages, quantifying over 3,000 proteins in total. The principles of SCoPE-MS have also been adapted by other groups, such as the single cell proteomics (scMS) platform developed by Schoof and colleagues ${ }^{317}$. Like SCoPE2, scMS also utilises TMT labelling and a carrier population. However, cell lysis was performed using trifluoroethanol buffer instead of purified water to increase protein yield. In addition, Schoof et al. proposed the use of index sort data, collected during FACS, and developed an integrated data analysis pipeline. Most intriguingly, scMS did not include a reference population and utilised all available TMT channels, thus increasing cell throughout to 14 cells per multiplex. Overall, Schoof et al. achieved similar sensitivity as SCoPE2, with $\sim 1,000$ quantified proteins per cell.

In parallel efforts, Brunner and colleagues proposed an alternative approach to singlecell proteomics without the requirement of peptide labelling ${ }^{318}$. In brief, the authors constructed a novel mass spectrometer by adapting multiple aspects of the ionisation process to facilitate 10 -fold higher instrument sensitivity. In turn, that enabled the identification of $\sim 800$ proteins from a single cell at a throughput of 40 cells per day. However, the custom adaptation of the mass spectrometer currently limits wider adaption.

Based on the principles of nanoPOTS, several studies demonstrated its applicability to achieve single-cell resolution. For instance, Cong and colleagues incorporated high field asymmetric ion mobility spectrometry (FAIMS) ${ }^{319}$ into the workflow, following sample processing by nanoPOTS and ion separation by nanoLC ${ }^{320,321}$. The introduction of FAIMS allowed selective exclusion of singly charged ions in favour of multiply charged ions being passed into the mas analyser. The authors demonstrated a 2 -fold increase in protein identifications, yielding $\sim 1,000$ protein per single cell ${ }^{321}$. In separate efforts by Tsai et al. and Dou et al., the authors utilised TMT labelling during the nanoPOTS workflow to achieve single cell resolution ${ }^{322,323}$. In particular, the authors optimised the automatic gain control and injection time to achieve over 1,100 quantified proteins ( 1,500 proteins across a set of 104 single cells) ${ }^{322}$.

Overall, both label-free and labelling-mediated multiplexing approaches shared similar rates of protein quantification, with the best performing techniques reaching as many
as 1,000-1,500 proteins in single cells. However, the protocols vary significantly in sample throughput and the requirement for bespoke apparatus. With the field of singlecell proteomics in its infancy, further technological advances will undoubtedly improve accessibility, throughput and sensitivity.

### 1.7.3. The HSC Proteome

The advent of single cell approaches has allowed extensive characterisation of transcriptional profiles and functional properties of HSCs. Nevertheless, little is known about how well such transcriptional profiles are recapitulated by the proteome. Similarly, it remains unknown how molecular profiles on the protein level differ between different HSC subtypes ${ }^{57}$. The low frequency of HSCs, the difficulty of isolating HSC subtypes, and the technical limitations of proteomic analyses prevented large scale inspection of molecular protein networks for a long period of time. However, the proteome holds the greatest functional complexity within the genotypephenotype paradigm ${ }^{324}$. Hence, it remains vital to inspect protein networks to fully understand the underlying molecular mechanisms of any biological system, including the regulation of HSC self-renewal and drivers of fate choices in different HSC subtypes.

Cabezas-Wallscheid and colleagues conducted the first comprehensive proteomic screen of HSCs to characterise molecular networks driving HSCs and downstream progenitor populations ${ }^{176}$. The above comparison of bulk HSCs (Lin ${ }^{\text {neg }}{ }^{\text {Sca }}-1^{+} \mathrm{c}$ -$\mathrm{Kit}^{+}$CD34-CD48-CD150 ${ }^{+}$CD135 ${ }^{-}$) and MPPs (Lin ${ }^{\text {neg }} \mathrm{Sca}-1^{+} \mathrm{C}-\mathrm{Kit}^{+} \mathrm{CD} 34^{+}{ }^{+} \mathrm{CD} 48^{-}$ CD150 ${ }^{+}$CD135 ${ }^{\circ}$ ) was conducted using 400,000 cells and identified 47 proteins as differentially expressed between the two populations ${ }^{176}$. As expected, the authors observed an overrepresented network of cell cycle regulators, components of the DNA replication machinery, and cell proliferation proteins within the MPP population. HSCs on the other hand, were enriched for glycolytic enzymes, hypoxic response factors and iron homeostasis regulators. More interestingly perhaps, the study also revealed the over-representation of downstream effector proteins of the Lin28-let7 signalling pathway in HSCs. This negative feedback loop pathway has been widely implicated in stem cell differentiation and self-renewal, and overexpression of Lin28 has been
implicated in tumorigenesis (extensively reviewed $\mathrm{in}^{325}$ ). Enrichment of downstream effectors in HSCs was thus proposed to indicate the involvement of the Lin28-let7 pathway in maintaining self-renewal of HSCs. A functional study by Copley et al., outlining Lin28 and Hmga2 as key regulators of self-renewal potential during development ${ }^{36}$, complements the findings by Cabezas-Wallscheid et al. Subsequent work by Jassinskaja et al. set out to compare the global proteome of adult HSPCs (Lin ${ }^{\text {neg }} \mathrm{Sca}-1^{+} \mathrm{c}-\mathrm{Kit}^{+}$) against their FL counterparts, which undergo extensive clonal expansion and proliferation ${ }^{326}$. As expected, key cell cycle regulators and proliferative drivers were enriched in FL HSPCs ${ }^{326}$. Most intriguingly, Jassinskaja and colleagues also reported increased complexity of adult HSPC proteomes and differences in protein complex stoichiometries between foetal and adult cells. Such differences provide an intriguing avenue to uncover molecular regulators of HSC fate choices. However, despite the advances outlined above, the role of HSC heterogeneity in these processes remains largely elusive due to cell number constrains (both studies required $4-4.5 \times 10^{5}$ cells per sample). Therefore, circumvention of the low cell number limitation is of paramount importance in order to uncover self-renewal drivers by screening HSC subtypes.

Recent efforts to uncover the HSC proteome have recognised the requirement of proteomic mass spectrometry approaches that require significantly fewer cells. Zaro et al. and Amon et al. have performed LC-MS/MS-based proteomic profiling using as few as 50,000 and 25,000 cells per sample, respectively ${ }^{327,328}$. Both approaches enabled identification of $\sim 4,000$ proteins in the HSC fraction. While Zaro et al. provided a comprehensive resource of various murine HSPC populations of young and aged mice, Amon et al. screened human HSPC populations. A comparison between a broad HSC population and MPPs revealed that HSCs were depleted of a plethora of differentiation-associated proteins but enriched for quiescent signatures ${ }^{328}$. Furthermore, aged HSCs expressed higher enrichment scores for DNA repair and cell cycle activity ${ }^{328}$. The human HSC and MPP proteome revealed enrichment of telomerase maintenance proteins, such as GAR1, DKC1, NHP2 and NOP10, and key quiescence regulators, including IDH1, IDH3A and IDH3B, in HSCs ${ }^{327}$. Most intriguingly, Amon et al. showed that the HSC-specific enrichment of quiescencemaintaining isocitrate dehydrogenases, including IDH1, IDH3A and IDH3B, were only detected by LC-MS/MS and not recapitulated by the transcriptome ${ }^{327}$.

Several studies reported discrepancy between the proteome and transcriptome of HSCs, with few overlapping proteins exhibiting correlation ${ }^{327,328}$. Interestingly, both Zaro et al. and Amon et al. highlighted that the proteome-transcriptome correlation notably increased for cell types further down the haematopoietic hierarchy ${ }^{327,328}$. Interestingly, HSCs were recently shown to exhibit rapid transcription but significantly reduced rate of protein synthesis ${ }^{329-331}$. Such observations corroborate the absence of correlation between the HSC proteome and transcriptome. These observations emphasise the importance of mapping the HSC proteome to understand cell function.

Proteomic studies provide great resources for the global HSC proteome and other HSPC populations, corroborating molecular pathways previously implicated in HSC biology. Despite the great value such studies provide, the scope of uncovering novel regulators of HSC fate, particularly self-renewal, remains limited to date. This can be largely attributed to cell number restrictions, which only permitted comparisons of a broad HSC pool with downstream progenitors. To uncover HSC fate regulators, including of processes such as self-renewal, it is of paramount importance to dissect HSC heterogeneity and the advent of rare cell proteomic screening technology is likely to greatly contribute.

### 1.8. Thesis Aims

Functional studies and molecular profiling, particularly using single cell -omics technologies, provide a detailed description of molecular HSC self-renewal drivers. However, the limited success in translating these concepts into reliable, reproducible ex vivo culturing systems and the functional heterogeneity of HSCs have highlighted the molecular complexity of HSC self-renewal.

Most cellular functions, ranging from structural integrity to signalling cascades in response to external stimuli, are performed and regulated by the proteome. Not surprisingly, the proteome also contains the greatest molecular complexity within the genotype-phenotype paradigm, with post-translational modifications, protein isoforms, protein-protein interactions all contributing to functional diversity of a single protein.

Hence, it is of paramount importance to inspect the proteome in order to uncover the molecular mechanisms driving HSC self-renewal. However, the rare nature of HSCs has limited our ability to perform unbiased proteomic studies.

At the onset of the PhD, no technologies were available for screening the proteome of rare cell populations. Results Chapter 1 outlines the efforts to develop a novel LCMS/MS protocol, capable of screening the proteome of rare cell at sufficient depth. The chapter covers the method development of said protocol and demonstrates its utility using a haematopoietic progenitor cell line. In Chapter 2, I applied the improved protocol to identify key HSC self-renewal regulators and integrated proteome with transcriptomic profiles. The LC-MS/MS screen was applied to primary HSCs and other multipotent cell types from wild-type and TET2-/ mice. As outlined above, ex vivo HSC expansion cultures are subject to significant heterogeneity. While significant efforts have been expended to develop reliable culturing systems, very little is known about the molecular pathways governing HSC self-renewal in the ex vivo setting. In Chapter 3 , we set out to characterise the underlying molecular pathways of ex vivo expanded HSCs and identify, for the first time, gene signatures distinguishing functional HSCs with long-term repopulation capacity from non-repopulating phenotypic HSCs.

## Chapter 2. Materials and Methods

### 2.1. Mice

TET2 loss-of-function (TET2 ${ }^{-/}$) mice were obtained from Prof. Anjana Rao and were bred at the Central Biomedical Service animal facility at the University of Cambridge (Ko et al., PMID: 21873190). Wild-type (WT) C57BL/6 mice were either purchased from Charles River or bred in-house as a pedigree line at the Central Biomedical Service (Cambridge) or the biomedical services facility (BSF) at the University of York and $T E T 2^{+/+}$littermates were used as often as possible for experiments. Fgd5 ${ }^{\text {ZsGreen•ZsGeenr/t }}$ knock in/knock out mice were purchased from Jackson Laboratories. Transplantation assays were performed using c-Kit mutant C57BL/6 ${ }^{\mathrm{W} 41 / \mathrm{W41}}$-Ly5.1 (W41) recipient mice. All mice were housed in microisolator cages under specified pathogen-free (SPF) conditions at animal facilities at either the University of Cambridge or University of York. Animals continuously received sterile water, food and bedding. All procedures were conducted in accordance with United Kingdom Home Office regulation and the Animal Scientific Procedure Act. Mice were sacrificed by one of the two following methods: 1) Primary: $\mathrm{CO}_{2}$ inhalation, secondary: Dislocation of the neck; 2) Primary: Dislocation of the neck, secondary: exsanguination.

### 2.2. Isolation of haematopoietic stem cells

### 2.2.1. Bone marrow harvesting

Bone marrow was harvested from the tibiae, femurs and hip bones of both hind legs, the sternum and spines. Harvested bones were crushed in PBS-2\%FCS [phosphate buffered saline, supplied with $2 \%$ foetal calf serum (Sigma or STEMCELL Technologies)] to extract the BM. Subsequently, extracted BM tissue was filtered using $20 \mu \mathrm{~m}$ sterile filter units and pelleted by centrifugation at 300 xg for 7 mins and the supernatant was removed.

### 2.2.2. Red cell lysis

Pelleted BM cells were resuspended in 3 mL PBS-2\%FCS per mouse. Red cell lysis was performed by adding 5 mL ammonium chloride $\left(\mathrm{NH}_{4} \mathrm{Cl}\right.$, supplied by STEMCELL Technologies) per mouse and incubating for 5 mins at $4^{\circ} \mathrm{C}$. Subsequently, samples were vortexed and incubated for a further 5 mins at $4^{\circ} \mathrm{C}$. Cells were washed by adding 20 mL of PBS-2\%FCS and pelleted by centrifugation at 300 x g for 5 mins . The supernatant was removed, bone marrow cell pellets were resuspended in $500 \mu \mathrm{~L}$ PBS$2 \%$ FCS per mouse and transferred to a new tube.

### 2.2.3. Lineage depletion

HSPCs were enriched by removal of mature cells (Lineage ${ }^{+}$). For this purpose, the EasySep Mouse Hematopoietic Progenitor Cell Enrichment Kit (STEMCELL Technologies) was used. Per $500 \mu L$ cell suspension, $10 \mu l$ of the biotinylated EasySep Mouse Hematopoietic Progenitor Cell Isolation cocktail was added and incubated for 15 min at $4^{\circ} \mathrm{C}$. Next, $15 \mu L$ EasySep Streptavidin Rapid Spheres 50001 were added to the cell suspension, mixed and incubated for a further 15 min at $4^{\circ} \mathrm{C}$. Following the final incubation step, the cell suspension was diluted with PBS-2\%FCS up to a volume of 2.5 mL . The tube was transferred into an EasySep magnet with the cap removed and incubated for 3 min at room temperature. Without removing the tube from the magnet, the cell suspension was tipped into a new tube. The tube in the magnet was topped up to 2.5 mL with PBS-2\%FCS, vortexed and subjected to a second 3 min incubation in the magnet. The collected supernatant was centrifuged at $300 \times \mathrm{g}$ for 5 min .

### 2.2.4. Isolation strategy of ESLAM-Sca1+ HSCs

Long-term HSCs (LT-HSCs) were defined as CD45+CD48-CD150+EPCR+Sca1+ (ESLAM-Sca1+) and isolated by FACS as previously described ${ }^{35}$. Such strategy resulted in $\sim 60 \%$ purity of functional HSCs, as defined by single cell transplantation assays ${ }^{332}$. Antibodies used for each marker are described in Table 1.

Table 1. ESLAM-Sca1+ panel antibodies

| Antibody | Clone | Fluorochrome | Supplier |
| :--- | :--- | :--- | :--- |
| CD45 | 30-F11 | BV421 or FITC | Biolegend |
| CD48 | HM48-1 | APC | Biolegend |
| CD150 | TC15-12F12.2 | PE-Cy7 | Biolegend |
| EPCR | RMEPCR1560 | PE | STEMCELL Tech |
| Sca1 | D7 | BV605 or BV421 | Biolegend |



Figure 2.1. Representative gating strategy for ESLAM-Sca1+ HSCs. CD45 ${ }^{+}$CD48CD150+EPCR ${ }^{+}$Sca1+ (ESLAM-Sca1+) were isolated from lineage-depleted BM cells using the gating strategy described herein. First, cells were gated based on and the forward (FSC) and side (SSC) scatter. Cells were then gates to exclude cell aggregates by gating FSC against the trigger pulse width. Next, dead cells were excluded based on the retention of the 7AAD dye. Finally, ESLAM-Sca1+ cells were acquired by sequential gating of CD48/CD150, CD45/EPCR and Sca-1/EPCR.

Cells were sorted using either a MoFlo Astrios (Beckman-Coulter) or Influx (BD). ESLAM-Sca1+ HSCs from Fgd5 ${ }^{\text {zsGreen•ZsGeenr/t }}$ mice were sorted using CD45 BV421, CD48 APC, CD150 PE-Cy7, EPCR PE, Sca1 BV421 and 7AAD. For all other genotypes, CD45 FITC and Sca1 BV421 were used instead. All genotypes were sorted using the following filters: 530/40 (ZsGreen, FITC), 585/29 (PE), 670/30 (APC), 460/50 (BV421), 670/30 (7AAD) and 610/20 (BV605). A representative gating strategy is outlined in Figure 1. For single-cell assays, scRNA-seq experiments and small bulk
cultures, ESLAM-Sca1+ HSCs were deposited into $100 \mu L$ of appropriate medium in 96 well plates.

### 2.2.5. Isolation strategy of CD150 ${ }^{+}$and CD150 ${ }^{-}$cell fractions

To conduct the proteomic screen and matching bulk RNA sequencing analysis, LTHSCs were defined as Lineage - CD $45^{+}$CD48 ${ }^{-}$CD150 ${ }^{+} \mathrm{cKit}^{+}$(called CD150+ hereafter) and short-term HSCs (ST-HSCs) were defined as Lineage-CD45+CD48-CD150-cKit ${ }^{+}$ (called CD150- hereafter). The phenotype was selected to obtain a highly enriched LT-HSC fraction ( $\sim 20-30 \%$ ) while increasing cell yield by $\sim 4$ fold in order to undertake proteomic assays. LSK cells were defined as Lineage-Sca1+cKit ${ }^{+}$. FACS was performed as outlined above, using antibodies listed in Table 2. The biotinylated HSPC cocktail was used in conjunction with Streptavidin.

Table 2. CD150+ and CD150- HSC panel antibodies

| Antibody | Clone | Fluorochrome | Supplier |
| :--- | :--- | :--- | :--- |
| CD45 | 30-F11 | BV785 | Biolegend |
| CD48 | HM48-1 | APC | Biolegend |
| CD150 | TC15-12F12.2 | PE-Cy7 | Biolegend |
| cKit | 2B6 | APC-Cy7 | Biolegend |
| Mouse HSPC <br> cocktail | - | Biotin | STEMCELL Tech |
| Streptavidin | - | BV510 | Biolegend |

Cells were sorted using either a MoFlo (Beckman-Coulter) or Influx (BD). CD150+ LTHSCs and CD150 ST-HSCs were first stained with the biotinylated lineage cocktail (STEMCELL Technologies) and subsequently stained with the complete antibody cocktail, containing CD45 BV785, CD48 APC, CD150 PE-Cy7, cKit APC-Cy7, Streptavidin BV510 (Table 2). FACS was conducted using the following settings: 405 520/35 (BV510), 405 750LP (BV785), 640 750LP (APC-Cy7), 640 670/30 (APC) and 561 750LP (PE-Cy7). Cells were subjected to a 2-way sort into $500 \mu L$ PBS.


Figure 2.2. Representative gating strategy for CD150 ${ }^{+}$and CD150 HSCs. Gating strategy for
 Cells were separated from debris by gating the side scatter (SSC) against the side scatter (FSC), followed by exclusion of cell doublets and other cell aggregates (Trigger pulse width against FSC). Lineage negative cells were separated (Lin), followed by a CD45 and cKit double positive gate. Finally, LT-HSCs and ST-HSCs were separated by the CD150 surface marker, plotted against CD48. Both populations were purified simultaneously, using a two-way sort.

### 2.2.6. Isolation strategy for ELSK cells post ex vivo expansion

Cells, cultured for 28 days in expansion media ${ }^{225,226}$, were re-sorted using the EPCR ${ }^{+}$Lin $-S c a 1^{+}$CKit $^{+}$(ELSK) sorting strategy to isolate phenotypic HSCs ${ }^{333}$. FACS was performed as previously described, using antibodies listed in Table 3.

Table 3. ELSK panel antibodies

| Antibody | Clone | Fluorochrome | Supplier |
| :--- | :--- | :--- | :--- |
| EPCR | RMEPCR1560 | PE | STEMCELL Tech |
| Sca1 | D7 | BV605 | Biolegend |
| CD11b | M1/70 | APC | Biolegend |
| Gr-1 | RB6-8C5 | PE-Cy7 | Biolegend |
| cKit | 2B6 | APC-Cy7 | Biolegend |

Cells were sorted using an Influx (BD) and deposited into $500 \mu L$ PBS-10\%FCS (Figure 3). Non-ELSK cells were pooled from three fractions: 1) $\mathrm{Lin}^{+}$, 2) $\mathrm{Lin}^{-} \mathrm{Sca1}{ }^{-c \mathrm{cKit}}{ }^{-}$ , 3) EPCR-Lin-Sca1 ${ }^{+} \mathrm{CKit}^{+}$. The sort was conducted using filters described above.


Figure 2.3. Representative gating strategy for ex vivo cultured ELSK HSCs. Gating strategy for ELSK and non-ELSK fractions. ELSK gating is marked in red (gate 4). The non-ELSK fraction contained a pool of gates 1-3, marked in black. In brief, the ELSK fraction was defined as lineage- by a doublenegative gate for Mac1 and Gr1. Subsequently, cKit ${ }^{+}$and Sca-1+ gates formed the LSK phenotype. Finally, EPCR was gated against Fgd5zsGreen:ZsGeen. ELSK gating included both Fgd5+ and Fgd5- cells. Where FELSK was specified, only the Fgd5 ${ }^{+} E P C R^{+}$fraction was collected. Simultaneously, the Lin ${ }^{+}$(1), $\mathrm{Lin}^{-} \mathrm{SK}^{-}$(2) and $E P C R-L S K$ (3) populations were isolated and pooled to yield the nonELSK cell fraction.

### 2.3. Ex vivo HSC expansion cultures

### 2.3.1. Stemspan 10-day HSC maintenance cultures

Bulk Fgd5-ZsGreen ${ }^{+} E P C R^{+}$CD45 ${ }^{+}$cells were cultured in Stemspan conditions, previously described by Kent et al. ${ }^{49}$. Cells were cultured in $100 \mu \mathrm{~L}$ StemSpan Serumfree Expansion Medium (STEMCELL Technologies), supplemented with 1\% LGlutamine (Sigma), 1\% Penicillin/Streptomycin (Sigma), 0.2\% $\beta$-Mercaptoethanol (Life Technologies), $300 \mathrm{ng} / \mathrm{mL}$ mouse SCF (STEMCELL Technologies or BioTechne) and $20 \mathrm{ng} / \mathrm{mL}$ IL-11 (STEMCELL Technologies or Bio-Techne). Cells were deposited in round-bottom 96 -well plates (Corning) and kept at $37^{\circ} \mathrm{C}$ and $5 \% \mathrm{CO}_{2}$.

### 2.3.2. F12-based 28-day expansion cultures

Single and bulk expansion cultures for a period of 28 days were conducted as previously described ${ }^{225,226}$. Cells were deposited into either BioCoat Fibronectin 96 well paltes (Corning) or in tissue culture-treated flat-bottom plates (Falcon), manually coated with human Fibronectin (Millipore). Here, each well was coated with $10 \mu \mathrm{~g} / \mathrm{cm}^{2}$ Fibronectin (diluted in PBS) for 1 h at $37^{\circ} \mathrm{C}$ and subsequently washed with $50 \mu L$ PBS. Cells were cultured in $200 \mu L$ Ham's F12 nutrient mix (Thermo), supplemented with 1\% Insulin-Transferrin-Selenium-Ethanolamine (ITSX, Gibco), 1\% Penicillin/Streptomycin/L-Glutamine (PSG, Gibco), 0.1\% Polyvinyl alcohol (PVA, Sigma), 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, Gibco), $100 \mathrm{ng} / \mathrm{mL}$ mouse TPO (Peprotech) and $10 \mathrm{ng} / \mathrm{mL}$ mouse SCF (Peprotech). Plates were incubated at $37^{\circ} \mathrm{C}$ and $5 \% \mathrm{CO}_{2}$. Complete media changes were conducted after the first 5-6 days of culture and performed every 2-3 days thereafter. Where indicated, a $10 \%$ aliquot at day 27 of the culturing period was extracted for flow cytometry analysis.

### 2.3.3. F12-based 28 -day cultures, supplemented with signalling pathway modulators

Selected bulk expansion cultures were supplemented with various small molecular inhibitors and activators. 50 ESLAM HSCs were deposited per well in a 96-well plate assay. Cells were expanded for 28 days, as described in 2.3.2. However, F12-based culturing media was additionally supplemented with the following set of RhoGTPasemodulating molecules: 1) NSC23766 (Tocris), 2) Rhosin (Tocris), 3) CASIN (Tocris), 4) ML-099 (Merck). The molecules were supplied in three different concentrations, as indicated in the results section. Complete media changes, supplied with the appropriate inhibitor/activator, were first conducted after 5-6 days and then every 2-3 days thereafter.

### 2.3.4. F12-based 10-day expansion cultures

To facilitate short term 10-day cultures, cells were deposited into 96 -well round bottom plates (Corning), without fibronectin coating. Media formulations were consistent with the recipe previously describe in 2.3.2. No media changes were undertaken.

### 2.3.5. Hibernation cultures

Single ESLAM-Sca1+ HSCs were deposited by FACS into round bottom 96-well culturing plates (Corning) containing $100 \mu L$ hibernation media, as previously described ${ }^{334}$. Hibernation cultures were maintained for 7 days, as previously described with no cells undergoing division ${ }^{334}$.

### 2.3.6. Functionalised Stembond cultures

Preparation and functionalisation of Stembond culture plates and subsequent culture were performed by Dr Juan Rubio-Lara in accordance with the original Stembond recipe ${ }^{335,336}$. CELLview Plates (Greiner) were activated inside a Diener plasma cleaner for 7.5 min at $70 \%$ power, at a flow of 5 standard cubic centimetres per minute (SCCM) $\mathrm{O}_{2}$ and a 0.43 mBar pressure. Functionalisation was performed using a $5 \%$ bind-silane
solution (GE Healthcare), diluted in 10\% acetic acid and 85\% pure ethanol. Each well was covered with $30 \mu L$ of bind-silane solution and incubated for 90 min . The solution was removed by rinsing with absolute ethanol and dried. Gel precursor solution was combined as follows: $35 \mu L 40 \%$ acrylamide, $27.2 \mu L 2 \%$ Bis-acrylamide, $20 \mu L 2 M 6-$ acrylamidohexanoic acid (AHA), $410.3 \mu L \mathrm{dH}_{2} \mathrm{O}, 2.5 \mu L$ TEMED and $5 \mu L 10 \%$ ammonium persulfate (APS). The formulation achieved the desired stiffness of 750 Pascal. $50 \mu L$ of the solution was added per well and polymerised for 35 min at room temperature. Wells were rinsed twice with $100 \%$ methanol, followed by a wash with PBS. Next, gels were washed with MES buffer (19.5g MES hydrate and 29.2 g NaCl in $1000 \mathrm{~mL} \mathrm{dH}_{2} \mathrm{O}, \mathrm{pH} 6.1$ ) and gels were activated with freshly prepared EDAC/NHS solution [0.2M EDAC ( N -(3-Dimethylaminopropyl)- N '-ethylcarbodiimide hydrochloride, Sigma-Aldrich) and 0.5M NHS (N-Hydroxysuccinimide, Sigma-Aldrich), diluted in MES buffer] for 30 min . The gels were rinsed with chilled $60 \%$ methanol (diluted in PBS) and subsequently with 50 mM HEPES buffer ( pH 8.5 ). Gels were then coated with the target protein at a concentration of $200 \mu \mathrm{~g} / \mathrm{mL}$ in 50 mM HEPES buffer and incubated for 2 hours at room temperature or overnight at $4^{\circ} \mathrm{C}$. The protein solution was removed and rinsed with 50 mM HEPES. The reaction was blocked using 0.5 M ethanolamine in 50 mM HEPES and incubated for 30 min at room temperature. A final wash with 50 mM HEPES was performed. The following ECM proteins were used for Stembond functionalisation: human type I Collagen (GTX27533, Insight Biotechnology), Hyaluronan (GLR001, Bio-Techne), human type VI Collagen (GTX27538, Insight Biotechnology), native human Fibrinogen (229-20019, Cambridge Bioscience), Fibulin 1 (RPC472Ju01-50ug, 2B Scientific), human Plasminogen (MD-14-0071P, Cambridge Bioscience), human Fibrin (F5386-1G, Merck Life Science UK). Prior to seeding, gels were washed with cell culture media, as described in 2.3.2. While in cell culture media, gels were allowed to equilibrate for 2 hours at $37^{\circ} \mathrm{C}$ and $5 \% \mathrm{CO}_{2}$. Media was replaced before seeding cells. Cells were deposited into cell culture media and expanded for 28 days. Media changes were conducted as described in section 2.3.2.

### 2.4. Flow cytometry analysis of ex vivo expansion cultures

All ex vivo expansion cultures, conducted on tissue culture plastic or functionalised Stembond hydrogels, were analysed by flow cytometry. The single cell 28-day
expansion cultures were conducted and analysed by Dr James Che and Dr Juan Rubio-Lara. Cells were stained with 7AAD and antibodies outlined in Table 4. ELSK cells were defined as EPCR ${ }^{+}$Lin - Sca $1^{+} \mathrm{CKit}^{+} \mathrm{CD} 45^{+}$. To better estimate absolute cell counts, fluorescent beads (Trucount Control Beads, BD) were added to each well as described previously ${ }^{337}$. Flow cytometry was performed using a Beckman Coulter CytoFlex (with plate reader) or BD LSRFortessa, equipped with a High Throughput Sampler. Representative gating strategies for cultures conducted on tissue culture plastic (Figure 4) and 750 Pa Stembond hydrogel (Figure 5) are provided. Data analysis was performed using FlowJo (version 10) software (Treestar Inc).

Table 4. ELSK panel antibodies for flow cytometry

| Antibody | Clone | Fluorochrome | Supplier |
| :--- | :--- | :--- | :--- |
| EPCR | RMEPCR1560 | PE | STEMCELL Tech |
| Sca1 | D7 | BV421 or BV605 | Biolegend |
| CD11b | M1/70 | APC | Biolegend |
| Gr-1 | RB6-8C5 | PE-Cy7 | Biolegend |
| cKit | 2B6 | APC-Cy7 | Biolegend |
| CD45 | 30-F11 | FITC or BV421 | Biolegend |



Figure 2.4. Representative gating strategy for flow cytometry analysis of ELSK cells cultured on plastic. Single cells were gated for forward scatter (FSC) against 7AAD to exclude dead cells. Linage ${ }^{-}$ cells were defined as Mac1- and GR1-. Next, the Lin cells were gated for the cKit+Sca-1+ phenotype. Finally, the double positive population of CD45 and EPCR were defined as ELSK.


Figure 2.5. Representative gating strategy for flow cytometry analysis of ELSK cells cultured on functionalised Stembond hydrogels. ELSK cell fractions were gated and recorded in accordance to the gating strategy outlined on Figure 4.

### 2.5. Gene expression profiling by RT-qPCR

To perform reverse transcription quantitative PCR (RT-qPCR) for candidate genes, $10 \%$ of 28 -day bulk ex vivo cultures were extracted. RNA was extracted using the Picopure RNA Isolation Kit (Thermo Fisher) according to manufacturer's protocol. In turn, complementary DNA (cDNA) was synthesised using the SuperScript III FirstStrand Synthesis system, according to manufacturer's protocol. RT-qPCR was performed using TaqMan probes (Thermo Fisher). The following TaqMan probes were used: Prdm16 (Mm00712556_m1), Fst11 (Mm00433371_m1), Prex2 (Mm02747802_s1), Mpdz (Mm00447870_m1), Cebpa (Mm00514283_s1), Rab44 (Mm01306199_m1), Siglecf (Mm00523987_m1), Klhl4 (Mm00555463_m1), Gapdh (Mm99999915_g1), Palld (Mm01341202_m1), Ptk2 (Mm00433209_m1). The
reactions were run using the TaqMan Fast Advanced Master Mix (Applied Biosystems) and the ViiA 7 Real-Time PCR System (Applied Biosystems), using conditions according to manufacturer's protocol.

### 2.6. Calcium imaging

600 ESLAM-Sca1+ HSCs were sorted into Ham's F12-based mediate, formulated as previously described in section 2.3.2. Cells were maintained in culture for 7 days prior to imaging. Adhesive solution was prepared using $1 \mu l$ Cell-Tak Cell and Tissue Adhesive (Corning), diluted in $28.5 \mu L 0.1 \mathrm{M}$ sodium bicarbonate (pH 8.0, Sigma) and $0.5 \mu \mathrm{~L} 1 \mathrm{M} \mathrm{NaOH}$ for a total volume of $30 \mu \mathrm{~L}$ per sample. Microscopy cover slides were covered with $30 \mu L$ Cell-Tak solution per slide and incubated for 30 min at room temperature. Subsequently, the Cell-Tak solution was aspirated and cover slides were washed with $30 \mu l \mathrm{dH}_{2} \mathrm{O}$. $\mathrm{Ca}^{2+}$ buffers were prepared at $0.3 \mathrm{mM}, 1.5 \mathrm{mM}$ and 5 mM $\mathrm{CaCl}_{2}$ in HEPES. Cells were seeded onto Cell-Tak-coated cover slips for 40 min at $37^{\circ} \mathrm{C}$ and $5 \% \mathrm{CO}_{2}$. After seeding, cells were washed with the appropriate $\mathrm{CaCl}_{2}$ buffer. $8 \mu L$ Fura-2 dye (50029, Insight Biotechnology) was diluted in 1 mL appropriate $\mathrm{CaCl}_{2}$ buffer and added to the cells. In turn, Fura-2 loaded cells were incubated for 40 min at room temperature, washed with appropriate $\mathrm{CaCl}_{2}$ buffer (without Fura-2) and incubated for an additional 20 min . The subsequent calcium imaging protocol was conducted by Dr Andrew James in the lab of Dr. Will Brackenbury. Following Fura-2 loading, cells were mounted onto an epifluorescent microscope within a perfusion chamber. First, the baseline Fura-2 fluorescence was measured. Subsequently, the concentration of $\mathrm{Ca}^{2+}$ was equilibrated across the cell membrane using 10 mM ionomycin. The minimum and maximum fluorescence levels were determined by sequential perfusion with zero $\left[\mathrm{Ca}^{2+}\right]$ and saturating $20 \mathrm{mM} \mathrm{Ca}{ }^{2+}$ buffer. These values were used to calibrate the baseline cellular $\left[\mathrm{Ca}^{2+}\right]$ using the Grynkiewicz equation ${ }^{338}$.

### 2.7. Transplantation assays

Transplantation assays and peripheral blood analysis were performed by Dr James Che, Dr. Alyssa Cull and Dr. David Kent. Briefly, recipient W41 mice were between 8 and 24 weeks old and subjected to sub-lethal irradiation with a single dose of Caesium
irradiation (400cGy) or X-ray irradiation (360cGy). All transplantation were performed by intravenous tail vein injection of cell suspensions in 200-300 $\mu \mathrm{L}$ PBS, using a 29.5G insulin syringe. Transplantation assays were conducted over a period of 16 weeks ${ }^{334,339}$.

### 2.7.1. Peripheral blood analysis

Peripheral blood samples were extracted via tail veins using EDTA coated microvette tubes (Sarstedt AGF \& Co, Nuembrecht, Germany). Samples were taken every 4 weeks until the final 16 -week timepoint post transplantation. The samples were subject to red cell lysis using ammonium chloride (STEMCELL Technologies), following the protocol described in 2.2.2. Flow cytometry analysis was performed on the LSRFortessa (BD) analyser. Antibodies used for flow cytometry analysis are outlined in Table 5. The gating strategy has previously been described ${ }^{332,340}$ and representative plots are depicted in Figure 6. Data was analysed using FlowJo (version 10) and repopulation was defined as $>1 \%$ donor chimerism and $>1 \%$ contribution to the granulocyte-macrophage (GM) lineage.

Table 5. Antibody panel for peripheral blood analysis

| Antibody | Clone | Fluorochrome | Supplier |
| :--- | :--- | :--- | :--- |
| CD45.1 | A20 | AF700 | eBioscience |
| CD45.2 | 104 | APC-Cy7 | Biolegend |
| CD11b | M1/70 | PE-Cy7 or BV605 | Biolegend |
| B220 | RA3-6B2 | APC | Biolegend |
| Ly6G | 1A8 | BV421 | Biolegend |
| CD3 | 17A2 | PE | Biolegend |
| 7AAD | - | - | Life Technologies |



Figure 2.6. Representative gating strategy for peripheral blood flow cytometry analysis. The gating strategy was adapted from Che et al. ${ }^{341}$. In brief, all samples were gated to first exclude debris, then isolate single cells and finally dead cells. Subsequently, cells were gated for Ly6g and Mac1 to isolate Mac1+ granulocyte-macrophage (GM) cells. The cell population was separated between donor and recipient mice based on CD45.1 and CD45.2 isoforms. In parallel, B cells and T cells were defined as $C D 3^{-} B 220^{+}$and $C D 3^{+} B 220^{-}$respectively. The respective contribution of donor and recipient cells was determined by CD45.1 and CD45.2 isoforms.

### 2.8. RNA sequencing

### 2.8.1. Bulk RNA-seq library preparation and sequencing

Freshly isolated CD150 ${ }^{+}$and CD150 ${ }^{-}$HSCs or 28-day expanded ELSK and nonELSK fractions were subjected to bulk RNA sequencing. RNA was extracted using the

Picopure RNA isolation kit (Thermo) in accordance with manufacturer's protocol. Libraries were constructed using the SMARTer Stranded Total RNA-seq kit v2 - Pico Input mammalian (Takara Bio) following manufacturer's instructions. Quality control of libraries was performed using the Qubit RNA HS Assay Kit (Thermo) and run on the Bioanalyzer system (Agilent). Constructed libraries were sequenced using the Illumina Novaseq 6000 system using 50 bp paired end reads. Sequencing was performed by the Cancer Research UK Cambridge Institute Genomics Core.

### 2.8.2. Bulk RNA-seq raw data processing

Raw fastq files were first subject to adapter trimming, using trim_galore (parameters: --paired --quality 30 --clip_R2 3). Reads were aligned against the Mus musculus genome build (mm10) using STAR (default parameters). Gene counts were computed using HTSeq (parameters: --format $=$ bam - stranded $=$ reverse - type $=$ exon - mode = intersection-nonempty --additional-attr = gene_name).

### 2.8.3. Single cell RNA-seq library preparation and sequencing

Single cell RNA sequencing (scRNA-seq) was performed for WT and $T E T 2^{-/}$ESLAMSca1+ cells, fetal liver and hibernating HSCs. scRNA-seq sample and library preparation was conducted as previously described in the Smart-Seq2 protocol ${ }^{342}$. Cell were deposited by FACS into lysis buffer [0.2\% Triton X-100 (Sigma), RNAse inhibitor (SUPERase, ThermoFisher), nuclease-free water (ThermoFisher)] in 96-well plates. Libraries were constructed using the Illumina Nextera XT DNA preparation kit (Illumina). Reverse transcription, PCR preamplification, PCR purification, tagmentation reaction, Tn5 transposase tripping, adapter-ligated fragment amplification, PCR purification, quality control and pooling were performed as previously described ${ }^{342}$. Library preparations were performed by Dr Alyssa Cull or Dr James Che. Single end, 50 bp libraries were pooled and sequenced using Illumina HiSeq 4000. Sequencing was performed by the Cancer Research UK Cambridge Institute Genomics Core.

### 2.8.4. Single cell RNA-seq raw data processing

Raw fastq files were first aligned to the Mus musculus genome build (mm10) using STAR (default setting) and the output was generated in SAM format. Aligned reads were sorted, converted to BAM format and gene counts were computed using featureCounts (default settings). Raw data processing was conducted by Dr John Davey, Dr Iwo Kucinski and Daniel Bode.

### 2.9. Proteomic Analysis

### 2.9.1. LC-MS/MS Sample Preparation

Hoxb8-FL cells were cultured in RPMI 1640 (Sigma R8758) media, supplemented with 10\% FBS (Gibco), 0.1\% mercaptoethanol (Invitrogen), 1\% penicillin-streptomycin (Sigma), $1 \%$ glutamine (Sigma), $1 \mu \mathrm{M}$ estradiol and 5\% FLT3L conditioned media from the B16-FL cell line. Cells were maintained in culture at concentrations of $10^{5}-10^{6}$ cells $/ \mathrm{mL}$. Prior to proteomic analysis, Hoxb8-FL cells were re-suspended in phosphate buffered saline (PBS). Cell aliquots ranged from 500 to 200,000 cells in $20 \mu \mathrm{~L}$ PBS. Cell lysis was performed using $2 \%$ sodium dodecyl sulphate (SDS) with subsequent boiling at $95^{\circ} \mathrm{C}$. Cell lysates were sonicated for 8 min and dried using a Savant SpeedVac Concentrator (Thermo). Samples were re-suspended in 100 mM triethylammonium bicarbonate (TEAB) buffer. Subsequently, disulphide bonds on cysteine residues were reduced with a final concentration of 5 mM tris-2carboxymethyl phosphine (TCEP) and incubated at $60^{\circ} \mathrm{C}$ for 1 h . The reaction was followed by incubation with a final concentration of 10 mM lodoacetamide (IAA) for 30 min in the dark. Prior to labelling, each TMT 10-plex reagent (Thermo) was resuspended in $41 \mu \mathrm{~L}$ anhydrous acetonitrile. Each replicates of reduced and alkylated cell lysates were labelled with $10 \mu \mathrm{~L}$ of a single TMT regent. Labelling was performed at room temperature for 1 h before quenching with $8 \mu \mathrm{~L} 5 \%$ hydroxylamine. At this point, the 10-plex samples were pooled. For spike-in experiments, $100 \mu \mathrm{~g}$ bovine serum albumin (BSA) was added at this stage. 100\% (w/v) Trichloroacetic acid (TCA) was added to the sample mixture at a ratio of 1 to 4 , followed by incubation for 10 min . The sample was centrifuged at $14,000 \mathrm{rpm}$ and the resulting protein pellet was resuspended in 100 mM TEAB buffer. Trypsin (Pierce, MS grade) was added for an
overnight digestion at $37^{\circ} \mathrm{C}$. Samples that underwent co-digestion with the endoproteinase gluC were incubated with gluC (New England Biolabs) overnight at $37^{\circ} \mathrm{C}$. Samples were dried using a Savant SpeedVac Concentrator for storage at -20 ${ }^{\circ} \mathrm{C}$. Prior to fractionation or LC-MS/MS analysis, dried peptide mixtures were resuspended in $40 \%$ acetonitrile.

### 2.9.2. Peptide fractionation

High pH Reversed-Phase (RP) fractionation was performed with the Waters XBridge C18 column ( $2.1 \times 150 \mathrm{~mm}, 3.5 \mu \mathrm{~m}, 120 \AA$ ) on a Dionex Ultimate 3000 HPLC system. Ammonium hydroxide at $0.1 \%$ was used as mobile phase $A$ and mobile phase $B$ was set as $100 \%$ acetonitrile / $0.1 \%$ ammonium hydroxide. The peptide mixture was reconstituted in $100 \mu \mathrm{~L}$ mobile phase A and subjected to gradient elution at $200 \mu \mathrm{~L} / \mathrm{min}$ as follows: 5 minutes isocratic at $5 \%$ B, for 15 min gradient to $35 \% \mathrm{~B}$, for 5 min gradient to $80 \% \mathrm{~B}$, isocratic for 5 minutes and re-equilibration to $5 \%$ (B). The chromatogram was recorded at 215 and 280 nm and fractions were collected every 1 minute.

### 2.9.3. LC-MS/MS Analysis

LC-MS/MS analysis was performed on the Dionex Ultimate 3000 UHPLC system coupled with the Orbitrap Lumos Mass Spectrometer (Thermo Scientific). Each peptide fraction was reconstituted in $10 \mu \mathrm{~L} 0.1 \%$ formic acid and $7 \mu \mathrm{~L}$ were loaded on the Acclaim PepMap 100, $100 \mu \mathrm{~m} \times 2 \mathrm{~cm}$ C18, $5 \mu \mathrm{~m}$, trapping column with the $\mu$ IPickUp method at flow rate $10 \mu \mathrm{~L} / \mathrm{min}$. The sample was then subjected to a multistep gradient elution on an EASY-Spray ( $75 \mu \mathrm{~m} \times 50 \mathrm{~cm}, 2 \mu \mathrm{~m}$ ) C18 capillary column (Thermo) at $45{ }^{\circ} \mathrm{C}$. Mobile phase A was $0.1 \%$ formic acid and mobile phase B was $80 \%$ acetonitrile / $0.1 \%$ formic acid. The gradient separation method at flow rate 300 $\mathrm{nL} / \mathrm{min}$ was as follows: for 90 min gradient $5 \%$ to $38 \% \mathrm{~B}$, for 5 min up to $95 \% \mathrm{~B}$, for 5 min isocratic at $95 \% \mathrm{~B}$, re-equilibration to $5 \% \mathrm{~B}$ in 5 min , for 10 min isocratic at $5 \% \mathrm{~B}$. Precursor ions were selected with mass resolution of 120 k , AGC $4 \times 10^{5}$ and max IT 50 ms in the top speed mode within 3 sec . Peptides were isolated for HCD fragmentation with quadrupole isolation width 0.7 Th and 50k resolution. Collision energy was set at
$38 \%$ with AGC $1 \times 10^{5}$ and max IT 105 ms . Targeted precursors were dynamically excluded for further isolation and activation for 45 seconds with 7 ppm mass tolerance.

### 2.9.4. Protein identification and quantification

The raw data were submitted to SequestHT database search in Proteome Discoverer 2.2 for protein identification and quantification. The precursor mass tolerance was 20 ppm and the fragment ion mass tolerance was 0.02 Da . Spectra were searched for fully tryptic peptides with no more than 2 miss-cleavages and minimum length of 6 amino acids. TMT6plex at N-termini/K and Carbamidomethyl at C were used as static modifications. Dynamic modifications were oxidation of $M$ and Deamidation of N/Q. Peptide FDR was set at 0.01 and validation was based on q-value and decoy database search in the Percolator node. All spectra were searched against a UniProt fasta file containing 16,945 reviewed mouse entries. The Reporter Ion Quantifier node included a custom TMT-10plex Quantification Method with integration window tolerance 15 ppm . Only peptides uniquely belonging to protein groups were used for quantification. Scaled quantitative values were obtained by dividing each TMT signal-to-noise (S/N) value with the mean TMT $(S / N)$ across samples per protein (divide by row mean $\times$ 100). All plots were derived using scaled quantitative values of proteins consistently quantified across the 10-plex. The coefficient of variation (CV) was computed as follows:

$$
C V(\%)=\left(\frac{\sigma}{\mu}\right) \times 100
$$

where $\sigma$ is the standard deviation, $\mu$ is the mean across the 10 -plex and the CV was expressed in percentile.

### 2.10. RNA-seq data analysis

Data analysis was performed using Python (version 3.8) and $R$ (version 4.0.3). Visualisation was performed using ggplot2 (version 3.3.5), base $R$ functions, EnhancedVolcano (version 1.4.0), heatmap3 (version 1.1.9), VennDiagram (version 1.6.2), Seurat (version 4.0.0), Limma (version 3.42.2), scanpy (1.6.1).

### 2.10.1. Published datasets

Beside newly generated data, multiple previously published dataset have been utilised throughout the analysis. These included:

1) A single cell landscape of haematopoietic stem and progenitor cells, generated by Nestorowa and colleagues ${ }^{22}$.
2) Comprehensive scRNA-seq profiling of mouse LK/LSK cells ${ }^{343}$.
3) scRNA-seq dataset of freshly isolated and hibernating HSCs, including their SCF-stimulated counterparts ${ }^{334}$.
4) Comprehensive scRNA-seq profiling of 28-day expanded bulk HSC culture (unpublished, courtesy of Dr Iwo Kucinski and Dr Adam Wilkinson).

### 2.10.2. Bulk RNA-seq quality control and processing

Downstream analysis of raw counts was conducted using edgeR ${ }^{344,345}$. First, the total number of read counts per sample was computed. Samples with library sizes < 30\% than the median library size were excluded from the analysis. Read counts were transformed to counts per million (cpm) and genes with fewer than two libraries expressing a minimum of 1 cpm were removed. Next, read counts were normalised using the trimmed mean of $M$ values (TMM) algorithm ${ }^{346}$.

### 2.10.3. Bulk RNA-seq batch correction

ELSK and nonELSK cell fractions were sequenced in three separate runs. To correct for batch effects, each sequencing run contained a technical replicate across at least two sequencing runs. Batch correction was computed using Limma (version 3.42.2) ${ }^{347}$. Following data integration, batch corrected counts were log-transformed and utilised for downstream processing.

### 2.10.4. scRNA-seq quality control and processing

Quality control of scRNA-seq profiles was performed according to the following set of criteria:

1) Threshold of mapped reads per cell: $10^{5}-3 \times 10^{7}$ including nuclear genes, mitochondrial genes and ERCC spike ins).
2) A minimum of $20 \%$ of reads mapping to protein-coding genes.
3) Threshold for the proportion of mitochondrial genes was set to $>0.075$, to ensure that at least $7.5 \%$ of reads were mapped to protein coding, nonmitochondrial genes.
4) Cutoff for the proportion of reads mapped to ERCC spike ins was set to $5 \%$.
5) Cells with more than 1800 gene identifications were excluded.

Cells passing quality control were subsequently processed using the Seurat package (version 4.0.0). Raw counts were normalised using a scaling factor of 10,000, a variable feature count of 7,500 was selected and data was scaled using default parameters.

### 2.10.5. Differential gene expression analysis

Differential gene expression (DGE) was computed using a likelihood ratio test approach. In brief, a negative binomial generalised linear model (GLM) was fitted and DGE was computed using Limma (version 3.42.2). For the screening of ELSK and nonELSK cell fractions, genes were considered differentially expressed when $\log _{2}$ FC $\log _{2} F C \geq 2$ and $F D R \leq 0.05$. For screening of CD150 ${ }^{+}$and CD150 cells from wildtype and TET2-deficient animals, DGE cut-offs were set to $\log _{2} F C \geq 0.5$ and $F D R \leq$ 0.05 to match the LC-MS/MS analysis. DGE analyses for scRNA-seq datasets were performed using DeSeq2, executed via Seurat (version 4.0.0). To match LC-MS/MS data, DGE of scRNA-seq profiles for ESLAM-Sca1+ WT and TET2-/ HSCs were considered differentially expressed according to the following cut-offs: $\log _{2} F C \geq 0.5$ and adj $p$ val $\leq 0.05$. p-values were adjusted for multiple testing using BenjaminiHochberg correction. Furthermore, the previously published LK/LSK transcriptional landscape was used to derive key marker genes for Hoxb8-FL cells. Significantly enriched genes in Leiden clusters 4, 8 and 15 were computed using scanpy (version 1.6.1).

### 2.10.6. Gene signature scoring

Gene set signature scoring for bulk RNA-seq samples was performed for the LT-HSC specific molecular overlap (MolO) signature ${ }^{332}$ and the newly derived repopulation signature (RepopSig). For this purpose, signature genes passing the threshold for expressed genes, defined as $\geq 1 \mathrm{cpm}$ in at least 2 samples, were extracted. Subsequently, the geometric mean was computed, using log-transformed expression values for signature genes. Gene signature scoring of the RepopSig and MoIO signature in scRNA-seq datasets was performed using Seurat (version 4.0.0). In brief, the average expression of the gene set was computed for individual cells and subtracted by aggregate expression values of a control set, comprised of randomly selected genes. Furthermore, cell cycle scoring was performed on scRNA-seq profiles of WT and TET2-/ ESLAM-Sca1+ HSCs using Seurat (4.0.0). Key cell cycle genes were previously defined by Tirosh et al. ${ }^{348}$.

### 2.10.7. Principal Component analysis

Principal component analyses (PCA) were performed using the PCAtools R package (version 1.2.0). Prior to the computation of the PCA, histogram plots of all bulk RNAseq profiles were performed to ensure a Gaussian distribution. Where appropriate, lowly expressed genes were removed by setting a cut-off 40 cpm (sum across all samples). Subsequently, the bottom 10\% of non-variable genes were excluded from PCA computations. In addition to principal components (PCs) plots, loading plots were computed to identify genes driving separation across the respective PCs. For this purpose, the top 200 variable genes were plotted. Genes were selected based on directionality along PC1-2 and a 0.05 cut-off. Furthermore, Eigen plots were computed to identify the proportion of variability explained by each PC. Finally, the correlation between PCs and repopulation metadata (repopulation outcome, donor chimerism, GM contribution, T cell contribution, B cell contribution) and the MolO score were computed. For this purpose, Pearson correlation coefficients and the respective $r^{2}$ values were derived.

### 2.10.8. Correlation scoring using SingleR

SingleR (version 1.0.6), a tool developed for cell type identification and deconvolution based on scRNA-seq profiles, was adapted to identify prevalent cell types in ELSK and nonELSK bulk transcriptomes ${ }^{349}$. The correlation of each ELSK and nonELSK cell fraction against the curated ImmGen reference dataset was computed using default parameters ${ }^{350-352}$. Correlation with a broad set of HSPC cell types was visualised.

### 2.10.9. GO term and KEGG pathway analysis

Gene ontology (GO) term enrichment and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis were performed using Entrez gene identifiers, derived from the mouse genome annotation database (org.Mm.eg.db, version 3.10.0). GO term enrichment analyses were computed using the Limma package (version 3.42.2) and biological processes with significant enrichment, defined as adj p val $\leq$ 0.05 , were extracted. Similarly, KEGG pathway analyses were derived using Limma (version 3.42.2) and the KEGG annotation database (version 3.2.3). Significance of pathway enrichment was defined as adj p val $\leq 0.05$.

### 2.10.10. Reactome pathway analysis

Reactome pathway analysis was performed using the Reactome PA tool (version 1.30.0). Pathway enrichment was computed using Entrez gene identifiers and a significance cut-off of $p$ val $\leq 0.05$ was selected.

### 2.10.11. Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) was computed with the GSEA software (US San Diego and Broad Institute) against previously defined gene sets of various haematopoietic cell types ${ }^{353,354}$. Hematopoietic cell gene sets were acquired from Chambers et al. and Venezia et al. ${ }^{355,356}$.

### 2.10.12. scRNA-seq landscape projections

Single cell RNA-seq profiles of fetal liver HSCs and hibernating HSCs were projected onto the scRNA-seq landscape of HSPCs. To enable data integration, both datasets were pre-processed in an identical manner, including normalisation, variable gene selection and scaling. Here, the same parameters were used as outlined in section 2.10.4. Subsequently, anchors between the reference and query datasets were computed to enable integration. Transfer anchors were computed using (Seurat, version 4.0.0) and the following parameters: 30 PCs for reference computation, L2 normalisation was performed, the annoy method was used to compute nearest neighbours, 30 dimensions utilised for definition of neighbour search space, 5 neighbours were used for anchor identification, 200 neighbours for anchor filtering, 30 neighbours for scoring anchors, maximum 200 features used for specifying the neighbourhood search space and 50 trees were used. The acquired UMAP coordinates were plotted for the reference dataset and the quesry dataset simultaneously. Where stated, a selection of FL HSCs mapping outside the stem cell cluster were removed.

### 2.10.13. Direction of Transition (DoT) Scoring

In order to deconvolute bulk RNA-seq profiles and identify specific cell types enriched in each complex bulkj transcriptome, direction of transition (DoT) scores were computed ${ }^{357}$. First, differentially expressed genes between 1) PosELSK vs NegELSK, 2) PosNonELSK vs PosELSK and 3) NegNonELSK vs PosELSK were extracted and subsequently utilised for DoT scoring. A previously published scRNA-seq dataset of mouse LK/LSK cells ${ }^{343}$ was utilised as the reference landscape. DoT scoring was performed as described by Kucinski et al. ${ }^{357}$. A naïve stem and progenitor cell cluster, adjacent to LT-HSCs, was set as the point of origin. Data analysis was performed using Python (version 3.8) and scRNA-seq landscape projections were computed and visualised using scanpy (version 1.6.1).

### 2.10.14. Logistic and linear regression models

Linear and logistic regression models were computed to facilitate selection of key repopulation signature genes. Three transplantation metadata were used for fitting models against gene expression profiles of the top 50 signature genes. Logistic regression models were fitted for the binary assignment of repopulation (defined as > $1 \%$ donor chimerism and > 1\% GM contribution). Such models were fitted using logistf (version 1.24) using Firth's penalized maximum likelihood and alpha $=0.05$. To model the numeric proportions of donor chimerism and GM contribution, linear regression models were applied using base R functions. The coefficients, standard errors and $p$ values were extracted. To curate the repopulation signature, coefficients were plotted and the standard error indicated. All genes with coefficients within the lower bound of the standard error of the gene with the highest coefficient for each transplantation parameter were identified as enriched. Genes fitting such criteria for at least 2 of 3 metadata parameters were included in the final repopulation signature.

### 2.10.15. Pairwise gene correlations

Pairwise Pearson correlations were derived for genes identified by bulk RNA-seq in ELSK and nonELSK cell fractions against the geometric mean of the repopulation signature. The acquired $p$-values were corrected for multiple testing using the Benjamini-Hochberg method.

### 2.11. Proteomic data analysis

Data analysis was performed using $R$ (version 4.0.3). Visualisation was performed using ggplot2 (version 3.3.5), base R functions, EnhancedVolcano (version 1.4.0), heatmap3 (version 1.1.9), VennDiagram (version 1.6.2), Limma (version 3.42.2).

### 2.11.1. Protein interaction network analysis and data visualisation

Protein interaction networks were computed using the String database (version $10.5)^{358}$, using a minimum interaction confidence of 0.4 based on text mining, direct
experimental evidence and database searches only. Only first order interactions were computed. The networks were exported to include node coordinates. The network was subsequently visualised in Cytoscape (version 3.6.1) and non-interacting proteins were removed from the network visualisation.

### 2.11.2. Differential protein expression

Data analysis was performed in R (version 4.0.3). Differential protein expression (DPE) was computed between two biological conditions (in duplicates). To determine the DPE between two cell fractions, the log fold change (logFC) was computed between each biological replicate (A1:B1, A1:B2, A2:B1, A2:B2). Subsequently, a One Sample T Test was computed to determine the significance of the logFC deviation. Finally, the average logFC was computed.

### 2.11.3. GO term, KEGG and Reactome pathway analysis

GO term enrichment and KEGG pathways analysis were performed as previously described in section 2.10.9. For this purpose differentially expressed proteins were converted to their respective ENTREZ gene identifiers. Enrichment analyses were subsequently performed using the Limma package (version 3.42.2). Similarly, Reactome pathway analysis was computed with ENTREZ gene identifiers and the Reactome PA tool (version 1.30.0), as previously described in 2.10.10.

### 2.11.4. Principal component analysis and hierarchical clustering

PCA was performed as previously described in 2.10.7, using the PCAtools package (version 1.2.0). Hierarchical clustering was performed using the base R functions and default parameters.

### 2.11.5. Proteome-Transcriptome correlation

To derive the dynamic range of the proteome and transcriptome of Hoxb8-FL cells, average expression values were extracted for genes, identified by both omic screens.

Subsequently, pairwise Pearson correlations were computed for LC-MS/MS and scRNA-seq data separately. To correct for multiple testing, $p$-values were adjusted using the Benjamini-Hochberg method. Pearson's $r$ values were displayed in form of a density plot.

To determine the correlation between the proteome and transcriptome, LC-MS/MS and bulk RNA-seq profiles of WT and TET2-deficient CD150+ and CD150- cell fractions were compared. To enable comparability, DPE and DGE (i.e. $\log _{2} \mathrm{FC}$ ) between two biological samples (e.g. CD150+WT against CD150-WT) were plotted. The correlation between the LC-MS/MS-derived $\log _{2} \mathrm{FC}$ s and RNA-seq-derived $\log _{2}$ FCs was computed using a linear model ( R , version 4.0.3).

### 2.11.6. Other statistical analyses

If not stated otherwise, student's t-test or one-way ANOVA were utilised as the standard statistical testing. If not otherwise specified, it can be assumed these statistical tests have been utilised.

## Results

## Chapter 3. Adapting multiplex LC-MS/MS to facilitate proteomic screening of rare cell populations

To build a comprehensive protein network to inform HSC fate we opted for an initial large-scale, unbiased approach. For this purpose, we aimed to characterise the cellular proteome using a mass spectrometry approach. The scarcity of HSCs in combination with the finite resource of primary murine cells has historically prevented widespread adoption of proteomic profiling. Cabezas-Wallscheid and colleagues pioneered one of the first unbiased proteome screens of long-term HSCs (LT-HSCs) by isolating phenotypic HSCs (Lin ${ }^{\text {neg }} \mathrm{Sca}-1^{+} \mathrm{C}-\mathrm{Kit}^{+} \mathrm{CD} 34-\mathrm{CD} 48^{-C D 150}{ }^{+} \mathrm{CD} 135^{-}$) and comparing these with multipotent progenitor cells that acquired CD34 cell surface expression ${ }^{359}$. The proteomic approach by Cabezas-Wallscheid et al. required $4 \times 10^{5}$ cells for each proteomic mass spectrometry replicate using stable isotope dimethyl labelling and the authors identified $\sim 6000$ proteins. The subsequent drive to further characterise the HSC proteome yielded intriguing insight into key proteins governing mouse and human HSCs ${ }^{326,328,360}$. However, these studies also relied on at least 100,000 cells per replicate, thus requiring a trade-off between cell frequency and HSCcontent purity. In brief, current LT-HSC cell isolation strategies achieve purities of ~60\% functional HSCs ${ }^{117,332}$. While such purity is beneficial for functional characterisation, only a maximum of 2,000 LT-HSC can be extracted per mouse. Consequently, approximately 200 mice would be required to acquire $4 \times 10^{5}$ LT-HSCs per replicate.

To identify key drivers of HSC self-renewal it is vital to dissect HSC heterogeneity. Improved HSC isolation strategies have allowed us to address some of the longstanding questions about the identity of key self-renewal drivers in heterogeneous HSC populations or their role in HSC fate choice regulation. These have been largely
addressed using single cell functional assays and single cell transcriptomics. It would thus be of great interest to complement such studies by introducing the additional dimension of the proteome of HSCs with varying self-renewal potency and functional stem cell frequency. However, such a goal was subject to technical limitations, as conventional liquid chromatography with tandem mass spectrometry (LC-MS/MS) protocols required hundreds of thousands or millions of cells per sample. To overcome this limitation, this Chapter sets out to develop a novel proteomic tandem mass spectrometry approach to significantly reduce the input cell number. The work to establish a reliable LC-MS/MS protocol for rare cell populations was conducted in close collaboration with Dr Theodoros Roumeliotis and Professor Jyoti Choudhary of the functional proteomics group at the Institute of Cancer Research, London where I undertook a portion of my laboratory bench work.

The lack of reliable proteomic screening tools for rare cell populations sparked broad efforts to develop new methods for LC-MS/MS. As outlined in the Introduction Chapter 1.6, pioneering work by Budnik et al., Specht et al. and Brunner et al. provided amongst the first single cell proteomics protocols ${ }^{309,318,361}$. These ground-breaking studies are paving the way for single-cell proteomics but few protocols exceed more than 1,000 identified proteins ${ }^{311,318,361}$. The similarity in gene expression profiles of functionally heterogeneous HSC populations has been well documented in both the mouse and human HSC systems ${ }^{332}$. To decipher HSC heterogeneity and identify key drivers of HSC self-renewal, we reasoned that the depth of the proteome would be of paramount importance and protocols yielding below 1,000 proteins would not suffice for our biological question. Hence, our method development efforts were focused on developing a protocol that enabled quantification of multiple thousands of proteins from small bulk samples. Of note, an elegant and recent study by Amon et al. ${ }^{327}$ revealed parallel efforts to reduce required cell input for proteomic screening of primary haematopoietic stem and progenitor cells. Using samples of 25,000 cells, the authors achieved similar levels protein identification and quantification to our protocol.

### 3.1. TMT labelling provides sufficient accuracy to enable down-scaling of cell numbers

To avoid using primary murine HSCs for technique optimisation, we selected the Hoxb8-FL cell line, a murine haematopoietic progenitor cell line with the capacity to differentiate into both myeloid and lymphoid cells ${ }^{362}$. Before optimising the current protocol, we set out to identify the stage at which reduction of cell number would disproportionally affect the total protein abundance. Hoxb8-FL cell pools ranging from 200,000 to 500 cells were multiplexed in two separate runs (sample preparation protocol is outlined in the methods). We used the isobaric tandem mass tag (TMT) ${ }^{313}$ to label 5 samples for each run: (I) 200,000 - 10,000 cells (Figure 1A) and (II) 10,000 - 500 cells (Figure 1B). In principle, each replicate/sample is labelled with a unique isobaric TMT tag, which consists of a reactive group, a mass reporter and a mass normaliser. The mass normaliser adjusts all tags to have the same size, thus ensuring that the same peptides from different samples (labelled with different tags) are isolated together in the precursor ion scan. Upon fragmentation of isolated peptide peaks (MS/MS), the tag is fragmented to release the mass reporter. The unique mass reporter size then allows quantification of sample-specific peptides. Interestingly, we observed a linear correlation between the cell number and the total protein abundance (Figure 1A/B). The correlation coefficient for the 200,000-10,000 cell multiplex ( $r^{2}=$ 0.964 ) was marginally higher than the $10,000-500$ cells multiplex ( $r^{2}=0.907$ ). We thus concluded that there was scope for reducing cell numbers and did not observe a distinct point at which protein abundance was disproportionally reduced.


Figure 3.1. Detected protein abundance is proportional to cell number. (A) Relative total protein abundance was computed in the form of a ratio to the highest abundance reading. 200k, 100k, 50k, $10 k, 5 k$ Hoxb8-FL cells were multiplexed. The line of best fit (red) and the associated Pearson $r^{2}$ value are indicated. Samples were multiplexed using TMT labelling. (B) Separate run of $10 k, 5 k, 2 k, 1 k, 500$ Hoxb8-FL cells using the same strategy outlined in (A).

### 3.2. Current TMT-based LC-MS/MS protocols do not support screening of small sample sizes

The above results confirmed that, theoretically, cell input could be reduced to as little as 500 cells without comprising the linear relationship between total extracted protein and the cell number. Consequently, we set out to establish the baseline of protein identification and sample recovery when subjecting small cell numbers to the standard LC-MS/MS protocol. For this purpose, we multiplexed 6 technical replicates (6-plex) of 5,000 Hoxb8-FL cells and performed the standard TMT labelling-based proteomic protocol (see methods). The variation in total protein abundance highlighted significant peptide losses throughout the sample preparation process for two fractions (Figure 2A). These can lead to inconsistency for protein group quantification and thus high variation. Most importantly, when inspecting the number of proteins reliably identified and quantified in the 6-plex replicate run, we acquired identifications for less than 150 proteins of which only 92 were quantified (Figure 2B). Surprisingly, the previous multiplex run of (I) 200,000 - 10,000 cells (High Run) also yielded fewer than 600 identified proteins (Figure 2B). These outcomes are several orders of magnitude below our target of 3,000 quantified proteins and we therefore concluded that the standard TMT multiplex protocol was not sufficient to inspect the proteome at low cell numbers.


Figure 3.2. The standard TMT multiplex protocol does not suffice for the application to low cell number experiments. (A) Multiplex of 6 technical replicates, each containing 5000 cells. Total protein abundance for each labelled replicate is indicated ( $\times 10,000$ ). Category names reflect unique TMT10 mass tag labels. (B) Number of proteins reliably identified (red) and quantified (blue) across three experiments: I) 200k, 100k, 50k, 10k, 5k Hoxb8-FL cell multiplex (High run), II) 10k, 5k, 2k, 1k, 500 Hoxb8-FL cell multiplex (Low run), III) 6 replicates of 5,000 Hoxb8-FL cells ( 5 k Multiplex).

### 3.3. TCA protein precipitation reduced protein loss throughout the sample handling process

The significant reduction in total protein abundance for several multiplexed samples suggested significant sample loss during the handling process. To reduce the impact of biased sample loss, we introduced protein level TMT-labelling and Trichloroacetic acid (TCA) protein precipitation to the sample preparation protocol. By labelling directly at the protein level (lysine residues), smaller individual protein amounts can be combined early in the protocol. In principle, this allows handling up to 10 times higher total protein amounts throughout the rest of the protocol.

Furthermore, TCA facilitates hydrophobic aggregation of proteins and thus allows to precipitate these from other solvents and contaminants. In addition to the advantage of minimising protein losses due to reduced sample handling, TCA precipitation allows removal of the sample buffer, thereby allowing for direct LC-MS analysis without peptide clean-up (which is require for the standard single-shot protocol). Since fluorescence activated cell sorting (FACS) would be required for isolating primary

HSCs, and these cells require a salt buffer (e.g. PBS) to maintain cellular integrity, salts would have to be removed prior to analysis as they would significantly interfere with mass spectrometry results ${ }^{363}$. Conventional methods, most commonly centrifugal pelleting of cells, were not appropriate for low cell numbers due to extensive loss of material (data not shown) and the use of TCA precipitation step was therefore additionally useful to help remove the PBS contaminants.

First, to assess the efficiency of TCA precipitation in our protocol using relatively small cell numbers, we performed a label-free LC-MS/MS experiment on 200,000 Hoxb8-FL cells (Figure 3A). The number of cells was selected to simulate the total amount obtained by 20,000 cells across 10 combined samples. We acquired identifications for nearly 4,500 proteins, of which $\sim 1,500$ had unique peptide counts of above 3 (Figure 3B). Similarly, over $50 \%$ of identified proteins had peptide spectral match (PSMs) counts above 3 (Figure 3B). Thus, we concluded that TCA protein precipitation achieves effective removal of PBS and lysis buffers and permits direct LC-MS analysis and adequate protein identification rates. Alternatively, other precipitation methods, such as acetone or methanol chloroform precipitation, could be used ${ }^{364}$. However, the superior performance of TCA precipitation in combination with sonication prompted us to utilise this approach ${ }^{364}$.


Figure 3.3. Bulk label-free LC-MS/MS shows efficacy of TCA protein precipitation. (A) Flowchart illustrating the label-free protocol used for a single $2 \times 10^{5}$ cell sample. Cells were sonicated and lysed using $2 \%$ SDS. The lysate was subject to TCA precipitation, followed by digestion and analysis by LCMS/MS. (B) Plot indicating the number of proteins identified from $2 \times 10^{5} \mathrm{Hoxb8}-\mathrm{FL}$ cells. The number of unique peptides for each respective protein identification and the number of peptide spectrum matches (PSMs) is indicated.

### 3.4. Introduction of a BSA spike-in improves protein quantification and reduces stochastic variability

Aside from the number of proteins identified and quantified, it is critical to assess the stochasticity of the quantification rate within a multiplexed run. Large variability of intramultiplexed replicates would render direct comparisons impossible. Using the initial 6plex analysis of 5,000 cell replicates, we computed the coefficient of variability (CV) (expressed in percentile, \%CV) across replicates within the same multiplex for each protein and we plotted the frequency in a histogram (Figure 4A). We found significant variability between identical replicates, suggesting that our first approach suffered from high levels of stochastic noise and large sample losses. In contrast to nucleic acid sequencing techniques, proteomics does not benefit from sample amplification
technologies. Therefore, it was clear that we needed to further minimise sample losses throughout handling to enable screening of LT-HSCs.


Figure 3.4. Introduction of a BSA spike-in reduces sampling stochasticity. (A) The coefficient of variability across the multiplexed replicates was computed for each quantified protein for the 5 k 6 -plex cell run. The variability coefficients (CVs) are expressed as percentiles and the histogram displays their frequency within bins of $1 \%$. (B) Variability coefficient histogram for 10 -plex of $5 k$ cells using a BSA spike-in. (C) Variability coefficient histogram for 10-plex of 10k cells using a BSA spike-in. (D) Variability coefficient histogram for 10-plex of 15 k cells using a BSA spike-in. (E) The number of proteins identified (red) and quantified (blue) for the original 5k multiplex, 5k BSA spike-in, 10k BSA spike-in and 15k BSA spike-in. All BSA spike-in samples were conducted in a 10-plex.

We hypothesised that introducing larger defined quantities of a single protein (i.e., a "spike-in") could reduce the effect of random sample losses on the consistency of protein quantification by saturating the surfaces and trapping the cell protein content within a bigger protein pellet after TCA precipitation. Such spike-ins can be comprised of a single purified protein, such as bovine serum albumin (BSA), or a cell lysate from
a different species. Here, an E. coli whole proteome extract and a BSA spike-in were utilised as a spike-in source. A direct comparison of both spike-in sources showed that BSA provided a superior identification rate (data not shown). Subsequently, a 10-plex of 5,000 Hoxb8-FL cell replicates was performed using a BSA spike-in. When inspecting the distribution of \%CV values, we observed a significant reduction in variability with the apex of the distribution situated at $\sim 10 \%$ CV (Figure 4B). Increasing the replicate cell numbers from 5,000 to 10,000 Hoxb8-FL cells further reduced variability by narrowing the distribution (Figure 4C). Interestingly, a further increase to 15,000 Hoxb8-FL cells did not significantly reduce \%CV (Figure 4D). These results further suggested that reducing required cell input from $10^{6}$ to $10^{4}$ did not negatively affect intra-experimental variability and thus provided a viable target range for small cell number proteomics.

When comparing protein identification and quantification between the above three experiments, we observed a 3-fold increase with the introduction of the BSA spike-in (Figure 4 E ). Notably, doubling the cell number to 10,000 for each replicate had the greatest effect, surpassing 1,000 identified proteins, while a further increase to 15,000 cells per replicate further increased the number of identified proteins to over 1,500 (Figure 4E). Interestingly, the number of proteins reliably quantified across all replicates did not increase proportionally to the number of identified proteins (Figure 4E). From this analysis, we conclude that bulking up a sample with greater total protein via the BSA spike-in during sample preparation significantly improved data reproducibility and protein recovery.

### 3.5. Introduction of carrier populations to enhance protein detection

While the above efforts to reduce variability achieved a 3-fold increase in protein identification, the total number remained below our set target of $>3000$ proteins and I therefore explored additional optimisation measures. Budnik et al., a study focused on developing single cell proteomics, had reported the use of a large carrier population to boost the number of identified proteins by providing additional proteins from a related cell type ${ }^{309}$. Taking a similar approach, we adapted our protocol to evaluate the impact of a carrier population on the rate of protein identification (Figure 5A). The
protocol developed by Budnik et al., utilised a 200-fold higher carrier population - in their case 200 cells for deriving the proteome of a single cells ${ }^{309}$. However, using a large dynamic range between multiplexed TMT channels might compromise quantification accuracy due to the isolation interference problem of isobaric labelling ${ }^{365}$. Hence, a 2-fold difference in abundance between sample and carrier was selected (Figure 5A). Three independent 10-plex carrier runs were performed, two of which were conducted with 10,000 cells (sample triplicate $+7 \times 20,000$ cells for the carrier channels) and one run with 15,000 cells (sample triplicate, carrier: $7 \times 30,000$ cells). The digestion strategy was also amended to have one 10 k carrier run subjected to sequential digestion with trypsin and gluC. The remaining two 10-plex experiments (10k and 15 k ) were digested using trypsin only. When comparing the 10k runs, the carrier population approach (trypsin only) yielded a 2-fold increase in identified proteins (Figure 5B). Surprisingly, sequential trypsin-gluC digest reduced the identification rate (Figure 5B). Increasing cell numbers to 15,000 showed marginal improvement (Figure 5B).

Together, these results indicated that the introduction of carrier populations, but not gluC-mediated digestion, improved the yield of identified and quantified proteins by approximately 2 -fold.


Figure 3.5. Carrier populations within a 10-plex experiment increase the number of identified proteins. (A) Detailed schematic outlining the sample preparation protocol and subsequent analysis by LC-MS/MS. 7 lanes with $2 x$ carrier populations were used. TMT labelling was performed at protein level, prior to digestion. Digests were performed with trypsin and gluC or trypsin only. (B) Overview of the total number of proteins identified (red) and quantified (blue) during each multiplex run, as indicated in the bar labels. $5 k$ Cell Run (old), $5 k$ Cell Run and 10k Cell Run did not contain a carrier population (consistent cell number across the multiplex). The cell numbers indicated for carrier runs refer to the sample cell number. The respective carrier population cell numbers are double of the sample cell number.

Next, we inspected the stochasticity of our approach, as the consistency of protein identifications across several multiplexes is critical for the efficacy of any proteomics protocol. The number of proteins was consistent across multiple experiments and the protein groups identified in the 10k carrier experiment and the 15k carrier experiment (both trypsin only) were highly overlapping (71\%, Figure 6A). The variation of quantification in all three runs was also examined to determine the effect of a carrier population on sample consistency. The distribution of \%CV values for 10 k and 15 k carrier samples (trypsin digests) mirrored previous observations, made for the 10 k and

15k BSA spike-in samples (Figure 6C/D). In line with the reduced levels of protein identification and quantification, the 10k multiplex sequentially digested with trypsin and gluC displayed a broader \%CV distribution (Figure 6B). These results further support the previous observation that trypsin+gluC digestion was subject to reduced efficiency.


Figure 3.6. The carrier approach permits reduction of the required cell numbers per 10-plex experiment. (A) Venn diagram to show the overlap of protein accession ID numbers for the 15k carrier run (labelled 15k) and the 10k carrier run (labelled 10k). (B) Variability coefficient (CV) histogram for the carrier population multiplex using 10k Hoxb8-FL for the sample triplicate and 20k Hoxb8-FL cells for the remaining 7 carrier channels, digested using trypsin and gluC. histogram displays \%CV frequency within bins of $1 \%$. (C) \%CV of 10-plex using 3x10k and 7x20k Hoxb8-FL cells. (D) \%CV of 10-plex using $3 \times 15 k$ and $7 \times 30 k$ Hoxb8-FL cells.

### 3.6. Offline fractionation maximised protein quantification and reproducibility

The protocol improvements outlined above demonstrate that proteomic screening of as few as 10 k cells per replicate can provide an adequate \%CV distribution. However, $\sim 30 \%$ of proteins in the 10 k carrier run were not identified in a subsequent 15 k carrier run (Figure 6 A ) and only $\sim 50 \%$ of identified proteins were reliably quantifiable across all replicates within the complete multiplex experiment (Figure 5B). These differences, and the relatively low rate of protein quantification, could be attributed to protein level TMT labelling. The initial protocol utilised protein level TMT labelling to minimise loss of lysate due to sequential sample transfers. In the process of optimising sample handling (Figure 5A), sample transfers could be avoided which negated the need for protein level TMT labelling and TCA precipitation. Omission of these steps and movement toward a single-tube protocol might therefore result in an increased proportion of quantified proteins.


Figure 3.7. Offline fractionation prior to LC-MS/MS boosts protein identification and reduces data variability. (A) Overview of the total number of proteins identified (red) and quantified (blue) during each multiplex run, as indicated in the bar labels. Includes previous carrier approaches and subsequent 10-plex of 3x10k and 7x20k Hoxb8-FL that underwent offline fractionation. (B) Venn diagram showing the overlap of protein accession ID numbers for the direct injection 15k carrier run (labelled Direct 15k), direct injection 10k carrier run (labelled Direct 10k) and the fractionated 10k carrier run (labelled Fract). (C) Venn diagram showing the overlap of protein IDs between both fractionation experiments. (D-E) \%CVs histogram for the first fractionated 10k carrier population run (D) and the second (E).

In parallel to the introduction of TMT labelling post-digestion to increase quantification efficiency, we set out to further improve the total protein yield. For this purpose, we utilised sample pre-fractionation prior to LC-MS/MS analysis using a single-tube protocol ${ }^{366,367}$. Exploiting the sensitivity of the Orbitrap Lumos mass spectrometer, the complexity of the peptide mixture was reduced by performing high pH reverse-phase peptide fractionation. We performed offline fractionation on a 10-plex of $3 \times 10 \mathrm{k}$ and $7 x 20 \mathrm{k}$ Hoxb8-FL cells. 6 fractions were collected, and each fraction was injected separately into the LC-MS/MS system. When comparing the resulting protein identification and quantification rate with the two best carrier runs, a modest increase in total identified proteins was observed but this was accompanied by a significant increase in quantification efficacy, with nearly 3,000 proteins quantified across all 10 replicates (Figure 7A). Next, we compared the overlap of proteins identified in samples that were fractionated and samples that were not fractionated (10k carrier experiment, 15k carrier experiment and 10k fractionation experiment). As expected, the substantially higher quantification rate of the fractionated experiment resulted in 1,199 unique proteins including higher coverage of the fractionated sample (Figure 7B). Next, we repeated the peptide fractionation using the same sample setup but increased the number of collected fractions. An increase in fraction collection was accompanied by an increase in total protein identification and the number of quantified proteins (Figure 7A). Most importantly, the second fractionation experiment was able to detect $87 \%$ of proteins, previously identified in the first fractionation experiment (Figure 7C). Furthermore, quantified proteins of the first fractionation run showed similar CV distribution to previous directly injected carrier runs (Figure 7D). Most intriguingly, the second fractionation experiment reduced the \%CV distribution even further, resulting in the successful development of a robust protocol with highly reproducible quantification across multiplexed replicates (Figure 7E).


Figure 3.8. Optimised sample handling reduces sampling errors across biological samples. (A) Total protein abundances per sample with the 10-plex 10x10k Hoxb8-FL with a BSA spike-in. (B) Total protein abundances for the 10-plex 10k carrier run with protein-level TMT labelling ( $3 \times 10 \mathrm{k}$ and $7 \times 20 \mathrm{k}$ Hoxb8-FL cells). (C) Total protein abundances for the 10-plex fractionation run ( $3 \times 10 \mathrm{k}$ and $7 \times 20 \mathrm{k}$ Hoxb8-FL cells).

To further illustrate the striking difference, we plotted the total protein abundance for the 10k BSA spike-in multiplex (Figure 8A), the 10k carrier approach (Figure 8B) and the 10k fractionation (II) multiplex (Figure 8C). The fractionation run was subject to the most consistent protein abundance across replicates and displayed a 5-fold and 2.5-
fold higher protein abundance when compared to previous runs (Figure 8). Overall, these data confirm that the finalised protocol provides reliable and reproducible quantification of approximately 3,500 proteins from a 10-plex with as little as 10,000 cells. The final protocol utilised a carrier population approach with a 2 -fold dynamic range, cell sorting and sample processing in a single tube, peptide TMT labelling posttrypsin digest and pH reverse-phase peptide fractionation prior to LC-MS/MS analysis (Figure 9).


Figure 3.9. Workflow of the final small cell number LC-MS/MS protocol. Diagram of the final small cell number proteomics protocol. 10 samples of as little as 10,000 cells are first lysed, reduced, alkylated and subject to trypsin digestion. The resulting peptides are labelled using unique TMT labels and pooled for offline fractionation. Acquired fractions are analysed by LC-MS/MS for protein identification and quantification.

### 3.7. Peptide fractionation improved identification of DNA-binding proteins

To interrogate the depth of the proteome covered, we examined the number of DNAbinding proteins to include transcription factors and epigenetic regulators. For this purpose, the consistency of identification between multiplexed samples was examined and a protein-protein interaction network for DNA-binding proteins was computed for the two 10k carrier runs using first order interactions - one using direct injection and the other using fractionation (Figure 10).


Figure 3.10. Protein Interaction network of DNA-binding proteins reveals consistent protein identification across samples. Protein network of DNA binding proteins, identified in the 10k carrier runs with direct injection and fractionation. Node colours indicate proteins shared between both runs (green), unique for the direct injection sample (blue) and proteins unique for the fractionated sample (red). Edges indicate direct protein-protein interaction.

A wide range of proteins were observed, including components of the DNA replication machinery (Mcm proteins), chromatin remodelling factors (NurD complex components, SWI/SNF complex) and haematopoiesis-specific transcription factors, such as Erg (Figure 10). As expected, the significantly higher protein yield from the fractionated run was accompanied by a large proportion of exclusive DNA-binding proteins. These included further components of basic functions described above, but also more general transcription factors such as STAT3, STAT5, ABT1 and ETV6. On the other hand, only a handful of proteins were exclusively identified in the directly injected sample. Overall, these data provide a powerful illustration of the increased depth of the final offline fraction-based protocol which allows reliable mapping of entire protein complexes, thus enabling new insight into key pathways.

### 3.8. Proteomics identifies ECM proteins and metabolic regulators, previously not detected by scRNA-seq

Next, the utility of the small cell number proteomics protocol was assessed by determining the depth of novel proteome-derived insights unable to be detected using transcriptomics. Data were compared to a single cell RNA sequencing (scRNA-seq) dataset for Hoxb8-FL cells generated by the Bertie Göttgens lab (unpublished) where raw scRNA-seq counts and protein abundances were normalised and log transformed. To identify proteins uniquely identified by LC-MS/MS, a pairwise plot was computed for all genes that appeared in both the proteomic and transcriptomic datasets (Figure 11A). Intriguingly, a subset of 33 proteins was identified, which was not detected by RNA sequencing (Figure 11B). Interestingly, these included multiple extracellular matrix (ECM) proteins, such as PLG, COL2A1 and COL3A1 ${ }^{368-370}$.

A


B

| Proteins uniquely identified in the Proteome |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- |
| Hist2h2a <br> a1 | Itih2 | Alb | Itih3 | F5 |
| Serpinc1 | Col1a1 | Thbs4 | Col3a1 | Lum |
| Fscn1 | Efemp1 | Apoa1 | A2m | Hgfac |
| Igfbp2 | Igf2r | Plg | Fgb | Mia2 |
| Fmod | Apom | Col6a1 | Mst1 | Rgn |
| Pzp | Dsg1b | Aoc3 | Amy1 | Col2a1 |
| Pclo | Kif27 | Pdcd2 |  |  |

Figure 3.11. The proteome of Hoxb8-FL cells correlates with the transcriptome, but permits identification of unique proteins, not detected by scRNA-seq. (A) Plot depicting mean normalised read counts and mean normalised protein abundance of Hoxb8-FL cells. Genes not identified in the proteomic screen were excluded. Green line represents theoretical uniform correlation ( $r=1$ ) and proteins not identified in the transcriptome are marked in orange. (B) Table of proteins uniquely identified in the proteomic screen.

To identify the set of genes selectively enriched in the proteome, k-means clustering was performed and the resulting clusters were visualised on a pairwise RNA-Protein
plot (Figure 12A). Next, gene ontology (GO) term enrichment was performed ${ }^{371}$ for biological processes for the subset of proteome enriched genes (Figure 12B). In line with the above observation of ECM protein enrichment, genes involved in the regulation of the extracellular structure were enriched in the proteome (Figure 12B). In addition, metabolic proteins were also enriched (Figure 12B). A subsequent Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis ${ }^{372}$ corroborated the GO term enrichment analysis (Figure 12C) with observed enrichment of cholesterol metabolism, ECM-receptor interactions and focal adhesion (Figure 12C). In addition, enrichment of PI3K-Akt signalling and Relaxin signalling pathways were also observed (Figure 12A).


Figure 3.12. Proteins overrepresented in the proteome are enriched for extracellular matrix proteins and metabolic regulators. (A) Proteome and transcriptome plot, as outlined in Figure 11A. K-means clusters were computed and visually depicted. (B) Gene Ontology (GO) term enrichment analysis for $k$ means cluster 1 (green). Top 20 biological process terms are displayed. (C) Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway enrichment for k means cluster 1 (green). Manually curated set of KEGG pathways with p-value $<0.05$.

Overall, these data indicate that the proteomic analysis revealed a subset of novel molecules not detected by scRNA-seq and identified different categories of molecules compared to transcriptomic data. Moreover, since the vast majority of transcriptomic screens rely on capture of polyA+ RNA, they are subject to an intrinsic sampling bias ${ }^{373}$
which is avoided in proteomic studies. Hence, the complementary nature of several omic strategies, such as gene expression and proteomic profiling, provide a promising avenue to broaden molecular characterisation and better define true molecular states. In addition, the enrichment of ECM-related and membrane proteins was particularly encouraging since the relative insolubility and low expression of such proteins traditionally results in underrepresentation of such membrane proteins in LC-MS/MS experiments ${ }^{374}$.

### 3.9. The proteome provides a broader dynamic range than the transcriptome

Next, we explored the distribution of the correlation between identified proteins and their corresponding transcriptomes. For this purpose, we computed pairwise Pearson correlations for all proteins identified by LC-MS/MS and all genes identified by scRNAseq (Figure 13A). The density plot of correlation coefficients presented a narrow distribution for scRNA-seq compared to a wide distribution of correlations for identified proteins (Figure 13A). These data indicate that the increased dynamic range of the protein is substantially wider and can therefore capture a broader range of differences in molecular states, providing deeper insight into the underlying mechanism of biological phenomena.

Transcriptomic and proteomic data are inherently derived using different quantification and detection methods and therefore pairwise comparison plots are commonly derived using differential expression between two conditions to bring both omics datasets onto the same dimension ${ }^{373}$. Unfortunately, such a comparison was not possible for this dataset. However, independent ranking, based on the abundance within each ome, can provide a valuable insight. Hence, all overlapping proteins/genes were plotted according to their expression rank. In brief, we first ranked all proteins based on total abundance and all genes by mean normalised expression. As previously outlined, we observed moderate correlation between protein and gene ranks (Figure 13B). Subsequently, we isolated the subset of proteins that ranked amongst the top 500 most abundant proteins, while ranking amongst the least expressed in the transcriptome (red dots in Figure 13B). The resulting set of 78 proteins was used to compute GO term enrichment and KEGG pathway analysis. Interestingly, strong
enrichment of several metabolic biological processes suggested strong underrepresentation of cellular metabolism in the transcriptome (Figure 13C). We also observed striking enrichment of ribosomal proteins (Figure 13D). It has previously been reported that ribosomal gene transcripts are subject to disproportionally extensive translation, leading to a disconnect between the number of proteins and transcripts ${ }^{375}$. Hence, the enrichment of ribosomal proteins in our analysis is unsurprising. However, these data also suggest that the proteome provided a more robust representation of the ribosomal machinery, complementing transcriptomic data. In addition, the transcriptomic analysis relied upon polyA tail pulldowns, thus preventing inspection of non-poly-adenylated transcripts, of which histones are a prominent example ${ }^{376,377}$. Instead, several studied demonstrated efficacy of ribosomal RNA depletion in absence of polyA capture to circumvent such biases and provide a likely more appropriate counterpart to the proteomic data ${ }^{378,379}$.


Figure 3.13. Dynamic range of LC-MS/MS and RNA-seq data. (A) Density plot of pairwise Pearson correlation coefficients (r) between all genes/proteins, identified in scRNA-seq (blue) and LC-MS/MS (red) datasets. (B) Rank plot, outlining all genes/proteins identified in scRNA-seq (Gene Rank) and LCMS/MS (Protein Rank). Within each dataset, genes were ranked according to abundance in descending order. Genes enriched in the proteome are marked in red (Protein rank $\leq 500$ and Gene Rank $\geq 1500$ ). A linear model was used to determine the correlation between scRNA-seq and LC-MS/MS. (C) GO term enrichment analysis for proteins overrepresented in the proteome (marked red in Figure 13B). (D) KEGG pathway enrichment analysis for proteins overrepresented in the proteome (marked red in Figure 13B).

### 3.10. The LC-MS/MS protocol, optimised for small cell number experiments, reliably identifies key marker genes of Hoxb8-FL cells

Finally, the proteomic screen was tested for its ability to validate key driver genes of Hoxb8-FL cells identified by scRNA-seq in Kucinski et al., where Hoxb8-FL cells were profiled to determine the interaction between lineage-specific transcription factors ${ }^{357}$. A previously published scRNA-seq dataset of primary mouse LK/LSK cells was used and the cell type annotations and Leiden clusters were plotted to create a reference gene expression landscape of stem and progenitor cells ${ }^{343}$ (Figure 14A/B). In accordance with Kucinski et al., the scRNA-seq profiles of Hoxb8-FL were enriched in the multipotent progenitor cluster between lymphoid and monocytes/neutrophils, monocyte and neutrophil clusters on the reference landscape ${ }^{357}$ (Figure 14C). These clusters corresponded to Leiden clusters 4,8 and 15, with the majority of cells falling within cluster 4 (Figure 14B/C).


Figure 3.14. Hoxb8-FL cells are primed for myeloid and lymphoid differentiation. (A) UMAP representation of a previously published scRNA-seq landscape of mouse LK/LSK cells ${ }^{343}$. Cell types are annotation, based on classification outlined by Kucinski et al. ${ }^{357}$ (B) UMAP representation of LK/LSK scRNA-seq landscape, displaying Leiden clusters (adapted from Kucinski et al.) ${ }^{357}$. (C) Projection of previously published scRNA-seq profiles of Hoxb8-FL cells onto the reference LK/LSK landscape. Adapted from Kucinski et al. ${ }^{357}$

In order to assess the robustness of the 10-plex proteomic screen using 10,000 Hoxb8-FL cells differential gene expression (DGE) analysis was performed and the top 25 marker genes were plotted for the three clusters, enriched for Hoxb8-FL cells (Leiden clusters 4, 8, 15) (Figure 15A). Of these 75 genes, 26 were also identified in the proteomic screen (Figure 15B). We ranked all 26 proteins according to their mean protein abundance across the 10-plex experiment, derived using the 10,000 cell fractions, the carrier population and offline fractionation (Figure 15B). Interestingly, we observed a dominant identification of genes enriched in cluster 4, closely followed by cluster 15 markers (Figure 15B). In contrast, few cluster 8 markers were identified and their protein abundances were significantly lower than other marker genes (Figure 15B). These data align with the scRNA-seq projections of Hoxb8-FL, showing high identification frequency of cluster 4 marker genes but only few cluster 8-specific genes (Figure 15B). A knock-out screen of 39 haematopoiesis-specific transcriptional factors (TFs) revealed a set of 23 TFs with mechanistic roles in Hoxb8-FL cells ${ }^{357}$ and these data show that 9 out of 23 (39\%) could be reliably quantified by both RNA-seq and LC-MS/MS (Figure 15C). Overall, these data provide further support of reproducible protein identification and sufficient depth of proteome coverage to partially corroborate with scRNA-seq data. In addition, the wider dynamic range of LC-MS/MS can provide a novel insight into HSC biology, otherwise missed by RNA-seq. Such was particularly evident for basic cellular machineries, such as the ribosome.


Figure 3.15. The small cell number LC-MS/MS protocol reliably identifies key marker genes of Hoxb8-FL cells. (A) Top 25 genes enriched in the indicated Leiden cluster, based of differential gene expression. Genes are ordered by rank according to the enrichment score. (B) Mean protein abundances for all cluster marker genes that were identified in the proteomic screen (10-plex: 10k Hoxb8-FL cells, carrier population strategy, offline fractionation). Colour indicates the corresponding Leiden cluster. (C) Protein-protein interaction network of 25 transcription factors, identified by Kucinski et al. as key regulators of Hoxb8-FL cells ${ }^{357}$. Proteins co-identified by RNA-seq and LC-MS/MS are marked in blue and transcriptional factors not identified in the proteome (RNA-seq only) are marked in orange.

### 3.11. Discussion

The optimised LC-MS/MS protocol presented in this chapter was specifically adapted for screening as few as 10,000 cells per sample within a multiplex of 10 samples. We abolished any form of protein lysate transfer, adapted all sample preparation steps into a single-microtube process, introduced 2-fold carrier population and extensive offline peptide fractionation to achieve a 100-fold input reduction, when compared to traditional gold-standard TMT multiplex protocol ${ }^{380}$. Despite this 100 -fold reduction in
sample input, we were able to identify 4,372 unique proteins, of which 3,663 were quantifiable. We subsequently demonstrated that the protocol yields highly reproducible protein identification between experiments and consistent quantification within the multiplex, thus enabling direct quantitative comparisons between biologically relevant samples. We also demonstrated the utility of the small cell number LC-MS/MS protocol in identifying key marker genes, transcriptional factors and molecular machineries often underrepresented in transcriptomic screens.

A notable example included the underrepresentation of the ribosomal machinery in the transcriptome. Regulation of ribosomal activity has previously been implicated in playing a crucial role in haematopoietic differentiation and lineage commitment ${ }^{381,382}$. Our small cell number LC-MS/MS protocol provided extensive coverage of ribosomal proteins, thus highlighting its utility for better dissecting the role of basic cellular machineries, such as ribosomes, splicing and metabolism, in haematopoiesis. Thus, complementing existing transcriptomic screening tools.

Since development of this method, the scope of TMT labelling has been greatly expanded. The addition of additional isobaric labels now provides the choice between 11 -plex, 16-plex and more recently even 27-plex proteomic screens ${ }^{312}$. The access to expanded multiplex tools will further reduce the required cell input per sample. Furthermore, the haematopoietic stem cell size of approximately $4.6 \mu m \pm 0.2 \mu \mathrm{~m}$ falls below the size of other prominent cell types, such as embryonic stem cells (between $9.6 \mu \mathrm{~m} \pm 3.5 \mu \mathrm{~m}$ and $30.1 \mu \mathrm{~m} \pm 1.5 \mu \mathrm{~m})^{383,384}$. The significant increase in cell size is accompanied by a proportional increase in total protein abundance per cell. Hence, it is conceivable that the required cell numbers are cell type specific and can be further reduced in other biological systems.

As previously outlined, recent efforts to reduce required cell for LC-MS/MS experiments has provided the first single-cell proteomic platforms ${ }^{310,318}$. Here, the primary trade-offs between single cell proteomics and the small cell number protocol presented in this chapter are sequencing depth and dynamic range (technological details are extensively discussed in introduction chapter 1.6), making the choice of technique dependent upon the biological question to be addressed. For example, systems with substantial molecular differences or perturbations can leverage single
cell proteomics, as elegantly demonstrated by Schoof et al. ${ }^{317}$. Here, the authors compared the proteomes of single leukemic stem cells, progenitors and blasts by acquiring hundreds and up to 1,000 proteins per cell. Due to significant differences between these cell types, the authors were able to capture differential protein expression amongst the up to 1,000 most abundant proteins. In contrast, singular phenotypic outcomes, such as cellular decisions such as self-renewal of HSCs, are governed by intricate molecular mechanisms. Deciphering such mechanisms often relies on the comparison of molecularly related cell populations, thus requiring deeper sequencing beyond to 1,000 most abundant proteins to identify key drivers. The small cell number LC-MS/MS protocol presented here provides an alternative, complementary tool for screening rare cell populations with increased depth, moving beyond the transcriptome for insight into novel biological mechanisms.

## Chapter 4. Simultaneous proteomic and transcriptomic profiling to characterise molecular networks governing HSC self-renewal

Following extensive expansion during early development, multipotent HSCs enter a predominantly quiescent state and undergo steady-state asymmetric self-renewal at the population level to maintain homeostatic numbers of HSCs ${ }^{385,386}$. In turn, bone marrow injury induces rapid, but reversible, proliferation of HSCs to produce progenitors that restore stable haematopoiesis ${ }^{387,388}$. Similarly, deregulation of HSC self-renewal has been widely implicated in tumorigenicity and tumour initiation ${ }^{389}$. Unsurprisingly, the importance of self-renewal in tissue homeostasis and disease has raised great interest in identifying key molecular drivers of HSC self-renewal. A plethora of molecular and functional studies have revealed numerous intrinsic and extrinsic molecular networks, driving HSC self-renewal ${ }^{112}$. However, the rare nature of HSCs and their functional heterogeneity have largely limited molecular profiling to transcriptomic screening.

More recently, several studies utilised novel multi-omic screening platforms to characterise the molecular pathways governing HSC function at single-cell resolution including transcriptomics, epigenetics, and Cellular indexing transcriptome and epitome by sequencing (CITE-seq) ${ }^{390,391}$. While these studies provide unprecedented insight into molecular profiles of HSC, the proteome, arguably a critical component of the molecular paradigm, remains largely uncharacterised. Pioneering studies by Cabezas-Wallscheid et al., Jassinskaja et al. and others provided the first insights into the underlying protein networks operative in HSCs ${ }^{326,359}$. However, technical limitations at the time coupled with the rare nature of HSCs limited the scope of possible comparisons due to the large amount of tissue required to obtain such numbers.

The optimised LC-MS/MS protocol for screening as few as 10,000 cells per sample, presented in Chapter 3, vastly increases the scope of addressable biological questions, allowing more comprehensive studies including the dissection of stem cell heterogeneity. To identify novel drivers of HSC self-renewal, we compared closely
related cell populations that only differed in their self-renewal durability. This included separating the stem cell compartment into long-term HSCs (LT-HSCs) with durable self-renewal potency and LT-HSCs which exhibit finite self-renewal potency and limited ability for bone marrow reconstitution ${ }^{392}$ (Figure 1).


Figure 4.1. Haematopoietic cell types and their self-renewal potency. Graphical depiction the cell populations, utilised for multi-omic screening. Long-term haematopoietic stem cells (LT-HSCs) with durable self-renewal (SR) or LT-HSCs with finite SR were isolated from wild-type (WT) and Tet2 knockout (TET2-1) animals. The self-renewal potency (blue) and stem cell frequency (orange) are indicated.

In combination with this approach, we also took advantage of a mouse model with enhanced self-renewal as a comparator. Loss-of-function mutations of the ten-eleven translocation 2 (TET2) gene, an epigenetic regulator catalysing DNA demethylation, have been shown to enhance HSC self-renewal ${ }^{180,393}$ and Ko et al. developed a mouse model harbouring loss-of-function mutant TET2 by knocking out the catalytic domain (TET2KO or TET2 ${ }^{-/}$hereafter) ${ }^{245}$. The authors and others subsequently reported increased frequency of TET2-/ LT-HSCs and competitive advantage over wild type LT-HSCs, due to an increased rate of self-renewal, proliferation and multilineage potency ${ }^{245,339}$. Consequently, we included $T E T 2^{-/}$LT-HSCs to generate a dataset with a continuum of self-renewal potency (Figure 1).

Overall, multiplexing durable and finite LT-HSCs from wild-type (WT) and TET2-/ animals provides a dataset of closely related HSC populations, only distinguishable by their self-renewal potency (Figure 1). We set out to identify key protein networks governing HSC self-renewal using such experimental strategy. We utilised our new

LC-MS/MS screen, optimised for rare cell populations, and integrated these with transcriptomic profiling to generate a multi-omic landscape of HSC self-renewal.

### 4.1. Experimental setup for multi-omic screening to decipher HSC self-renewal

To establish a comprehensive molecular landscape of HSC self-renewal, we profiled the proteome and transcriptome of WT and TET2-/ HSCs. Long-term HSCs (LT-HSCs) with durable self-renewal were defined as $\mathrm{Lin}^{-} \mathrm{cKit}^{+}$CD45 ${ }^{+}$CD48-CD150+ ${ }^{+}$(CD150+ hereafter) and LT-HSCs with finite self-renewal were defined as LincKit ${ }^{+}$CD45 ${ }^{+}$CD48-CD150- (CD150- hereafter). To complement the dataset, we also isolated an LT-HSC population of greater purity, defined as CD45+CD48-EPCR ${ }^{+}$CD150 ${ }^{+}$Sca1 ${ }^{+}$ (abbreviated as ESLAM hereafter). Using the LC-MS/MS protocol for rare cell populations, we performed an 11-plex, comprising the following cell populations: 1) TET2 ${ }^{-/}$CD150+; 2) TET2 ${ }^{-/}$CD150-; 3) WT ESLAM; 4) WT CD150+; 5) WT CD150-; 6) WT LSK (Lin-Sca1+cKit ${ }^{+}$) (Figure 2A). Each cell population was comprised of two biological replicates and contained between 10,000 and 30,000 freshly isolated cells. To acquire a matching bulk transcriptome, we freshly isolated 1,000 cells in biological triplicates for CD150+ and CD150- cell fractions of WT and TET2-/ animals (Figure 2B). While bulk RNA-seq provides a crucial element for a mult-iomic, analysis including LC-MS/MS, we also performed single cell RNA-seq (scRNA-seq) to determine whether identified molecular profiles can be recapitulated at the single cell level. Hence, we performed scRNA-seq using the SmartSeq2 protocol for WT and TET2-/ ESLAM LT-HSCs.


Figure 4.2. Sample acquisition strategy for LC-MS/MS and bulk RNA-seq. (A) Sample acquisition for LC-MS/MS. The number of WT and TET2 ${ }^{-/}$animals pooled for each sample and the number of biological replicates (arrows) are indicated. Fluorescence activated cell sorting (FACS) was used to acquire CD150 ${ }^{+}$(Lin-cKit ${ }^{+}$CD45 ${ }^{+}$CD48-CD150 ${ }^{+}$), CD150 ( Lin $^{-}$cKit ${ }^{+}$CD45 ${ }^{+}$CD48-CD150), ESLAM $\left(C D 45^{+} C D 48 E P C R^{+} C D 150^{+}\right.$Sca1 ${ }^{+}$) and LSK (Lin-Sca1+ ${ }^{+}$KKit ${ }^{+}$) cell populations. (B) Experimental setup for bulk RNA-seq. CD150 ${ }^{+}$and CD150 cell populations of WT and TET2-/ were collected in triplicates. Bone marrow of one animal was harvested for each biological replicate and matching CD150 / CD150 cell populations were isolated.

### 4.2. Small cell number LC-MS/MS provides reproducible proteomic profiles of HSC populations

Prior to performing downstream pathway analyses and integration with RNA-seq data, the quality and reproducibility of the 11-plex LC-MS/MS dataset was inspected.

Overall, we were able to identify 4133 protein groups, of which 3989 were reliably quantified across all cell fractions. Next, the number of unique peptides and peptide spectral matches (PSMs) was determined and, unsurprisingly, most proteins were identified using a single unique peptide (Figure 3A). In addition, a large proportion of identifications were also based on few PSMs (Figure 3B). Hence, a false discovery rate (FDR) cut-off of $5 \%$ was assigned to ensure reliability of protein identifications (data not shown).

Next, the depth and consistency of the proteome captured by rare cell LC-MS/MS was compared to one of the most comprehensive HSC proteomes available to date ${ }^{359}$. We were encouraged by a significant overlap in identified proteins between both datasets, with $88 \%$ of proteins quantified by rare cell LC-MS/MS also identified by CabezasWallscheid et al. (Figure 3C). Simultaneously, the overlap between both datasets represents $56 \%$ of proteins identified by Cabezas-Wallscheid et al. despite our protocol using 93\% fewer cells as the initial protocol (Figure 3C).


Figure 4.3. Small cell number proteomics captures most of the previously described HSC proteome. (A) Histogram depicting the frequency of unique peptides per protein group detected by LCMS/MS. Each bin comprises 1 unique peptide. (B) Histogram of the frequency of peptide spectrum matches (PSMs) per protein group. Each bin contains 1 PSM. (C) Venn diagram of identified, shared between the first proteome ${ }^{359}$ and the small cell number multiplex.

These data suggested robust identification of proteins by LC-MS/MS, but the reproducibility of data across biological replicates is another key requirement that needed to be assessed. A one-way ANOVA was performed and a false discover rate (FDR) cut-off < $5 \%$ was set. Subsequently, hierarchical clustering was performed to determine the stochasticity of LC-MS/MS-based protein sampling. Biological replicates clustered closely together, and the genotype provided the greatest separating variable (Figure 4). Overall, these data indicate that rare cell LC-MS/MS reproducibly captures the proteome at sufficient depth to infer biological function.


Figure 4.4. LC-MS/MS and RNA-seq profiles are reproducible across biological replicates. (A) Hierarchical clustering plot of all LC-MS/MS samples. The distance between samples is indicated by height.

### 4.3. The CD150- LT-HSC compartment holds most molecular differences between WT and TET2 ${ }^{-/}$animals

Having established the reliability of the LC-MS/MS data, we focused on the comparison of the bulk proteome and bulk transcriptome. First, unsupervised dimensionality reduction was performed for each dataset separately. For this purpose, we utilised principal component analysis (PCA), outlining distinct differences between the proteome and transcriptome (Figure 5A/B). The PCA revealed that sample clusters were largely distinguished by genotype, with $T E T 2^{-/}$samples clustering away from WT samples (Figure 5A). While TET2 $^{-/}$CD150 ${ }^{+}$and TET2 ${ }^{-/}$CD150 cell fractions also exhibited some variation along principal component 2 (PC2), these did not exceed the separation by genotype (Figure 5A/C). These data are in accordance with hierarchical clustering (Figure 4). In contrast, the transcriptome revealed proximity between CD150 ${ }^{+}$cell fraction of WT and TET2-/ genotypes, while CD150- fractions clustered at considerable distance (Figure 5B/D). Collectively, these data indicate that the
transcriptome and proteome likely captured differing molecular states, setting the stage for novel biological insight.


Figure 4.5. Proteomic and transcriptomic profiling reveals distinct molecular states between WT and TET2-/ HSCs. (A) Principal component analysis (PCA) of all LC-MS/MS samples. Plot depicts separation by principal component 1 (PC1) and PC2. (B) PCA analysis of bulk RNA-seq samples. As above, PC1 and PC2 are depicted. (C) Bar plot representing the percentage variance explained by each PC (PC1-10) for LC-MS/MS profiles. (D) Bar plot representing the percentage variance explained by each PC (PC1-10) for bulk RNA-seq profiles. Sums of all previous PCs in descending order are depicted using a line plot (red line).

In order to identify genes with corresponding protein and transcriptional profiles genes identified in both datasets were extracted. Using normalised protein and transcript values for extracted genes, values were scaled based on the average expression across all samples (individually for each -ome). An ANOVA was performed ( $p$-value < 0.05 ) to identify genes with significantly overlapping profiles across the proteome and transcriptome (Figure 6). The scaled and log-transformed profiles were displayed in form of a heatmap (Figure 6). Intriguingly, we observed that genes with consistent enrichment across both -omes were predominantly differentially regulated between WT and TET2-/ CD150- fractions (Figure 6). In contrast, few differentially expressed
genes distinguished LT-HSCs of both phenotypes. These data suggest that the loss of TET2 function has a more profound impact on the molecular profile, and consequently the function, of LT-HSCs with finite self-renewal durability. Alternatively, such a result might indicate discrepancy between the transcriptome and proteome or be derived from different levels of intrinsic heterogeneity between LT-HSC populations with durable or finite self-renewal potency.


Figure 4.6. TET2 knockout predominantly alters protein and mRNA profiles of finite HSCs. Heatmap of scaled LC-MS/MS and RNA-seq profiles for genes, detected in both screens. Scaled values are depicted using Z-scoring. An ANOVA was computed to identify differentially expressed genes. Only genes with significant differences ( $p$-value $<0.05$ ) were depicted and subject to hierarchical clustering.

### 4.4. Poor correlation between the proteome and transcriptome of LT-HSCs

To identify drivers of HSC self-renewal drivers, a direct comparison of WT and TET2${ }^{\text {- }}$ CD150 ${ }^{+}$HSCs was undertaken. First, differential protein expression (DPE) and differential gene expression (DGE) were computed separately. In order to facilitate reliable scoring of differentially expressed proteins using two replicates, pairwise DPE was undertaken between each replicate individually and One Sample T-Tests were
subsequently computed (see methods). To ensure consistency, DE for both the proteomic and transcriptomic data was classified as fold change $\left(\log _{2} \mathrm{FC}\right) \geq 0.5$ with an associated $p$-value $\leq 0.05$ (Figure 7A/B). When comparing the set of differentially expressed genes (DEGs), enriched in CD150+ TET2 ${ }^{-/}$HSCs across both molecular dimensions, we observed a strikingly low overlap (Figure 7C). Only 3 genes, Myh6, Gm2a and Hp, were identified as differentially expressed by bulk RNA-seq and had their protein products also differentially expressed in the LC-MS/MS data (Figure 7C). Similarly, gene enrichment in CD150 ${ }^{+}$WT HSCs revealed little overlap, with only 7 genes (Vcl, Pdcd4, Hist1h2ak, Psat1, Mt2, Mt1, Sub1) consistently identified across both datasets (Figure 7D).

In addition to the identification of key self-renewal drivers, a DEG-based comparative analysis between two biological states brings both -omic paradigms onto a single, directly comparable dimension. We isolated genes identified by proteomics and transcriptomics and plotted the corresponding $\log _{2} \mathrm{FC}$ values between CD150+ WT and CD150 ${ }^{+}$TET2 ${ }^{-1}$ HSCs (Figure 7E). Strikingly, we observed no correlation between the proteome and transcriptome (Figure 7E), indicating that proteomics and transcriptomic screens capture distinct molecular repertoires. Limited correlation between transcriptomes and proteomes has been widely recognised across a variety of biological systems ${ }^{394-396}$. The limited correlation between the HSC proteome and transcriptome is thus in line with previous observations.


Figure 4.7. LC-MS/MS and RNA-seq capture unique molecular profiles of WT and TET2-/- HSCs. (A) Differential protein expression for CD150 ${ }^{+}$HSCs from WT against TET2 ${ }^{-/-}$animals. Significant enrichment was classified as $\log _{2} F C>0.5$ and p-value $<0.05$. Proteins enriched in the CD150 ${ }^{+} W T$ fraction are marked in orange and proteins enriched in CD150+ TET2-/ cell fractions are marked in blue. (B) Differential gene expression (DGE) for CD150+ HSCs from WT against TET2-/ samples. DGE classified and marked as outlined for the proteome in (A). (C) Venn diagram highlighting the overlap between CD150 ${ }^{+}$TET2-/ fraction-enriched genes, identified by LC-MS/MS and bulk RNA-seq. (D) Venn diagram highlighting the overlap between CD150+ WT fraction-enriched genes, identified by LC-MS/MS and bulk RNA-seq. (E) Correlation between proteomic and transcriptomic profiling. Corresponding $\log _{2} F C$ values between CD150+ WT and CD150 ${ }^{+}$TET2-/ samples for genes identified in both datasets were utilised.

### 2.5. Proteomics identifies new pathways enriched in TET2 $^{-/}$HSCs

As mentioned above, a disconnect between the proteome and transcriptome has been widely recognised. The inherently different capture methods (e.g., polyA-based capture and amplification versus non-amplified sequencing), protein turnover and the divergence between transcription and translation are likely causes of such discrepancies. However, such discrepancies do not necessarily mean that both datasets should be mutually exclusive. Hence, we set out to identify whether both techniques have capturing biases of specific biological functions or rather highlight different components of the same molecular machineries.

We computed the enrichment of curated Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways in CD150 ${ }^{+}$TET2 ${ }^{-/}$HSCs (Figure 8). Targets identified by LCMS/MS and bulk RNA-seq were kept separate for the analysis. We focused on the CD150 ${ }^{+}$TET2 ${ }^{-1-}$ HSC cell fraction, as there is a quantified functional difference in selfrenewal compared to WT HSC counterparts. Interestingly, we observed LC-MS/MSspecific enrichment of several functions associated with extracellular proteins, including leukocyte trans-endothelial migration, extracellular trap formation, proteoglycans, focal adhesion and ECM-receptor interactions (Figure 8). While the transcriptome also identified enrichment of cell adhesion molecules and cytokinecytokine receptor interactions, multiple signalling pathways were particularly prevalent (Figure 8).


Figure 4.8. Transcriptomic profiling, but not LC-MS/MS, preferentially identifies signalling pathways in TET2-/- HSCs. KEGG pathway analysis (manually curated list) for genes enriched in TET2-/ HSCs. Enrichement analysis was computed separately for genes identified by LC-MS/MS and bulk RNA-seq. A significance cut-off (p-value) was selected (red line).

The consistent enrichment of extracellular matrix (ECM) and cell-cell interaction pathways across both omics datasets was particularly encouraging. The importance of niche signals for extrinsic regulation of HSC self-renewal has been widely recognised ${ }^{112}$ and several signalling pathways, namely MAPK and PI3K-mediated signalling, have been identified downstream of integrin activation in response to niche signals and ECM interactions ${ }^{397}$. MAPK and PI3K signalling enrichment in CD150+ TET2 ${ }^{-/}$HSCs thus complements these findings (Figure 8). Furthermore, several additional signalling pathways, including MAPK, PI3K, JAK-STAT and FoxO signalling, have also been implicated in HSC proliferation and elevated selfrenewal ${ }^{398-401}$ and these are also evident in the proteomics and transcriptomic datasets. Overall, these data indicate that while proteomic and transcriptomic analysis of HSCs identified distinct molecules, the biological function represented by these molecules appear to be complementary and, in the case of regulating HSC selfrenewal potential, raise a particular interest in the role of the ECM in this process.

### 4.6. Intracellular calcium regulators, inflammatory response, extracellular matrix and cytoskeletal proteins are deregulated in TET2 ${ }^{-/}$HSCs

As outlined above, KEGG pathway analysis of CD150 ${ }^{+}$TET2 ${ }^{-/}$HSCs raised interest in the role of the ECM in HSC self-renewal. To identify a comprehensive set of key biological functions underlying HSC self-renewal, a Reactome pathway analysis was undertaken comparing CD150+ WT against CD150+ ${ }^{+}$TET2 ${ }^{-/}$HSCs. As expected, multiple ECM related pathways (marked with a red asterisk) were enriched in the CD150 ${ }^{+}$TET2-/ cell fraction (Figure 9A). Interestingly, multiple ECM-related functions were also enriched in the CD150+ WT fraction (Figure 9B). These results suggest potential remodelling of the ECM space in response to TET2 knockout, which could impact HSC localisation and/or self-renewal potency.

Furthermore, we also observed that cytosolic calcium concentration ( $\left[\mathrm{Ca}^{2+}\right]_{c}$ ) regulation was the most significantly enriched pathway in CD150 ${ }^{+}$WT HSCs, but not in TET2-/ HSCs (Figure 9B). Such an observation provides some preliminary evidence of reduced $\left[\mathrm{Ca}^{2+}\right]_{c}$ regulation in $T E T 2^{-/}$HSCs which is in line with previous reports of stringent $\left[\mathrm{Ca}^{2+}\right]_{c}$ regulation in HSCs being a key factor for HSC maintenance, with low $\left[\mathrm{Ca}^{2+}\right]_{c}$ levels being supportive ${ }^{402}$. Interestingly, low $\left[\mathrm{Ca}^{2+}\right]_{c}$ levels were shown to stabilise TET2 function to help suppress HSC self-renewal in low $\left[\mathrm{Ca}^{2+}\right]_{c}$ conditions ${ }^{403}$. The relative absence of calcium concentration regulators in our TET2-/ dataset would therefore imply that a TET2 knockout desensitises HSCs to $\left[\mathrm{Ca}^{2+}\right]_{c}$ regulation and may play a role in regulating the increased self-renewal.


Figure 4.9. WT and TET2-/- HSCs differ in extracellular matrix proteins and intracellular calcium regulators. (A) Reactome pathway analysis for proteins enriched in CD150+ TET2 $^{-1}$. Significance cutoff of $p$-value $<0.05$. (B) Reactome pathway analysis for proteins enriched in CD150 ${ }^{+}$WT. Significance cut-off of $p$-value $<0.05$. Gene ratios (proportion of genes from total enrichment set) are depicted.

To identify protein complexes regulating HSC self-renewal, a protein-protein interaction network was constructed using proteins enriched in CD150 ${ }^{+}$TET2-- HSCs. We did not display singular proteins without connecting edges (Figure 10). Next, proteins were highlighted according to their involvement in selected biological functions (Figure 10). Interestingly, three major biological processes were overrepresented in CD150 ${ }^{+}$TET2 ${ }^{-/}$HSCs: cell cycle regulators (green), inflammatory response proteins (red), and actin cytoskeletal proteins (blue) (Figure 10).


Figure 4.10. Protein network analysis reveals enrichment of cytoskeletal, inflammatory and cell cycle proteins in TET2-- HSCs. Protein-protein interaction network of proteins enriched in the CD150+ fraction of TET2 cells. Node colours indicate proteins implicated in actin filament-based processes (blue), inflammatory responses (red) and cell cycle (green). Edges represent protein-protein interactions, based on curated experimental data, literature and co-expression scoring.

Intriguingly, the actin cytoskeleton has been previously shown to connect with integrins and induce mechanotransduction, downstream of ECM-integrin-mediated signalling of HSCs $^{404}$. However, the underlying molecular mechanisms of mechanosensation remain nearly completely unexplored ${ }^{404}$. The proteomics screen described above provides a set of key proteins potentially involved in governing the HSC response to mechanical stimulation. Most prominently, there was an enrichment of ACTN2, ACTA1, MYL2, MYH7, MYH8 and MYH10 (Figure 10) and these structural proteins might enable TET2 $^{-/}$HSCs to tolerate mechanical forces ${ }^{405}$. Another potential target protein, XIRP2, has previously been implicated in the response to mechanical stress, but its function in HSCs currently remains unknown ${ }^{406}$.

Several key cell cycle regulators, such as CDK6, DSN1 and TAF10, were enriched in TET2-/ HSCs. All three targets protein have been identified as potent drivers of cell cycle progression and proliferation ${ }^{407-409}$. While DSN1 and TAF10 have not yet been implicated in HSC biology, CDK6 was shown to accelerate quiescence exit of human HSC ${ }^{408}$. Finally, it has been long described that various inflammatory signals have roles in regulating HSC self-renewal and proliferation ${ }^{145,410,411 \text {. The inflammatory }}$ response proteins S100A8 and S100A9 are involved in a variety of fucntions, including cell proliferation and cytokine secretion in response to inflammatory stimulation ${ }^{412,413}$. Most intriguiongly, S100A8/A9 have been identified as biomarkers for acute myeloid leukaemia $(A M L)^{412}$.

Overall, these data agree with previous findings which implicate inflammatory signalling and cell cycle progression and build additional evidence that mechanosensation and the ECM might be involved in the regulation of HSC selfrenewal. Most importantly, the proteomic screen provides a novel set of candidate proteins not identified by transcriptomic studies, particularly for functions with poorly characterised molecular mechanisms. Most notably, LC-MS/MS is best positioned to interrogate cell surface and structural proteins, allows to explore self-renewal regulators beyond DNA-binding and transcription factor targets.

### 4.7. Single cell RNA-seq revealed enrichment of cell division and SUMOylation in TET2 ${ }^{-/}$HSCs

So far, my thesis has focused on a direct comparison of CD150+ WT and CD150+ TET2-/ HSCs with bulk proteomics and bulk RNA-seq. In order to gain additional information at the single cell level, I undertook scRNA-seq profiling of WT and TET2-/ LT-HSCs from male mice (CD45 ${ }^{+}$CD48-EPCR ${ }^{+}$CD150 ${ }^{+}$Sca1 ${ }^{+}$) to determine whether candidate gene profiles could be recapitulated at single cell resolution. Following removal of cells not passing quality control (see methods), cells were plotted on a UMAP representation and WT and TET2-/ HSCs formed distinct clusters (Figure 11A). To determine the difference between WT and TET2-/ HSCs, the previously described LT-HSC Molecular Overlap (MoIO) gene signature described by Wilson et al., was utilised ${ }^{332}$. MolO signature scoring revealed that WT and TET2-/ HSCs displayed
indistinguishable MolO scores (Figure 11B) and furthermore, cell cycle scores computed for each cell were also not different, with most cells having the molecular profile of cells in the $\mathrm{G}_{0} / \mathrm{G}_{1}$ phase (Figure 11C).


Figure 4.11. Single cell transcriptomics reveals distinct molecular profiles of WT and TET2-/HSCs. (A) UMAP representation of scRNA-seq profiles for WT (blue) and TET2/- (red) HSCs (ESLAM cell fractions of sex-matched mice). (B) Violin plot summarising computed gene set scores for the LT-HSC-specific MolO signature (MolO Score). (C) Proportions of cells in G1/G0, G2M and S cell cycle phases, grouped by genotype (WT and TET2-/). Scores were computed based on correlation with curated cell cycle gene sets. (D) Differential gene expression (DGE) between WT and TET2/ HSCs (ESLAM cell fraction). Significance cut-offp-value $<0.05$ and $\log _{2} F C>0.5$. (E) Violin plots of four genes, enriched in WT HSCs (Fos, Fosb, Junb, Jund).

Since global differences were not observed, differential gene expression was next performed between both genotypes and DEGs determined based on the identical cutoffs set for LC-MS/MS and bulk RNA-seq ( $p$-value $\leq 0.05$ and $\log _{2}$ FC $\geq 0.5$ ). Interestingly, we observed significant enrichment of histone genes (H4c4, H4c3, H1f4) and epigenetic regulators (Uty, Gtf3c4, Kdm5d). While these histones and epigenetic regulators have not yet been directly implicated in TET2 ${ }^{-/}$HSC function, broad alteration of the epigenome in response to the loss of TET2 are conceivable. In addition, we also observed the several mitochondrial genes (Acacb, Lars2), protein synthesis (Eif2s3y) and most intriguingly, the cytoskeletal gene Actb (Figure 11D). Increased mitochondrial activity and protein synthesis in HSCs are consistent with increased proliferation ${ }^{414,415}$. In contrast, we observed downregulation of key members of the AP-1 complex, namely Fos, Fosb, Junb and Jund, in TET2-/ HSCs ${ }^{416}$ (Figure 11E). Interestingly, the loss of AP-1 has been shown to increase HSC proliferation, while preserving self-renewal potency and multilineage capacity ${ }^{417}$. However, another recent report highlighted that the AP-1 complex was dispensable for HSC quiescence, as hibernating HSCs lost AP-1 expression over a period of 7 days of ex vivo culture ${ }^{334}$. Despite the data presented in this thesis identifying the downregulation of the AP-1 complex, the expression profiles suggest that expression is not fully abrogated.

Next, a Reactome pathway analysis was performed to identify key functions associated with HSC self-renewal. Interestingly, enrichment of SUMOylation and SUMO target proteins was observed (Figure 12A). While there is not much literature on SUMOylation in HSCs, the SUMO ligase PIAS1 has previously been implicated in the regulation of self-renewal ${ }^{418}$. Pias1 itself was not significantly differentially expressed $\left(\log _{2} \mathrm{FC}=0.51\right.$; adj $p$-value $\left.=1\right)$, but the enrichment of SUMOylation target proteins in TET2 $^{-/}$HSCs suggests a potential role of SUMOylation in supporting selfrenewal decisions. In contrast, WT HSCs were enriched for Toll like receptor (TLR) signalling and cell cycle checkpoint regulators, which have been implicated in stress responses, differentiation and other HSC functions ${ }^{419-421}$ (Figure 12B). Overall, these data provide an insight into the possible mechanism of HSC self-renewal. Several functions, such as cytoskeletal genes and cell cycle regulators, are in line with bulk RNA-seq and proteomics. As the experimental setup was based on a direct comparison of WT and TET2-deficient HSCs, these data also provide an intriguing insight into the disease mechanism of TET2-mediated MPNs.


Figure 4.12. Knockout of TET2 induces mitosis and SUMOylation in HSCs. (A) Reactome pathway analysis of differentially expressed genes (DEGs), enriched in scRNA-seq profiles of TET2-/ HSCs (ESLAM). (B) Reactome pathway analysis of DEGs, enriched in scRNA-seq profiles of WT HSCs (ESLAM). Enrichment cut-off: p-value < 0.05.

### 4.8. TET2 knockout induces remodelling of HSC extracellular matrix proteins

Unbiased analyses of the biological functions and molecular pathways across both the proteome and transcriptome indicated a change in extracellular matrix proteins between TET2 and WT HSCs. Whether or not these molecules are related to the functional differences in HSC self-renewal was next investigated. First, all ECM proteins identified by LC-MS/MS were plotted on a volcano plot which revealed enrichment of multiple collagen proteins in TET2 $^{-/}$HSCs (Figure 13A). In contrast, 16 ECM proteins were shown to be downregulated in TET2 $2^{-/}$HSCs (Figure 13A), including transglutaminase 2 (TGM2 or TG2), VWF, ITIH2, ITIH3, SERPINC1 and a plethora of structural proteins. TGM2 also showed high expression in the most primitive stem cell cluster containing LT-HSCs from a previously published scRNAseq landscape of haematopoietic stem and progenitor cell types ${ }^{22}$ (Figure 13B). A recent study showed that deletion of Tgm2 increased the frequency of HSCs, inhibited differentiation and depleted multipotent progenitor populations ${ }^{422}$, again suggesting that ECM proteins play a role in regulating HSC self-renewal.

```
A Not Significant - CD150+ WT - CD150+ TET2KO
```




Figure 4.13. Differentially expressed ECM proteins between WT and TET2-/ HSCs. (A) DEGs between CD150 ${ }^{+}$WT and CD150 ${ }^{+}$TET2 $2^{-/}$cells, as identified by LC/MS/MS. ECM proteins are selectively depicted. (B) UMAP reference map of mouse HSPC scRNA-seq profiles²2. Cell type annotations (left) and Tgm2 expression (right) are depicted.

Next, a protein-protein interaction network was constructed for all identified ECMassociated proteins where proteins enriched in TET2-/ HSCs are marked in blue and WT enriched ECM proteins are marked in orange (Figure 14A). A striking enrichment of type I collagen and type VI collagen was observed in TET2-/ HSCs (Figure 14A) compared to a fibrinogen network of molecules including fibulin (FBLN1), plasminogen (PLG), von Willebrand factor (VWF) and fibrinogen (FGG) upregulated in WT HSCs
(Figure 14A). Additionally, WT-enriched ITIH2 has been shown to form protein complexes with hyaluronan ${ }^{423}$. Together these data indicate that the loss of TET2 leads to a remodelling of the ECM could alter HSC niche interactions.


Figure 4.14. TET2 knockout HSCs are enriched for collagens and their respective integrin binding partners. (A) Protein-protein interaction network of ECM proteins identified by proteomic profiling in CD150+ HSCs. ECM protein enriched in TET2-/ HSCs are highlighted in blue and WTenriched ECM proteins are marked in orange. (B) Average scaled protein abundances for integrins, identified by LC-MS/MS. Protein abundances are grouped by WT and TET2-/ phenotype.

ECM proteins are recognised and anchored by varying types of integrins which induce downstream signalling cascades ${ }^{424}$. When scaled protein abundances were plotted for identified integrins for WT and TET2-/ HSCs (Figure 14B), ITGB2 and ITGB2L were found to be enriched in TET2-/ HSCs. ITGB2 (CD18) has been previously shown to preferentially bind type I and type VI collagen ${ }^{425,426}$ and ITGB1 (CD29) has been shown to preferentially bind to hyaluronan and fibronectin, while also forming heterodimers with ITGA6 (CD49f) ${ }^{427-429}$. Furthermore, ITGB3 (CD61) was shown to selectively bind fibrinogen, plasminogen, fibronectin and VWF ${ }^{430,431}$. Similarly, ITGA2B (CD41) binds fibrinogen and dimerises with CD61 ${ }^{432}$. Collectively, these data show remarkable consistency of complementary enrichment of various integrins and their ECM ligands in the TET2-/ setting.

Next, we hypothesised that the protein abundances of self-renewal regulators would correspond with the continuum of self-renewal potency across screened cell types. A continuum of self-renewal potency in ascending order was defined as follows: 1) CD150+ WT, 2) ESLAM WT, 3) CD150+ TET2-\%. First CD150+ WT-enriched ECM proteins were plotted along the continuum and, strikingly, all proteins excluding CTSS followed the continuum (Figure 15A). On the other side, the protein abundance of key structural ECM proteins and cell:cell adhesion molecules, enriched in TET2 ${ }^{-/} \mathrm{HSCs}$, increased in line with self-renewal potency (Figure 15B). These included type I collagen, type VI collagen, LGALS1, S100A8, MGP and MMP9 (Figure 15B). Thus, we further curated the list of novel potential key self-renewal regulators (Figure 15C).

A



B


C
Downregulated candidate proteins

| Comp | Ctss | Fbln1 | Fgg |
| :--- | :--- | :--- | :--- |
| Plg | Plxnc1 | S100a13 | Serpinc1 |
| Anxa4 | Anxa5 | Gdf3 | Itih2 |
| Ith3 | Serpina1a | Tgm2 | Vwf |

Upregulated candidate proteins

| Col1a1 | Col1a2 | Col6a1 |
| :--- | :--- | :--- |
| Lgals1 | S100a8 | Mgp |
| Mmp9 |  |  |

Figure 4.15. ECM protein abundance changes are proportional to the self-renewal potency. (A) Line plots for ECM proteins enriched in WT HSCs (CD150+ fractions). Average protein abundance across 2 biological replicates for CD150+ WT, ESLAM ET and CD150+ $T E T 2^{-/}$fractions, ordered from lowest to highest stem cell frequency respectively (left to right). (B) Line plots for ECM proteins enriched in TET2- HSCs (CD150 ${ }^{+}$fractions). Average protein abundance across 2 biological replicates for CD150 ${ }^{+}$WT, ESLAM ET and CD150 ${ }^{+}$TET2 ${ }^{-/}$fractions, ordered from lowest to highest stem cell frequency respectively (left to right). (C) Curated list of candidate proteins for HSC self-renewal regulation. Contains a list of genes downregulated (left) and upregulated (right).

### 4.9. Bulk and scRNA-seq profiles of ECM-associated genes

After establishing a curated set of candidate ECM proteins, we wondered whether these candidates were recapitulated by the transcriptome and whether new ECM
targets could be identified. Transcriptomic profiles of ECM-associated genes were extracted to compute DGE between CD150+ WT and TET2-deficient HSCs (Figure 16A). Interestingly, only 2 DEGs overlapped with the proteome, namely Col1a2 and Tgfb1 (Figure 15C and 16A). However, Col1a2 and Tgfb1 transcripts were overrepresented in WT LT-HSCs, while the proteome indicated enrichment in TET2deficient LT-HSCs. Inspection of ECM candidate protein expression in the scRNA-seq landscape of WT and TET2-deficient LT-HSCs found that none of the proteins downregulated upon increasing self-renewal potency were differentially expressed in the scRNA-seq dataset (Figure 16B). Several candidates were not even detected by scRNA-seq, including: Serpinc1, Serpina1a, Itih2, Itih3, Fbln1, Fgg and Plg. One molecule, Lgals1, enriched in the TET2-deficient HSC proteome, was also identified as a DEG in single cell transcriptomes (Figure 16C) but all remaining upregulated proteins in the TET-deficient HSCs were either not differentially expressed or not detected (Figure 16C). Overall, the transcriptome poorly captured differentially expressed ECM proteins, in particular with respect to structural proteins, such as type I collagen, type VI collage, plasminogen and fibrinogen which were not detected or poorly captured by transcriptomics. These data suggest that these proteins could stem from neighbouring niche cells and not be inherently expressed by HSCs or that the proteins are sufficiently long-lived that measuring the active transcriptomic has little bearing on the amount of protein remaining. Such possibilities highlight the complementary nature of LC-MS/MS for deciphering the molecular mechanisms governing HSC self-renewal, especially as a surrogate for understanding the molecules attached to the HSCs themselves.


Figure 4.16. ECM genes are differentially expressed between WT and TET2-/- HSCs. (A) DGE of ECM genes in WT and TET2 ${ }^{-1}$ CD150+ HSCs. Significance cut-offs: p-value $<0.05$ and $\log _{2} F C>0.5 \mid$ $\log _{2} F C<-0.5$. (B) Violin plots of candidate ECM proteins in scRNA-seq profiles of WT and TET2-/ HSCs. (C) Violin plots of ECM proteins, enriched in the CD150 ${ }^{+}$TET2-/ fraction.

### 4.10. Ex vivo cultures of LT-HSCs on singular ECM proteins did not significantly alter HSC expansion

The observation that structural changes in ECM molecules were present between TET2-/ and WT HSCs raised the question of whether or not direct interaction of HSCs with different ECMs helped determine the self-renewal potency of HSCs. To test this initial hypothesis, HSC expansion cultures were prepared by coating tissue culture plastic with identified ECM components and culturing HSCs (Figure 17A). 200 WT and TET2-/ ${ }^{-1}$ T-HSCs (CD45 ${ }^{+}$CD48-EPCR ${ }^{+}$CD150 ${ }^{+}$Sca1 ${ }^{+}$) were cultured for a period of 28 days on tissue culture plastic, coated with type I collagen, type VI collagen, plasminogen, hyaluronan, fibrinogen and fibronectin (Figure 17A). At the end of the expansion period, we profiled the cultures for ELSK HSCs (EPCR ${ }^{+}$Lin ${ }^{-}$Sca1 ${ }^{+}$cKit $^{+}$) by flow cytometry ${ }^{333}$ (see Chapter 5) to quantify HSC expansion. Interestingly, we did not observe significant differences in HSC expansion between WT and TET2-/ genotypes on coated plates with none of the ECM proteins outperforming the Fibronectin-coated control (Figure 17A). These data suggest that simple addition of a tissue culture plastic coating with different ECMs was not sufficient to alter HSC self-renewal output.

ECM protein, and particularly its structural components, are required for cell tethering in the in vivo microenvironment. Tissue culture plastic does not adequately recapitulate such environment. Instead, hydrogel-based culture systems provide a surface that enables tethering, representative of native microenvironments. The StemBond hydrogel provides a tuneable system that can be functionalised with ECM proteins ${ }^{336}$. To closely resemble the bone marrow niche, Soft (750 Pa) hydrogels were functionalised with the identical set of ECM proteins and ESLAM HSCs were cultured for 28 days (Figure 17B). Interestingly, we observed significantly higher levels of ELSKs in TET2 $^{-/}$cells, cultured in type I collagen and hyaluronan (Figure 17B). While increased expansion of $T E T 2^{-/}$HSCs in type I collagen cultures is consistent with the enrichment of Itgb2 in TET2 $2^{-/}$HSCs, TET2 $^{-/}$-specific expansion in hyaluronan cultures was not in line with the proteome. Of note, none of the ECM conditions outperformed the fibronectin control in proportion of LSK and ELSK cells (Figure 17B). While these data provide preliminary evidence of some impact of ECM proteins on HSC selfrenewal potency, the similarity with the fibronectin control indicates that single ECM factors are possibly not sufficient to elicit an effect. More robust combinatorial screens
are required to further determine the differential impact of the ECM on WT and TET2${ }^{\text {- HSCs. However, these preliminary finding provide a promising nature for a functional }}$ difference between both genotypes.


Figure 4.17. In vitro expansion of HSCs in presence of ECM proteins. (A) Expansion of 50 WT and TET2 ${ }^{-/}$HSCs (ESLAM-Sca1) for 28 days in PVA/F12 culturing conditions. Selected ECM proteins were immobilised on tissue culture plastic prior to HSC culture. Flow cytometry analysis following 28-day expansion was used to determine cell type distribution of cell types, indicated as percentage of total cells. Cell types included: Lin-, LSK (Lin-Sca1+ ${ }^{+}$Kit $^{+}$) and ELSK (EPCR ${ }^{+}$Lin-Sca1 $\left.{ }^{+} c K i t^{+}\right)$cell fractions. The experiment was conducted using WT mice only and 10 bulk ( 50 cell) HSC cultures per ECM protein. (B) Bulk expansion ( 50 cells per replicate) of WT and TET2-/ ESLAM-Sca1 HSCs for 28 days in PVA/F12 culturing conditions. Soft ( 750 Pa ) hydrogels were functionalised with ECM proteins prior to 28 -day expansion culture. Flow cytometry analysis of stem and progenitor cell types was conducted as outlined in $(A)$.

### 4.11. Intracellular calcium binding proteins are deregulated in TET2 $^{-/}$HSCs

The Reactome pathway analysis identified $\left[\mathrm{Ca}^{2+}\right]_{c}$ regulation as the most significantly overrepresented pathway in WT LT-HSCs (Figure 9B). In contrast, TET2-deficient LTHSCs did not exhibit any enrichment of $\left[\mathrm{Ca}^{2+}\right]_{c}$ pathways. Based on these observations, the proteome was inspected for differential expression of calciome proteins, which include $\mathrm{Ca}^{2+}$ regulators, ion channels and $\mathrm{Ca}^{2+}$-dependent proteins. Collectively, 9 calciome proteins were overrepresented in WT HSCs, while TET2deficient HSCs showed enrichment of 3 calciome proteins (Figure 18A). In brief, the voltage-dependent anion channels VDAC1 and VDAC3, which mediate $\mathrm{Ca}^{2+}$ transport between the cytosol and mitochondria, were enriched in CD150 ${ }^{+}$WT HSCs (Figure 18A). Particularly VDAC1 has been shown as a key mediator of apoptotic intracellular $\left[\mathrm{Ca}^{2+}\right]$ signalling, whereby high $\left[\mathrm{Ca}^{2+}\right]_{c}$ induce apoptosis ${ }^{433,434}$. Similarly, ATP2B4, an active $\mathrm{Ca}^{2+}$ efflux ion pump to clear excessive $\left[\mathrm{Ca}^{2+}\right]_{c}$ levels and avoid apoptotic signalling ${ }^{435}$, was enriched in CD150 ${ }^{+}$WT HSCs (Figure 18A). Interestingly, two isoforms of the potent intracellular $\left[\mathrm{Ca}^{2+}\right]$ regulator sarcoendoplasmic reticulum calcium transport ATPase (SERCA), namely ATP2A2 and ATP2A3, were also enriched in CD150 ${ }^{+}$WT HSCs ${ }^{436}$ (Figure 18A). Collectively, these data indicate that reduced expression of such proteins in TET2-deficient HSCs might alter or desensitise TET2 ${ }^{-/} \mathrm{HSCs}$ to elevated $\mathrm{Ca}^{2+}$ concentrations.


Figure 4.18. Intracellular Ca2+ regulators are deregulated in TET2-/- HSCs. (A) Differential protein expression for WT and TET2 CD150 ${ }^{+}$HSCs, using LC-MS/MS data. Only calciome ${ }^{437}$ proteins are displayed. Significance cut-offs: $p$-value $<0.05$ and $\log _{2} F C>0.5$ (TET2-1 enriched) | $\log _{2} F C<-0.5$ (WT enriched). (B) Differential gene expression for WT and TET2 ${ }^{-1}$ CD150 ${ }^{+}$HSCs, using bulk transcriptome (RNA-seq) data. Only calciome proteins are displayed. Significance cut-offs: $p$-value $<0.05$ and $\log _{2}$ FC $>0.5$ (TET2 enriched) $\mid \log _{2}$ FC $<-0.5$ (WT enriched). (C) Line plots for calciome proteins enriched in WT HSCs (CD150+ fractions). Average protein abundance across 2 biological replicates for CD150+ WT, ESLAM ET and CD150+ TET2 $^{-/}$fractions, ordered from lowest to highest stem cell frequency respectively (left to right). (D) Line plots for calciome proteins enriched in TET2-/ HSCs (CD150+ fractions). Average protein abundance across 2 biological replicates for CD150 $W$, ESLAM ET and CD150 ${ }^{+}$TET2-/ fractions, ordered from lowest to highest stem cell frequency respectively (left to right).

In line with previous observations, the bulk RNA-seq did not recapitulate the proteome, with just 7 genes differentially expressed between CD150+ WT and TET2-deficient HSCs (Figure 18B). All CD150 ${ }^{+}$WT-associated proteins were continuously downregulated as self-renewal increased (Figure 18C) and TET2 ${ }^{-}$-enriched ATP2A1 and TNNC2 were in line with the continuum, while CAMK2D plateaued (Figure 18D). When proteome-derived calciome proteins were specifically assessed in the scRNAseq profiles of WT and TET2 ${ }^{-/}$LT-HSCs there were also no differentially expressed
genes (Figure 19). These data further indicate that LC-MS/MS is useful for identifying novel networks of targets specific to mutant HSCs with different self-renewal properties.


Figure 4.19. Single cell transcriptomes do not corroborate LC-MS/MS-derived calciome deregulation in TET2-/ HSCs. Violin plots of calciome genes in WT and TET2-/ scRNA-seq profiles. Calciome target were selected based on enrichment in the LC-MS/MS dataset.

In order to test whether $\left[\mathrm{Ca}^{2+}\right]_{\mathrm{c}}$ regulation is disrupted in $\mathrm{TET2}^{-1} \mathrm{HSCs}$ and rendering them less sensitive to elevated $\mathrm{Ca}^{2+}$ concentrations, intracellular $\mathrm{Ca}^{2+}$ imaging was performed to identify the uptake of extracellular $\mathrm{Ca}^{2+}$ by WT and TET2-deficient HSCs. ESLAM-Sca1+ HSCs were cultured at three extracellular $\left[\mathrm{Ca}^{2+}\right]$ prior to imaging. As expected, elevated extracellular $\left[\mathrm{Ca}^{2+}\right]$ resulted in a continuously increasing $\left[\mathrm{Ca}^{2+}\right]_{c}$ across both genotypes (Figure 20). However, TET2-deficient cells consistently had higher levels of intracellular calcium concentration (Figure 20). These data correspond to the proteomic profiles, which outlined deficiency of $\mathrm{Ca}^{2+}$ regulating proteins in TET2deficient HSCs. Overall, the consistency of protein profiles raised confidence in the identified candidates for the underlying molecular mechanism of $\left[\mathrm{Ca}^{2+}\right]_{c}$ regulation in

HSC self-renewal decisions. In addition, the proteome-transcriptome dichotomy highlights the importance of multi-omic screening to uncover more complete molecular mechanisms operative in HSCs and governing cellular outputs.


Figure 4.20. TET2-deficient HSCs are desensitised towards extracellular calcium. Expansion of 50 WT or TET2-/ HSCs (ESLAM-Sca1) for 7 days in PVA/F12 culturing conditions, supplied with additional $\mathrm{CaCl}_{2}$ to achieve $\mathrm{Ca}^{2+}$ concentrations of $0.3 \mathrm{mM}, 1.5 \mathrm{mM}$ and 5 mM . Cells were plated on glass cover slips and subjected to calcium dye prior to imaging. Error bars represent standard error.

### 4.12. HSC and progenitors express distinct sets of calcium regulators

Finally, the implication of $\left[\mathrm{Ca}^{2+}\right]_{c}$ in HSC fate and more specifically self-renewal raises the question whether LT-HSCs exhibit distinct regulation of $\left[\mathrm{Ca}^{2+}\right]_{c}$ within the haematopoietic hierarchy. To address this, t-SNE plots were computed for calciome genes using a dataset of $\sim 1600$ haematopoietic stem and progenitor cells ${ }^{22}$ (Figure 21A). Distinct clustering was observed of LT-HSCs away from progenitor cell populations (Progs) indicating that LT-HSCs are likely to express a unique set of $\left[\mathrm{Ca}^{2+}\right]_{c}$ regulators. Differential gene expression identified two ion channels with preferential specificity for $\mathrm{Ca}^{2+}$ transport to regulate intracellular $\mathrm{Ca}^{2+}$ concentrations (Figure 21B). P2rx1 is an ATP receptor responsible for $\mathrm{Ca}^{2+}$ influx ${ }^{438}$ and extracellular ATP release activates $P 2 r x 1$ to induce $\mathrm{Ca}^{2+}$ influx, a process also implicated in neutrophil chemotaxis ${ }^{439}$. A similar role was also attributed to another candidate, the $\mathrm{Ca}^{2+}$ channel molecule Trpc6, whereby inflammatory responses trigger TRPC6mediated $\mathrm{Ca}^{2+}$ influx to trigger leukocyte migration ${ }^{440}$. The enrichment of $P 2 r x 1$ and Trpc6 in LT-HSCs might suggest a role of $\left[\mathrm{Ca}^{2+}\right]_{c}$ in the cellular response to broad range of stimuli.


Figure 4.21. The calciome of HSCs differs from progenitor cell populations. (A) $t$-SNE representation of the calciome of HSC and progenitor cells, extracted from previously publishe scRNAseq profiles ${ }^{22}$. Dimnesionality reduction (t-SNE) was computed using the calciome only. Dimension 1 and Dimension 2 were plotted, and cell types depicted. (B) Dotplot of differentially expressed calciome genes from scRNA-seq profiles, outlined in (A). dot size indicates the proportion of cells expressing the gene in question and colour intensity indicates scaled average gene expression.

### 4.13. Haematopoietic transcription factors are deregulated in TET2 $^{-/}$HSCs

To complement the findings in proteomics, transcriptional data were also analysed to identify key transcription factors regulating HSC self-renewal and disease progression of TET2-deficient HSCs. For this purpose, DNA-binding genes were isolated and identified by LC-MS/MS (Figure 22A) and bulk RNA-seq (Figure 22B). Interestingly, the transcriptome only provided a limited set of differentially expressed DNA-binding
genes (Figure 22B) whereas differential protein expression was observed for several key transcription factors and chromatin modellers (Figure 22A). MEIS1, HMGA1, HMGA2, ERG and FLI1 were downregulated in TET2-/ HSCs (Figure 22A). These factors being directly implicated in HSC maintenance and self-renewal potency, albeit at various developmental stages ${ }^{441-445}$. While these genes are key regulators of WT HSC maintenance, our data suggest that these do not drive the $T E T 2^{-} \%$-mediated selfrenewal advantage.


Figure 4.22. Key haematopoietic transcription factors are downregulated in TET2\%- HSCs. (A) Volcano plot of all DNA-binding proteins, identified by LC-MS/MS in WT and TET2-1 CD150+ HSCs. DNA-binding proteins enriched in WT HSCs are marked in orange and TET2-1 HSC-enriched proteins are marked in blue. Significance cut-offs: $p$-value $<0.05$ and $\log _{2} \mathrm{FC}>0.5$ (TET2-1- enriched) $\mid \log _{2} \mathrm{FC}<$ -0.5 (WT enriched). (B) Volcano plot of all DNA-binding genes, identified by bulk RNA-seq in WT and TET2-/ CD150 ${ }^{+}$HSCs. DNA-binding genes enriched in WT HSCs are marked in orange and TET2-/ HSC-enriched genes are marked in blue. Significance cut-offs: p-value < 0.05 and $\log _{2}$ FC $>0.5$ (TET2${ }^{-}$enriched) $\mid \log _{2}$ FC $<-0.5$ (WT enriched). (C) Expression profiles of Elf1 and Tle2 transcription factors in scRNA-seq reference landscape ${ }^{22}$. Left: UMAP representation of the scRNA-seq reference landscape, depicting cell types. Middle: UMAP depiction of Elf1 expression. Right: UMAP depiction of Tle2 expression.

To identify which transcriptional regulators might drive TET2 $2^{-}$-mediated HSC selfrenewal expansion, the $T E T 2^{-/}$HSC proteome and transcriptome were inspected for likely candidates. Firstly, enrichment of the ETS transcription factor ELF1 was observed by LC-MS/MS (Figure 22A). Previous studies revealed that ELF1 localisation at Meis1 and Runx1 promoter regions induces expression ${ }^{446-448}$. In turn, both MEIS1 and RUNX1 are potent activators of HSC self-renewal ${ }^{446,447}$. While LC-MS/MS revealed downregulation of MEIS1 in TET2-- HSCs, RUNX1 protein was not detected (Figure 22A). However, Runx1 transcripts were moderately, but not significantly, enriched in scRNA-seq profiles of $T E T 2^{-/} \mathrm{HSCs}\left(\log _{2} \mathrm{FC}=0.51\right.$ and adj $p$-value $=1$; data not shown). Interestingly, Elf1 was ubiquitously expressed across stem and progenitor populations (Figure 22C). These data indicate that ELF1 could contribute towards the hyperproliferative phenotype of TET2-deficient HSCs.

Transcription factors enriched in in the transcriptome of TET2-deficient HSCs included several known regulators of HSC function, such as Sox4, Sox6, Stat1 and HIf ${ }^{\text {f49-452 }}$ (Figure 22B). Interestingly, we also observed enrichment of the transcriptional cofactor Tle2. Tle2 deletion has been shown to be embryonic lethal ${ }^{453}$ and its expression was confined to the LT-HSC and HSPC compartment (Figure 22C). Javed et al. identified TLE2 as a direct binding partner and co-factor of RUNX transcription factors ${ }^{454}$. In summary, these data provide an intriguing set of candidate genes, previously implicated in HSC self-renewal, as drivers of TET2-mediated expansion of HSCs. In addition, we highlight TLE2 as a possible potent, and largely novel, driver of HSC self-renewal.

### 4.14. Shear stress and ECM proteins are enriched in CD150 ${ }^{+}$TET2 ${ }^{-1}$ cells but not their CD150 counterparts

So far, our analysis has focused on the acquired self-renewal advantage of TET2deficient HSCs against their WT counterparts. However, since the process is likely only regulated by a subset of known and novel self-renewal regulators, we also sought to leverage the diverse set of HSC subtypes screened by LC-MS/MS and bulk RNAseq. First, DGE was performed comparing TET2-deficient CD150- finite self-renewal cells and CD150 ${ }^{+}$durable HSC cell fractions (Figure 23A/B). As above, the correlation
between the proteome and transcriptome based on $\log _{2} \mathrm{FC}$ was relatively poor (Figure 23C). Overall, 15 genes overlapped between the proteome and transcriptome, with $95 \%$ of identified genes and proteins not corresponding well (Figure 23D).


Figure 4.23. Comparison of proteome and transcriptome for CD150+ and CD150- TET2-/- cells. (A) Volcano plot of differentially expressed proteins between CD150 ${ }^{+}$and CD150 $\mathrm{TET2}^{-1}$ cells. Significance cut-offs: p-value $<0.05$ and $\log _{2} F C>0.5$ (CD150 ${ }^{+}$enriched) | $\log _{2} F C<-0.5$ (CD150 enriched). (B) Volcano plot of DEGs between CD150 ${ }^{+}$and CD150 TET2-/ cells. Significance cut-offs: $p$-value < 0.05 and $\log _{2} F C>0.5$ (CD150+ enriched) $\mid \log _{2} F C<-0.5$ (CD150 enriched). (C) Correlation between proteomic and transcriptomic profiling. Corresponding Log $_{2} F C$ values between CD150- TET2-- and CD150+ TET2-- samples for genes identified in both datasets were utilised. (D) Venn diagram of genes enriched in CD150+ TET2 $^{-/-}$HSCs, as indetified by LC-MS/MS and bulk RNA-seq.

Based on the previous comparison of CD150+ WT against CD150 ${ }^{+}$TET2-/ HSCs, a limited gene overlap was expected. Hence, we computed GO term enrichment in the CD150+ TET2 ${ }^{-/}$cell population to identify overlapping molecular machineries between the proteome and transcriptome. Here, the overlap accounted for $26 \%$ of transcriptome-enriched and 40\% of proteome-enriched GO terms (Figure 24A). In brief, consistent enrichment of developmental processes, cell cycle regulation, cell
adhesion, GTPase-mediated signalling, and stress responses were observed in CD150 ${ }^{+}$TET2 ${ }^{-/}$HSCs (Figure 24B). These processes have previously been well described in HSC maintenance and function ${ }^{421,455,456}$. Interestingly, overrepresentation of cytoskeletal organisation, ribosome biogenesis and translational functions was exclusively identified by LC-MS/MS (Figure 24C). These results are in line with our previous analysis of CD150+ WT and CD150+ TET2 $^{-/-}$HSCs suggesting that corresponding transcriptional profiles do not always capture changes in these machineries.


Figure 4.24. GO term enrichment of genes enriched in TET2-/- CD150+ HSCs. (A) Venn diagram of GO terms enriched in CD150+ TET2 $^{-/}$cells, when compared to the CD150- TET2-/ fraction. GO terms were first separately computed for LC-MS/MS and bulk-RNAseq data. (B) Manually selected GO terms, consistently identified by RNA-seq and LC-MS/MS. Computed enrichment adjusted p-values (log-transformed) were plotted for the bulk RNA-seq (blue, left-hand side) and LC-MS/MS (red, righthand side). Significance cut-off: p-value < 0.05 (red line). (C) Curated set of GO terms, uniquely enriched in the proteome (red, right-hand side) and bulk RNA-seq (blue, left-hand-side). Significance cut-off: p-value < 0.05 (red line).

We performed a similar analysis for KEGG pathways, which yielded 7 overlapping KEGG pathways for CD150+ ${ }^{+}$TET2 ${ }^{-/}$HSCs (Figure 25A). As expected, cell adhesion and niche interaction functions, namely gap junction, focal adhesion and ECMreceptor interactions, were identified by both -omic screens (Figure 25B). Intriguingly, fluid shear stress was also enriched. Fluid shear stress and the associated biomechanical forces have been implicated in haematopoietic development throughout embryogenesis, during which HSCs undergo extensive self-renewal ${ }^{457-459}$. In contrast, the role of shear stress and mechanosensation in adult HSCs remains poorly understood, with the underlying molecular drivers remaining particularly elusive. Our observations thus provide a novel set of candidate genes for adult HSC responses to mechanical stimuli. Furthermore, bulk RNA-seq appeared to preferentially identify signalling pathways, such as Wnt, VEGF, chemokine, JAKSTAT, calcium, Ras, PI2k-Akt and Rap1 signalling (Figure 25C). Overall, these data show consistent enrichment of niche interaction functions in CD150 ${ }^{+}$TET2--1, when compared to CD150 ${ }^{+}$WT and CD150 TET2 $^{-/-}$cell populations. In addition, this comparison of TET2-deficient durable LT-HSCs against finite LT-HSCs revealed a possible function of fluid shear stress in regulating adult HSC self-renewal.


Figure 4.25. KEGG pathway enrichment of genes enriched in TET2-/ CD150+ HSCs. (A) Venn diagram of KEGG pathways enriched in CD150 ${ }^{+}$TET2 $2^{-1}$ cells, when compared to the CD150- TET2-/ fraction. KEGG pathway enrichment was first separately computed for LC-MS/MS and bulk-RNAseq data. (B) KEGG pathways consistently identified by RNA-seq and LC-MS/MS. Computed enrichment adjusted p-values (log-transformed) were plotted for the bulk RNA-seq (blue, left-hand side) and LCMS/MS (red, right-hand side). Significance cut-off: p-value $<0.05$ (red line). (C) Curated set of KEGG pathways, uniquely enriched in the proteome (red, right-hand side) and bulk RNA-seq (blue, left-handside). Significance cut-off: $p$-value $<0.05$ (red line).

### 4.15. Niche interactors are enriched in CD150+ ${ }^{+}$WT cells but not their CD150 counterparts

As previously conducted for the TET2-deficient cell populations, we set out to compare CD150- WT against CD150+ WT cells to identify molecular pathways, governing HSC self-renewal. Following the computation of DEGs for LC-MS/MS and bulk RNA-seq profiles, we observed significantly fewer proteins enriched (50 proteins in total) in CD150+ WT HSCs (Figure 26A/B/D). Intriguingly, the correlation between LC-MS/MS and bulk RNA-seq was improved, when compared to the previous comparisons (CD150- TET2-- against CD150+ TET2--; CD150+ WT against CD150+ TET2--)
(Figure 26C). We also performed GO term and KEGG pathway enrichment analysis, which revealed that $15 \%$ of RNA-seq-derived and $60 \%$ of LC-MS/MS-derived GO terms were identified in both omics screens, while niche interaction functions (cell adhesion, cell-matrix adhesion, cell junction assembly) and developmental functions we ubiquitously enriched (Figure 27). In total, 14 CD150+ WT-enriched KEGG pathways were shared between the proteome and transcriptome (Figure 28). In summary, these data indicated consistent identification of core mechanisms and pathways across all comparative analyses, thus raising the likelihood of their involvement in HSC self-renewal regulation.


Figure 4.26. Comparison of proteome and transcriptome for CD150+ and CD150- WT cells. (A) Volcano plot of differentially expressed proteins between CD150 ${ }^{+}$and CD150 WT cells. Significance cut-offs: $p$-value $<0.05$ and $\log _{2} F C>0.5$ (CD150 enriched) | $\log _{2} F C<-0.5$ (CD150 enriched). (B) Volcano plot of DEGs between CD150+ and CD150 WT cells. Significance cut-offs: $p$-value $<0.05$ and $\log _{2} F C>0.5\left(C D 150^{+}\right.$enriched) $\mid \log _{2} F C<-0.5$ (CD150- enriched). (C) Correlation between proteomic and transcriptomic profiling. Corresponding $\log _{2}$ FC values between CD150 WT and CD150+ WT samples for genes identified in both datasets were utilised. (D) Venn diagram of genes enriched in CD150+ WT HSCs, as identified by LC-MS/MS (orange) and bulk RNA-seq (blue).


Figure 4.27. GO term enrichment of genes enriched in WT CD150+ HSCs. (A) Venn diagram of GO terms enriched in CD150 ${ }^{+}$WT cells, when compared to the CD150 ${ }^{-}$WT fraction. (B) GO terms consistently identified by RNA-seq and LC-MS/MS. Computed enrichment adjusted p-values (logtransformed) were plotted for the bulk RNA-seq (blue, left-hand side) and LC-MS/MS (red, right-hand side). Significance cut-off: p-value < 0.05 (red line). (C) Curated set of GO terms, uniquely enriched in the proteome (red, right-hand side) and bulk RNA-seq (blue, left-hand-side). Significance cut-off: pvalue < 0.05 (red line).


Figure 4.28. KEGG pathway enrichment of genes enriched in WT CD150+ HSCs. (A) Venn diagram of KEGG pathways enriched in CD150+ WT cells, when compared to the CD150- WT fraction. (B) KEGG pathways consistently identified by RNA-seq and LC-MS/MS. Computed enrichment adjusted p-values (log-transformed) were plotted for the bulk RNA-seq (blue, left-hand side) and LC-MS/MS (red, righthand side). Significance cut-off: p-value < 0.05 (red line). (C) Curated set of KEGG pathways, uniquely enriched in the proteome (red, right-hand side) and bulk RNA-seq (blue, left-hand-side). Significance cut-off: p-value < 0.05 (red line).

### 4.16. Fluid shear stress and ECM regulation are ubiquitously enriched in cell populations harbouring increased self-renewal potency

The multi-omic screen of cell types within a self-renewal continuum provides a unique opportunity to identify core biological functions driving HSC self-renewal. We hypothesised that core drivers of HSC self-renewal would be consistently identified across all DGE comparison conducted previously. Hence, we set out to identify biological functions and pathways consistently enriched in CD150 ${ }^{+}$cell fractions across the proteome, transcriptome and both genotypes.

We extracted GO terms and KEGG pathways that were enriched in CD150+ WT HSCs (LC-MS/MS and RNA-seq) and CD150+ TET2-/ HSCs (LC-MS/MS and RNA-seq),
computed based of DGE against their respective CD150- counterparts (Figure 29A/B). While we observed a significant range of GO terms (254 terms in total), the KEGG pathways analysis revealed a set of 3 ubiquitously enriched pathways (Figure 29B). These included fluid shear stress, focal adhesion and ECM-receptor interaction (Figure 29C).

A GO terms - CD150+ overlap



| Overlapping KEGG pathways |
| :--- |
| Focal adhesion |
| ECM-receptor interaction |
| Fluid shear stress and <br> atherosclerosis |

Figure 4.29. Overlap between WT-derived and TET2-/- -derived CD150+ HSC functions revealed enrichment of shear stress and ECM function. (A) Venn diagram, depicting the overlap between GO terms, uniformly enriched in LC-MS/MS and bulk RNA-seq of CD150 ${ }^{+}$cell fractions. This includes WT (comparison: CD150 $W T$ versus CD150 ${ }^{+}$WT) and TET2-/ (comparison: CD150- TET2 ${ }^{-/-}$versus CD150+ TET2 $^{-/-}$) HSCs. (B) The corresponding overlap of KEGG pathways for the identical comparison, outlined in (A). (C) Table of KEGG pathways identified in WT and TET2-deficient CD150+ HSC fractions.

The identification of ECM-receptor interactions and focal adhesion are in line with our previous observations in the TET2 proteome, highlighting the implication of ECM remodelling in HSC self-renewal. In contrast, the overrepresentation of fluid shear stress and atherosclerosis did appear in previous analyses. In the context of atherosclerosis, cellular responses to shear stress play a crucial role in atheroprotection and normal blood vessel physiology ${ }^{460}$. In turn, TET2 loss-of-function mutations have been implicated in atherosclerosis ${ }^{461}$. Despite inclusion of atherosclerosis, the enrichment of such KEGG pathway is indicative of a role of the
cellular response to shear stress in haematopoietic stem cells. The impact of shear stress on adult haematopoiesis has been recognised before ${ }^{462}$; however, the underlying molecular mechanisms and the role in HSC biology remains unknown. Thus, we set out to identify a set of shear stress-specific candidate proteins, possibly involved in HSC self-renewal. We extracted all genes and proteins associated with the shear stress KEGG pathway that were enriched in CD150+ WT and TET2-deficient HSCs (Table 1). The analysis revealed a total of 52 candidates (Table 1). Interestingly, only ITGA2B was consistently identified by all 4 comparisons. Overall, these data provide a novel set of candidate genes for the regulation of the cellular response to shear stress in HSC with increased self-renewal potency.

Table 6. Genes associated with fluid shear stress and enriched in WT and TET2deficient CD150+ HSCs, as identified by the proteome and transcriptome.

|  | WT CD150+ HSCs |  | TET2 ${ }^{-/}$CD150+ HSCs |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Protein | mRNA | Protein | mRNA |
| Actg1 |  | + |  |  |
| Acvr1 |  | + |  |  |
| Acvr2b |  | + |  |  |
| Akt3 |  | + |  |  |
| Arhgef2 |  | + |  |  |
| Ass 1 |  | + |  |  |
| Bcl2 |  | + |  | + |
| Bmpr2 |  | + |  |  |
| Cyba |  | + |  |  |
| Dusp1 |  | + |  |  |
| Fos |  | + |  |  |
| Gstm1 |  | + |  |  |
| Gstm2 |  | + |  | + |
| Gstm5 |  |  | + |  |
| Gstm7 |  | + |  |  |
| Gstp1 |  |  | + |  |
| Gstt1 |  | + |  |  |
| Gstt3 |  | + |  |  |
| Icam1 |  | + | + |  |
| lkbkg |  |  | + |  |
| II1b |  | + |  |  |
| Itga2b | + | + | + | + |
| Itgav |  | + |  |  |
| Itgb3 | + | + |  | + |
| Jun |  | + |  |  |


| Kdr |  | + |  | + |
| :---: | :---: | :---: | :---: | :---: |
| KIf2 |  | + |  |  |
| Map2k5 |  | + |  |  |
| Map3k5 |  | + |  |  |
| Mapk11 |  | + |  |  |
| Mapk12 |  | + |  |  |
| Mapk14 |  | + |  |  |
| Mapk7 |  | + |  |  |
| Mapk9 |  |  | + |  |
| Mef2a | + | + |  |  |
| Mef2c |  | + |  | + |
| Ncf2 |  | + |  |  |
| Nfkb1 |  | + |  |  |
| Pdgfb |  | + |  | + |
| Pecam1 |  | + |  |  |
| Pik3cd |  | + |  |  |
| Pik3r1 |  | + |  |  |
| Prkaa2 |  | + |  |  |
| Rac1 |  | + |  |  |
| Rac2 |  | + |  |  |
| Sdc2 |  | + |  |  |
| Sdc4 |  | + |  |  |
| Src |  | + |  | + |
| Thbd |  | + |  |  |
| Tnf |  | + |  |  |
| Tnfrsf1a |  | + |  |  |
| Vegfa |  | + |  |  |

### 4.17. Discussion

Here, we provide the first results of the small cell number LC-MS/MS protocol and demonstrate its utility in identifying novel molecular pathways underlying HSC fate choices. The ability to multiplex as few as 10,000 now provides experimental feasibility to dissect HSC heterogeneity and screen functionally distinct HSC populations in normal and malignant haematopoiesis. To provide a comprehensive landscape of HSC populations, only differing in their self-renewal potency, we set out to integrate proteomic data with transcriptomic profiles. Intriguingly, we observed no correlation between both -omes. Multiple reports identified that HSCs exhibit rapid rates of transcription but harbour a significantly lower rate of protein synthesis ${ }^{330,331,463}$. Such
discrepancy is a potential contributing factor to the absence of mRNA-protein correlation. Many molecular mechanisms of HSC fate choices remain poorly understood despite the abundance of transcriptomic data at single cell resolution. Such an observation, in conjunction with the discrepancy in transcription and protein synthesis, further highlights the importance of proteomic screening to complement our abundant understanding of HSC transcriptional landscapes.

In this work, we set out to integrate proteomic and transcriptomic profiles of HSC populations only differing in their self-renewal potency to uncover the underlying mechanism. This analysis has focused on utilising existing curated pathways databases and construction of protein interaction networks to provide a common set of biological functions associated with HSC self-renewal. In turn, this provided an intriguing subset of key target genes for potential HSC self-renewal regulation.

Using the continuum of CD150 ${ }^{+}$WT, ESLAM WT and CD150 ${ }^{+}$TET2-/, plus direct comparisons between CD150 $^{+}$vs CD150 cell fractions unanimously revealed enrichment of structural extracellular matrix proteins and ECM regulators in cell population with increasing self-renewal potency. Efforts to replicate the niche ex vivo remain in their infancy and our inability to produce homogenous ex vivo HSC selfrenewing cultures suggests that we are lacking complete understanding of selfrenewal regulation ${ }^{112,464,465}$. Here, we provide a set of candidate ECM proteins, including structural components, secreted regulatory factors, and integrins, that correlate with HSC self-renewal potency. In addition, the ex vivo expansion cultures of WT and TET2-deficient HSCs on 2D surfaces, functionalised with individual candidate structural proteins, did not show changes in HSC self-renewal rates, indicating that one-dimensional localisation of HSCs is not sufficient to alter HSC fate. Alternatively, the remodelling of the ECM landscape between WT and TET2-deficient HSCs could be reflective of a change in niche localisation. In turn, altered cytokine and niche signalling would induce different HSC self-renewal outcomes between WT and TET2 ${ }^{-/}$genotypes. Overall, such candidate genes provide further insight in the underlying mechanism for niche-mediated HSC self-renewal regulation and can be utilised to inform efforts in developing ex vivo 3D cultures, replicating the bone marrow niche.

In addition, the identification of ECM and niche proteins provides an intriguing avenue to uncover novel TET2 disease biology. Deletion of TET2 has been widely implicated in the formation of myeloproliferative neoplasms (MPN) and AML ${ }^{258,466}$, whereby the bone marrow microenvironment has been implicated in the progression of said diseases ${ }^{467-469}$. Interestingly, the loss of TET2 in mesenchymal stem cell has been shown to disrupt the bone marrow niche ${ }^{470}$. Furthermore, TET2 depletion altered the cellular composition of the bone marrow and induced secretion of pro-inflammatory cytokine ${ }^{471}$. Transplantation of TET2-deficient bone marrow cells into recipient mice, harbouring an AML microenvironment, revealed that such TET2-deficient cells exhibited MPN characteristics. Collectively, these data highlight the importance of the bone marrow microenvironment in disease progression and HSC deregulation. However, the molecular mechanisms that facilitate HSC:niche interaction and its deregulation during MPN and AML formation remain elusive. Here, we provide an intriguing set of ECM proteins, integrin, secreted regulatory enzymes and downstream effectors to help uncover the disease mechanism.

The discrepancy between the proteome and transcriptome indicates that structural ECM proteins, identified by LC-MS/MS, could stem from the surrounding bone marrow niche, rather than being actively synthesised by HSCs. Such observation further highlights the merit of proteomic analyses to provide an additional dimension otherwise not captured by RNA-seq. In addition, such phenomenon raises the importance of post-translational and post-transcriptional modifications. Posttranslational modifications directly regulate the localisation, folding, function and stability of proteins ${ }^{291-295}$. In contrast, post-transcriptional regulators, such as RNAbinding proteins, $\mathrm{N}^{6}$-methyladenosine, microRNAs, alter the expression of a given transcript by regulating its localisation, rate of translation splicing and beyond ${ }^{472-474}$. In brief, such modifications often form post-transcriptional RNA operons, which consist of RNA-binding proteins, microRNA and metabolites ${ }^{473}$. In turn, such operons provide a cellular mechanism for the dynamic regulation of gene expression. Pioneering work work by Abdelmohsen et al. showed that phosphorylation of the RNA-binding protein HuR altered the expression and stability of SIRT1 transcripts ${ }^{475}$. Collectively, Abdelmohsen and others demonstrated effective regulation of gene expression by modulating RNA operation via phosphorylation, downstream of signalling cascades ${ }^{475-477}$.

The role of intracellular $\mathrm{Ca}^{2+}$ in mediating HSC self-renewal and maintenance have previously been described ${ }^{403}$ where the authors showed that low cytoplasmic [ $\mathrm{Ca}^{2+}$ ] improves in vitro HSC maintenance and that HSCs express multiple $\mathrm{Ca}^{2+}$ efflux pumps to actively maintain low intracellular $\left[\mathrm{Ca}^{2+}\right]$. In turn, low $\left[\mathrm{Ca}^{2+}\right]$ inhibit calpains to prevent the inhibition of TET2-mediated HSC maintenance. By utilising a $T^{-1} 2^{-1}$ mouse model, the authors also provided evidence that TET2 deficiency desensitises HSCs to the phenotypic effect of low $\mathrm{Ca}^{2+}$ conditions. While these data provide evidence for the importance of $\mathrm{Ca}^{2+}$ in HSC self-renewal, the underlying mechanism of $\mathrm{Ca}^{2+}$ sensing and particularly the molecular impact of TET2-deficiency remain poorly described. Here, we provide a set of novel $\mathrm{Ca}^{2+}$ regulators and their deregulation upon TET2 knockout.

The 10-plex of 5 ST- and LT-HSC populations allowed us to perform multiple independent analyses, such as CD150+ against CD150- cell fractions of each genotype, to identify novel self-renewal regulators. By overlaying biological pathways, identified across all comparison and omes, we observed consistent enrichment of fluid shear stress in cell fractions with self-renewal capacity. Mechanosensing in response to fluid shear stress has been shown to regulate differentiation throughout adult haematopoiesis and differentiation of embryonic stem cells into the haematopoietic lineage ${ }^{462,478}$. Nevertheless, the role of mechanotransduction and fluid shear forces in HSC biology remain elusive, with underlying molecular mechanisms remaining undescribed. In addition, we also observed enrichment of cytoskeletal proteins. Interestingly, cytoskeletal remodelling has previously been implicated in the cellular response to fluid shear stress ${ }^{479}$. The set of cytoskeletal proteins and candidate gene list of shear stress response factors provide a promising tool for uncovering the molecular mechanisms of shear stress responses and their role in HSC self-renewal. For instance, we observed enrichment of MEF2 and KLF2 in cell types with selfrenewal potency. MEF2 activity has been shown to be upregulated in endothelial cells subjected to fluid shear stress. In turn, the transcription factor MEF2 increases KLF2 expression, a potent regulator of self-renewal in embryonic stem cells ${ }^{480-482}$. Collectively, these data highlight that our candidate gene list provides a powerful tool for hypothesis generation to define the role of mechanosensation of bone marrow niche shear forces and their role in HSC self-renewal regulation.

## Chapter 5. Identifying drivers of HSC self-renewal and repopulation potency in ex vivo expansion

Pioneering studies over the past two decades have revealed a plethora of cell surface markers for the identification of phenotypic HSCs with self-renewal capacity (extensively discussed in Chapter 1.5). Various combinations of such markers are routinely applied for the isolation of fresh bone marrow HSCs, yielding up to $\sim 50-60 \%$ purity ${ }^{332}$. Despite such progress, these immunophenotyping strategies do not support the isolation of a homogeneous and fully functional HSC population, leaving the underlying mechanism of key properties such as HSC self-renewal not fully characterised. The scarcity of primary HSCs limits the range of experimental tools that can be applied for uncovering such molecular mechanisms, thus further contributing to the difficulty of deciphering HSC self-renewal. Beside the needs for uncovering the underlying biology of HSCs, reliable ex vivo expansion protocols are of paramount importance in gene therapy and bone marrow transplantations. In brief, gene therapies for acquired and inherited genetic disorders often include ex vivo editing of CD34+ cells, which requires subsequent expansion prior to patient administration ${ }^{483}$.

Limited supply of HSCs raised the need for a reliable culturing system to facilitate ex vivo expansion of functional HSCs, retaining self-renewal potency. In 2019, Wilkinson and colleagues pioneered a serum-free culturing system, allowing $>200$-fold expansion of murine HSCs over a period of 1 month ${ }^{225}$. Most intriguingly, this breakthrough culturing system could be used for single cell expansion to enable clonal expansion of phenotypic HSCs. However, similar to freshly isolated HSC ${ }^{387,484}$, ex vivo expanded phenotypic HSCs expressed significant functional heterogeneity in subsequent transplantation assays ${ }^{225}$. With approximately 1 in 34 expanded cells retaining HSC function, the authors demonstrated that the novel culture system (termed "F12 cultures" hereafter) also supported the expansion and differentiation of various progenitor cell types, providing a complex cellular composition throughout expansion. Furthermore, ex vivo expansion has been shown to alter the immunophenotype of mouse and human HSCs ${ }^{485,486}$, thus rendering flow cytometric strategies for the purification of functional HSCs from culture largely ineffective. To date, EPCR has been identified as the most promising in vitro marker of expanded
human CD34 ${ }^{+}$cord blood stem cells ${ }^{486}$. Hence, culture-specific HSC purification strategies remain large unknown, requiring systematic profiling of immunophenotypes and molecular pathways of expanded HSCs.

The aim of this chapter was to first identify a reliable purification strategy for ex vivo expanded HSCs and to then undertake molecular characterisation of the clonal heterogeneity of cultured HSCs. Linking cell function with the molecular profile thus enables the identification of molecular drivers of HSC self-renewal potency and derive a strategy for the prospective isolation of functional HSCs in ex vivo expansion cultures.

### 5.1. EPCR and Fgd5 mark functional HSCs in ex vivo expansion cultures

To identify robust markers of expanded HSCs, we explored the utility of previously identified HSC-specific cell surface markers and HSC reporter models. The discovery of EPCR as a highly specific marker for HSCs in vivo greatly improved the purity of freshly isolated bone marrow $\mathrm{HSCs}^{35,332}$ and has also been identified as a potent in vitro marker of human $\mathrm{HSCs}^{486}$. In parallel, a plethora of HSC reporters have been developed ${ }^{300,487-492}$. For this thesis, we primarily explored the Fgd5 mouse where Gazit et al. identified the selective expression of $F g d 5$ in adult BM HSCs ${ }^{488}$ and deveoped Fgd5zsGreen•ZsGreen/+ and Fgd5 ${ }^{\text {mCherry/+ }}$ reporter strains to demonstrate that BM-derived mCherry ${ }^{+}$cells express common HSC cell surface markers, including EPCR, and retain repopulation capacity in transplantation assays ${ }^{488}$. Notably, Rabe et al. confirmed the co-expression of Fgd5 and EPCR in primary HSCs in vivo ${ }^{493}$, including during periods of inflammatory stress.

To identify putative markers of ex vivo expanded HSCs, Fgd5 and EPCR were explored to determine whether they would mark functional HSCs following expansion using multiple culturing conditions. To test the utility of Fgd5-ZsGreen ${ }^{+}$as a marker of actively cycling HSCs, ESLAM cells (CD48-CD150 ${ }^{+}$CD45 $\left.{ }^{+} E P C R^{+}\right)^{35,50}$ were isolated from BM and FL tissues (Figure 1A) and Fgd5 co-expression was assessed. Both BM and FL ESLAM HSCs (the latter being an actively cycling HSC population) ubiquitously expressed Fgd5, suggesting that the EPCR ${ }^{+}$Fgd5 ${ }^{+}$phenotype might serve as a
promising strategy for marking ex vivo expanded, cycling HSCs (Figure 1B). To determine the merit of the Fgd5-ZsGreen ${ }^{+} \mathrm{EPCR}^{+}\left(\mathrm{F}^{+} \mathrm{E}^{+}\right)$stategy, a comprehensive set of single cell-derived HSC clones were expanded for 10 days and assessed for their $\mathrm{F}^{+} \mathrm{E}^{+}$phenotype alongside other markers that can enrich for HSCs in culture (LinSca1+cKit ${ }^{+}$, LSK). When comparing the proportions of $\mathrm{F}^{+} \mathrm{E}^{+}$and LSK cells within the same clone, a signfincant correlation between $\mathrm{F}^{+} \mathrm{E}^{+}$and LSK became evident (Figure 1C). Nevertheless, a subset of \%LSK-high clones did not express high $\% \mathrm{~F}^{+} \mathrm{E}^{+}$, while all clones with high \% $\mathrm{F}^{+} \mathrm{E}^{+}$also expressed high \%LSK (Figure 1C). This suggests that $\mathrm{F}^{+} \mathrm{E}^{+}$subfractionates the cells and might provide a more acurate strategy for isolating functional HSCs that retain repopulation and self-renewal potency. Furthermore, additional 10-day cultures were screened for proliferation and differentiation using alternative culturing conditions (described in ${ }^{339}$ ) and again, the $\% \mathrm{~F}^{+} \mathrm{E}^{+}$correlated with the \%LSK in smaller, less differentiated clones, thus according with our previous observation (Figure 1D).


Figure 5.1. EPCR and Fgd5-ZsGreen allow identification of cycling HSCs. (A) Flow cytometry plots, outlining the sorting strategy for Fgd5 ${ }^{+} E P C R^{+} H S C s\left(C D 48-C D 150^{+} C D 45^{+} E P C R^{+} F g d 5^{+}\right)$for adult bone marrow (BM) and foetal liver (FL) tissues. (B) Proportion of ESLAM (CD48-CD150 $C D 45^{+} E P C R^{+}$) with positive Fgd5 phenotype. (C) Single cell-derived 10-day expansion clones ( $n=2$ ), displaying the Pearson correlation between the LSK proportion of total cells (\%LSK) and Fgd5 ${ }^{+} E P C R^{+}$proportion (\%FE) of total cells from their respective clones. (D) 10-day expansion cultures ${ }^{339}$. Gating strategy for cultured cells for small, medium and large colonies. The corresponding proportion of LSK cells of total live cells is shown for $\left(F^{l o} E^{l o}\right),\left(F^{h i} E^{h i}\right)$ and all cultured cells. One-way ANOVA was computed: ${ }^{*}=<0.05 ;{ }^{* *}=<0.01$.

Next, transplantation assays were undertaken to determine the functional capacity of ex vivo expanded $\mathrm{F}^{+} \mathrm{E}^{+}$cells. Fgd5-ZsGreen ${ }^{+} E P C R^{+} C D 45^{+}$cells were sorted and cultured in previously described conditions (Stemspan $+300 \mathrm{ng} / \mathrm{ml}$ SCF $+20 \mathrm{ng} / \mathrm{ml}$ IL$11)^{337}$ for 3 days, followed by FACS to isolate Fgd5 ${ }^{\text {hi }} E P C R^{\text {hi }}\left(F^{\text {hi }} E^{\text {hi }}\right)$, Fgd5 ${ }^{\text {lo }} E P C R^{\text {lo }}$ ( $F^{10} \mathrm{E}^{10}$ ) and a proportion of total live cells for transplantation into irradiated mice (Figure 2A). Following 16-week transplantation assays, $\mathrm{F}^{\text {hi }} \mathrm{E}^{\text {hi }}$ cells displayed significantly higher levels of donor chimerism and were able to contribute to multi-lineage reconstitution (Figure 2B/C). These results indicate that Fgd5 and EPCR expression is selectively retained in vitro by functional HSCs with long-term repopulation capacity.

As previously outlined in section 1.6, the F12 culture conditions which expand HSCs substantially over 28 days are subject to significant clonal heterogeneity, both within and between clones. Hence, we set out to determine whether Fgd5 and EPCR could provide a viable strategy for prospective isolation of functional HSCs following longterm culture. We performed 28-day HSC expansion cultures from which $90 \%$ were subject to transplantation assays and the remaining $10 \%$ were analysed by flow cytometry to determine the immunophenotype of individual clones. As expected, transplanted clones exhibited significant heterogeneity in donor chimerism and multilineage contribution (Figure 2D/E). We then compared the correlation of donor chimerism with the proportion of $\mathrm{F}^{+} \mathrm{E}^{+}$or $\mathrm{F}^{+} \mathrm{E}^{+} \mathrm{LSK}$ cells in the respective clones (Figure $2 F / G)$. Donor chimerism was significantly correlated with the proportion of $\mathrm{F}^{+} \mathrm{E}^{+}$cells within individual clones ( $r^{2}=0.67$ ) (Figure 2F). The addition of traditional ex vivo HSC markers (LSK) further improved the correlation with donor chimerism (Figure 2G). However, the addition of the LSK cell surface phenotype only moderately improved the correlation, suggesting that the two-factor $\mathrm{F}^{+} \mathrm{E}^{+}$strategy provided a reliable strategy on its own to prospectively identify ex vivo expanded clones that retained functional HSCs.


Figure 5.2. EPCR and Fgd5-ZsGreen mark ex vivo expanded functional HSCs. (A) Schematic for the experimental design of transplantation assays. $F g d 5^{+} E P C R^{+} C D 45^{+}$cells were cultured for 3 days and split by FACS into I) Fgd5 ${ }^{\text {ow }} E P C R^{\text {low }}\left(F^{\circ} E^{\prime}\right)$ ), II) Fgd5 ${ }^{\text {high }} E P C R^{\text {high }}\left(F^{\text {hi }} E^{\text {hi }}\right)$ or III) live cells. Sorted cells were subject to transplantation assays. (B) Proportion of donor chimerism in primary recipients following transplantation of ex vivo expanded HSC cultures, including I) Fgd5 ${ }^{\text {hi }} E P C R^{\text {hi }}\left(F^{\text {hi }} E^{h i}\right)$ cell portions ( $n=3$ ), II) Fgd5 ${ }^{\circ} E P C R^{\prime o}\left(F^{\circ} E^{\prime o}\right)$ cell portions $(n=3)$ and III) bulk live cultures $(n=2)$. (C) Lineage output of transplantation assays in $(B)$, displaying the proportion of $T$ cells ( $T$ ), B cells (B) and granulocytes/macrophages (GM) of total donor cells 16 weeks post-transplantation. (D) Donor Chimerism of HSC-derived clones that underwent 28-day expansion using the F12 culturing strategy. (E) Lineage output of clones with >1\% donor chimerism from F12 cultures outlined in (D). (F) Correlation between donor chimerism and the proportion of $\mathrm{Fgd5}^{+} E P C R^{+}(\% F E)$ cells within clonally expanded 28 day HSC cultures. (G) Correlation between donor chimerism and the proportion of Fgd5 ${ }^{+} E P C R^{+} L S K$ (\%FELSK) cells within clonally expanded 28-day HSC cultures. Significance levels for all panel figures: * $=<0.05$; $^{* *}=<0.01$; $^{* * *}=<0.001$; ${ }^{* * * *}=<0.0001$.

### 5.2. Deciphering clonal heterogeneity of long-term HSC expansion cultures

Characterising the underlying causes of the heterogeneity observed in clonal HSC expansion cultures requires both functional and molecular characterisation. Previous evidence has suggested that secreted factors and the cellular composition of
differentiated, non-HSC cell types in clonal cultures might influence the self-renewal potency and repopulation capacity of ex vivo expanded HSCs ${ }^{225,494,495}$. Hence, an experimental strategy was devised to simultaneously interrogate the functional output and molecular profiles of phenotypic HSCs and non-HSCs of individual 28-day expanded clones. First, single ESLAM HSCs were sorted and maintained in culture for 28 days, using the F12 culture conditions (Figure 3A). At the end of the culturing period, 20 single cell-derived clones were re-sorted into phenotypic HSCs [EPCR+LinSca1+cKit+ (ELSK)] and the remaining non-HSCs were pooled. This included Linㄹ, $\mathrm{Lin}^{-}$ cKit-Sca1- and Lin ${ }^{-}$Kit $^{+}$Sca1+EPCR', collectively termed the "nonELSK" cell fraction (Figure 3B). 50\% of each cell fraction was subjected to transplantation assays using irradiated recipients, while the remaining $50 \%$ of cells had gene expression profiling performed using RNA-sequencing (Figure 3A). Prior to re-sorting, 10\% of each clone was removed to assess the immunophenotype of each close by flow cytometry. Of note, we opted to utilise the ELSK gating strategy as opposed to the FELSK strategy for isolating functional HSCs from ex vivo cultures. Firstly, the reliance on a genetically engineered reporter strain greatly limits the scope and utility of the ex vivo functional HSC marker strategy in a broad range of experimental designs, particularly in the study of disease models. Secondly, the inclusion of LSK markers and close correlation of Fgd5 and EPCR in expanded phenotypic HSCs permits the exclusion of Fgd5 in the experimental setup.


Figure 5.3. Approach to inspect clonal Heterogeneity of clonally expanded HSCs. A) Schematic outlining the experimental setup for screening the cellular output and heterogeneity of single cell-derived HSC expansion cultures. Single HSCs were isolated and maintained in culture for 28 days. 20 clones were harvested and phenotypic HSCs [EPCR ${ }^{+}$Lin ${ }^{-S c a-1+}{ }^{+}$-kit (ELSK)] were isolated from the remaining nonELSK cells. $50 \%$ of each cell fraction was subject to transplantation assays and the remaining $50 \%$ were used for gene expression profiling (bulk RNA-seq). (B) Representative sorting strategy for ELSK and nonELSK cell fractions. The non-ELSK fraction contains all pooled fraction (labelled 1-3).

As expected, primary transplantation assays of the 20 single HSC-derived clones yielded a heterogeneous functional output, with the ELSK cell fraction from 8 out of 20 clones (40\%) able to achieve donor chimerism $>20 \%$ (Figure 4A). Similarly, multilineage reconstitution was limited to the same set of 8 clones (Figure 4B). The remaining clones lacked sufficient donor chimerism or were not able to contribute to the reconstitution of the granulocyte-monocyte (GM) lineage (Figure 4A/B). These observations were in line with previous studies using the F12 culture conditions (Wilkinson et al 2019). Next, we inspected the correlation of the transplantation outcomes with the immunophenotype of the respective clones. Interestingly, the total number of cells ( $r^{2}=0.01$ ) and the absolute number of phenotypic HSCs ( $r^{2}=0.19$ ), as defined by the ELSK phenotype, did not correlate with donor chimerism of the clone (Figure 4C/D). Instead, the proportion of ELSK (\%ELSK) within a clone was significantly correlated with both donor chimerism ( $r^{2}=0.83$ ) and GM lineage contribution ( $r^{2}=0.85$ ), indicative of the capacity for long-term multilineage reconstitution ${ }^{496}$ (Figure 4E/F). These results indicate that rapid ex vivo proliferation is not associated with the expansion of functional HSCs. Furthermore, the importance of
\%ELSK for repopulation indicates that the content and level of differentiation in ex vivo cultures is critical for HSCs to retain self-renewal potency.


Figure 5.4. Proportion of phenotypic HSCs in heterogeneous clonal cultures. (A) Donor chimerism (represented as percentile proportion) for ELSK cell fractions (dose: 45-50\%) throughout primary transplantation assays. (B) The proportion of donor cells attributed to $T$ cell ( $T$ ), B cell (B) and granulocyte-macrophage (GM) lineages 16 weeks post-transplantation. Displaying clones with $>1 \%$ donor chimerism. (C) Pearson correlation between the donor chimerism and total number of live cells in the respective single HSC-derived 28-day clones. (D) Pearson correlation between donor chimerism and the total number of ELSK cells in respective clones. (E) Pearson correlation between donor chimerism and the proportion of ELSK cells in clones. (F) Pearson correlation between GM contribution and the proportion of ELSK cells in clones. Significance levels for all panel figures: $n s=>0.05$; * $=$ $<0.05 ;{ }^{* *}=<0.01 ;{ }^{* * *}=<0.001 ;{ }^{* * * *}=<0.0001$.

To test whether the Fgd5 phenotype is critical to the identification of functional HSCs from ex vivo cultures, we determined the correlation of donor chimerism and the contribution to the GM lineage with the proportion of Fgd5+ELSK (\%FELSK) in the
corresponding clones. As expected, \%FELSK was significantly correlated with donor chimerism ( $r^{2}=0.86$ ) and GM contribution ( $r^{2}=0.88$ ), indicating that Fgd5 provides limited improvement when compared to the ELSK phenotypes on its own ( $r^{2}=0.83$ ) (Figure 5A/B). Interestingly, donor chimerism ( $r^{2}=0.65$ ) and GM contribution ( $r^{2}=$ 0.62 ) were also significantly correlated with \%F+E+ (Figure 5C/D). These results indicate that the Fgd5 and EPCR molecules are closely correlated and are both able to identify functional HSCs in expansion cultures. Overall, these data show that the ELSK immunophenotyping strategy reliably marks functional HSCs that are undergoing ex vivo expansion. Consequently, this strategy provides a tool to prospectively select single HSC-derived clones containing a high proportion of HSCs with self-renewal potency.


Figure 5.5. FELSK correlates with the repopulation outcome of ex vivo expanded HSCs. (A-B) Correlation of donor chimerism or contribution to the GM lineage with the proportion of Fgd5 ${ }^{+} E P C R^{+}$LinSca1+ ${ }^{+}$Kit ${ }^{+}$(\%FELSK) following 28-day expansion. (C-D) Correlation of donor chimerism or contribution to the GM lineage with the proportion of Fgd5 ${ }^{+} E P C R^{+}$(\%FE+) cell. Pearson correlations displayed, significance levels: $n s=>0.05 ;{ }^{*}=<0.05 ;{ }^{* *}=<0.01 ;^{* * *}=<0.001 ;{ }^{* * * *}=<0.0001$.

### 5.3. Repopulating clones have distinct gene expression profiles

As previously outlined, $50 \%$ of ELSK and $50 \%$ nonELSK cells were isolated from the 20 clones that underwent primary transplantation assays. To characterise the underlying molecular profiles governing self-renewal potency and repopulation of HSCs following ex vivo expansion, bulk RNA sequencing was performed (RNA-seq) on a selected subset of these clones. We selected 6 repopulating and 6 nonrepopulating clones, defined by $>1 \%$ donor chimerism and $>1 \% \mathrm{GM}$ contribution. We classified all samples into four cell types: I) ELSKs from repopulating clones (PosELSK); II) nonELSKs from repopulating clones (PosNonELSK); III) ELSKs from non-repopulating clones (NegELSK); IV) nonELSKs from non-repopulating clones (NegNonELSK) (Figure 6A).

As the data was acquired over multiple batches, we included technical replicates of two identical samples per sequencing batch to account for any possible batch effects. Multidimensional scaling (MDS) plots of all samples revealed disjointed clustering of technical replicates (Figure 6B). Consequently, we corrected for the batch effect using the technical replicates (Figure 6C). Following normalisation and batch correction, the MDS plot revealed that ELSK and nonELSK clustered separately. Interestingly, ELSK cell fractions from repopulating and non-repopulating clones also clustered separately (Figure 6C). The same trend was also evident for nonELSK cell fraction. Surprisingly, we also observed an overlap between some NegELSK and PosNonELSK fractions (Figure 6C). Such similarity indicates a likely similar cellular composition of the ELSK fraction of non-repopulating clones and the nonELSK fraction of repopulating clones. In turn, this raises the question whether clones that lost repopulation potential are comprised of cell types further downstream in the haematopoietic hierarchy, while repopulating clones contain a higher proportion of naïve cell types.


Figure 5.6. Batch correction and pre-processing of bulk RNA-seq data. (A) Schematic and nomenclature for samples used for bulk RNA-seq analysis. Positive repopulation was defined as $>1 \%$ donor chimerism and $>1 \%$ contribution to GM. ELSK and nonELSK fractions represent $50 \%$ of clones, used for primary transplantation studies (Figure 2/3). Consistent colour scheme for all following figures. (B) Raw multidimensional scaling (MDS) plot for acquired data in three separate batches. Unique sample IDs and technical replicates of selected samples for batch correction indicated in the right-hand plot. (C) MDS plot following correction for batch effects.

### 5.4. Non-repopulating clones are largely comprised of differentiated progenitors

Next, in order to assess which cell types were overrepresented in single cell-derived HSC clones, the clonal composition of successful and unsuccessful expansion cultures were compared. Since bulk RNA-seq libraries for pooled cell fractions were prepared, the acquired expression profiles were comprised of an average of all cell types captured within the fraction. To deconvolute the averaged RNA-seq profiles and identify dominant gene signatures, a previously developed tool called SingleR was utilised ${ }^{349}$. SingleR was originally developed to identify and assign cell identities, based on gene signatures extracted from single cell RNA-seq (scRNA-seq) profiles. In brief, SingleR derives correlation scores between query scRNA-seq profiles and a reference gene expression set, based on bulk transcriptomes of pure populations of interest. We applied the SingleR algorithm to query the correlation of our bulk transcriptomes against the comprehensive and manually curated Immunological Genome Project (ImmGen) database, which is comprised of microarray and next generation sequencing (NGS) data for 253 purified haematopoietic cell types ${ }^{497}$.

We reasoned that dominant cell types within each cell fraction would lead to a proportional overrepresentation of the associated gene signatures in the bulk RNAseq data. The correlation analysis against all 253 ImmGen cell types revealed that all ELSK and nonELSK fractions were predominantly correlated with stem and progenitor cell types (data not shown; Figure 7). These results indicate that, independent of their repopulation potential, single cell-derived HSC expansion clones are largely comprised of immature cell types. These results are also in line with previous observations made by Wilkinson and colleagues, outlining that F12 cultures did not support the expansion and survival of terminally differentiated cell types ${ }^{225,226}$.


Figure 5.7. Repopulating clones are enriched for LT-HSC gene signatures. Heatmap outlining the correlation of gene expression profiles for each sample with single-cell RNA-seq-derived profiles of haematopoietic stem and progenitor cell populations, curated as part of the Immgen database. Sample IDs correspond to Figure 6A. Cell types are arranged according to hierarchical clustering.

Next, we extracted all stem and progenitor cell types from the ImmGen database and computed the correlation with our bulk RNA-seq data. We first inspected PosELSK and NegELSK cell fractions. Intriguingly, PosELSK cell fractions were more closely correlated with long-term HSCs (LT-HSCs), while NegELSKs most closely resembled short-term HSCs (ST-HSCs; CD34+FIk2-LSK phenotype) (Figure 7). Of note, PosELSK were also enriched for ST-HSC gene signatures. These results corroborated the functional screen, outlining that ELSK cell fractions of non-repopulating clones only retained ST-HSCs and lacked sustained self-renewal potency. The inspection of NonELSK cell fractions revealed that non-repopulating clones were uniformly enriched for multipotent lymphoid progenitors (MLP), defined by Lineage-CD19-lgM ${ }^{-}$ CD43 ${ }^{+}$CD24-AA4.1+CD45RA${ }^{+}$CD117+IL7R${ }^{+}$) (Figure 7). In contrast, NonELSK fractions from repopulating clones were predominantly enriched for ST-HSCs and only two out of six fractions were enriched for MLP gene signatures. Interestingly, the four samples enriched for ST-HSC signatures correspond to the four samples that clustered closely with NegELSK and in proximity of the PosELSK cluster (Figure 6C). Overall, these results indicate that ex vivo expanded HSCs with sustained self-renewal potency retain molecular similarity to freshly isolated HSCs, while non-repopulating clones have lost long-term self-renewal capacity and express increased differentiation toward lymphoid progenitors.

### 5.5. Ex vivo expanded ELSKs retain the core HSC molecular profiles

Following the unbiased correlation analysis of all samples against purified reference cell types, we set out to characterise the molecular differences between ELSK and nonELSK cells following ex vivo expansion. We first computed the differential gene expression (DGE) between all ELSK and nonELSK fractions, pooling repopulating and non-repopulating clones. This revealed a set of 1158 genes overrepresented in ELSKs and 1250 genes overrepresented in nonELSKs ( $\log _{2} \mathrm{FC} \geq 1$ and $p$-value $\leq 0.05$ ) (Figure 8).


Figure 5.8. Differential gene expression between ELSKs and nonELSKs. Differential gene expression between ELSK and nonELSK fractions. Repopulating and non-repopulating clones were pooled for the analysis. NegELSKs (cut-offs: $\log _{2} F C=1 ; p$-value $=0.05$ ).

Next, gene ontology (GO) term enrichment analysis was performed to identify key molecular processes enriched in both cell fractions. In line with the previous observation that ELSKs from repopulating and non-repopulating clones predominantly express LT-HSC and ST-HSC gene signatures, we observed enrichment of developmental and cellular organisation processes (Figure 9A). Interestingly, we also observed ELSK-specific enrichment of metabolic pathways, governing the synthesis of precursors and intermediates throughout glycolysis (GO terms: I) phosphorus metabolic process; II) phosphate-containing compound metabolic process) (Figure 9A). A multitude of studies have now discussed the role of cellular metabolism in regulating HSC fate and it has been shown that adult HSCs predominantly utilise
glycolysis within their native hypoxic BM niche to sustain quiescence and support HSC maintenance ${ }^{498-500}$. Subsequently, Vannini et al. also demonstrated a causal relationship between glycolysis and HSC self-renewal ${ }^{501}$. In brief, the authors demonstrated that low mitochondrial activity and thus reliance of glycolysis was essential for HSCs to retain self-renewal capacity ex vivo. In contrast, differentiation induces a rapid shift towards mitochondrial oxidative phosphorylation amongst progenitors and terminally differentiated cells ${ }^{502}$. Intriguingly, FL HSCs, which extensively proliferate and undertake symmetric self-renewal divisions, were shown to utilise oxidative phosphorylation. While our ex vivo cultured HSCs, captured within the ELSK fraction, display phenotypic characteristics of rapidly dividing FL HSCs, the enrichment for glycolytic intermediates indicates that glycolytic metabolism remains relevant for expanded functional HSCs. However, GO term enrichment analysis is subject to a multitude of well-described sampling and annotation biases ${ }^{503}$, thus limiting the above observation to a tentative indication. Thorough cell cycle analysis, hypoxic screening and metabolic profiling of ex vivo expansion cultures using the F12 strategy are required to address the question. In contrast to ELSKs, nonELSKs were enriched for a broad range of immune response and immune regulation GO terms (Figure 9B). Furthermore, nonELSK were enriched for inflammatory terms and cell activation in response to ligands or other activating factors (Figure 9B). These observations corroborate the previous findings that nonELSK cells are comprised of progenitor cell types (Figure 7).


Figure 5.9. Go term enrichment of ELSK and nonELSK fractions. Gene Ontology (GO) term enrichment analysis computed using differentially expressed genes from Figure 8. Top 40 biological process terms are displayed for the ELSK fractions $(A)$ and nonELSK fractions $(B)$.

Following GO term enrichment, we performed a pathway analysis using differentially expressed genes (DEGs) between ELSK and nonELSK fractions. For this purpose, we utilised the curated Reactome database ${ }^{504}$. The analysis revealed the enrichment of an interconnected cluster of Rho GTPase signalling and cell cycle regulation in ELSKs (Figure 10A). In contrast, nonELSK were enriched for ribosomal activity pathways and the response to cytokine signalling (Figure 10B). Rho GTPases have been widely implicated in HSC biology (extensively reviewed in ${ }^{505}$ ). Most prominently, Cdc42 has been shown to regulate HSC quiescence and self-renewal by modulating cell cycle progression of $\mathrm{HSCs}^{506}$. In absence of Cdc42, HSCs underwent extensive proliferation and differentiation towards ST-HSCs. Furthermore, Gu et al. demonstrated a critical role of Rac1 and Rac2 signalling in regulating HSC proliferation and HSC engraftment in transplantation assays ${ }^{507}$. The enrichment of Rho GTPases and cell cycle regulation in ELSKs further indicates that ex vivo expansion using the F12 culturing strategy retains phenotypic HSCs. The enrichment of a strong ribosomal cluster is in line with the increased proliferative phenotype of downstream progenitor populations (Figure 10B). Intriguingly, we also observed enrichment of Interleukin-3, Interleukin-5, GM-CSF and SCF-KIT signalling (Figure 10B). Previous studies described key roles of IL-3, IL-5 and GM-CSF in a broad range of progenitor and mature cell types ${ }^{508,509}$. Similarly, SCF and KIT-mediated signalling pathways have been identified as key proliferative drivers throughout haematopoiesis ${ }^{510,511}$.

Overall, these data indicate that a broad range of core molecular pathways governing self-renewal and maintenance of freshly isolated, functional HSCs were retained throughout ex vivo expansion as well.


Figure 5.10. Pathway analysis of ELSK and nonELSK fractions. Reactome pathway enrichment analysis for ELSKs (A) and nonELSK (B). A p-value cut-off of 0.05 was selected. The size of nodes represented the number of pathway-associated genes enriched in the cell fraction. Node colour indicates the adjusted p-value. Functionally related pathways are connected.

### 5.6. Non-repopulating clones lose expression of HSC-specific genes

Following a global comparison of ELSK against nonELSK fractions and identification of dominant cell types within each sample, we set out to identify molecular differences between repopulating and non-repopulating clones. For this purpose, we first computed DGE between PosELSKs and NegELSKs. Interestingly, only a small subset of differentially expressed genes was statistically significant (Figure 11A). To obtain an overview of gene function, we performed a GO term enrichment analysis, which revealed that PosELSK-enriched genes were involved in development and niche interaction (namely integrin binding, cell surface receptor signalling pathway and cell adhesion) (Figure 11B). In contrast, NegELSKs were enriched for differentiationassociated terms, in particular for the myeloid lineage (Figure 11B). Despite the relatively small number of genes, ELSK from non-repopulating clones appeared to be primed for differentiation.


Figure 5.11. Gene expression profiling of PosELSKs and NegELSKs. (A) Differential gene expression (DGE) volcano plot of PosELSKs and NegELSKs (cut-offs: $\log _{2} F C=1$; $p$-value $=0.05$ ). (B) Selected gene ontology (GO) terms, enriched in PosELSK and NegELSK cell fractions ( $p$-value < 0.05). (C) Heatmap depicting expression of differentially expressed genes (DEGs) of PosELSK against NegELSKs. Unique clone IDs are referenced to Figure 6A.

Next, we inspected individual DEGs to identify key drivers of HSCs, retaining HSC function post-expansion. Firstly, expression of von Willebrand Factor (Vwf), a well described marker of HSCs ${ }^{35}$, was enriched in PosELSKs (Figure 11C). Similarly, Plxnd2, Tcf7l1 and Lmna, were each enriched in PosELSKs, and have been shown to play a critical role in maintaining functional properties of HSCs ${ }^{512-514}$. Serpinb6 has previously been identified to be selectively expressed in HSCs with long term selfrenewal capacity, but its function remains unknown ${ }^{515}$. The fibronectin and osteopontin-binding integrin ltgb5 has previously been identified as part of a quiescence gene signature for hematopoietic cells ${ }^{516-519}$. In addition, elegant work by Sagar et al. outlined a critical role of fibronectin for the retention of self-renewal and
functional capacity of ex vivo expanded HSCs ${ }^{106}$. Hence, enrichment of Itgb5 in PosELSKs might provide a promising marker for the identification of functional HSCs following expansion.

On the other side, NegELSK were enriched for Cebpa, Siglecf and Rab44 (Figure 11C). Cebpa and Siglecf have been widely implicated in hematopoietic differentiation and progenitor function ${ }^{520,521}$. Rab44 has been particularly implicated in myeloid differentiation ${ }^{522}$. Together these data support the notion of increased expression of differentiation marker genes within the non-repopulating ELSK fraction.

In summary, the differential gene expression analysis yielded a narrow set of genes that distinguish functional HSCs from non-functional phenotypic HSCs. The majority of genes enriched in functional HSCs (PosELSK fraction) are part of a broad set of genes with well-described functions in fresh HSCs. The above comparison thus provides an intriguing subset of such HSC-specific genes with key function for ex vivo expanded functional HSCs.

### 5.7. Erythrocytes and Megakaryocytes are enriched in non-repopulating clones

The molecular analysis of ELSK and nonELSK cell fractions revealed enrichment of both lymphoid and myeloid pathways. A comparative analysis for each sample suggested an enrichment of lymphoid progenitors amongst nonELSKs, with the repopulating clones exhibiting transcriptional profiles of less-differentiated ST-HSCs. While these results support the notion of increased differentiation in non-repopulating clones, they also raise the subsequent question of which mature cell types are associated with the loss of self-renewal potency and repopulation capacity of ex vivo expanded HSCs. Hence, we set out to compare the molecular profiles of NonELSKs from repopulating and non-repopulating fractions.

First DEGs were computed between NegNonELSK (the non-HSCs from clones that did not contain functional HSCs) and PosNonELSKs (the non-HSCs from clones that contained functional HSCs) (cut-offs: $\log _{2} \mathrm{FC}=1$; $p$-value $=0.05$ ). The analysis revealed 426 genes overexpressed in NegNonELSKs and 530 genes enriched in

PosNonELSKs (Figure 12A). As the previous results indicated that clones contained multiple progenitor cell types, we performed a gene set enrichment analysis (GSEA), using well described gene signatures for a broad range of mature cell types. For this purpose, we utilised a comprehensive transcriptomic profiling dataset of HSCs, natural killer (NK) cells, T cells, B cells, macrophages, granulocytes and erythrocytes, as described by Chambers et al. ${ }^{355}$. Using DEGs between NegNonELSKs and PosNonELSKs, we observed strong enrichment of the myeloid lineage in NegNonELSK, specifically monocytes and granulocytes (Figure 12B). These observations mirror our findings above, which also indicate an increase in a broad set myeloid cell signatures in non-repopulating clones. In contrast, PosNonELSK fractions were weakly correlated to NK cells, erythrocytes and B cells (Figure 12B). These results indicate that repopulating clones are also subject to differentiation pressures but have weaker associations with specific lineages.



Figure 5.12. Repopulating and non-repopulating clones are enriched for distinct NonELSK cell types. (A) DGE volcano plot for NegNonELSK and PosNonELSK cell fractions (cut-offs: $\log _{2}$ FC = 1; pvalue $=0.05)$. (B) Gene set enrichment analysis (GSEA) based on DEGs of PosNonELSK and NegNonELSK cell fractions, using previously defined gene sets of haematopoietic cell types (see methods for details). NK = Natural killer cells; N.ER = Erythrocytes; BCELL = B cells; MYELO = Myeloid cells; MONO = Monocytes; GRAN = Granulocytes.

The results above provide further evidence of increased myeloid differentiation in nonrepopulating clones. However, cell type-specific gene signatures are defined by binary inclusion or exclusion of genes, often based on arbitrary cut-off selection following differential gene expression. Hence, we opted to utilise the recently developed direction of transition (DoT) algorithm to decipher which myeloid lineages are particularly enriched in non-repopulating clones ${ }^{357}$. In brief, the DoT score method projects differentiation trajectories onto a single cell RNA-seq (scRNA-seq) reference landscape ${ }^{357}$. By setting a point of origin on the references landscape and providing a set of genes of interest, the algorithm computes the likely differentiation trajectory away from the origin. The drive towards a particular lineage is indicated by positive DoT scores (red), while negative DoT scores (blue) are indicative of differentiation away from that lineage. The DoT score method thus provides a more unbiased approach, as it utilises the continuum of gene expression profiles across the scRNAseq landscape.

To capture a broad range of stem and progenitor populations, we selected a previously generated mouse LK/LSK scRNA-seq dataset to serve as a reference landscape ${ }^{343}$ (Figure 13A). Firstly, we selected the point of origin based on Leiden clustering of the reference landscape (Figure 13A). Here, we reasoned that ex vivo expanded clones with repopulation capacity are largely comprised of naïve cells and express similar transcriptomic profiles to long-term and short-term HSCs. Hence, we selected cluster 2 as the point of origin, as it is situated closest to the HSC cluster and poised to contain the ST-HSC cell group (Figure 13B).


Figure 5.13. Clonally expanded HSCs resemble freshly isolated HSCs. (A) UMAP representation of murine LK/LSK single cell gene expression profiles ${ }^{343}$, highlighting Leiden clusters. Clusters are numerated in descending order by cluster size. (B) UMAP representation as outlined in (A), colouring identified cell types. Dotted box (black dotted line) indicates the cluster, selected as the point of origin (Leiden cluster 2) for the Direction of Transition (DoT) Score computation ${ }^{357}$. (C) UMAP representation of single LK/LSK transcriptomes, outlining the DoT score for each cell. DoT scores were computed using DEGs of I) PosELSK vs PosNonELSK (left); II) PosELSK vs NegELSK (middle); III) PosELSK vs NegNonELSK (right). Enriched cell populations and trajectories are highlighted by positive DoT scores (red) and underrepresented cell types have negative DoT scores (blue). Red arrows indicate enriched cell populations and blue arrows indicate underrepresented cell populations.

Next, we used the gene set enriched in the PosELSKs to establish the trajectory of functional HSCs. As expected, we observed a strong transition towards the HSC cluster, while all other cell types were strongly distanced from the molecular profiles of PosELSKs (Figure 13C). As our ex vivo expansion cultures were derived from single phenotypic HSCs, we reasoned that differentiation towards mature lineages would originate from naïve compartment, captured by the PosELSK fraction. Hence, we computed DGE between I) PosELSKs against PosNonELSKs and II) PosELSK against NegNonELSKs. The projections of PosNonELSK revealed strong transition towards monocytes, neutrophils and basophils (Figure 13C). NegNonELSKs showed similar levels of transition towards monocytes, neutrophils and basophils. However, higher DoT scores for these lineages indicated a stronger differentiation force. In contrast, only NegNonELSKs showed differentiation towards erythrocytes and megakaryocytes (Figure 13C). These results are also in line with visual observations
of expansion clones, whereby high megakaryocyte content was indicative of the loss of functional HSCs with durable self-renewal potency (data not shown). Both repopulating and non-repopulating clones appeared to also express comparable affinity towards the lymphoid lineage (Figure 13C).

Collectively, these results indicate that the differentiation drive towards erythrocyte and megakaryocyte lineages in 28-day expansion cultures inhibits the self-renewal and proliferation of functional HSCs. Instead, only largely undifferentiated clones with moderate transition towards monocytes, neutrophils, basophils and the lymphoid lineage are able to support expansion of functional HSCs with self-renewal capacity.

### 5.8. Deriving a gene signature for repopulating HSCs

The ability to connect molecular profiles with the functional output from the same clone provides a powerful tool for deriving a set of signature genes for ex vivo cultured HSCs with robust repopulation capacity.

Wilson and colleagues previously described a gene signature shared between fresh HSCs, isolated using multiple FACS strategies (Wilson et al 2015). The resulting curated gene signature, termed molecular overlap (MoIO), was subsequently used to identify fresh HSCs in scRNA-seq experiments ${ }^{334,523}$. We computed the geometric mean of MolO signature genes and projected the scores onto the MDS plot of all cell fractions (Figure 14A). The MolO scores correlated with increased repopulation potency of clones, with PosELSKs scoring highest for the MolO signature (Figure 14A). Surprisingly, NonELSK from repopulating clones scored higher than ELSKs of non-repopulating clones, thus further supporting the notion of increased differentiation in non-repopulating clones and correlation of the MoIO signature with functional HSCs. (Figure 14A).

Next, all ELSK cell fractions were visualised on an MDS plot and each sample marked with its respective transplantation parameters, namely donor chimerism and GM contribution (Figure 14B). These results highlighted that differential transplantation outcomes are accompanied by transcriptional changes. Together with the apparent
correlation of MoIO signature scores with the transplantation outcomes, these observations raised the prospect of deriving a gene signature for functional HSCs with repopulation potential following ex vivo expansion.





Figure 5.14. Repopulating and non-repopulating cell fraction cluster separately. (A) MDS plot depicting clustering of batch-corrected bulk RNA-seq samples, excluding technical replicates (left-hand plot). Accompanying MDS plot representing the geometric mean across the previously defined LT-HSC molecular overlap (MoIO) gene signature (MoIO Score). (B) Isolated MDS plots for ELSK cell fractions (NegELSK and PosELSK), including I) cell fractions, II) percentage donor chimerism for each sample, III) percentage GM contribution for each sample.

To derive a repopulation signature, a principal component analysis (PCA) using all collected samples was performed (Figure 15A). Pearson correlation coefficients of principal components were derived using the following core transplantation parameters: 1) binary repopulation outcome (yes/no); 2) proportion of donor chimerism; contribution to 3) granulocyte-macrophage (GM), 4) T cells, 5) B cells lineages. Finally, the loading plots for principal components significantly correlated with repopulation metadata were used to derive the set of signature genes.

First the PCA was computed and this largely mirrored the pattern observed in the previous MDS plots (Figure 15B). Next all samples with successful BM reconstitution in primary transplantation assays were highlighted (Figure 15C). Of note, several PosNonELSK samples produced positive transplantation outcomes.


Figure 5.15. Deriving a molecular signature for repopulation potency using PCA. (A) Schematic outlining the strategy for deriving a repopulation signature. Principal compontent analysis (PCA) was performed using gene expression profiles. Transplantation metadata was used to derive a correlation with principal components (PCs). Loading plots of PCs associated with the repopulation outcome provided a set of target genes. (B) PCA plot for all batch corrected bulk RNA-seq samples (excluding technical replicates). Colour scheme matches previous panel figures. (C) PCA plot highlighting the binary repopulation outcome of primary transplantation assays at the 16-week timepoint.

Next, Pearson correlations for the first 10 principal components (PCs) were derived. The $R$-values for each parameter revealed a strong correlation with PC1, while all other PCs showed weak correlation (Figure 16A). To confirm the observation, we also plotted the corresponding $r^{2}$ values (Figure 16B). PC1 showed strongest correlation with the percentage GM contribution and the binary repopulation outcome (Figure 16A/B). Interestingly, PC1 was also highly correlated with the MolO score of each sample.


Figure 5.16. Transplantation metadata is correlated with principal components. (A) Pearson correlation of the MolO score and transplantation metadata for each sample, including Donor chimerism (float), GM lineage contribution (float), T cell lineage contribution (float), B cell lineage contribution (float), repopulation outcome (boolean, positive $>1 \%$ donor chimerism and $>1 \%$ GM contribution). Displaying Pearson r values and significance: ${ }^{*}=p<0.05,{ }^{* *}=p<0.01,{ }^{* * *}=p<0.001,{ }^{* * * *}=p<0.0001$. The prefix outlines the directionality and the first 10 PCs are displayed. (B) Pearson $r^{2}$ values, depicting the correlation of metadata outlined in (A) with the first 7 PCs. Significance: ${ }^{*}=p<0.05,{ }^{* *}=p<0.01$, *** $=p<0.001,{ }^{* * * *}=p<0.0001$.

To identify key coefficients driving PC1, we derived the loading plot for PC1 and PC2 and displayed the top 200 genes contributing to both principal components (Figure 17). The top 50 genes in the negative range of PC1 were then selected to form the repopulation gene signature. The signature included multiple genes with welldescribed functions in HSC biology. In particular, Esam, Prdm16, Slamf1, Procr and Mecom are amongst the most prominent genes ${ }^{35,72,75,524-526}$.


Figure 5.17. PCA loading plot reveals genes correlated with the transplantation outcome. PCA loading plot for PC1 and PC2. Top 200 genes are displayed and the top 50 genes associated with a positive repopulation outcome were selected, based on the importance for PC1.

### 5.9. Curation of the repopulation signature using linear models

While a broad gene signature for functional HSCs provides a powerful tool for identifying driver genes of repopulation potency, it does not indicate which genes are best suited for prospectively identifying clones with capacity to repopulate. In order to curate the set of 50 repopulation signature genes, regression models were utilised. For this purpose, three repopulation metadata parameters with the highest correlation with PC1 were assessed: I) binary repopulation outcome, II) donor chimerism, III) percentage GM contribution. Logistic curves for each signature gene were fit against the repopulation outcome and plotted the Firth penalised regression coefficients (Figure 18A). The process was repeated for donor chimerism and GM contribution, fitting linear regression models for the same set of signature genes (Figure 18B/C). The resulting set of regression coefficients identified the most significant driver genes for each metadata parameter. Next, we reasoned that the genes consistently scoring highest across all three parameters would be the most promising candidates for a stringent repopulation gene signature. A cut-off for regression coefficient at the lower range of the standard deviation of the gene with the highest coefficient (Figure 18A-
C). We identified 11 genes that were consistently identified amongst the clones with the most functional HSC activity and most significantly correlated with all three metadata parameters (Figure 18D). Several genes were previously implicated in HSC biology, including Esam and Fgfr1 ${ }^{524,527}$.

D

| Overlapping Genes |  |  |  |
| :--- | :--- | :--- | :--- |
| Esam | Fgfr1 | Fst11 | Gm38066 |
| Insyn1 | Lzts1 | Pim2 | Ptk2 |
| Skint3 | Tgfbr3 | Timp3 |  |

Figure 5.18. Regression analysis reveals a core repopulation gene signature. (A) Logistic regression models for the top 50 genes, associated with repopulation. Firth penalised logistic regression coefficients were computed for the binary repopulation outcome of each sample. (B-C) Linear regression coefficients computed for the correlation of genes in (A) with the percentage donor chimerism and GM contribution of each sample. The significance cut-offs for all regression models are indicated by a red dotted line and the ranges represent the standard deviation. (D) List of genes above the regression coefficient cut-off, shared between all three regression analyses.

Amongst the 11 shared signature genes, seven have previously been described in the context of HSC biology and haematopoiesis. As previously mentioned, ESAM has been recognised as a promising marker of HSCs ${ }^{524}$. A functional screen identified Fstl1 as a potential key driver for repopulation activity of $\mathrm{HSCs}^{528}$. Furthermore, the expression of Tgfbr3 has been shown to increase significantly following ex vivo
expansion ${ }^{529}$. With the role of TGF- $\beta$ in HSC self-renewal widely characterised, Tgfbr3 provides an intriguing target gene ${ }^{139}$.

Fgfr1 was shown to play a critical role in maintaining multilineage potential of LT$\mathrm{HSCs}^{530}$. It has also been demonstrated how the provision of the complementary agonist, fibroblast growth factor 1 (FGF1), enhanced expansion of functional HSCs with long-term BM reconstitution capacity ex vivo ${ }^{527}$. Pim2 has also been identified as a potent activator of HSC proliferation in myelodysplastic syndrome patients ${ }^{531}$.

Lu et al. demonstrated that the deletion of Ptk2 [alternative name: focal adhesion kinase (FAK)] induced increased cycling of HSCs, while not altering their repopulation potency ${ }^{532}$. In a subsequent study, Batista et al. reported that the loss of Ptk2 induced myeloid lineage bias during HSC commitment ${ }^{533}$. Interestingly, we observed that nonrepopulating clones, subject to increased myeloid differentiation, expressed low levels of Ptk2, while repopulating ELSK fractions were subject to significantly higher Ptk2 levels. Hence, our results complement the study by Batista et al. and collectively raise the prospect of Ptk2 playing a crucial role in HSC self-renewal.

Overexpressed of the matrix metalloproteinase inhibitor Timp3 by BM niche cells has been shown to result in increased proliferation of stem and progenitor cells in vivo ${ }^{534}$. The expression of Timp3 was localised in the endosteal niche, while HSCs themselves expressed low levels of Timp3 in vivo. Another study corroborated the above findings, demonstrating that supplementing recombinant Timp3 to ex vivo HSC cultures played a critical role in inducing LT-HSC proliferation and expansion of naïve multipotent hematopoietic cells ${ }^{535}$. Here we provide the first evidence that Timp3 expression increased in proliferating stem and progenitor cells (ELSKs) following ex vivo expansion. These ex vivo observations mirror the published in vivo data above, thus prompting the question of whether ex vivo expansion cultures, and particularly the functional HSC compartment itself, is able to mimic the niche signals that would favour HSC self-renewal.

While the majority of signature genes were previously implicated in HSC biology, particularly in HSC proliferation and self-renewal, the newly derived synergy between in vivo and ex vivo data provides evidence for the core molecular machinery of HSC
self-renewal and multilineage capacity. Furthermore, these data also instil confidence that ex vivo expanded functional HSCs share a wide range of molecular characteristics with their in vivo counterparts.

In addition to the conserved genes with previously described in vivo functions in HSCs, the RepopSig included several novel genes (not previously described in HSC function). The biological function of Insyn1 was first described in 2016, when Uezu et al. identified a critical role of Insyn1 in synaptic inhibition ${ }^{536}$. Beside its role in synaptic transmission, Insyn1 remain uncharacterised in all other tissues and none of its protein domains have currently known functions. The clear enrichment of Insyn1 in ex vivo expanded HSCs thus provides preliminary evidence for a role in haematopoiesis. Similarly, the uncharacterised gene Gm38066 was enriched in the repopulation gene signature (Figure 18D). A previous transcriptomic screen also revealed enrichment of Gm38066 in stem and progenitor cells ${ }^{537}$. Furthermore, Lzts1 has been implicated in neuronal delamination, while Skint3 was found to regulate wound repair ${ }^{538,539}$. Together, the RepopSig specifies a set of genes that reliably track with HSC function, thus providing an intriguing avenue to discover novel regulators of HSC self-renewal and multilineage reconstitution across a multitude of cell origins.

### 5.10. Extension of the repopulation gene signature using highly correlated genes

The core repopulation signature, hereafter referred to as RepopSig, was derived using a stringent overlap between three metadata parameters. However, multiple genes scored highly for 2 parameters and narrowly missed the cut-off for the third. Hence, we plotted the expression for all 50 signature genes against each repopulation parameter and projected the computed linear or logistic regression models. We then selected genes that showed high $p$-values and were amongst the hits above the cuoff for at least 2 parameters (Figure 19A-C). The resulting 12 signature genes displayed significant differential expression between repopulating and nonrepopulating clones (Figure 19A-C).


Figure 5.19. Manual curation of regression models reveals additional repopulation-assoiated genes. (A) Logistic regression models of selected individual repopulation signature genes and the repopulation outcome of respective samples (p-value displayed). (B-C) Linear regression plots outlining correlation of selected signature genes with GM contribution and donor chimerism. $R^{2}$ and $p$-values are indicated.

In total, the RepopSig contained 23 genes, with 14 previously implicated in HSC biology and the remaining 9 genes not previously described in haematopoiesis (Figure 20A). To determine whether the RepopSig was able to selectively mark repopulating clones and particularly the PosELSK fraction, we computed the geometric mean and projected the RepopSig score onto the MDS landscape (Figure 20B). As expected, the RepopSig scores distinguished ELSK from nonELSK fractions and repopulating from non-repopulating clones (Figure 20B). Within each sample group, the scores appeared consistent and provided clear distinction to each other group.
A

| Repopulation Signature |  |  |  |
| :--- | :--- | :--- | :--- |
| Esam | Fgfr1 | Fst11 | Gm38066 |
| Insyn1 | Lzts1 | Pim2 | Ptk2 |
| Skint3 | Tgfbr3 | Timp3 | Arx |
| Dennd2a | Klhl4 | Mpdz | Myof |
| Nrk | Palld | Prdm16 | Prex2 |
| Ryk | Slamf1 | Zfp532 |  |



Figure 5.20. Curated repopulation signature for HSCs. (A) List of final Repopulation gene signature. Includes genes identified in at least 2 of 3 metadata regression models (as outlined in Figure 14A-C), based on regression coefficients and p-values. (B) MDS plots of all samples, coloured by cell type (left) and the geometric mean of repopulation signature (RepopSig) genes (right).

The resulting set of 23 RepopSig genes contained several novel targets and some genes with well-described functions in HSC biology. Firstly, Prdm16 has widely been recognised as a key regulator of HSC quiescence, self-renewal and differentiation ${ }^{525,540}$. The WNT receptor RYK modulates HSC survival and quiescence, by modulating downstream WNT signalling ${ }^{541-543}$. Previous work by Bernatchez et al. outlined a key role of Myof in regulating signalling pathways mediated by vascular endothelial growth factor (VEGF), with VEGF previously being implicated in mediating HSC survival ${ }^{544,545}$. In turn, the multi-omic screen of HSCs and multipotent progenitors, conducted by Cabezas-Wallscheid et al., revealed HSC-specific enrichment of Myof ${ }^{176}$. Slamf1 (CD150) is amongst the most commonly used cell surface markers used for the isolation of pure LT-HSC populations ${ }^{75}$. Together with Prdm16, Esam, Ryk, Tgfbr3 and Myof, the enrichment of these key HSC-associated genes in the RepopSig raises confidence in the validity und utility of the novel set of marker genes.

Next, we inspected the set of RepopSig genes previously not directly implicated in HSC function:

1) The transcription factor Arx was widely implicated in neuronal function and pancreatic cell identity ${ }^{546-548}$. Interestingly, Arx was shown to regulate and induce Meis1 expression in cortical migrating interneurons ${ }^{548}$. Meis1 also being a key regulator of reactive oxygen species (ROS) and quiescence in LT-HSCs could indicate a novel Arx-Meis1 regulatory axis of HSC function ${ }^{441}$.
2) The GDP-GTP exchange factor Dennd2a forms part of the Ras superfamilymediated signalling pathways, with particular prevalence in Rab signalling ${ }^{549,550}$. The Ras superfamily and specifically Rho GTPases, such as Cdc42, have been widely implicated in HSC function and fate choices ${ }^{551}$. However, Ras-mediated signalling remains largely unknown.
3) The function of KIhl4 remains poorly described, but has previously been identified as homolog of the Drosophila Kelch gene, a member of the ubiquitinproteasome pathway ${ }^{552}$.
4) Mpdz has been well described to play a role in cell-to-cell adhesion and the central nervous system, by co-localising to intracellular domains of cell junction in neuronal synapses, epithelial and endothelial cells ${ }^{553-556}$. While Mpdz has not been previously described in HSC biology, these results indicate a potential role of Mpdz in facilitating cell-cell interaction of HSCs in vitro.
5) While Nrk has not been directly implicated in HSC biology, its involvement in the tumor necrosis factor alpha (TNF- $\alpha$ ) suggests a possible function in HSC maintenance ${ }^{557,558}$.
6) Palld plays a crucial role in the formation of the actin cytoskeleton in various cell types, including the regulation of axon growth ${ }^{559,560}$. Interestingly, a transcriptomic screen that identified expansion of LT-HSCs in response to sex steroid ablation (SSA) also revealed enrichment of Palld in the LT-HSC compartment ${ }^{561}$. In addition, the loss of Palld was associated with erythropoietic defects in FLs but did not affect the intrinsic multilineage and self-renewal potency of HSCs ${ }^{562}$. These results suggest that Palld is not essential for HSC function, but the correlation with actively cycling LT-HSCs makes Palld a promising marker of functional HSCs.
7) Intriguingly, RNA-seq profiling of human stem and progenitor cells revealed enrichment of Prex2 in actively cycling progenitor cell populations ${ }^{563}$. As a guanine nucleotide exchanger (GEF), Prex2 thus warrants further attention to decipher its role in HSC biology.
8) The transcriptional regulator $Z f p 532$ is a known downstream target gene in response to TPO signalling and its expression was downregulated in a TPOdesensitised mouse model ${ }^{564}$. TPO-mediated support of HSC quiescence has been well documented and the majority of ex vivo HSC culturing strategies rely on the supply of exogenous TPO, including the F12 culturing system used in this study ${ }^{132}$. Hence, the observation of a TPO response gene in the RepopSig could thus indicate that TPO-responsive HSCs are selectively enriched amongst ex vivo expanded functional HSCs.

### 5.11. The repopulation signature can stratify clones by HSC content

Next, we compared the 2015 MolO signature with RepopSig for the ability to mark functional HSCs with long-term repopulation capacity. The geometric means for both gene signatures (Figure 21A/B) were plotted and, as previously mentioned, both signatures were able to significantly separate repopulating and non-repopulating clones (including ELSK and nonELSK fractions). However, RepopSig provided greater resolution between repopulating ELSK and non-repopulating ELSKs (Figure 21C) and also better defined NegNonELSKs from NegELSKs (Figure 21B). Overall, these results indicate that the RepopSig is well-suited for attributing functional capacity (in form of repopulation potential) of HSCs from molecular profiles. However, the same data that was used to derive the RepopSig, was utilised to test stratification. Hence, application in other independently generated RNA-seq datasets is required to validate whether the RepopSig can prospectively identify and stratify clones with functional HSC content (see section 3.13.).


Figure 5.21. Repopulation signature provides an improved separation of repopulating samples. (A) Boxplot of MolO scores (geometric mean) for each population. (B) Boxplot of RepopSig scores (geometric mean) for each population. Significance: $n s=>0.05 ; *=<0.05 ; * *=<0.01 ; * * *=<0.001$; **** $=<0.0001$. (C) Line plot depicting geometric means for MolO signature genes and RepopSig. Corresponding samples are connected via a line.

### 5.12. Inhibition of Rho GTPases collapsed HSC clones, but over-stimulation did not enhance HSC proliferation

While the RepopSig provides an intriguing insight into marker genes and possibly novel driver genes of functional HSCs, it does not give mechanistic insight into the molecules governing HSC self-renewal. To identify key signalling pathways, enriched in ex vivo expanded functional HSCs, we first computed pairwise Pearson correlations of each identified gene against the geometric mean of the RepopSig and ranked all genes according to their Pearson correlation coefficients and highlighted all MoIO signature genes (Figure 22A). As expected, the vast majority of MoIO signature genes were highly correlated with the RepopSig geometric mean (Figure 22A). However, Fads3, Ctsf and Pde1b displayed correlation values of $r<0.5$ and Ifitm1, Metti7a1, Sult1a1, Gstm1, Limd2 and Ub/3 were nearly entirely uncorrelated with the RepopSig ( $r<0.1$ ) (Figure 22A). These results indicate that these genes, despite their enrichment in fresh HSCs, are not required for HSC function following ex vivo culture.

To identify key pathways involved in HSC self-renewal, two gene lists were prepared: 1) RepopSig genes (all genes with a correlation coefficient $r \geq 0.7$ and $p$-value $\leq 0.01$ (Benjamini-Hochburg corrected)) and anti-correlated genes (Anti) using $r \leq-0.7$ and $p$-value $\leq 0.01$. The resulting gene sets, which either correlated (RepopSig) or anticorrelated (Anti) with repopulation potential, were used to derive enrichment of KEGG pathways (Kyoto Encyclopedia of Genes and Genomes) ${ }^{565}$. All signalling pathways with significant enrichment ( $p$-value $\leq 0.05$, indicated by red line) were visualised using either RepopSig-correlated or Anti-correlated gene sets (Figure 22A). A number of pathways emerged including Sphingolipid, FoxO and VEGF signalling. Notably, recent work by Xie et al. revealed that modulation of sphingolipid signalling in human HSCs altered their functional output and particularly the self-renewal potency ${ }^{566}$. Interestingly, several other pathways, namely FoxO, Hippo, Ras and VEGF, have not only been implicated in HSC biology but also share elements of the Rho GTPase signalling family ${ }^{544,545,567-569}$. In addition, the Rho GTPases CDC42 and ARHGAP5 have also been directly implicated in HSC function, thus raising the question whether the modulation of Rho GTPase signalling in ex vivo cultures could enhance the yield of functional HSCs ${ }^{570-572}$.


Figure 5.22. Signalling pathways enriched in single cell-derived clones with repopulation capacity. (A) Rank plot of Pearson correlation coefficients for all individual genes against the RepopSig. MolO genes are highlighted in red. MolO gene names are displayed for genes falling below $r=0.5$. (B) KEGG pathway analysis for genes correlated with the RepopSig (Pearson correlation; $r>0.7$ ) and genes inversely correlated (Anti; $r<-0.7$ ). All signalling pathway KEGG terms and the associated adjusted p-values for RepopSig and anti-RepopSig are displayed. Red line indicates a p-value cut-off $<0.05$.

To address the above question, we set up 28-day expansion cultures for bulk HSCs ( 50 HSCs per well) and supplied various activators and inhibitors of RhoGTPases. These included the activator ML099 and the following inhibitors: I) CASIN, II) NSC23766, III) Rhosin ${ }^{572-575}$. Each small molecule was supplied in three different concentrations to establish any relevant dose response (see methods for more details). While all concentrations of Casin resulted in a complete collapse of cultures, NSC23766 and Rhosin reduced viability in a dose-dependent manner (Figure 23A/B). This included the overall survival of all live cells and the proportion of ELSKs in each culture (Figure 23A/B). These results suggest that Rho GTPases are essential for the survival and expansion of ex vivo cultured HSCs and their downstream progeny. In contrast, constitutive activation of Rho GTPase signalling via ML099 did not enhance the proportion of ELSKs in each culture and did not affect the total number of live cells (Figure 23A/B) thus suggesting that additional activation of Rho GTPases does not induce increased proliferation and self-renewal of HSCs. In summary, these results corroborate previous studies outlining the importance of Rho GTPases in in vivo HSC function, by demonstrating a similar importance in the ex vivo setting. However, overstimulation did not appear to induce HSC proliferation.


Figure 5.23. Modulation of signalling pathways in 28-day expansion cultures using extrinsic factors. (A-B) Proportion ELSK cells and the total number of live cells after 28 days of bulk (input: 50 cells) HSC expansion cultures, exposed to the following molecules: I) CASIN (2uM, 10uM, 20uM), II) NSC237D66 (5uM, 50uM, 200uM), III) Rhosin (1uM, 10uM, 50uM) and IV) ML099 (1uM, 10uM, 50uM). Statistical analysis: One-way ANOVA with significance levels ${ }^{*}=<0.05 ;{ }^{* *}=<0.01 ;{ }^{* * *}=<0.001 ;{ }^{* * * *}=$ <0.0001.

### 5.13. Prospective identification of successful ex vivo expansion cultures using RepopSig

In order to explore whether genes in the RepopSig might be useful for prospectively identifying cultures containing large numbers of HSCs, a series of experiments were undertaken to assess both positive and negative markers of HSC content by qPCR. First, linear regression models were computed to determine genes that anti-correlated with GM contribution and donor chimerism. Amongst these, three possible targets were particularly prominent, namely Cebpa, Siglecf and Rab44 (Figure 24A/B), each of which has been previously identified as regulators of hematopoietic progenitors and mature cell lineages ${ }^{520,522,576}$. Alongside these three potential negative markers,

RepopSig genes Ptk2, Prex2, Prdm16, Palld, Mpdz, Klhl4 and Fstl1 were selected as possible positive markers of functional HSC content.

A series of 28-day single HSC-derived expansion cultures were undertaken in individual clones to include simultaneous expression profiling by qPCR, phenotypic profiling by flow cytometry and functional profiling by transplantation assays. Previous data (Chapter 3.2.) indicated that clones with high ELSK percentages (\%ELSK) in single cell-derived clones were highly correlated with positive repopulation outcomes so all clones with >20\% ELSK and all clones < 1\% ELSK (Figure 24C) were grouped for these assays. Five of seven RepopSig genes showed significant enrichment in clones with $>20 \%$ ELSK content (Figure 24C), including in many cases exclusive expression in >20\% ELSK clones. All three negative marker genes were significantly enriched in <1\% ELSK-containing clones (Figure 24C).


Figure 5.24. Repopulation signature genes are uniquely expressed in clones with functional HSCs. (A-B) Linear regression models to compute the correlation of selected genes, enriched in nonrepopulating clones, with GM lineage contribution and donor chimerism. (C) Gene expression profiling (RT-qPCR) of RepopSig genes and genes, correlated with non-repopulating cell populations. Single HSC-derived 28 -day expansion cultures were harvested for flow cytometry analysis. Clones with $>20 \%$ ELSK proportion ( $n=13$ ) and clones with $<1 \%$ ELSK proportion of live cells ( $n=15$ ) were selected for gene expression profiling. Significance levels: ${ }^{*}=<0.05 ;{ }^{* *}=<0.01$; ${ }^{* * *}=<0.001$. Error bars indicate standard deviation.

Nine clones with >20\% ELSK and ten clones $<1 \%$ ELSK content from the above pool of single HSC-derived 28-day clones were selected for transplantation assays (Figure 25A). Clones with $>20 \%$ ELSK content exhibited donor chimerism $\geq 50 \%$ and showed robust multilineage contribution, while all other clones fell $<1 \%$ donor chimerism (Figure 25A/B). Individual gene expression of the RepopSig genes and negative markers were plotted against the \%ELSK of the corresponding clones (Figure 25C).

Transplanted clones with $>20 \%$ ELSK are marked in red and clones with $<1 \%$ are marked in blue. Siglecf, Rab44, Cebpa, Prdm16, Fstl1, Prex2, Palld, Mpdz and Klhl4 all showed significant correlation with ELSK percentage and consequently the repopulation potency of the clone (Figure 25C). Overall, these data confirm that \%ELSK correlates with the repopulation outcome of 28-day HSC clones and that the expression of RepopSig genes predicts the presence of functional HSCs. In addition, these data also independently validate the correlation of the RepopSig with repopulation, thus showing robust identification of function HSCs by the RepopSig.


Figure 5.25. Repopulation Signature genes selectively mark clones with repopulation potential. (A) 28-day expanded HSC cultures were first inspected for their ELSK proportion. Clones with > 20\% $E L S K(r e d, n=9)$ and clones with $<1 \%$ ELSK (blue, $n=10$, pooled into a single recipient mouse) were subject to transplantation assays. The proportion of donor chimerism up to 16 weeks is displayed. (B) Lineage output of the transplantation assays in (A). Contribution of T, B and GM lineages 16 weeks post-transplantation is expressed as the percentage of donor cells. The corresponding donor chimerism (\%) is indicated below each bar. (C) Gene expression profiling (RT-qPCR) of RepopSig genes and negative markers, identified in Figure 18C. Computed Pearson correlation ( $r^{2}$ displayed) between gene expression and the proportion of ELSK in respective clones. Clones transplanted in (A) are indicated in red (> $20 \%$ ELSK) and blue ( $<1 \%$ ELSK). Significance: ${ }^{*}=<0.05 ;{ }^{* *}=<0.01 ;{ }^{* * *}=<0.001 ;{ }^{* * * *}=$ <0.0001.

Prospective screening of the marker genes outlined above permits early detection of repopulating clones. Consequently, such pre-screening might streamline the culturing process, improve time- and cost- efficiency by reducing the length of time unsuccessful clones are maintained. Furthermore, current screening strategies rely on cell surface phenotyping, requiring more cell input and manual labour than targeted qPCR. Hence, a reliable and simple prospective screen, based on the expression of defined marker genes, enables fast identification of functional HSCs without the need of time-intensive transplantation assays. This provides a powerful tool to enable a myriad of other multiomic profiling methods that rely on destructive assays, including proteomics, phosphor-proteomics, metabolomics and others.

### 5.14. RepopSig marks functional HSCs across multiple cellular states

Having established that the RepopSig correlates with transplantation outcomes, selectively marks functional HSCs ex vivo and can serve as a prospective screening tool, we set out to determine whether the RepopSig might be more broadly applicable for the identification of HSCs from multiple cellular states. For this purpose, we interrogated previously generated scRNA-seq datasets to determine the utility of the RepopSig by comparing its signature scores with the MolO signature.

First, we utilised a scRNA-seq dataset of $\sim 1600$ freshly isolated long-term HSCs (LTHSCs), hematopoietic stem and progenitor cells (HSPCs) and Progenitors (Prog) ${ }^{22}$ (Figure 26A). We scored each cell for the RepopSig and MoIO signature, which revealed that both signatures were able to significantly distinguish the LT-HSC cluster from HSPCs and Progenitors (Figure 26B). The MolO signature provided a more striking resolution of the cell type clusters and outperformed the RepopSig, largely due to the inclusion of cell cycle genes in the MoIO signature (Figure 26C). In contrast, positive RepopSig scores were confined to the LT-HSC cluster (Figure 26D). In particular, the expression of Fstl1, Klhl4, Esam, Skint3 and Timp3 was restricted to the LT-HSC cluster (Figure 26E). Overall, these data indicate that the RepopSig is also able to selectively mark freshly isolated HSCs.


Figure 5.26. Repopulation signature selectively marks freshly isolated LT-HSC. (A) UMAP representation of previously published murine haematopoietic stem and progenitor (HSPC) scRNA-seq profiles ${ }^{22}$, coloured by the respective broad cell types. (B) Boxplots of MolO and RepopSig scores for each respective cell type. Pearson correlation p-values displayed. (C-D) UMAP representation mirroring (A), coloured by MoIO signature and Repopulation signature. (E) UMAP representations of mouse HSPCs ${ }^{22}$. Normalised gene expression intensity for RepopSig genes depicted.

The results above indicate that the RepopSig holds the potential to provide a universal signature for cultured and proliferating LT-HSCs since it could identify both in vitro and in vivo HSCs. To explore whether it could also detect actively cycling cells in vivo and quiescent HSCs in vitro we generated new single cell gene expression profiling (scRNA-seq) of actively cycling fetal liver HSCs (FL HSCs) and hibernating HSCs (hibHSCs), which were kept in cytokine-depleted media for 7 days ${ }^{177,334}$. We projected FL HSC and hibHSC profiles onto the reference landscape (dataset previously generated by Nestorowa et al. ${ }^{22}$ ) and computed RepopSig and MolO signature scores (Figure 27A/B). As expected, both cell populations mapped to the LT-HSC cluster and
expressed similar gene expression profiles (Figure 27A). However, a subset of FL HSCs mapped to the HSPC and Progenitor clusters (Figure 27A). Such phenomenon is not surprising, as freshly isolated FL HSCs have a lower purity than hibHSCs (~25\% versus $\sim 50 \%$ ). Exclusion of FL HSCs mapping outside the LT-HSC cluster was also performed (Figure 27B). Despite the overlapping global molecular profiles, the MoIO signature significantly separated hibHSCs and FL HSCs (Figure 27B) whereas the RepopSig did not distinguish FL HSCs and hibHSCs, despite their distinct cell cycle status.


Figure 5.27. Repopulation signature identifies foetal liver and hibernating HSCs. (A) UMAP reference map of mouse HSPC single cell gene expression profiles ${ }^{22}$, coloured in grey. scRNA-seq profiles for fetal liver (green) and hibernating HSCs (orange) are projected onto the reference map. (B) MolO (top left) and RepopSig (bottom left) gene signature scores for fetal liver HSCs (FL) and hibernating HSCs (hibHSC). T-test p-values indicated. FL HSCs, mapped outside the reference LTHSC cluster, were excluded and boxplots for MolO (top right) and RepopSig (bottom right) computed. $T$-test $p$-values indicated.

Next, a scRNA-seq dataset of freshly isolated and hibernating HSCs, previously generated in our lab ${ }^{334}$, was used to test the utility of the RepopSig (Figure 28A). In brief, both the MoIO and RepopSig marked HSCs (Figure 28B-D). However, the MoIO
score exhibited bias towards freshly isolated HSCs, while the RepopSig expressed some bias towards cultured HSCs (Figure 28B).


Figure 5.28. Repopulation signature reduces HSC identification bias towards freshly isolated cell populations. (A) UMAP representation of freshly isolated unstimulated HSCs (HSC), unstimulated hibernating HSCs (hibHSCs) and their SCF-stimulated counterparts (HSC_SCF and hibHSC_SCF) ${ }^{334}$. (B) Boxplots of MoIO and RepopSig gene scores for the respective cell populations. T-test p-values indicated. (C-D) UMAP representation as outlined in (A), coloured by the respective MolO and RepopSig scores.

More recently, we compared our data with a large scale scRNA-seq screen of a 28day bulk HSC expansion culture using the F12 strategy (data provided by Dr Iwo Kucinski, Prof Bertie Gottgens, and Dr Adam Wilkinson, Figure 29A). This dataset provides an intriguing tool to identify the subset of cell marked by the RepopSig at single-cell resolution. Signature scores were computed for the Molo and RepopSig. Interestingly, the MoIO signature marked HSCs, a subset of HSPCs and several progenitor populations (Figure 29B). In contrast, RepopSig showed high specificity for the HSC cluster, but also some enrichment of late megakaryocyte progenitors (Figure 29C). These results provide an independent test dataset, confirming high specificity of the RepopSig for HSCs in ex vivo expansion cultures.


Figure 5.29. Single cell gene expression profiling of 28 -day HSC expansion cultures reveals that the repopulation Signature marks the HSC cluster. (A) Comprehensive scRNA-seq profiling of a single bulk HSC culture expanded for 28 days (data generated and processed by Dr Kucinski and Dr Wilkinson). UMAP representation coloured by cell type annotation (Dr Kucinski). (B-C) UMAP depicting MoIO and RepopSig gene scores.

To identify key driver genes for gene set scoring, we plotted average expression profiles and the associate cell frequency for each cell type. Interestingly, the majority of MolO signature genes were neither unique expressed, nor significantly enriched in HSCs (Figure 30A). In contrast, the majority of RepopSig genes were most highly expressed in HSCs and often largely undetected in other cell types (Figure 30B). However, the frequency within the HSC cluster was below $50 \%$ for genes, excluding Pim2 and the mean expression across the HSC cluster was also moderate (Figure 30B). The expression plots also revealed that late megakaryocyte progenitors express high levels of Timp3 and Slamf1 (Figure 30B), suggesting that the observed RepopSig scoring for that cell cluster is largely driven by two signature genes. When gene signature scoring for the RepopSig was computed excluding the highly expressed marker Timp3 the enrichment of late megakaryocyte progenitors was significantly
reduced and the RepopSig score in the HSC cluster (Figure 30C). Refinement of the gene signature with newly generated datasets may allow a more precise definition of molecular cell identities.


Figure 5.30. Repopulation signature genes are enriched in the HSC cluster. (A) Gene expression of MolO signature genes. Mean expression per cluster and fraction of cell in each cluster expressing the gene are indicated. (B) Gene expression of RepopSig genes. (C) UMAP plot of RepopSig scores, excluding Timp3 in the gene signature computation.

Overall, these data suggest that the RepopSig marks HSCs from multiple states. This included I) freshly isolated and ex vivo cultured, II) actively cycling and quiescent or III) adult and fetal liver HSCs. The destructive nature of molecular screening tools, such as scRNA-seq, do not permit simultaneous molecular and functional profiling at single cell resolution. Hence, reliable gene expression markers are vital to infer the functional capacity of profiled cells. The RepopSig was derived from matched samples, linking gene expression profiling with transplantation assays. Thus, we are confident the RepopSig provides a versatile tool for identifying functional HSCs with durable self-renewal potency and multilineage contribution in a variety of experimental setups.

### 5.15. Discussion

Here, we report a novel repopulation gene signature, selectively marking functional HSCs following ex vivo culture. Most intriguingly, screening for such signature genes allowed for prospective identification of single cell-derived clones with long-term repopulation potency. In addition, we showed that the repopulation gene signature was able to mark LT-HSCs uniformly and selectively across a multitude of cellular states, including FL HSCs, hibernating HSCs, freshly isolated HSCs and ex vivo expanded HSCs. Hence, in contrast to the more commonly utilised MolO signature, our repopulation signature identifies HSCs irrespective of their cell cycle status. Furthermore, molecular characterisation of long-term ex vivo cultures revealed that non-repopulating HSC clones were enriched for erythrocyte and megakaryocyte gene signatures. Finally, Fgd5 and EPCR provide potent markers for functional HSCs, undergoing ex vivo expansion.

To date, the limited availability of long-term ex vivo expansion protocols has hindered comprehensive molecular characterisation of cultured HSCs. While the underlying molecular states of freshly isolated HSCs have been extensively profiled using scRNAseq, little is known about changes in the molecular machinery throughout the ex vivo culturing process ${ }^{225}$. Here, we provide comprehensive bulk and scRNA-seq profiling of 28-day expanded, single cell-derived HSC clones.

To be able to identify key drivers of ex vivo expansion of functional HSCs, it is of paramount importance to develop a set of reliable in vitro markers. Here, we report that EPCR and Fgd5, in combination with the traditional LSK phenotype, improved the identification of functional HSCs, as confirmed by transplantation assays. Previous efforts by Fares and colleagues identified EPCR as a potent in vitro marker of human HSCs, thus corroborating our observations using murine HSCs and illustrating high relevance to understanding the molecular regulators of human HSC expansion where sufficient discrimination of HSCs and non-HSCs has not been achieved to be able to undertake robust molecular studies ${ }^{486}$.

Extensive clonal heterogeneity remains a common limitation shared between all ex vivo culturing strategies, including the most recent strategy described by Wilkinson et al. ${ }^{225}$. Such heterogeneous outcome between individual HSC clones and particularly its unpredictability are key unmet needs for ex vivo gene therapies, targeting haematological diseases ${ }^{577,578}$. To separate differentiating clones from clones that retained expanded functional HSCs, we performed transplantation assays, flow cytometry analysis for cell surface markers and RNA-seq profiling all from the same single-cell derived culture. Linking the molecular profiles with functional outcomes enabled the curation of a repopulation-specific gene signature (RepopSig), which correlated with a positive transplantation outcome. Most intriguingly, screening for the expression of individual signature genes enabled prospective identification of repopulating clones.

Several RepopSig genes, such as Esam, Prdm16 and Fst/1, have previously been characterised as potent HSC self-renewal drivers and regulators of long-term repopulation potency ${ }^{579-584}$. As outlined in section 3.3.10, several novel genes were selectively enriched in functional HSCs and identified as part of the RepopSig. Interestingly, the neuronal genes $A r x$ and $M p d z$ were amongst such genes. As discussed previously, Mpdz facilitates cell-to-cell adhesion in neuronal synapses and Arx induces Meis1 expression to regulate interneuron migration ${ }^{547,548,553,585-587}$. While the role of the sympathetic nervous system in the bone marrow niche has been widely described, these data suggest an HSC-intrinsic, previously unrecognised, role of molecules linked to neuronal signalling ${ }^{588}$.

The importance of RhoGTPases signalling for in vivo HSC self-renewal has been widely recognised ${ }^{570,571,589}$. Particularly as signalling effectors of VEGF, Hippo or FoxO pathways ${ }^{570,571,589}$. Our observations of abrogated HSC proliferation in cultures supplemented with inhibitors of RhoGTPases confirmed that these pathways are essential for HSC proliferation ex vivo. However, the notable absence of improved HSC yield in presence of pathway agonists suggests that while the pathway requires activation, manipulation of its signalling activity does not provide a proportional increase in self-renewal.

Previous reports have hypothesised that negative feedback signals, secreted by differentiating progenitor cell types, could perpetuate the progression of differentiation ${ }^{494,590}$. However, which cell types are responsible for such collapse of HSC cultures remain unknown. Intriguingly, non-repopulating single cell-derived clones were enriched for megakaryocyte and erythrocyte gene signatures. The increased level of megakaryocyte differentiation in clones that lost functional HSCs was also observed by Wilkinson et al. ${ }^{225}$. Megakaryocytes have been attributed a dual role within the HSC niche ${ }^{591,592}$. Bruns et al. observed megakaryocyte-mediated secretion of TFG- $\beta$ and CXCL4 to maintain HSC quiescence ${ }^{591}$. In contrast, Zhao et al. proposed that megakaryocytic secretion of Fgf1 induced HSC proliferation in response to injury ${ }^{592}$. Others also observed a critical function of Fgf1 in HSC proliferation ${ }^{527}$. Most intriguingly, the complementary receptor Fgfr1 has been identified as a RepopSig signature gene. While the role of megakaryocytes in the in vivo HSC niche has been described, the role of megakaryocytes in ex vivo cultures remains poorly understood. Prolonged or excessive proliferative signalling has been shown induce HSC exhaustion ${ }^{593,594}$. Speculatively, excessive proliferative signalling, induced by increased differentiation into megakaryocytes, could stimulate ex vivo HSC exhaustion and restrict multilineage contribution. While not fully revealed the underlying molecular mechanism, these data indicate a critical role of the erythrocyte and megakaryocyte lineages in providing feedback signals to induce differentiation.

Finally, the vast majority of other LT-HSC gene signatures are dominated by cell cycle regulatory genes, likely due to the distinctly quiescent state of HSCs in vivo ${ }^{332,595}$. Illustrating this further, signatures such as the MoIO score exhibited a strong bias towards freshly isolated adult HSCs ${ }^{332}$. However, our inability to reliably dissect
functionally heterogeneous HSCs limits the universal applicability of such gene signatures. Hence, it is vital to link molecular profiles with functional assays to identify a universal gene signature for functional HSCs with durable multilineage contribution across such heterogeneous pool of HSCs. With all unbiased molecular screening tools relying on destructive assays, such an approach is currently not possible. In this thesis, the use of single cell-derived clones has provided the closest approximation, as it permits simultaneous gene expression and transplantation assays using the progeny of the same cell. By linking transcriptomic profiles with a positive repopulation outcome, we were able to derive the RepopSig which, in turn, was able to identify LTHSCs irrespective of their cell cycle status and developmental origin.

## Chapter 6. Discussion

### 6.1. Summary of major findings

During this thesis, I set out to identify and characterise the molecular machinery of HSC self-renewal. These efforts were focused on both, in vivo resident HSCs and ex vivo expanded functional HSCs. While, functional assays or perturbation studies, coupled with transcriptomic profiling greatly enhanced our understanding of the in vivo molecular pathways governing HSC self-renewal, the limited success in replicating long-term symmetric HSC self-renewal ex vivo suggests that the process remains only partially understood. In the absence of such long-term ex vivo culturing protocols, the limits of our understanding of HSC self-renewal regulation in the ex vivo setting are further exacerbated.

To address the above questions, we set out to address three distinct unmet needs: 1) Develop a new method to allow characterisation of the HSC proteome using lower cell input (Chapter 3); 2) characterise the proteome of HSCs differing only in their selfrenewal capacity (Chapter 4); and 3) develop a novel reporter system to allow dissection of clonal heterogeneity in HSC expansion cultures (Chapter 5).

In Chapter 3, we developed an LC-MS/MS protocol to screen as few as 10,000 cells in a multiplex of up to 11 samples. Using the protocol, over 4,000 proteins can be reliably quantified, presenting a 20 -fold reduction of protein input against conventional protocols.

In Chapter 4, the above LC-MS/MS protocol was applied to LT-HSCs, freshly isolated from wild-type and TET2-deficient mice and only differing in their self-renewal potency. The analysis revealed progressive remodelling of the extracellular matrix and their complementary integrins. In addition, TET2-deficient cells were desensitised towards increasing $\mathrm{Ca}^{2+}$ concentrations, which was accompanied by the loss of key $\mathrm{Ca}^{2+}$ regulatory proteins. The transcriptome and proteome of both wild-type and TET2deficient HSCs were not correlated. Concurrent analysis of both -omes indicated that
shear stress regulators were ubiquitously enriched in cell fractions with higher selfrenewal potency.

Finally, in Chapter 5, we undertook molecular profiling of ex vivo expanded HSCs to generate a universal gene signature, with selective enrichment in repopulation and self-renewing HSCs. This signature enabled prospective isolation of self-renewing HSCs and was further complemented by the EPCR ${ }^{+}$Lin - Sca1 ${ }^{+}$cKit- (ELSK) cell surface phenotype. Importantly, the resultant gene signature also identified LT-HSCs irrespective of cell cycle status and origin.

### 6.2. The advent of rare cell and single cell proteomics

At the onset of our work, reliable LC-MS/MS protocols for screening rare cells were not available. To perform such screens, hundreds of thousands, or millions, of cells were required. To acquire such cell numbers using purified LT-HSC populations ${ }^{49}$, hundreds of mice would need to be pooled. Hence, we developed a protocol that reduced this required cell input 100 -fold. During my PhD, I also explored the development of a custom closed-loop microfluidic platform (data not shown), but for the purposes of this thesis, I focused on a developing a protocol that could be widely, and immediately, replicated by others across the globe.

We were encouraged to see the emergence of other rare cell protocols and most excitingly single cell proteins in recent years ${ }^{308,310,311,318}$. The proliferation of such protocols highlights the extent of the unmet need. Our protocol is on par with, and in many cases outperforms, these rare cell protocols ${ }^{308,310,311,318}$. Interestingly, since the inception of single cell transcriptomics, a plethora of novel single cell -omic screens have emerged and been integrated across multiple dimensions (extensively reviewed in our recent review ${ }^{596}$ ) but very few of these are currently capable of integrating protein information. Of these, CITE-seq, ECCITE-seq and Perturb-CITE-seq record protein information by capturing a selection of cell surface proteins ${ }^{597-599}$. The advent of rare cell LC-MS/MS and most importantly single cell proteomics, such as SCoPE2 ${ }^{310}$, will undoubtedly revolutionise the landscape of multi-omics and is a very exciting space to watch develop.

### 6.3. The HSC proteome

The results in Chapter 3 demonstrate a striking absence of correlation between the proteome and transcriptome across all stem/progenitor cell fractions, which is in line with recent reports by Amon et al. and Zaro et al. ${ }^{327,328}$. Interestingly, HSCs have been shown to have a high rate of transcription but a disproportionally lower rate of protein synthesis ${ }^{330,331,463}$, suggesting that HSCs might reside in a transcriptionally primed state, capable of rapidly responding to external stimuli. However, this simultaneously suggests that the complete transcriptome of HSCs is comprised of numerous molecules that are not involved in the underlying molecular machinery of immediate HSC functions, such as maintenance and self-renewal. The likelihood of such a disconnect is further supported by the field's inability to reliably establish symmetric HSC self-renewal ex vivo. Collectively, these insights emphasise the critical need to inspect the proteome of HSCs to further resolve the molecular regulators of core HSC functions.

Further in this vein, and in contrast to transcriptomics, LC-MS/MS can detect proteins, irrespective of their source (i.e., not just what the HSCs are making). As an example, the work in Chapter 4 shows that the proteome of both wild-type and TET2-deficient LT-HSCs included a multitude of ECM proteins and their corresponding integrins. As extensively discussed in Chapter 4.2.16, the ECM of wild-type and TET2-deficient LTHSCs underwent significant remodelling. Interestingly, several key structural ECM proteins have not been shown to be expressed in HSCs. Their presence in the LCMS/MS data could indicate that the niche-secreted ECM proteins were captured by integrins on the HSC surface and raises the intriguing question of whether differential niche localisation between wild-type and TET2-deficient HSCs might be a contributing factor to the increased number of HSCs observed in TET2 KO mice. In addition, the screen in Chapter 4 revealed a set of putative targets for HSC self-renewal regulation, previously not implicated in the process and not visible using transcriptomic datasets. Despite evidence of $\mathrm{Ca}^{2+}$-mediated self-renewal regulation in TET2-deficient cells, the molecular mechanisms of the increased HSC number remain nearly completely unknown ${ }^{403}$. Further supporting additional thinking in the ECM space, mechanosensation and shear stress response have emerged as key pathways enriched in HSCs with enhanced self-renewal potency. The molecular machinery of
both such pathways has previously not been described in HSCs ${ }^{462,478,479}$ and the data in Chapter 4 provides a defined set of potential regulators. In addition, this provides a novel set of pathways and direct target proteins to uncover previously unknown mechanisms of clonal evolution and malignancy, induced by TET2 loss-of-function mutations.

This thesis identified a plethora of signalling pathways and biological functions for selfrenewal regulation. Several such pathways, for instance calcium-mediated signalling, have previously been implicated in HSC biology, but their mechanisms remain largely unknown. For others, such as shear stress and mechanosensation, their role in HSC biology remain elusive, with their molecular drivers being particularly unknown. In this thesis, I identified a defined set of putative target proteins to enable hypothesis-driven studies to uncover the underlying mechanism of enriched pathways. Notable examples include the calcium ion channels VDAC1, VDAC3 and ATP2B4, which are responsible to clear cytosolic calcium ${ }^{433-435}$. VDAC1 is of particular interest, due to its ability to induce apoptosis under sustained elevated intracellular calcium levels ${ }^{433,434}$. Similarly, several shear stress proteins such as KLF2 and MEF2 are of particular interest. MEF2 induces KLF2 expression, which in turn regulates self-renewal in embryonic stem cells ${ }^{480-482}$. These data provide a testable hypothesis to determine whether MEF2 and KLF2 can regulate HSC self-renewal or have a previously unknown function in the $T E T 2^{-1} \mathrm{HSCs}$.

### 6.4. Applications of molecularly defined ex vivo expanded HSCs

The identification of in vitro markers that track with functional HSC content, combined with a defined molecular signature that is consistent across multiple cellular states, raises the attractive prospect of performing prospective quality control of HSC expansion cultures.

Previous prospective isolation strategies have utilised cell surface marker profiles to dissect functionally distinct HSC subtypes but these have nearly exclusively focused on freshly isolated HSCs ${ }^{35}$. The efforts in Chapter 5 to define cell surface markers for functional HSC has revealed that the percentage, but not the total cell number,
retrospectively correlated with self-renewal capacity. Thus, these results indicate that hyperproliferating ex vivo cultures are largely comprised of differentiating progenitor cell populations. In turn, increased secretion of progenitor-mediated signals could potentiate the differentiation drive of an expanding HSC clone ${ }^{494,590}$. These observations are in line with previous reports of stem cell exhaustion and loss of longterm self-renewal potency upon extensive HSC proliferation ${ }^{388,593,594}$.

The observation that the proportion of FELSK (Fgd5 ${ }^{+} \mathrm{EPCR}^{+} \mathrm{Lin}^{-} \mathrm{Sca} 1^{+}{ }^{+} \mathrm{CKit}^{+}$) or ELSK (EPCR ${ }^{+}$Lin-Sca1 ${ }^{+}$cKit ${ }^{+}$) cells within a clone correlates with repopulating HSC content now enables the accumulation of large populations, enriched for functional HSCs. In turn, large pools of functional HSCs will enable the application of bulk molecular profiling and large multiplexed screens, previously viewed as impossible due to cell number restrictions. Most notably, high-throughput small molecule screening platforms, containing hundreds of thousand compounds, have proven immensely valuable in drug discovery ${ }^{600}$. Many severe haematological diseases, including malignancies, can be caused by HSC defects ${ }^{483,601}$. With the advent of HSC expansion, coupled with the enrichment of functional HSCs, can now facilitate the application of high-throughput screening for HSC disorders.

Additionally, a number of sequencing-based molecular profiling tools have been adapted for single-cell application ${ }^{596}$. However, these inevitably involve a compromise between cell number input and depth of -omic capture ${ }^{596}$. The ability to generate large pools of functional HSCs greatly enhances the scope of possible molecular profiling, such as proteomics, metabolomics and beyond.

### 6.5. Therapeutic utility of ex vivo expanded HSCs

As outlined previously, a plethora of acquired and inherited genetic disorders of the haematopoietic system can be address by gene therapies, thus raising the prospect of utilising corrected HSCs in the clinical setting ${ }^{483,601}$. To date, therapeutic efficacy has been limited by the heterogeneous output, requirements to freeze/thaw and QC cultures, and the progressive differentiation drive of ex vivo culture systems.

As demonstrated in this thesis, prospective qPCR-based monitoring of the expression of RepopSig genes reliably identifies which clones retained self-renewal potency and which did not. In the clinical setting, reliable quality control strategies could significantly improve therapeutic efficiency, while simultaneously reducing waste during the ex vivo expansion process. To test the prospect of translating the RepopSig as an effective quality control measure in the human system, we have begun to explore the enrichment of the RepopSig across a range of human haematopoietic stem and progenitor cell types, including LT-HSCs. The preliminary findings so far have suggested that the RepopSig is highly and selectively enriched in human LT-HSCs (data not shown).

In addition, the ELSK phenotype and its ability to mark ex vivo expanded functional HSCs could also become translationally valuable. Fares et al. have previously demonstrated that EPCR is a robust marker of human HSCs in vitro although it remains unable on its own to purify functional HSCs for the types of molecular assays described above ${ }^{486}$. Furthermore, preliminary data (unpublished) showed that an F12based culture system, developed by Wilkinson et al., can also expand human HSCs.

Limited access to matching donors and high risk of graft-versus-host disease, have limited the ability to treat severe inherited genetic disorder ${ }^{596}$. Pioneering work by Bordignon et al. in 1995 explored the use of corrected human HSPC as a replacement for tradition bone marrow donations ${ }^{602}$. However, such process required ex vivo culturing capabilities of human HSCs. The concept rapidly gained traction for a variety of inherited disease, raising the need for reliable tools to expand small pools of corrected HSPCs ${ }^{596}$. Collectively, the rapid progression of human HSC culturing strategies, paired with qPCR quality control and cell surface phenotyping, provide prospect for routine human HSC expansion at clinical scale.

### 6.6. Future Studies and concluding remarks

The characterisation of the molecular mechanisms driving self-renewal of expanded HSCs have not only provided an insight into regulatory pathways but also informed several practical tools to identify self-renewing HSCs. However, additional studies are
required to advance the utility of the FELSK/ELSK flow phenotype and RepopSig qPCR quality control screen.

Firstly, transplantation assays have been conducted using a constant percentage of a given clone. In order to more accurately quantify the content of functional HSCs, limiting dilution assays would be required. Next, the studies to test the utility of RepopSig genes as a prospective quality control tool have been conducted at a single timepoint. It will be intriguing to perform time-course experiments to determine the evolution of signature gene expression throughout the 28 -day process and to determine whether earlier identification of successfully expanding clones could be identified earlier. Early detection of differentiating clones could greatly increase efficiency of the culturing process and reduce waste. The enrichment of the RepopSig in human LT-HSCs raised the promise of applying the qPCR quality control approach in human cultures.

Next, as prefaced above, the disconnect between the transcriptome and proteome has emphasised the need for profiling the molecular machinery beyond gene expression. Several concurrent approaches can help achieve this goal, including performing bulk proteomic screens of ex vivo expanded HSCs which would take advantage of larger LT-HSC pools. In addition, utilising newly developed single cell proteomics technologies, such as SCoPE2 ${ }^{310}$, can provide additional information on the heterogeneity within a self-renewing HSC population. Furthermore, ribosome sequencing (Ribo-seq) would allow direct screening of protein synthesis, identification of open reading frames and deciphering of the dynamic nature of translation ${ }^{603}$. Until recently, bulk Ribo-seq approaches required significant cell input ${ }^{604}$. However, VanInsBerghe et al. have now pioneered the first single cell Ribo-seq approach, without significant loos of resolution ${ }^{605}$. In the context of HSC protein synthesis, single cell Ribo-seq could provide an immensely valuable tool to screen HSC translation and integrate with the proteome to decipher HSC self-renewal.

The multi-omic characterisation of wild-type and TET2-deficent HSCs provided a novel set of putative self-renewal drivers. However, it remains of paramount importance to distinguish true driver genes from correlated, but functionally unrelated, target genes. Functional knockout screens paired with transplantation assays provide a robust
framework to identify true self-renewal regulators, as demonstrated by the in vivo CRISPR screening protocol described by Rodriguez-Fraticelli et al. ${ }^{595}$. First, LT-HSCs with doxycycline-inducible Cas9 expression are transduced with a sgRNA library, targeting the set of putative self-renewal drivers. In turn, transduced LT-HSCs are subject to transplantation and doxycycline treatment, inducing Cas9 expression to disrupt target gene expression. Ultimately, single cell RNA-seq profiling of harvested bone marrow cells enables the identification of sgRNAs and link their functional consequences with the transcriptome to identify true self-renewal regulators amongst the putative targets. Once self-renewal drivers are identified, their importance in human LT-HSC self-renewal can be inspected using xenotransplantation-based functional screening, as outlined by Wagenblast et al. ${ }^{606}$ In addition, the proteomic characterisation of wild-type and TET2-deficent HSCs possibly implicated niche location in the progression of MPN diseases. Hence, it would be critically important to inspect the localisation of LT-HSCs in wild-type and TET2-l- mice. Existing confocal imaging protocols for screening BM cross sections can be utilised to simultaneously stain ECM proteins and HSCs ${ }^{487,607}$. Furthermore, functional assays to test HSC migration and adhesion will further provide an insight into the cellular mechanism. Such experiments can help determine whether TET2-- HSCs occupy a different niche and acquire a stem cell advantage due to a change in stimulation by the niche.

Overall, this PhD thesis explored the molecular pathways of HSC self-renewal across multiple -omic dimensions and wide range of cell origins. By integrating wild-type and TET2-deficnent genotypes, screening several closely related HSC populations (differing in their self-renewal capacity only) and recording their proteomic and transcriptomic profiles, I was able to identify novel putative regulatory pathways of HSC self-renewal. In addition, the characterisation of ex vivo self-renewal not only provided an insight into the underling regulation of HSC proliferation but also yielded several practical applications, most notably a set of robust surface markers for ex vivo expanded HSCs and a repopulation-specific gene signature capable of prospectively identifying clones with long-term self-renewal durability. Most intriguingly, several preliminary studies have raised the prospect of translating the prospective screening tool into the human system. Ultimately, such quality control system for reliably identifying truly functional human HSCs can help achieve the goal of limitless access to HSCs for gene therapies, BM transplantations and beyond.

## References

1. Orkin, S. H. \& Zon, L. I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell 132, 631-644 (2008).
2. Doulatov, S. et al. Hematopoiesis: a human perspective. Cell Stem Cell 10, 120-36 (2012).
3. Kondo, M. Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. Immunol. Rev. 238, 37-46 (2010).
4. Seita, J. \& Weissman, I. L. Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip. Rev. Syst. Biol. Med. 2, 640-53 (2010).
5. Suda, J., Suda, T. \& Ogawa, M. Analysis of differentiation of mouse hemopoietic stem cells in culture by sequential replating of paired progenitors. Blood 64, 393-9 (1984).
6. Wilson, A. \& Trumpp, A. Bone-marrow haematopoietic-stem-cell niches. Nat. Rev. Immunol. 6, 93-106 (2006).
7. Ford, C. E., Hamerton, J. L., Barnes, D. W. \& Loutit, J. F. Cytological identification of radiation-chimaeras. Nature 177, 452-4 (1956).
8. Jacobson, L. O., Simmons, E. L., Marks, E. K. \& Eldredge, J. H. Recovery from radiation injury. Science 113, 510-11 (1951).
9. Thomas, E. D., Lochte, H. L., Cannon, J. H., Sahler, O. D. \& Ferrebee, J. W. SUPRALETHAL WHOLE BODY IRRADIATION AND ISOLOGOUS MARROW TRANSPLANTATION IN MAN*? J. Clin. Invest. 38, 1709-1716 (1959).
10. Thomas, E. D., Lochte, H. L., Lu, W. C. \& Ferrebee, J. W. Intravenous Infusion of Bone Marrow in Patients Receiving Radiation and Chemotherapy. N. Engl. J. Med. 257, 491-496 (1957).
11. Till, J. E. \& McCulloch, E. A. A Direct Measurement of the Radiation Sensitivity of Normal Mouse Bone Marrow Cells. Radiat. Res. 14, 213 (1961).
12. TILL, J. E., MCCULLOCH, E. A. \& SIMINOVITCH, L. A STOCHASTIC MODEL OF STEM CELL PROLIFERATION, BASED ON THE GROWTH OF SPLEEN COLONY-FORMING CELLS. Proc. Natl. Acad. Sci. U. S. A. 51, 29-36 (1964).
13. Abramson, S., Miller, R. G. \& Phillips, R. A. The identification in adult bone marrow of pluripotent and restricted stem cells of the myeloid and lymphoid systems. J. Exp. Med. 145, 1567-79 (1977).
14. Osawa, M., Hanada, K., Hamada, H. \& Nakauchi, H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science 273, 242-5 (1996).
15. Woolthuis, C. M. \& Park, C. Y. Hematopoietic stem/progenitor cell commitment to the megakaryocyte lineage. Blood 127, 1242-8 (2016).
16. Akashi, K., Traver, D., Miyamoto, T. \& Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature 404, 193-197 (2000).
17. Kondo, M., Weissman, I. L. \& Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. Cell 91, 661-72 (1997).
18. Sanjuan-Pla, A. et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. Nat. 20135027470 502, 232-236 (2013).
19. Dykstra, B. et al. Long-Term Propagation of Distinct Hematopoietic Differentiation Programs In Vivo. Cell Stem Cell 1, 218-229 (2007).
20. Yoshida, T., Yao-Ming Ng, S., Zuniga-Pflucker, J. C. \& Georgopoulos, K. Early
hematopoietic lineage restrictions directed by Ikaros. Nat. Immunol. 7, 382391 (2006).
21. Roch, A., Trachsel, V. \& Lutolf, M. P. Brief Report: Single-Cell Analysis Reveals Cell Division-Independent Emergence of Megakaryocytes From Phenotypic Hematopoietic Stem Cells. Stem Cells 33, 3152-3157 (2015).
22. Nestorowa, S. et al. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood 128, (2016).
23. Laurenti, E. \& Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. Nature 553, 418-426 (2018).
24. Stegle, O., Teichmann, S. A. \& Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. Nat. Rev. Genet. 16, 133-145 (2015).
25. Phipson, B., Zappia, L. \& Oshlack, A. Gene length and detection bias in single cell RNA sequencing protocols. F1000Research 6, (2017).
26. Laurenti, E. \& Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. Nature vol. 553 418-426 (2018).
27. Medvinsky, A. \& Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. Cell 86, 897-906 (1996).
28. Moore, M. A. \& Metcalf, D. Ontogeny of the haemopoietic system: yolk sac origin of in vivo and in vitro colony forming cells in the developing mouse embryo. Br. J. Haematol. 18, 279-96 (1970).
29. Müller, A. M., Medvinsky, A., Strouboulis, J., Grosveld, F. \& Dzierzak, E. Development of hematopoietic stem cell activity in the mouse embryo. Immunity 1, 291-301 (1994).
30. Taoudi, S. et al. Extensive Hematopoietic Stem Cell Generation in the AGM Region via Maturation of VE-Cadherin+CD45+ Pre-Definitive HSCs. Cell Stem Cell 3, 99-108 (2008).
31. Medvinsky, A., Rybtsov, S. \& Taoudi, S. Embryonic origin of the adult hematopoietic system: advances and questions. Development 138, 10171031 (2011).
32. Bowie, M. B. et al. Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. J. Clin. Invest. 116, 2808-2816 (2006).
33. Mikkola, H. K. A. \& Orkin, S. H. The journey of developing hematopoietic stem cells. Development 133, (2006).
34. McKinney-Freeman, S. et al. The transcriptional landscape of hematopoietic stem cell ontogeny. Cell Stem Cell 11, 701-14 (2012).
35. Kent, D. G. et al. Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. Blood 113, (2009).
36. Copley, M. R. et al. The Lin28b-let-7-Hmga2 axis determines the higher selfrenewal potential of fetal haematopoietic stem cells. Nat. Cell Biol. 15, 916-25 (2013).
37. He, S., Kim, I., Lim, M. S. \& Morrison, S. J. Sox17 expression confers selfrenewal potential and fetal stem cell characteristics upon adult hematopoietic progenitors. Genes Dev. 25, 1613 (2011).
38. Dick, J. E., Magli, M. C., Huszar, D., Phillips, R. A. \& Bernstein, A. Introduction of a Selectable Gene into Primitive Stem Cells Capable of Long-Term Reconstitution of the Hemopoietic System of W/W v Mice. Cell 42, 71-79 (1985).
39. Keller, G., Paige, C., Gilboa, E. \& Wagner, E. F. Expression of a foreign gene
in myeloid and lymphoid cells derived from multipotent haematopoietic precursors. Nature 318, 149-154 (1985).
40. Lemischka, I. R., Raulet, D. H. \& Mulligan, R. C. Developmental potential and dynamic behavior of hematopoietic stem cells. Cell 45, 917-27 (1986).
41. Goldschneider, I., Gordon, L. K. \& Morris, R. J. Demonstration of Thy-1 antigen on pluripotent hemopoietic stem cells in the rat. J. Exp. Med. 148, (1978).
42. Muller-Sieburg, C. E., Whitlock, C. A. \& Weissman, I. L. Isolation of two early B lymphocyte progenitors from mouse marrow: a committed pre-pre-B cell and a clonogenic Thy-1-lo hematopoietic stem cell. Cell 44, 653-62 (1986).
43. Uchida, N. \& Weissman, I. L. Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. J. Exp. Med. 175, 175-84 (1992).
44. Spangrude, G. J., Heimfeld, S. \& Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. Science 241, 58-62 (1988).
45. Chabot, B., Stephenson, D. A., Chapman, V. M., Besmer, P. \& Bernstein, A. The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. Nature 335, 88-89 (1988).
46. Okada, S. et al. In vivo and in vitro stem cell function of c-kit- and Sca-1positive murine hematopoietic cells. Blood 80, 3044-50 (1992).
47. Kiel, M. J. et al. SLAM Family Receptors Distinguish Hematopoietic Stem and Progenitor Cells and Reveal Endothelial Niches for Stem Cells. Cell 121, 1109-1121 (2005).
48. Balazs, A. B., Fabian, A. J., Esmon, C. T. \& Mulligan, R. C. Endothelial protein C receptor (CD201) explicitly identifies hematopoietic stem cells in murine bone marrow. Blood 107, 2317-21 (2006).
49. Kent, D. G., Dykstra, B. J., Cheyne, J., Ma, E. \& Eaves, C. J. Steel factor coordinately regulates the molecular signature and biologic function of hematopoietic stem cells. Blood 112, (2008).
50. Benz, C. et al. Cell Stem Cell Hematopoietic Stem Cell Subtypes Expand Differentially during Development and Display Distinct Lymphopoietic Programs. Stem Cell 10, 273-283 (2012).
51. Wilson, N. K. et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell 16, 712-724 (2015).
52. Lin, K. K. \& Goodell, M. A. Purification of Hematopoietic Stem Cells Using the Side Population. in Methods in enzymology vol. 420 255-264 (2006).
53. Benveniste, P. et al. Intermediate-Term Hematopoietic Stem Cells with Extended but Time-Limited Reconstitution Potential. Cell Stem Cell 6, 48-58 (2010).
54. van Pel, M., Fibbe, W. E. \& Schepers, K. The human and murine hematopoietic stem cell niches: are they comparable? Ann. N. Y. Acad. Sci. 1370, 55-64 (2016).
55. Notta, F. et al. Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment. Science (80-. ). 333, 218-221 (2011).
56. Ema, H., Morita, Y. \& Suda, T. Heterogeneity and hierarchy of hematopoietic stem cells. Exp. Hematol. 42, 74-82.e2 (2014).
57. Crisan, M. \& Dzierzak, E. The many faces of hematopoietic stem cell heterogeneity. Development 143, (2016).
58. Müller-Sieburg, C. E., Cho, R. H., Thoman, M., Adkins, B. \& Sieburg, H. B. Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. Blood 100, 1302-9 (2002).
59. Yamamoto, R. et al. Clonal Analysis Unveils Self-Renewing LineageRestricted Progenitors Generated Directly from Hematopoietic Stem Cells. Cell 154, 1112-1126 (2013).
60. Muller-Sieburg, C. E. \& Sieburg, H. B. Clonal diversity of the stem cell compartment. Curr. Opin. Hematol. 13, 243-248 (2006).
61. Sieburg, H. B. et al. The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets. Blood 107, 2311-2316 (2006).
62. Muller-Sieburg, C. E., Cho, R. H., Karlsson, L., Huang, J.-F. \& Sieburg, H. B. Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. Blood 103, 4111-4118 (2004).
63. Dykstra, B. et al. High-resolution video monitoring of hematopoietic stem cells cultured in single-cell arrays identifies new features of self-renewal. Proc. Natl. Acad. Sci. U. S. A. 103, 8185-90 (2006).
64. Uchida, N., Dykstra, B., Lyons, K. J., Leung, F. Y. K. \& Eaves, C. J. Different in vivo repopulating activities of purified hematopoietic stem cells before and after being stimulated to divide in vitro with the same kinetics. Exp. Hematol. 31, 1338-47 (2003).
65. Copley, M. R., Beer, P. A. \& Eaves, C. J. Hematopoietic Stem Cell Heterogeneity Takes Center Stage. Cell Stem Cell 10, 690-697 (2012).
66. Sudo, K., Ema, H., Morita, Y. \& Nakauchi, H. Age-associated characteristics of murine hematopoietic stem cells. J. Exp. Med. 192, 1273-80 (2000).
67. Cho, R. H., Sieburg, H. B. \& Muller-Sieburg, C. E. 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, 5553-5561 (2008).
68. Weksberg, D. C., Chambers, S. M., Boles, N. C. \& Goodell, M. A. CD150- side population cells represent a functionally distinct population of long-term hematopoietic stem cells. Blood 111, 2444-51 (2008).
69. Schulte, R. et al. Index sorting resolves heterogeneous murine hematopoietic stem cell populations. Exp. Hematol. 43, 803-811 (2015).
70. Morita, Y., Ema, H., Yamazaki, S. \& Nakauchi, H. Non-side-population hematopoietic stem cells in mouse bone marrow. Blood 108, 2850-2856 (2006).
71. Morita, Y., Ema, H. \& Nakauchi, H. Heterogeneity and hierarchy within the most primitive hematopoietic stem cell compartment. J. Exp. Med. 207, 11731182 (2010).
72. Beerman, I. et al. Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proc. Natl. Acad. Sci. 107, 5465-5470 (2010).
73. Goodell, M. A., Brose, K., Paradis, G., Conner, A. S. \& Mulligan, R. C. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. J. Exp. Med. 183, 1797-806 (1996).
74. Challen, G. A., Boles, N. C., Chambers, S. M. \& Goodell, M. A. Distinct Hematopoietic Stem Cell Subtypes Are Differentially Regulated by TGF- $\beta 1$. Cell Stem Cell 6, 265-278 (2010).
75. Oguro, H., Ding, L. \& Morrison, S. J. SLAM family markers resolve functionally
distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell Stem Cell 13, 102-16 (2013).
76. Huang, X., Cho, S. \& Spangrude, G. J. Hematopoietic stem cells: generation and self-renewal. Cell Death Differ. 14, 1851-1859 (2007).
77. $\mathrm{R}, \mathrm{S}$. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. Blood Cells 4, 7-25 (1978).
78. Pinho, S. \& Frenette, P. S. Haematopoietic stem cell activity and interactions with the niche. Nat. Rev. Mol. Cell Biol. 2019205 20, 303-320 (2019).
79. Morrison, S. J. \& Scadden, D. T. The bone marrow niche for haematopoietic stem cells. Nat. 20145057483 505, 327-334 (2014).
80. A, B. \& PS, F. Niche heterogeneity in the bone marrow. Ann. N. Y. Acad. Sci. 1370, 82-96 (2016).
81. M, M., S, T. \& PS, F. Neural Regulation of Bone and Bone Marrow. Cold Spring Harb. Perspect. Med. 8, (2018).
82. Madel, M.-B. et al. Immune Function and Diversity of Osteoclasts in Normal and Pathological Conditions. Front. Immunol. 0, 1408 (2019).
83. Y, K. et al. Arteriolar niches maintain haematopoietic stem cell quiescence. Nature 502, 637-643 (2013).
84. EC, W. \& RH, A. Biology of Bone: The Vasculature of the Skeletal System. Cold Spring Harb. Perspect. Med. 8, (2018).
85. Winkler, I. G. et al. Bone marrow macrophages maintain hematopoietic stem cell (HSC) niches and their depletion mobilizes HSCs. Blood 116, 4815-4828 (2010).
86. K, Y. \& TD, A. Ultrastructural morphometric study of efferent nerve terminals on murine bone marrow stromal cells, and the recognition of a novel anatomical unit: the 'neuro-reticular complex'. Am. J. Anat. 187, 261-276 (1990).
87. S, M.-F. et al. Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. Nature 466, 829-834 (2010).
88. Katayama, Y. et al. Signals from the Sympathetic Nervous System Regulate Hematopoietic Stem Cell Egress from Bone Marrow. Cell 124, 407-421 (2006).
89. Taichman, R. S. Blood and bone: two tissues whose fates are intertwined to create the hematopoietic stem-cell niche. Blood 105, 2631-2639 (2005).
90. Moore, K. A. \& Lemischka, I. R. Stem Cells and Their Niches. Science (80-. ). 311, 1880-1885 (2006).
91. Arai, F. et al. Tie2/Angiopoietin-1 Signaling Regulates Hematopoietic Stem Cell Quiescence in the Bone Marrow Niche. Cell 118, 149-161 (2004).
92. Wilson, A. et al. c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. Genes Dev. 18, 2747-63 (2004).
93. Sugiyama, T., Kohara, H., Noda, M. \& Nagasawa, T. Maintenance of the Hematopoietic Stem Cell Pool by CXCL12-CXCR4 Chemokine Signaling in Bone Marrow Stromal Cell Niches. Immunity 25, 977-988 (2006).
94. De Grandis, M., Lhoumeau, A.-C., Mancini, S. J. C. \& Aurrand-Lions, M. Adhesion receptors involved in HSC and early-B cell interactions with bone marrow microenvironment. Cell. Mol. Life Sci. 2015734 73, 687-703 (2015).
95. Domingues, M. J., Cao, H., Heazlewood, S. Y., Cao, B. \& Nilsson, S. K. Niche Extracellular Matrix Components and Their Influence on HSC. J. Cell. Biochem. 118, 1984-1993 (2017).
96. SK, N. et al. Osteopontin, a key component of the hematopoietic stem cell
niche and regulator of primitive hematopoietic progenitor cells. Blood 106, 1232-1239 (2005).
97. SK, N. et al. Immunofluorescence characterization of key extracellular matrix proteins in murine bone marrow in situ. J. Histochem. Cytochem. 46, 371-377 (1998).
98. S, K. \& C, V. The role of novel and known extracellular matrix and adhesion molecules in the homeostatic and regenerative bone marrow microenvironment. Cell Adh. Migr. 8, 563-577 (2014).
99. JS, C. \& BA, H. Marrow-inspired matrix cues rapidly affect early fate decisions of hematopoietic stem and progenitor cells. Sci. Adv. 3, (2017).
100. Dao, M. A., Hashino, K., Kato, I. \& Nolta, J. A. Adhesion to Fibronectin Maintains Regenerative Capacity During Ex Vivo Culture and Transduction of Human Hematopoietic Stem and Progenitor Cells. Blood 92, 4612-4621 (1998).
101. Klein, G., Müller, C. A., Tillet, E., Chu, M. L. \& Timpl, R. Collagen Type VI in the Human Bone Marrow Microenvironment: A Strong Cytoadhesive Component. Blood 86, 1740-1748 (1995).
102. Oswald, J. et al. Gene-Expression Profiling of CD34+ Hematopoietic Cells Expanded in a Collagen I Matrix. Stem Cells 24, 494-500 (2006).
103. J, G. et al. Thrombin-cleaved osteopontin regulates hemopoietic stem and progenitor cell functions through interactions with alpha9beta1 and alpha4beta1 integrins. Blood 114, 49-59 (2009).
104. H, C. et al. Osteopontin is An Important Regulative Component of the Fetal Bone Marrow Hematopoietic Stem Cell Niche. Cells 8, (2019).
105. Gu, Y.-C. et al. Laminin isoform-specific promotion of adhesion and migration of human bone marrow progenitor cells. Blood 101, 877-885 (2003).
106. Sagar, B. M. M., Rentala, S., Gopal, P. N. V., Sharma, S. \& Mukhopadhyay, A. Fibronectin and laminin enhance engraftibility of cultured hematopoietic stem cells. Biochem. Biophys. Res. Commun. 350, 1000-1005 (2006).
107. Susek, K. H. et al. Bone marrow laminins influence hematopoietic stem and progenitor cell cycling and homing to the bone marrow. Matrix Biol. 67, 47-62 (2018).
108. SK, N. et al. Hyaluronan is synthesized by primitive hemopoietic cells, participates in their lodgment at the endosteum following transplantation, and is involved in the regulation of their proliferation and differentiation in vitro. Blood 101, 856-862 (2003).
109. Zöller, M. CD44, Hyaluronan, the Hematopoietic Stem Cell, and LeukemiaInitiating Cells. Front. Immunol. 0, 235 (2015).
110. Gordon, M. Y. \& Amos, T. A. S. Stochastic effects in hemopoiesis. Stem Cells 12, 175-179 (1994).
111. Alenzi, F. Q. et al. The haemopoietic stem cell: between apoptosis and self renewal. Yale J. Biol. Med. 82, 7-18 (2009).
112. Wilkinson, A. C., Igarashi, K. J. \& Nakauchi, H. Haematopoietic stem cell selfrenewal in vivo and ex vivo. Nat. Rev. Genet. 2020219 21, 541-554 (2020).
113. Ding, L., Saunders, T. L., Enikolopov, G. \& Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. Nature 481, 457-462 (2012).
114. Barker, J. E. Early transplantation to a normal microenvironment prevents the development of Steel hematopoietic stem cell defects. Exp. Hematol. 25, 5427 (1997).
115. Hassan, H. T. \& Zander, A. Stem cell factor as a survival and growth factor in human normal and malignant hematopoiesis. Acta Haematol. 95, 257-62 (1996).
116. Ikuta, K. \& Weissman, I. L. Evidence that hematopoietic stem cells express mouse c-kit but do not depend on steel factor for their generation. Proc. Natl. Acad. Sci. U. S. A. 89, 1502-6 (1992).
117. MB, B., DG, K., MR, C. \& CJ, E. Steel factor responsiveness regulates the high self-renewal phenotype of fetal hematopoietic stem cells. Blood 109, 5043-5048 (2007).
118. Nocka, K. et al. Molecular bases of dominant negative and loss of function mutations at the murine c-kit/white spotting locus: W37, Wv, W41 and W. EMBO J. 9, 1805-1813 (1990).
119. Thorén, L. A. et al. Kit Regulates Maintenance of Quiescent Hematopoietic Stem Cells. J. Immunol. 180, (2008).
120. Y, T. et al. Spred1 Safeguards Hematopoietic Homeostasis against DietInduced Systemic Stress. Cell Stem Cell 22, 713-725.e8 (2018).
121. CC, Z. \& HF, L. Cytokines regulating hematopoietic stem cell function. Curr. Opin. Hematol. 15, 307-311 (2008).
122. de Sauvage, F. J. et al. Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand. Nature 369, 533-538 (1994).
123. Sungaran, R., Markovic, B. \& Chong, B. H. Localization and Regulation of Thrombopoietin mRNA Expression in Human Kidney, Liver, Bone Marrow, and Spleen Using In Situ Hybridization. Blood 89, 101-107 (1997).
124. N, B.-V. et al. Cytokines regulate postnatal hematopoietic stem cell expansion: opposing roles of thrombopoietin and LNK. Genes Dev. 20, 2018-2023 (2006).
125. S, K., AW, R., D, M. \& WS, A. Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin. Proc. Natl. Acad. Sci. U. S. A. 95, 1195-1200 (1998).
126. N, F., G, P., T, P. \& K, K. Thrombopoietin expands hematopoietic stem cells after transplantation. J. Clin. Invest. 110, 389-394 (2002).
127. JL, A. \& J, C. Studies of c-Mpl function distinguish the replication of hematopoietic stem cells from the expansion of differentiating clones. Blood 109, 5186-5190 (2007).
128. Solar, G. P. et al. Role of c-mpl in Early Hematopoiesis. Blood 92, 4-10 (1998).
129. Seita, J. \& Weissman, I. L. Hematopoietic stem cell: self-renewal versus differentiation. Wiley Interdiscip. Rev. Syst. Biol. Med. 2, 640-53 (2010).
130. Ema, H. et al. Quantification of Self-Renewal Capacity in Single Hematopoietic Stem Cells from Normal and Lnk-Deficient Mice. Dev. Cell 8, 907-914 (2005).
131. J, S. et al. Lnk negatively regulates self-renewal of hematopoietic stem cells by modifying thrombopoietin-mediated signal transduction. Proc. Natl. Acad. Sci. U. S. A. 104, 2349-2354 (2007).
132. Graaf, C. A. de \& Metcalf, D. Thrombopoietin and hematopoietic stem cells. http://dx.doi.org/10.4161/cc.10.10.15619 10, 1582-1589 (2011).
133. Arai, F. et al. Tie2/Angiopoietin-1 Signaling Regulates Hematopoietic Stem Cell Quiescence in the Bone Marrow Niche. Cell 118, 149-161 (2004).
134. Arai, F., Ohneda, O., Miyamoto, T., Zhang, X. Q. \& Suda, T. Mesenchymal stem cells in perichondrium express activated leukocyte cell adhesion molecule and participate in bone marrow formation. J. Exp. Med. 195, 154963 (2002).
135. N, A. et al. Differential cytokine contributions of perivascular haematopoietic stem cell niches. Nat. Cell Biol. 19, 214-223 (2017).
136. L, D. \& SJ, M. Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. Nature 495, 231-235 (2013).
137. A, G. et al. CXCL12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. Nature 495, 227-230 (2013).
138. Massagué, J. TGF $\beta$ signalling in context. Nat. Rev. Mol. Cell Biol. 20121310 13, 616-630 (2012).
139. Blank, U. \& Karlsson, S. TGF- $\beta$ signaling in the control of hematopoietic stem cells. Blood 125, 3542-3550 (2015).
140. Garbe, A. et al. Transforming growth factor-beta 1 delays formation of granulocyte-macrophage colony-forming cells, but spares more primitive progenitors during ex vivo expansion of CD34+ haemopoietic progenitor cells. Br. J. Haematol. 99, 951-958 (1997).
141. Batard, P. et al. TGF-(beta) 1 maintains hematopoietic immaturity by a reversible negative control of cell cycle and induces CD34 antigen upmodulation. J. Cell Sci. 113, 383-390 (2000).
142. Transforming Growth Factor $\beta i$ Directly and Reversibly Inhibits the Initial Cell Divisions of Long-Term Repopulating Hematopoietic Stem Cells. Blood 88, 82-88 (1996).
143. Yamazaki, S. et al. TGF- $\beta$ as a candidate bone marrow niche signal to induce hematopoietic stem cell hibernation. Blood 113, 1250-1256 (2009).
144. Capron, C. et al. A major role of TGF- $\beta 1$ in the homing capacities of murine hematopoietic stem cell/progenitors. Blood 116, 1244-1253 (2010).
145. Pietras, E. M. Inflammation: a key regulator of hematopoietic stem cell fate in health and disease. Blood 130, 1693-1698 (2017).
146. I, P. et al. G-CSF induces stem cell mobilization by decreasing bone marrow SDF-1 and up-regulating CXCR4. Nat. Immunol. 3, 687-694 (2002).
147. Miller, C. L. \& Eaves, C. J. Expansion in vitro of adult murine hematopoietic stem cells with transplantable lympho-myeloid reconstituting ability. Proc. Natl. Acad. Sci. U. S. A. 94, 13648-53 (1997).
148. Musashi, M., Clark, S. C., Sudo, T., Urdal, D. L. \& Ogawa, M. Synergistic Interactions Between Interleukin-11 and Interleukin-4 in Support of Proliferation of Primitive Hematopoietic Progenitors of Mice. Blood 78, 14481451 (1991).
149. S, D. V. et al. The gp130 Cytokine Interleukin-11 Regulates Engraftment of Vav1 -/- Hematopoietic Stem and Progenitor Cells in Lethally Irradiated Recipients. Stem Cells 36, 446-457 (2018).
150. Hawley, R. G. et al. Thrombopoietic potential and serial repopulating ability of murine hematopoietic stem cells constitutively expressing interleukin 11. Proc. Natl. Acad. Sci. 93, 10297-10302 (1996).
151. A, B. et al. Interleukin-6 is required in vivo for the regulation of stem cells and committed progenitors of the hematopoietic system. Immunity 1, 725-731 (1994).
152. K, Y. et al. Targeted disruption of gp130, a common signal transducer for the interleukin 6 family of cytokines, leads to myocardial and hematological disorders. Proc. Natl. Acad. Sci. U. S. A. 93, 407-411 (1996).
153. Audet, J., Miller, C. L., Rose-John, S., Piret, J. M. \& Eaves, C. J. Distinct role of gp130 activation in promoting self-renewal divisions by mitogenically stimulated murine hematopoietic stem cells. Proc. Natl. Acad. Sci. 98, 1757-

1762 (2001).
154. H, T. et al. Pathogen-Induced TLR4-TRIF Innate Immune Signaling in Hematopoietic Stem Cells Promotes Proliferation but Reduces Competitive Fitness. Cell Stem Cell 21, 225-240.e5 (2017).
155. Zhang, C. C. \& Lodish, H. F. Cytokines regulating hematopoietic stem cell function. Curr. Opin. Hematol. 15, 307-11 (2008).
156. Duncan, A. W. et al. Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. Nat. Immunol. 6, 314-322 (2005).
157. Chiba, S. Concise Review: Notch Signaling in Stem Cell Systems. Stem Cells 24, 2437-2447 (2006).
158. Calvi, L. M. et al. Osteoblastic cells regulate the haematopoietic stem cell niche. Nature 425, 841-846 (2003).
159. Nemeth, M. J., Topol, L., Anderson, S. M., Yang, Y. \& Bodine, D. M. Wnt5a inhibits canonical Wnt signaling in hematopoietic stem cells and enhances repopulation. Proc. Natl. Acad. Sci. 104, 15436-15441 (2007).
160. Kabiri, Z. et al. Whts are dispensable for differentiation and self-renewal of adult murine hematopoietic stem cells. Blood 126, 1086-94 (2015).
161. Pimanda, J. E. et al. Gata2, Fli1, and Scl form a recursively wired generegulatory circuit during early hematopoietic development. Proc. Natl. Acad. Sci. U. S. A. 104, 17692-7 (2007).
162. Tsai, F.-Y. et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature 371, 221-226 (1994).
163. Robb, L. et al. The scl gene product is required for the generation of all hematopoietic lineages in the adult mouse. EMBO J. 15, 4123-9 (1996).
164. Porcher, C. et al. The T Cell Leukemia Oncoprotein SCL/tal-1 Is Essential for Development of All Hematopoietic Lineages. Cell 86, 47-57 (1996).
165. Gottgens, B. et al. Establishing the transcriptional programme for blood: the SCL stem cell enhancer is regulated by a multiprotein complex containing Ets and GATA factors. EMBO J. 21, 3039-3050 (2002).
166. Maeno, M. et al. The role of BMP-4 and GATA-2 in the induction and differentiation of hematopoietic mesoderm in Xenopus laevis. Blood 88, 196572 (1996).
167. Oren, T., Torregroza, I. \& Evans, T. An Oct-1 binding site mediates activation of the gata2 promoter by BMP signaling. Nucleic Acids Res. 33, 4357-4367 (2005).
168. Walmsley, M., Ciau-Uitz, A. \& Patient, R. Adult and embryonic blood and endothelium derive from distinct precursor populations which are differentially programmed by BMP in Xenopus. Development 129, 5683-95 (2002).
169. Pimanda, J. E. \& Gttgens, B. Gene regulatory networks governing haematopoietic stem cell development and identity. Int. J. Dev. Biol. 54, 12011211 (2010).
170. Hamey, F. K. et al. Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. Proc. Natl. Acad. Sci. U. S. A. 114, 58225829 (2017).
171. Kosan, C. \& Godmann, M. Genetic and Epigenetic Mechanisms That Maintain Hematopoietic Stem Cell Function. Stem Cells Int. 2016, 1-14 (2016).
172. He, S. et al. Transient CDK4/6 inhibition protects hematopoietic stem cells from chemotherapy-induced exhaustion. Sci. Transl. Med. 9, (2017).
173. Maurer, B., Brandstoetter, T., Kollmann, S., Sexl, V. \& Prchal-Murphy, M. Inducible deletion of CDK4 and CDK6 - deciphering CDK4/6 inhibitor effects in
the hematopoietic system. Haematologica 106, 2624-2632 (2021).
174. Kamminga, L. M. et al. The Polycomb group gene Ezh2 prevents hematopoietic stem cell exhaustion. Blood 107, (2006).
175. Lessard, J. \& Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature 423, 255-260 (2003).
176. Cabezas-Wallscheid, N. et al. Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis. Cell Stem Cell 15, 507-522 (2014).
177. Manesia, J. K. et al. Highly proliferative primitive fetal liver hematopoietic stem cells are fueled by oxidative metabolic pathways. Stem Cell Res. 15, 715-721 (2015).
178. Bowie, M. B. et al. Identification of a new intrinsically timed developmental checkpoint that reprograms key hematopoietic stem cell properties. www.pnas.org/cgi/content/full/ (2007).
179. Kim, I., Saunders, T. L. \& Morrison, S. J. Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells. Cell 130, 470-83 (2007).
180. Moran-Crusio, K. et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell 20, 11-24 (2011).
181. M, J. et al. Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. Cell Rep. 23, 1-10 (2018).
182. Delhommeau, F. et al. Mutation in TET2 in Myeloid Cancers. N. Engl. J. Med. 360, 2289-2301 (2009).
183. Langemeijer, S. M. C. et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat. Genet. 41, 838-842 (2009).
184. Pardanani, A. et al. IDH1 and IDH2 mutation analysis in chronic- and blastphase myeloproliferative neoplasms. Leuk. 2010246 24, 1146-1151 (2010).
185. Stegelmann, F. et al. DNMT3A mutations in myeloproliferative neoplasms. Leuk. 2011257 25, 1217-1219 (2011).
186. V, G.-B. et al. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. J. Hematol. Oncol. 5, (2012).
187. Buscarlet, M. et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. Blood 130, 753-762 (2017).
188. DW, B., CE, F., PL, I., PC, K. \& JF, L. Tissue transplantation in the radiation chimera. J. Cell. Physiol. Suppl. 50, 123-138 (1957).
189. EA, C. Hematopoietic stem-cell transplantation. N. Engl. J. Med. 354, 18131826 (2006).
190. Remberger, M. et al. Effect of Total Nucleated and CD34+ Cell Dose on Outcome after Allogeneic Hematopoietic Stem Cell Transplantation. Biol. Blood Marrow Transplant. 21, 889-893 (2015).
191. Sagoo, P. \& Gaspar, H. B. The transformative potential of HSC gene therapy as a genetic medicine. Gene Ther. 2021 1-19 (2021) doi:10.1038/s41434-021-00261-x.
192. RA, M., D, G., A, L. \& DB, K. Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned. Cell Stem Cell 21, 574-590 (2017).
193. Tucci, F., Scaramuzza, S., Aiuti, A. \& Mortellaro, A. Update on Clinical Ex Vivo Hematopoietic Stem Cell Gene Therapy for Inherited Monogenic Diseases. Mol. Ther. 29, 489-504 (2021).
194. Pavani, G. et al. Ex vivo editing of human hematopoietic stem cells for
erythroid expression of therapeutic proteins. Nat. Commun. 2020 111 11, 1-13 (2020).
195. Li, C. et al. In Vivo HSC Gene Therapy for Hemoglobinopathies: A Proof of Concept Evaluation in Rhesus Macaques. Blood 136, 46-47 (2020).
196. Nolta, J. et al. The AFT024 stromal cell line supports long-term ex vivo maintenance of engrafting multipotent human hematopoietic progenitors. Leuk. 2002163 16, 352-361 (2002).
197. Moore, K. A., Ema, H. \& Lemischka, I. R. In Vitro Maintenance of Highly Purified, Transplantable Hematopoietic Stem Cells. Blood 89, 4337-4347 (1997).
198. F, N. et al. Engineering a haematopoietic stem cell niche by revitalizing mesenchymal stromal cells. Nat. Cell Biol. 21, 560-567 (2019).
199. RA, O. et al. Long-term maintenance of hematopoietic stem cells does not require contact with embryo-derived stromal cells in cocultures. Stem Cells 23, 842-851 (2005).
200. Overexpression of HOXB4 Enhances the Hematopoietic Potential of Embryonic Stem Cells Differentiated In Vitro. Blood 87, 2740-2749 (1996).
201. Antonchuk, J., Sauvageau, G. \& Humphries, R. K. HOXB4-Induced Expansion of Adult Hematopoietic Stem Cells Ex Vivo. Cell 109, 39-45 (2002).
202. T, R. et al. A role for Wnt signalling in self-renewal of haematopoietic stem cells. Nature 423, 409-414 (2003).
203. A, K. et al. HES-1 preserves purified hematopoietic stem cells ex vivo and accumulates side population cells in vivo. Blood 101, 1777-1783 (2003).
204. KY, C. et al. Enforced expression of NUP98-HOXA9 in human CD34(+) cells enhances stem cell proliferation. Cancer Res. 66, 11781-11791 (2006).
205. E, K., U, T., N, M., T, N. \& G, S. NUP98-HOXA9 expression in hemopoietic stem cells induces chronic and acute myeloid leukemias in mice. EMBO J. 20, 350-361 (2001).
206. Bourgine, P. E. et al. In vitro biomimetic engineering of a human hematopoietic niche with functional properties. Proc. Natl. Acad. Sci. 115, E5688-E5695 (2018).
207. A, T. et al. Ex vivo expansion of haematopoietic stem/progenitor cells from human umbilical cord blood on acellular scaffolds prepared from MS-5 stromal cell line. J. Tissue Eng. Regen. Med. 7, 871-883 (2013).
208. Bai, T. et al. Expansion of primitive human hematopoietic stem cells by culture in a zwitterionic hydrogel. Nat. Med. 20192510 25, 1566-1575 (2019).
209. Y, T. et al. Spred1 Safeguards Hematopoietic Homeostasis against DietInduced Systemic Stress. Cell Stem Cell 22, 713-725.e8 (2018).
210. CC, Z. \& HF, L. Cytokines regulating hematopoietic stem cell function. Curr. Opin. Hematol. 15, 307-311 (2008).
211. CL, M. \& CJ, E. Expansion in vitro of adult murine hematopoietic stem cells with transplantable lympho-myeloid reconstituting ability. Proc. Natl. Acad. Sci. U. S. A. 94, 13648-13653 (1997).
212. T, U. et al. Expansion of human NOD/SCID-repopulating cells by stem cell factor, FIk2/FIt3 ligand, thrombopoietin, IL-6, and soluble IL-6 receptor. J. Clin. Invest. 105, 1013-1021 (2000).
213. Audet, J., Miller, C. L., Rose-John, S., Piret, J. M. \& Eaves, C. J. Distinct role of gp130 activation in promoting self-renewal divisions by mitogenically stimulated murine hematopoietic stem cells. Proc. Natl. Acad. Sci. U. S. A. 98, 1757-62 (2001).
214. J, A., CL, M., CJ, E. \& JM, P. Common and distinct features of cytokine effects on hematopoietic stem and progenitor cells revealed by dose-response surface analysis. Biotechnol. Bioeng. 80, 393-404 (2002).
215. Sitnicka, E. et al. The Effect of Thrombopoietin on the Proliferation and Differentiation of Murine Hematopoietic Stem Cells. Blood 87, 4998-5005 (1996).
216. V, R. et al. Thrombopoietin, but not erythropoietin, directly stimulates multilineage growth of primitive murine bone marrow progenitor cells in synergy with early acting cytokines: distinct interactions with the ligands for ckit and FLT3. Blood 88, 4481-4492 (1996).
217. H, K., Y, Y., K, K. \& M, O. Thrombopoietin, the ligand for the Mpl receptor, synergizes with steel factor and other early acting cytokines in supporting proliferation of primitive hematopoietic progenitors of mice. Blood 87, 45444551 (1996).
218. N, B.-V. et al. FLT3 receptor and ligand are dispensable for maintenance and posttransplantation expansion of mouse hematopoietic stem cells. Blood 113, 3453-3460 (2009).
219. Sheng Ding, $\dagger$, Nathanael S. Gray, ${ }^{*}, \ddagger, X u W u, \ddagger$, Qiang Ding, $\ddagger$ and \& Peter G. Schultz*, $\dagger, \ddagger$. A Combinatorial Scaffold Approach toward Kinase-Directed Heterocycle Libraries. J. Am. Chem. Soc. 124, 1594-1596 (2002).
220. Boitano, A. E. et al. Aryl hydrocarbon receptor antagonists promote the expansion of human hematopoietic stem cells. Science (80-. ). 329, 13451348 (2010).
221. Murray, L. J. et al. Thrombopoietin, flt3, and kit ligands together suppress apoptosis of human mobilized CD34+ cells and recruit primitive CD34+Thy-1+ cells into rapid division. Exp. Hematol. 27, 1019-1028 (1999).
222. I, F. et al. Cord blood expansion. Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal. Science 345, 1509-1512 (2014).
223. J, C. et al. UM171 induces a homeostatic inflammatory-detoxification response supporting human HSC self-renewal. PLoS One 14, (2019).
224. Cohen, S. et al. Hematopoietic stem cell transplantation using single UM171expanded cord blood: a single-arm, phase 1-2 safety and feasibility study. Lancet Haematol. 7, e134-e145 (2020).
225. Wilkinson, A. C. et al. Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. Nat. 20195717763 571, 117-121 (2019).
226. Wilkinson, A. C., Ishida, R., Nakauchi, H. \& Yamazaki, S. Long-term ex vivo expansion of mouse hematopoietic stem cells. Nat. Protoc. 2020 152 15, 628648 (2020).
227. JS, C. \& BA, H. Marrow-inspired matrix cues rapidly affect early fate decisions of hematopoietic stem and progenitor cells. Sci. Adv. 3, (2017).
228. SK, N. et al. Immunofluorescence characterization of key extracellular matrix proteins in murine bone marrow in situ. J. Histochem. Cytochem. 46, 371-377 (1998).
229. MR, C., HY, C. \& R, M. Preleukemic Hematopoietic Stem Cells in Human Acute Myeloid Leukemia. Front. Oncol. 7, (2017).
230. Laverdière, I. et al. Leukemic stem cell signatures identify novel therapeutics targeting acute myeloid leukemia. Blood Cancer J. 2018 86 8, 1-16 (2018).
231. Schieber, M., Crispino, J. D. \& Stein, B. Myelofibrosis in 2019: moving beyond JAK2 inhibition. Blood Cancer J. 2019 99 9, 1-11 (2019).
232. S, F., C, M., A, F., T, D. G. \& N, S. Myeloproliferative neoplasms (MPNs) - Part 1: An overview of the diagnosis and treatment of the 'classical' MPNs. Can. Oncol. Nurs. J. 28, 262-268 (2018).
233. J, N. et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. N. Engl. J. Med. 369, 2391-2405 (2013).
234. R, Z. et al. Identification of an acquired JAK2 mutation in polycythemia vera. J. Biol. Chem. 280, 22788-22792 (2005).
235. James, C. et al. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. Nat. 20054347037 434, 1144-1148 (2005).
236. J, S. et al. An amphipathic motif at the transmembrane-cytoplasmic junction prevents autonomous activation of the thrombopoietin receptor. Blood 107, 1864-1871 (2006).
237. MA, J. \& CW, C. Recent insights regarding the molecular basis of myeloproliferative neoplasms. Korean J. Intern. Med. 35, 1-11 (2020).
238. Belcic Mikic, T., Pajic, T. \& Sever, M. CALR mutations in a cohort of JAK2 V617F negative patients with suspected myeloproliferative neoplasms. Sci. Reports 2019 919, 1-9 (2019).
239. How, J., Hobbs, G. S. \& Mullally, A. Mutant calreticulin in myeloproliferative neoplasms. Blood 134, 2242-2248 (2019).
240. Kameda, T. et al. Loss of TET2 has dual roles in murine myeloproliferative neoplasms: disease sustainer and disease accelerator. Blood 125, 304-315 (2015).
241. Ortmann, C. A. et al. Effect of Mutation Order on Myeloproliferative Neoplasms. N. Engl. J. Med. 372, 601-612 (2015).
242. Tefferi, A. Novel mutations and their functional and clinical relevance in myeloproliferative neoplasms: JAK2, MPL, TET2, ASXL1, CBL, IDH and IKZF1. Leuk. 2010246 24, 1128-1138 (2010).
243. X, X. et al. JAK2(V617F): Prevalence in a large Chinese hospital population. Blood 109, 339-342 (2007).
244. Lundberg, P. et al. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood 123, 2220-2228 (2014).
245. Ko, M. et al. Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc. Natl. Acad. Sci. 108, 14566-14571 (2011).
246. Li, Z. et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood 118, 45094518 (2011).
247. Quivoron, C. et al. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. Cancer Cell 20, 25-38 (2011).
248. K, M.-C. et al. Tet2 loss leads to increased hematopoietic stem cell selfrenewal and myeloid transformation. Cancer Cell 20, 11-24 (2011).
249. CA, O. et al. Effect of mutation order on myeloproliferative neoplasms. N. Engl. J. Med. 372, 601-612 (2015).
250. Globisch, D. et al. Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. PLoS One 5, e15367 (2010).
251. Li, W. et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. Cell Res. 20172710 27, 1243-1257 (2017).
252. MR, B., G, F. \& W, R. Uncovering the role of 5 -hydroxymethylcytosine in the epigenome. Nat. Rev. Genet. 13, 7-13 (2011).
253. E, S., OA, B., A, T., F, F. \& W, V. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. Leukemia 28, 485-496 (2014).
254. Tefferi, A. et al. TET2 mutations and their clinical correlates in polycythemia vera, essential thrombocythemia and myelofibrosis. Leuk. 2009235 23, 905911 (2009).
255. A, T. et al. Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFRA correlates. Leukemia 23, 900-904 (2009).
256. JS, H., DS, J., JR, K., NH, R. \& JS, S. Analysis of the Ten-Eleven Translocation 2 (TET2) gene mutation in myeloproliferative neoplasms. Ann. Clin. Lab. Sci. 44, 173-179 (2014).
257. Ko, M. et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nat. 20104687325 468, 839-843 (2010).
258. Tefferi, A. et al. Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. Leuk. 2009 237 23, 1343-1345 (2009).
259. T, K. et al. Loss of TET2 has dual roles in murine myeloproliferative neoplasms: disease sustainer and disease accelerator. Blood 125, 304-315 (2015).
260. E, C. et al. Distinct effects of concomitant Jak2V617F expression and Tet2 loss in mice promote disease progression in myeloproliferative neoplasms. Blood 125, 327-335 (2015).
261. A, T. \& A, P. Myeloproliferative Neoplasms: A Contemporary Review. JAMA Oncol. 1, 97-105 (2015).
262. Moulard, O. et al. Epidemiology of myelofibrosis, essential thrombocythemia, and polycythemia vera in the European Union. Eur. J. Haematol. 92, 289-297 (2014).
263. Tefferi, A. Polycythemia Vera: A Comprehensive Review and Clinical Recommendations. Mayo Clin. Proc. 78, 174-194 (2003).
264. E, C. \& T, T. Complications and causes of death in polycythaemia vera. Acta Med. Scand. 172, 513-523 (1962).
265. N, G. et al. Leucocytosis in polycythaemia vera predicts both inferior survival and leukaemic transformation. Br. J. Haematol. 138, 354-358 (2007).
266. F, P. et al. Prognostic factors for thrombosis, myelofibrosis, and leukemia in essential thrombocythemia: a study of 605 patients. Haematologica 93, 16451651 (2008).
267. F, P. et al. A prospective study of 338 patients with polycythemia vera: the impact of JAK2 (V617F) allele burden and leukocytosis on fibrotic or leukemic disease transformation and vascular complications. Leukemia 24, 1574-1579 (2010).
268. N, G. et al. Risk stratification for survival and leukemic transformation in essential thrombocythemia: a single institutional study of 605 patients. Leukemia 21, 270-276 (2007).
269. HL, G. \& RA, M. Therapy for myeloproliferative neoplasms: when, which agent, and how? Blood 124, 3529-3537 (2014).
270. Patnaik, M. M. et al. Prognostic interaction between ASXL1 and TET2 mutations in chronic myelomonocytic leukemia. Blood Cancer J. 2016 61 6, e385-e385 (2016).
271. Chou, W.-C. et al. TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. Blood 118, 3803-3810 (2011).
272. Patterson, S. D. \& Aebersold, R. H. Proteomics: the first decade and beyond. Nat. Genet. 33, 311-323 (2003).
273. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E. \& Kent, S. B. Internal amino acid sequence analysis of proteins separated by one- or twodimensional gel electrophoresis after in situ protease digestion on nitrocellulose. Proc. Natl. Acad. Sci. U. S. A. 84, 6970-6974 (1987).
274. Edman, P., Högfeldt, E., Sillén, L. G. \& Kinell, P.-O. Method for Determination of the Amino Acid Sequence in Peptides. Acta Chem. Scand. 4, 283-293 (1950).
275. Laursen, R. A. Solid-Phase Edman Degradation. An Automatic Peptide Sequencer. Eur. J. Biochem. 20, 89-102 (1971).
276. Tempst, P., Link, A. J., Riviere, L. R., Fleming, M. \& Elicone, C. Internal sequence analysis of proteins separated on polyacrylamide gels at the submicrogram level: Improved methods, applications and gene cloning strategies. Electrophoresis 11, 537-553 (1990).
277. Anderson, N. L., Hofmann, J. P., Gemmell, A. \& Taylor, J. Global approaches to quantitative analysis of gene-expression patterns observed by use of twodimensional gel electrophoresis. Clin. Chem. 30, 2031-2036 (1984).
278. Kind, T. \& Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. Bioanalytical Reviews vol. 2 23-60 (2010).
279. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. \& Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. Science vol. 246 64-71 (1989).
280. Patterson, S. D. \& Aebersold, R. Mass spectrometric approaches for the identification of gel-separated proteins. ELECTROPHORESIS vol. 16 17911814 (1995).
281. Mann, M., Kulak, N. A., Nagaraj, N. \& Cox, J. The Coming Age of Complete, Accurate, and Ubiquitous Proteomes. Molecular Cell vol. 49 583-590 (2013).
282. Wilm, M. S. \& Mann, M. Electrospray and Taylor-Cone theory, Dole's beam of macromolecules at last? Int. J. Mass Spectrom. Ion Process. 136, 167-180 (1994).
283. Wilm, M. \& Mann, M. Analytical properties of the nanoelectrospray ion source. Anal. Chem. 68, 1-8 (1996).
284. Mann, M. \& Wilm, M. Error-Tolerant Identification of Peptides in Sequence Databases by Peptide Sequence Tags. Anal. Chem. 66, 4390-4399 (1994).
285. Shevchenko, A., Tomas, H., Havliš, J., Olsen, J. V. \& Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat. Protoc. 1, 2856-2860 (2007).
286. Wilm, M. et al. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature 379, 466-469 (1996).
287. Link, A. J. et al. Direct analysis of protein complexes using mass spectrometry. Nat. Biotechnol. 17, 676-682 (1999).
288. Wolters, D. A., Washburn, M. P. \& Yates, J. R. An automated multidimensional protein identification technology for shotgun proteomics. Anal. Chem. 73, 5683-5690 (2001).
289. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M. C. \& Yates, J. R. Protein analysis by shotgun/bottom-up proteomics. Chemical Reviews vol. 113 2343-

2394 (2013).
290. Harper, J. W. \& Bennett, E. J. Proteome complexity and the forces that drive proteome imbalance. Nature 537, 328-38 (2016).
291. Reddy, P. J., Ray, S. \& Srivastava, S. The quest of the human proteome and the missing proteins: Digging deeper. Omi. A J. Integr. Biol. 19, 276-282 (2015).
292. Kim, M. S. et al. A draft map of the human proteome. Nature 509, 575-581 (2014).
293. Farrah, T. et al. State of the human proteome in 2013 as viewed through peptideatlas: Comparing the kidney, urine, and plasma proteomes for the biology- and disease-driven human proteome project. J. Proteome Res. 13, 60-75 (2014).
294. Farrah, T. et al. The state of the human proteome in 2012 as viewed through peptideatlas. J. Proteome Res. 12, 162-171 (2013).
295. Uhlén, M. et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. Mol. Cell. Proteomics 4, 1920-1932 (2005).
296. Wang, D. et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. Mol. Syst. Biol. 15, (2019).
297. Beck, M. et al. The quantitative proteome of a human cell line. Mol. Syst. Biol. 7, (2011).
298. Nagaraj, N. et al. Deep proteome and transcriptome mapping of a human cancer cell line. Mol. Syst. Biol. 7, (2011).
299. Palii, C. G., Cheng, Q., Gillespie, M. A., Morrissey, E. \& Higgs, D. R. SingleCell Proteomics Reveal that Quantitative Changes in Co-expressed LineageSpecific Transcription Factors Determine Cell Fate. Cell Stem Cell 24, 812820 (2019).
300. S, P. et al. Lineage-Biased Hematopoietic Stem Cells Are Regulated by Distinct Niches. Dev. Cell 44, 634-641.e4 (2018).
301. Doerr, A. Single-cell proteomics. Nature Methods vol. 1620 (2019).
302. Gutstein, H. B., Morris, J. S., Annangudi, S. P. \& Sweedler, J. V. Microproteomics: Analysis of protein diversity in small samples. Mass Spectrom. Rev. 27, 316-330 (2008).
303. Stoeckli, M., Chaurand, P., Hallahan, D. E. \& Caprioli, R. M. Imaging mass spectrometry: A new technology for the analysis of protein expression in mammalian tissues. Nat. Med. 2001 74 7, 493-496 (2001).
304. M, S., TB, F. \& RM, C. Automated mass spectrometry imaging with a matrixassisted laser desorption ionization time-of-flight instrument. J. Am. Soc. Mass Spectrom. 10, 67-71 (1999).
305. EB, M. et al. Massively parallel sample preparation for the MALDI MS analyses of tissues. Anal. Chem. 78, 6826-6832 (2006).
306. Taylor, M. J., Lukowski, J. K. \& Anderton, C. R. Spatially Resolved Mass Spectrometry at the Single Cell: Recent Innovations in Proteomics and Metabolomics. J. Am. Soc. Mass Spectrom. 32, 872-894 (2021).
307. Russell, C. L. et al. Combined tissue and fluid proteomics with Tandem Mass Tags to identify low-abundance protein biomarkers of disease in peripheral body fluid: An Alzheimer's Disease case study. Rapid Commun. Mass Spectrom. 31, 153-159 (2017).
308. Zhu, Y. et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. Nat. Commun. 201891 9, 110 (2018).
309. Budnik, B., Levy, E., Harmange, G. \& Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. 2018191 19, 1-12 (2018).
310. Specht, H. et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. Genome Biol. 2021221 22, 1-27 (2021).
311. Budnik, B., Levy, E., Harmange, G. \& Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. 19, 161 (2018).
312. Z, W. et al. 27-Plex Tandem Mass Tag Mass Spectrometry for Profiling Brain Proteome in Alzheimer's Disease. Anal. Chem. 92, 7162-7170 (2020).
313. A, T. et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895-1904 (2003).
314. Cheung, T. K. et al. Defining the carrier proteome limit for single-cell proteomics. Nat. Methods 2020181 18, 76-83 (2020).
315. Chen, A. T., Franks, A. \& Slavov, N. DART-ID increases single-cell proteome coverage. PLOS Comput. Biol. 15, e1007082 (2019).
316. H, S. \& N, S. Optimizing Accuracy and Depth of Protein Quantification in Experiments Using Isobaric Carriers. J. Proteome Res. 20, 880-887 (2021).
317. Schoof, E. M. et al. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. Nat. Commun. 2021121 12, 1-15 (2021).
318. Brunner, A.-D. et al. Ultra-high sensitivity mass spectrometry quantifies singlecell proteome changes upon perturbation. bioRxiv 2020.12.22.423933 (2021) doi:10.1101/2020.12.22.423933.
319. Guevremont, R. High-field asymmetric waveform ion mobility spectrometry: A new tool for mass spectrometry. J. Chromatogr. A 1058, 3-19 (2004).
320. Cong, Y. et al. Improved Single-Cell Proteome Coverage Using Narrow-Bore Packed NanoLC Columns and Ultrasensitive Mass Spectrometry. Anal. Chem. 92, 2665-2671 (2020).
321. Cong, Y. et al. Ultrasensitive single-cell proteomics workflow identifies $>1000$ protein groups per mammalian cell. Chem. Sci. 12, 1001-1006 (2021).
322. Tsai, C.-F. et al. An Improved Boosting to Amplify Signal with Isobaric Labeling (iBASIL) Strategy for Precise Quantitative Single-cell Proteomics *. Mol. Cell. Proteomics 19, 828-838 (2020).
323. M, D. et al. High-Throughput Single Cell Proteomics Enabled by Multiplex Isobaric Labeling in a Nanodroplet Sample Preparation Platform. Anal. Chem. 91, 13119-13127 (2019).
324. Harper, J. W. \& Bennett, E. J. Proteome complexity and the forces that drive proteome imbalance. Nat. 20165377620 537, 328-338 (2016).
325. Balzeau, J., Menezes, M. R., Cao, S. \& Hagan, J. P. The LIN28/let-7 Pathway in Cancer. Front. Genet. 8, 31 (2017).
326. Jassinskaja, M. et al. Comprehensive Proteomic Characterization of Ontogenic Changes in Hematopoietic Stem and Progenitor Cells. Cell Rep. 21, 32853297 (2017).
327. S, A. et al. Sensitive Quantitative Proteomics of Human Hematopoietic Stem and Progenitor Cells by Data-independent Acquisition Mass Spectrometry. Mol. Cell. Proteomics 18, 1454-1467 (2019).
328. BW, Z. et al. Proteomic analysis of young and old mouse hematopoietic stem cells and their progenitors reveals post-transcriptional regulation in stem cells.

Elife 9, 1-29 (2020).
329. Mansell, E. et al. Mitochondrial Potentiation Ameliorates Age-Related Heterogeneity in Hematopoietic Stem Cell Function. Cell Stem Cell 28, 241$256 . e 6$ (2021).
330. V, S. et al. Bile Acids Protect Expanding Hematopoietic Stem Cells from Unfolded Protein Stress in Fetal Liver. Cell Stem Cell 18, 522-532 (2016).
331. RA, S., JA, M., A, S. \& SJ, M. Haematopoietic stem cells require a highly regulated protein synthesis rate. Nature 509, 49-54 (2014).
332. Wilson, N. K. et al. Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. Cell Stem Cell 16, (2015).
333. Che, J. L. C. et al. A highly efficient reporter system for identifying and characterizing in vitro expanded hematopoietic stem cells. bioRxiv 2021.06.18.448972 (2021) doi:10.1101/2021.06.18.448972.
334. Oedekoven, C. A. et al. Hematopoietic stem cells retain functional potential and molecular identity in hibernation cultures. Stem Cell Reports (2021) doi:https://doi.org/10.1016/j.stemcr.2021.04.002.
335. Segel, M. et al. Niche stiffness underlies the ageing of central nervous system progenitor cells. Nat. 20195737772 573, 130-134 (2019).
336. Labouesse, C. et al. StemBond hydrogels optimise the mechanical microenvironment for embryonic stem cells. bioRxiv 768762 (2019) doi:10.1101/768762.
337. Kent, D. G. et al. Self-Renewal of Single Mouse Hematopoietic Stem Cells Is Reduced by JAK2V617F Without Compromising Progenitor Cell Expansion. PLoS Biol. 11, e1001576 (2013).
338. Grynkiewicz, G., Poenie, M. \& Tsienb, R. Y. A New Generation of Ca2+ Indicators with Greatly Improved Fluorescence Properties. J. Biol. Chem. 260, 1385 (1985).
339. Shepherd, M. S. et al. Single-cell approaches identify the molecular network driving malignant hematopoietic stem cell self-renewal. Blood 132, (2018).
340. Kent, D. G., Dykstra, B. J. \& Eaves, C. J. Isolation and Assessment of Single Long-Term Reconstituting Hematopoietic Stem Cells from Adult Mouse Bone Marrow. in Current Protocols in Stem Cell Biology 2A.4.1-2A.4.24 (John Wiley \& Sons, Inc., 2016). doi:10.1002/cpsc. 10.
341. Che, L. C. J. MOLECULAR CHARACTERISATION OF EXPANDED MOUSE HAEMATOPOIETIC STEM CELLS USING A NOVEL IN VITRO REPORTER STRATEGY. (University of Cambridge, 2020).
342. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods 10, 1096-1100 (2013).
343. JS, D. et al. A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice. Blood 131, e1-e11 (2018).
344. Robinson, M. D., McCarthy, D. J. \& Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-40 (2010).
345. McCarthy, D. J., Chen, Y. \& Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40, 4288-97 (2012).
346. Robinson, M. D. \& Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25 (2010).
347. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-
sequencing and microarray studies. Nucleic Acids Res. 43, e47-e47 (2015).
348. I, T. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352, 189-196 (2016).
349. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. 20, 163-172 (2019).
350. Jojic, V. et al. Identification of transcriptional regulators in the mouse immune system. Nat. Immunol. 14, 633-43 (2013).
351. Shay, T. \& Kang, J. Immunological Genome Project and systems immunology. Trends Immunol. 34, 602-9 (2013).
352. S, C.-A. et al. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res. 48, (2020).
353. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. 102, 15545 LP - 15550 (2005).
354. Mootha, V. K. et al. PGC-1a-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. 34, 267-273 (2003).
355. Chambers, S. M. et al. Hematopoietic Fingerprints: An Expression Database of Stem Cells and Their Progeny. Cell Stem Cell 1, 578-591 (2007).
356. TA, V. et al. Molecular signatures of proliferation and quiescence in hematopoietic stem cells. PLoS Biol. 2, (2004).
357. Kucinski, I. et al. Interactions between lineage-associated transcription factors govern haematopoietic progenitor states. EMBO J. 39, e104983 (2020).
358. Szklarczyk, D. et al. The STRING database in 2017: quality-controlled proteinprotein association networks, made broadly accessible. Nucleic Acids Res. 45, D362-D368 (2017).
359. Cabezas-Wallscheid, N. et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. Cell Stem Cell 15, 507-522 (2014).
360. Hennrich, M. L. et al. Cell-specific proteome analyses of human bone marrow reveal molecular features of age-dependent functional decline. Nat. Commun. 201891 9, 1-18 (2018).
361. Specht, H. et al. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity using SCoPE2. Genome Biol. 22, 50 (2021).
362. Redecke, V. et al. Hematopoietic progenitor cell lines with myeloid and lymphoid potential. Nat. Methods 10, 795-803 (2013).
363. McMillan, A., Renaud, J. B., Gloor, G. B., Reid, G. \& Sumarah, M. W. Postacquisition filtering of salt cluster artefacts for LC-MS based human metabolomic studies. J. Cheminformatics 2016 81 8, 1-5 (2016).
364. E, F., S, K.-K., U, J., A, P. \& M, D.-W. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. Electrophoresis 31, 3573-3579 (2010).
365. SY, O. et al. iTRAQ underestimation in simple and complex mixtures: 'the good, the bad and the ugly'. J. Proteome Res. 8, 5347-5355 (2009).
366. TS, B. \& JV, O. Offline High pH Reversed-Phase Peptide Fractionation for Deep Phosphoproteome Coverage. Methods Mol. Biol. 1355, 179-192 (2016).
367. TS, B., C, F. \& JV, O. Off-line high-pH reversed-phase fractionation for indepth phosphoproteomics. J. Proteome Res. 13, 6176-6186 (2014).
368. N, Z., R, G., A, D. P., S, B. \& M, C. Human fibroblasts with mutations in COL5A1 and COL3A1 genes do not organize collagens and fibronectin in the
extracellular matrix, down-regulate alpha2beta1 integrin, and recruit alphavbeta3 Instead of alpha5beta1 integrin. J. Biol. Chem. 279, 18157-18168 (2004).
369. SP, G. et al. Influence of cartilage extracellular matrix molecules on cell phenotype and neocartilage formation. Tissue Eng. Part A 20, 264-274 (2014).
370. Ny, L. et al. The plasminogen receptor, Plg-RKT, plays a role in inflammation and fibrinolysis during cutaneous wound healing in mice. Cell Death Dis. 2020 1112 11, 1-15 (2020).
371. Mi, H., Muruganujan, A., Ebert, D., Huang, X. \& Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47, D419-D426 (2019).
372. M, K. \& S, G. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27-30 (2000).
373. Chen, G., Ning, B. \& Shi, T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front. Genet. 0, 317 (2019).
374. Vit, O. \& Petrak, J. Integral membrane proteins in proteomics. How to break open the black box? J. Proteomics 153, 8-20 (2017).
375. L, X. \& A, G. Coordination of Ribosomal Protein and Ribosomal RNA Gene Expression in Response to TOR Signaling. Curr. Genomics 10, 198-205 (2009).
376. Lima, S. A. et al. Short Poly(A) Tails are a Conserved Feature of Highly Expressed Genes. Nat. Struct. Mol. Biol. 24, 1057 (2017).
377. Marzluff, W. F., Wagner, E. J. \& Duronio, R. J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat. Rev. Genet. 2008 911 9, 843-854 (2008).
378. Zhao, S., Zhang, Y., Gamini, R., Zhang, B. \& Von Schack, D. Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. Sci. Reports 2018 81 8, 1-12 (2018).
379. Kim, I. V. et al. Efficient depletion of ribosomal RNA for RNA sequencing in planarians. BMC Genomics 20, 1-12 (2019).
380. Roumeliotis, T. I. et al. Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. Cell Rep. 20, 2201-2214 (2017).
381. Athanasiadis, E. I. et al. Single-cell RNA-sequencing uncovers transcriptional states and fate decisions in haematopoiesis. Nat. Commun. 201781 8, 1-11 (2017).
382. RK, K. et al. Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. Cell 173, 90-103.e19 (2018).
383. S, S. et al. Electronic volume of CD34 positive cells from peripheral blood apheresis samples. Cytometry B. Clin. Cytom. 74, 182-188 (2008).
384. Wadkin, L. E. et al. Dynamics of single human embryonic stem cells and their pairs: a quantitative analysis. Sci. Reports 2017 71 7, 1-12 (2017).
385. SH, C., SJ, M., X, L. \& IL, W. In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. Proc. Natl. Acad. Sci. U. S. A. 96, 3120-3125 (1999).
386. GB, B., B, W., R, R. \& I, B. Quiescence, cycling, and turnover in the primitive hematopoietic stem cell compartment. Exp. Hematol. 25, 445-453 (1997).
387. R, Y. et al. Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment. Cell Stem Cell 22, 600-607.e4 (2018).
388. A, W. et al. Hematopoietic stem cells reversibly switch from dormancy to self-
renewal during homeostasis and repair. Cell 135, 1118-1129 (2008).
389. AA, W. et al. Phf6 Loss Enhances HSC Self-Renewal Driving Tumor Initiation and Leukemia Stem Cell Activity in T-ALL. Cancer Discov. 9, 436-451 (2019).
390. Sommarin, M. N. E. et al. Single-Cell Multiomics Reveals Distinct Cell States at the Top of the Human Hematopoietic Hierarchy. bioRxiv 2021.04.01.437998 (2021) doi:10.1101/2021.04.01.437998.
391. JD, B. et al. Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape of Human Hematopoietic Differentiation. Cell 173, 1535-1548.e16 (2018).
392. Cheng, H., Zheng, Z. \& Cheng, T. New paradigms on hematopoietic stem cell differentiation. Protein Cell 2019111 11, 34-44 (2019).
393. Kunimoto, H. et al. Tet2 disruption leads to enhanced self-renewal and altered differentiation of fetal liver hematopoietic stem cells. Sci. Reports 201221 2, 1-10 (2012).
394. Franks, A., Airoldi, E. \& Slavov, N. Post-transcriptional regulation across human tissues. PLOS Comput. Biol. 13, e1005535 (2017).
395. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. Nat. 20114737347 473, 337-342 (2011).
396. Liu, Y., Beyer, A. \& Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell 165, 535-550 (2016).
397. Hastings, J. F., Skhinas, J. N., Fey, D., Croucher, D. R. \& Cox, T. R. The extracellular matrix as a key regulator of intracellular signalling networks. Br. J. Pharmacol. 176, 82-92 (2019).
398. J, S. \& SN, C. The JAK-STAT pathway and hematopoietic stem cells from the JAK2 V617F perspective. JAK-STAT 1, 184-190 (2012).
399. S, H. et al. PI3 kinase alpha and delta promote hematopoietic stem cell activation. JCI insight 5, (2019).
400. K, M. et al. Foxo3a is essential for maintenance of the hematopoietic stem cell pool. Cell Stem Cell 1, 101-112 (2007).
401. CG, L. \& MA, K. Cell responses to FGFR3 signalling: growth, differentiation and apoptosis. Exp. Cell Res. 304, 417-431 (2005).
402. T, U., M, H., T, M., A, N.-I. \& T, S. Ca 2+-mitochondria axis drives cell division in hematopoietic stem cells. J. Exp. Med. 215, 2097-2113 (2018).
403. Luchsinger, L. L. et al. Harnessing Hematopoietic Stem Cell Low Intracellular Calcium Improves Their Maintenance In Vitro. Cell Stem Cell 25, 225-240.e7 (2019).
404. Zhang, P. et al. The physical microenvironment of hematopoietic stem cells and its emerging roles in engineering applications. Stem Cell Res. Ther. 2019 101 10, 1-13 (2019).
405. C, Z. et al. Ageing characteristics of bone indicated by transcriptomic and exosomal proteomic analysis of cortical bone cells. J. Orthop. Surg. Res. 14, (2019).
406. SA, M. et al. Modulation of angiotensin II-mediated cardiac remodeling by the MEF2A target gene Xirp2. Circ. Res. 106, 952-960 (2010).
407. WS, M., E, S., O, W., D, M. \& L, T. TAF10 (TAF(II)30) is necessary for TFIID stability and early embryogenesis in mice. Mol. Cell. Biol. 23, 4307-4318 (2003).
408. E, L. et al. CDK6 levels regulate quiescence exit in human hematopoietic stem cells. Cell Stem Cell 16, 302-313 (2015).
409. Q, L., J, L. \& B, C. Identification of CDCA8, DSN1 and BIRC5 in Regulating

Cell Cycle and Apoptosis in Osteosarcoma Using Bioinformatics and Cell Biology. Technol. Cancer Res. Treat. 19, (2020).
410. H, T. et al. Pathogen-Induced TLR4-TRIF Innate Immune Signaling in Hematopoietic Stem Cells Promotes Proliferation but Reduces Competitive Fitness. Cell Stem Cell 21, 225-240.e5 (2017).
411. Challen, G. A., Boles, N. C., Chambers, S. M. \& Goodell, M. A. Distinct Hematopoietic Stem Cell Subtypes Are Differentially Regulated by TGF- $\beta 1$. Cell Stem Cell 6, 265-278 (2010).
412. Brenner, A. K. \& Bruserud, Ø. S100 Proteins in Acute Myeloid Leukemia. Neoplasia 20, 1175-1186 (2018).
413. Wang, S. et al. S100A8/A9 in Inflammation. Front. Immunol. 9, 1298 (2018).
414. Magee, J. A. \& Signer, R. A. J. Developmental Stage-Specific Changes in Protein Synthesis Differentially Sensitize Hematopoietic Stem Cells and Erythroid Progenitors to Impaired Ribosome Biogenesis. Stem Cell Reports 16, 20-28 (2021).
415. $M, M . \& D, C$. The mitochondrial metabolic checkpoint and aging of hematopoietic stem cells. Curr. Opin. Hematol. 23, 318-324 (2016).
416. Gazon, H., Barbeau, B., Mesnard, J.-M. \& Peloponese, J.-M. J. Hijacking of the AP-1 Signaling Pathway during Development of ATL. Front. Microbiol. 0, 2686 (2018).
417. Santaguida, M. et al. JunB Protects against Myeloid Malignancies by Limiting Hematopoietic Stem Cell Proliferation and Differentiation without Affecting Self-Renewal. Cancer Cell 15, 341-352 (2009).
418. Liu, B. et al. PIAS1 SUMO ligase regulates the self-renewal and differentiation of hematopoietic stem cells. EMBO J. 33, 101-113 (2014).
419. ML, C. Toll-like receptor signaling in hematopoietic stem and progenitor cells. Curr. Opin. Hematol. 26, 207-213 (2019).
420. Höpner, S. S. et al. LIGHT/LT $\beta$ R signaling regulates self-renewal and differentiation of hematopoietic and leukemia stem cells. Nat. Commun. 2021 121 12, 1-16 (2021).
421. EM, P., MR, W. \& E, P. Cell cycle regulation in hematopoietic stem cells. J. Cell Biol. 195, 709-720 (2011).
422. Oliverio, S. et al. Transglutaminase Type 2 is Involved in the Hematopoietic Stem Cells Homeostasis. Biochem. 20208510 85, 1159-1168 (2020).
423. J, X., Z, W., X, X., G, L. \& J, X. Screening and identification of key genes and pathways in metastatic uveal melanoma based on gene expression using bioinformatic analysis. Medicine (Baltimore). 99, e22974 (2020).
424. MA, S. Integrins and extracellular matrix in mechanotransduction. Cold Spring Harb. Perspect. Biol. 2, (2010).
425. Walzog, B. et al. The Leukocyte Integrin Mac-1 (CD11b/CD18) Contributes to Binding of Human Granulocytes to Collagen. Exp. Cell Res. 218, 28-38 (1995).
426. Garnotel, R. et al. Human Blood Monocytes Interact with Type I Collagen Through $\alpha x \beta 2$ Integrin (CD11c-CD18, gp150-95). J. Immunol. 164, 5928-5934 (2000).
427. Ridger, V. C., Wagner, B. E., Wallace, W. A. H. \& Hellewell, P. G. Differential Effects of CD18, CD29, and CD49 Integrin Subunit Inhibition on Neutrophil Migration in Pulmonary Inflammation. J. Immunol. 166, 3484-3490 (2001).
428. Aldridge, V. et al. Human mesenchymal stem cells are recruited to injured liver in a $\beta 1$-integrin and CD44 dependent manner. Hepatology 56, 1063-1073
(2012).
429. PH, K. \& LG, V.-D. The Role of Integrin $\alpha 6$ (CD49f) in Stem Cells: More than a Conserved Biomarker. Stem Cells Dev. 26, 1090-1099 (2017).
430. V, C., M, P., KJ, H. \& NA, C. An Evaluation of CD61 Immunohistochemistry in Identification of Vascular Invasion in Follicular Thyroid Neoplasms. Head Neck Pathol. 14, 399-405 (2020).
431. Patel, Y. et al. Functional characterization of PM6/13, a $\beta 3$-specific (GPIIIa/CD61) monoclonal antibody that shows preferential inhibition of fibrinogen binding over fibronectin binding to activated human platelets. Thromb. Haemost. 79, 177-185 (1998).
432. MJ, F. et al. CD41 expression defines the onset of primitive and definitive hematopoiesis in the murine embryo. Development 130, 4393-4403 (2003).
433. Keinan, N., Pahima, H., Ben-Hail, D. \& Shoshan-Barmatz, V. The role of calcium in VDAC1 oligomerization and mitochondria-mediated apoptosis. Biochim. Biophys. Acta - Mol. Cell Res. 1833, 1745-1754 (2013).
434. De Stefani, D. et al. VDAC1 selectively transfers apoptotic Ca2+ signals to mitochondria. Cell Death Differ. 2012192 19, 267-273 (2011).
435. Stafford, N., Wilson, C., Oceandy, D., Neyses, L. \& Cartwright, E. J. The Plasma Membrane Calcium ATPases and Their Role as Major New Players in Human Disease. https://doi.org/10.1152/physrev.00028.2016 97, 1089-1125 (2017).
436. Periasamy, M. \& Kalyanasundaram, A. SERCA pump isoforms: Their role in calcium transport and disease. Muscle Nerve 35, 430-442 (2007).
437. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. Nat. Genet. 2000251 25, 25-29 (2000).
438. Wang, X. et al. Identification of a subset of immunosuppressive P2RX1negative neutrophils in pancreatic cancer liver metastasis. Nat. Commun. 2021 121 12, 1-17 (2021).
439. Wang, X. et al. Endotoxin-induced autocrine ATP signaling inhibits neutrophil chemotaxis through enhancing myosin light chain phosphorylation. Proc. Natl. Acad. Sci. 114, 4483-4488 (2017).
440. EW, W. et al. TRPC6 is the endothelial calcium channel that regulates leukocyte transendothelial migration during the inflammatory response. J. Exp. Med. 212, 1883-1899 (2015).
441. Z, U. et al. Meis1 preserves hematopoietic stem cells in mice by limiting oxidative stress. Blood 120, 4973-4981 (2012).
442. S, B. et al. Loss of Hmga1 gene function affects embryonic stem cell lymphohematopoietic differentiation. FASEB J. 17, 1496-1498 (2003).
443. MR, C. et al. The Lin28b-let-7-Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells. Nat. Cell Biol. 15, 916-925 (2013).
444. Badwe, C. R., Lis, R., Barcia Duran, J. G., Kunar, B. \& Rafii, S. Fli1 Is Essential for the Maintenance of Hematopoietic Stem Cell Homeostasis and Function. Blood 130, 3769-3769 (2017).
445. KJ, K. et al. ERG promotes the maintenance of hematopoietic stem cells by restricting their differentiation. Genes Dev. 29, 1915-1929 (2015).
446. P, X. et al. Identification of E74-like factor 1 (ELF1) as a transcriptional regulator of the Hox cofactor MEIS1. Exp. Hematol. 38, 798-808.e2 (2010).
447. Nottingham, W. T. et al. Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. Blood 110, 4188-4197 (2007).
448. Wilson, N. K., Calero-Nieto, F. J., Ferreira, R. \& Göttgens, B. Transcriptional regulation of haematopoietic transcription factors. Stem Cell Res. Ther. 2011 21 2, 1-8 (2011).
449. T, Y. et al. Hlf marks the developmental pathway for hematopoietic stem cells but not for erythro-myeloid progenitors. J. Exp. Med. 216, 1599-1614 (2019).
450. Zhang, H. et al. Sox4 Is a Key Oncogenic Target in C/EBPa Mutant Acute Myeloid Leukemia. Cancer Cell 24, 575-588 (2013).
451. Forsberg, E. C. et al. Differential Expression of Novel Potential Regulators in Hematopoietic Stem Cells. PLOS Genet. 1, e28 (2005).
452. Dorritie, K. A., Mccubrey, J. A. \& Johnson, D. E. STAT transcription factors in hematopoiesis and leukemogenesis: opportunities for therapeutic intervention. Leukemia 28, 248-257 (2014).
453. JK, M. et al. Distinct Molecular Signature of Murine Fetal Liver and Adult Hematopoietic Stem Cells Identify Novel Regulators of Hematopoietic Stem Cell Function. Stem Cells Dev. 26, 573-584 (2017).
454. Javed, A. et al. Groucho/TLE/R-esp proteins associate with the nuclear matrix and repress RUNX (CBF(alpha)/AML/PEBP2(alpha)) dependent activation of tissue-specific gene transcription. J. Cell Sci. 113, 2221-2231 (2000).
455. DA, W., Y, Z. \& JA, C. Rho GTPases and regulation of hematopoietic stem cell localization. Methods Enzymol. 439, 365-393 (2008).
456. Clapes, T., Lefkopoulos, S. \& Trompouki, E. Stress and Non-Stress Roles of Inflammatory Signals during HSC Emergence and Maintenance. Front. Immunol. 0, 487 (2016).
457. TE, N. et al. Hematopoietic stem cell development is dependent on blood flow. Cell 137, 736-748 (2009).
458. L, A. et al. Biomechanical forces promote embryonic haematopoiesis. Nature 459, 1131-1135 (2009).
459. N, L., MF, D. \& PL, W. Application of fluid mechanical force to embryonic sources of hemogenic endothelium and hematopoietic stem cells. Methods Mol. Biol. 1212, 183-193 (2015).
460. Cunningham, K. S. \& Gotlieb, A. I. The role of shear stress in the pathogenesis of atherosclerosis. Lab. Investig. 2005851 85, 9-23 (2004).
461. Fuster, J. J. et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. Science 355, 842 (2017).
462. D, T. Going with the flow: how shear stress signals the emergence of adult hematopoiesis. J. Exp. Med. 212, 600 (2015).
463. Mansell, E. et al. Mitochondrial Potentiation Ameliorates Age-Related Heterogeneity in Hematopoietic Stem Cell Function. Cell Stem Cell 28, 241256.e6 (2021).
464. Nakahara, F. et al. Engineering a haematopoietic stem cell niche by revitalizing mesenchymal stromal cells. Nat. Cell Biol. 2019215 21, 560-567 (2019).
465. S, K. \& H, G. HSC Niche Biology and HSC Expansion Ex Vivo. Trends Mol. Med. 23, 799-819 (2017).
466. Weissmann, S. et al. Landscape of TET2 mutations in acute myeloid leukemia. Leuk. 2012 265 26, 934-942 (2011).
467. Sanchez-Correa, B. et al. Cytokine profiles in acute myeloid leukemia patients at diagnosis: Survival is inversely correlated with IL-6 and directly correlated with IL-10 levels. Cytokine 61, 885-891 (2013).
468. Duan, C.-W. et al. Leukemia Propagating Cells Rebuild an Evolving Niche in

Response to Therapy. Cancer Cell 25, 778-793 (2014).
469. Forman, S. J. \& Rowe, J. M. The myth of the second remission of acute leukemia in the adult. Blood 121, 1077-1082 (2013).
470. Zhao, Z. et al. Tet2 Loss Dysregulates the Behavior of Mesenchymal Stem Cells and Increases Their Ability to Promote Tet2-/--Driven Myeloid Malignancy Progression. Blood 128, 1510-1510 (2016).
471. Ramdas, B. et al. Driver Mutations in Leukemia Promote Disease Pathogenesis through a Combination of Cell-Autonomous and Niche Modulation. Stem Cell Reports 15, 95-109 (2020).
472. Filipowicz, W., Bhattacharyya, S. N. \& Sonenberg, N. Mechanisms of posttranscriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. 200892 9, 102-114 (2008).
473. Keene, J. D. RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. 8, 533-543 (2007).
474. Zhao, B. S., Roundtree, I. A. \& He, C. Post-transcriptional gene regulation by mRNA modifications. Nat. Rev. Mol. Cell Biol. 2016181 18, 31-42 (2016).
475. Abdelmohsen, K. et al. Phosphorylation of HuR by Chk2 Regulates SIRT1 Expression. Mol. Cell 25, 543-557 (2007).
476. Garbarino-Pico, E. et al. Immediate early response of the circadian polyA ribonuclease nocturnin to two extracellular stimuli. RNA 13, 745-755 (2007).
477. Intine, R. V., Tenenbaum, S. A., Sakulich, A. L., Keene, J. D. \& Maraia, R. J. Differential phosphorylation and subcellular localization of La RNPs associated with precursor tRNAs and translation-related mRNAs. Mol. Cell 12, 1301-1307 (2003).
478. RP, W. \& T, A. Shear stress during early embryonic stem cell differentiation promotes hematopoietic and endothelial phenotypes. Biotechnol. Bioeng. 110, 1231-1242 (2013).
479. Osborn, E. A., Rabodzey, A., C. Forbes Dewey, J. \& Hartwig, J. H. Endothelial actin cytoskeleton remodeling during mechanostimulation with fluid shear stress. https://doi.org/10.1152/ajpcell.00218.2005 290, 444-452 (2006).
480. W, W. et al. Fluid shear stress stimulates phosphorylation-dependent nuclear export of HDAC5 and mediates expression of KLF2 and eNOS. Blood 115, 2971-2979 (2010).
481. JC, Y. et al. KIf2 is an essential factor that sustains ground state pluripotency. Cell Stem Cell 14, 864-872 (2014).
482. Park, C. S., Lewis, A., Chen, T. \& Lacorazza, D. Concise Review: Regulation of Self-Renewal in Normal and Malignant Hematopoietic Stem Cells by Krüppel-Like Factor 4. Stem Cells Transl. Med. 8, 568-574 (2019).
483. RA, M., D, G., A, L. \& DB, K. Hematopoietic Stem Cell Gene Therapy: Progress and Lessons Learned. Cell Stem Cell 21, 574-590 (2017).
484. Yamamoto, R. et al. Clonal Analysis Unveils Self-Renewing LineageRestricted Progenitors Generated Directly from Hematopoietic Stem Cells. Cell 154, 1112-1126 (2013).
485. E, T. et al. Integrin-a3 Is a Functional Marker of Ex Vivo Expanded Human Long-Term Hematopoietic Stem Cells. Cell Rep. 28, 1063-1073.e5 (2019).
486. Fares, I. et al. EPCR expression marks UM171-expanded CD34+ cord blood stem cells. Blood 129, 3344-3351 (2017).
487. M, A. et al. Deep imaging of bone marrow shows non-dividing stem cells are mainly perisinusoidal. Nature 526, 126-130 (2015).
488. R, G. et al. Fgd5 identifies hematopoietic stem cells in the murine bone
marrow. J. Exp. Med. 211, 1315-1331 (2014).
489. K, B. et al. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature 518, 542-546 (2015).
490. JY, C. et al. Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. Nature 530, 223-227 (2016).
491. Cabezas-Wallscheid, N. et al. Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. Cell 169, 807-823.e19 (2017).
492. Tajima, Y. et al. Continuous cell supply from Krt7-expressing hematopoietic stem cells during native hematopoiesis revealed by targeted in vivo gene transfer method. Sci. Reports 2017 71 7, 1-10 (2017).
493. JL, R. et al. CD34 and EPCR coordinately enrich functional murine hematopoietic stem cells under normal and inflammatory conditions. Exp. Hematol. 81, 1-15.e6 (2020).
494. Csaszar, E. et al. Rapid Expansion of Human Hematopoietic Stem Cells by Automated Control of Inhibitory Feedback Signaling. Cell Stem Cell 10, 218229 (2012).
495. A, I. et al. An All-Recombinant Protein-Based Culture System Specifically Identifies Hematopoietic Stem Cell Maintenance Factors. Stem cell reports 8, 500-508 (2017).
496. Dykstra, B. et al. Long-Term Propagation of Distinct Hematopoietic Differentiation Programs In Vivo. Cell Stem Cell 1, 218-229 (2007).
497. Heng, T. S. P. et al. The Immunological Genome Project: networks of gene expression in immune cells. Nat. Immunol. 2008910 9, 1091-1094 (2008).
498. T, S. et al. The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. Cell Stem Cell 7, 380-390 (2010).
499. K, T. et al. Regulation of glycolysis by Pdk functions as a metabolic checkpoint for cell cycle quiescence in hematopoietic stem cells. Cell Stem Cell 12, 49-61 (2013).
500. T, S., K, T. \& GL, S. Metabolic regulation of hematopoietic stem cells in the hypoxic niche. Cell Stem Cell 9, 298-310 (2011).
501. N, V. et al. Specification of haematopoietic stem cell fate via modulation of mitochondrial activity. Nat. Commun. 7, (2016).
502. Yu, W.-M. et al. Metabolic Regulation by the Mitochondrial Phosphatase PTPMT1 Is Required for Hematopoietic Stem Cell Differentiation. Cell Stem Cell 12, 62-74 (2013).
503. Tomczak, A. et al. Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. Sci. Reports 2018 818,110 (2018).
504. B, J. et al. The reactome pathway knowledgebase. Nucleic Acids Res. 48, D498-D503 (2020).
505. Mulloy, J. C. et al. Rho GTPases in hematopoiesis and hemopathies. Blood 115, 936-947 (2010).
506. Yang, L. et al. Rho GTPase Cdc42 coordinates hematopoietic stem cell quiescence and niche interaction in the bone marrow. Proc. Natl. Acad. Sci. 104, 5091-5096 (2007).
507. Gu, Y. et al. Hematopoietic Cell Regulation by Rac1 and Rac2 Guanosine Triphosphatases. Science (80-. ). 302, 445-449 (2003).
508. Crooks, G. M., Hao, Q.-L., Petersen, D., Barsky, L. W. \& Bockstoce, D. IL-3 Increases Production of B Lymphoid Progenitors from Human CD34+CD38Cells. J. Immunol. 165, 2382-2389 (2000).
509. Johnson, B. S., Mueller, L., Si, J. \& Collins, S. J. The cytokines IL-3 and GMCSF regulate the transcriptional activity of retinoic acid receptors in different in vitro models of myeloid differentiation. Blood 99, 746-753 (2002).
510. J, A., CD, H., DJ, H. \& WS, A. Synergistic effects on erythropoiesis, thrombopoiesis, and stem cell competitiveness in mice deficient in thrombopoietin and steel factor receptors. Blood 104, 1306-1313 (2004).
511. Smith, A. L., Ellison, F. M., McCoy, J. P. \& Chen, J. c-Kit Expression and Stem Cell Factor-Induced Hematopoietic Cell Proliferation Are Up-Regulated in Aged B6D2F1 Mice. Journals Gerontol. Ser. A 60, 448-456 (2005).
512. Grigoryan, A. et al. Attrition of $X$ Chromosome Inactivation in Aged Hematopoietic Stem Cells. Stem Cell Reports 16, 708-716 (2021).
513. Huls, G., van Es, J., Clevers, H., de Haan, G. \& van Os, R. Loss of Tcf7 diminishes hematopoietic stem/progenitor cell function. Leuk. 2013277 27, 1613-1614 (2012).
514. Yu, W. et al. Plexin-B2 Mediates Physiologic and Pathologic Functions of Angiogenin. Cell 171, 849-864.e25 (2017).
515. Zhong, J. F. et al. Gene expression profile of murine long-term reconstituting vs. short-term reconstituting hematopoietic stem cells. Proc. Natl. Acad. Sci. 102, 2448-2453 (2005).
516. Morandi, E. M. et al. ITGAV and ITGA5 diversely regulate proliferation and adipogenic differentiation of human adipose derived stem cells. Sci. Reports 201661 6, 1-14 (2016).
517. Qiu, J., Papatsenko, D., Niu, X., Schaniel, C. \& Moore, K. Divisional History and Hematopoietic Stem Cell Function during Homeostasis. Stem Cell Reports 2, 473-490 (2014).
518. Martino, M. M. et al. Controlling integrin specificity and stem cell differentiation in 2D and 3D environments through regulation of fibronectin domain stability. Biomaterials 30, 1089-1097 (2009).
519. Nilsson, S. K. et al. Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells. Blood 106, 1232-1239 (2005).
520. Quintana-Bustamante, O. et al. Overexpression of wild-type or mutants forms of CEBPA alter normal human hematopoiesis. Leuk. 2012267 26, 1537-1546 (2012).
521. Bolden, J. E. et al. Identification of a Siglec-F+ granulocyte-macrophage progenitor. J. Leukoc. Biol. 104, 123-133 (2018).
522. Tokuhisa, M. et al. Expression and localisation of Rab44 in immune-related cells change during cell differentiation and stimulation. Sci. Reports 2020101 10, 1-13 (2020).
523. Hamey, F. K. \& Göttgens, B. Machine learning predicts putative hematopoietic stem cells within large single-cell transcriptomics data sets. Exp. Hematol. 78, 11-20 (2019).
524. Yokota, T. et al. The endothelial antigen ESAM marks primitive hematopoietic progenitors throughout life in mice. Blood 113, 2914-2923 (2009).
525. Aguilo, F. et al. Prdm16 is a physiologic regulator of hematopoietic stem cells. Blood 117, 5057-5066 (2011).
526. Kataoka, K. et al. Evi1 is essential for hematopoietic stem cell self-renewal, and its expression marks hematopoietic cells with long-term multilineage repopulating activity. J. Exp. Med. 208, 2403-2416 (2011).
527. Haan, G. de et al. In Vitro Generation of Long-Term Repopulating

Hematopoietic Stem Cells by Fibroblast Growth Factor-1. Dev. Cell 4, 241-251 (2003).
528. Holmfeldt, P. et al. Functional screen identifies regulators of murine hematopoietic stem cell repopulation. J. Exp. Med. 213, 433-449 (2016).
529. Wang, X. et al. Stem Cell Reports Article TGF-b1 Negatively Regulates the Number and Function of Hematopoietic Stem Cells. (2018) doi:10.1016/j.stemcr.2018.05.017.
530. Lee, Y. et al. FGF signalling specifies haematopoietic stem cells through its regulation of somitic Notch signalling. Nat. Commun. 201451 5, 1-13 (2014).
531. Liu, Z. et al. High expression of PIM2 induces HSC proliferation in myelodysplastic syndromes via the IDH1/HIF1- $\alpha$ signaling pathway. Oncol. Lett. 17, 5395-5402 (2019).
532. Lu, J. et al. Fak depletion in both hematopoietic and nonhematopoietic niche cells leads to hematopoietic stem cell expansion. Exp. Hematol. 40, 307$317 . e 3$ (2012).
533. Batista, S. et al. Haematopoietic focal adhesion kinase deficiency alters haematopoietic homeostasis to drive tumour metastasis. Nat. Commun. 2014 51 5, 1-12 (2014).
534. Shen, Y. et al. Tissue Inhibitor of Metalloproteinase-3 (TIMP-3) Regulates Hematopoiesis and Bone Formation In Vivo. PLoS One 5, e13086 (2010).
535. Nakajima, H. et al. TIMP-3 recruits quiescent hematopoietic stem cells into active cell cycle and expands multipotent progenitor pool. Blood 116, 44744482 (2010).
536. Uezu, A. et al. Identification of an elaborate complex mediating postsynaptic inhibition. Science (80-. ). 353, 1123-1129 (2016).
537. Klimmeck, D. et al. Transcriptome-wide Profiling and Posttranscriptional Analysis of Hematopoietic Stem/Progenitor Cell Differentiation toward Myeloid Commitment. Stem Cell Reports 3, 858-875 (2014).
538. Kawaue, T. et al. Lzts1 controls both neuronal delamination and outer radial glial-like cell generation during mammalian cerebral development. Nat. Commun. 2019101 10, 1-18 (2019).
539. Keyes, B. E. et al. Impaired Epidermal to Dendritic T Cell Signaling Slows Wound Repair in Aged Skin. Cell 167, 1323-1338.e14 (2016).
540. Gudmundsson, K. O. et al. Prdm16 is a critical regulator of adult long-term hematopoietic stem cell quiescence. Proc. Natl. Acad. Sci. 117, 31945-31953 (2020).
541. Jeong, S.-Y., Lyu, J., Kim, J.-A. \& Oh, I.-H. Ryk modulates the niche activity of mesenchymal stromal cells by fine-tuning canonical Wnt signaling. Exp. Mol. Med. 2020527 52, 1140-1151 (2020).
542. Famili, F. et al. The non-canonical Wnt receptor Ryk regulates hematopoietic stem cell repopulation in part by controlling proliferation and apoptosis. Cell Death Dis. 2016711 7, e2479-e2479 (2016).
543. Povinelli, B. J., Srivastava, P. \& Nemeth, M. J. Related-to-receptor tyrosine kinase receptor regulates hematopoietic stem and progenitor sensitivity to myelosuppressive injury in mice. Exp. Hematol. 43, 243-252.e1 (2015).
544. Bernatchez, P. N. et al. Myoferlin Regulates Vascular Endothelial Growth Factor Receptor-2 Stability and Function *. J. Biol. Chem. 282, 30745-30753 (2007).
545. Gerber, H.-P. et al. VEGF regulates haematopoietic stem cell survival by an internal autocrine loop mechanism. Nat. 20024176892 417, 954-958 (2002).
546. Dhawan, S., Georgia, S., Tschen, S., Fan, G. \& Bhushan, A. Pancreatic $\beta$ Cell Identity Is Maintained by DNA Methylation-Mediated Repression of Arx. Dev. Cell 20, 419-429 (2011).
547. Courtney, M. et al. The Inactivation of Arx in Pancreatic $\alpha$-Cells Triggers Their Neogenesis and Conversion into Functional $\beta$-Like Cells. PLOS Genet. 9, e1003934 (2013).
548. Friocourt, G. M. \& Parnavelas, J. G. Identification of Arx targets unveils new candidates for controlling cortical interneuron migration and differentiation. Front. Cell. Neurosci. 0, 28 (2011).
549. Yoshimura, S., Gerondopoulos, A., Linford, A., Rigden, D. J. \& Barr, F. A. Family-wide characterization of the DENN domain Rab GDP-GTP exchange factors. J. Cell Biol. 191, 367-381 (2010).
550. Allaire, P. D. et al. The Connecdenn DENN Domain: A GEF for Rab35 Mediating Cargo-Specific Exit from Early Endosomes. Mol. Cell 37, 370-382 (2010).
551. Cancelas, J. A. \& Williams, D. A. Rho GTPases in hematopoietic stem cell functions. Curr. Opin. Hematol. 16, 249-254 (2009).
552. Braybrook, C. et al. Identification and Characterization of KLHL4, a Novel Human Homologue of the Drosophila Kelch Gene That Maps within the XLinked Cleft Palate and Ankyloglossia (CPX) Critical Region. Genomics 72, 128-136 (2001).
553. Feldner, A. et al. Loss of Mpdz impairs ependymal cell integrity leading to perinatal-onset hydrocephalus in mice. EMBO Mol. Med. 9, 890-905 (2017).
554. Yang, J. et al. Murine MPDZ-linked hydrocephalus is caused by hyperpermeability of the choroid plexus. EMBO Mol. Med. 11, e9540 (2019).
555. Hamazaki, Y., Itoh, M., Sasaki, H., Furuse, M. \& Tsukita, S. Multi-PDZ Domain Protein 1 (MUPP1) Is Concentrated at Tight Junctions through Its Possible Interaction with Claudin-1 and Junctional Adhesion Molecule *. J. Biol. Chem. 277, 455-461 (2002).
556. Ernkvist, M. et al. The Amot/Patj/Syx signaling complex spatially controls RhoA GTPase activity in migrating endothelial cells. Blood 113, 244-253 (2009).
557. Yamashita, M. \& Passegué, E. TNF-a Coordinates Hematopoietic Stem Cell Survival and Myeloid Regeneration. Cell Stem Cell 25, 357-372.e7 (2019).
558. Meyer, S. U. et al. Tumor Necrosis Factor Alpha and Insulin-Like Growth Factor 1 Induced Modifications of the Gene Expression Kinetics of Differentiating Skeletal Muscle Cells. PLoS One 10, e0139520 (2015).
559. Parast, M. M. \& Otey, C. A. Characterization of Palladin, a Novel Protein Localized to Stress Fibers and Cell Adhesions. J. Cell Biol. 150, 643-656 (2000).
560. Boukhelifa, M. et al. A Role for the Cytoskeleton-associated Protein Palladin in Neurite Outgrowth. https://doi.org/10.1091/mbc.12.9.2721 12, 2721-2729 (2017).
561. Khong, D. M. et al. Enhanced Hematopoietic Stem Cell Function Mediates Immune Regeneration following Sex Steroid Blockade. Stem Cell Reports 4, 445-458 (2015).
562. Liu, X.-S. et al. Disruption of palladin leads to defects in definitive erythropoiesis by interfering with erythroblastic island formation in mouse fetal liver. Blood 110, 870-876 (2007).
563. Avilés-Vázquez, S. et al. Global gene expression profiles of hematopoietic stem and progenitor cells from patients with chronic myeloid leukemia: the
effect of in vitro culture with or without imatinib. Cancer Med. 6, 2942-2956 (2017).
564. Graaf, C. A. de et al. Regulation of hematopoietic stem cells by their mature progeny. Proc. Natl. Acad. Sci. 107, 21689-21694 (2010).
565. Kanehisa, M. \& Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27-30 (2000).
566. Xie, S. Z. et al. Sphingolipid Modulation Activates Proteostasis Programs to Govern Human Hematopoietic Stem Cell Self-Renewal. Stem Cell 25, 639$653 . e 7$ (2019).
567. Tothova, Z. et al. FoxOs Are Critical Mediators of Hematopoietic Stem Cell Resistance to Physiologic Oxidative Stress. Cell 128, 325-339 (2007).
568. Jansson, L. \& Larsson, J. Normal Hematopoietic Stem Cell Function in Mice with Enforced Expression of the Hippo Signaling Effector YAP1. PLoS One 7, e32013 (2012).
569. Sasine, J. P. et al. Wild-type Kras expands and exhausts hematopoietic stem cells. JCI Insight 3, (2018).
570. Hinge, A. et al. p190-B RhoGAP and intracellular cytokine signals balance hematopoietic stem and progenitor cell self-renewal and differentiation. Nat. Commun. 8, 14382 (2017).
571. Florian, M. C. et al. Aging alters the epigenetic asymmetry of HSC division. PLOS Biol. 16, e2003389 (2018).
572. Liu, W. et al. Rational identification of a Cdc42 inhibitor presents a new regimen for long-term hematopoietic stem cell mobilization. Leuk. 2018333 33, 749-761 (2018).
573. Gao, Y., Dickerson, J. B., Guo, F., Zheng, J. \& Zheng, Y. Rational design and characterization of a Rac GTPase-specific small molecule inhibitor. Proc. Natl. Acad. Sci. 101, 7618-7623 (2004).
574. Shang, X. et al. Small-molecule inhibitors targeting G-protein-coupled Rho guanine nucleotide exchange factors. Proc. Natl. Acad. Sci. 110, 3155-3160 (2013).
575. Surviladze, Z. et al. Three small molecule pan activator families of Ras-related GTPases. Probe Reports from NIH Mol. Libr. Progr. (2010).
576. Hey, Y.-Y., Tan, J. K. H. \& O'Neill, H. C. Redefining Myeloid Cell Subsets in Murine Spleen. Front. Immunol. 0, 652 (2016).
577. KL, W., J, A. \& HP, K. Hematopoietic stem cell expansion and gene therapy. Cytotherapy 13, 1164-1171 (2011).
578. Epah, J. \& Schäfer, R. Implications of hematopoietic stem cells heterogeneity for gene therapies. Gene Ther. 2021289 28, 528-541 (2021).
579. E, D. et al. A functional screen to identify novel effectors of hematopoietic stem cell activity. Cell 137, 369-379 (2009).
580. S, C., BP, L., ML, S. \& SJ, M. Prdm16 promotes stem cell maintenance in multiple tissues, partly by regulating oxidative stress. Nat. Cell Biol. 12, 9991006 (2010).
581. KO, G. et al. Prdm16 is a critical regulator of adult long-term hematopoietic stem cell quiescence. Proc. Natl. Acad. Sci. U. S. A. 117, 31945-31953 (2020).
582. T, Y. et al. The endothelial antigen ESAM marks primitive hematopoietic progenitors throughout life in mice. Blood 113, 2914-2923 (2009).
583. AG, O. et al. The adhesion molecule esam1 is a novel hematopoietic stem cell marker. Stem Cells 27, 653-661 (2009).
584. P, H. et al. Functional screen identifies regulators of murine hematopoietic stem cell repopulation. J. Exp. Med. 213, 433-449 (2016).
585. J, Y. et al. Murine MPDZ-linked hydrocephalus is caused by hyperpermeability of the choroid plexus. EMBO Mol. Med. 11, (2019).
586. Y, H., M, I., H, S., M, F. \& S, T. Multi-PDZ domain protein 1 (MUPP1) is concentrated at tight junctions through its possible interaction with claudin-1 and junctional adhesion molecule. J. Biol. Chem. 277, 455-461 (2002).
587. M, E. et al. The Amot/Patj/Syx signaling complex spatially controls RhoA GTPase activity in migrating endothelial cells. Blood 113, 244-253 (2009).
588. Cosentino, M., Marino, F. \& Maestroni, G. J. M. Sympathoadrenergic modulation of hematopoiesis: a review of available evidence and of therapeutic perspectives. Front. Cell. Neurosci. 0, 302 (2015).
589. Liu, W. et al. Rational identification of a Cdc42 inhibitor presents a new regimen for long-term hematopoietic stem cell mobilization. Leukemia 33, 749761 (2019).
590. DC, K. et al. Dynamic interaction networks in a hierarchically organized tissue. Mol. Syst. Biol. 6, (2010).
591. Bruns, I. et al. Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion. Nat. Med. 2014 2011 20, 1315-1320 (2014).
592. Zhao, M. et al. Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. Nat. Med. 20142011 20, 1321-1326 (2014).
593. S, S., B, J. \& JR, K. Protection of hematopoietic stem cells from stress-induced exhaustion and aging. Curr. Opin. Hematol. 27, 225-231 (2020).
594. Ma, Z. et al. Hes1 deficiency causes hematopoietic stem cell exhaustion. Stem Cells 38, 756-768 (2020).
595. Rodriguez-Fraticelli, A. E. et al. Single-cell lineage tracing unveils a role for TCF15 in haematopoiesis. Nat. 20205837817 583, 585-589 (2020).
596. Bode, D., Cull, A. H., Rubio-Lara, J. A. \& Kent, D. G. Exploiting Single-Cell Tools in Gene and Cell Therapy. Front. Immunol. 0, 2775 (2021).
597. Stoeckius, M. et al. Simultaneous epitope and transcriptome measurement in single cells. Nat. Methods 14, 865-868 (2017).
598. Mimitou, E. P. et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat. Methods 16, 409412 (2019).
599. CJ, F. et al. Multimodal pooled Perturb-CITE-seq screens in patient models define mechanisms of cancer immune evasion. Nat. Genet. 53, 332-341 (2021).
600. NP, C. et al. Small-Molecule Screens: A Gateway to Cancer Therapeutic Agents with Case Studies of Food and Drug Administration-Approved Drugs. Pharmacol. Rev. 69, 479-496 (2017).
601. Sagoo, P. \& Gaspar, H. B. The transformative potential of HSC gene therapy as a genetic medicine. Gene Ther. 2021 1-19 (2021) doi:10.1038/s41434-021-00261-x.
602. Bordignon, C. et al. Gene Therapy in Peripheral Blood Lymphocytes and Bone Marrow for ADA- Immunodeficient Patients. Science (80-. ). 270, (1995).
603. Brar, G. A. \& Weissman, J. S. Ribosome profiling reveals the what, when, where and how of protein synthesis. Nat. Rev. Mol. Cell Biol. 20151611 16, 651-664 (2015).
604. J, Z., B, Q., R, N., CMT, S. \& G, Z. Translatomics: The Global View of

Translation. Int. J. Mol. Sci. 20, (2019).
605. VanInsberghe, M., van den Berg, J., Andersson-Rolf, A., Clevers, H. \& van Oudenaarden, A. Single-cell Ribo-seq reveals cell cycle-dependent translational pausing. Nat. 20215977877 597, 561-565 (2021).
606. Wagenblast, E. et al. Functional profiling of single CRISPR/Cas9-edited human long-term hematopoietic stem cells. Nat. Commun. 2019 101 10, 1-11 (2019).
607. Kiel, M. J. et al. SLAM Family Receptors Distinguish Hematopoietic Stem and Progenitor Cells and Reveal Endothelial Niches for Stem Cells. Cell 121, 1109-1121 (2005).

