



UNIVERSITY OF  
CAMBRIDGE

# The Effects of Methyltransferase Genes on Human Hematopoietic Stem Cell Function

**ADITI VEDI**

**Trinity College**

**Submission Date: January 2023**

This thesis is submitted for the degree of Doctor of Philosophy

Primary Supervisor: Dr Elisa Laurenti

Secondary Supervisor: Prof Berthold Göttgens



# 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 relevant Degree Committee.

# Abstract

## The Effects of Methyltransferase Genes on Human Haematopoietic Stem Cell Function

Dr Aditi VEDI

The highly structured and hierarchical haematopoietic system is defined by the existence of a pool of haematopoietic stem cells (HSCs) at the apex, which possess both multipotent differentiation and self-renewal properties. The HSC pool is heterogenous in terms of self-renewal and differentiation capacity, and HSCs with various genetic alterations lie intermingled with wild type (WT) HSCs. Clonal expansion of HSCs is an ineluctable feature of advancing age, and age related clonal haematopoiesis (ARCH) is associated with an increased risk of haematological malignancy, atherosclerosis and cardiovascular disease. The functional effects of these clonal expansions remain to be thoroughly investigated in primary human HSCs. In this thesis, I have focused on 2 methyltransferase genes, *PRMT5* and *DNMT3A*, which are both epigenetic modulators of HSC function.

Pharmacological inhibition of *PRMT5* in healthy human cord blood (hCB) haematopoietic stem and progenitor cells (HSPCs) leads to reduced HSPC proliferation and a relative differentiation block at the CD34+ progenitor level in a dose dependent manner. While *PRMT5* has been identified by others as an important therapeutic target in the context of haematological malignancies, my work indicates its clinical applicability may be limited by the likely toxicity of *PRMT5* inhibition on HSPCs, which could lead to significant myelosuppression.

Mutations in the gene *DNMT3A* are the most common in ARCH and are associated with haematological and non-haematological diseases with significant morbidity and mortality. The *R882* mutation is the most common within *DNMT3A*, especially in the context of haematological disease. I used single cell biology tools to obtain functional data from 12 primary human samples from ARCH individuals and acute myeloid leukaemia (AML) patients. A total of 2990 single HSC/multi-potent progenitor (MPP) derived colonies from *DNMT3A* mutant HSCs/MPPs and internal WT control HSC/MPPs from within each individual were compared. *DNMT3A R882* mutation confers an intrinsic differentiation phenotype, with *DNMT3A R882* mutant HSC/MPPs displaying more efficient differentiation towards neutrophils and less efficient monocyte differentiation *in vitro* compared to internal WT control HSC/MPPs in 6 independent individuals. No overall lineage bias is observed in *DNMT3A R882* mutant HSC/MPPs, and no selective differentiation advantage. RNA sequencing of single HSC/MPP derived colonies containing mature monocytes confirmed that *DNMT3A R882* mutant HSC/MPPs produce less mature monocytes, and are marked by upregulation of metabolic and inflammatory gene pathways.

Overall, these functional phenotypes provide critical insight into the inherent functional effects of *DNMT3A R882* mutation at the HSC/MPP level in humans, beyond the previously described transcriptional priming towards more quiescent HSCs and increased self-renewal phenotypes. Further mechanistic understanding will be required to elucidate the link between the myeloid differentiation phenotypes observed in my study and the inflammatory conditions associated with ARCH.

# Acknowledgements

I dedicate this work to my amazing husband, Dr Brent O’Carrigan, and my 3 most wonderful children, Imogen, Oliver and Elliot, without whom this PhD would not be complete, nor anywhere near what it is today. Brent is the very drive and motivation for all that I have achieved over the past 20 years that I have known him. It is because of him that we have both dedicated ourselves to the study of human cancer in Cambridge. Without Brent, Imogen, Oliver and Elliot, I would not be anywhere near the person I am today with the experiences I have gained and all that I have become. These most amazing people have given me the courage, motivation, and strength to persist and achieve all that I have and make me and this work whole. I am so grateful for your unconditional love, support, hugs, ongoing motivation and for believing in me all the way through this marathon and beyond!

Of course, this work would not exist without my supervisor, Dr Elisa Laurenti, who has taught me a great deal over the course of this PhD, not all of which is science! She has dedicated countless hours to teaching me practical laboratory techniques, bioinformatic skills, and interpersonal skills. She read and re-read my thesis, abstracts, manuscripts, and presentations countless times. For her patience, kindness, and scientific acumen I am forever grateful.

I also appreciate all contributions from my fellow past and present lab colleagues, Daniel Hayler, Emily Mitchell, Aleksandra Krzywon, Antonella Santoro and Kendig Sham, who all contributed to experiments and bioinformatic analysis in this thesis. Much appreciation also belongs to the core facilities at the Cambridge Stem Cell Institute, in particular the phenotyping hub, who spent many weary hours sorting thousands of cells into immeasurably small volumes of liquid with me! Andy, Natalia, Daniel and Nika were always so patient with my requests to gently pander my cells! Also, to the sequencing and iPS facilities for helping with this work.

Colleagues in Dr Liran Shlush’s lab at the Weizmann Institute also contributed significantly to the genetic analyses included in this thesis: Tamir Biezuner, Amos Tuval (who was particularly kind and taught me much of what I know about reading raw sequencing data!), Yoni Moskovitz and Dr Shlush himself! I also thank them for their hospitality during my visit to Rehovot halfway through this PhD, in 2019 and for making me and my family feel so welcome.

I would also like to thank my clinical colleagues, friends and mentors for the ongoing encouragement. Dr Shivani Bailey, an incredibly talented clinician and close friend provided me with just the right support and advice to complete this thesis and see the past 5 years of labour come to fruition. Dr Amos Burke, Prof Matthew Murray and Dr Sam Behjati have all been great mentors throughout this process. Thank you to Dr Lynley Marsall and Dr Donna Lancaster for believing I could do this in the first place! And of course, to all the wonderful patients who provide me with the motivation and reason to keep going. No child should die of cancer and hopefully our research will achieve this goal one day.

Without the support of the Gates Cambridge Foundation, I could never have achieved this PhD. Aside from the financial support throughout my course, Prof Barry Everitt, Gates Cambridge Provost, supported me in taking the necessary leave to grow my family whilst undertaking my PhD.

Thank you to my lovely friends Dr Mariya Chhatriwala and Dr Elian Fink, who talked me through thesis writing and academia over many a glass of wine and dinners out.

Finally, I would like to mention my parents and wonderful sister, who have been my cheerleaders all the way. Thank you for believing in me and keeping motivated throughout this long journey.

*Dedicated to BOC, IOC, OOC and EOC*



## Table of Contents

<b>Declaration</b> .....	<b>2</b>
<b>Abstract</b> .....	<b>3</b>
<b>Acknowledgements</b> .....	<b>4</b>
<b>Abbreviations</b> .....	<b>11</b>
<b>Chapter 1 Introduction</b> .....	<b>13</b>
<b>1.1 Background</b> .....	<b>13</b>
<b>1.2 Haematopoietic Stem Cells (HSCs)</b> .....	<b>13</b>
<b>1.3 The haematopoietic hierarchy</b> .....	<b>15</b>
1.3.1 HSC heterogeneity .....	16
1.3.2 HSC quiescence .....	17
1.3.3 Regulators of HSC quiescence .....	19
<b>1.4 HSCs across the lifespan</b> .....	<b>20</b>
<b>1.5 Clonal Haematopoiesis (CH)</b> .....	<b>21</b>
1.5.1 Definitions – ARCH, CHIP and pre-leukaemia.....	22
1.5.2 Epidemiology of clonal haematopoiesis .....	22
1.5.3 Non-haematological associations of CH.....	23
<b>1.6 Leukemic stem cells</b> .....	<b>24</b>
<b>1.7 Clinical significance</b> .....	<b>25</b>
<b>1.8 Overall Aims</b> .....	<b>27</b>
<b>Chapter 2 Materials and Methods</b> .....	<b>28</b>
<b>2.1 Sample collection and preparation</b> .....	<b>28</b>
2.1.1 Human umbilical cord blood (hCB) samples .....	28
2.1.2 Peripheral blood (PB) samples .....	29
2.1.2.1 Patient PB samples .....	29
2.1.2.2 Patient selection .....	29

<b>2.1.3 Healthy donor PB samples .....</b>	<b>30</b>
2.1.3.1 Venesection .....	30
2.1.3.2 Leukocyte cones .....	30
2.1.3.3 Bone marrow sample .....	31
<b>2.2 In vivo engraftment .....</b>	<b>31</b>
<b>2.3 Cell preparation and sorting strategy.....</b>	<b>31</b>
<b>2.3.1 HSPC isolation from hCB CD34+ cells .....</b>	<b>31</b>
<b>2.3.2 HSPC isolation from peripheral blood and BM samples .....</b>	<b>33</b>
2.3.2.1 CD34+ selection .....	33
2.3.2.2 Sorting pre-leukemic stem cells.....	35
2.3.2.3 Methylcellulose (MC) assays .....	37
2.3.2.4 Liquid culture assays for hCB CD34+ cells.....	38
2.3.2.5 Apoptosis Assays .....	40
2.3.2.6 Cell cycle analysis assays .....	40
<b>2.3.3 Tex cells .....</b>	<b>41</b>
<b>2.3.4 Peripheral blood and bone marrow samples .....</b>	<b>44</b>
<b>2.4 Treatment in vitro .....</b>	<b>45</b>
<b>2.4.1 Pharmacological inhibitors.....</b>	<b>45</b>
<b>2.4.2 Lentiviral overexpression .....</b>	<b>45</b>
2.4.2.1 Cloning lentiviral overexpression plasmids .....	45
2.4.2.2 Producing lentiviral vectors .....	46
2.4.2.3 Transduction of HSCs with PRMT5 lentiviral vectors.....	47
<b>2.5 Induced pluripotent stem cells (iPSC).....</b>	<b>48</b>
<b>2.5.1 Derivation of iPSCs.....</b>	<b>48</b>
<b>2.5.2 Genotyping iPSC clones.....</b>	<b>48</b>
<b>2.6 Sequencing .....</b>	<b>49</b>
<b>2.6.1 Targeted genotyping .....</b>	<b>49</b>
2.6.1.1 Bulk primary PB samples .....	49
2.6.1.2 Single cell derived colonies.....	49
<b>2.6.2 Whole genome sequencing .....</b>	<b>52</b>
<b>2.6.3 RNA sequencing .....</b>	<b>53</b>

<b>2.7 Data analysis .....</b>	<b>53</b>
<b>2.7.1 Flow cytometry phenotyping analysis .....</b>	<b>53</b>
<b>2.7.2 Comparing phenotype and genotype .....</b>	<b>54</b>
2.7.2.1 FlowPAC.....	54
2.7.2.2 CytoTree .....	55
<b>2.7.3 RNA sequencing analysis.....</b>	<b>57</b>
2.7.3.1 Sequencing reads analysis .....	57
2.7.3.2 Gene expression analysis.....	57
<b>2.7.4 Literature review .....</b>	<b>58</b>
<b>2.7.5 Statistical analyses and software .....</b>	<b>58</b>
<b>2.7.6 Nomenclature for genetic mutations.....</b>	<b>59</b>
<b>Chapter 3     <i>HSC function in health</i>.....</b>	<b>60</b>
<b>3.1 Introduction.....</b>	<b>60</b>
<b>3.1.1 Definitions .....</b>	<b>60</b>
<b>3.1.2 Regulators of HSC function.....</b>	<b>60</b>
3.1.2.1 Methyltransferases.....	60
3.1.2.2 Spliceosome genes .....	62
<b>3.1.3 Clinical applications .....</b>	<b>63</b>
<b>3.2 Aims .....</b>	<b>64</b>
<b>3.3 Results .....</b>	<b>65</b>
<b>3.3.1 Genes with altered expression early in LT-HSC activation.....</b>	<b>65</b>
3.3.1.1 Genes upregulated on quiescence exit.....	68
3.3.1.2 Genes downregulated on quiescence exit.....	69
<b>3.3.2 Targeting genes with altered expression on quiescence exit .....</b>	<b>70</b>
3.3.2.1 PRMT5 .....	70
3.3.2.2 CLK1.....	85
<b>3.4 Conclusions.....</b>	<b>86</b>
<b>3.4.1 PRMT5 .....</b>	<b>87</b>
<b>3.4.2 CLK1.....</b>	<b>87</b>

<b>Chapter 4</b>	<b><i>HSC function in clonal haematopoiesis and disease</i></b>	<b>88</b>
<b>4.1</b>	<b><i>Introduction</i></b>	<b>88</b>
4.1.1	Genetic mutations in ARCH	88
4.1.2	DNA methyltransferases	89
4.1.3	Transcriptional characterisation of ARCH mutations	92
4.1.3.1	Mouse	92
4.1.3.2	Human	92
<b>4.2</b>	<b><i>Aims</i></b>	<b>93</b>
<b>4.3</b>	<b><i>Results</i></b>	<b>94</b>
4.3.1	Analysis of the clonal dynamics of DNMT3A R882 mutant HSC/MPPs in an ARCH individual	94
4.3.2	Establishing a cohort to study the effects of DNMT3A R882 mutation on HSC/MPP differentiation	101
4.3.3	Isolating HSCs distinct from LSCs in AML samples	104
4.3.4	Myeloid skewing of PB HSC/MPP differentiation in AML patients compared with healthy age matched controls	107
4.3.5	No broad selective advantage on HSC/MPPs during in vivo differentiation conferred by DNMT3A R882 mutation	108
4.3.6	Assessment of differentiation phenotype of DNMT3A R882 mutant HSC/MPPs vs WT	110
4.3.7	DNMT3A R882 mutation confers no major alteration in HSC/MPP cell surface markers	111
4.3.8	DNMT3A R882 mutation has no effect on the proliferation capacity of HSC/MPPs	111
4.3.8.1	DNMT3A R882 mutation confers no overall erythroid/ myeloid lineage bias	112
4.3.8.2	Changes in differentiation observed within the myeloid branch	115
<b>4.4</b>	<b><i>Conclusions and future work</i></b>	<b>129</b>
<b>Chapter 5</b>	<b><i>Discussion</i></b>	<b>130</b>
<b>5.1</b>	<b><i>PRMT5</i></b>	<b>130</b>
5.1.1	Clinical implications	130
5.1.2	PRMT5 inhibition induced decreased proliferation and differentiation block	132

5.1.3 Mechanistic insights.....	133
5.1.4 Validation of phenotype by genetic modulation in hHSPCs .....	134
<b>5.2 Future studies on normal HSCs and progenitors .....</b>	<b>135</b>
<b>5.3 DNMT3A mutations in Clonal Haematopoiesis .....</b>	<b>135</b>
5.3.1 Early acquisition in life and the haematopoietic hierarchy .....	136
5.3.2 Equal contribution to erythroid/myeloid lineages .....	137
5.3.3 Novel monocyte/granulocyte phenotype .....	138
5.3.4 No selective advantage during in vivo differentiation .....	139
<b>5.4 ARCH associated disease .....</b>	<b>140</b>
<b>5.5 Future studies in clonal haematopoiesis.....</b>	<b>141</b>
<b>References.....</b>	<b>143</b>

## Abbreviations

AML	Acute Myeloid Leukaemia
ANOVA	Analysis of Variance
ARCH	Age Related Clonal Haematopoiesis
BM	Bone Marrow
CAR-T	Chimeric Antigen Receptor T cell
CB	Cord Blood
CBSB	Cambridge Blood and Stem Cell
CFC	Colony Forming Cells
CH	Clonal Haematopoiesis
CHIP	Clonal Haematopoiesis of Indeterminate Potential
CLL	Chronic Lymphocytic Leukaemia
CMML	Chronic Monomyelocytic Leukaemia
COPD	Chronic Obstructive Pulmonary Disease
CSCI	Cambridge Stem Cell Institute
CSCI	Cancer Stem Cell
DESeq	Differential Expression in Sequencing
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNMT	DNA Methyltransferase
EDTA	Ethylenediaminetetraacetic acid
ESC	Embryonic Stem Cell
ET	Essential Thrombocythaemia
FACS	Fluorescence Activated Cell Sorting
FBS	Fetal Bovine Serum
GCSF	Granulocyte Colony Stimulating Factor
GSEA	Gene Set Enrichment Analysis
GSVA	Gene Set Variation Analysis
GWAS	Genome Wide Association Study
hCB	Human Cord Blood
HDL	High Density Lipoprotein
hHSC	Human Haematopoietic Stem Cells
HR	Hazard Ratio
HSC	Haematopoietic Stem Cells
hSCF	human Stem Cell Factor
HSCT	Haematopoietic Stem Cell Transplant
HSPC	Haematopoietic Stem and Progenitor Cells
IMDM	Isco Modified Dulbecco Media
iPSC	Induced Pluripotent Stem Cells
KO	Knock Out
LDL	Low Density Lipoprotein
LiHep	Lithium Heparin
LSC	Leukaemic Stem Cells

LSPC	Leukaemic Stem and Progenitor Cells
LT-HSC	Long Term Haematopoietic Stem Cells
LV	Lentiviral
MC	Methylcellulose
MDS	Myelodysplastic Syndrome
MEM	Myeloerythroid medium
MFI	Median Fluorescence Index
MM	Multiple Myeloma
MNC	Mononuclear cells
MOI	Multiplicity of Infection
mPB	Mobilised Peripheral Blood
MPN	Myeloproliferative Neoplasm
MPP	Multipotent Progenitors
MRD	Minimal Residual Disease
mRNA	Messenger Ribonucleic Acid
NGS	Next Generation Sequencing
NSG	NOD SCID Gamma mice
PB	Peripheral Blood
PBS	Phosphate Buffered Solution
PCR	Polymerase Chain Reaction
pL-HSC	Pre-Leukaemic HSCs
pL-HSPC	Pre-Leukemic Haematopoietic Stem and Progenitor Cells
pLM	Pre-Leukaemic Mutation
PRMT	Protein Methyltransferase
qPCR	Quantitative Polymerase Chain Reaction
RDW	Red Cell Distribution Width
RNA	Ribonucleic Acid
RNAseq	RNA Sequencing
RT	Room Temperature
scRNAseq	Single Cell RNA Sequencing
shRNA	short hairpin Ribonucleic Acid
snRNA	small nuclear Ribonucleic Acid
snRNP	small nuclear Ribonucleic Proteins
ST-HSC	Short Term Haematopoietic Stem Cells
TF	Transcription Factor
UK	United Kingdom
VAF	Variant Allele Frequency
WGS	Whole Genome Sequencing
WT	Wild Type

# Chapter 1 Introduction

## 1.1 Background

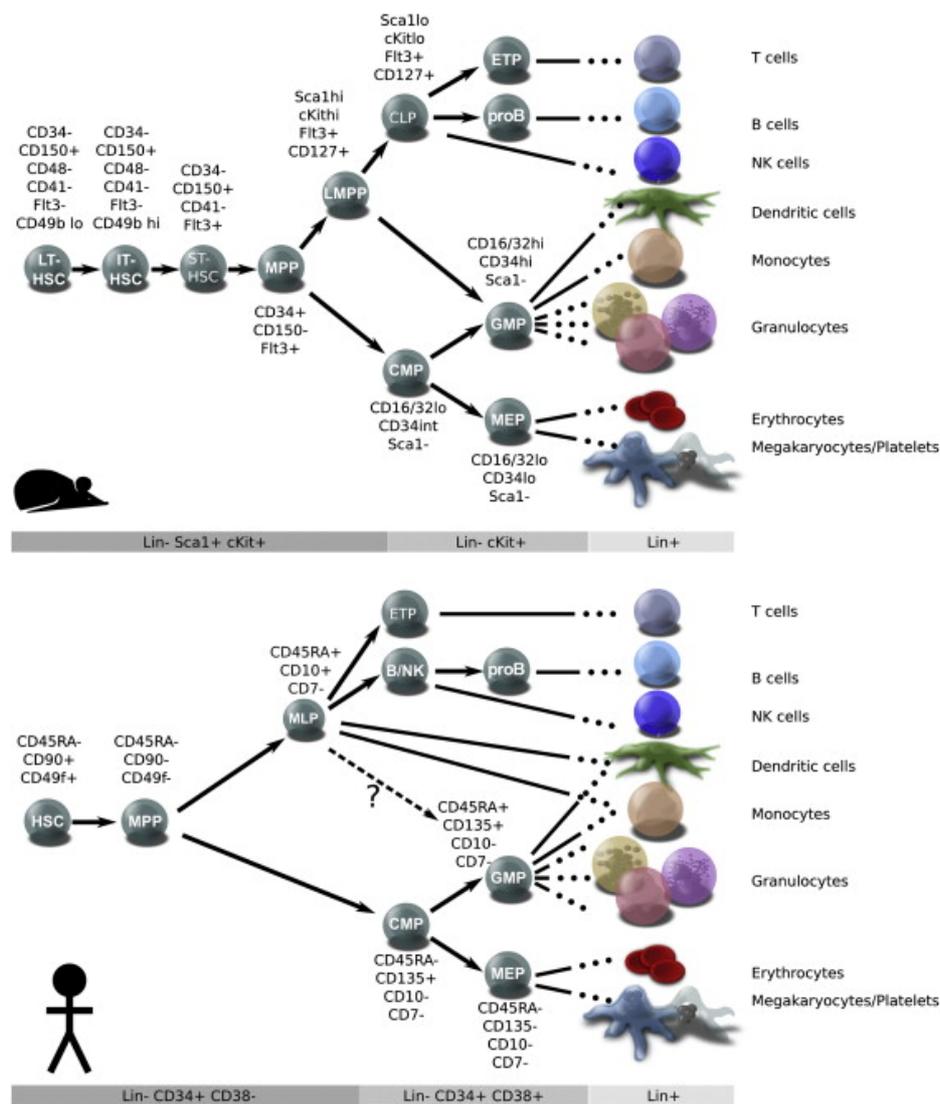
The term stem cell (or “stammzelle”) appears in the scientific literature as early as 1868, with German biologist Ernst Haeckel being credited with coining this word, inspired by his Darwinist beliefs to describe common ancestor cells from which multicellular organisms descended<sup>1</sup>. This concept was subsequently extrapolated to the fertilised egg, and later was applied by histopathologists to normal and leukemic haematopoiesis. The notion of a common ancestor giving rise to differentiated cell lineages was postulated in 1909 by Russian biologist A. Maximow<sup>2</sup>. The hierarchical structure of the haematopoietic system was first described almost half a century ago by Till and McCulloch<sup>3</sup>, when the myeloid and lymphoid branches were identified and blood was recognised as a highly heterogeneous system, with a small stem cell pool that was able to give rise to multilineage colonies<sup>4</sup>. The first proof of concept of the repopulation capacity of the haematopoietic system was seen in mice when an appropriate number of bone marrow (BM) cells injected into isologous hosts, previously exposed to supra-lethal doses of total body irradiation, were able to generate splenic colony forming units<sup>5</sup>. Our knowledge and understanding of haematopoiesis has increased considerably over the last century.

The haematopoietic system is among the most complex in human biology and is one of the most well researched. It is a highly regenerative structure, with carefully balanced cellular proliferation and differentiation, producing in excess of a trillion cells each day<sup>6</sup>, making it a tractable system to study the clonal haematopoiesis process. Clonal expansions, which are implicated both in diseases of overproduction, such as malignancies and myeloproliferative disorders; as well as syndromes of underproduction or bone marrow failure such as aplastic anaemia. Detailed study of the mechanisms and regulation of haematopoiesis, cell cycling and the careful balance between cellular renewal and quiescence is crucial for the prevention and treatment of haematopoietic disorders.

## 1.2 Haematopoietic Stem Cells (HSCs)

While resident stem cells are a feature of most mammalian adult tissues, HSCs are the most well-characterised and studied, largely due to the relative ease by which they can be acquired from BM, cord blood (CB), and mobilised peripheral blood (mPB). However, the rarity of pure HSCs poses a major challenge in their study, as they represent less than 1 in  $10^6$  cells in human BM and CB<sup>7</sup>, which are the richest sources of HSCs in humans. Study of human HSCs (hHSCs) currently relies on purification techniques based on canonical cell surface markers, initially described over 30 years ago using positive and negative selection<sup>8</sup>. Over the past two decades these have become sophisticated enough to isolate hHSCs by fluorescence activated cell sorting (FACS) using differential expression of cell surface markers on HSC subsets<sup>9,10</sup>.

The initial functional and phenotypic characterisation of HSCs were derived from mouse models, however recent advances in single cell isolation<sup>11–13</sup> and transplantation<sup>10</sup> has furthered our understanding of hHSCs significantly. There are several critical differences between mouse and human stem cell biology including genetic diversity, size, ecology, lifespan and potential for accumulation of genetic mutations, which have implications for the behaviour and malignant potential of HSCs<sup>9</sup>, hence studies in hHSCs are required to adequately understand the evolution of HSCs and disease in humans. Cell surface marker definitions of HSCs also differ significantly between mouse and human (Fig 1.2-1)<sup>9</sup>.



**Fig 1.2-1 Canonical cell surface markers used to identify mouse and human hematopoietic cells across the hierarchy.** Differences in cell surface markers between mouse and human HSCs and mature blood cells. *Doulatov et al, Cell Stem Cell, 2012.*

Over the past 4-5 decades, our understanding of HSCs and their role in haematological disease has increased substantially, mainly with the advent of FACS<sup>14–16</sup>, which has allowed multi-parametric analysis and isolation of

blood cells and their subtypes based on their cell-surface markers for *in vitro* and *in vivo* assays to study their functional biology.

More recently, gene expression studies, particularly in single cells, have allowed comparison of transcriptional differences within this heterogeneous pool of cells down to the single-cell level<sup>17</sup>. Modulation of gene expression has been achieved with lentiviral vector (LV) transduction permitting overexpression or short hairpin RNA-based (shRNA) gene silencing. Experimental models generated *in vivo* have been used to transplant experimentally manipulated CD34+ HSCs into sub-lethally irradiated NOD SCID gamma (NSG) mice, or *in vitro* with seeding in surrogate stromal environments to study the functional effects of gene perturbations on cell fate decisions of HSCs<sup>9</sup>. Recent advances in single cell genomic studies and novel gene editing technology such as the CRISPR/Cas9<sup>18,19</sup> system have facilitated investigations into the genetic basis of HSC regulation<sup>20,21</sup>.

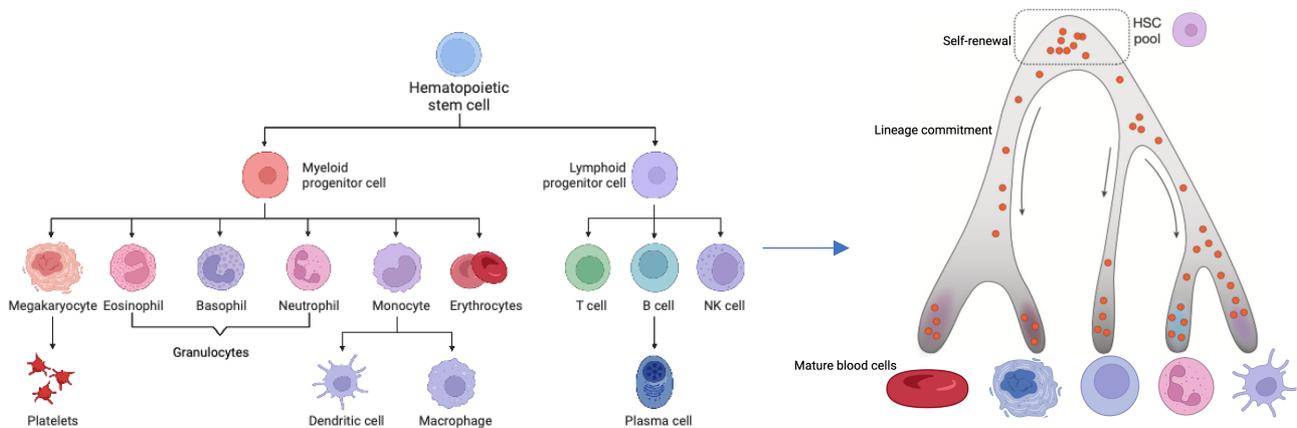
## 1.3 The haematopoietic hierarchy

The tree-like model of haematopoiesis originates from the stem cell concept in which multipotent stem cells give rise to their progeny through an ordered and predictable series of branches. Haematopoiesis is recognised as a highly coordinated hierarchical process that regulates blood cell production and maintenance under homeostatic conditions. At the apex of this structured hierarchy reside HSCs, with terminally differentiated mature cells at the bottom and several stages of intermediate progenitor cells in between<sup>9</sup>. HSCs are defined by their self-renewal and multipotent differentiation capacity, with each daughter cell becoming more lineage restricted as their progeny progress down the hierarchy. Loss of HSC self-renewal capacity has long been believed to precede lineage commitment in HSCs, as evidenced by the existence of multipotent progenitors (MPPs), a subpopulation of HSCs that maintain multipotent potential with transient repopulation capacity<sup>22</sup>. Progenitors have greater proliferative ability than stem cells and rapidly proceed towards differentiation, but with limited to no self-renewal and restricted lineage differentiation capacity.

Characterisation of progenitors downstream from HSCs led to the development of the classic haematopoietic tree with lymphoid progenitors branching above myeloid, erythroid, and megakaryocytic progenitors, progressing to more lineage restriction up to the mature circulating blood cells that we can characterise using well-recognised cell surface markers using FACS. The tree has developed significantly over the past two decades with the introduction of other surface markers suggesting that myeloid and lymphoid cell fates remain linked until further down the hierarchy<sup>12</sup>, while megakaryocytes branch earlier<sup>23,24</sup>. Our understanding of MPPs has also significantly developed, as the HSC pool is increasingly recognised as being heterogeneous (Fig 1.3-1)<sup>25</sup>.

Cell fate decisions are usually associated with changes in gene expression, which are often preceded by epigenetic changes in gene regulatory regions marking them as active, silent or poised. To understand cell fate decisions and the sequence of developmental events, a good understanding of epigenetic and transcriptional changes between closely related HSCs and progenitors is important. Many groups have systematically profiled the transcriptome of human HSC and progenitor cell populations and identified a continuous landscape of

transcriptional programs that spans across population and lineage boundaries<sup>26,27</sup>. This supports the hypothesis that the haematopoietic hierarchy is in fact a continuum of differentiation rather than a system of distinctly demarcated stem-cell and lineage-commitment circuits (Fig 1.3-1).



**Fig 1.3-1 Evolution of the haematopoietic hierarchy.** The haematopoietic hierarchy is increasingly being understood as a continuum between stem cells, progenitors and mature blood cells, with decreasing self-renewal and increasing lineage commitment, as transcriptional data identify gene expression that spans across population and lineage boundaries. The HSC pool is also regarded as highly heterogeneous; *adapted from Laurenti and Gottgens, Nature 2018. Made with BioRender.*

A single HSC can give rise to a large number of progenies, exponentially increasing with each cell division. Studies of the divisional history of HSCs have also demonstrated that only a limited pool of the most dormant HSCs can provide robust, life-long reconstitution after transplantation. Long-term regenerative potential is linked to the period between cell divisions, with this potential lost following the fifth division<sup>28</sup>. In humans, the transition from fetal to adult haematopoiesis occurs within the first year of life, as measured by the rapid decline in telomere length in granulocytes as a surrogate marker for the division rate of HSCs<sup>29</sup>.

### 1.3.1 HSC heterogeneity

Mature blood cells have a rapid turnover and hence high levels of production at the progenitor level, while the HSC pool is more stable. Stem cells have long been recognised as distinct from progenitor cells due to their inherent capacity both for self-renewal and the generation of differentiated progeny<sup>30</sup>. The demarcation between stem and progenitor cell populations is increasingly being recognised as being less distinct with the help of technologies such as single cell RNA sequencing (scRNAseq)<sup>31,32</sup> that have been used to outline the broad molecular landscape haematopoietic cells within the differentiation hierarchy<sup>33</sup>. HSC pools feature more heterogeneity than previously understood, with several subpopulations within long term HSCs (LT-HSCs) and short term HSCs (ST-HSCs) identified by functional, surface marker and transcriptional analyses<sup>34</sup>. Although these

subpopulations share many genes when isolated under homeostatic quiescent states<sup>10,33,35</sup>, they diverge transcriptionally during cell cycle activation *in vitro* and *in vivo*<sup>36</sup>.

HSCs are functionally heterogeneous in their cell cycle properties, self-renewal, differentiation and their durability of engraftment<sup>25</sup>. The hHSC pool is characterised based on cell surface marker expression as Lin-CD34+CD38-CD45RA<sup>-37,38</sup>, with LT-HSCs (CD90+CD49f+) and ST-HSCs (CD90-CD49f-) being a subset of this pool<sup>10,37</sup>. Human HSCs are generally categorised as CD34+ LT-HSCs<sup>10,36</sup> and ST-HSCs or MPPs. HSC subsets are defined by their ability to repopulate in transplantation assays, with LT-HSCs being able to generate robust long-term engraftment for more than 16 weeks in primary transplantation and at least in secondary transplantation<sup>11,39</sup>, mainly due to their unique self-renewal capacity. ST-HSCs/MPPs can produce transient multilineage engraftment<sup>40-42</sup>.

Heterogeneity of differentiation outputs within the HSC and MPP compartments has also been well described<sup>43,44</sup>, with platelet-biased HSCs having been identified<sup>23,24</sup>, as well as CD71+ HSCs/MPPs that exclusively produce erythrocytes and megakaryocytes<sup>45</sup>. HSC heterogeneity is maintained in some cases over serial transplantation<sup>46,47</sup>. HSC location also contributes to heterogeneity, with phenotypic HSCs being significantly more abundant in the BM compared to PB and spleen, although HSCs from these sites all have similar long-term repopulation capacity<sup>48</sup>.

Recent advances in scRNAseq have dramatically changed our understanding of the haematopoietic stem and progenitor cell (HSPC) landscape, including core regulatory networks<sup>49</sup>, new progenitor sub-populations<sup>50,51</sup> and the interplay of molecular markers and surface immunophenotypes of HSPCs<sup>51</sup>. This approach has also highlighted the importance of single cell approaches to studying HSPC heterogeneity and the complex interactions between HSPCs at the functional and molecular level<sup>31</sup>.

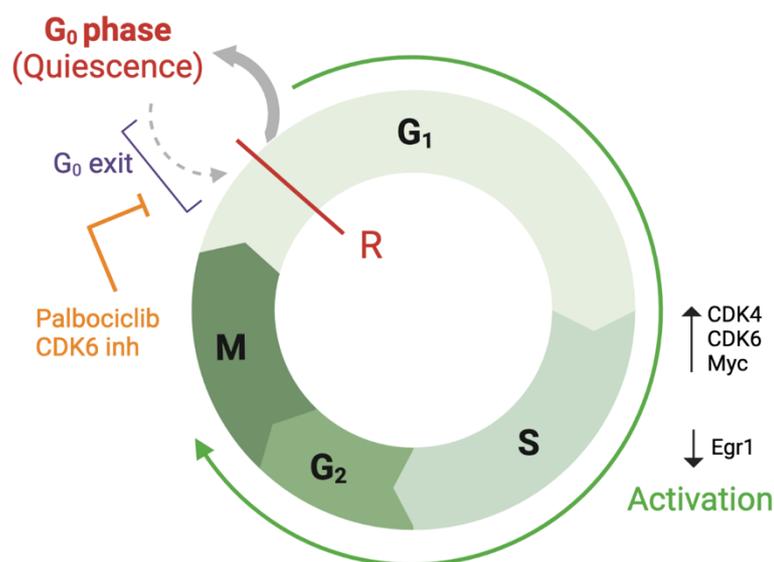
### **1.3.2 HSC quiescence**

The vast majority of HSCs reside in a quiescent<sup>52,53</sup>, glycolytic state with low mitochondrial activity compared to all other cells in the haematopoietic system<sup>25</sup>. They mostly reside outside the cell cycle in G<sub>0</sub> and divide infrequently, although they can be efficiently activated in response to stress<sup>53,54</sup>. HSCs are highly restricted in their cell cycling to maintain a regenerative cell pool and prevent haematopoietic consumption. This stem cell pool is a critical component of the haematopoietic system and maintains homeostasis through several factors, of which self-renewal capacity and quiescence are important features. Meanwhile, progenitors are highly metabolically active, proliferative, and dependent on oxidative metabolism and mitochondrial function.

Quiescence is well recognised as a highly regulated and actively maintained state<sup>32,55</sup> and a defining characteristic of steady state adult HSCs, involving signalling pathways that maintain a controlled state ready for rapid activation and differentiation<sup>56</sup>. It is closely correlated with self-renewal capacity, with more dormant HSCs being able to produce the most durable and robust grafts on transplantation<sup>53,57</sup>. It has been proposed as a

poised state with ability to activate and re-enter the cell cycle in response to external physiological stimuli<sup>56</sup>, rather than stochastically entering the cell cycle. Current speculation is that at any given time, HSCs may be quiescent, actively differentiating or in a transitional state between these on a spectrum of quiescence exit or activation.

HSC activation is defined as the process of exit from quiescence and entering the cell cycle (Fig 1.3-2). The process of HSC activation is thought to be a continuous transition between  $G_0$  and  $G_1$  rather than a binary switch. The HSC transcriptional programme has specific metabolic and cellular properties that are not necessarily linked with multipotency, since the majority of changes in gene expression between HSCs and early progenitors occur independent of lineage fate decisions<sup>35</sup>. This concept is being increasingly defined with the use of transcriptomic<sup>35,40,58</sup> and mouse genetic studies.



**Fig 1.3-2 The cell cycle.**  $G_0$  is defined as quiescence, while HSCs from  $G_0$  exit to first division are in activation. Many genes have altered expression on quiescence exit including *CDK4*, *CDK6*, *Myc* and *Egr1*. Pharmacological inhibition of *CDK6* with Palbociclib can maintain HSCs in quiescence before the restriction point R. *Made with BioRender.*

HSC quiescence is regarded to be of critical biological importance in protecting the stem cell compartment from exhaustion and accumulation of genetic mutations that can lead to bone marrow failure and/or haematological malignancies<sup>54,59</sup>. Particularly under stress conditions such as malignancy and inflammation, maintenance of the HSC pool is crucial to prevent stem cell depletion and haematopoietic failure. It is, however, under these conditions that the molecular mechanisms regulating stem cell quiescence are most likely to become perturbed, often leading to the pathological clinical phenotype seen in these conditions. Links between stress programs and self-renewal properties of HSCs are particularly important to understand during quiescence exit, when HSCs also undergo a dynamic shift in metabolic state<sup>60</sup>. Autophagy has been identified as an important mediator for HSC

stress response<sup>61</sup>. Detailed understanding of the molecular cues that regulate HSC fate is important to improve our knowledge of the regulation of haematopoiesis in health and disease<sup>62</sup>.

### **1.3.3 Regulators of HSC quiescence**

Molecular regulation of human HSCs has been previously studied. HSCs are known to have a unique transcriptional programme compared to lineage restricted progenitors, with almost 70% of genes and transcription factors expressed by HSCs being significantly changed upon exit from the HSC compartment<sup>35</sup>.

Critical regulators of HSC quiescence have been identified by several groups, many of which include enzymes such as kinases that regulate the entry of stem cells into the cell cycle. The cyclin dependent kinase CDK6 has been shown to govern quiescence exit of both LT- and ST-HSCs, as an important regulator of division kinetics. Lower levels of CDK6 correspond to higher dormancy of HSCs<sup>36</sup>. Thus, pharmacological inhibition of CDK6 can be used to maintain LT-HSCs in quiescence and prolong the quiescence exit phase of ST-HSCs, to allow close examination of altered genes during early quiescence exit of hHSCs (Fig 1.3-2). In the absence of the G<sub>1</sub> checkpoint regulator, cyclin-dependent kinase inhibitor, p21<sup>cip1/waf1</sup>, self-renewal of HSCs is impaired, leading to stem cell exhaustion and haematopoietic failure<sup>63</sup>.

Dietary vitamin A is another factor that has been shown to have significant impact on the regulation of cell cycle mediation of stem cell behaviour, with functional impairment of the HSC pool and irreversible loss of long-term self-renewal in mice depleted of vitamin A. All-trans retinoic acid (ATRA) maintains HSCs in quiescence *in vitro* through maintenance of long-term self-renewal, decreased CDK6 levels associated with lower proliferation, reduced protein synthesis and low Myc protein levels<sup>52</sup>. The tyrosine kinase receptor Tie2 and its interaction with its ligand Ang1 has also been shown to maintain HSC quiescence<sup>64</sup>.

Several master regulators of HSC exit from quiescence have been identified, including *CDK4*<sup>65</sup>, *CDK6*<sup>36</sup> and *Myc*<sup>66</sup>, which are steadily upregulated during transition towards activation, while *Egr1* is downregulated<sup>67</sup> (Fig 1.3-2). *Ifitm1* and *Ly6a* are also observed to be downregulated during quiescence exit, though causality of quiescence exit with these has not yet been established<sup>52</sup>. Activated HSCs that have not yet entered the cell cycle are genetically and metabolically primed for active differentiation. Genes related to DNA replication are differentially expressed between quiescent and active HSCs including *Cdc45* and *Mcm4*, which are upregulated in active HSCs, along with *Cdk4* and *Cdk6*<sup>52</sup>. It has also been shown that p53 deficiency in HSCs leads to a reduction in the number of quiescent HSCs and promotes cell cycle entry<sup>63</sup>. Similarly, the tumour suppressor gene *RB* also inhibits cell cycle progression as one of its major roles<sup>68</sup>, with loss of the quiescent HSC pool and expansion of early haematopoietic progenitors observed when RB family proteins are depleted<sup>69</sup>.

There exist transcriptional signatures that are specific to certain sub-populations of quiescent HSCs, which are being explored in mice<sup>70</sup>, and remain to be explored in humans. Despite extensive study of the regulators of HSC quiescence, the molecular control of this carefully balanced and actively maintained state in the haematopoietic

system remains to be completely elucidated. Some of the challenges faced in identifying these factors are the techniques used to isolate purified HSCs, which currently rely on cell-surface markers and functional biology, with FACS sorting having now been adopted as the standard technique used to isolate HSCs<sup>9,55</sup>. There are currently no techniques to functionally isolate pure HSCs.

Knowledge of the genetic regulators of HSC exit from quiescence provides further insight into the mechanisms of HSC maintenance, exhaustion and possible development of disease. They also provide a potentially novel therapeutic approach based on enhancing stem cell function as part of supportive treatment to minimise toxicity, particularly myelotoxicity, of cytotoxic chemotherapy used in the standard treatment of solid and liquid malignancies. Identification of genes that maintain stem cells in a state poised for activation may provide insight into the mechanisms by which HSCs rapidly respond to environmental changes and switch from quiescence to cell cycling<sup>56</sup>.

Furthermore, understanding the precise role of these genes that regulate quiescence exit on HSC differentiation and proliferation can provide strategies to manipulate HSC function in health and disease to allow targeted alteration of HSC properties that may lead to diseased states. For example, the ability to target leukaemic cells while maintaining HSC self-renewal capacity would significantly reduce the myelosuppressive adverse effects very commonly seen with chemotherapy. Whereas manipulating HSC differentiation capacity could expedite engraftment post haematopoietic stem cell transplant (HSCT) or myeloablative treatment in haematology/oncology patients.

Extrinsic factors that regulate HSC development and homeostasis including the cellular and molecular components of the microenvironment have also been studied. The stem cell niche has been shown to have critical importance in HSC regulation and is influenced by several factors. LT-HSCs that express the thrombopoietin receptor MPL are highly quiescent in adult BM, and are closely associated with the osteoblastic niche<sup>71</sup>. Adhesion molecules, which are immunoglobulins, regulate HSC quiescence, proliferation and commitment in adult BM by retaining HSCs in their niche micro-environment<sup>72</sup>, where they maintain quiescence supported by cells including mesenchymal cells, osteoblasts and vascular endothelial cells<sup>73-75</sup>. Certain extrinsic factors have been shown to oppose HSC quiescence and skew them towards differentiation including interferon-gamma (IFN- $\gamma$ )<sup>76</sup> and IFN- $\alpha$ <sup>77</sup>.

## 1.4 HSCs across the lifespan

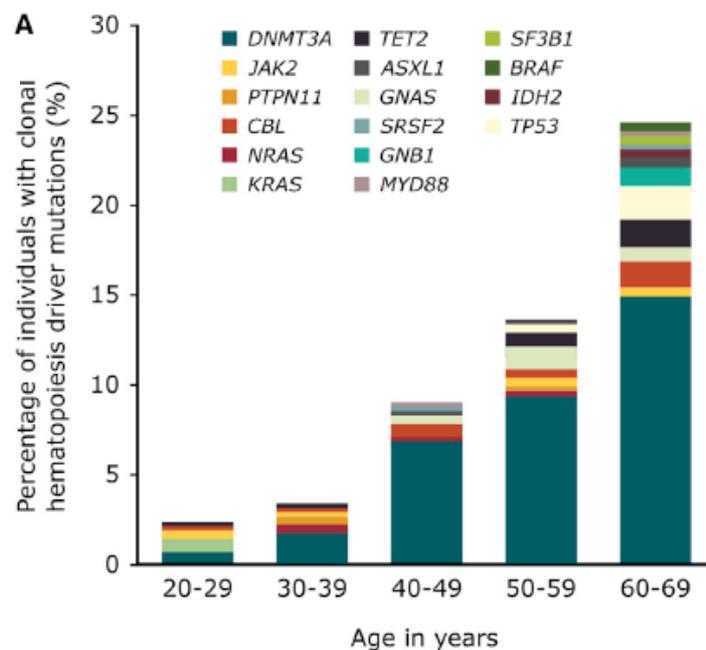
Compared to adult HSCs, foetal and neonatal HSC have significantly distinct properties including differences in cell surface markers, proliferative state, self-renewal capacities and differentiation potential. Foetal HSCs are highly proliferative and have high self-renewal properties, while adult HSCs are much more quiescent with reduced self-renewal capacity<sup>78</sup>. Foetal HSCs also exhibit an erythroid differentiation bias relative to the more balanced differentiation in adults, with a myeloid bias occurring in late adulthood. Haematopoiesis and HSCs vary across the lifespan, changing organ sites several times throughout life to meet physiological demands.

During embryonic development sites of development of HSCs are highly dynamic, localising primarily to the BM by adulthood<sup>79</sup>.

## 1.5 Clonal Haematopoiesis (CH)

Further to our understanding of variations within the HSC pool, age associated skewing in blood cells, particularly within the myeloid compartment has previously been observed<sup>80,81</sup>. HSC expansion can be driven by certain genetic clones, known as CH, contributing to the overall heterogeneity of the HSC pool. This phenomenon occurs with increasing age<sup>82</sup>, disease<sup>83,84</sup> and following certain insults such as cytotoxic therapy<sup>85,86</sup>. The effect of clonality on HSC function including proliferation and differentiation capacity is not well understood.

HSCs carrying clonally expanded mutations can contribute disproportionately to blood production without inducing leukaemia and have been shown to exist in healthy individuals<sup>87-89</sup>. This entity has now been widely recognised as Clonal Haematopoiesis of Indeterminate Potential (CHIP), or Age related Clonal haematopoiesis (ARCH) since the incidence correlates strongly with age<sup>90</sup> (Fig 1.5-1). ARCH is defined by somatic mutations, which were conventionally assumed to be acquired after birth. Using recent modelling techniques, we now know that these somatic mutations can indeed be acquired in utero and remain dormant for decades<sup>84,91</sup>. For example, *DNMT3A* mutant HSCs can be maintained *in vivo* for decades before the onset of clonal expansion and/or clinically apparent disease<sup>92</sup>.



**Fig 1.5-1 Frequency of the mutations seen in age related clonal haematopoiesis (ARCH).** The prevalence of known driver mutations driving haematopoiesis rises steeply with age, especially over

60 years. *DNMT3A* is the most common ARCH mutation, accounting for >15% of ARCH mutations (VAF >2%) in individuals >60 years. *Acuna-hidalgo R et al., Am J Hum Gen 2017; 101:1-15.*

### **1.5.1 Definitions – ARCH, CHIP and pre-leukaemia**

Early in life, all HSCs contribute equally to blood production. With age, certain clones almost inevitably predominate, driving CH<sup>82</sup>. ARCH or CHIP is defined as the presence of one or more expanded somatic haematopoietic clones with a variant allele frequency (VAF)  $\geq 2\%$ , in the absence of haematological disease<sup>80,90</sup>. Clonal expansion is commonly seen in recurrent genes, which are also seen in leukaemia, particularly myeloid malignancies<sup>88</sup>, hence these are frequently termed pre-leukemic mutations (pLM)<sup>93</sup>. In this thesis, pLM refer specifically to mutations that occur in HSCs in patients who have developed leukaemia, hence it is a retrospective annotation. Mutations related to CH will be referred to as ARCH mutations from here on.

### **1.5.2 Epidemiology of clonal haematopoiesis**

CH was first recognised as an entity over 10 years ago when healthy individuals were noted to have copy number changes in chromosomal loci associated with haematological malignancy (20q, 5q, 11q, 17p) on genome wide association studies (GWAS)<sup>94</sup>. The exact prevalence of CH is not known, though large studies have indicated that 5-10% of healthy individuals over 65 years of age carry ARCH mutations<sup>87,89</sup>. The vast majority of these mutations are accounted for by the 3 recurrent genes: *DNMT3A*, *TET2* and *ASXL1*<sup>93</sup>.

We now recognise that oligoclonality is a feature of age and the prevalence rises exponentially in the elderly population<sup>82</sup>. ARCH is strongly associated with increased risk of haematological malignancy (hazard ratio, HR 11-12), and all-cause mortality (HR 1.4) compared to age matched individuals without ARCH mutations<sup>87,88,90</sup>. Recently ARCH has been associated with other non-malignant complications including atherosclerosis, cardiovascular disease<sup>88,95,96</sup>, stroke and metabolic syndrome, clinical entities that are linked by inflammation<sup>97</sup>. More recently ARCH has also been linked to chronic obstructive pulmonary disease (COPD)<sup>98</sup> and osteoporosis<sup>99</sup>. In the elderly population, cardiometabolic events pose a far greater public health risk than haematological malignancy due to their higher incidence<sup>100</sup>.

It is believed that the widespread availability of high-throughput DNA sequencing technology in recent times has allowed the recognition of CH and until very recently the nature of clonal progression was not known<sup>87</sup>. Sequencing of blood derived DNA by targeted and whole genome sequencing (WGS) is being increasingly used to screen for inherited risk factors for common diseases<sup>87</sup>, especially malignancies. As we understand the role of CH related genes in pathophysiology, there may be a role for screening for these mutations in high-risk populations such as the elderly. However, the validity of screening may be limited by availability of subsequent monitoring and treatment strategies. Further clinical trials are required to reduce the risk of progression to cancer and other CH associated conditions in a clinically applicable way.

*DNMT3A* is the most common mutation observed in CH, accounting for 15-20% of ARCH cases<sup>88,93</sup>. It is a 130kDa protein encoded by 23 exons on human chromosome 2p23<sup>101</sup>. It is one of a group of DNA methyltransferase enzymes (*DNMT1*, *DNMT3A* and *DNMT3B*), that govern DNA methylation patterns in mammals. DNA methylation primarily occurs at the C-5 position of cytosine in 70-80% of CpG sites throughout the human genome<sup>102</sup>, although *DNMT3A* and *DNMT3B* introduce non-CpG methylation (mainly CpA) in certain cells including oocytes, embryonic stem cells<sup>103</sup> and neural cells<sup>102,104</sup>. To mediate *DNMT3A* binding to DNA, a continuous DNA-binding surface is created by three *DNMT3A* domains: a loop from the target recognition domain (TRD, residues R831-F848), the catalytic loop (residues G707-K721) and the *DNMT3A* homodimeric interface<sup>105</sup>.

### **1.5.3 Non-haematological associations of CH**

ARCH has been clearly linked to cardiovascular disease, independent of other cardiovascular risk factors including age, sex, type 2 diabetes status, total cholesterol, high density lipoprotein (HDL) cholesterol, hypertension and smoking status<sup>95</sup>. Risk is highest in individuals with *DNMT3A*, *TET2*, *ASXL1* and *JAK2* mutations, with the latter 3 mutations being further associated with early-onset myocardial infarctions<sup>95</sup>. In asymptomatic individuals without coronary heart disease, risk of coronary-artery calcification increases significantly with VAF  $\geq 10\%$  for an ARCH mutation clone compared with those with a smaller clone<sup>95</sup>.

Inflammation has been established as the main systemic link between the various manifestations of cardiovascular disease including coronary artery disease, atherosclerosis and ischaemic stroke<sup>106</sup>. Besides protein and small molecule mediated inflammation, leukocytes have been recognised as playing a role in this inflammatory pathogenesis and promotion of atherosclerosis, in particular macrophages<sup>107</sup>, which appear along the monocyte differentiation axis<sup>108</sup>. Proinflammatory monocytes expressing *Ly6c* have been noted to play a significant role in hypercholesterolemia in mice<sup>106</sup>.

There is compelling evidence from murine models linking cardiovascular disease in the *Tet2*-mutated CH context to inflammasome-mediated endothelial injury (NLRP3, IL-1b secreted) in *Tet2*-null low density lipoprotein (LDL) receptor knockout (KO) mice<sup>109</sup>. Macrophages derived from circulating clonal monocytes undergo pro-inflammatory interactions with endothelium to generate injury and inflammation in coronary artery endothelium and other tissues including spleen, liver, lung and kidneys<sup>95</sup>. Furthermore, inflammation initiated by macrophages in the atherosclerosis and cardiovascular disease context has been well established in terms in causality<sup>110,111</sup>. While causality has been established between *TET2* mutations and cardiovascular disease<sup>95,109</sup>, *DNMT3A* has not yet been causally linked despite the broader epidemiological associations, and it being the most commonly mutated gene in ARCH.

There is also evidence that elevated red cell distribution width (RDW) is associated with higher mortality in the context of an expanded CH clone (*DNMT3A/TET2/ASXL1*), but not on its own<sup>88</sup>. This may be a marker of perturbed erythropoiesis.

Clinical management of ARCH currently consists of reduction of modifiable cardiovascular risk factors and monitoring for haematological changes. However, in the absence of available interventions and the low rate of malignant transformation, routine monitoring and screening is not currently recommended<sup>112</sup>. Eventually, armed with further understanding of the mechanism of inflammatory and functional processes underlying these complications, management could potentially include anti-inflammatories and targeted approaches.

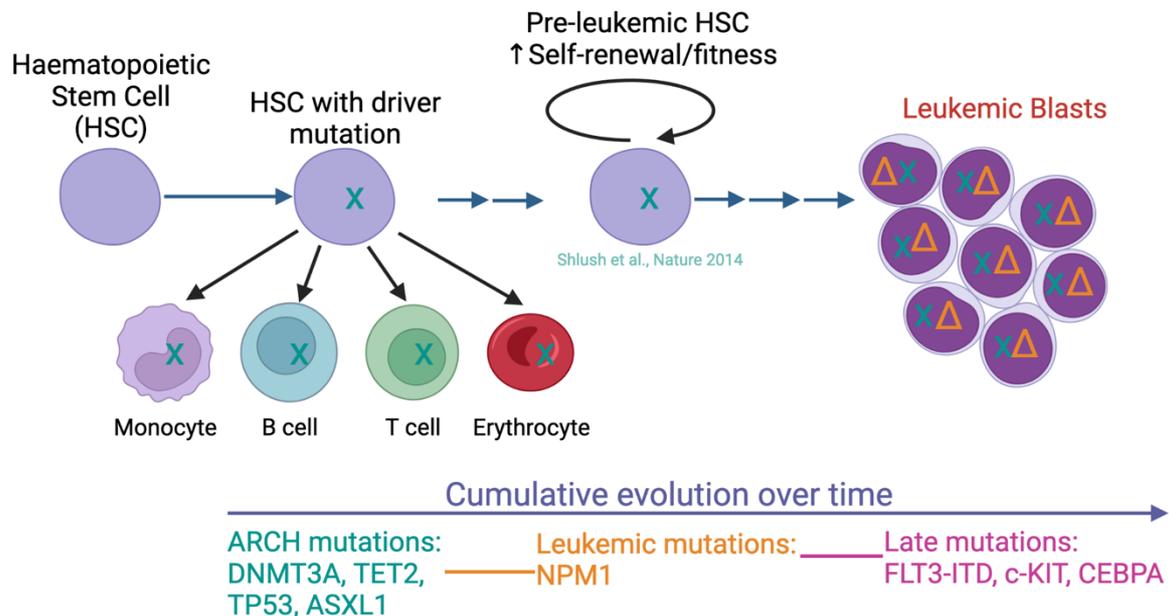
## 1.6 Leukemic stem cells

HSPCs, pre leukemic-HSCs (pL-HSPCs) and leukemic stem cells (LSCs) are considered to be distinct populations, harbouring no identifiable clonal mutations, preL-mutations (pLM) and leukemic mutations respectively. The former two populations are able to differentiate into all lineages, as pLM are observed in all mature blood cells<sup>113,114</sup>, while LSCs are restricted in their differentiation capacity and only produce leukemic blasts.

Leukemia is an extreme example of CH, in which one clonally expanded stem cell (LSC) drives all the (abnormal) haematopoiesis, leading to blast production, and suppresses normal haematopoiesis<sup>112</sup>. The cancer stem cell (CSC) model is now widely recognised as a single cell of origin, clonally expanding to give rise to all progeny, i.e. malignant cells<sup>115–117</sup>. It is best characterised in acute myeloid leukaemia (AML), with the discovery of the ability of LSCs to generate leukaemia upon transplantation in mouse models<sup>118</sup>. This model postulates a hierarchical organisation and differentiation of LSCs with self-renewal capacity at the apex alongside the ability to perpetuate malignant cell differentiation<sup>119–124</sup>. The hierarchy within AML LSCs has been demonstrated by a number of groups with LSCs enriched in the CD34+CD38- fraction giving rise to CD34+CD38+ leukaemia progenitors, which further differentiate into leukemic blasts<sup>13,125</sup>. LSCs are, similarly to HSCs, thought to reside predominantly in the quiescent state, and thus inherently resistant to most forms of chemotherapy, with response being influenced by somatic mutations<sup>126</sup>. These are thus the very cells hypothesised to be responsible for relapsed disease following anti-neoplastic therapy<sup>125,127–129</sup>. Since relapsed leukaemia carries the greatest morbidity and mortality, it is prudent to develop an intimate understanding of the role of these cells and their ability to survive standard chemotherapy.

LSCs are defined functionally by their ability to re-establish leukaemia in immunodeficient mice rather than by expression of cell surface markers since they cannot be phenotypically distinguished from HSCs<sup>130–132</sup>. The relationship between hHSCs and AML LSCs is now being investigated at the transcriptional, proteomic and epigenetic levels with the aim of identifying a common molecular state of AML LSCs independent from cell surface markers. There is emerging evidence that HSC and LSC signatures have prognostic implications beyond the current standard risk stratification factors: age, cytogenetics and molecular mutations<sup>133–135</sup>. Patient survival can be improved through the ability to identify patients who currently classify as ‘low-risk’ at diagnosis who would benefit from more aggressive, but toxic therapy such as HSCT or Chimeric Antigen Receptor T-cell (CAR-T) therapy, based on stem cell signatures. The role of leukemic stem and progenitor cells (LSPCs) is also widely recognised in the pathogenesis of haematological malignancies, especially myeloid, with identification of the

cell of origin, which acquires the first somatic mutation, leading to the multistep progression and acquisition of subsequent mutations towards frank malignancy (Fig 1.6-1)<sup>25</sup>.



**Fig 1.6-1 Stepwise acquisition of somatic mutations leading to the development of leukaemia.**

Schematic diagram showing that healthy HSCs can produce all blood lineages, even when a pre-leukemic or ARCH mutation (marked with X) is acquired at the HSC level, such as *DNMT3A*, *TET2*, *TP53* or *ASXL1*. These mutations confer self-renewal, followed by the acquisition of further, more proliferative mutations (marked with Δ) such as *NPM1*, *FLT3-ITD*, *c-KIT* and *CEBPA*, which lead to leukaemia. *Made with BioRender.*

Further evaluation of HSC function and behaviour in the context of leukaemia would provide useful insights into the relationship between hHSCs and LSCs. There are several limitations in the current methodologies applied to studying hHSC function, since they are a rare population and cannot replicate the length and variability of the human lifespan during functional studies<sup>54</sup>. However, recent developments in novel humanised mouse models<sup>136,137</sup>, single cell technology and genome editing techniques such as CRISPR should allow a better understanding of the regulation of hHSC function, particularly in the context of leukaemia.

## 1.7 Clinical significance

Genetic and functional changes at the HSPC level precede malignancy as observed in pre-malignant clinical states such as myelodysplastic syndrome (MDS) and development of haematological neoplasms. Development of haematologic malignancy has now been recognised as a stepwise acquisition of mutations rather than a sudden acute event (Fig 1.6-1). The timing of acquisition of somatic mutations associated with haematological malignancy is becoming increasingly important, leading to our understanding of the underlying heterogeneity in the malignant phenotypes between patients carrying the same mutation, since there can be great inter-patient variation in the cellular states and times at which these mutations were acquired<sup>25</sup>.

There are inherent differences between ARCH and other precursor pre-malignant phases. Other pre-malignant phases rarely lead to morbidity/mortality on their own, with malignant transformation being a prerequisite, e.g. Barrett's oesophagus and colonic polyps. ARCH however, can lead to significant morbidity and mortality via cardiovascular disease in the absence of malignancy. It is unclear whether ARCH/CHIP is on a trajectory to malignancy (pre-malignant condition) or a stand-alone clinical state, or indeed both.

ARCH is significantly more likely to be diagnosed in individuals being investigated for cytopenias of unknown cause that do not meet diagnostic criteria for a malignant or pre-malignant condition such as MDS than in healthy individuals with normal blood counts. While the risk of developing malignancy is 12-13 fold higher in ARCH than the normal population without detectable somatic mutations, the absolute risk is low with annual rate of 0.5-1%<sup>87,88,90</sup>. This may be partially explained because individuals with ARCH tend to be older and may die of other causes before developing haematological malignancy given the long latency between CH and malignant transformation<sup>138</sup>; we also know that additional cooperating gene mutations are almost universally required for ARCH progression to leukaemia.

Typically patients with MDS/AML harbour more than 1 recurrent somatic mutation in the leukemic blasts, while individuals with ARCH generally only have 1, and those with >1 ARCH mutation have a greater risk of progression to disease<sup>88</sup>. The size of the clone measured by VAF also correlates with risk of malignancy, with a VAF  $\geq 10\%$  being associated with a significantly greater risk of haematologic malignancy (HR 49)<sup>88</sup>. Malignancy is predominantly myeloid (60%) and includes AML, MDS, chronic monomyelocytic leukaemia (CMML), myeloproliferative neoplasms (MPN) and 40% lymphoid, including chronic lymphocytic leukaemia (CLL), multiple myeloma (MM) and B-cell lymphoma. This is suggestive of the fact that these somatic mutations occur at the stem cell level before lineage commitment occurs, i.e. at the HSC level<sup>87,100,139-143</sup>. Overall mortality even in the absence of malignancy is increased in individuals with ARCH (HR 1.4,  $p = 0.02^{87}$ ), which is mainly linked to cardiovascular events<sup>88</sup> (HR 2.6,  $p = 0.003$ ).

While CH usually precedes AML, most CH clones do not spontaneously transform to AML and the factors contributing to leukemic progression, besides acquisition of secondary mutations, are not well understood. Although clonality is a hallmark of cancer, it is not sufficient to be causal in isolation. Altered epigenetic programs induced by CH associated mutations may facilitate clonal instability and predispose carriers to acquire further leukemogenic mutations including *FLT3* and *NPM1*.

Some ARCH mutations can help cells evade conventional chemotherapy, forming the basis of molecular relapse<sup>144</sup>, which is the leading cause of morbidity and mortality in haematological malignancies, whereby most patients can achieve first remission but succumb to relapsed disease. Proof of this lies in the persistence of such mutations (especially *DNMT3A*, but also *TP53* and *PTPN1*) in patients with AML even after achieving complete remission<sup>112,126</sup>

Since relapsed disease is notoriously difficult to treat in almost all types of malignancies, eradicating LSCs or indeed CSCs as part of first line treatment following initial diagnosis may provide an important therapeutic

strategy to prevent disease relapse and achieve cure. However, this approach is no doubt fraught with difficulties including our current inability to specifically target LSCs distinct from HSCs, since no unique cell surface marker for LSCs has been identified to date. Also, given the inherent self-renewal capacity of these cells, by definition, complete eradication is challenging<sup>54</sup>. One approach may be the indefinite maintenance of these cells in the quiescent state, leading to effective aplasia of mature leukemic cells. Any such therapy targeting the maintenance of LSCs in quiescence needs to be selective, effective and tolerable on a long-term basis. Targeted treatment of *BCR-ABL* mutant LSCs in CML has been one successful application of this approach<sup>145</sup> and form part of standard treatment for all patients with *BCR-ABL* and *ABL* class mutations<sup>146</sup>. Maintaining LSCs in quiescence may form an important therapeutic approach as a bridge to more definitive treatment such as HSCT.

## 1.8 Overall Aims

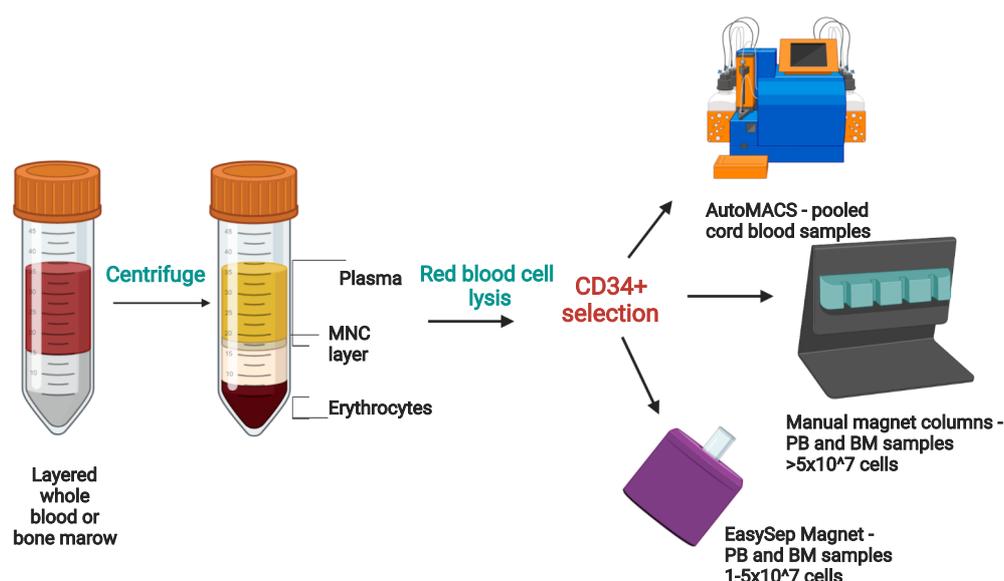
Developing on the current understanding of human HSC/MPPs, the overall aim of my thesis was to further delineate differentiation dynamics of HSPCs in health and disease. To explore these concepts, I will address the following overall aims in my thesis:

- a) **HSCs in health:** understanding the role of genes altered during quiescence exit in normal healthy hHSCs. Here the focus was on specific genes known to have differential expression between quiescence and early activation, and to explore their role in HSC function.
- b) **Clonal haematopoiesis:** characterising the role of pre-leukemic mutations on the dynamics of pre-leukemic HSC differentiation, specifically in the context of *DNMT3A R882* mutations.

# Chapter 2 Materials and Methods

## 2.1 Sample collection and preparation

All whole blood or bone marrow (BM) samples were processed to isolate CD34+ cells for future use as shown in Fig 2.1-1. The human biological samples were sourced ethically and their research use was in accord with the terms of the informed consents under an IRB/EC approved protocol.



**Fig 2.1-1: CD34+ selection process from hCB, PB and BM samples.** Whole blood or bone marrow aspirate sample was layered on pancoll using density gradient and centrifuged to generate layers as indicated. Following red blood cell lysis, CD34+ selection was performed using one of the 3 methods indicated: AutoMACS, manual magnetic column selection or EasySep Magnet. *Made with BioRender.*

### 2.1.1 Human umbilical cord blood (hCB) samples

hCB samples were obtained from healthy neonatal donors at the Rosie Hospital by Cambridge Blood and Stem Cell Biobank (CBSB) trained nursing staff, following informed consent in accordance with regulated procedures approved by the relevant Research and Ethics Committee (07/MRE05/44 research study and IRAS Ref: 149581). hCB samples acquired on the same day were pooled and processed as an individual sample.

Cord blood (CB) was diluted 1:1 with Phosphate Buffered Solution (PBS) and layered using Lymphoprep or Pancoll, and density gradient centrifugation was used to obtain a mononuclear cell (MNC) layer, which was manually isolated using pipettes. MNCs were depleted of red blood cells (RBC) by incubating with RBC Lysis Buffer for 15 minutes at 4°C, then washed with PBS + 3% foetal bovine serum (FBS). Positive selection of CD34+ cells was then achieved by incubating with CD34 Micro Beads (30µL/10<sup>8</sup> cells), FcR Blocking Reagent (30µL/10<sup>8</sup> cells) and PBS + 3% FBS (90µL/10<sup>8</sup> cells) for 30 minutes at 4°C. Cells were then washed and resuspended in 2ml AutoMACS running buffer prior to selection on the AutoMACS machine using 'Posseld2' program for >1x10<sup>8</sup>

MNCs, or manual magnetic selection columns (Miltenyi) for  $5 \times 10^7 - 1 \times 10^8$  MNCs (Fig 2.1-1). Isolated hCB CD34+ cells were then stored at  $-150^\circ\text{C}$  until use.

## **2.1.2 Peripheral blood (PB) samples**

### **2.1.2.1 Patient PB samples**

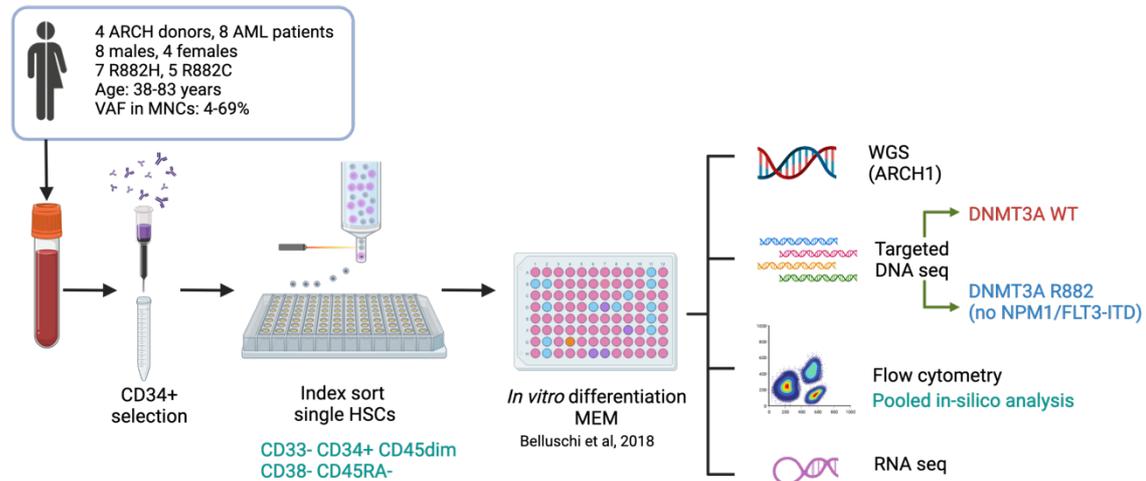
Patient PB samples were obtained from the Weizmann Institute for Medical Science under a materials transfer agreement (MTA). Primary samples were received from Princess Margaret Cancer Centre, University Health Network (UHN), Canada (UHN IRB protocol 01-0573) and from Rambam Health Care Campus, Israel (IRB protocol #283-1). 1 BM sample was obtained from Newcastle Hospitals NHS Foundation Trust under an MTA. Frozen MNCs were originally sourced from Granulocyte Colony Stimulating Factor (GCSF)-mobilised (mPB) and non-mobilised PB and BM. All samples were collected by phlebotomy or apheresis for mPB samples, Ficoll separated and MNCs viably frozen without RBC lysis. Mobilised PB MNCs were also further enriched for CD34+ cells using the Miltenyi microbeads kit.

### **2.1.2.2 Patient selection**

Patients were selected based on targeted sequencing performed on PB MNCs using Miseq for known pre-leukemic mutations (pLM) including *DNMT3A*, *ASXL1* and *TET2*. Samples with *DNMT3A R882 VAF* >2% in PB were selected for the study. One healthy control sample (HC1) was selected based on sequencing data confirming <2% VAF in the *DNMT3A R882* gene, while the remaining HC were obtained from clinically healthy age matched controls without genetic confirmation.

Samples from 4 females and 8 males, 4 ARCH and 8 AML patients (Table 4.3-1) were assessed using our experimental platform (Fig 2.1-2). Of the ARCH individuals, 1 sample (ARCH1) was taken from a mPB collection in preparation for allogeneic transplant donation for a matched unrelated recipient with relapsed AML, which was aborted due to premature patient death prior to HSCT. The ARCH2 individual was identified as a carrier of *DNMT3A R882H* clone on screening as part of the Wizeageing study at the Weizmann Institute of Science, Israel, which set out to screen healthy individuals for CH related mutations in non-mobilised PB. A follow up sample was collected for our study once the clone was identified during screening. ARCH3 was identified as a carrier of *DNMT3A R882H* when a mPB collection was screened prior to planned autologous stem cell infusion following chemotherapy treatment for Non-Hodgkin Lymphoma. ARCH4 was a patient receiving hip replacement, and the BM from the removed femoral head was used to detect a significant (42%) *DNMT3A R882H* clone, and the BM sample donated for research with informed consent. All AML PB samples were from patients with *de novo* malignancy, collected at diagnosis alongside BM samples. These were all non-mobilised PB samples. Phenotyping information on circulating peripheral blasts was available for all patients. DNA from PB samples

from all patients was sequenced and specific CH related mutations identified with the allele frequency in bulk MNCs.



**Fig 2.1-2: Experimental schema used to correlate phenotype and genotype information from single HSC derived colonies from individuals with *DNMT3A* R882 clone in PB.** Bone Marrow (BM) or peripheral blood (PB) from individuals with a known *DNMT3A* R882 clone was used to select CD34+ cells and isolate HSCs/MPPs for *in vitro* differentiation into erythroid/myeloid/NK lineages. Colonies were genotyped using WGS, targeted DNA sequencing, RNAseq and phenotyped using high throughput flow cytometry. *Made with BioRender.*

## 2.1.3 Healthy donor PB samples

### 2.1.3.1 Venesection

PB samples were collected from healthy volunteers based on the Addenbrookes research campus through venesection performed by trained phlebotomists in accordance with CBSB Biobank ethics (IRAS Ref: 149581). Donors were not screened for genetic mutations prior to sample collection and did not receive GCSF stimulation. 20-25ml whole blood samples were collected in LiHep containing tubes. Blood was diluted and layered for density gradient centrifugation using the same method as hCB samples. MNCs obtained were depleted of RBCs as above and positively selected for CD34+ cells using the MicroBeads CD34+ selection kit prior to being manually selected using MACS cell separation columns (Fig 2.1-1). CD34+ cells were either stained and sorted as single cells for culture on the same day or frozen at -150°C for future use.

### 2.1.3.2 Leukocyte cones

Leukocyte cones residual from platelet donation were obtained from National Health Service Blood and Transfusion Unit (NHSBT), which are a by-product of the apheresis process that would otherwise be discarded<sup>147</sup>. Blood from the cones were first treated with RBC lysis buffer at 4°C for 5min, and positively selected for CD34+

cells using the MicroBeads CD34+ selection kit prior to being manually selected using MACS cell separation columns and frozen at -150°C for future use.

### 2.1.3.3 Bone marrow sample

A BM sample was obtained from a phenotypically healthy individual following hip replacement in an effort to establish a biobank of CHIP individuals in Newcastle, UK. The sample was layered and MNCs isolated before being frozen and transferred to Cambridge under the agreed MTA.

## 2.2 In vivo engraftment

Primary CD3 depleted PB samples from patients with AML were injected into SGM3, NSG or hSCF mice to assess engraftment capacity. Xenotransplantation assays were performed at the Weizmann Institute by Amos Tuval, in accordance with the institutional guidelines approved by the Weizmann Institute of Science Animal Care Committee (11790319-2). Engrafting human cells were then assessed by flow cytometry; engraftment is defined as myeloid if >90% cells are myeloid (CD45dimCD33+), and multi-lineage if  $\geq$ 10% cells are non-myeloid/lymphoid (CD19+). As per Shlush et al<sup>126</sup>, leukemic engraftment is characteristically myeloid predominant, while non-leukemic engraftment is seen as multilineage myeloid/lymphoid.

Engrafted human cells were further assessed by targeted Amplicon DNA sequencing to determine the functional phenotype of the stem cell of origin that outcompeted other stem cells in that *in vivo* assay and defined as in Shlush et al, 2014<sup>126</sup>:

- **Leukemic:** *DNMT3A* and *NPM1* mutant. Engraftment is always myeloid
- **Pre-leukemic:** *DNMT3A* mutant, *NPM1* WT. Engraftment can be myeloid or myelo-lymphoid
- **Neither** of the above: *DNMT3A* and *NPM1* WT. Engraftment can be myeloid or myelo-lymphoid

Even in the absence of leukemic engraftment, the final engraftment is not defined as “healthy” as the HSC of origin was from a patient with AML.

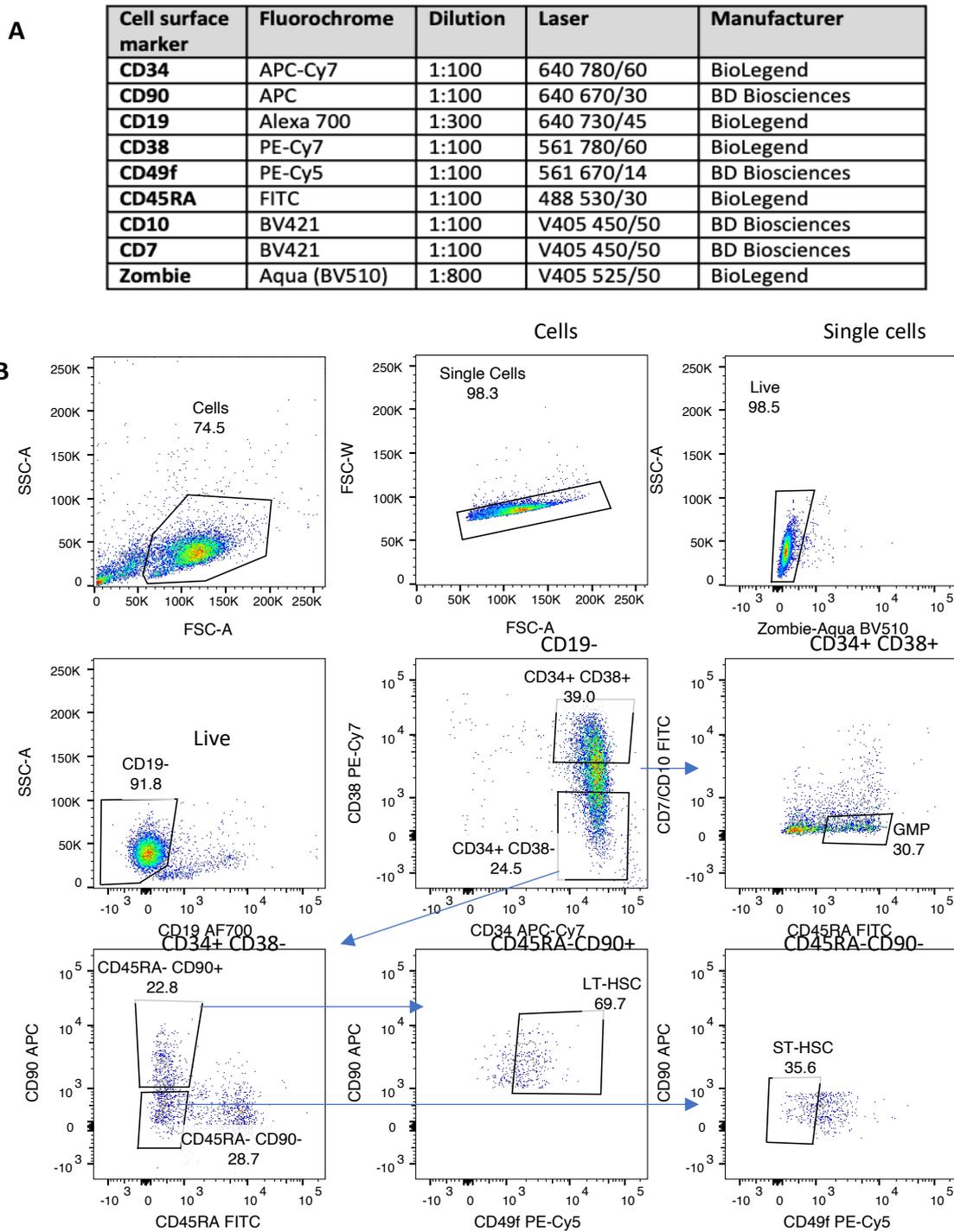
Engraftment capacity is defined as the final proportion of human cells identified amongst the viable cells from the mouse BM.

## 2.3 Cell preparation and sorting strategy

### 2.3.1 HSPC isolation from hCB CD34+ cells

For all experiments, frozen CD34+ cells from the above procedure were initially thawed using pre-warmed 50% IMDM + 50% FBS + 1:100 DNase, added dropwise. The solution was centrifuged and resuspended in PBS + 3% FBS and stained with the pre-designed flow cytometry antibody panel as outlined in Fig 2.3-1A. Cells were

incubated at room temperature (RT) in the dark for 20 minutes prior to being washed with 2 volumes of PBS + 3% FBS. This stained sample was then sorted into LT-HSCs (500 cells per condition), ST-HSCs (350 cells per condition) and GMPs (granulocyte-macrophage progenitors, 350 cells per condition) using the BD Aria-Fusion sorter available at the National Institute of Health and Care Research (NIHR) Cambridge Biomedical Research Centre (BRC) Cell Phenotyping Hub facility. These cells were sorted into PBS + 3% FBS, centrifuged and resuspended prior to being added to the appropriate culture medium for the experiment. A representative example of the gating strategy used to isolate HSCs and progenitors is shown Fig 2.3-1B.



**Fig 2.3-1 Isolation of LT-HSCs, ST-HSCs and GMPs from hCB CD34+ cells on FACS based cell sorter.** (A) Staining antibody panel used for CD34+ hCB cells. (B) LT-HSCs = live/CD19-/CD34+/CD38-/CD45RA-/CD90+/CD49f+; ST-HSCs = CD90-/CD49f-; GMPs = CD34+CD38+/CD45RA+/CD7-/CD10-. Cells were FACS sorted into the appropriate culture medium for 3 weeks to allow differentiation all major blood lineages and treated with pharmacological inhibitors as necessary.

## ***2.3.2 HSPC isolation from peripheral blood and BM samples***

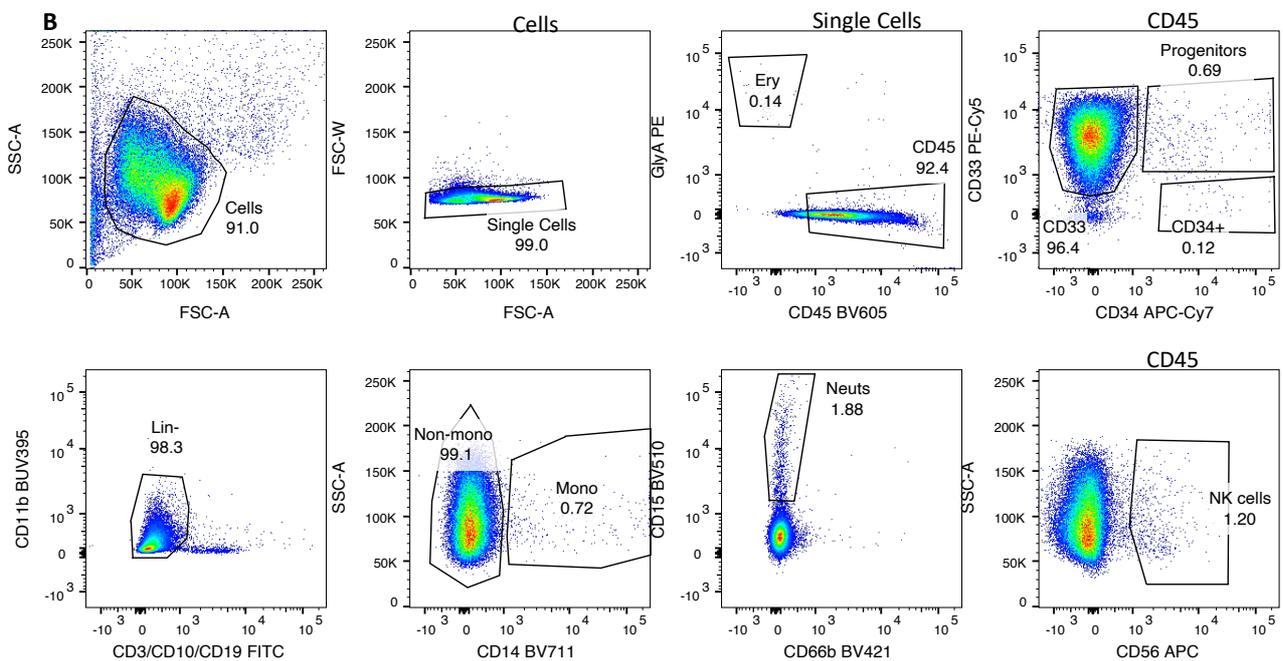
### ***2.3.2.1 CD34+ selection***

All samples were received as frozen MNCs isolated from non-mobilised PB or BM, except for 2 samples (ARCH1 and ARCH3), which were mPB from patients treated with GCSF stimulation prior to PB collection. MNCs from each individual were thawed using 50% IMDM/50% FCS with 1:100 DNase. After suspension in PBS + 3% FCS, a small sample was stained with markers of mature PB cells for pre-selection analysis to assess mature cell composition of these samples (Fig 2.3-2A) and analysed on a Fortessa FACS analyser as per the gating strategy in Fig 2.3-2B.

Positive selection of CD34+ cells from the remaining sample was achieved by incubating with CD34 Micro Beads (30 $\mu$ L/10<sup>8</sup> cells), FcR Blocking Reagent (30 $\mu$ L/10<sup>8</sup> cells) and PBS + 3% FBS (90 $\mu$ L/10<sup>8</sup> cells) for 30 minutes at 4°C. Cells were then washed and resuspended in MACS buffer and applied to a prepared LS magnetic column for manual selection as per the Miltenyi user manual (Fig 2.1-1). Samples with < 1 x 10<sup>7</sup> MNCs after thawing were stained and sorted into single cells without CD34+ magnetic selection, while samples with 1-5x10<sup>7</sup> MNCs were selected using EasySep. EasySep was performed by incubating in the selection cocktail for 15min and with magnetic particles for 10min at room temperature. Once in the magnet, supernatant containing CD34- cells was poured out and frozen separately, while CD34+ cells were separated for staining and selection (Fig 2.1-1). CD34- cells from all samples were frozen in aliquots for future studies.

A

Cell surface marker	Fluorochrome	Dilution	Laser	Manufacturer
CD56	APC	1:200	640 670/30	BioLegend
CD34	APC-Cy7	1:100	640 780/60	BD Biosciences
GlyA	PE	1:1000	561 582/15	BioLegend
CD33	PE-Cy5	1:1000	561 670/14	BioLegend
CD38	PE-Cy7	1:100	561 780/60	BD Biosciences
CD3; CD10; CD19	FITC	1:100; 1:200; 1:500	488 530/30	BioLegend
CD66b	BV421	1:100	V405 450/50	BD Biosciences
CD41a	BV510	1:200	V405 525/50	BD Biosciences
CD15	BV605	1:100	V405 610/20	BioLegend
CD14	BV711	1:1000	V405 780/60	BioLegend
CD11b	BUV395	1:1000	355 379/20	BD Biosciences



**Fig 2.3-2 Pre-selection analysis of MNCs from *DNMT3A R882* samples.** (A) Cell surface markers and (B) gating strategy used for pre-CD34+ selection analysis of MNCs from *DNMT3A R882* samples. Cells from human PB/BM samples were analysed using this strategy to understand the overall phenotype of MNCs prior to CD34+ selection.

### 2.3.2.2 Sorting pre-leukemic stem cells

CD34+ cells were stained using the staining panel and sorting strategy shown in Fig 2.3-3, as single HSC/MPPs into 96 well plates containing 100µL/well MEM media (Myeloerythroid medium, Table 2.3-1): Cells/singlets/live cells/CD19- or Lin- (CD3/CD19/CD20/CD14/CD56/CD11c)/CD33-/CD34+/CD45dim/CD38-/CD45RA-. For some samples GMPs were also sorted using the above strategy until CD45dim, followed by CD38+/CD10-/CD7-/CD45RA+ (Fig 2.3-3).

To isolate pre-leukemic HSC/MPPs (pL-HSCs) without including leukemic cells of any maturity from leukemic stem cells (LSCs) to leukaemic blasts in our analysis, we used the following 3 strategies:

1. Using a published sorting strategy to isolate pL-HSCs using FACS on CD34+ enriched PB<sup>126</sup>. After excluding CD33+ and CD45hi cells, which are likely to be leukemic, we isolated CD45dim/CD34+/CD38-/CD45RA- cells (Fig 2.3-3)
2. Culture medium and conditions that do not support growth of LSCs in vitro (unpublished), and are optimised for HSCs<sup>50</sup> (Table 2.3-1).
3. Targeted DNA sequencing of all HSC/MPP derived colonies to identify only those that were derived from stem cells and progenitors without the leukemic mutations *NPM1* and *FLT3-ITD*. This was done using the Miseq panel as described in Section 2.6.1.2.

Component	Volume (ml)	Ratio	Manufacturer
StemPro Media	50		Life Technology
StemPro Nutrients	1.4		Life Technology
L-Glutamine	0.5	1:100	Life Technology
Penicillin/ Streptomycin	0.5	1:100	Life Technology

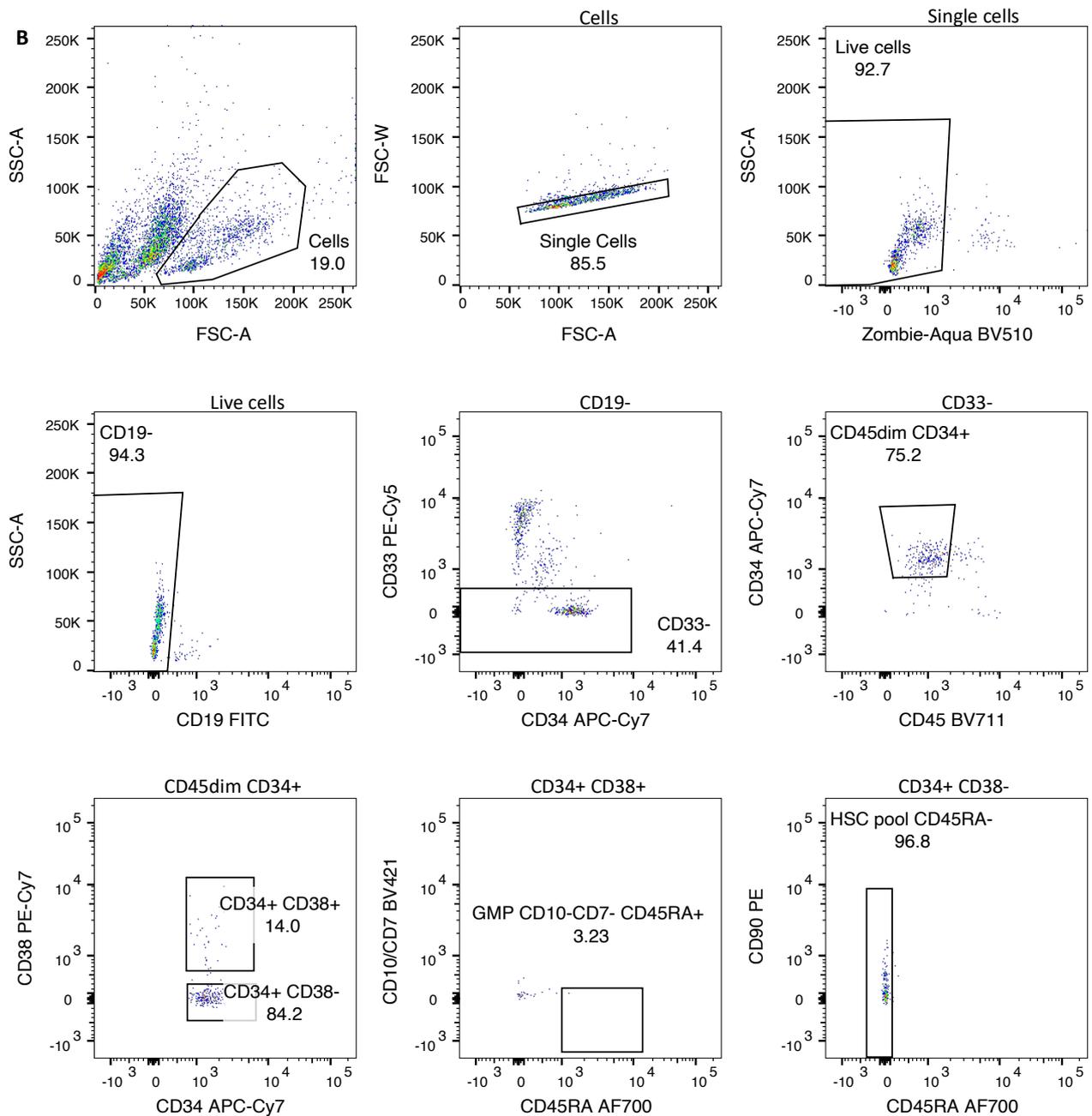
  

Cytokine	Concentration (ng/ml)	Ratio	Manufacturer
SCF	100		Miltenyi
FLT-3	20		Miltenyi
TPO	100		Miltenyi
EPO	3 Units/ml		Eprex, Janssen-Cilag
IL-6	50		Miltenyi
IL-3	10		Miltenyi
IL-2	10		Miltenyi
IL-7	20		Miltenyi
GM-CSF	20		Miltenyi
hLDL	50		Stem Cell Technologies

**Table 2.3-1 Components of cytokine enriched myeloerythroid medium (MEM) medium to support HSPC differentiation into myeloid and erythroid lineages.**

**A**

Cell surface marker	Fluorochrome	Dilution	Laser	Manufacturer
CD33	APC	1:200	640 670/30	BD BioSciences
CD45RA	Alexa 700	1:300	640 730/45	BioLegend
CD34	APC-Cy7	1:100	640 780/60	BioLegend
CD90	PE	1:50	561 582/15	BioLegend
CD49f	PE-Cy5	1:100	561 670/30	BD BioSciences
CD38	PE-Cy7	1:100	561 780/60	BD BioSciences
CD19	FITC	1:100	488 530/30	BioLegend
CD7/CD10	BV421	1:100	V405 450/50	BD BioSciences
CD45	BV785	1:100	V405 780/60	BD BioSciences
Zombie	Aqua (BV510)	1:800	V405 525/50	BioLegend



**Fig 2.3-3 Isolation of HSC/MPPs and GMPs from PB and BM samples from patients with *DNMT3A R882* mutations.** (A) Sorting panel and (B) gating strategy. HSC pool (singlets/live/CD19-/CD33-/CD45dim/CD34+/CD38-/CD45RA-) cells and GMPs (singlets/live/CD19-/CD33-/CD45dim/CD34+/CD38+/CD45RA+/CD7-/CD10-) were isolated using FACS into MEM to allow differentiation into all major blood lineages.

### 2.3.2.3 Methylcellulose (MC) assays

MC is a semi-solid medium shown to be useful for human colony forming cell (CFC) assays<sup>148,149</sup>, particularly valuable for enumerating and characterising colonies formed in response to cytokines, thus evaluating the differentiation capacity of stem and progenitor cells.

Culture medium was prepared as Methocult Optimum supplemented with 2 cytokines: FLT-3 and IL-6 (Table 2.3-2). The sorted populations of cells (LT-HSCs, ST-HSCs and GMPs) were centrifuged and resuspended in 50µL PBS + 3% FBS before being added to each corresponding tube of prepared medium. The desired dose of the small molecule inhibitor being assessed (2.5µL of stock or diluted solution, 1:1000 final concentration) was also added to each tube with one control for each cell type containing the same volume of DMSO (2.5µL). Each tube was mixed using a 1ml syringe and allowed to settle before being plated 1ml/well in a 6-well Smart Dish, which in turn was placed in a 24cm<sup>2</sup> dish with surrounding small dishes containing PBS. The cells were then incubated for 14 days at 37°C and 5% CO<sub>2</sub>.

Component	Concentration	Manufacturer
Methocult Optimum	2.5ml	StemCell Technologies
FLT3	10mg/ml	Miltenyi
IL-6	10mg/ml	Miltenyi

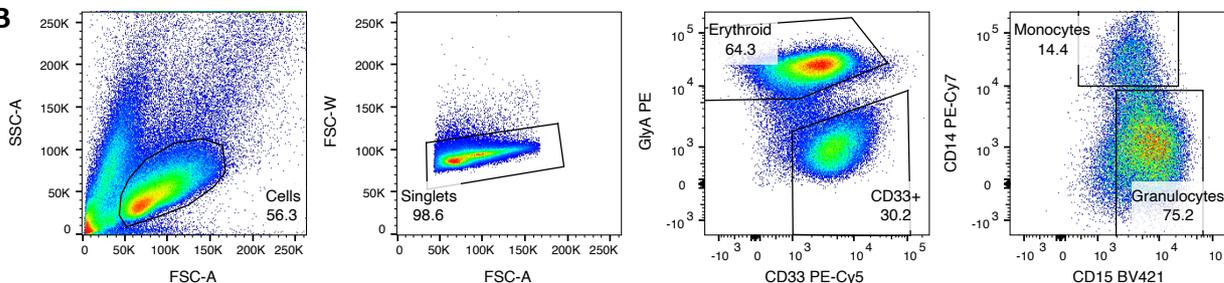
**Table 2.3-2: Components of Methylcellulose (MC) colony forming cell (CFC) assays.** Volume of Methocult is per well in a 6 well smartdish.

Colonies were visualised on days 11 and 14 using the StemVision instrument, and photographs recorded of each well. The colonies were then manually counted and categorised as erythroid (Ery), granulocytic (G), myeloid (M) or mixed (GM and Ery).

All colonies from each well were harvested on day 14 and washed twice before being resuspended in 400µL PBS + 3% FBS. 50µL cells were placed in a 96-well round bottom plate with 50µL antibody mix shown in Fig 2.3-4A and stained for 20 minutes at RT. After another wash with 100µL PBS + 3% FBS, the cells were resuspended in 200µL and analysed by high-throughput flow cytometry using the LSR II HTS Analyser. All data was combined with results of total, Ery, G, M and mixed colonies for analysis. A representative gating strategy for lineage determination in MC assays is shown in Fig 2.3-4B.

**A**

Cell surface marker	Fluorochrome	Dilution	Laser	Manufacturer
GlyA	PE	1:1000	561 582/15	BD Biosciences
CD33	PE-Cy5	1:1000	561 670/14	BD Biosciences
CD14	PE-Cy7	1:1000	561 780/60	BioLegend
CD15	BV421	1:200	V405 450/50	BioLegend

**B**

**Fig 2.3-4: Phenotyping colonies in MC assays.** (A) Antibody staining panel and (B) gating strategy for lineage determination of HSPCs cultured in MC CFC assays. Erythroid colonies were identified as GlyA+, monocytes as CD33+CD14+ and granulocytes as CD33+CD15+.

### 2.3.2.4 Liquid culture assays for hCB CD34+ cells

Thawed CD34+ cells were cultured in a cytokine enriched medium designed to support myeloid and erythroid cell differentiation (MEM, Table 2.3-1). Equal numbers of cells (10,000 – 50,000 depending on available cell count after thawing) were added to each well in a 24-well plate containing 500µL media. One control condition and the desired number of dose concentrations were generated by adding either Dimethyl Sulfoxide (DMSO) or diluted concentrations of the small molecule inhibitor at 1:1000 concentration to obtain the final desired dose of each inhibitor. Each condition was created in triplicate. Plates were placed inside a 24cm<sup>2</sup> dish and incubated at 37°C and 5% CO<sub>2</sub>.

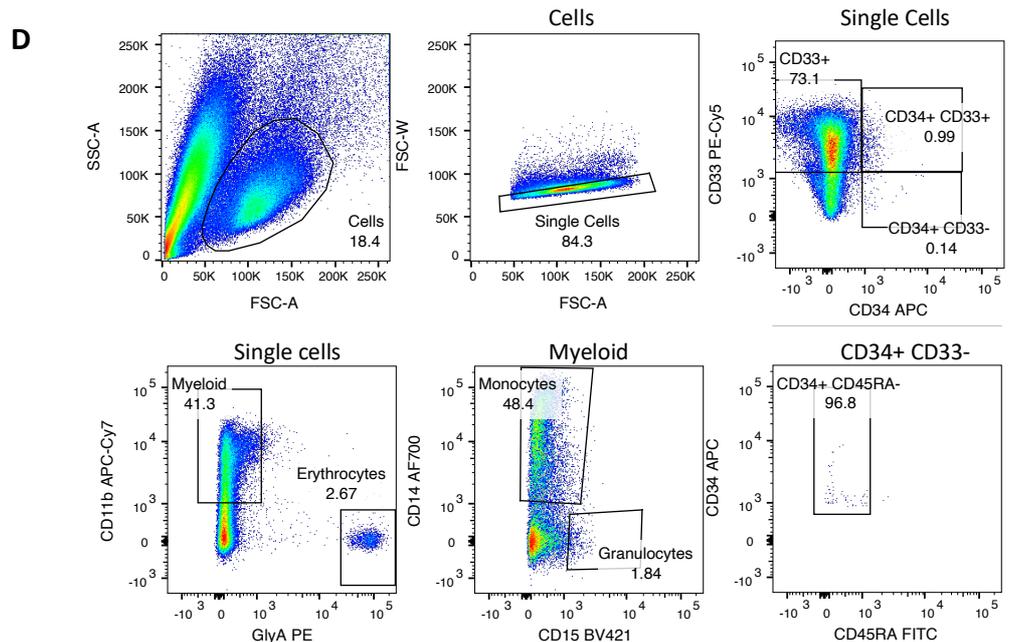
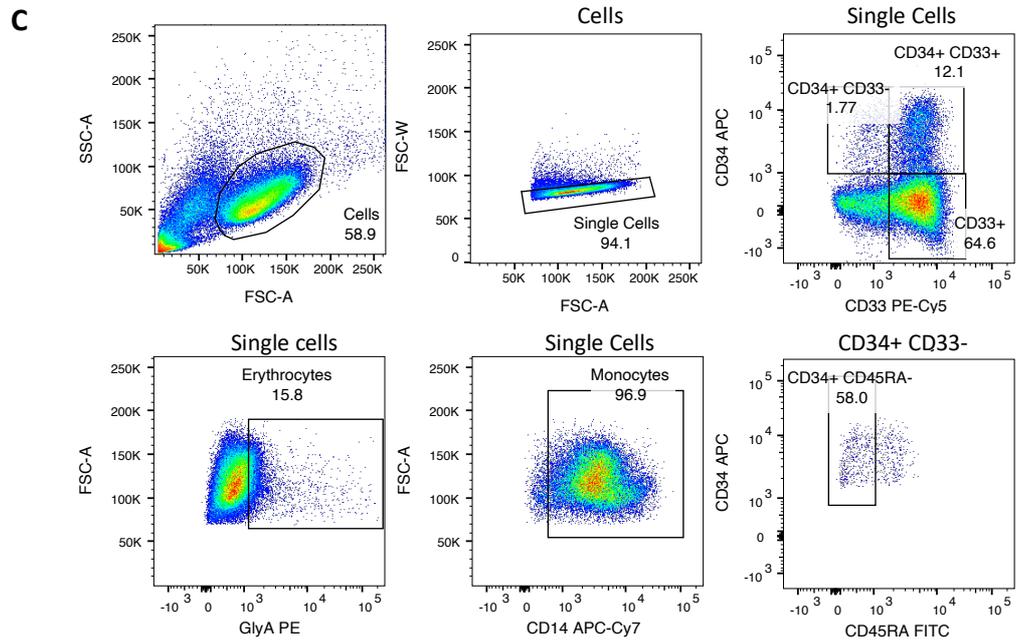
Half the volume of each well was harvested after 7 days in culture and each well was topped up to 500µL using the same media (Table 2.3-1). Harvested cells were centrifuged and resuspended in 500µL PBS + 3% FBS, of which 50µL were placed in a 96-well round-bottom plate and stained with 50µL antibody mix (Fig 2.3-5A) for 20 minutes at RT before being washed, centrifuged, and resuspended in 200µL PBS + 3% FBS for analysis by high-throughput flow cytometry using the LSR II HTS Analyser or a Fortessa analyser. The same procedure was followed for day 14, except that the entire well was harvested for analysis by the same method. The antibody panel was also altered to include cell surface markers for the detection of some additional myeloid markers (Fig 2.3-5B). Representative gating strategies for lineage determination after 7 and 14 days are shown in Fig 2.3-5C and Fig 2.3-5D respectively.

**A**

Cell surface marker	Fluorochrome	Dilution	Laser	Manufacturer
CD34	APC	1:200	640 670/30	BD Biosciences
CD33	PE-Cy5	1:1000	561 670/14	BD Biosciences
GlyA	PE	1:1000	561 582/15	BD Biosciences
CD14	PE-Cy7	1:1000	561 780/60	BioLegend
Zombie	Aqua (BV510)	1:800	V405 525/50	BioLegend

**B**

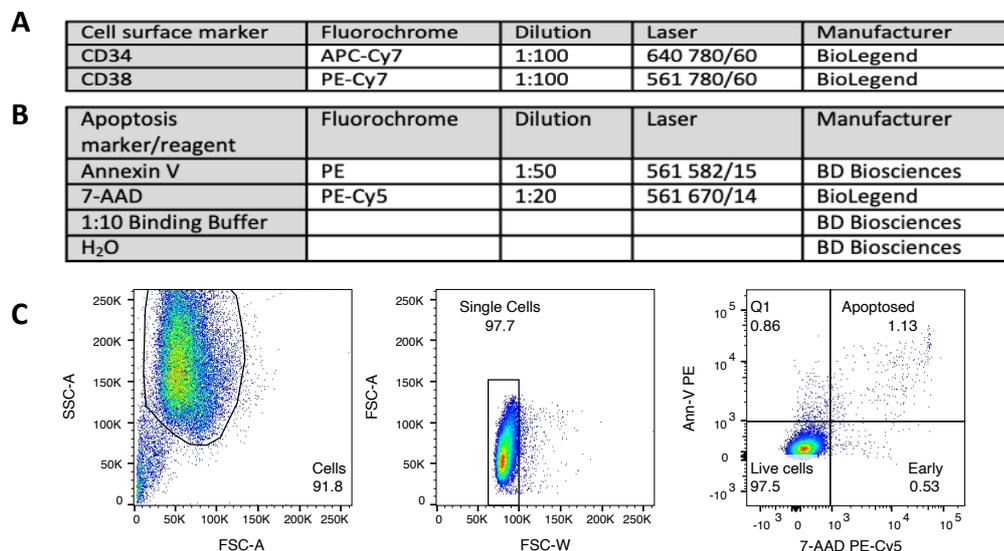
Cell surface marker	Fluorochrome	Dilution	Laser	Manufacturer
CD34	APC	1:200	640 670/30	BD Biosciences
CD14	Alexa 700	1:50	640 730/45	BioLegend
CD11b	APC-Cy7	1:300	640 780/60	BioLegend
CD33	PE-Cy5	1:1000	561 670/14	BD Biosciences
GlyA	PE	1:1000	561 582/15	BD Biosciences
CD15	BV421	1:200	V405 525/50	BioLegend
Zombie	Aqua (BV510)	1:800	V405 525/50	BioLegend



**Fig 2.3-5 Analysis of hCB CD34+ cells *in vitro*.** Antibody staining panel for harvested CD34+ derived colonies after (A) 7 and (B) 14 days. Gating strategy for lineage determination and total cell count analysis after (C) 7 days and (D) 14 days.

### 2.3.2.5 Apoptosis Assays

Thawed CD34+ cells were cultured in a MEM (Table 2.3-1). After the appropriate time in culture (24 hours, 72 hours or 7 days), cells were harvested into tubes and centrifuged prior to being resuspended in 100µL antibody mix in Fig 2.3-6A and allowed to stain for 20 minutes at RT. Cells were then washed and stained with the apoptosis staining mastermix in Fig 2.3-6B for a further 20 minutes at RT. Binding buffer was then added and cells were analysed by flow cytometry using the LSR II HTS Analyser. Representative gating strategy for determination of proportion of apoptosed cells is shown in Fig 2.3-6C.

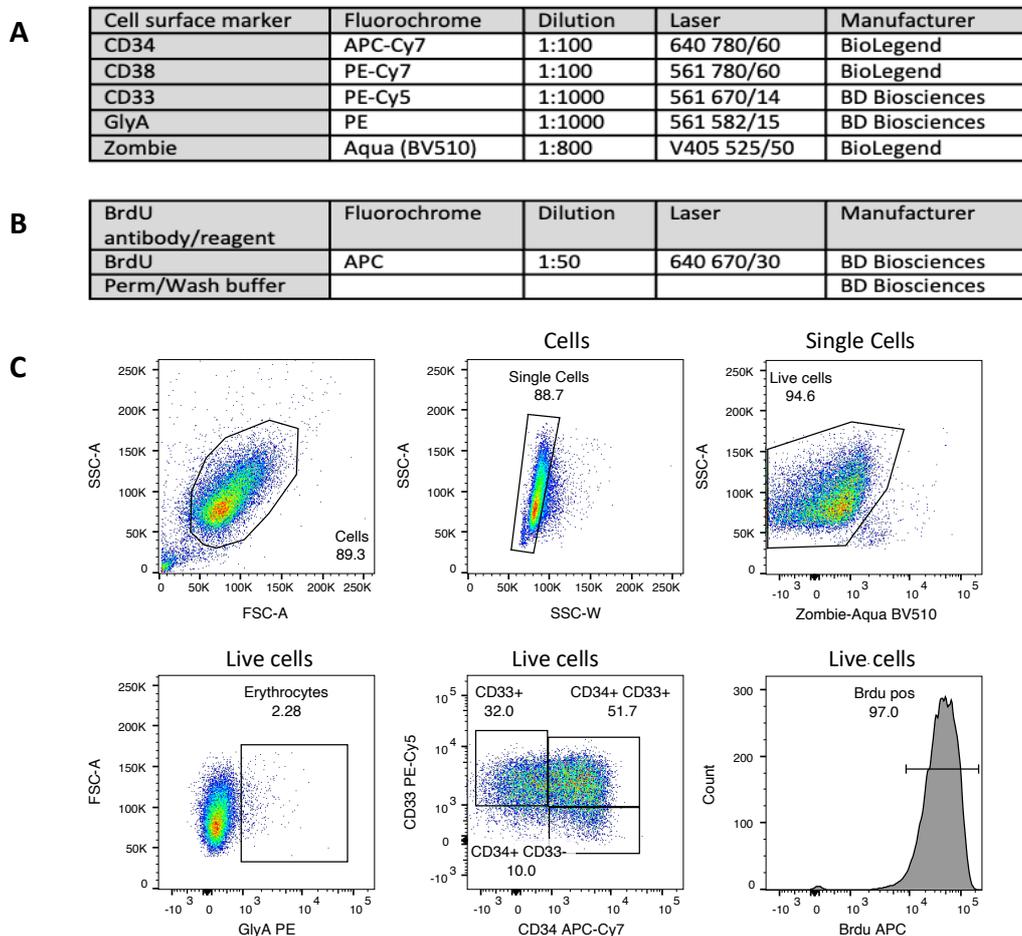


**Fig 2.3-6 Apoptosis assay.** (A & B) Antibody staining panel used to analyse hCB CD34+ cells in MEM culture treated with small molecule inhibitors after 24, 48 and 72 hours and (C) gating strategy for apoptosis determination and total cell count analysis of hCB CD34+ cells in liquid culture after 24 hours.

### 2.3.2.6 Cell cycle analysis assays

Thawed CD34+ cells were cultured in a MEM (Table 2.3-1) 8 hours prior to cell harvesting, 10µM BrdU was added to all wells except one control well to obtain a negative unlabelled control. After 72 hours in culture, cells were harvested into tubes, centrifuged, and stained for cell surface markers with the antibody panel in Fig 2.3-7A for 20 minutes at RT. Cells were then washed with PBS + 3% FBS and fixed and permeabilised using the BrdU kit, which involved 2 fixation and permeabilisation steps with washing after each step. Cells were then incubated with DNase (30µg per sample) for 1 hour at 37°C. BrdU staining was then performed with the mastermix shown

in Fig 2.3-7B at 4°C overnight. Cells were then washed, resuspended in 300µL PBS + 3% FBS and analysed by flow cytometry using the LSR II HTS Analyser. Representative gating strategy for determination of BrdU incorporation is shown in Fig 2.3-7C.



**Fig 2.3-7: Cell cycle analysis assay.** (A) Antibody staining panel, (B) BrdU incorporation assay and (C) gating strategy for cell cycle analysis of hCB CD34+ cells in MEM after 72 hours.

### 2.3.3 Tex cells

Tex cells<sup>150</sup> stored at -180°C in liquid nitrogen were thawed using pre-warmed 95% IMDM + 5% FBS + 1:100 DNase, added dropwise. Cells were then manually counted and cultured in a flask with IMDM enriched with cytokines (**Error! Reference source not found.**) for 5 days.

Cells were recounted and cultured in the same medium (**Error! Reference source not found.**) with one control condition containing 1:1000 DMSO only and 5 dose concentrations containing diluted doses of inhibitor at 1:1000 concentration in the culture well to maintain equal volumes of DMSO between conditions. Cells were cultured in a 24 well dish at 37°C and 5% CO<sub>2</sub>.

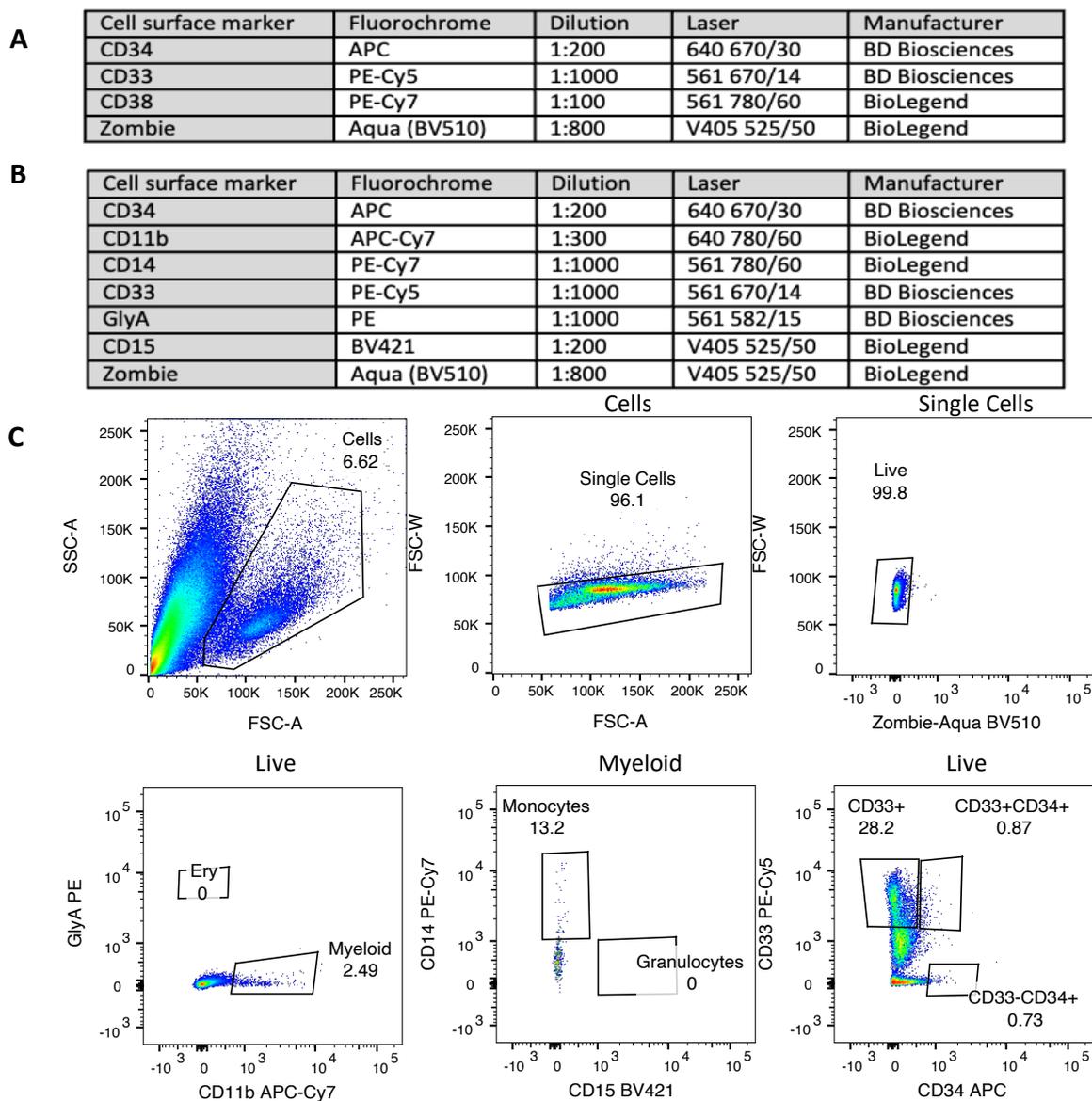
After 7 days in culture, half the volume from each well was harvested and the wells topped up with the same medium (**Error! Reference source not found.**). Cells were washed, centrifuged, and stained with the antibody mix in Fig 2.3-8A for 20 minutes at RT before being washed, centrifuged and resuspended in 200µL PBS + 3% FBS for analysis by flow cytometry using the LSR II HTS Analyser. The same procedure was followed for day 14, except that the entire well was harvested for analysis by the same method. The antibody panel was also altered to include cell surface markers for the detection of some additional myeloid markers (Fig 2.3-8B). Representative gating strategies for lineage determination after 14 days in culture are shown in Fig 2.3-8C.

Cytokine	Concentration (ng/ml)	Manufacturer
SCF	20	Miltenyi
IL-3	2	Miltenyi

Component	Volume (ml)	Manufacturer
IMDM Media	42.5	Life Technology
FBS	7.5	PanBiotech
L-Glutamine	0.5	Life Technology
Penicillin/Streptomycin	0.5	Life Technology

**Table 2.3-3: Components of cytokine enriched medium used for Tex cell culture assays.**



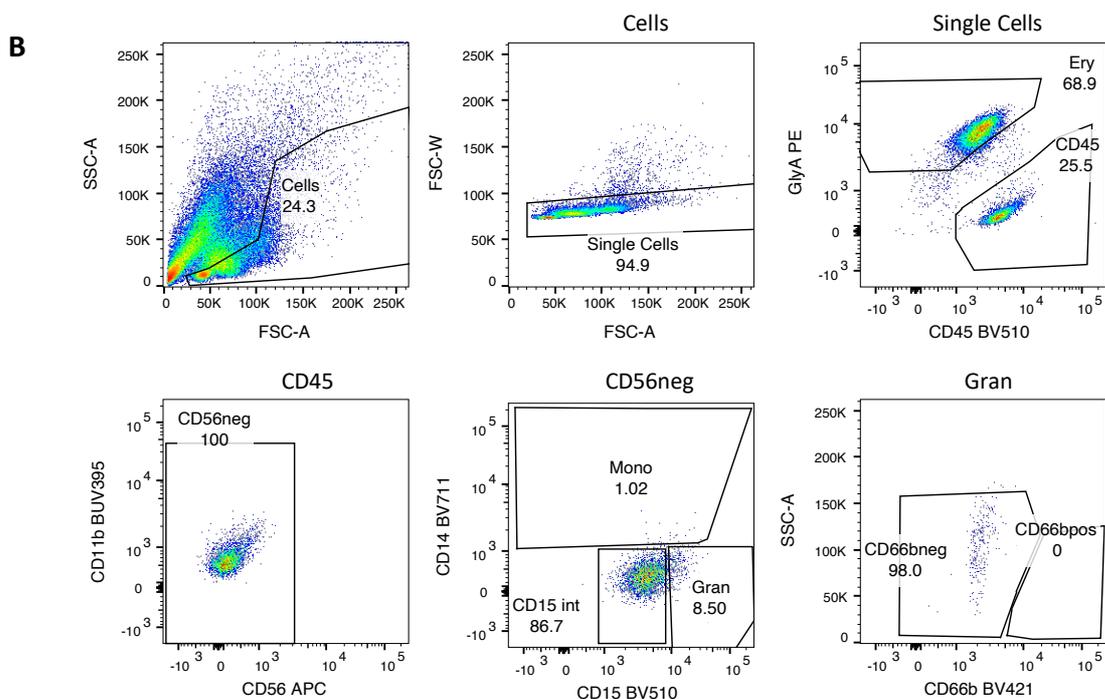
**Fig 2.3-8: Analysis of Tex cells following in vitro culture.** Antibody staining panel for Tex cells cultured in IMDM + cytokines medium used after (A) 7 days and (B) 14 days. (C) Gating strategy for lineage determination and total cell count analysis of Tex cells after 14 days in culture.

## 2.3.4 Peripheral blood and bone marrow samples

Single cell culture was performed using sterile technique and all wells containing visible colonies were harvested after 3 weeks in culture. Media containing all listed cytokines (MEM, Table 2.3-1) was replenished as required during the 3-week *in vitro* culture period as required based on manual visualisation. All colonies were harvested into a new 96 well plate and split 2/3 for flow cytometry phenotyping analysis and 1/3 for genotyping. Colonies for phenotyping were stained with the antibody panel in Fig 2.3-9A using 50µL/well for 20 minutes at RT. After washing, colonies were resuspended in 100µL PBS + 3% FCS for analysis using a Fortessa analyser. Gating strategy is shown in Fig 2.3-9B. All remaining cells from each colony were frozen in 50µL PBS for future DNA extraction.

**A**

Cell surface marker	Fluorochrome	Dilution	Laser	Manufacturer
CD56	APC	1:200	640 670/30	BioLegend
CD34	APC-Cy7	1:100	640 780/60	BD Biosciences
GlyA	PE	1:1000	561 582/15	BioLegend
CD33	PE-Cy5	1:1000	561 670/14	BioLegend
CD38	PE-Cy7	1:100	561 780/60	BD Biosciences
CD3; CD10; CD19	FITC	1:100; 1:200; 1:500	488 530/30	BioLegend
CD66b	BV421	1:100	V405 450/50	BD Biosciences
CD41a	BV510	1:200	V405 525/50	BD Biosciences
CD15	BV605	1:100	V405 610/20	BioLegend
CD14	BV711	1:1000	V405 780/60	BioLegend
CD11b	BUV395	1:1000	355 379/20	BD Biosciences



**Fig 2.3-9: Analysis of HSC/MPP and GMP derived colonies in MEM after 21 days.** (A) Staining panel and (B) example of gating strategy for HSC/MPP and GMP colonies derived from ARCH and AML samples

## 2.4 Treatment in vitro

### 2.4.1 Pharmacological inhibitors

Five pharmacological inhibitors were used in this study: KH-CB19 (*CLK1* and *CLK4* inhibitor), TG003 (*CLK1*, *CLK2* and *CLK4* inhibitor), EPZ015666, CMP5 and GSK3023591 (*PRMT5* inhibitors). GSK3023591 was provided via Material Transfer Agreement by Glaxo-Smith-Kline (GSK), while the other inhibitors were commercially purchased. All 5 small molecule inhibitors were resuspended in DMSO prior to being frozen in 5µL aliquots. Volumes of DMSO were based on the molecular weight of the inhibitors to obtain the desired molar concentration. Each suspended solution was frozen at 1000x desired concentration such that upon dilution at 1:1000 in the media containing cells for culture, the final desired dose would be obtained in each well. Details of manufacturer and suspension volumes is shown in Table 2.4-1.

Lower dose concentrations of the *PRMT5* inhibitors were obtained by diluting the stock solution of 5mM in DMSO before being further diluted 1:1000 in the culture medium to achieve the final concentration in the assay.

Small molecule inhibitor	KH-CB19	TG003	EPZ015666	GSK3203591A	CMP5
Molecular targets	CLK1, CLK4	CLK1, CLK2, CLK4	PRMT5	PRMT5	PRMT5
Molecular formula	C <sub>15</sub> H <sub>13</sub> Cl <sub>2</sub> N <sub>3</sub> O <sub>2</sub>	C <sub>13</sub> H <sub>15</sub> NO <sub>2</sub> S	C <sub>20</sub> H <sub>25</sub> N <sub>5</sub> O <sub>3</sub>	C <sub>22</sub> H <sub>28</sub> N <sub>4</sub> O <sub>2</sub>	C <sub>10</sub> H <sub>15</sub> N <sub>3</sub> O <sub>12</sub> P <sub>2</sub>
Molecular weight (g/mol)	338.19	249.33	383.44	380.48	388.34
Mass	1mg	1mg	5mg	10mg	10mg
Desired concentration	5mM	10mM	5mM	5mM	5mM
Dosage in culture	5µM	10µM	5µM	0.5µM	40µM
Manufacturer	Santa Cruz Biotechnology	Selleck Chem	Selleck Chem	GSK	Merck

**Table 2.4-1: Suspension and preparation of small molecule inhibitors of *CLK1* and *PRMT5*.** Aliquots of the molecule suspended in the desired volume of DMSO were frozen at -20°C; DMSO = Dimethyl Sulfoxide.

### 2.4.2 Lentiviral overexpression

#### 2.4.2.1 Cloning lentiviral overexpression plasmids

Plasmids containing *PRMT5* were purchased as a naked plasmid (GeneCopoeia), while the Luciferase control and empty were a donation from the Dick Laboratory, Toronto. Plasmids were transformed into One Shot Stbl3

Chemically Competent E. Coli as per supplier's instructions including 1h 37°C incubation. Transformed bacteria were incubated overnight on imMedia Growth medium agar ampicillin plates. An individual colony of each transformed bacteria was scraped and cultured overnight in a 15ml tube containing 5ml 1X broth + 1X ampicillin (100µg/ml) at 37°C. Once solutions were verified to be cloudy, 500µL was frozen at -80°C in 500µL 50% glycerol and Milli-Q water (25% final glycerol concentration).

The remaining transformed bacteria was processed using a QIAprep Spin Miniprep kit. NanoDrop 2000 Spectrophotometer was used to quantify concentrations of purified minipreps. The shuttle plasmids were then transferred to the pSMAL-BFP-empty plasmid with Gateway LR Clonase II kit as per the manufacturer's instructions. The lentiviral plasmids were analysed by Nhe1 and Xba1 (NEB R3131S and R0145L) or BamH1-HF and Xho1 (NEB R3136S and R0146S) restriction enzyme double digest. Presence of digest bands of correct sizes was verified by 1% agarose gel electrophoresis at 120V for 45-75'. The pSMAL-BFP-PRMT5 plasmids were also Sanger sequenced with primers flanking the insertion site, and sequencing-verified clones were maxipreped from frozen bacterial samples with a QIAfilter Plasmid Maxi Kit. Corning Costar Spin-X Centrifuge Tube Filters were used to purify maxipreps of lentiviral expression plasmids with spectrophotometry both before and after purification. Plasmids were stored at -20°C until use.

### ***2.4.2.2 Producing lentiviral vectors***

Lentiviral vectors were produced using a modified TransIT system protocol. Lenti-X 293T cells at low-passage were thawed and plated at 4 million cells in 20ml HEK medium (Table 2.4-2) per 15cm dish x 5 dishes and incubated at 37°C for 4 days. After aspirating medium and washing with 10ml pre-warmed PBS, 5ml trypsin was added at room temperature, then incubated at 37°C for a maximum of 1min. Trypsin was neutralised with 5ml warm HEK medium and cells collected in a single 50ml conical tube and manually counted using a haemocytometer with trypan blue staining. 2 million cells were separately plated for MOI calculations after 3 days, while 8 million Lenti-X 293T cells were plated on each of twelve 15cm dishes for incubation at 37°C for 24h. Modified TransIT solution was prepared (Table 2.4-3) and incubated at RT for 20-30min before adding 1ml evenly over the surface of each dish containing Lenti-X 293T, then cells were incubated at 37°C overnight. 500µL of 0.5M sodium butyrate was then added for 6h and removed before rinsing cells with pre-warmed PBS. Cells were further incubated for 22h with an additional pre-warmed 14ml HEK medium. Supernatant was then transferred into Parafilm-sealed 50ml conical tubes, centrifuged 5min 500g and filtered through 0.45µm PES vacuum attached conical filters for centrifugation in the ultra-centrifuge for 2h x 25,000g. Virus was then resuspended in X-vivo +1% BSA and pooled with identical viral constructs prior to being frozen in 10µL aliquots in an isolated freezer. A 5µL aliquot was retained for MOI calculations.

Component	Concentration	Manufacturer
IMDM	88%	Pan Biotech
FCS	10%	Pan Biotech
Penicillin/ Streptomycin	1%	Life Technologies
L-glutamine	1%	Life Technologies

**Table 2.4-2: Components of HEK medium for lentiviral vector production.**

Lenti-X 293T cells plated for MOI calculations were trypsinised and plated at 50,000 cells/well in a 24-well plate of HEK medium x 6 well per virus and incubated at 37°C for 24h. The 5µL aliquot of virus was used to transduce the plated cells with several dilutions ranging from 1:200 to 1:20,000 for 72h, with supernatant being replaced by HEK medium after 24h. Cells were then trypsinised, centrifuged 5min x 500g, resuspended in 250µL PBS + 3% FCS for analysis in the V405 450/50 channel on a FACS analyser for MOI calculations.

A	Component	Volume for 12 dishes (µL)	Manufacturer
	TransIT-LT1 Transfection Reagent	1152	Mirus Bio, MIR 2304
	Opti-MEM Reduced Serum Medium	6000	Thermo Fisher Scientific 31985070

B	Component	Volume for 12 dishes (µL)	Manufacturer
	RRE plasmid	1152	Maxiprepped in Laurenti lab with Qiagen 12263 kit
	VSVG plasmid		
	Rev plasmid		
	Adv plasmid		
	pSMAL-BFP		
	Opti-MEM Reduced Serum Medium	6000	Thermo Fisher Scientific 31985070

**Table 2.4-3: Modified TransIT solution for the production of lentiviral vectors.** (A) solution A, incubated for 5min prior to being added to (B) solution B.

### **2.4.2.3 Transduction of HSCs with PRMT5 lentiviral vectors**

LT-HSCs, ST-HSCs and GMPs were sorted into 200µL media (X-vivo + cytokines, Table 2.4-4) and transferred into 96-well plates where either PMRT5 LV vector or Empty LV vector were added to obtain an MOI of 250 and cultured at 37°C for 3 days. CD34+ cells from hCB were transduced in a similar way except wells were incubated with retronectin (33.3µg/ml) for 2h at RT in the dark followed by a PBS wash prior to transferring cells in X-vivo medium (Table 2.4-4) for lentiviral transduction with PRMT5, Luciferase or Empty vectors. Transduced cells were then sorted based on Blue Fluorescence Protein (BFP) positivity alongside BFP- controls either into 1.5ml Eppendorf tubes then transferred into MC culture (Table 2.3-2) for 14 days or directly into MEM media (Table 2.3-1) in 96-well plates for 21 days.

Component	Volume (ml)	Manufacturer
X-Vivo	50	Life Technology
L-Glutamine	0.5	Life Technology
Penicillin/ Streptomycin	0.5	Life Technology
Cytokine	Concentration (ng/ml)	Manufacturer
SCF	100	Miltenyi
FLT-3	20	Miltenyi
TPO	100	Miltenyi
IL-7	100	Miltenyi

**Table 2.4-4: X-vivo media for lentiviral transduction.** Cells were transduced in this media for 3 days prior to being sorted on BFP positivity and cultured further in methylcellulose or MEM media.

## 2.5 Induced pluripotent stem cells (iPSC)

### 2.5.1 Derivation of iPSCs

Using PB MNCs from 3 separate individuals (ARCH1, AML1 and AML4), the Cambridge Stem Cell Institute (CSCI) iPSC facility successfully produced a total of 27 cell lines, including *DNMT3A R882* mutant and isogenic WT controls within each individual. I successfully expanded several cell lines using TeSR E8 medium in Vitronectin XF coated 6-well plates. Embryoid bodies were seeded in clumps in plates to allow adherence to the plate surface at 37°C and 5% CO<sub>2</sub>, removed using 0.5mM Ethylenediaminetetraacetic acid (EDTA) and passaged weekly. Y-27362 (ROCK inhibitor) was added to all wells and media was changed every 24 hours during the expansion protocol. These will be used to further study the functional and metabolic effects of *DNMT3A R882* mutation on mature monocyte and neutrophil function.

### 2.5.2 Genotyping iPSC clones

Clones from within each individual sample were genotyped using Sanger sequencing for *DNMT3A R882* to determine whether each cell line is mutant or WT for this gene of interest. DNA was extracted by incubating with 40µL isopropanol at RT for 15 minutes, centrifugation at 2000rcf for 45 minutes, adding 25µL 70% ethanol and further centrifugation at 2000rcf for 10 minutes. After resuspension in 20µL dH<sub>2</sub>O, nucleic acid is quantified using NanoDrop. Primers were used to flank the site of *DNMT3A R882* mutation as listed in section 2.6.1.2, and sent to SourceBioScience for Sanger sequencing. Data returned was analysed using SnapGene Viewer in which the DNA sequence of the clone of interest was compared to the reference *DNMT3A R882* sequence. Clones were identified at *DNMT3A R882* mutant or WT.

## 2.6 Sequencing

### 2.6.1 Targeted genotyping

#### 2.6.1.1 Bulk primary PB samples

Primary PB samples from ARCH and AML patients were sequenced by Dr Tamir Biezuner in the Shlush laboratory at the Weizmann Institute of Science, Israel, as published as in Tuval et al<sup>151</sup>. DNA from CD3 depleted (or CD34+ enriched) cells from primary PB samples was extracted using the Qiagen Dneasy Blood and Tissue kit as per the manufacturer's protocol. Samples with low viable cell numbers had DNA extraction by a different method: centrifugation of 50,000 cells, lysis by incubation with 50µL NaOH 50mM at 99°C for 10min, cooling to RT and addition of 5µL Tris 1M pH8; 4.5µL of this solution was taken for library preparation.

Simultaneously, expanded T cells from each sample were used to extract DNA as matched germline control, which carries inherent issues. Since T cells and other haematopoietic cells originate from a common HSC ancestor, hence may result in somatic haematopoietic mutations being incorrectly assigned as germline. However, in the absence of available tissue from an alternate embryonic layer, e.g., mesodermal skin fibroblasts, T cells are a reasonable alternative, since the myeloid and lymphoid lineages separate relatively early in the haematopoietic tree, indicating that they arise from different HSCs/progenitors. Using this approach, mutations observed in the myeloid lineage only, and absent in T cells could be regarded as somatic. A difference of >20% VAF between T cells and myeloid cells can be considered somatic rather than germline<sup>144</sup>. In this project, DNMT3A R882 mutation detected by targeted sequencing of mature myeloid cells only and absent in T cells could be considered sufficiently somatic. While incorrectly assigning this mutation as germline remains a risk, in none of the patients in our cohort was DNMT3A R882 mutation detected in the T cells.

Libraries were prepared for Next Generation Sequencing (NGS) using single molecule Molecular Inversion Probes (smMIPs)<sup>152,153</sup>, designed with MIPgen software<sup>154</sup>. The final smMIP panel was designed to identify recurrently mutated AML hotspots in 33 genes as previously outlined in Tuval et al<sup>151</sup>. DNA was sequenced in duplicates directly following extraction without whole genome amplification.

#### 2.6.1.2 Single cell derived colonies

Genotyping was performed on DNA amplified using the Qiagen RepliG amplification kit. Approximately 1/3 of each colony was transferred to a new 96-well PCR plate. Depending on the overall clonogenic efficiency of each experiment, colonies were either transferred to the same corresponding well in the new plate, or a new map was generated correlating the new location with the originally sorted location of that HSC from which the colony was derived. After centrifugation at 2000rpm for 5min, cells were resuspended in 3µL PBS. As per the Qiagen RepliG kit, 3.5µL Buffer D2 was added to each well, which was constituted fresh as per the number of colonies to be amplified. After 10min incubation on ice, 3.5µL Stop solution was added to each well and mixed by

vortexing. Freshly prepared mastermix containing mini reaction buffer, DNA polymerase and nuclease free water (Table 2.6-1) for each colony was added as 40µL per well. Plates were incubated at 30°C for 16h, heated to 65°C for 3 minutes to inactivate DNA polymerase and kept at 4°C prior to transferring to -20°C for transport to Weizmann Institute, Israel, on dry ice for targeted DNA sequencing.

Component	Volume per reaction (µL)	Manufacturer
Nuclease free water	10	Qiagen
RepliG mini reaction buffer	29	Qiagen
RepliG mini DNA polymerase	1	Qiagen

**Table 2.6-1: Components of RepliG mastermix for DNA amplification using Qiagen kit.**

The RepliG DNA amplification kit was tested on 5 phenotypically different colonies produced by HSCs sorted from healthy donor PB cones to assess whether sufficient DNA would be yielded from a small proportion of cells taken from each colony. A range of colony types and sizes were used, ranging from 4000 cells to 93,000 cells and included myeloid, erythroid and lymphoid (Natural Killer, NK) colonies. Approximately 30% of each colony's cells were remaining after FACS analysis and were transferred to a separate 96-well plate for DNA amplification using the Qiagen RepliG kit. Nucleic acid content was assessed using Nanodrop analyser and sufficient DNA content was detected in even the smallest colonies (Table 2.6-2Table 2.6-1).

An amplicon-based approach was used to generate libraries to sequence for targeted mutations including *DNMT3A R882*, *NPM1* and *FLT3-ITD*. Following 2 PCR amplification cycles, samples were pooled. PCR purification was then performed using DNA clean and concentrator-5, and library evaluation was done on TapeStation as per manufacturer's protocol. Cleaning was performed by Blue Pippin and Qubit 4 Fluorometer was used to measure library concentration of each well using iQuant dsDNA Assay kit as per manufacturer's protocol. All sequencing was performed with MiSeq, MiniSeq and NovaSeq sequencers.

Well	Col type	Col size (cells)	Nuc acid (ng/µL)
A6	MyE	4,000	447.5
A7	MyNK	32,000	638.1
A8	My	93,000	392.1
A9	My	4,000	1523.4
A10	MyENK (multi)	16,000	396.4

**Table 2.6-2: Estimation of nucleic acid per well following DNA amplification using Qiagen RepliG kit in preparation for targeted DNA sequencing.** Sufficient DNA was present following RepliG amplification for even the smallest colonies.

Primary PCR for *DNMT3A R882* (exon 23) was performed with the following primers:

- Forward: CTACACGACGCTCTCCGATCTTAACTTTGTGTCGCTACCTC
- Reverse: CAGACGTGTGCTCTCCGATCTTTTTCTCCCCAGGGTATTTG

Secondary PCR was performed with the following primers:

- Forward: AATGATACGGCGACCACCGAGATCTACAC
- c) [Fw\_Index\_D5XX]AACTCTTTCCCTACACGACGCTCTTCCG;
- Reverse primer:
- d) CAAGCAGAAGACGGCATAACGAGAT [Rev\_Index\_D7XX]GTGACTGGAGTTCAGACGTGTGCTCTTCCG;

Paired-end 2X151bp sequencing data were converted to fastq format. Reads were merged using Bbmerge v38.62<sup>155</sup> with default parameters, followed by trimming of the ligation and extension arm using Cutadapt v2.10<sup>156</sup>. Unique Molecular Identifiers were trimmed and assigned to each read header. Processed reads were aligned using BWA-MEM<sup>157</sup> to a custom reference genome, comprised of the appropriate smMIP panel sequences  $\pm 150$  bases extracted from GRCh37 hg19. Aligned files were sorted, converted to BAM<sup>158</sup> followed by Indel realignment using IndelRealigner (GATK v.3.7<sup>159</sup>).

Variant calling was performed using the same method as in Tuval et al<sup>151</sup>. Genotype analysis was based on targeted sequencing data using 3 independent variant callers: MuTect2 in a tumor-only mode, VarScan2 v2.3.9<sup>160</sup> and Platypus<sup>161</sup> for indels. This bioinformatics pipeline can identify variants as low as 0.005<sup>162</sup>.

Colonies with adequate sequencing depth were assessed using all 3 variant callers to determine VAF and classified as WT (<2% VAF) or *DNMT3A R882* mutant ( $\geq 2\%$  VAF)<sup>90</sup>. The lower limit of detection for variants depended on the depth of coverage, which was determined by the median depth of each run (Table 2.6-3). Colonies with sequencing depth <2% of the median depth of the run were excluded. In addition, colonies with sequencing depth below which there was significant discrepancy between the 3 variant callers were also excluded. This lower limit was manually determined by examining the entire data set for each sequencing run. Data was then correlated with the phenotyping data available from FACS.

Sample	Median depth of run	Lower threshold for depth for analysis
ARCH1	4400	30
ARCH2	10014	3500
ARCH3	40090	120
ARCH4	1557	50
AML1	1275	34
AML2	5343	100
AML3	620	14
AML4	685	35
AML5	1813	85
AML6	3693	85
AML7	NA	NA
AML8	1268	5

**Table 2.6-3: Details of targeted sequencing of HSC/MPP derived colonies from ARCH and AML individuals.** Median depth of sequencing of all the samples in the same run was used to calculate the lower threshold of depth that was included in the overall analysis. Colonies with depth of sequencing below which there was discordance between 3 variant callers, or <2% of the median depth of the run were excluded from the overall analysis. AML7 only produced 4 colonies, hence insufficient data for depth/variant calling is available.

## 2.6.2 Whole genome sequencing

DNA was extracted from single HSPC derived colonies from ARCH1 for WGS using the PicoPure DNA extraction kit, which allows DNA to be recovered from a minimum of 10 cells. Colonies were transferred to a PCR plate and 17µL Proteinase K (prepared with buffer as per kit) added per well. Cells were incubated on a G-Storm thermocycler at 65°C for 6h, 75°C for 30min and maintained at 4°C until transfer to -20°C for sequencing analysis at Wellcome Trust Sanger Institute.

WGS was performed at average sequencing depths of 13x on 127 colonies derived from 1 ARCH individual. A phylogenetic tree was constructed by Dr Emily Mitchell, and terminal branch length was corrected for sequencing depth within each individual colony. The bioinformatic approach used to construct the phylogenetic tree was the same as in Mitchell et al<sup>82</sup>. Heatmaps of fluorochrome density correlating to each colony as represented on the phylogenetic tree was performed by Dr Joe Lee using R.

## 2.6.3 RNA sequencing

Larger colonies (>10,000 cells) from a total of 5 samples (ARCH1, ARCH2, ARCH3, AML1, AML4) were used to extract RNA using the Qiagen Rneasy kit. RNAseq was performed on bulk cells within individual colonies containing only mature monocytes (CD45+/CD56-/CD11b+/CD14+/CD15-/GlyA-) from a total of 44 HSC/MPP and GMP derived colonies from ARCH1. RNA was extracted from colonies using the Qiagen RNEasy kit as per manufacturer's instructions and sent to the CSCI core sequencing facility for library preparation and sequencing using the NovaSeq6000 platform. 33 colonies were included in the final analysis of bulk monocyte RNAseq results as colonies produced by GMPs and those containing few erythroid cells were excluded from the final analysis.

scRNAseq was attempted from 5 samples (AML2, AML3, AML5, AML6 and ARCH4) by leaving larger colonies (>10,000 cells) in culture for an additional week (total time *in vitro* culture 28 days) and identifying the colonies containing mature monocytes and neutrophils using phenotyping data obtained at 21 days. These colonies were then harvested, stained as per panel in Fig 2.3-9A and sorted as single mature monocytes and neutrophils into 96 well plates containing lysis buffer (Table 2.6-4). Following RNA extraction, Novaseq was attempted. A total of 3108 single monocytes and neutrophils were sorted from 499 colonies from 5 individuals, all of which did not contain sufficient nucleic acid for RNAseq.

Component	Volume (µl)
0.4% Triton X-100	2
Rnase inhibitor 20U/µL	0.1
100mM DTT	0.5
10mM dNTP	1
Nuclease Free Water	0.2

**Table 2.6-4: Lysis buffer for sorting cells in preparation for single cell RNAseq.** Volumes given per well.

## 2.7 Data analysis

### 2.7.1 Flow cytometry phenotyping analysis

Colonies derived from single HSCs from each sample were assessed for phenotype using the relevant antibody fluorochromes in FlowJo™ v10.8 Software (BD Life Sciences) whilst blinded to the genotype as this data became available only after the phenotype analysis. An example gating strategy is shown in Fig 2.3-9B. Colonies were classified as multipotent (erythroid/myeloid), erythroid (GlyA+), myeloid (CD45+/CD11b+/CD56-) or undifferentiated (C45+). Myeloid colonies were further subclassified as monocyte (CD14+/CD15-), granulocyte (CD15+/CD14-) or mono-gran (CD14+/CD15+). Granulocyte containing colonies (granulocyte or mono-gran) were sub-classified again based on CD66b expression for categorical analysis of cells in the CD66b+/- gates. P-values for all categorical colony analyses was calculated using Fisher tests. Alongside categorical classification

based on gates, the proportion of each cell type within each colony (erythroid, myeloid, multipotent, monocyte, granulocyte and neutrophils (CD66b+)) as a percentage of all singlets was compared between *DNMT3A R882* mutant and WT using Wilcoxon signed rank test for significance. The MFI for each marker of mature cells (GlyA, CD14, CD15, CD66b) was also examined and compared between WT and mutant colonies using the Wilcoxon signed-rank test for significance.

## **2.7.2 Comparing phenotype and genotype**

### **2.7.2.1 FlowPAC**

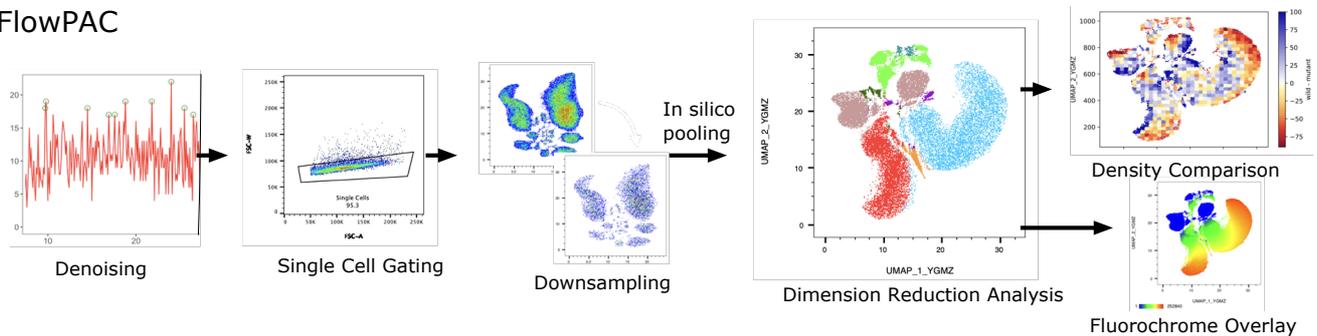
To compare phenotype and genotype in an unbiased manner, we developed a novel analysis pipeline called FlowPAC, which utilises pooled in silico analysis of all cells from mature colonies from each individual sample (Fig 2.7-1). This is a powerful method that allows us to capture subtle but relevant differentiation phenotypes. Analysis of all samples and colonies using this bioinformatic approach was performed by Daniel Hayler.

Several factors can introduce noise into fcs files generated by flow cytometry analysers, including variation in flow rate, outlier events that lie outside the instrument sensors' dynamic range and laser instability. These can be visually inspected and corrected or gated out during manual analysis of fcs files in FlowJo™ v10.8 Software (BD Life Sciences). However, since FlowPAC is an unbiased tool that automates analysis of individual cell fluorochrome markers beyond the singlets gating step, we used FlowAI (Monaco 2010) to mitigate technical anomalies. FlowAI uses an automated algorithm to return fcs files containing only those events that have passed its thresholds.

Following this initial denoising step, fcs files are manually gated for singlets based on forward and side scatter from all colonies derived from single HSCs. Each colony is then randomly downsampled to 1000 cells using the downsample plugin for FlowJo™ v10.8 Software, which maintains data structure while reducing total number of events within each colony and minimises the effect of larger colonies dominating the analysis. Colonies from within each sample are then organised into 2 pools: *DNMT3A R882* WT or mutant as identified by targeted DNA sequencing and concatenated to produce a single fcs file each for WT and mutant within each sample, maintaining reference to the source fcs file allowing integration with the traditional analysis of flow cytometry data. Two-dimensional Uniform Manifold Approximation and Projection (UMAPs) are then generated based on fluorochrome intensity across all mature cells from each sample using the UMAP plugin for FlowJo™ v10.8 Software, then WT and mutant cells are identified within the clusters. The FlowSOM package is used to generate labels for clusters within the MAP using the automatic settings. When mapped, WT and *DNMT3A R882* mutant cells overlap on the UMAP but are unequally distributed. A density distribution map is hence generated, showing the ratio of WT:mutant cells within each region of the UMAP. Mutant and WT colonies are downsampled to equal numbers within the UMAPs to better distinguish density distribution across the map using a quadrant-based approach. Both WT and mutant maps are divided into 1600 equally sized quadrants, a number based on several trials to optimise visualisation and quantification. The number of WT and mutant events in each quadrant

are represented as a percentage difference. Finally, fluorochrome intensity is overlaid onto the same UMAP to assess correlation between genotype and phenotype.

## FlowPAC



**Fig 2.7-1: FlowPAC analysis pipeline.** An unbiased analytical tool that pools all mature cells from all colonies within each individual and is used to compare phenotype and genotype of HSC derived colonies within each *DNMT3A R882* sample. The automated algorithm FlowAI is used to denoise all cells, then fcs files from the single cell gate are downsampled to 1000 cells per colony. Dimension Reduction produces UMAPs containing clusters of cells based on fluorochrome expression. *DNMT3A R882:WT* ratios are overlaid on this UMAP alongside individual fluorochrome intensities.

### 2.7.2.2 CytoTree

In another in silico pooled analysis approach, we pooled all mature from all colonies within each individual sample and generated PCA plots based on cell surface marker expression, analysing each sample separately. In this approach, no downsampling was involved and an average of 457749 cells were analysed per sample. Data is represented as log normalised and further normalised on all UMAP figures (Fig 2.7-2). This bioinformatic analysis was performed by Aleksandra Krzywon.

Flow cytometry data was used to analyse mature cells within HSC/MPP derived colonies from all 6 individuals with sufficient numbers of colonies within each genotype (>20 WT and *DNMT3A R882* mutant HSC/MPP derived colonies, consisting of >10% of the total colonies within each sample), and with healthy multilineage engraftment in immunocompromised mice. Data from the live singlets gate of each colony within each individual sample was exported as FCS files. Using CytoTree, a pre-existing R/Bioconductor package, FCS files from all colonies derived from each donor were pooled and converted into a cell expression matrix for selected markers of mature blood cell production (CD45, GlyA, CD56, CD11b, CD15, CD14 and CD66b). No normalization was performed. Metadata such as cell genotype (WT or *DNMT3A R882*) were included.

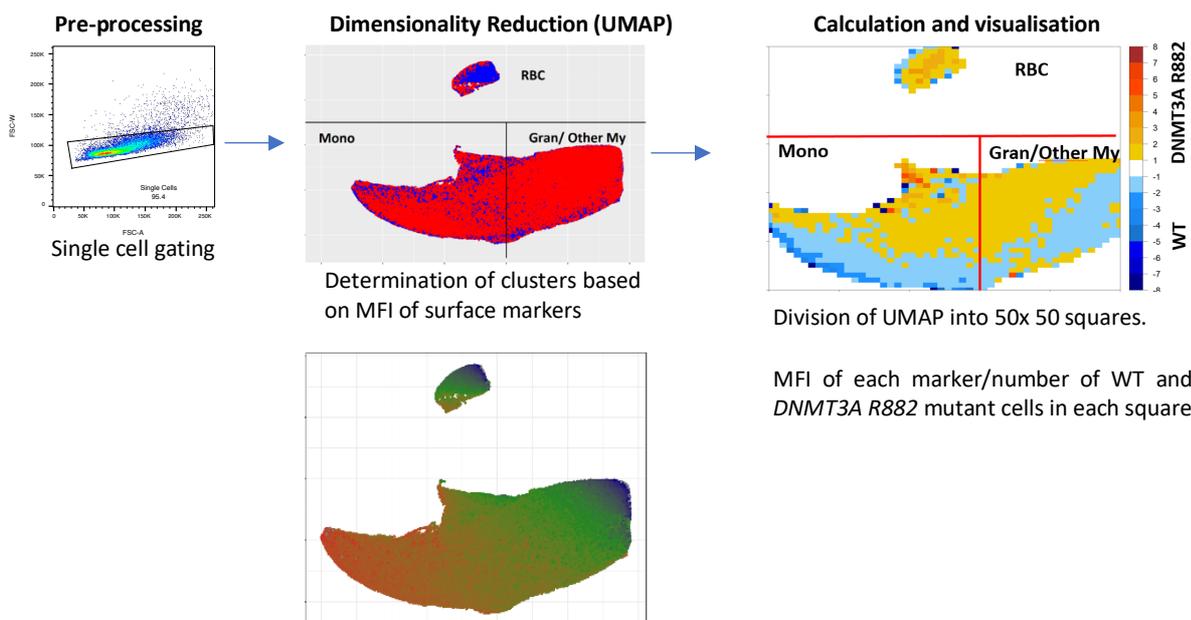
Dimensionality reduction (UMAP) was performed using all single cells without downsampling. Determination of clusters was based on visual inspection of the fluorescence intensity distribution of markers on the UMAPs. Up to 5 clusters were identified and labelled as Ery (GlyA hi, CD14 lo, CD15lo), Mono (GlyA- CD14hi, CD15lo/med), Gran (GlyA- CD14lo/med, CD15hi, CD66b hi [where available]), Other My (GlyA- CD14lo CD15lo), Gran/Other My

(GlyA- CD14lo/med, CD15med/hi, CD66b hi [where available]). These steps were processed with the help of a high performance computing system.

For determining the relative ratios of *DNMT3A R882: WT* cells across this landscape, UMAP coordinates were used. The UMAP was divided into 50x 50 squares and for each square we calculated the following:

- i) MFI of each cell surface marker as listed above, calculated independently for WT and *DNMT3A R882* mutant cells in each square
- ii) Number of WT and *DNMT3A R882* mutant cells in each square.

The ratio of the number of *DNMT3A R882* mutant cells/number of WT cells was calculated for each square and visualized on the UMAP. The ratio of MFI for *DNMT3A R882* mutant cells / MFI for WT cells was calculated for each manually annotated cluster and visualized in a dotplot. A Wilcoxon test was used to determine whether the expression of a specific marker was significantly different between *DNMT3A R882* mutant and WT cells in each cluster.



**Fig 2.7-2: CytoTree analysis pipeline.** An unbiased analytical tool that is used to compare phenotype and genotype within each sample by pooling all mature cells from all HSC/MPP derived colonies. In the pre-processing stage all cells in the single cell gate from all colonies within an individual sample are pooled together to generate a UMAP in which clusters are based on MFI of 11 different fluorochromes representing cell surface markers. Colony types determined by conventional gating are

overlaid on to this map. Once the map is divided bioinformatically into 50x50 equal squares, the ratio of *DNMT3A R882:WT* cells are overlaid alongside the MFI of relevant individual fluorochromes.

## 2.7.3 RNA sequencing analysis

All RNA sequencing analysis was performed by Kendig Sham, bioinformatician in Dr Elisa Laurenti's group.

### 2.7.3.1 Sequencing reads analysis

RNAseq reads were aligned and mapped against Ensembl genes<sup>163</sup> (GRCh37) using STAR<sup>164</sup> (version 2.5.0a). Duplicate reads were marked and removed from the BAM files using picard (version 2.20.8, Broad Institute 2019). Cell counts were quantified using HTSeq<sup>165</sup> (version 0.6.0). After quality control, <5 cells were removed from the analysis of each experiment. Highly variable genes were calculated based on method described in Brennecke et al<sup>166</sup>.

### 2.7.3.2 Gene expression analysis

Differential gene expression was used to generate a PCA map identifying 3 clusters, which were further explored for other factors including proportion of WT and mutant colonies within each cluster, monocyte maturity, colony size and differential gene expression between the clusters. Genes and transcription factors upregulated in mutant vs WT colonies were identified overall and within the defined clusters. Cytoscape was used to identify gene networks with significant differential expression between clusters and Reactome was also used to define interactions between differentially expressed genes and transcription factors.

Single LT-HSCs were sorted and cultured *in vitro*. Genes with significantly differential expression during early quiescence exit, i.e. 6 hours or 72 hours with Palbociclib (*CDK6* inhibitor) in culture compared to baseline (0 hours) were identified from scRNA-seq data. These were selected on the basis of an adjusted p value  $\leq 0.05$  (calculated with DESeq2 algorithm). Genes significantly differentially expressed after both 6 hours *and* 72 hours with Palbociclib were overlaid by generating a Venn diagram using the online tool Ugent (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) to shortlist those genes common to both these groups of data.

The bioinformatics resource DAVID (Database for annotation, Visualisation and Integrated Discovery, v6.8)<sup>167</sup> was used to analyse genes with statistically significant differential expression in early quiescence exit compared to baseline. The DAVID Functional Annotation Tool was used to find the biological categories enriched in our list of genes. To visualise these categories, the Cluego plugin<sup>168</sup> for Cytoscape<sup>169</sup> was used.

The package 'bglab' was used to get log10 normalised counts (Jawaid, W., 2017, <https://github.com/wjawaid/bglab>). Genes were filtered out if not expressed in at least 3 cells with at least 1

count. The filtered count matrix has 44819 genes and were used as input for gene set variation analysis (GSVA). Gene set variation analysis (GSVA) was performed in R using the package 'GSVA'<sup>170</sup>. The function 'gsva' was used to calculate GSVA scores. C2 curated gene set downloaded from <http://www.gsea-msigdb.org/gsea/downloads.jsp> was used.

Ranked lists of genes were prepared based on the 'stat' value of the Deseq2 results, they were sorted in a descending order. Ranked list was used as input for gene set enrichment analysis (GSEA). C2 curated gene set downloaded from <http://www.gsea-msigdb.org/gsea/downloads.jsp> was used.

## **2.7.4 Literature review**

The genes shortlisted above were examined within the significant clusters identified using the bioinformatics tools DAVID and Cytoscape. Each gene was studied using the following publicly available gene databases to understand their known function and role in haematopoiesis and haematopoietic malignancies:

- <http://www.genecards.org/>
- <http://www.genecopoeia.com/>
- <https://www.ncbi.nlm.nih.gov/gene>

Further literature review was conducted searching PubMed and reference lists of journal articles for relevant published literature. Splicing genes and methyltransferases emerged as gene clusters of particular interest given their extensively studied role in haematopoiesis and malignancy. Genes were prioritised based on the commercial availability of small molecule inhibitors.

## **2.7.5 Statistical analyses and software**

All flow cytometry data was acquired with FACSDiva version 6.2 (BD Biosciences, San Jose, CA, USA) on an LSR II or Fortessa cytometer (BD Biosciences, San Jose, CA, USA) and analysed with FlowJo™ v10.8 Software (BD Life Sciences) to devise the appropriate gating strategy and statistics for the relevant cell populations. All statistical analyses and data visualisation were performed using GraphPad Prism v7 (GraphPad Software, La Jolla, CA, USA) and the R statistical package ([www.r-project.org](http://www.r-project.org)). Comparisons within the FlowPAC pipeline were performed using Python, while CytoTree is a statistical package in R. Statistical comparisons are based on Fisher tests for categorical colony analyses, one-way ANOVA for mature cell populations, t-test and Wilcoxon for cell counts/colony sizes. P values on all graphs are represented as \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001. Some figures created with BioRender.com.

## ***2.7.6 Nomenclature for genetic mutations***

All genetic mutation nomenclature will follow the standard as described in Ogino et al<sup>171</sup>, with mutations in genes written in italics, capitals for human and lower case for mouse.

# Chapter 3 HSC function in health

## 3.1 Introduction

### **3.1.1 Definitions**

*Quiescence* is defined as a specific haematopoietic stem cell (HSC) state, with cells residing in the G<sub>0</sub> state of the cell cycle, not actively proliferating or differentiating. *Pharmacological quiescence* is an induced state *in vitro* or *in vivo* produced artificially by the administration of drugs, usually small molecule inhibitors of known molecular regulators of stem cell quiescence. *Quiescence exit* is the phase of entering the cell cycle prior to the restriction point (R, marked by the G<sub>1</sub>/S checkpoint, Fig 1.3-2), also termed as G<sub>0</sub> exit. It is considered part of *activation*, which is defined here as the transition from quiescence to the end of the first division *in vitro* (Fig 1.3-2).

### **3.1.2 Regulators of HSC function**

Certain groups of genes and genetic pathways have been identified as carrying prognostic and therapeutic significance in malignancy and have also been implicated in regulation of HSPC function. These genes should be further studied in the normal, healthy HSC context to understand their potential role in HSC quiescence and differentiation. This understanding has the subsequent potential to then be applied in the study of the role of these genes in cancer. Some selected groups of such genes are reviewed in more detail below.

#### **3.1.2.1 Methyltransferases**

This group of enzymes control protein, amino acid, DNA and RNA methylation, a process affecting transcription and epigenetic modification<sup>172</sup>. They have been strongly implicated in haematological malignancies as well as stem cell biology. Arginine methylation is a common post-translational modification and is altered by DNA, RNA and protein methyltransferases<sup>173</sup>.

DNA methyltransferases (DNMTs) are the group of enzymes responsible for DNA methylation (DNAm), an epigenetic process essential for normal development<sup>174</sup>. They include *DNMT1*, *DNMT3A* and *DNMT3B*, which are essential for de novo methylation during embryonic development<sup>175–177</sup>. Embryonic Stem Cells (ESCs) have been shown to maintain stem cell properties and chromosomal stability in the absence of these genes<sup>178</sup>. They have also been described in cancer including haematological malignancies<sup>179</sup> and alteration of differentiation potential of HSCs when DNMT activity is reduced<sup>180</sup>. DNMT1 has been shown to be required for HSC self-renewal in mice as demonstrated by serial transplantation, and loss of function mutations impair serial transplantation and result in loss of retention of the BM niche<sup>181</sup>. *DNMT3A* is discussed in further detail in Chapter 4.

Protein methyltransferases (PRMTs) are responsible for post-transcriptional modifications, particularly acting on arginine and lysine, with mRNA splicing identified as a key role of *PRMT5* in various species<sup>182–186</sup>. There are 9 known PRMTs encoded in mammalian genomes, which are ubiquitously expressed and affect important cellular processes controlling cell growth, proliferation and differentiation<sup>173</sup>. They have also been shown to play a crucial role in maintaining pluripotency or inducing differentiation of ESCs<sup>187</sup>. *CARM1/PRMT4*<sup>188</sup>, *PRMT5*<sup>189</sup> and *PRMT7*<sup>190</sup> have been implicated in the regulation of pluripotent states and ESC development in mouse models. *PRMT1* and *PRMT5* are the major asymmetric and symmetric arginine methyltransferases, respectively, and complete loss of either is not compatible with embryonic mouse or cell viability<sup>191</sup>.

*PRMT5* is the type II and major symmetric arginine methyltransferase, which has been implicated in maintenance of pluripotency and development of mouse ESCs, with *PRMT5*-KO mice being embryonically lethal<sup>189</sup>. *PRMT5*-dependent methylation has also been shown to suppress differentiation and to conserve pluripotent states in mouse ESCs<sup>189</sup>. However, these results have not been replicated in human ESCs, where pluripotency appears to be independent of *PRMT5*<sup>192</sup>. In cancer cell lines *PRMT5* is considered a common essential gene, i.e. it is one of the highest ranking most depleting genes in at least 90% of cancer cell lines as per data from the Cancer Dependency Map portal ([www.depmap.org](http://www.depmap.org)).

The precise role of the *PRMT5* gene in haematopoiesis has not yet been established, though there is evidence that human and mouse HSCs do not respond in the same manner to its suppression. In mouse HSCs it has been shown to be critical for maintaining normal haematopoiesis<sup>186</sup>, with its deletion associated with rapid HSC exhaustion<sup>186,193</sup>, and in some cases p53-induced apoptosis. *PRMT5* depletion results in increased HSC size, protein synthesis rate and HSC loss, which is partly a result of *AKT/mTOR* pathway activation<sup>194</sup>. However, another group showed that in hCB CD34+ cells, knockdown of *PRMT5* results in increased cell differentiation and proliferation<sup>195</sup>. Hence, the role of *PRMT5* in human haematopoiesis is incompletely understood and the data is inconsistent. The applicability of haematopoietic dynamics seen in mice is constantly questioned as a suitable model for human haematopoiesis, which not only functions over a much longer lifespan, but is also frequently exposed to infectious and other stressful stimuli, compared with laboratory mice maintained under sterile, steady state conditions.

*PRMT5* was originally identified as a transcriptional repressor<sup>196</sup>, and has since been implicated as an oncogene due to its abilities to suppress the expression of tumour suppressor genes<sup>197,198</sup>. Overexpression of PRMTs has been observed in a number of malignant conditions including breast cancer (*PRMT1*, *PRMT2*, *PRMT3*, *PRMT4/CARM1*), leukaemia/lymphoma (*PRMT1*, *PRMT5*), and prostate cancer (*PRMT1*, *PRMT4/CARM1*)<sup>173</sup>. There are suggestions from pre-clinical data that inhibition of some PRMTs may have a therapeutic benefit in leukaemia<sup>199</sup>. *PRMT1* interacts with the *RUNX1-ETO* fusion protein in AML and is critical for the transcription activation of AE9a, which is a splicing isoform of *RUNX1-ETO*, suggesting a potential therapeutic target in *RUNX1-ETO* AML<sup>200</sup>.

Given the implications of this gene in haematological<sup>201</sup> and solid malignancies, potent and selective *PRMT5* inhibitors have been developed and shown to have activity in a number of cancer cell lines<sup>203,204</sup> (Table 3.1-1). These have also been used *in vivo* in mouse models to assess effects on HSC proliferation and differentiation<sup>194</sup>, although the direct effects of *PRMT5* inhibition on human HSCs in steady state has not been established. This is important pre-clinical data if *PRMT5* inhibition with EPZ015666, CMP5 or GSK3203591<sup>203</sup> are to be considered for treatment of haematological and solid malignancies, in order to minimise BM and haematopoietic toxicity, since *PRMT5* has also been shown to have a critical role in haematopoiesis.

EPZ015666 (GSK3235025) (Chan-Penebre 2015, Kaushik 2017)	GSK3203591 (Gerhart 2018)	GSK3326595 (Gerhart 2018)	CMP5 (Alinari 2015)
Competitively binds in with substrate peptide of S-adenosyl-L-methionine (SAM)  IC50 6.2 ± 0.8nM in various cell lines  Potent, selective PRMT5 inhibitor	In vitro tool molecule  Highly potent and selective PRMT5 inhibitor	In clinical development  Highly potent and selective PRMT5 inhibitor	Selective PRMT5 inhibitor  Prevents EBV driven B cell immortalization Induces lymphoma cell death > normal B cell death

**Table 3.1-1** Currently available *PRMT5* inhibitors.

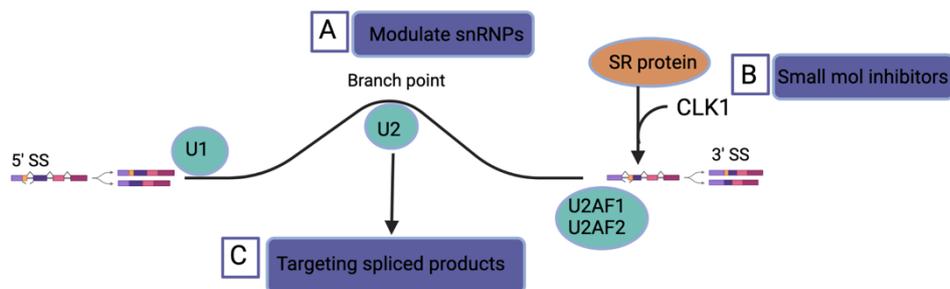
### 3.1.2.2 Spliceosome genes

The spliceosome is a large complex molecular machine found in the cell nucleus, composed of 5 small nuclear ribonucleic acids (snRNAs) and protein complexes. Transcriptional diversity is highly regulated and controlled by the spliceosome, and its main function is alternative splicing (AS), which is a post-transcriptional regulation mechanism involving removal of introns from a transcribed pre-mRNA<sup>205</sup>. AS of core transcription factors (TFs) have been shown to affect pluripotency control, self-renewal and lineage specification in ESC<sup>206,207</sup>, and HSCs<sup>208,209</sup>, especially during development, although the role of splicing factors in HSC cell fate decisions during steady state is not well understood.

Somatic mutations in genes encoding core subunits of the spliceosome are of particular interest, as they have been identified in several different cancer types including myelodysplastic syndrome (MDS)<sup>210</sup>, acute myeloid leukaemia (AML) and chronic lymphocytic leukaemia (CLL) as well as solid tumours<sup>211</sup>. Genes regulating pre-mRNA splicing occur in 50% of patients with MDS, with variable prognostic significance in patients with myeloid malignancies. More recently these genes have been identified in CH<sup>212,213</sup>. Most frequently mutated genes in this group include *SF3B1*, *U2AF1* and *SRSF2*<sup>121,123,214</sup>. These genes control the supply of full-length, functional mRNAs coding for proteins essential to cell growth and survival in cancer cells. Understanding their precise role

in tumorigenesis may provide novel therapeutic targets in these conditions. However, their role in normal HSC biology and function is yet to be elucidated and improved understanding may aid in minimising toxicity to the normal HSC pool when targeting these genes as part of therapy for haematological and other malignancies. Several groups are currently developing pre-clinical and clinical trials to study the effect of pharmacological perturbation of splicing on the outcomes of mutant haematological malignancies<sup>214</sup>.

Spliceosome function may be altered by targeting various components of the machinery from upstream mRNA binding affinity of small nuclear ribonuclear proteins (snRNPs), to downstream depletion of spliced products (Fig 3.1-1)<sup>211</sup>. Cdc2-like kinases (CLKs) are dual specificity protein kinases, that phosphorylate serine/arginine in the nucleus and modulate the selection of splice sites during pre-mRNA processing<sup>211,215</sup>. There have been significant attempts at generating compounds that inhibit the CLK family of genes with the aim of targeting and inhibiting cancer cell growth<sup>215,216</sup>. Inhibition of *CLK1* was also recently found to induce autophagy<sup>217</sup>, an important characteristic that may be manipulated to target cancer cell growth. However, most *CLK1* inhibitors are non-selective, making them difficult to use for targeting particular CLK genes.



**Fig 3.1-1 Potential targets in the spliceosome machinery.** (A) Upstream alteration of binding affinity of small nuclear ribonuclear proteins (snRNPs), (B) Small molecule inhibitors targeting SR protein phosphorylation e.g., CLK inhibitors, (C) Depletion of downstream spliced products. *Adapted from Saez et al, Blood 2016.*

### 3.1.3 Clinical applications

A common cause of morbidity and mortality in patients with haematologic malignancy is disease relapse. Strategies to maintain HSC/leukemic stem cells (LSC) quiescence and/or eradicate LSCs can also be therapeutically used for the prolongation or maintenance of remission in these patients to facilitate haematopoietic stem cell transplant (HSCT) and achieve cure. However, a major constraint to this approach is bone marrow (BM) and HSC toxicity, which limits the tolerability of maximal doses of treatment. Hence a deep understanding of the effects of therapeutic agents on steady state human HSCs is a critical component of pre-clinical studies. Maintenance of a robust HSC pool is critical in health and even more critical in disease, particularly in the context of patients undergoing myelosuppressive treatment for malignancy. Through detailed

analysis and understanding of the molecular regulators of HSC quiescence, quiescence exit and differentiation, pharmacological and genetic modification of these regulators is possible for therapeutic purposes.

## 3.2 Aims

To identify genes, transcription factors and pathways that are involved in HSC function, our group studied the transcriptional changes that occur during HSC exit from quiescence. scRNAseq of LT-HSCs was performed previously by the group at specific time points after these cells were cultured *in vitro*: 0 hours, 6 hours, 24 hours, and 72 hours. LT-HSCs maintained in quiescence by treatment with Palbociclib, a *CDK6* inhibitor, were also studied. This analysis generated a unique data set to identify genes with significantly altered expression during exit from quiescence. These genes are hypothesised to play an important role in regulation of stemness and cell fate decisions in HSCs.

The specific aims of this chapter are to:

1. Utilise the above unique data set to identify candidate genes with altered expression during quiescence exit in normal human cord blood (hCB) LT-HSCs and prioritise those that are pharmacologically targetable and belong to gene families that have known involvement in malignancy.
2. Study the role of these genes in cell fate decisions of hCB HSPCs by pharmacologically modulating their expression by using small molecule inhibitors.

## 3.3 Results

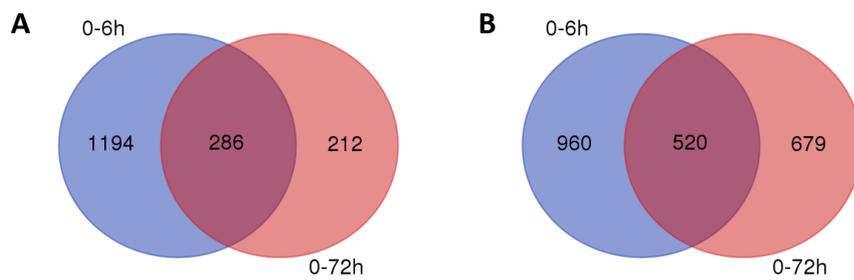
### **3.3.1 Genes with altered expression early in LT-HSC**

#### **activation**

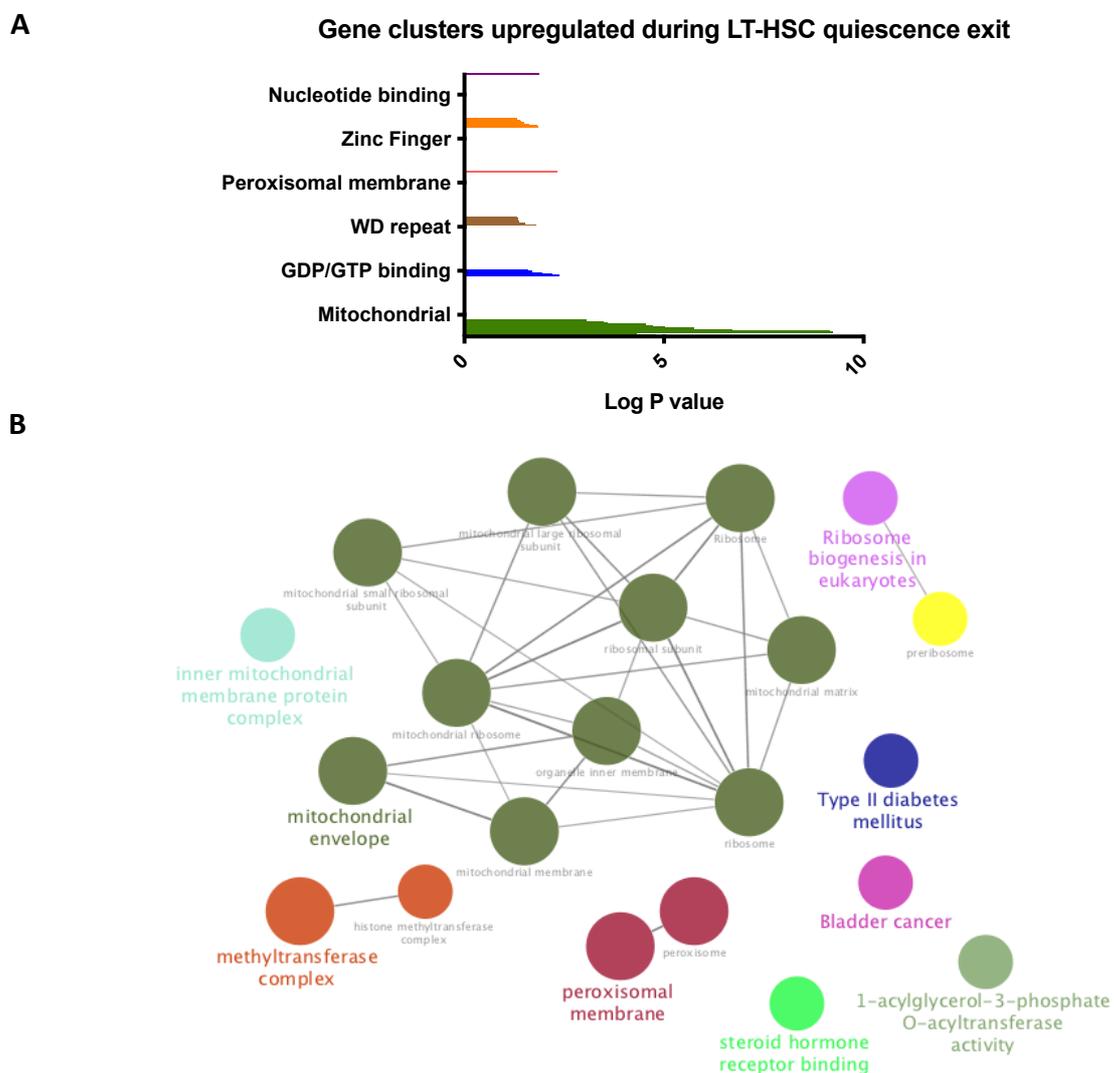
To understand the molecular regulators of human LT-HSC activation, pharmacological inhibition of the enzyme *CDK6* via Palbociclib was used to maintain these cells in quiescence in the pre-R point of the cell cycle. Dr Serena Belluschi and Dr Carys Johnson, 2 PhD students in the Laurenti group have verified that Palbociclib reversibly arrests hCB and mPB HSCs before the R point marked by phosphorylation of RB (before entry into late G<sub>1</sub>, Fig 1.3-2). Single LT-HSCs from hCB were analysed by scRNAseq in at various stages after culture initiation: 0 hours, 6 hours 24 hours, 48 hours, 72 hours and 72 hours with *CDK6* inhibition (200 nM Palbociclib, performed by Dr Serena Belluschi). Analysis of scRNAseq data obtained from these single LT-HSCs, bulk quiescent LT-HSCs, pharmacologically quiescent LT-HSCs and activated LT- and ST-HSCs or MPPs identified thousands of genes that were differentially expressed between these phases (performed by Dr Serena Belluschi and Dr Elisa Laurenti).

My aim was to identify a subset of genes for which expression changes significantly early in the activation process relative to the quiescent baseline (0 hours). I therefore subsetted the genes differentially expressed between 0 hours and 6 hours in culture and those differentially expressed between 0 hours and 72 hours in culture with the *CDK6* inhibitor Palbociclib, hence maintained in pharmacological quiescence before entry into late G<sub>1</sub>. Overall, using the Ugent online tool, a Venn diagram was generated showing that 286 genes were significantly upregulated under both these conditions compared to baseline (Fig 3.3-1A), and a further 520 genes were downregulated compared to the quiescent baseline (Fig 3.3-1B). I then further shortlisted these genes using publicly available gene databases to identify the major gene clusters present in these groups. Using the DAVID Functional Annotation Tool<sup>167</sup>, 34 significant gene clusters were identified within the 286 genes with relative upregulation compared to baseline. These clusters are based on known gene functions and organised by statistical significance (Fig 3.3-2A). To visually represent the enriched gene clusters into a network, I used the Cluego<sup>168</sup> plugin for Cytoscape<sup>169</sup>. This revealed over-representation of mitochondrial envelope and inner mitochondrial membrane protein complex genes (Fig 3.3-2B). Both functional annotation tools also identified splicing genes, which are of particular interest given their role in the malignant context.

Further shortlisting and analysis of these gene networks was performed by examining the published literature and publicly available gene databases to understand their function and role in HSCs and malignancies. 20 genes were selected as of interest based on their known function, the degree of differential expression in quiescence exit compared to baseline and the availability of experimental tools such as small molecule inhibitors, CRISPR guides or lentiviral vectors. Amongst this list were *PRMT5*, *GEMIN6*, *BATF* (Fig 3.3-3A), *DEGS1* (Fig 3.3-4), *SRSF2* and *CYP1B1*. The list of genes of interest whose expression is downregulated in quiescence exit include *CLK1*, *MAFF* and *DUSP1* (Fig 3.3-3B).

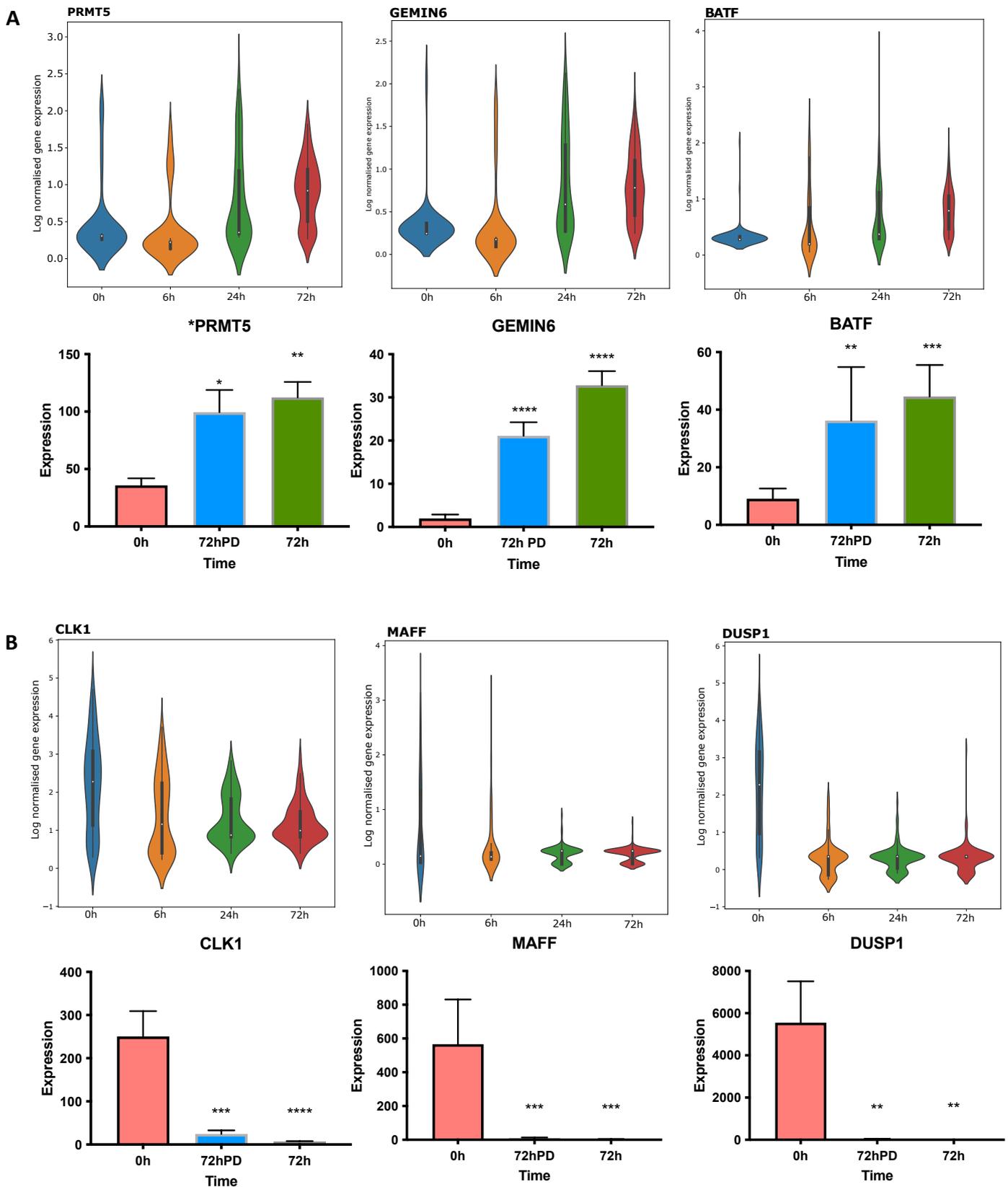


**Fig 3.3-1 Differentially expressed genes between baseline (0 hours) and early activation (6 hours or 72 hours with *CDK6* inhibition – Palbociclib).** Venn diagram showing (A) number of genes upregulated or (B) downregulated compared to baseline, generated using Ugent online tool.



**Fig 3.3-2 Identification of statistically significant gene clusters of gene sets with upregulated expression in early quiescence exit of LT-HSCs.** (A) Each bar represents a gene cluster as identified using the DAVID functional annotation tool, organised into broad groups as listed. (B) Visual

representation of network of gene clusters using Cytoscape with upregulation on quiescence exit in LT-HSCs. Each node represents a gene set, which are connected if they contain genes in common.



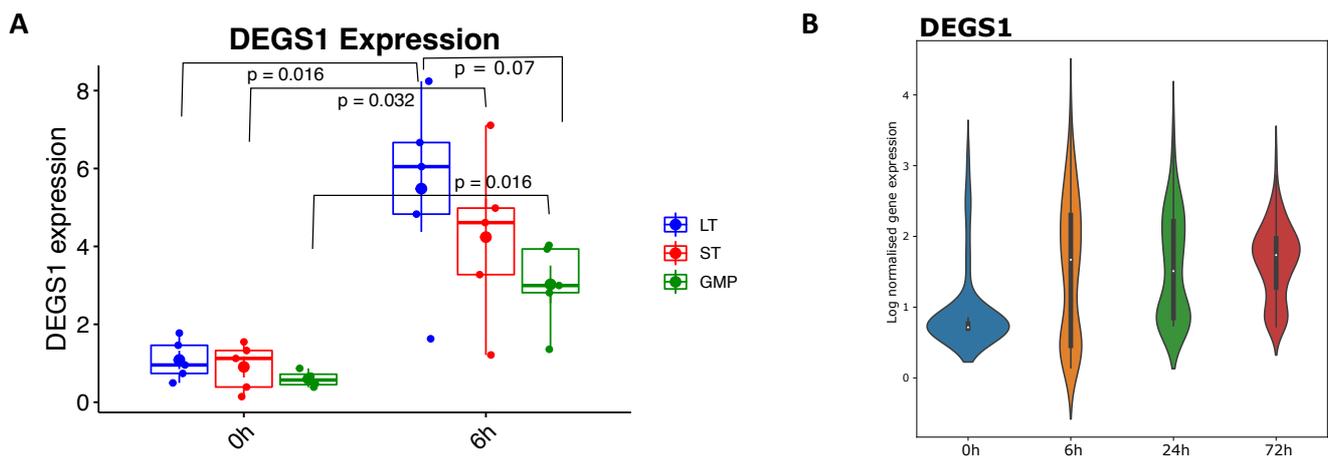
**Fig 3.3-3 Expression of selected genes at defined timepoints.** (A) Violin plots showing expression of *PRMT5*, *GEMIN6* and *BATF*; (B) *CLK1*, *MAFF* and *DUSP1* at defined timepoints. All data derived from single cell (violin plots) and bulk RNAseq results (20 cells, bar charts) from LT-HSCs from hCB expanded *in vitro*. Violin plots generated by Kendig Sham. P-values represent one sample t-test, and in all figures p-values are represented as \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

### 3.3.1.1 Genes upregulated on quiescence exit

- a) ***PRMT5*** – this gene is significantly upregulated with activation of LT-HSCs (Fig 3.3-3A). This type II protein arginine methyltransferase is an important transcriptional regulator, implicated in pluripotency and tumorigenesis. It is known to be upregulated in leukaemia, lymphoma and solid tumours. PRMTs are ubiquitously expressed and regulate important cellular processes affecting cell growth, proliferation and differentiation. This is predominantly through the processes of methylation of histones, transcription factors and indirect modification of splicing<sup>173</sup>. *Prmt5* is also essential for the maintenance of adult mouse HSCs<sup>218</sup>.
- b) ***GEMIN6*** – Its expression was significantly upregulated in early quiescence exit of LT-HSCs in single cell (Fig 3.3-3A top) and bulk RNAseq analysis (Fig 3.3-3A bottom). Gem nuclear organelle associated protein 6 is part of the spliceosome machinery and involved in cytoplasmic assembly of snRNPs. It is not well studied in malignancy, but is known to be required for the stability of the motor neurone complex from studies in spinal muscular atrophy<sup>219</sup>.
- c) ***DEGS1*** – an endoplasmic-reticulum spanning protein functioning as an enzyme in the de novo sphingolipid synthesis pathway.

Lipid homeostasis has been recognised as being critical for cellular health and maintaining cellular membrane dynamics for cell division in cell lines<sup>220</sup>. The unfolded protein response and autophagy have been identified as being important mechanisms for mediating the HSC stress response<sup>221</sup>, and the metabolic requirements of HSCs are highly adaptive to their cellular state<sup>40</sup>. Work in Prof John Dick's group had shown that the differential expression of lipid signalling and metabolism genes on highly purified HSPC subsets as assessed by transcriptome analysis suggests that the sphingolipid pathway plays an important role in HSC repopulation and maintaining the HSC hierarchy. Xie et al<sup>60</sup> demonstrated that *DEGS1* inhibition with the synthetic retinoid fenretinide/N-(4-hydroxyphenyl) retinamide (4HPR) alters HSC function and lineage balance *in vitro* and results in HSC autophagy, distinct from function in progenitors. 4HPR activates stress pathways that maintain HSCs in stemness immunophenotypically and functionally in *ex vivo* culture. This study showed that *DEGS1* is functionally required for HSC repopulation, shown transcriptionally and also by pharmacological inhibition of *DEGS1* with 4HPR, showing that lineage balance is altered in LT-HSCs, selectively enhancing their clonogenic efficiency *in vitro*.

I contributed to this broader investigation into the role of *DEGS1* and sphingolipid modulation in HSC function by quantifying *DEGS1* expression in highly purified HSPC subsets using FACS sorting of LT-HSCs, ST-HSCs and GMPs. Its expression was measured by Quantitative polymerase chain reaction (qPCR) in these HSPCs separately at baseline (0 hours) and 6 hours after treatment with cytokines in culture to stimulate HSPC activation. Expression of *DEGS1* is significantly increased in all 3 HSPC subpopulations at 6 hours, suggesting its upregulation is an early event in HSPC activation/quiescence exit (Fig 3.3-4A). *DEGS1* expression is also higher in LT-HSCs compared with GMPs ( $p=0.07$ ) after 6 hours in culture. Furthermore, expression of *DEGS1* is significantly upregulated upon LT-HSC exit from quiescence as assessed by scRNAseq of LT-HSCs in culture compared to baseline (Fig 3.3-4B). This quantifies the differential expression of *DEGS1* within HSPC subsets upon quiescence exit, as well as between different HSPC subsets within the HSPC hierarchy. Upregulation of *DEGS1* on quiescence exit confirms that this gene is required for HSC activation and inhibition of *DEGS1* leads to increased HSC stemness. The increased expression in LT-HSCs compared to GMPs after 6 hours shows the critical role *DEGS1* specifically plays at the stem cell level.



**Fig 3.3-4 *DEGS1* expression in HSC/MPPs and GMPs.** (A) *DEGS1* expression in LT-HSCs, ST-HSCs and GMPs after 6 hours *in vitro* compared with baseline (0 hours), 20 cells per condition. P-values represent two-tailed t-tests comparing means of technical replicates within each biological triplicate, box plots represent median and interquartile range. (B) Violin plot showing expression of *DEGS1* measured by scRNAseq in single LT-HSCs at defined timepoints during the HSC quiescence exit process. Violin plot generated by Kendig Sham.

### 3.3.1.2 Genes downregulated on quiescence exit

- d) ***CLK1*** – CDC-like kinase (CLK) 1, a dual specificity protein kinase, was found to be downregulated with HSC activation in our data set (Fig 3.3-3D). CLKs and serine arginine protein kinases (SRPKs) are components of the splicing machinery and are important for exon selection. They are important

potential therapeutic targets as inhibition of the CLK family of genes leads to depletion of cancer related proteins<sup>222</sup>.

- e) **MAFF** – encodes for the transcription factor protein MAFF and is downregulated upon HSC quiescence exit. It is associated with fibrosarcoma and autosomal dominant adult-onset proximal spinal muscular atrophy. It belongs to a family of transcription factors that are crucial regulators of mammalian gene expression<sup>223</sup>.
- f) **DUSP1** – dual specificity phosphatase 1 is a protein encoding gene that is significantly downregulated on HSC quiescence exit. It is related to the ERK signalling pathway and is associated with ovarian cancer. It is recognised as having a central role in the resolution of inflammation<sup>224</sup>.

### **3.3.2 Targeting genes with altered expression on quiescence exit**

#### **3.3.2.1 PRMT5**

*PRMT5* is a transcription regulation gene that was observed to have significantly upregulated expression upon stem cell activation. Thus, I hypothesise that inhibition of *PRMT5* at the HSPC level may contribute to differentiation and proliferation.

*PRMT5* can be pharmacologically inhibited by a number of small molecule inhibitors. I used the following inhibitors to assess the functional role of *PRMT5* on HSPC differentiation.

1. EPZ015666: potent and selective small molecule *PRMT5* inhibitor, with a recommended dose of 5µM (SelleckChem)
2. GSK3203591/EPZ015866: a potent, selective, reversible inhibitor of *PRMT5* (GSK)
3. CMP5: cell permeable carbazole based compound, highly selective inhibitor of *PRMT5* (Merck). This inhibitor has been shown to induce lymphoma cell death in a dose dependent manner, selectively over normal resting B cells<sup>225</sup>.

##### **3.3.2.1.1 Inhibition of PRMT5 in hCB HSPCs results in reduced cell proliferation**

To evaluate the effects of *PRMT5* inhibition on hCB HSPC proliferation, CD34+ cells were cultured in MEM media (Table 2.3-1), allowing differentiation to all major blood lineages<sup>50</sup>. *PRMT5* inhibition was performed with increasing doses of EPZ015666 (ranging from 0.025µM to 5µM) and assessed by flow cytometry at 24 hours, 72 hours, 7 days and 14 days. LT-HSCs, ST-HSCs and GMPs were also sorted using previously specified cell surface markers<sup>9</sup> and separately evaluated at 5 days of inhibition with low (0.2µM) and high (2.5µM) doses of EPZ015666 compared with control in MEM. Total cell count decreased with increasing doses of EPZ015666 after 7 and 14 days of culture (p<0.001 at all dose levels) (Fig 3.3-5A-E). This effect was not observed at the early 24h timepoint,

with only a modest reduction in cell counts after 72h at the higher doses, 1 $\mu$ M and 2.5 $\mu$ M EPZ015666. When separately assessed, LT-HSCs, ST-HSCs and GMPs all had reduced cell counts *in vitro* after 5 days, which reached statistical significance at the higher dose of EPZ015666 (Fig 3.3-5C). Hence, *PRMT5* inhibition with EPZ015666 results in dose dependent reduction in cell numbers in hCB CD34+ cells and HSC/MPPs and GMPs over 5, 7 and 14 days.

To confirm this finding, *PRMT5* inhibition with 2 other independent inhibitors, GSK3023591 and CMP5 was performed on hCB CD34+ cells. Cells treated with increasing doses of GSK3023591 were evaluated at 7 days by flow cytometry, at which timepoint additional media was added to the cells for a further 7 days (total 14 days) of culture. Conditions were separated into further addition of GSK3023591 or not to cells after 7 days *in vitro* to understand whether the effect of *PRMT5* inhibition is reversible. Cells treated with GSK3023591 were then evaluated after total 14 days in culture, while CD34+ cells treated with CMP5 were only assessed after 7 days in culture. Both inhibitors replicated the dose dependent reduction in cell numbers in culture after 7 and 14 days compared to control for GSK3023591 (Fig 3.3-5B) and after 7 days compared to control for CMP5 (Fig 3.3-5C). Cell numbers were similarly reduced at day 14 regardless of whether the inhibitor was added on day 7 of culture. This indicates that the effect of *PRMT5* inhibition on cell proliferation is not reversible in hCB CD34+ cells treated with GSK3023591.

The effect of *PRMT5* inhibition was then separately assessed on LT-HSCs, ST-HSCs and GMPs using MC CFC assays to ascertain whether the effect was restricted to any specific HSPC subtype. This was evaluated at 5 $\mu$ M EPZ015666 in biological and technical duplicates, and at escalating doses of EPZ015666.

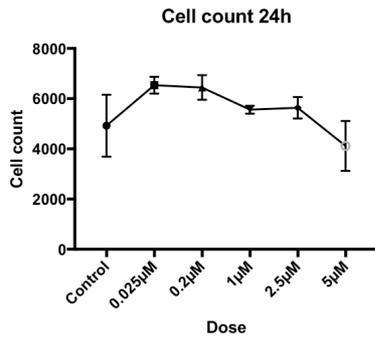
When LT-HSCs, ST-HSCs and GMPs were cultured in MC CFC assays with DMSO control and the *PRMT5* inhibitor EPZ015666 at the recommended manufacturer's dose (5 $\mu$ M) for 14 days, there was significant paucity of colony formation in all cell types, with very limited differentiation capacity. Significant reductions were seen in overall colony formation in treated HSPCs compared to controls with very few or no colonies visualised using the StemVision in cells treated with the inhibitor EPZ015666 (Fig 3.3-6A & B). This effect is seen across all seeded cell types and colonies yielded, with the decreased total number of colonies in treated ST-HSCs and GMPs compared to controls being statistically significant on two-tailed t-test ( $p=0.03$  and  $0.02$  respectively). Colony size was also significantly reduced as measured by cells per colony assessed on FACS phenotyping (Fig 3.3-6B). There were significantly fewer granulocytic and monocytic colonies produced by ST-HSCs ( $p=0.0009$  and  $0.002$  respectively), and GMPs ( $p=0.0006$  and  $0.01$  respectively, data not shown).

Dose dependent reduction in colony formation of all hCB HSPC subsets was also seen in response to *PRMT5* inhibition by EPZ015666 as measured by the number of colonies produced per 100 cells seeded and size of the colonies as measured by cell count on flow cytometry / number of colonies counted per condition (Fig 3.3-6C). The effect was most pronounced in LT-HSCs ( $p=0.02$ ,  $p=0.03$  and  $p=0.06$  for 0.2 $\mu$ M, 1 $\mu$ M and 2.5 $\mu$ M EPZ015666 respectively), while ST-HSCs did not reach statistical significance. The effect was seen even at low doses of the inhibitor. However, the relative proportion of cell types upon differentiation remained consistent between

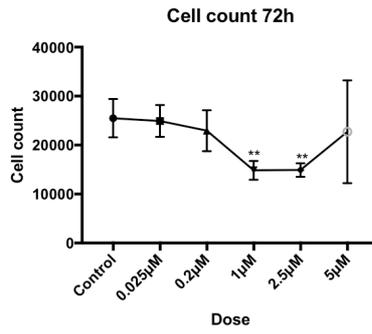
doses, such that the proportion of differentiated erythroid, granulocytic and monocytic cells within each dose concentration were similar within each cell type cultured.

Taken together, pharmacological inhibition of *PRMT5* in hCB HSPCs using 3 independent small molecule inhibitors results in dose dependent reduction in cell counts, an effect that is not reversible when further doses of 1 inhibitor (GSK3023591) is not added again at a mid-time point. The effect is observed at various time intervals from 5 days to 14 days.

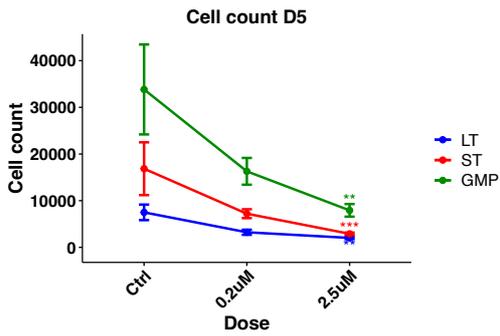
**A** EPZ015666



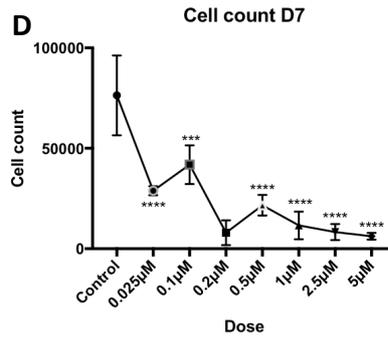
**B**



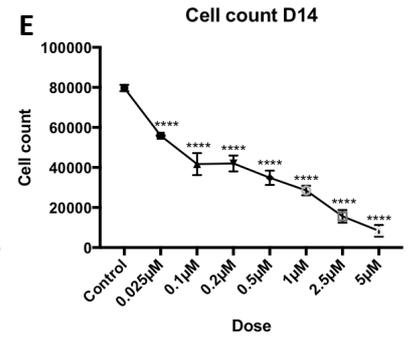
**C**



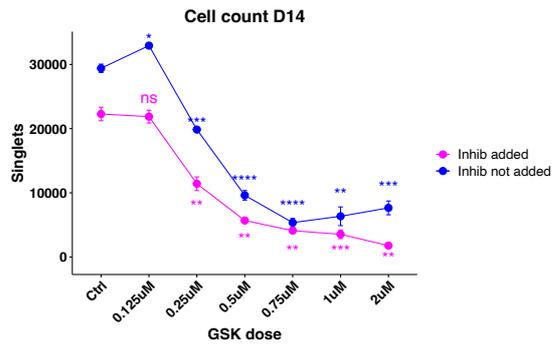
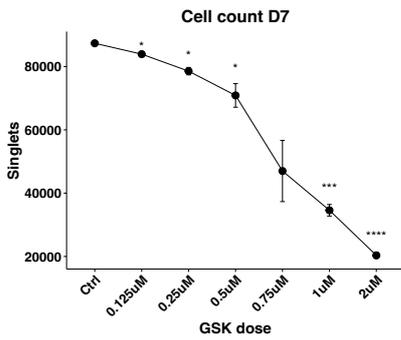
**D**



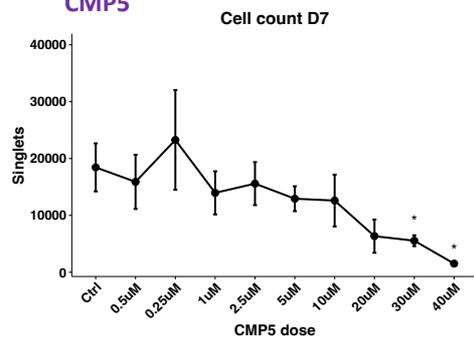
**E**



**F** GSK3203591



**G** CMP5



**Fig 3.3-5 Effect of dose dependent *PRMT5* inhibition on proliferation capacity of human cord blood (hCB) CD34+ HSPCs.** Dose dependent viable cell count from flow cytometry analysis of CD34+ HSPCs after (A) 24 hours, (B) 72 hours (comparing means of technical triplicates of each biological duplicate), (D) 7 and (E) 14 days (comparing means of 6 technical replicates from 4 biological replicates) and (C) LT-HSCs, ST-HSCs and GMPs after 5 days (comparing means of technical triplicates from 4 biological replicates), treated with escalating doses of EPZ015666 in MEM (F) CD34+ HSPCs treated with GSK3023591 for 7 and 14 days in MEM. Inhibitor was either added again or not after 7 days and assessed separately to assess reversibility of *PRMT5* inhibition after total 14 days in culture (comparing means from technical triplicates from a single biological replicate). (G) CD34+ HSPCs treated with CMP5 after 7 days in MEM; comparing single measurements from biological duplicates. Error bars represent mean +/- SD. P-values represent one way ANOVA, and in all figures p-values are represented as \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

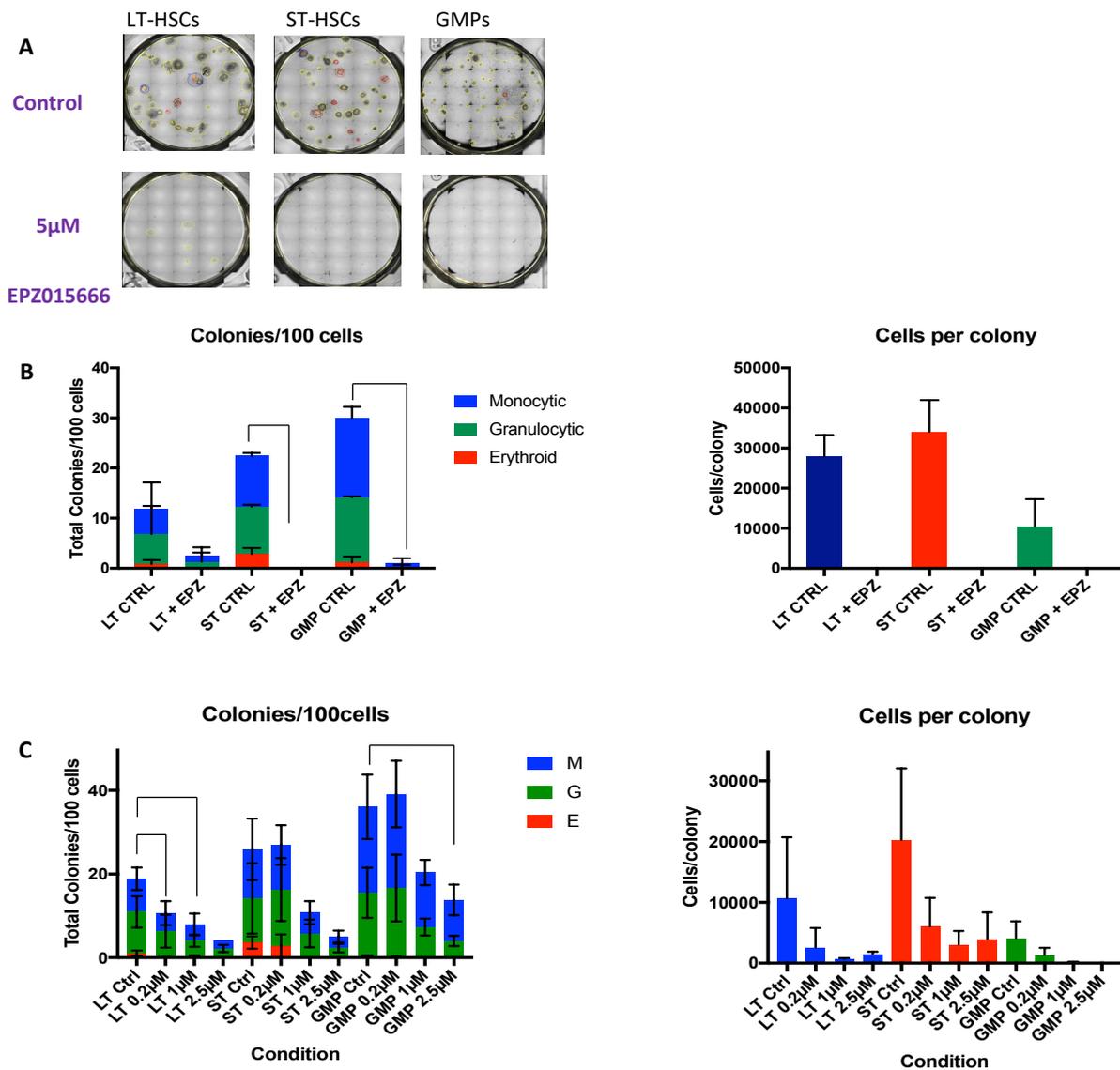
### *3.3.2.1.2 PRMT5 inhibition leads to a relative preservation of immature cells*

To assess the effects of *PRMT5* inhibition on differentiation, data from the above experiments was used to analyse HSPC markers on flow cytometry. Higher relative proportions of CD34+ cells were seen in LT-HSCs and ST-HSCs after 5 days in culture with low (0.2µM) and higher (2.5µM) doses of EPZ015666 (Fig 3.3-7A). There was also an observed increase in CD71+ erythroid progenitors in the ST-HSCs after 5 days in MEM at low and high doses of EPZ015666 (Fig 3.3-7B).

CD34+ cells also proportionally increased when hCB CD34+ HSPCs were cultured for 7 and 14 days with  $\geq 1\mu\text{M}$  EPZ015666 (Fig 3.3-7C). The effect was more marked in the CD34+/CD33+ population, particularly at the higher doses  $\geq 1\mu\text{M}$  at both time points (p<0.05). The less differentiated CD34+/CD33- population also increased with increasing *PRMT5* inhibition, which was more evident after 14 days in culture, although this population represented <5% of all cells in the colonies (p<0.01 at all doses  $\geq 0.2\mu\text{M}$  EPZ015666). This phenotype of a relative preservation of the CD34+ cells was evident as early as 72 hours in culture, but not at 24 hours (Fig 3.3-7D). Thus, HSPCs retaining cell surface markers of immaturity represented a higher relative proportion in cultures treated with *PRMT5* inhibitors than controls, especially at higher treatment doses in MEM.

The relative preservation of the less mature CD34+/CD33+ compartment was validated using GSK3023591 after 7 and 14 days in MEM, with the proportion of CD34+/CD33+ cells relatively increasing with increasing doses of the inhibitor after 7 days (Fig 3.3-7E).

In summary, while pharmacological *PRMT5* inhibition with 3 independent small molecules results in reduced overall proliferation of hCB HSPCs, the less mature CD34+/CD33+ compartment is relatively preserved in comparison to all cells in culture.



**Fig 3.3-6 Pharmacological *PRMT5* inhibition with EPZ015666 of human cord blood (hCB) LT-HSCs, ST-HSCs and GMPs in methylcellulose (MC) colony forming cell (CFC) assays.** (A) Representative MC CFC assays visualised using the StemVision after 14 days, treated with 5µM EPZ015666 compared with control. Yellow circles represent granulocytic (G), monocyte (M) or G/M colonies, red circles represent erythroid (E) colonies and blue circles represent mixed colonies. (B-C) Colonies visualised on MC CFC assays per 100 cells seeded and average cells counted by flow cytometry per counted colony after 14 days MC CFC assay, in LT-HSCs, ST-HSCs and GMPs treated with (B) 5µM EPZ015666 comparing means of technical duplicates from biological duplicates or (C) escalating doses of EPZ015666, comparing means of technical duplicates from biological triplicates. Error bars represent mean +/- SD. P-values

represent two-tailed t-tests, and in all figures p-values are represented as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

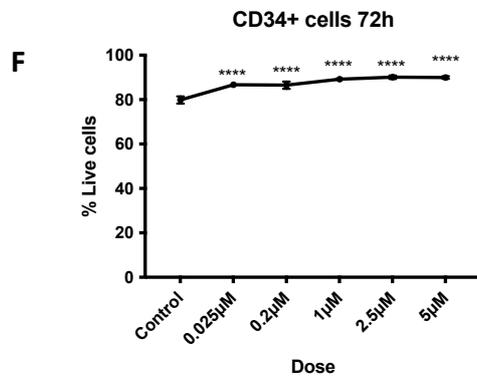
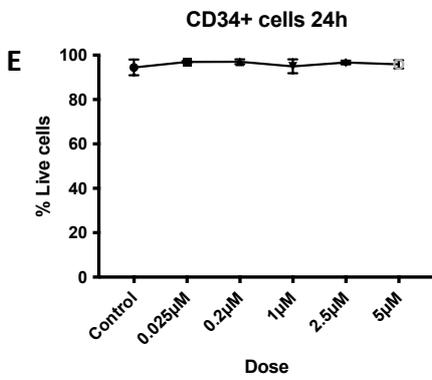
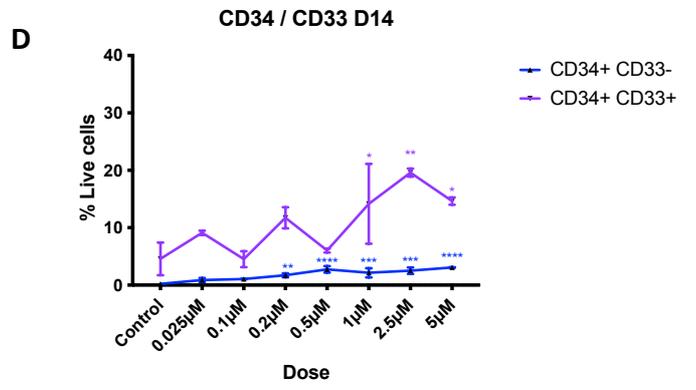
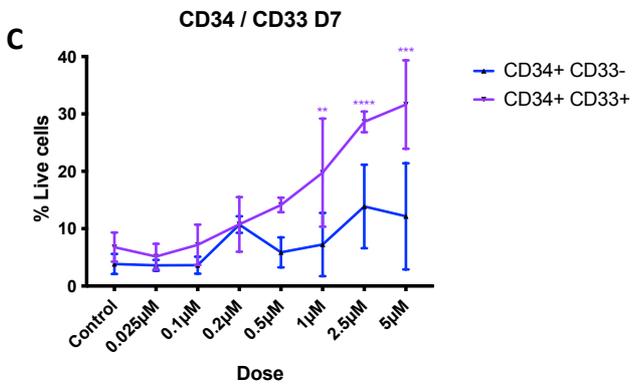
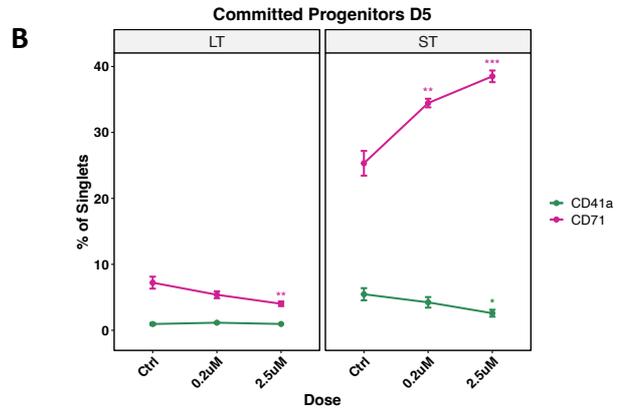
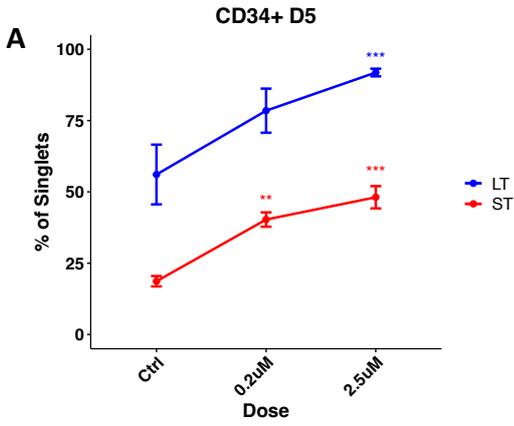
### 3.3.2.1.3 *PRMT5 inhibition reduces mature cell differentiation*

The effects of *PRMT5* inhibition on HSPC differentiation to mature blood cells using flow cytometry was also assessed using data from the above experiments. There was a reduction in mature monocytic cells (CD33+/CD14+) after 7 and 14 days in culture for hCB CD34+ HSPCs treated with EPZ015666 compared to controls ( $p < 0.05$  after 7 days for doses between 0.2 $\mu$ M – 2.5 $\mu$ M, inclusive; and for all doses  $> 0.1\mu$ M after 14 days, Fig 3.3-8A). The reduction was more significant after 14 days. Few erythroid (GlyA+) and granulocytic cells (CD33+/CD15+) were produced at both time points, though a significant reduction in erythroid cells compared to control was observed at all dose levels after 7 days in culture ( $p < 0.05$ ). A statistically significant increase in erythroid and granulocytic cells was noted only at the highest dose after 14 days, although the absolute percentages were low ( $p = 0.04$ , 1.7% erythrocytes in control c.f. 8.3% at 5 $\mu$ M EPZ015666;  $p = 0.003$ , 0.08% granulocytes in control c.f. 9.8% at 5 $\mu$ M EPZ015666).

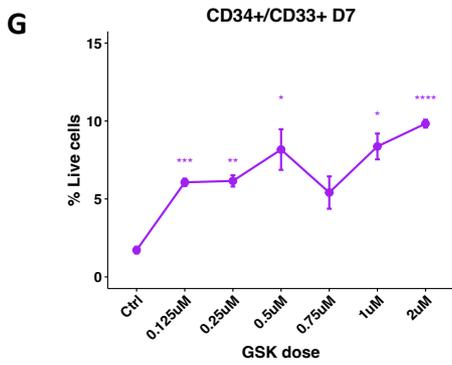
Reduced differentiation to mature blood cells was validated using 2 other independent small molecule inhibitors of *PRMT5*. Effects of CMP5 were analysed after 7 days in MEM, while effects of inhibition with GSK3023591 were analysed after 7 days, followed by a separate analysis of washout vs treated cells after total 14 days in culture, whereby at the 7 day timepoint further GSK3023591 was either added or not to the cells in MEM. Proportions of mature monocytes decreased with increasing doses of GSK3023591 after 7 days in MEM, while the low proportions of granulocytes and erythroid cells were unaffected. The mature erythroid population decreased after 14 days with or without inhibitor addition on day 7. Interestingly, the granulocyte (CD15+/CD66- and CD15+/CD66b+) populations increased after 14 days in culture with or without inhibitor addition on day 7. Proportions of mature monocytes were not significantly different after 14 days with this inhibitor, although the absolute proportion of monocytes was small at all doses including in controls ( $< 3\%$ , Fig 3.3-8B). This differentiation phenotype was largely replicated using CMP5. A relative reduction in monocyte and erythrocyte maturation was observed after 7 days in culture, particularly at higher doses (Fig 3.3-8C).

Overall, differentiation to mature blood cells was reduced by *PRMT5* inhibition using 3 independent small molecules, with the effect being most consistent in mature monocyte differentiation.

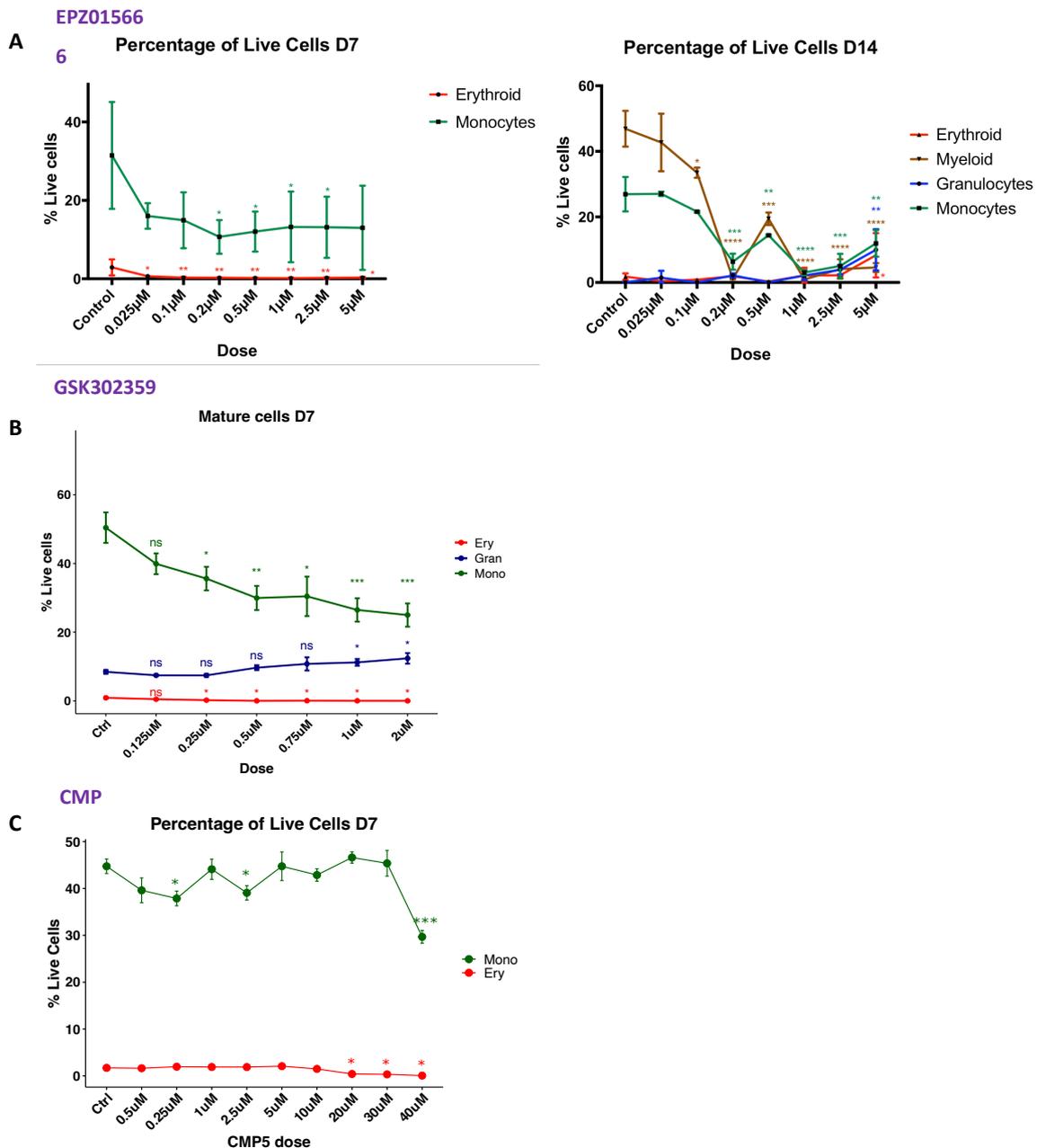
**EPZ015666**



**GSK3203591**



**Fig 3.3-7 Effects of *PRMT5* inhibition on progenitor populations during *in vitro* culture in MEM of hCB CD34+ HSPCs.** (A) Proportion of CD34+ cells as a percentage of singlets (see Methods Fig 2.3-5) seen in LT and ST-HSCs cultured with 2 doses of EPZ015666 after 5 days analysed by flow cytometry; (B) proportion of CD71+ erythroid committed progenitors as a percentage of singlets in LT-HSCs and ST-HSCs after 5 days in MEM; comparing means of 6 technical replicates from 4 biological replicates. CD33+/CD34+ and CD33-/CD34+ cell proportions as a percentage of live cells after inhibition with EPZ015666 after (C) 7 days and (D) 14 days in MEM, comparing means of technical triplicates from 4 biological replicates; and after (G) 24 hours and (F) 72 hours in MEM. (E) CD34+/CD33+ cell proportions as a percentage of live cells with *PRMT5* inhibition with GSK32035921 after 7 days in MEM, comparing means of technical triplicates from biological triplicates. Error bars represent mean +/- SD. P-values represent one way ANOVA, and in all figures p-values are represented as \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.



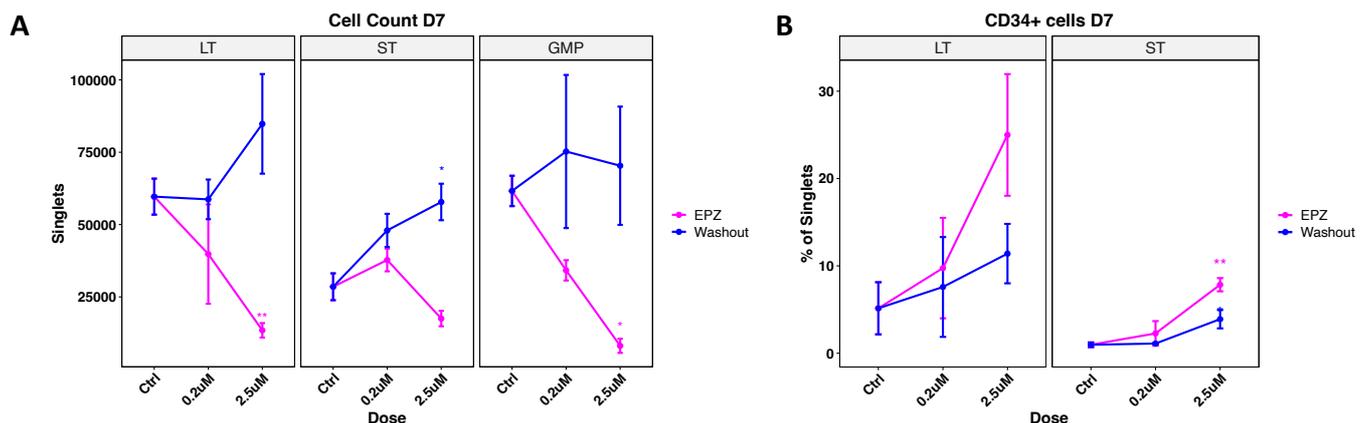
**Fig 3.3-8 Effect of *PRMT5* inhibition on hCB CD34+ HSPC differentiation to mature blood cells.** (A) Percentages of mature blood cell populations (Erythroid = GlyA+, Myeloid = CD33+/CD14+ or CD33+/CD15+, Monocytes = CD14+, Granulocytes = CD15+) after *PRMT5* inhibition with increasing doses of EPZ015666 after 7 and 14 days in MEM, comparing means of technical triplicates from 4 biological replicates. (B) Percentages of mature blood cells (Mono = CD14+, Gran = CD15+, Ery = GlyA+) after *PRMT5* inhibition with increasing doses of GSK3203591 after 7 and 14 days in MEM, comparing means of technical triplicates from biological triplicates. (C) Percentages of Mono (CD14+) and Ery (GlyA+) cells after *PRMT5* inhibition with CMP5 after 7 days in MEM, comparing single measurements

from biological duplicates. Error bars represent mean  $\pm$  SD. P-values represent one way ANOVA, and in all figures p-values are represented as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

### 3.3.2.1.4 *PRMT5* inhibition phenotype is reversible with EPZ015666

To assess reversibility of *PRMT5* inhibition with EPZ015666, LT-HSCs, ST-HSCs and GMPs were cultured for 5 days in MEM with increasing doses of *PRMT5* inhibitor EPZ015666, then for a further 7 days with and without the addition of EPZ015666 in MEM. The effect on proliferation was reversible as seen by decreased cell count after a further 7 days in the inhibitor treated cells only, but not the washout group (no EPZ015666 added post day 5; Fig 3.3-9A). The relative preservation of the CD34+ compartment is also a reversible effect as it is only seen in the treated LT-HSCs and ST-HSCs, not in the washout cells after 7 days in culture (Fig 3.3-9B).

Hence, the effect of *PRMT5* inhibition with EPZ015666 on hCB HSPC proliferation and relative preservation of the less mature CD34+ compartment is reversible when treated for 5 days.



**Fig 3.3-9 Effects of *PRMT5* inhibition with EPZ015666 on hCB HSPCs in MEM following washout of inhibitor.** hCB HSPCs were treated in MEM for 5 days with EPZ015666, then either washed out or treated with EPZ015666 for a further 14 days and analysed after 7 days and 14 days in MEM. (A) Cell proliferation in treated and washout cells after 7 days in MEM. (B) Proportions of CD34+ cells in treated and washout cells after 7 days in MEM. All results comparing means of 6 technical replicates from 4 biological replicates, p-values represent one way ANOVA, and in all figures p-values are represented as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

### 3.3.2.1.5 *PRMT5* inhibition phenotype is not related to apoptosis

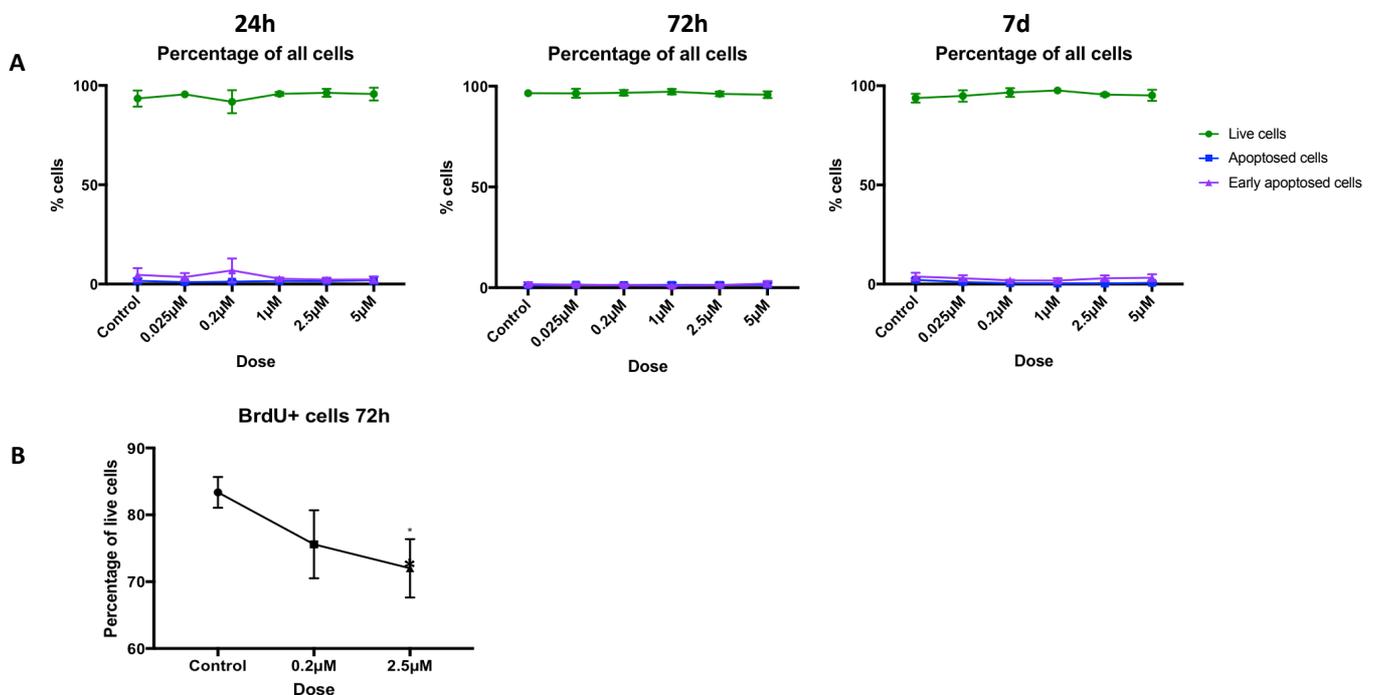
To investigate whether the observed reduction in cell numbers with *PRMT5* inhibition had any contribution from cellular apoptosis, hCB CD34+ HSPCs were treated with increasing doses of EPZ015666 for 24 hours, 72 hours and 7 days followed by fixation and apoptosis staining with Annexin-V and 7-AAD by flow cytometry.

Despite a total reduction in cell count, particularly at the higher doses as previously observed, no increase in apoptosis was observed at any of the doses tested at any time point. This was true for all cells (Fig 3.3-10A) and separately for cells within the CD34+ compartment only (data not shown). Hence the reduction in cell counts resulting from *PRMT5* inhibition cannot be attributed to apoptosis.

### 3.3.2.1.6 High doses of *PRMT5* inhibitor led to fewer actively cycling cells

Next, I tested if reduced cell counts observed with *PRMT5* inhibition *in vitro* are due to decreased cell proliferation. To assess whether *PRMT5* inhibition affects DNA synthesis as a marker of cell cycle activity, BrdU incorporation assay was performed on hCB C34+ HSPCs treated with increasing doses of EPZ015666 after 72 hours in MEM. BrdU was added to cells 8h hours prior to analysis and cells were fixed, permeabilised and stained for BrdU. There was a significant decrease in BrdU+ cells with 2.5 $\mu$ M EPZ015666 compared to control ( $p = 0.02$ , Fig 3.3-10B). Thus, reduced cell cycling contributes to the reduced growth of hCB HSPCs following *PRMT5* inhibition.

Taken together, the above data indicates that *PRMT5* inhibition induces decreased proliferation of hCB HSPCs, with relative preservation of the less mature progenitor compartment and decreased differentiation towards mature blood cells. HSPCs most likely undergo a differentiation block rather than cell death as a result of *PRMT5* inhibition in a dose dependent manner.



**Fig 3.3-10 Mechanistic insights into the effects of *PRMT5* inhibition on hCB HSPCs.** (A) Apoptosis assay analysis of hCB HSPCs treated with EPZ015666 using Annexin-V (AnnV) and 7-AAD. Percentage of Live (AnnV-/7AAD-), Apoptosed (AnnV+/7AAD+) and Early apoptosed (7AAD+/AnnV-) cells after 24 hours, 72 hours and 7 days in MEM. (B) Proportion of BrdU+ cells as a percentage of live cells after 72h in MEM, treated with EPZ015666. All results comparing means of technical triplicates from biological triplicates. Error bars represent mean +/- SD. P-values represent one way ANOVA, and in all figures p-values are represented as \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

### *3.3.2.1.7 Targeting *PRMT5* in Tex cell line*

Tex cells are a human leukemic cell line derived from normal CB haematopoietic cells transduced with TLS-ERG<sup>150</sup>. These cells have self-renewal properties similar to human HSCs, share similar cytokine dependencies but have limited differentiation potential, hence represent a surrogate model for hHSC self-renewal.

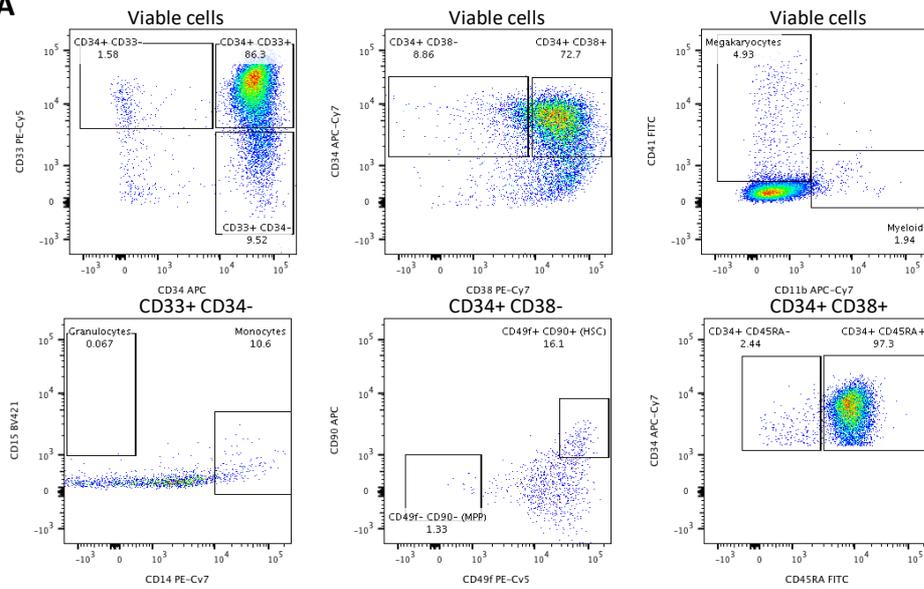
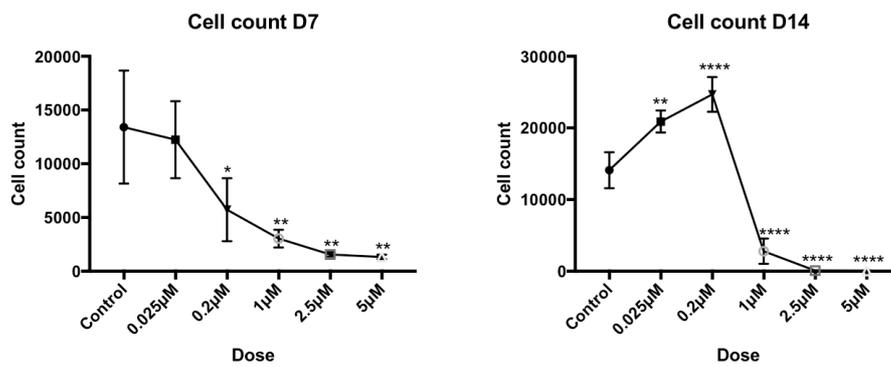
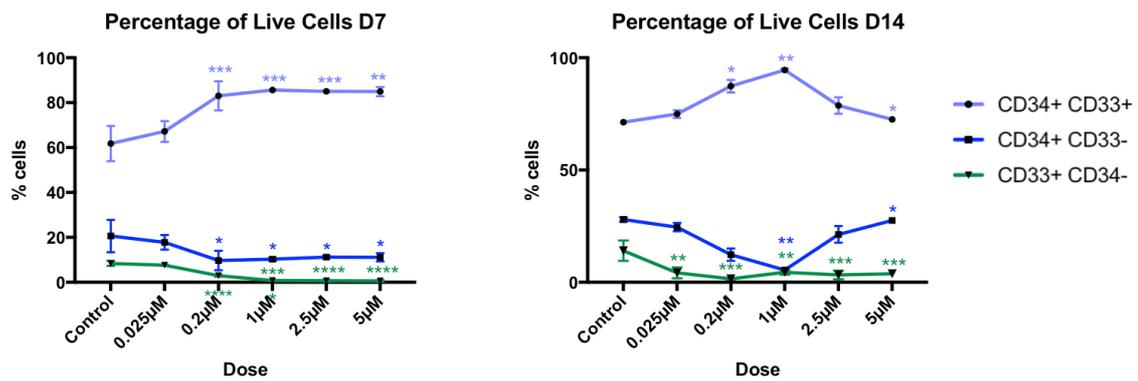
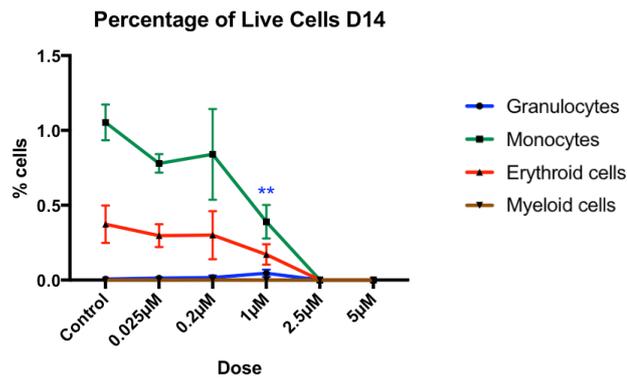
At baseline homeostatic conditions, Tex cells predominantly express CD34 and CD33 on their cell surface (87%), with a small (10%) CD33+/CD34- population (Fig 3.3-11A). They mostly reside in an intermediate differentiation stage between HSC/MPPs and progenitors, with 70% of cells expressing both CD34 and CD38 on their surface. Approximately 15% reside in the HSC pool (CD34+/CD38-/CD49f+/CD90+), consistent with hCB. There are very few differentiated cells including monocytes (CD14+), granulocytes (CD15+), megakaryocytes (CD41+) and myeloid cells (CD11b+). After 7-14 days in culture, they still have very few differentiated cells (<1.5%), which are similar in proportion to these baseline results (data not shown).

Tex cells were cultured in liquid medium (IMDM with 15% FBS, antibiotics and cytokines, see Methods **Error! Reference source not found.**) for 7 and 14 days with DMSO control and escalating doses of EPZ015666 to assess whether the effects of *PRMT5* inhibition on proliferation are equivalent to those seen in hCB CD34+ cells. Tex cells could represent a valuable alternative to donated hCB HSPCs to evaluate the molecular mechanisms of *PRMT5* inhibition with biochemical techniques requiring large number of cells.

Inhibition of *PRMT5* led to a dose dependent reduction in total cell count after 7 and 14 days in culture of Tex cells (Fig 3.3-11B; p<0.01 for all doses  $\geq 1\mu\text{M}$  EPZ015666). The increase in cell count with 0.025 $\mu\text{M}$  and 0.2 $\mu\text{M}$  EPZ015666 after 14 days in culture are most likely a result of drug exhaustion at these lower doses after 2 weeks as the inhibitor was not replaced when half of each well was harvested at 7 days. The effect was maintained at the moderate and higher doses after 14 days without replacement.

Consistent with the results of hCB CD34+ cells in culture, the relative proportion of CD34+ cells increased with increasing dose of EPZ015666 in Tex cells after 7 and 14 days of culture (Fig 3.3-11C). This was mainly observed in the CD34+/CD33+ population (p<0.01 for doses  $\geq 0.2\mu\text{M}$  after 7 days, p<0.05 for doses  $\geq 0.2\mu\text{M}$  after 14 days in culture). *PRMT5* inhibition did not induce any Tex cell differentiation (Fig 3.3-11D).

I conclude that Tex cells could be a valuable resource for evaluating the molecular mechanisms responsible for the effect of *PRMT5* inhibition on cell proliferation and maintenance of the progenitor compartment, though they cannot be used to evaluate the effects on HSPC differentiation.

**A****B****C****D**

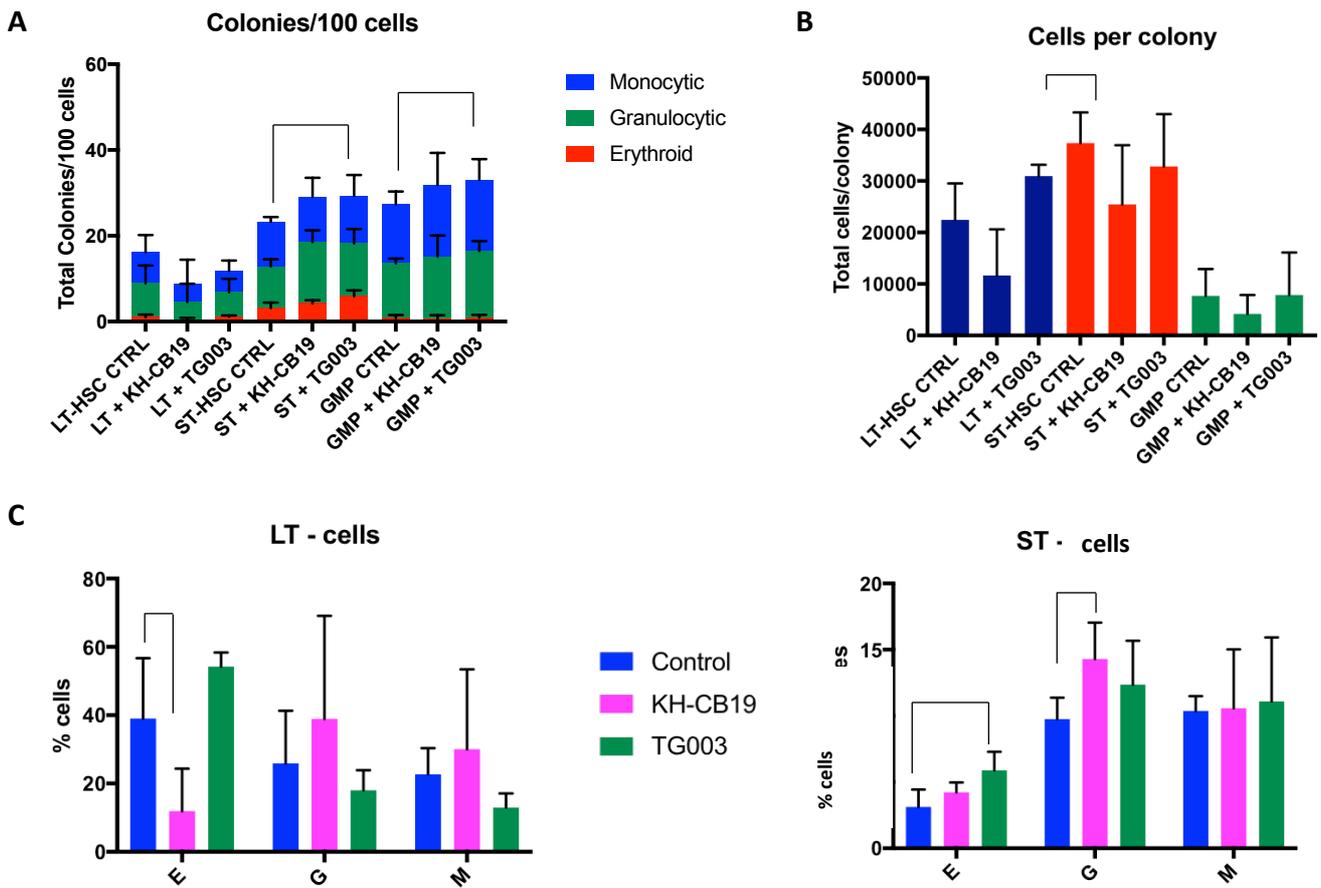
**Fig 3.3-11 Validating the effects of PRMT5 inhibition in Tex cell line.** (A) Baseline characterisation of surface markers expressed on Tex cells. (B) Dose dependent viable cell count after 7 and 14 days in liquid culture of Tex cells with DMSO control and escalating doses of EPZ015666. (C) Percentage of CD34+/CD33+ cells after 7 and 14 days and (D) percentage of mature cells after 14 days in culture. All results comparing means technical triplicates from triplicates of the experiment. Error bars represent mean +/- SD. P-values represent one way ANOVA, and in all figures p-values are represented as \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

### 3.3.2.2 CLK1

Here I assessed the activity of the spliceosome modifier gene *CLK1* in human HSCs by pharmacological alteration with two different non-selective small molecule inhibitors of *CLK1*: TG003 (*CLK1*, *CLK2* and *CLK4* inhibitor) and KH-CB19 (*CLK1* and *CLK4* inhibitor). Doses of 10µM TG003 and 5µM KH-CB19 were based on manufacturers guidelines and published literature. These are doses at which mammalian cells have been shown to have suppressed CLK1-mediated phosphorylation *in vitro*<sup>226</sup>. 500 LT-HSCs, 350 ST-HSCs and 350 GMPs were sorted from hCB CD34+ cells and cultured for 14 days in methylcellulose (MC) with added cytokines (see Methods, Table 2.3-2) with each of the above mentioned inhibitors and under control conditions with DMSO only. The experiment had 5 biological replicates, each with technical duplicates.

Overall, variable responses to *CLK1* inhibition were observed within each seeded cell type. The overall number and types of differentiated erythroid, monocytic or granulocytic colonies were similar between control and the 2 inhibitors within each HSPC cell type (**Fig 3.3-12A**). GMPs treated with TG003 produced more total colonies than control on average (21.4 vs 18 colonies, p=0.02), while ST-HSCs treated with TG003 produced more erythroid colonies on average than control (5.8 vs 3.1 colonies, p = 0.04). The size of differentiated colonies produced by HSPCs treated with either of the 2 *CLK1* inhibitors compared to controls were similar when photographed in MC culture (data not shown), with only one significant difference as measured by total number of cells per colony (as counted visually from MC CFC assays) assessed on FACS phenotyping (**Fig 3.3-12B**): ST-HSCs treated with KH-CB19 produced smaller colonies overall than controls (19,000 cells per colony vs 34,715, p = 0.02).

Proportions of erythroid (GlyA+), monocytic (CD33+/CD14+) and granulocytic (CD33+/CD15+) cells in each colony were analysed using flow cytometry after harvesting all cells within each condition (**Fig 3.3-12C**). ST-HSCs treated with *CLK1* inhibitors did produce more erythroid and granulocytic cells, but in each case only one inhibitor showed a statistically significant difference: increased erythroid cells with TG003 (p=0.01) and increased granulocytic cells with KH-CB19 (p=0.03). LT-HSCs produced fewer erythroid cells with KH-CB19 (p=0.04) but increased erythroid cells with TG003 (p=0.15). Given the inconsistent phenotypes observed in LT-HSC, ST-HSC and GMP following pharmacological inhibition of *CLK1* by 2 independent small molecule inhibitors, I conclude that *CLK1* does not have a significant role in the proliferation and differentiation of hCB HSPCs *in vitro*.



**Fig 3.3-12 Pharmacological inhibition of *CLK1* with 2 independent small molecule inhibitors (5µM KH-CB19 and 10µM TG003) has no effect on colony forming cell (CFC) assays of human cord blood (hCB) LT-HSCs, ST-HSCs and GMPs after 14 days.** (A) Colonies visualised on MC CFC assays per 100 cells seeded and (B) average total cell count by FACS analysis per colony counted in (A). (C) Percentage of live cells represented by E (GlyA+), G (CD15+) and M (CD14+) cells in LT-HSCs and ST-HSCs. All graphs derived from 5 biological replicates, comparing means of technical duplicates. Error bars represent mean +/- SD. P values represent two-tailed t-tests, and in all figures p-values are represented as \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

### 3.4 Conclusions

There are thousands of genes within hCB LT-HSCs that have differential expression between their quiescent and active states, with a small subset being associated with haematological malignancies. Genes that significantly alter their expression as HSCs transition through the quiescence exit process are likely to play an important role in their cell fate decisions and deserve further evaluation at the functional level. These include spliceosome genes, methyltransferases, and sphingolipid modulators, including DEGS1, which is upregulated early (after 6 hours in culture with cytokines) in LT-HSCs, ST-HSCs and GMPs as seen by qPCR and RNAseq.

### **3.4.1 PRMT5**

*PRMT5* is a methyltransferase gene that is significantly upregulated upon HSC quiescence exit. Pharmacological inhibition of this gene in hCB HSPCs results in decreased cell growth in a dose dependent manner, unrelated to apoptosis. There is also a relative increase in more immature stem and progenitor cells (CD34+) with *PRMT5* inhibition and a decrease in mature cell differentiation in a dose dependent manner. These results combined with reduced cell cycle activity in hCB progenitors treated with *PRMT5* inhibitors suggest that *PRMT5* inhibition results in reduced overall cell proliferation with a likely differentiation block at the progenitor cell level. Several aspects of this phenotype were confirmed by 3 independent small molecule inhibitors of *PRMT5* (EPZ015666, EPZ015866 and CMP5).

### **3.4.2 CLK1**

*CLK1*, a dual specificity protein kinase that plays an important role in the spliceosome, is significantly downregulated upon HSC quiescence exit as seen on single cell and bulk RNAseq analysis of LT-HSCs. However, pharmacological inhibition of this gene has no reproducible effect on hCB HSPC differentiation capacity as seen in MC CFC assays with 2 different non-selective *CLK1* inhibitors. Hence at these doses, *CLK1* inhibition does not appear to have a significant effect on the proliferation and differentiation capacity of HSCs and progenitors.

This gene forms part of the splicing machinery, perturbations of which have been observed in cancer, and in particular have been identified as a hallmark of some haematological malignancies<sup>211</sup>. Splicing factor mutations have been noted in the founding clone of myeloid malignancies, suggesting their role in disease initiation<sup>227</sup>. This was the first study exploring the role of *CLK1* in normal, healthy hHSCs, not in disease or cell lines.

Although there are some trends in the data arising from *CLK1* inhibition in hHSCs in my study, these should be interpreted with caution as they were usually limited to one of the two *CLK1* inhibitors and not reproducible in both colony formation and flow cytometry analysis in each condition. These results may be explained by the non-selective nature of the small molecule inhibitors used in this assay. TG003 also targets *CLK2* and *CLK4*, while KH-CB19 also targets *CLK4*. Some of the discordant results seen with TG003 and not KH-CB19 may be attributed to *CLK2* inhibition, rather than *CLK1* inhibition.

The above analyses provide insight into the role of the genes identified as being altered during HSC activation and hence hypothesised to play a role in HSC proliferation and cell fate decisions. My data indicates that the methyltransferase *PRMT5* has a role in regulating cell proliferation and differentiation in hCB HSPCs, which is likely through decreasing cell proliferation and maintaining a relative stem-like state.

# Chapter 4 HSC function in clonal haematopoiesis and disease

## 4.1 Introduction

Several genes are recurrently mutated in clonal haematopoiesis (CH), and these are also frequently observed in haematological malignancies<sup>93</sup>. Age-related CH (ARCH) carries an increased risk of haematological malignancy, atherosclerosis, cardiovascular disease<sup>88</sup>, and other non-haematological conditions<sup>99,228</sup>. Although the relative risk of haematological malignancy in individuals with CH is high, the absolute risk is low in mice and humans, with a long latency period<sup>138,229</sup>. Due to the virtue of most ARCH individuals being asymptomatic, the exact epidemiology of such CH mutations is difficult to specify. Establishing CH screening clinics is one strategy currently being used to develop a better understanding of these mutations and their clinical significance in the healthy population. Here I will review the published literature outlining the epidemiological and mechanistic insights of the role of ARCH mutations, in particular *DNMT3A* in haematological and non-haematological conditions.

### 4.1.1 Genetic mutations in ARCH

ARCH related mutations are usually involved in global chromatin changes such as DNA methylation (DNAm), histone modification and chromatin looping. The most commonly mutated genes in CH are proposed to function as epigenetic regulators and are commonly seen in myeloid malignancies<sup>230-232</sup>. *DNMT3A* is the most commonly observed somatic mutation in ARCH, accounting for the vast majority of ARCH cases<sup>88,93</sup> with *ASXL1* and *TET2* also being frequently observed<sup>87,88</sup>. *DNMT3A* is responsible for DNAm. *TET2* is responsible for DNA hydroxymethylation resulting in a hypermethylation phenotype in AML patients with this mutation<sup>80</sup>. *ASXL1* is responsible for histone methylation and histone ubiquitination.

*TET2* and *ASXL1* are the 2<sup>nd</sup> and 3<sup>rd</sup> most common mutations observed in CH following *DNMT3A*, and together account for approximately 20% of ARCH mutations<sup>88,93</sup>. *TET2* mutations are associated with DNAm, and are highly enriched in elderly individuals with age related skewing of blood cells<sup>80</sup>. *TET2* catalyses the hydroxylation of 5-methylcytosines, resulting in loss of DNAm. *ASXL1* is thought to play the role of tumour suppressor gene in myeloid malignancies<sup>232</sup>.

Other less frequent ARCH related mutations include *TP53*, *JAK2*, *SF3B1*, *GNB1*, *CBL*, *SRSF2*, *GNAS*, *PTPN11*, *MYD88*, *IDH2*, *NRAS*, *KRAS* and *BRAF*<sup>88,93</sup> (Fig 1.5-1, Introduction). *IDH1* has also been implicated as a pre-leukemic event as it has been observed in progenitor and mature cell populations of patients with AML in the absence of *NPM1c* mutations<sup>126</sup>. This is important in the clinical context of potentially targeted *IDH1/IDH2* inhibitors being available currently in clinical trials<sup>233</sup>.

These mutations are seen in individuals with completely normal blood counts or with only mild perturbations (e.g. increased red cell distribution width, RDW<sup>234</sup>). In contrast to AML, where patients carry a median of 5 recurrently mutated genes, in the vast majority of ARCH individuals, only 1 mutated gene is present<sup>88</sup>. While these recognised drivers of haematological malignancies are commonly seen in ARCH, activating mutations in more proliferative genes such as *FLT3* and *NPM1* are only observed in leukaemia, and are as such mutually exclusive with ARCH and subclinical clonal expansions<sup>235</sup>.

Although ARCH is an age related entity, it has now become evident that ARCH related mutations are actually acquired early in life and remain latent in a subclinical state for prolonged periods<sup>84</sup>. They are extremely rarely observed in paediatric malignancy<sup>236–238</sup>. Their clonal expansion and clinical significance has a strong correlation with age and earlier acquisition correlates with an increased risk of blood disorders in later life<sup>239</sup>.

### **4.1.2 DNA methyltransferases**

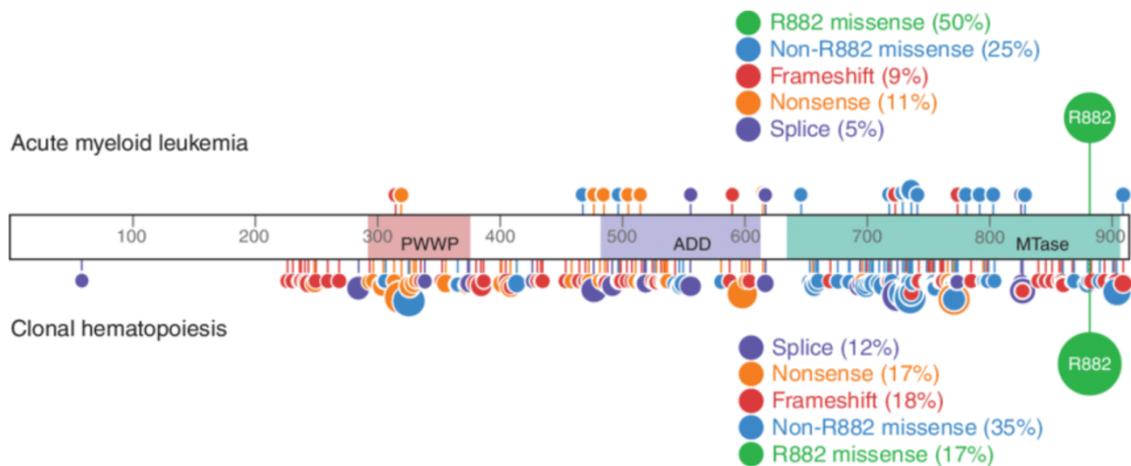
DNAme is an important epigenetic function that undergoes extensive alteration during HSC differentiation. It plays an important role in haematopoiesis, as suggested by the overrepresentation of modifiers of DNAme in haematological malignancies<sup>240</sup>. DNAme also has an important role in several disease states in humans<sup>241</sup>. It is mediated by a group of DNA methyltransferase enzymes including *DNMT1*, *DNMT3A* and *DNMT3B*<sup>242</sup>.

*Dnmt3a*, *Dnmt3b* and *Dnmt1* have been established as important regulators of mouse embryonic stem cell (ESC) development<sup>243</sup>, as their absence leads to embryonic or neonatal death in mice<sup>177</sup>. Loss of *de novo Dnmt3a* and *Dnmt3b* in ESCs leads to a differentiation block with maintenance of self-renewal potential<sup>243</sup>.

*Dnmt3a* mutations in mouse HSCs lead to progressive impairment in differentiation and HSC expansion in mouse BM over serial transplantation in *Mx1-Cre<sup>+</sup>* mice with induced *Dnmt3a* deletion<sup>138</sup>. Engraftment data from secondary mouse recipients of *Dnmt3a-null* or control HSCs shows that *Dnmt3a-null* HSCs have a significantly greater contribution to PB production than control, while maintaining multilineage differentiation. A significant increase in HSC numbers was observed during serial engraftment of *Dnmt3a-null* HSCs compared with control HSCs, up to quaternary recipients. Furthermore, the regenerative ability of *Dnmt3a-null* HSCs has been shown to last at least up to 12 serial transplants<sup>244</sup>. Taken with the normal multilineage engraftment and equal representation of *Dnmt3a* deletion and WT in downstream progenitors, the authors conclude that the effect on proliferation of *Dnmt3a* deletion was restricted to the most primitive HSCs.

Further, *Dnmt3a-null* HSCs transcriptionally displayed upregulation of HSC multipotency genes and incomplete repression of HSC-specific genes; and downregulation of differentiation factors in *Dnmt3a-null* mouse HSCs<sup>138</sup>. These findings suggest that *Dnmt3a* mutation in HSC/MPPs maintains them in a stem-like state and confer self-renewal advantage and impairs differentiation. This identified role of *Dnmt3a* in mouse HSCs is different from that of *Dnmt1*, which leads to premature HSC exhaustion and lymphoid-deficient differentiation, suggesting that the DNMTs have distinct roles in HSCs.

*DNMT3A* is seen in 15-20% of ARCH cases<sup>88,93</sup> and 18-22% of adult de novo AML<sup>230,245,246</sup>. It is a 130kDa protein encoded by 23 exons on human chromosome 2p23<sup>101</sup>. 50-60% of all *DNMT3A* mutations occur at a single hotspot, arginine 882 (*R882*)<sup>247</sup>, which is mutated to histidine (*R882H*) or cysteine (*R882C*) and is a dominant negative mutation. The other half of all AML-associated *DNMT3A* mutations are monoallelic nonsense or frameshift alterations<sup>92,248,249</sup> (Fig 4.1-1).



**Fig 4.1-1 *DNMT3A* mutations in acute myeloid leukaemia (AML) and clonal haematopoiesis (CH).**

Data from Jaiswal *et al*, *NEJM* 2014, 57 *DNMT3A* mutations from 51 AML patients and 403 *DNMT3A* mutations from 398 CH individuals. Each lollipop represents a non-synonymous mutation, and the size of the lollipops represent the frequency of that mutation within their group and the colour represents the type of mutation. From Brunetti *et al*, *Cold Spring Harb Perspect Med* 2017.

Its role in AML is being extensively studied and mutation of this gene in haematological malignancies has been established as a poor prognostic factor by a number of groups<sup>245,250-252</sup> but is contested by others<sup>253</sup>. *DNMT3A* mutant HSCs from patients with AML have been shown to have a higher rate of engraftment than *DNMT3A* WT HSCs<sup>254</sup>. These *DNMT3A* mutant HSCs with greater engraftment potential, also termed pre-leukemic HSCs (pL-HSCs), have been found in remission samples of AML patients, indicating they may have inherent resistance to chemotherapy<sup>254</sup>.

*DNMT3A* mutations have also been observed in patients with other haematological diseases including MDS (approximately 10%)<sup>238,255</sup>, MPN including polycythaemia vera (PV), essential thrombocytosis (ET)<sup>84,256</sup> and rarely in lymphoid malignancies<sup>139,143,257-259</sup>. In MDS, *DNMT3A* *R882* mutations confer a poorer prognosis and earlier progression to AML than patients without *DNMT3A* mutations<sup>260</sup>. *DNMT3A* *R882* is less commonly associated with lymphoid malignancies, with <20% of *DNMT3A* mutations affecting this region in peripheral T cell lymphoma<sup>141,257,261</sup>. The prevalence of *non-R882* is higher than *R882* in *DNMT3A* related ARCH that does not progress to leukaemia, suggesting *R882* mutations are particularly leukemogenic<sup>114,262</sup>.

*DNMT3A* mutations commonly co-occur with more proliferative gene mutations such as *NPM1* (60%) and *FLT3-ITD* in patients with *de novo* AML<sup>249,263</sup>. *IDH1* mutations are also enriched in patients with *DNMT3A* mutant AML,

with suggestion that this combination of mutations may contribute to progression from MDS to AML<sup>249,263</sup>. Interestingly, no association between *DNMT3A* mutations and other pLM such as *TET2* and *ASXL1* has been reported in the context of AML<sup>92</sup>. *DNMT3A* mutations have also been associated with splicing mutations *SF3B1* and *U2AF1* in patients with MDS<sup>255</sup>.

The prognostic implication of *DNMT3A* in patients with AML is inconsistent between studies<sup>249,253,263–266</sup>, and the variation in patient cohorts and treatment regimens poses significant challenges when making comparisons between these reports. Some studies indicate that *DNMT3A R882* confers poorer prognosis in patients with myeloid leukaemia than those without this mutation, with evidence of persistence of pL- HSCs harbouring this mutation following standard dose daunorubicin based chemotherapy<sup>254,263,267,268</sup>. Other studies suggest improved response to high-dose daunorubicin only in patients with *DNMT3A* mutant AML and not in *DNMT3A-WT* AML<sup>266</sup>. *DNMT3A R882* mutations also predict minimal residual disease (MRD), further implicating their role in resistance to standard chemotherapy<sup>248</sup>. However, there is also indication that *DNMT3A* mutant haematological malignancies have a better response to hypomethylating agents<sup>269</sup>, but this is not yet confirmed in large patient cohorts in prospective studies.

*DNMT3A* has been detected in progenitors without being present in HSCs<sup>126</sup>, raising the question of precise timing and cell of acquisition of this somatic mutation. While a phylodynamic approach has suggested that *DNMT3A* mutations occur very early in life, at least in individuals who subsequently progress to haematological malignancy later in life<sup>84</sup>, further studies are required to understand the clonal expansion in blasts and phenotypically normal mature cell populations.

*DNMT3A* is considered a pLM in the context of human disease for 3 main reasons:

1. *DNMT3A* mutations usually have higher VAF than other accompanying mutations in patients with leukaemia<sup>270,271</sup>,
2. *Dnmt3a* mutant HSCs from mice have a self-renewal advantage, leading to their expansion and serving as a preL lesion<sup>138,244</sup>
3. *DNMT3A* mutations are observed in all mature blood lineages<sup>114</sup> in individuals without haematological malignancy

pLM such as *DNMT3A R882* are not sufficient to induce an AML phenotype in mouse models and cooperation with other proliferative mutations, *Npm1* and/or *Flt3-ITD*, result in either fully penetrant leukemic phenotype when all 3 mutations are present, or a prolonged latency, incomplete or absent phenotype when any single or pair of disease alleles are present<sup>248</sup>.

Persistence of *DNMT3A* mutation in pL-HSCs that have preserved multilineage differentiation capacity is likely the harbinger of chemotherapy evasion and subsequent relapse in patients with this mutation<sup>126,129</sup>. Hence efforts to understand the functional effects of *DNMT3A* at the pL-HSC level are necessary to identify features that differentiate *DNMT3A*-mutant HSCs from *DNMT3A-WT* “normal” HSCs within the same individual.

### 4.1.3 Transcriptional characterisation of ARCH mutations

Since HSCs and mature blood cells with and without ARCH related mutations appear morphologically and phenotypically similar and co-exist in the same individual, they need to be distinguished by genotyping at the single cell level. Murine models and single cell sequencing of human blood cells has been used to characterise transcriptional changes conferred by these mutations at the HSC/MPP level. There are currently 2 published methods that allow simultaneous scRNAseq and genotyping in the same single cell: TARGET-seq<sup>272</sup> and Genotyping of Transcriptomes (GoT)<sup>273</sup>. The latter technique has recently been used to study transcriptional differences between *DNMT3A R882* and WT cells, using samples from multiple myeloma (MM) patients<sup>274</sup> and reviewed below.

#### 4.1.3.1 Mouse

The effects of ARCH mutations, specifically *TET2*, *DNMT3A* and *IDH2*, on HSC transcriptional priming in murine models have been described<sup>275</sup>. Transcriptional changes in murine models of ARCH mutations such as *Dnmt3a* KO complement functional studies, which have shown that *Dnmt3a* KO HSCs can persist, giving rise to multiple serial transplantations in mice, up to 12<sup>244</sup>.

Transcriptional priming of *Dnmt3a* mutant mouse models towards specific HSC subsets has been observed. Izzo et al<sup>275</sup> interrogated murine HSPCs by single cell sequencing of BM Lin- HSPCs from *Mx1-Cre Dnmt3a KO* mice to investigate transcriptional changes conferred by this mutation at the HSPC level. The HSC-1 cluster has been defined by this group as having increased cell quiescence and reduced cell cycle activity when expanded through differential gene expression, making this a dormant HSC subtype<sup>275</sup>. Relative reduction in the number of cells displaying transcriptional features of the HSC-1 cluster in *Dnmt3a* KO mice compared with WT mice was observed. Increased self-renewal conferred by *Dnmt3a* KO at the HSC level causes these HSCs to dramatically outcompete their WT counterparts, leading to their accumulation in the BM<sup>138</sup>.

In mice, the HSC phenotype of *Tet2* mutant HSCs has been shown as an expansion of the HSC-1 pool with decreased cell cycle activity and increased quiescence, which may lead to the expansion of these mutated HSCs and risk of secondary mutations leading to malignant transformation<sup>275</sup>. Mature lineage biases have also been noted in mouse HSCs with CH mutations, with an increase in myelomonocytic progenitors and decrease in erythroid progenitors described in *Tet2* KO mice<sup>275</sup>. The reverse was shown in *Dnmt3a* KO mice, with erythroid skewing over myelomonocytic lineage as observed from scRNA-seq of *Mx1-Cre Dnmt3a* KO mice<sup>275</sup>.

#### 4.1.3.2 Human

Some of the above findings from mouse models have been explored transcriptionally in hHSCs by studying single cell sequencing data from HPSCs of MM patients with *DNMT3A R882* mutations. Interestingly, GoT data

analysing pseudotime differentiation of HSPCs showed no global significant difference in *DNMT3A* mutant and WT cells, suggesting the absence of a differentiation block conferred by *DNMT3A R882* mutation in hHSPCs. Regarding differentiation, *DNMT3A R882* mutations in hHSCs resulted in a myeloid differentiation bias over lymphoid as suggested by an enrichment in myeloid biased cells versus early lymphoid progenitors when examining lympho-myeloid primed progenitors and common lymphoid progenitors from MM patients with known *DNMT3A R882* mutation. This group also observed a significant correlation between *DNMT3A R882* mutation and transcriptionally primed immature myeloid progenitors towards erythroid-megakaryocyte differentiation<sup>274</sup>. Proinflammatory signatures and dysregulated TF activity are also observed in *DNMT3A* mutated HSCs. Furthermore, genes associated with LSCs are enriched in *DNMT3A* mutant HSCs, including *PRSS21*, *FCER1G*, *TYROBP* and *TNFRSF4*, the latter 3 also being involved in pro-inflammatory signalling. Oncogenic targets including *PIM2*, downstream of STAT signalling, have also been identified as being upregulated in *DNMT3A* mutant progenitors<sup>274</sup>. In summary, transcriptional changes related to *DNMT3A R882* mutation in hHSPCs have been observed, which have not been investigated functionally to date.

Since *DNMT3A* mutant and WT cells, including HSPCs are morphologically and phenotypically similar, our ability to link phenotype with genotype has been limited. Other groups have overcome this challenge by characterising HSPCs from patients with haematological disorders carrying this mutation<sup>274</sup>. I have used primary human samples from both ARCH individuals and AML patients to isolate HSC/MPPs distinct from leukemic cells to functionally characterise differentiation of *DNMT3A R882* mutant cells using their own *DNMT3A WT* cells as the comparison.

## 4.2 Aims

Functional effects of pLM in mouse models have been described, with increased self-renewal at the HSC level being a consistent finding<sup>138,229</sup>, however the directly comparable effects in human HSCs has not been defined.

Focusing on the most common ARCH related mutation, the aims of this chapter are to answer the following questions:

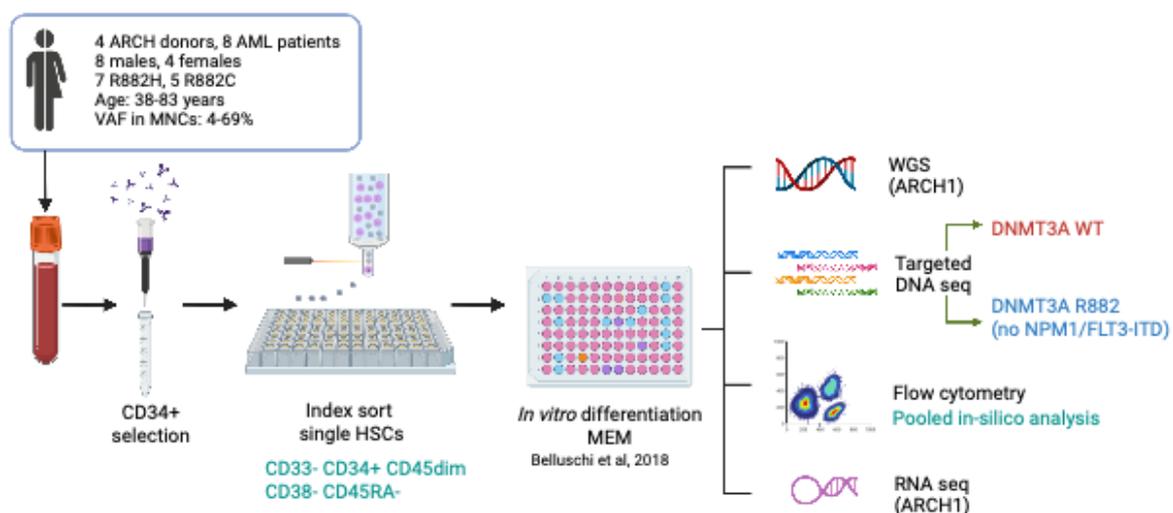
1. Does *DNMT3A R882* confer selective proliferative advantage in human HSC/MPPs over *DNMT3A WT* HSC/MPPs?
2. Are there any effects on proliferation and differentiation capacity and dynamics of *DNMT3A R882* mutation on human HSC/MPPs?

## 4.3 Results

### 4.3.1 Analysis of the clonal dynamics of *DNMT3A* R882 mutant HSC/MPPs in an ARCH individual

To reconstruct hHSC clonal dynamics and assess the role of *DNMT3A* R882 mutation in HSCs in healthy individuals, we obtained a sample of mobilised peripheral blood (mPB, originally collected for planned allogeneic HSCT for a patient with AML) from a 63-year-old female with a significant *DNMT3A* R882H clone (variant allele frequency, VAF, 35% in bulk peripheral blood mononuclear cells, PB MNCs) detected at the time of mobilised PB collection. The sample was not used for its intended purpose due to premature death of the recipient. No other mutations were detectable by targeted DNA sequencing of PB MNCs. At the time of collection, the donor was clinically well with normal blood counts, but went on to develop clinically apparent essential thrombocythemia (ET) 4 years after this sample was collected.

The overall experimental outline used to analyse this sample is illustrated in Fig 4.3-1. We performed CD34+ selection on frozen mPB MNCs from this individual, isolated 1280 single HSCs/MPPs (Live/CD19-/CD33-/CD34+/CD45dim/CD38-/CD45RA-) and 624 GMPs (Live/CD19-/CD33-/CD34+/CD45dim/CD38+/CD7-/CD10-) by FACS (Fig 4.3-1) and cultured them *in vitro* for 3 weeks under conditions that permit simultaneous differentiation of erythroid, monocytic, granulocytic and natural killer (NK) cells *in vitro*<sup>50</sup>. 498 single HSCs/MPPs/GMPs produced colonies suitable for phenotyping analysis using flow cytometry and targeted DNA genotyping for the *DNMT3A* R882H mutation. In addition, whole genome sequencing (WGS) was performed on 127 of these mature single-cell derived colonies at an average sequencing depth of 13X. We used this data to generate a phylogenetic tree using a population genetics approach<sup>276</sup> as in Mitchell et al<sup>82</sup> (Fig 4.3-2A).



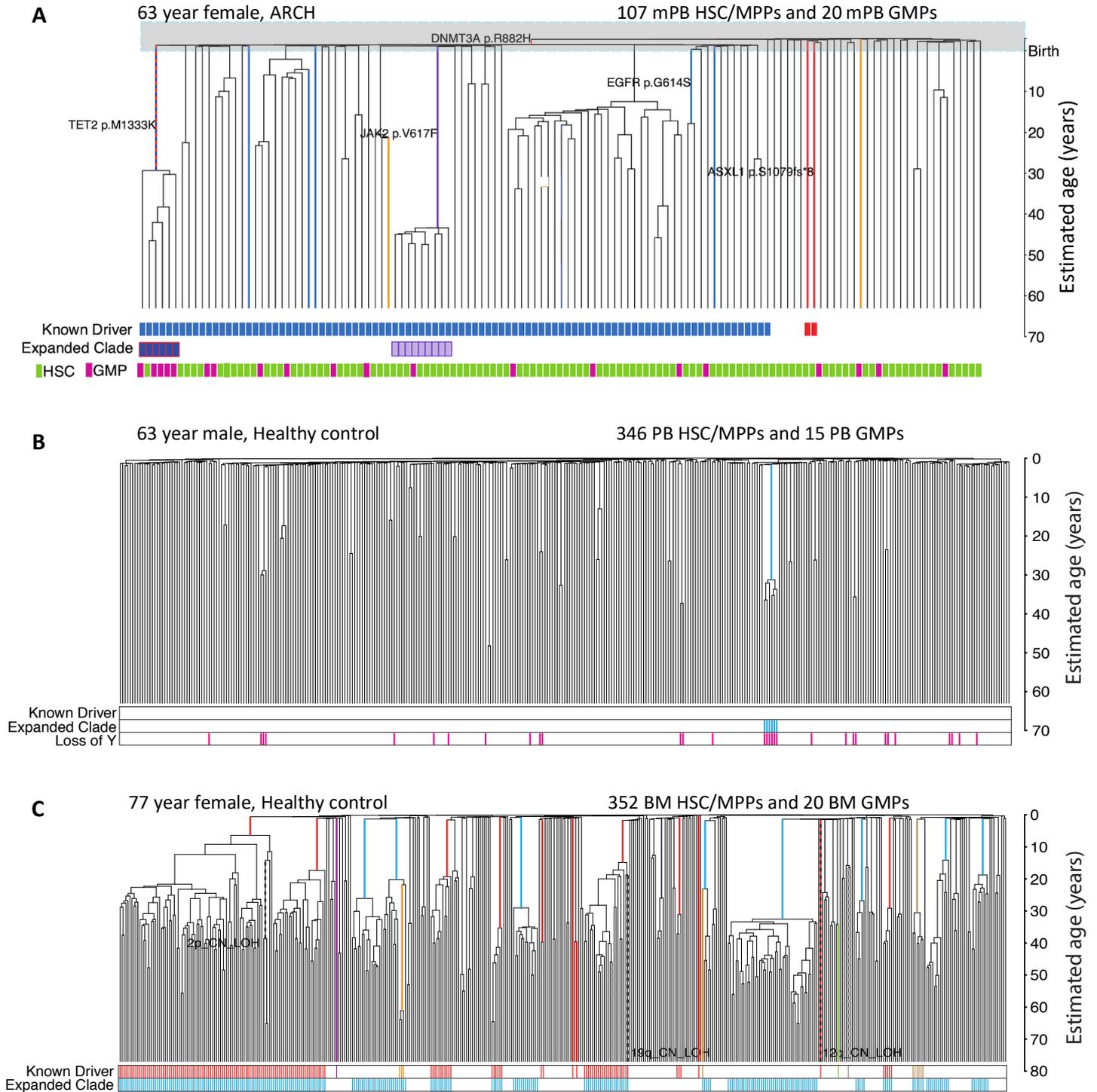
**Fig 4.3-1 Experimental schema used to correlate phenotype and genotype information from single HSC derived colonies from individuals with *DNMT3A* R882 clone in PB.** Bone Marrow (BM) or peripheral blood (PB) from individuals with a known *DNMT3A* R882 clone was used to select CD34+ cells and isolate HSCs/MPPs for *in vitro* differentiation into erythroid/myeloid/NK lineages. Colonies were genotyped using WGS, targeted DNA sequencing, RNAseq and phenotyped using high throughput flow cytometry. WGS and RNAseq were only performed on colonies from 1 individual (ARCH1).

The resulting phylogenetic tree displays an extensive *DNMT3A* R882H clade (74% of all colonies), established very early in life, followed by additional clonal expansions driven by mutations acquired later in adulthood (Fig 4.3-2A). Among these, we identified *TET2* M1133K and *JAK2* V617F (referred to as *TET2* and *JAK2* clades from here on), with the latter likely being responsible for this patient's subsequent diagnosis of ET, as this mutation is the hallmark of this disease<sup>277</sup>. Estimation of the time of mutation acquisition was performed as in Mitchell et al. and showed that *DNMT3A* R882 mutation acquisition in this individual was an early event, likely to have been acquired in utero. Williams et al<sup>239</sup> recently showed that in patients with MPNs, *DNMT3A* R882 mutation is frequently acquired early in life, and earlier acquisition correlates with increased risk of blood disorders in later life. This phylogenetic tree exemplifies *DNMT3A* R882H mutation acquisition early in life in the context of clinically normal blood counts in a healthy individual, hence captured at a time of transition from health to disease.

Furthermore, compared to a healthy control's age matched phylogenetic tree<sup>82</sup>, which shows high polyclonality and only 1 slightly expanded clade (Fig 4.3-2B), it is evident that our individual's clonal dynamics are more oligoclonal than would be expected for her age. In fact it is closer in oligoclonality to a 77 year old gender matched healthy control<sup>82</sup> (Fig 4.3-2C). The Shannon diversity index, is a mathematical measure of diversity within a species that takes into account the abundance and evenness of the species present. This measure was used to quantify clonal diversity, calculated using the number of lineages present at 100 mutations of molecular time (equivalent to the first few years after birth) as the number of species and the number of HSC/MPP derived colonies arising from that lineage as a measure of abundance<sup>82</sup>. The Shannon diversity index of ARCH1 across the whole phylogenetic tree is 6.09, similar to trees individuals >70 years as shown in Mitchell et al<sup>82</sup>, while that of an age matched control (63 year old male) without *DNMT3A* R882 mutation or any other known driver mutation is 14.80 (unpublished). This oligoclonality seems however to be restricted to the *DNMT3A* R882 clade, as WT HSC-derived colonies shared few mutations with one another, indicative of higher polyclonality, as expected from healthy individuals of that age<sup>82</sup>.

We then associated flow cytometry phenotyping information from the 107 HSC/MPP derived colonies used to generate the phylogenetic tree above and observed variations in the levels of median expression and proportions of the myeloid markers CD14 and CD15 across colonies and clades including *DNMT3A* R882 and

*DNMT3A* WT, as well as the *TET2* clade (Fig 4.3-3A). When analysed quantitatively, mature myeloid cells derived from single HSC/MPPs with *DNMT3A* R882H mutation had lower proportions of monocytes (percent of all myeloid cells within a colony being CD14+,  $p=0.06$ , Fig 4.3-3B) and lower CD14 MFI within CD45+/CD56-/CD11b+ cells ( $p=0.008$ , Fig 4.3-3C) than their WT counterparts, indicating fewer and less mature monocytes derived from HSC/MPPs with *DNMT3A* R882H mutation. Proportions of granulocytes (percent of all myeloid cells within a colony being CD15+) were higher in the *DNMT3A* R882H mutant HSC/MPP derived colonies than WT, ( $p=0.04$ , Fig 4.3-3D) with similar CD15 MFI between *DNMT3A* R882H mutant and WT colonies amongst the CD45+/CD56-/CD11b+ cells (Fig 4.3-3E), suggesting more granulocytes being produced from *DNMT3A* R882H mutant HSCs/MPPs.



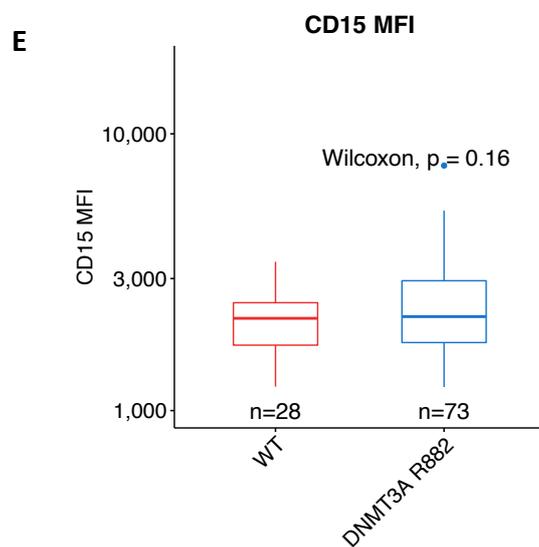
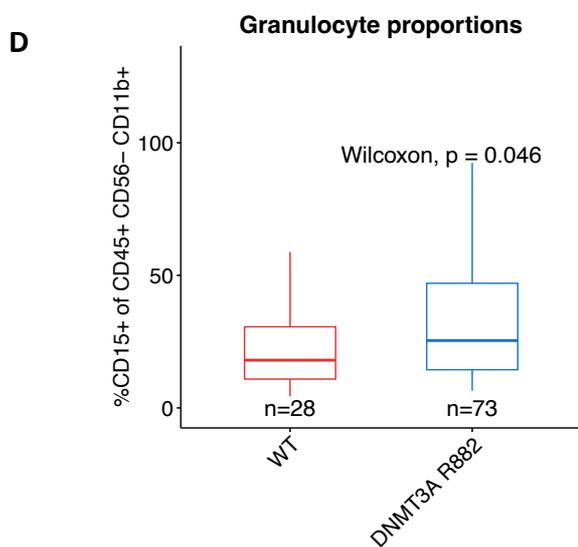
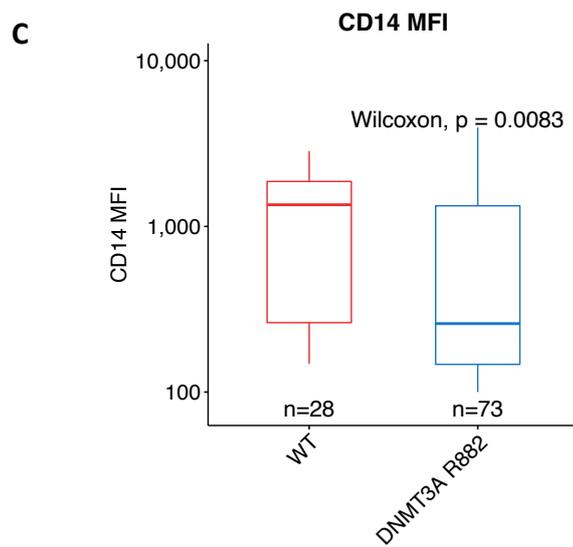
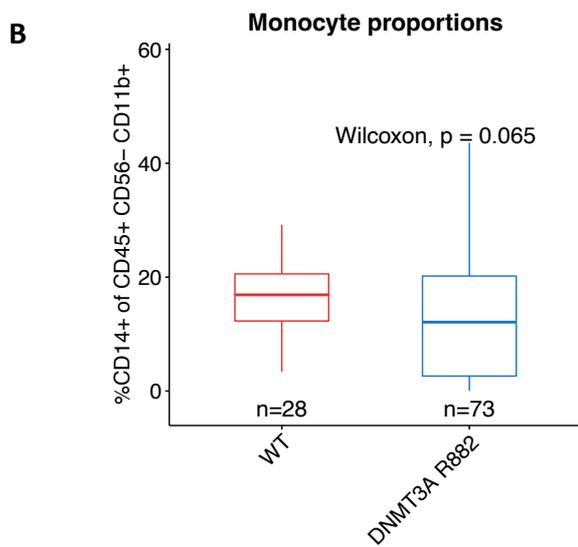
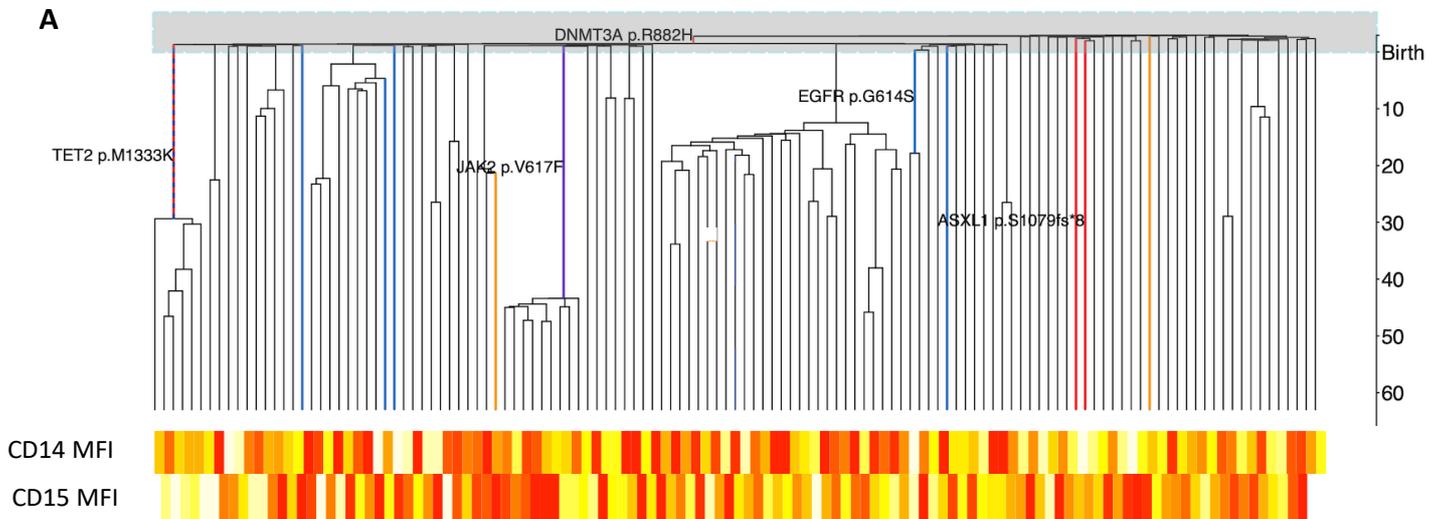
**Fig 4.3-2 Clonal dynamics of *DNMT3A* R882H mutation within an ARCH individual and comparison with healthy controls.** Phylogenetic trees are generated from single HSC/MPP derived colonies, with each branch representing a single colony. (A) Phylogenetic tree generated from single HSC/MPP/GMP derived colonies from the ARCH1 individual, displaying timing of mutation acquisition and relationships between mutations. (B) Age matched phylogenetic tree from a 63 year old male. (C) A 77 year old healthy control's phylogenetic tree. All phylogenetic trees bioinformatically generated by Dr Emily Mitchell (B and C generated by Dr Emily Mitchell and reported in Mitchell et al, Nature 2022).

When separately analysed within the *DNMT3A* R882H mutant clade, the *JAK2 + DNMT3A* (n=8 colonies) and *TET2 + DNMT3A* (n=6 colonies) clades had significantly different phenotypes for CD14 and CD15 expression compared with the broader *DNMT3A* clade, excluding the *JAK2* and *TET2* clades as shown in Fig 4.3-3A. The reduced monocyte differentiation phenotype was more prominent in the *TET2 + DNMT3A* clade. Percentage of CD14+ cells within the myeloid cells (CD45+/CD56-/CD11b+) was lower in the *TET2 + DNMT3A* clade than *DNMT3A* alone or WT (p=4.3e-6 and p=6.5e-7, Fig 4.3-4A). It was also lower in the *TET2 + DNMT3A* clade compared with the *JAK2 + DNMT3A* clade (p = 0.005, Fig 4.3-4A). CD14 MFI in the myeloid cells was not significantly altered between *TET2* and *JAK2* clades compared with each other or with *DNMT3A* (Fig 4.3-4B).

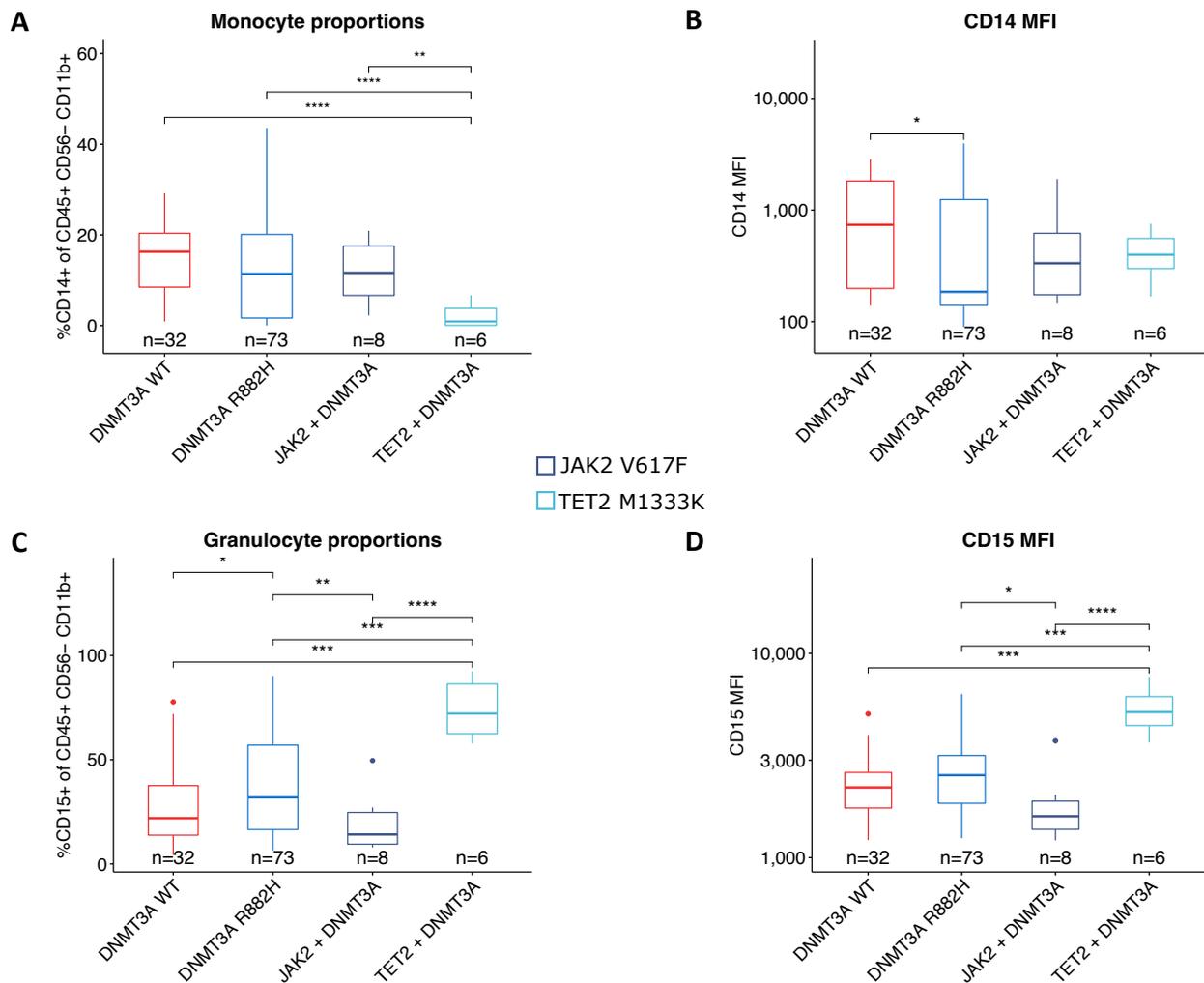
Similarly, the increased granulocyte phenotype was also more prominent within the *TET2 + DNMT3A* clade. Percent of CD15+ cells within myeloid cells in each colony was higher in the *TET2 + DNMT3A* clade compared with *DNMT3A* alone and WT (p= 2.7e-7 and p=7.5e-9 respectively, Fig 4.3-4B). Analysis of CD15 MFI of the myeloid cells yielded similar observations, with the *TET2 + DNMT3A* clade expressing a higher CD15 MFI than *DNMT3A* alone or WT (p=0.0005 and p=0.0001 respectively, Fig 4.3-4C). CD15 MFI of the myeloid cells was lower within each colony in the *JAK2 + DNMT3A* clade compared with *DNMT3A* alone (p=0.02, Fig 4.3-4C), and the same for percentage of CD15+ cells within the myeloid cells of each colony (p=0.02, Fig 4.3-4B). *JAK2 + DNMT3A* colonies contained fewer CD15+ cells and lower CD15 MFI amongst myeloid cells than the *TET2 + DNMT3A* clade (p=0.0001 and p=2.7e-5, Fig 4.3-4B and C respectively).

In summary, in this 63 year old individual with a significant *DNMT3A* R882H clone, which was acquired very early in life, there is evidence of oligoclonality within the *DNMT3A* R882H mutant HSPCs beyond what would be expected at her age within the *DNMT3A* R882H mutant HSC/MPPs and progenitors. Further, there is alteration in differentiation dynamics within the myeloid branch as measured by monocyte and granulocyte proportions

and maturity between *DNMT3A* R882H mutant and WT HSC/MPPs, as well as between *TET2* and *JAK2* mutant HSC/MPPs and progenitors compared with *DNMT3A* R882H mutant or WT HSCs/MPPs and progenitors.



**Fig 4.3-3 Variations in myeloid differentiation between clades of the phylogenetic tree derived from ARCH1.** (A) Observed variations in CD14 and CD15 MFI between *DNMT3A R882H* and WT colonies in the phylogenetic tree generated from ARCH1. (B) %CD14+ cells and (C) CD14 MFI (D) %CD15+ cells and (E) CD15 MFI within the myeloid (CD45+/CD56-/CD11b+) cells of HSC/MPP derived colonies. Box plots represent median and inter-quartile ranges, p values calculated using Wilcoxon test.



**Fig 4.3-4 Variations in myeloid differentiation between clades of the phylogenetic tree derived from ARCH1.** Observed variations in (A) CD14+ proportions, (B) CD14 MFI, (C) CD15+ proportions and (D) CD15 MFI between *DNMT3A R882H*, *JAK2* and *TET2* clades in the phylogenetic tree generated from ARCH1. *JAK2* and *TET2* clades refer to *JAK2 V617F* and *TET2 M1333K* mutations respectively. All proportions and MFI are within the myeloid (CD45+/CD56-/CD11b+) cells of HSC/MPP derived

colonies. Box plots represent median and interquartile range, significance as per two-tailed t-tests. P values on all graphs are represented as \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

### **4.3.2 Establishing a cohort to study the effects of *DNMT3A R882* mutation on HSC/MPP differentiation**

In view of the longstanding presence of the *DNMT3A R882H* clone in this individual and the phenotypic signals within the myeloid differentiation branch, I sought to further evaluate the role of this mutation on differentiation dynamics of mutant vs WT HSC/MPPs. As well as examining these changes within this individual, I also broadened the cohort to a further 11 patients with a known *DNMT3A R882* clone in PB MNCs.

I assessed a total of 12 independent PB, mPB and BM samples from carriers of *DNMT3A R882* mutations (7 with *R882H* and 5 with *R882C* mutations, including ARCH1, the individual discussed above), all with VAF  $> 2\%$  (Table 4.3-1). Samples were obtained from 4 ARCH donors and 8 AML patients (4 females and 8 males, age range 38-83 years, median 63 years). In addition, I also evaluated HSC/MPP differentiation from PB HSC/MPPs of 5 healthy age matched controls (HC) without a known *DNMT3A R882* clone. For all samples I used a single cell assay that allows HSC/MPP differentiation towards all major blood lineages as described in methods<sup>50</sup>, including Ery (erythroid, GlyA+), Mye (myeloid, CD45+/CD56-/CD11b+) and NK (CD45+/CD56+) cells. The overall differentiation phenotype and within the myeloid branch, further differentiation into monocytes (Mono, CD45+/CD56-/CD11b+/CD14+) and granulocytes (Gran, CD45+/CD56-/CD11b+/CD15+) can be distinguished using flow cytometry analysis.

Sample	Age (years)	Gender	Diagnosis	Source	DNMT3A mutation	DNMT3A VAF in PB mature cells	Engraftment in mice from PB	Leukemic blast cytogenetics	
HC1	63	Male	Healthy	Non-mPB	None	NA	NA	NA	
HC2	37	Male	Healthy	Non-mPB	None	NA	NA	NA	
HC3	30-40	Male	Healthy	Non-mPB	None	NA	NA	NA	
HC4	25-30	Male	Healthy	Non-mPB	None	NA	NA	NA	
HC5	35-40	Male	Healthy	Non-mPB	None	NA	NA	NA	
ARCH1	63	Female	ARCH	mPB	R882H	35%	Multi-lineage	NA	
*ARCH2	82	Male	ARCH	Non-mPB	R882H	4%	Unknown	NA	
*ARCH3	58	Female	ARCH (chemotherapy related)	mPB	R882H	69%	Multi-lineage	NA	
ARCH4	80	Male	ARCH	BM	R882C	43%	Unknown	NA	
AML1	38	Male	AML	Non-mPB	R882H	20%	Pre-leukemic, multilineage	NPM1 TCTG insertion	FLT3 WT
AML2	78	Male	AML	Non-mPB	R882H	48%	Multi-lineage	NPM1 TCTG insertion	FLT3 WT
AML3	70	Male	AML	Non-mPB	R882C	47%	Multi-lineage	NPM1 GTGC insertion	FLT3 ITD
AML4	63	Female	AML	Non-mPB	R882C	44%	Multi-lineage	NPM1 TCTG insertion	FLT3 ITD
*AML5	83	Male	AML	Non-mPB	R882H	44%	Myeloid	NPM1 WT	FLT3 ITD
*AML6	69	Male	AML	Non-mPB	R882C	48%	Unknown	NPM1 WT	FLT3 ITD
*AML7	64	Female	AML	Non-mPB	R882C	43%	Leukemic	NPM1 TCTG insertion	FLT3 ITD, low allelic ratio
*AML8	62	Male	AML	Non-mPB	R882H	42%	Unknown	NPM1 WT	FLT3 WT

**Table 4.3-1 Demographics of all samples.** \*= excluded from phenotypic differentiation analysis due to WT or *DNMT3A R882* mutant colonies being statistically underpowered (<20 colonies of either mutant or WT), or lack of multilineage engraftment; VAF = variant allele frequency; AML = Acute Myeloid Leukaemia; ARCH = Age-related Clonal Haematopoiesis; WT = wild type; PB = peripheral blood, mPB = mobilised PB; Non-mPB = non-mobilised peripheral blood; ITD = internal tandem duplication.

CD3 depleted MNCs from 6 of the 8 AML samples and 2 of the 4 ARCH samples were transplanted in immunocompromised mice (NSG-SGM3 and hSCF mice) to assess engraftment capacity (performed in Liran Shlush's lab, Weizmann Institute). Transplantation of AML1, AML2, AML3 and AML4 MNCs resulted in healthy multi-lineage engraftment, while that of 2 other AML samples resulted in leukaemic or myeloid only engraftment in mice. Transplantation of both ARCH1 and ARCH3 also resulted in multi-lineage engraftment as expected. Engraftment capacity of 2 ARCH and 2 AML samples is unknown.

Samples were assessed using an experimental platform combining single cell differentiation assays with genotyping. CD34+ cells from BM, mobilised and non-mobilised PB samples were FACS sorted to isolate single HSCs/MPPs and cultured *in vitro* for 3 weeks using the same strategy as the index sample, ARCH1 (Table 4.3-2). Targeted genotyping for *DNMT3A R882*, *NPM1* and *FLT3* mutations were performed using Miseq on amplified DNA from each individual HSC-derived colony and genotypes were assigned using 3 independent variant callers (VarScan, Platypus and Mutect, see methods 2.6.1.2).

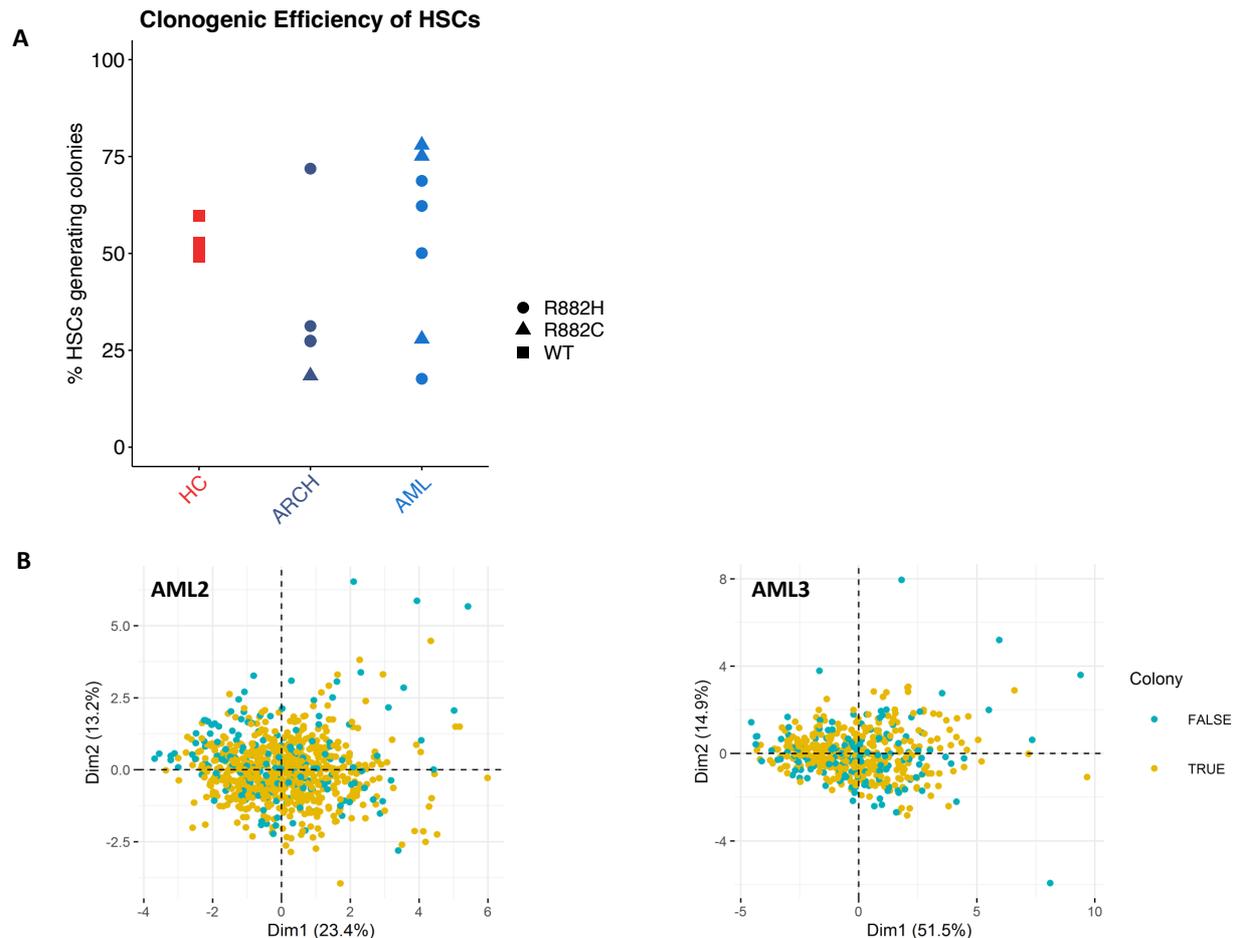
Overall, 8,946 single HSPCs were sorted from 12 *DNMT3A R882* mutation carriers for culture *in vitro*, yielding 3,167 single cell derived colonies, of which 2,990 were from HSCs/MPPs (Table 4.3-2). GMP derived colonies within each sample were statistically underpowered for phenotype/genotype correlation, hence were excluded from the analysis. Overall, median clonogenic efficiency ranged from 17-77% across all samples and was similar between HC, AML and ARCH samples (

Fig 4.3-5A). There was no observed difference in cell surface marker expression of HSCs/MPPs that did and did not produce colonies *in vitro* amongst all samples (

Fig 4.3-5B). Of the HSC/MPP derived colonies, 2,557 had phenotype and genotype data of reliable sequencing depth available. In the final differentiation analysis, 1766 colonies were included (see below for inclusion/exclusion criteria 4.3.6 ), obtained from 6 independent individuals with sufficient intra-individual statistical power to detect differences in phenotype between *DNMT3A R882* mutant and WT HSC/MPP derived colonies. Phenotyping analysis was performed prior to genotype information being available for all samples, in a blinded manner. To assess the effect of *DNMT3A R882* mutation on HSC/MPP differentiation, I always compared colonies derived from *DNMT3A R882* mutant HSC/MPPs to WT ones within the same individual to effectively control for other variable including age, gender and environmental exposures to unknown toxins.

Sample	HSCs/MPPs seeded	Colonies Harvested	Genotype + Phenotype available	DNMT3A R882	DNMT3A WT	NPM1c Mut	NPM1 WT	FLT3 ITD	FLT3 WT
HC1	129	77	77	0	77	NA	NA	NA	NA
HC2	336	175	175	NA	NA	NA	NA	NA	NA
HC3	96	6	6	NA	NA	NA	NA	NA	NA
HC4	177	87	87	NA	NA	NA	NA	NA	NA
HC5	93	49	49	NA	NA	NA	NA	NA	NA
ARCH1	1280	373	351	222	129	0	351	0	351
*ARCH2	96	69	69	4	65	NA	NA	NA	NA
*ARCH3	663	181	157	151	6	NA	NA	NA	NA
ARCH4	960	177	177	155	22	NA	NA	NA	NA
AML1	445	277	258	116	142	0	258	0	258
AML2	486	334	328	244	84	0	328	0	328
AML3	835	651	385	222	163	0	385	0	385
AML4	474	356	267	202	65	0	267	0	267
*AML5	627	314	314	24	290	0	314	0	314
*AML6	394	110	103	1	102	0	103	0	103
*AML7	753	4	4	0	4	0	4	0	4
*AML8	816	144	144	103	41	0	144	0	144

**Table 4.3-2 Clonogenic efficiency and genotyping across samples.** Table shows numbers of HSC/MPPs seeded and colonies harvested for genotype and phenotype analysis. \*= excluded from phenotypic differentiation analysis due to WT or *DNMT3A R882* mutant colonies being statistically underpowered (<20 colonies) or myeloid/leukaemic engraftment in mice, indicating LSCs, not HSCs present.



**Fig 4.3-5 Clonogenic efficiency of all samples.** (A) Distribution of clonogenic efficiency (the percentage of HSC/MPPs that produced mature colonies amongst all single HSC/MPPs that were seeded *in vitro*) of all samples, ranging from 17-77%. 4 healthy controls (HC), 5 age related clonal haematopoiesis (ARCH) individuals and 7 acute myeloid leukaemia (AML) patients included. (B) PCA plots derived from 13 index cell surface markers (see Methods Fig 2.3-9) on HSC/MPPs, showing distribution between those cells that produced colonies (True) and those that did not (False). 2 representative examples, AML2 and AML3 shown here.

### 4.3.3 Isolating HSCs distinct from LSCs in AML samples

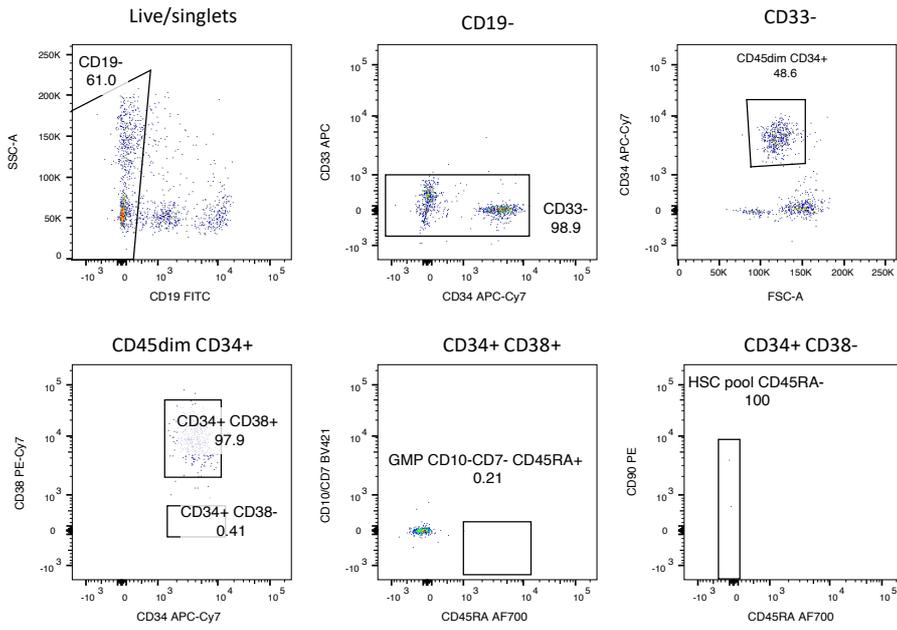
Since CH is an asymptomatic and subclinical state, identifying individuals with ARCH and obtaining samples for analysis is challenging. Hence patients with AML and a significant *DNMT3A* R882 clone in PB were included in our cohort. I first examined whether proportions of phenotypic cell populations in PB vary significantly between AML patients and HC, especially pertaining to the CD34+ compartment. AML samples had significantly fewer CD33- cells and far greater proportions of CD34+/CD38+/CD45RA- cells in CD34+ enriched PB cells compared

with HC and ARCH individuals (Fig 4.3-6). For the purposes of this project, it was critical to isolate pre-leukemic HSC/MPPs (pL-HSCs) and exclude leukemic cells of any maturity. To this end I used 3 strategies:

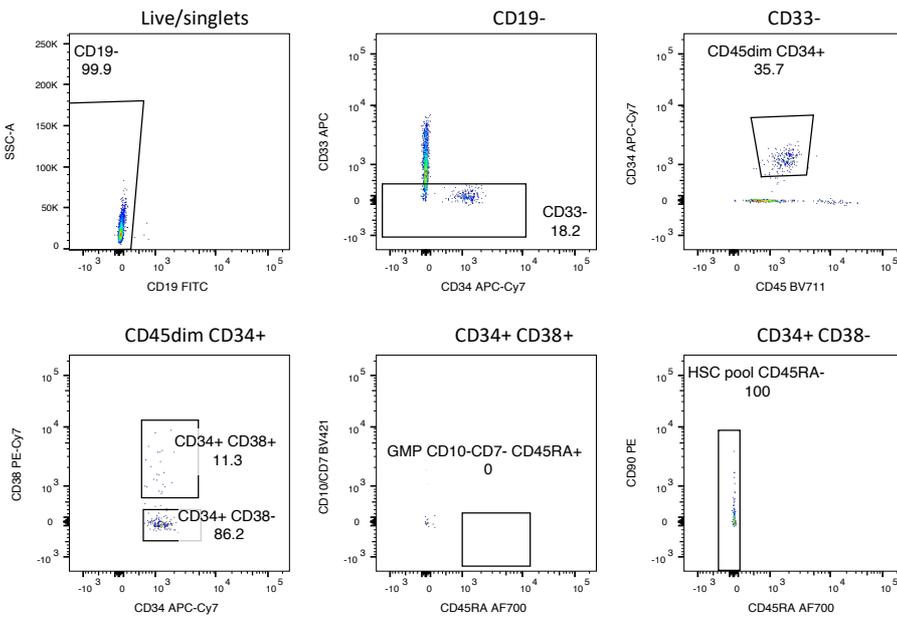
- g) A published sorting strategy to isolate pL-HSCs using FACS on CD34+ enriched PB<sup>126</sup>. After excluding CD33+ and CD45hi cells, which are likely to be leukemic, we isolated CD45dim/CD34+/CD38-/CD45RA- cells (Fig 2.3-3).
- h) We also used *in vitro* conditions that do not support growth of LSCs *in vitro* (Shlush laboratory, unpublished), and are optimised for HSCs<sup>50</sup> (Table 2.3-1).
- i) Finally, targeted sequencing of all colonies derived from all samples confirmed the absence of *NPM1* mutation and *FLT3*-ITD mutations in all the single HSC/MPP derived colonies (Table 4.3-1). These mutations are characteristic of leukaemia and mutually exclusive with health and pre-leukaemic HSC/MPPs.

Using the above multi-layered strategy, we can be confident that the HSC/MPP derived colonies that have been used in the subsequent analyses are in fact derived from true HSC/MPPs and not LSCs, in the AML samples included in the analyses.

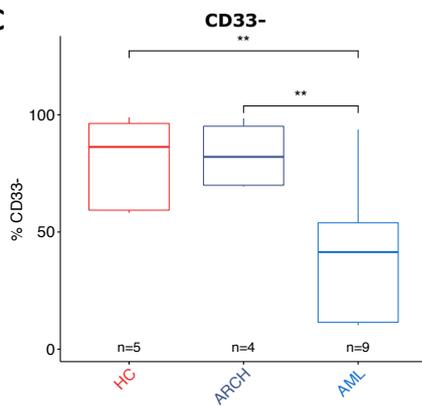
**A HC4**



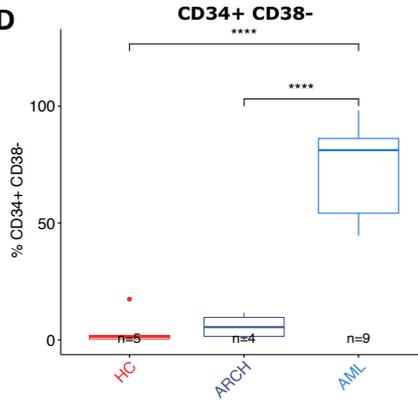
**B AML3**



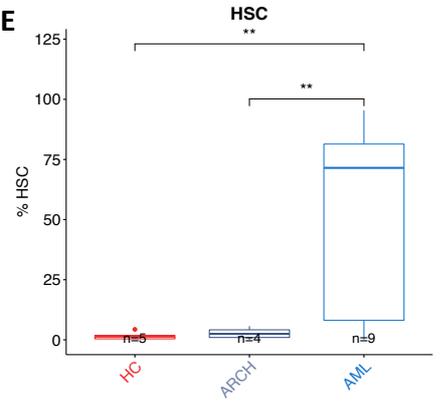
**C**



**D**



**E**

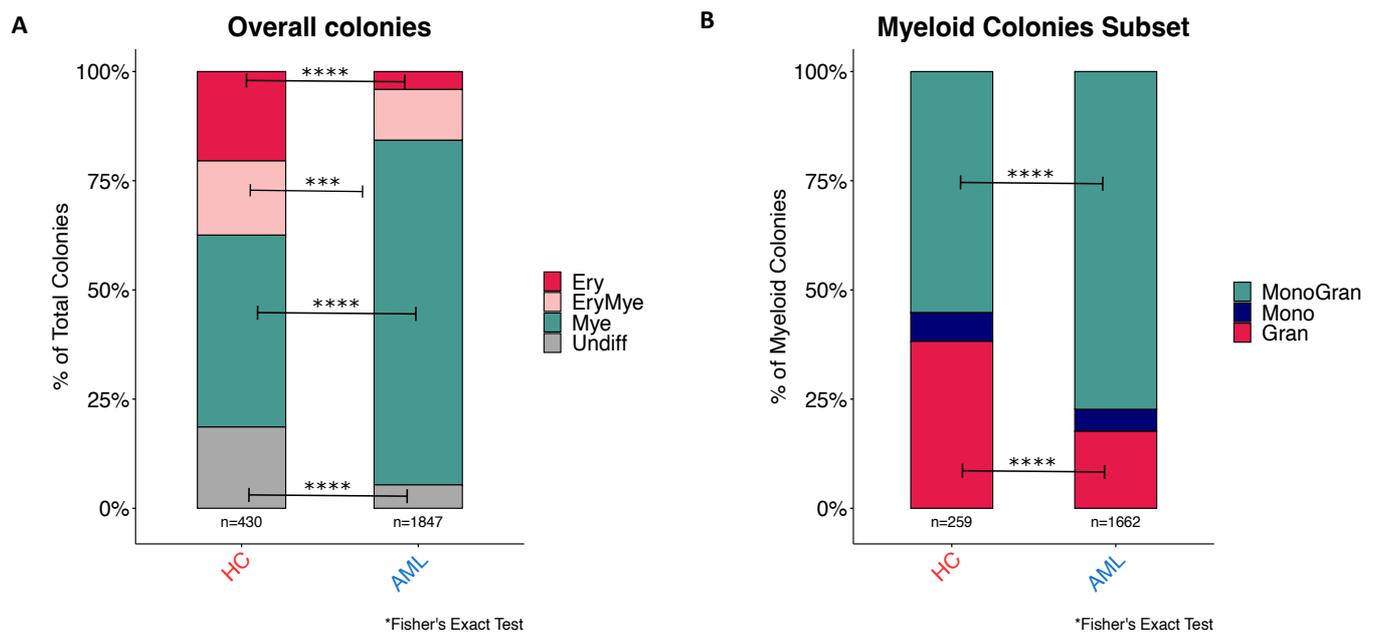


**Fig 4.3-6 Differences in cell populations in PB from AML vs HC after CD34+ enrichment.** 2 representative examples shown here from healthy control (HC4, A) and an AML patient (AML3, B). (C) AML patients have lower proportions of CD33- cells, (D) greater proportions of CD34+/CD38- and (E) CD34+/CD38-/CD45RA- cells from PB MNCs compared with HC and ARCH individuals. Box plots represent median and interquartile range, significance as per two-tailed t-tests.

#### ***4.3.4 Myeloid skewing of PB HSC/MPP differentiation in AML patients compared with healthy age matched controls***

First, I investigated whether the composition of the circulating HSC/MPP pool in PB from AML patients was similar to age matched HC. Single HSC/MPPs from 5 PB HC were cultured for 21 days using the same *in vitro* assay as for all the AML PB samples for 21 days and colonies yielded were assessed by flow cytometry phenotyping as described above in 4.3.2 .

HSC/MPPs from AML patients, including *DNMT3A* R882 mutant and WT HSCs/MPPs analysed together, produced more Mye colonies and fewer Ery colonies than HC (Fig 4.3-7A). In AML patients, these myeloid colonies contained mixed Mono-Gran (CD14+/CD15+) colonies at the expense of Gran only (CD14-/CD15+) colonies compared with HC (Fig 4.3-7B). In keeping with previously published data on circulating HSPCs in myeloproliferative neoplasms (MPNs), beta-thalassemia and chronic lymphocytic leukaemia (CLL) patients<sup>45,278</sup>, HSC/MPPs circulating in the PB of AML patients had myeloid skewing and restricted erythroid differentiation potential compared with those of HC. This data is consistent with myeloid-skewing of circulating HSC/MPPs as a phenotype seen in various haematological diseases, likely indicative of BM dysfunction.



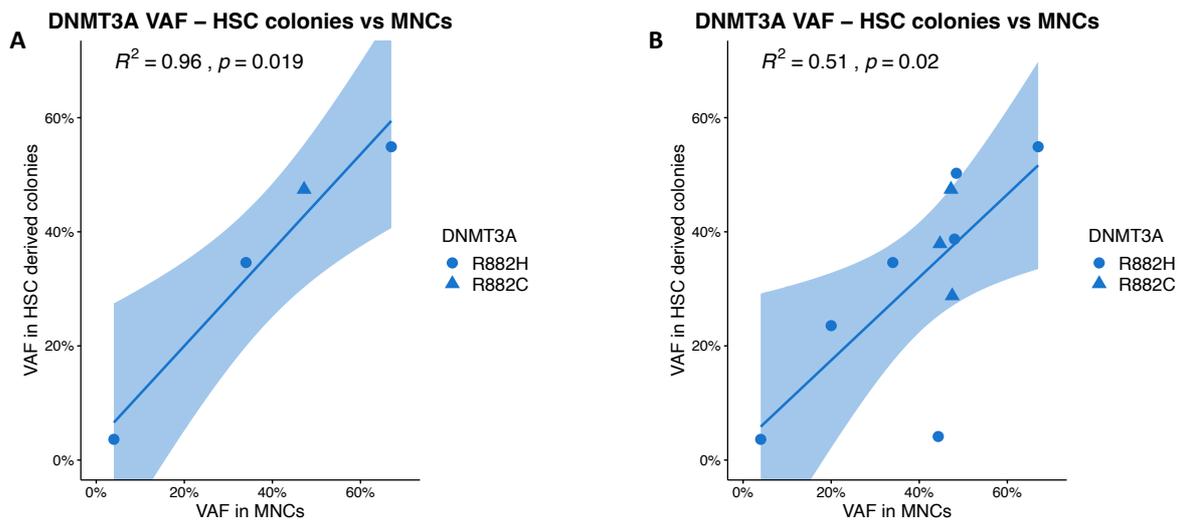
**Fig 4.3-7 HSC/MPP differentiation in AML patients (n=5) compared with healthy controls (HC, n=5) *in vitro* in MEM assay allowing differentiation towards all major blood lineages.** (A) Overall distribution of colony types produced by PB HSC/MPPs derived from AML patients (*DNMT3A R882* and WT together) compared with HC PB HSC/MPPs. Ery = erythroid (GlyA+), EryMye = mixed erythroid myeloid (GlyA+/CD45+/CD56-/CD11b+) and Mye = myeloid colony (GlyA-/CD45+/CD56-/CD11b+). (B) Distribution of myeloid-containing colony types produced by HSC/MPPs derived from AML patients compared with HC. MonoGran = mixed monocyte/granulocyte (CD45+/CD56-/CD11b+/CD14+/CD15+), Mono = monocyte only (CD45+/CD56-/CD11b+/CD14+/CD15-) and Gran = granulocyte only (CD45+/CD56-/CD11b+/CD14= /CD15+) colony. N within figures indicates total colony numbers included in analysis, significance as per two-tailed Fisher's exact test.

### ***4.3.5 No broad selective advantage on HSC/MPPs during *in vivo* differentiation conferred by *DNMT3A R882* mutation***

In our samples, there was a broad distribution of *DNMT3A R882* mutation burden in mature MNCs ranging from 4-48% VAF. To understand the effect of *DNMT3A R882* mutation on HSC/MPP differentiation capacity, we compared the VAF of *DNMT3A R882* mutation in HSC/MPP derived colonies as a marker of *DNMT3A R882* VAF at the HSC/MPP level, with the VAF in the mature MNCs from PB of the same individuals. We hypothesised that a higher VAF in the HSC/MPP derived colonies compared to PB MNCs would indicate a selective disadvantage in

*DNMT3A* R882 mutant HSPCs during *in vivo* differentiation. 10 of our individuals had sufficient data from the HSC/MPP derived colonies to infer the VAF at HSC/MPP level. On direct comparison of *DNMT3A* R882 mutation burden in bulk MNCs and single HSC/MPP derived colonies within each ARCH individual, we observed a significantly positive correlation between VAF in MNCs and colonies (Fig 4.3-8A). This observation remains true even when AML samples, which contain circulating blasts amongst the MNCs and which may artificially skew the VAF in this cohort, are included in the analysis (Fig 4.3-8B). This finding suggests that there is no differentiation block nor a significant selection advantage for cells carrying *DNMT3A* R882 mutation over WT cells during differentiation from HSC/MPPs.

In order to directly measure *DNMT3A* R882 mutation burden at the HSC/MPP level, we also attempted to sort single and bulk (10-50 cells) HSC/MPPs into lysis buffer and amplify DNA for targeted sequencing using Replig, the same technique as used for the HSC/MPP derived colonies. This approach was not successful or reliable in 4 separate individual samples, hence the above approach of measuring VAF in HSC/MPP derived colonies was adopted as a surrogate marker of *DNMT3A* R882 mutation burden in HSC/MPPs.



**Fig 4.3-8 *DNMT3A* R882 variant allele frequency (VAF) in HSC/MPPs as measured in HSC/MPP derived colonies vs MNCs from PB of the same individual.** VAF in single HSC/MPP derived colonies vs PB MNCs across all (A) ARCH samples (n=4), without circulating blasts and (B) including AML samples (n=6), which contain circulating blasts with known *DNMT3A* R882 mutation.  $R^2$  = coefficient of

determination, showing that observed correlation fits the regression model. P-value of  $R^2$  statistic representing fit with linear regression model.

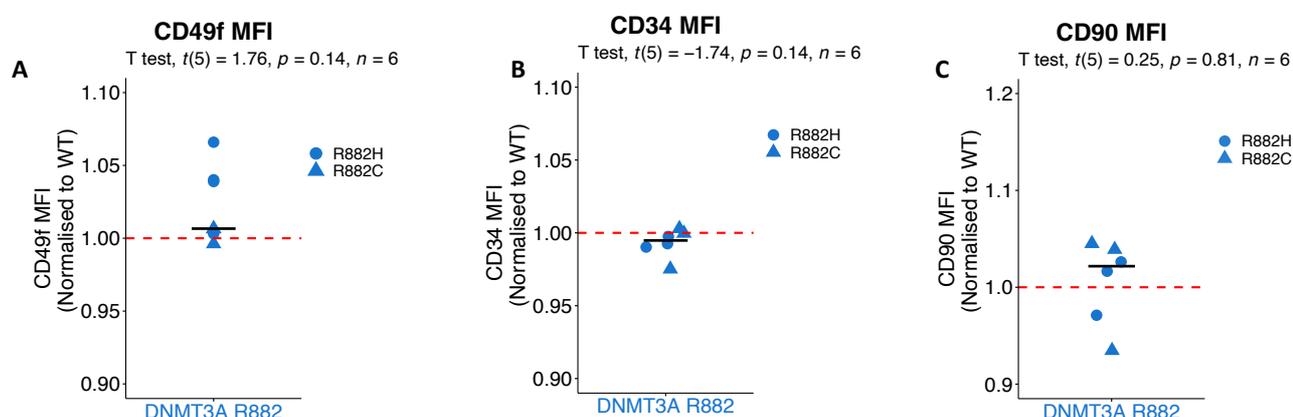
### **4.3.6 Assessment of differentiation phenotype of *DNMT3A R882* mutant HSC/MPPs vs WT**

For 6 individuals with healthy multi-lineage engraftment in mice, there was sufficient statistical power to compare *DNMT3A R882* mutant and WT HSC/MPP-derived colonies within each individual, hence these were selected for an extended analysis of how HSC/MPP differentiation capacity is affected by *DNMT3A R882* mutation. 2 ARCH and 4 AML patients ranging from 38-78 years (median 63 years), including 2 females and 4 males were included, with VAF ranging from 20-48% for *DNMT3A R882* mutation in the bulk peripheral mature MNCs (Table 4.3-1).

6 other samples from the original cohort were excluded based on the following criteria. Samples with <20 colonies, or <10% of the total colonies of the least abundant genotype (*DNMT3A R882* mutant or WT) were excluded from the phenotype analysis. Samples that had myeloid or leukemic engraftment in immunocompromised mice were also excluded from the phenotype analysis to ensure our samples contained HSCs distinct from LSCs. Of the 6 excluded samples, 4 were from AML patients and 2 from ARCH individuals. AML5 had 24 mutant colonies out of 314 total, and AML6 had only 1 mutant colony out of 103 total (VAF 44% and 48% respectively). 1 AML sample (AML7) failed to grow colonies on 2 separate occasions. CD3 depleted PB MNCs from this patient resulted in leukemic engraftment in NSG-SGM3 and hSCF mice. AML8 grew 144 very small colonies, for which no flow cytometry phenotyping data was able to be obtained, while genotyping data was available. Of the 2 ARCH individuals that were excluded from the phenotype analysis, ARCH2 had 4 mutant colonies out of 69 total (VAF in MNCs 4%) and the ARCH3, which was from an individual with treatment related clonal haematopoiesis (CH), had 11 WT colonies out of 153 total.

### 4.3.7 *DNMT3A R882 mutation confers no major alteration in HSC/MPP cell surface markers*

To understand whether *DNMT3A R882* mutation is more frequently observed within a specific HSC/MPP subtype, the median fluorescence intensity (MFI) of all cell surface markers used to isolate HSC/MPPs at FACS were recorded during index sorting. This information was retrospectively compared between the genotypes (*DNMT3A R882* mutant vs WT) of colonies derived from each single HSC/MPP once this data was available. We observed no significant differences, specifically in the expression of antibody markers associated with stemness: CD49f, CD34 or CD90 MFI of *DNMT3A R882* mutant HSCs/MPPs were not significantly different to those of WT HSC/MPPs in our cohort (Fig 4.3-9). This suggests that *DNMT3A R882* mutation does not have a propensity to select for any particular HSC/MPP subtype based on this limited set of cell surface markers.



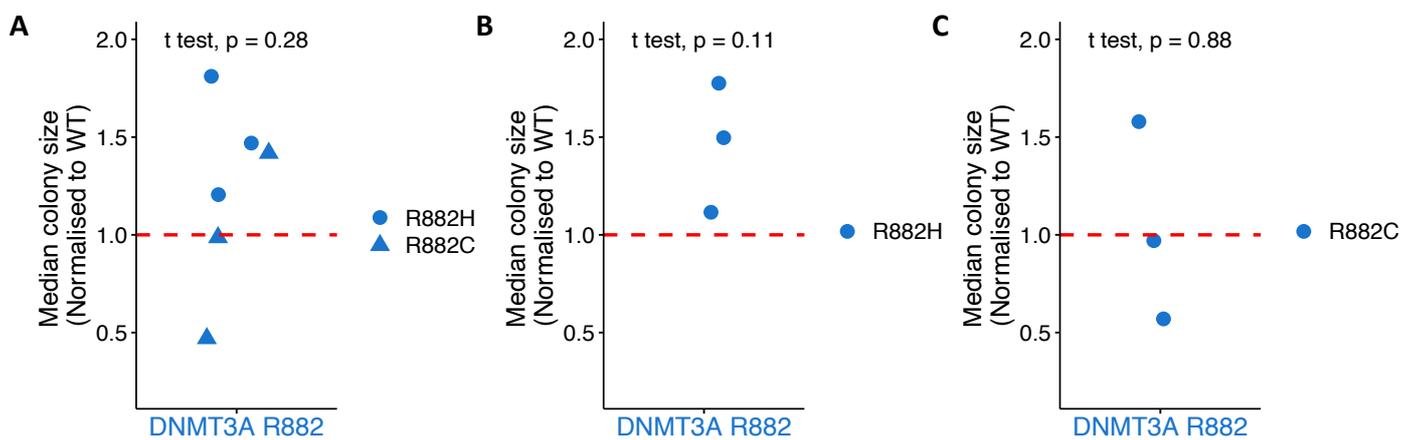
**Fig 4.3-9** Cell surface expression of markers of stemness in *DNMT3A R882* HSC/MPPs relative to WT HSC/MPPs within the same individual. (A) CD49f, (B) CD34 and (C) CD90 median fluorescence intensity (MFI) in *DNMT3A R882* HSC/MPPs relative to WT HSC/MPPs. Within each sample, MFI of cell surface markers of all single HSC/MPPs with *DNMT3A R882* mutation is shown relative to WT HSC/MPPs, significance as per one sample t-tests.

### 4.3.8 *DNMT3A R882 mutation has no effect on the proliferation capacity of HSC/MPPs*

To assess whether the overall proliferation capacity of HSC/MPPs is affected by *DNMT3A R882* mutation, the overall size of colonies produced by *DNMT3A R882* mutant HSC/MPPs was compared with WT HSC/MPP colonies using the number of cells in the live singlets gate as a measure of colony size. Since equal volumes of cells per colony were analysed, this was a uniform and reliable measure of colony size. Colonies were of similar size between WT and *DNMT3A R882* mutant HSC/MPP derived colonies (Fig 4.3-10A) across all ARCH and AML individuals. HSC/MPPs from individuals with *DNMT3A R882H* mutation produced relatively larger colonies

compared with WT HSC/MPPs from within the same individuals but no statistical significance was reached (Fig 4.3-10B), while *DNMT3A R882C* mutant HSC/MPP derived colonies are similar in size to their WT counterparts from within the same individual (Fig 4.3-10C). *DNMT3A R882* mutation thus does not affect the proliferation capacity of HSC/MPPs in this differentiation assay.

**Fig 4.3-10 Colony size of *DNMT3A R882* HSC/MPP derived colonies relative to WT HSC/MPP derived**



**colonies within the same individual.** (A) Colony size of *DNMT3A R882* HSC/MPP derived colonies relative to WT counterparts from within the same individual. Same analysis with (B) *DNMT3A R882H* and (C) *DNMT3A R882C* HSC/MPPs separated out. Total n=6, *R882H* n=3, *R882C* n=3, significance as per one sample t-tests.

### 4.3.8.1 *DNMT3A R882* mutation confers no overall erythroid/myeloid lineage bias

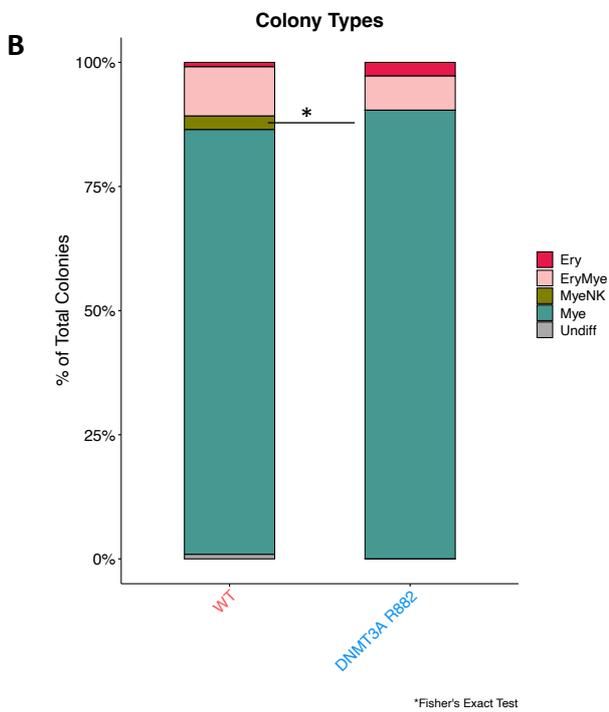
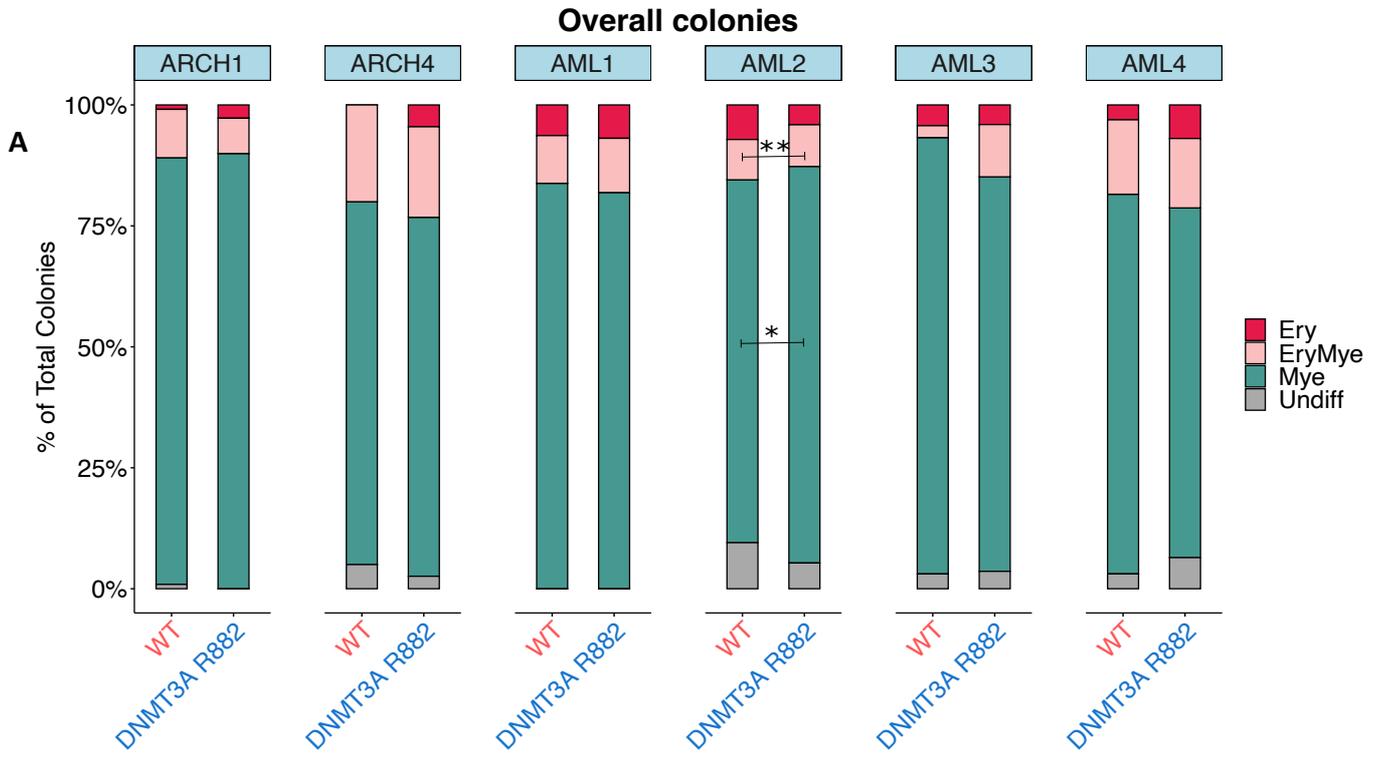
Overall differentiation capacity of *DNMT3A R882* mutant vs WT HSC/MPPs from within the same individual was assessed using flow cytometry based phenotyping of cell surface marker expression of mature cells in the single HSC/MPP derived colonies. Markers for all mature cells produced by our assay were included in the analysis (see methods, Fig 2.3-9). As it became evident that a subtle differentiation phenotype was observed within the myeloid branch (see below), more specific markers for distinguishing between increasingly mature granulocytes were added to the phenotyping panel later on in the study. The first 2 samples (ARCH1 and AML1) were assessed using a simple panel including markers for erythroid cells (GlyA), myeloid cells (CD11b), monocytes (CD14) and granulocytes (CD15). A lymphoid marker (CD56) and a megakaryocyte marker (CD41) was also included. Several panels including more specific markers for neutrophils at varying stages of maturity were included in a number of iterations of the phenotyping panel. It became evident that the differentiation assay did not yield certain mature cell types, including megakaryocytes (CD41), pre-neutrophils (CD49d) and eosinophils (CD101), as HSC/MPPs do not differentiate towards these cells under our MEM differentiation conditions. Hence the final differentiation panel was developed based on a number of test samples both from PB and CB, designed to assess

erythroid (GlyA/CD71), myeloid (CD11b), monocyte (CD14), granulocyte (CD15), mature neutrophil (CD66b) and lymphoid (CD19, CD56, CD3, CD10) differentiation markers (see methods Fig 2.3-9).

All samples predominantly produced myeloid colonies (Mye, CD45+/CD56-/CD11b+), with the proportion of colonies containing erythroid cells (Ery, CD45-/GlyA+) ranging from 10-50% in both *DNMT3A* R882 mutant and WT HSC/MPP derived colonies. In all samples, there were significantly higher numbers of Mye only or Mye-containing colonies than Ery-containing colonies, which is consistent with the observed differences between AML and healthy individuals Fig 4.3-7A). One *DNMT3A* R882C mutant AML sample (AML3) produced more Erythroid-Myeloid colonies and thus fewer Mye only in the *DNMT3A* R882 HSCs compared with WT. Within each individual in our cohort, *DNMT3A* R882 mutant and WT HSCs/MPPs produced similar proportions of colonies containing Ery and Mye cells as assessed by conventional FACS gating strategies Fig 4.3-11A, see Methods Fig 2.3-9 for gating strategies).

This indicates that HSC/MPPs with *DNMT3A* R882 mutation can differentiate towards all major blood lineages as measured by our assay, to the same capacity as WT HSC/MPPs. There does not appear to be a bias or skewing towards any major lineage in our *in vitro* assay.

In ARCH1, 3% of WT HSCs/MPPs produced lymphoid (Mye-NK, CD11b+/CD56+) colonies, while none of the *DNMT3A* R882 HSC/MPP derived colonies were lymphoid from within the same individual, a statistically significant difference ( $p=0.03$ , Fig 4.3-11B). None of the other samples produced sufficient lymphoid colonies for statistical comparison. There is insufficient data to conclude that there may be a difference in lymphoid differentiation between *DNMT3A* R882 mutant and WT HSC/MPPs.



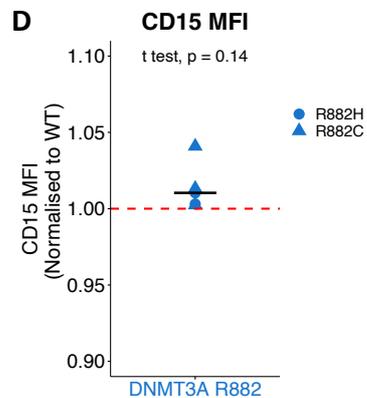
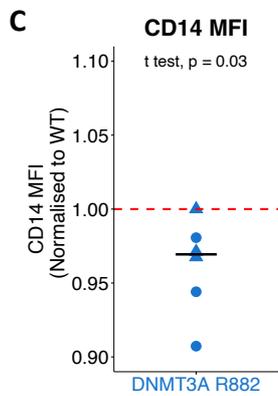
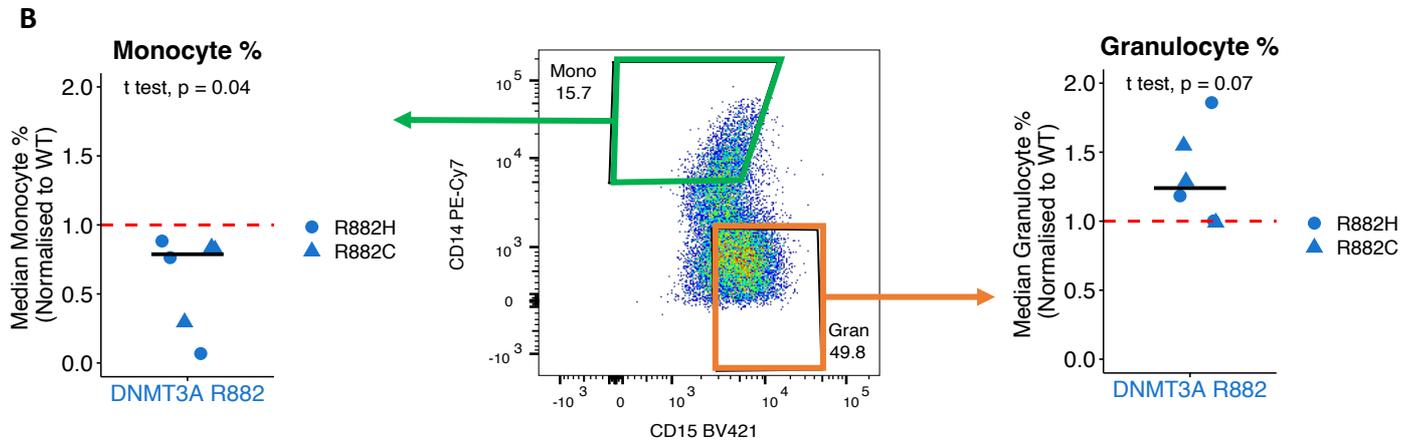
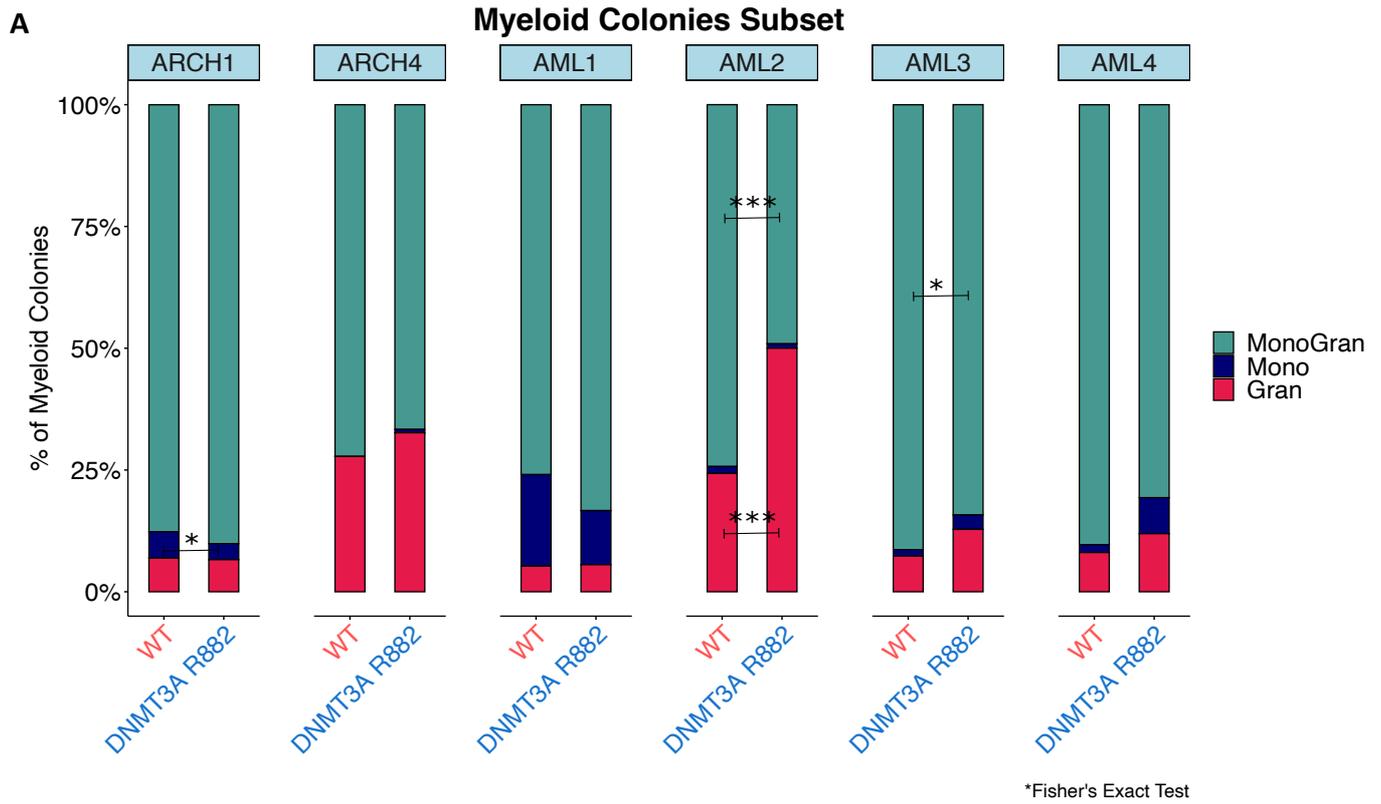
**Fig 4.3-11 Erythroid/myeloid differentiation capacity of *DNMT3A R882* vs WT HSC/MPPs.** (A) Bar graphs showing the proportions of total colonies that are erythroid/myeloid/undifferentiated colonies comparing *DNMT3A R882* mutant and WT HSC/MPPs (B) NK containing colonies (CD56+) were only produced by WT HSCs/MPPs from ARCH1 ( $p=0.03$ ), hence excluded from overall analysis in (A). Bar graphs represent proportions of all colonies containing at least 30 cells of GlyA+ (Ery), GlyA+/CD45+/CD56-/CD11b+ (EryMye), GlyA-/CD45+/CD56-/CD11b+ (Mye) or GlyA-/CD45+/CD56-/CD11b- (Undiff) cells. Significance as per Fisher's Exact test.

### ***4.3.8.2 Changes in differentiation observed within the myeloid branch***

Given the alterations in differentiation within the myeloid branch from the index ARCH1 individual, the expanded cohort of individuals were assessed for myeloid differentiation with flow cytometry using markers of monocyte (CD14) and granulocyte (CD15 and CD66b) differentiation and maturity.

#### ***4.3.8.2.1 DNMT3A R882 mutation does not affect HSC/MPP ability to produce both monocytes and granulocytes***

Using flow cytometry data, all colonies containing myeloid cells (CD45+/CD56-/CD11b+) were assigned as either monocytic (Mono, CD14+), granulocytic (Gran, CD15+) or mixed (Mono-Gran, CD14+/CD15+) colony types. Similar proportions of these colony types were produced from *DNMT3A R882* mutant and WT HSCs/MPPs from 3 of 6 samples, indicating no major differences in the capacity of *DNMT3A R882* mutant HSCs/MPPs to commit to either the Mono or Gran lineage (**Fig 4.3-12A**). In the other 3 samples, *DNMT3A R882* mutant HSC/MPPs could still commit to both Mono and Gran lineages but with some alterations in the proportions of myeloid colony types. AML2 and AML3 generated fewer mixed Mono-Gran colonies ( $p=0.0003$  and  $p=0.05$  respectively) from *DNMT3A R882* mutant HSC/MPPs compared with WT, with a greater proportion being Gran only colonies, reaching statistical significance in AML2 ( $p=0.0001$ ). The proportion of Mono colonies was lower in *DNMT3A R882* mutant HSC/MPP derived colonies compared with WT in 2 of 6 samples (ARCH1,  $p=0.02$ , and AML1,  $p=0.1$ ). These data warranted further investigation of differentiation capacity within the myeloid branch driven by the *DNMT3A R882* mutation.



**Fig 4.3-12 Differentiation capacity within the myeloid branch of *DNMT3A R882* vs WT HSC/MPPs.**

(A) HSC/MPPs with or without *DNMT3A R882* mutation can produce Mono (CD14+) and Gran (CD15+) containing colonies. Bar graphs represent proportions of myeloid (CD45+/CD56-/CD11b+) colonies containing at least 30 cells of CD14+ (Mono), CD15+ (Gran) or CD14+/CD15+ (MonoGran). (B) Percentage of monocytes (CD14+ cells) and granulocytes (CD15+ cells) within myeloid colonies from *DNMT3A R882* HSC/MPPs normalised to WT colonies from the same individual. (C) CD14 and (D) CD15 median fluorescence intensity (MFI) within myeloid cells of *DNMT3A R882* colonies normalised to WT colonies from within the same individual. N= 6 for all graphs, significance as per Fisher's exact test in (A) and one sample t-test for (B-D).

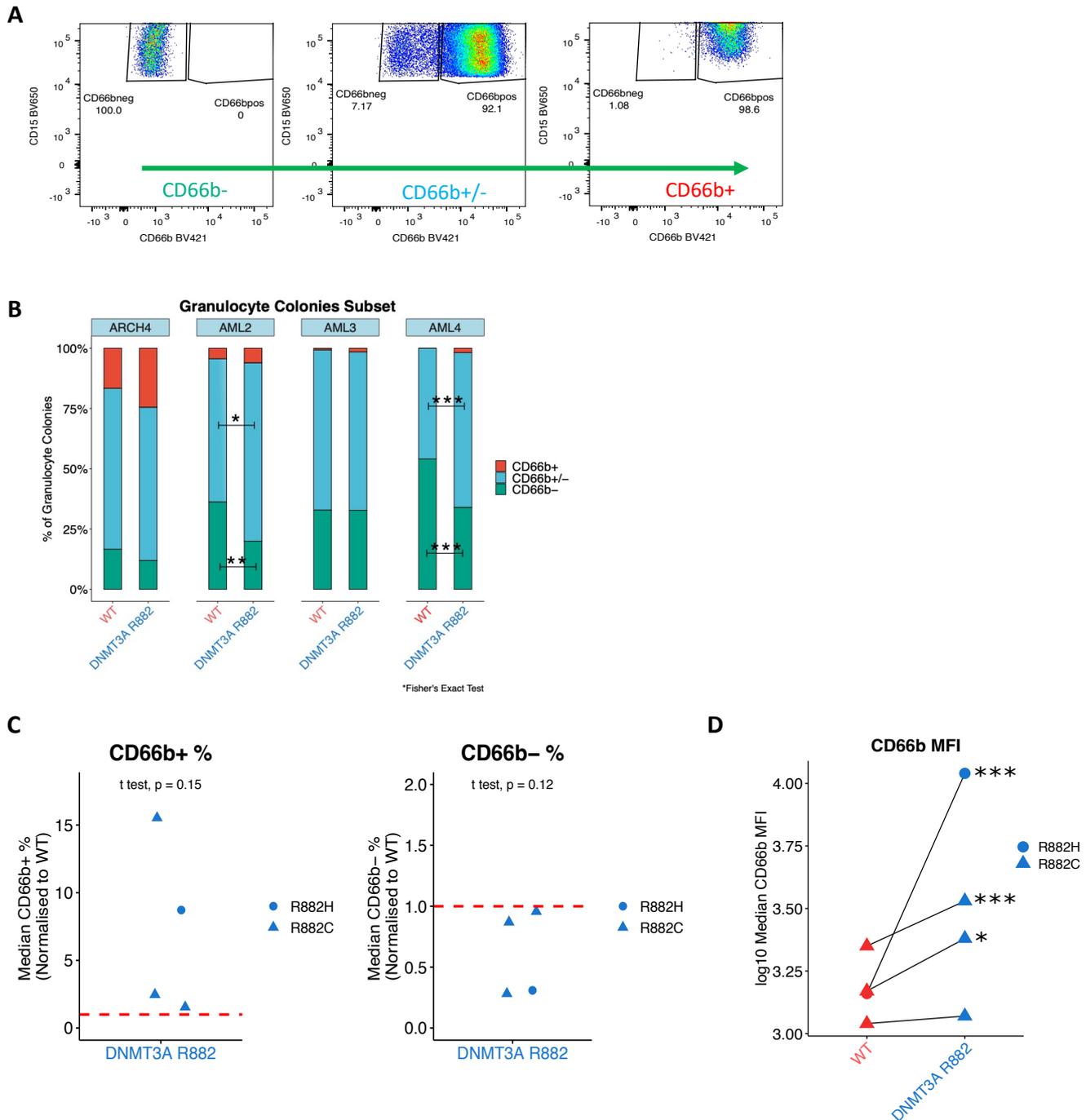
***4.3.8.2.2 DNMT3A R882 mutant HSCs/MPPs produce fewer and less mature monocytes and more neutrophils of higher maturity***

Myeloid differentiation was further explored by initially quantifying the proportions and MFI of monocytic (CD14+) and granulocytic (CD15+) cells within each colony. The percentage of CD14+ cells (Mono) within Mye colonies (CD45+/CD56-/CD11b+) was overall lower in *DNMT3A R882* mutant HSC/MPP derived colonies compared with their WT counterparts; in contrast, the percentage of CD15+ cells (Gran) within Mye colonies was higher (**Fig 4.3-12B**).

Maturity of blood cells correlates with median fluorescence intensity (MFI) of cell surface markers, especially CD14 and CD66b for monocytes and neutrophils respectively<sup>279,280</sup>. Differences in MFI of relevant cell surface markers within HSC derived colonies may identify changes in differentiation dynamics within the myeloid branch conferred by *DNMT3A R882* mutation at the HSC/MPP level. Analysis of MFI of myeloid cell surface markers showed lower CD14 MFI within Mye cells (CD45+ CD56- CD11b+) derived from *DNMT3A R882* mutant HSC/MPPs compared to those derived from WT HSC/MPPs, reaching statistical significance in 4 of 6 individuals, and globally significant when the CD14 MFI of *DNMT3A R882* mutant HSC derived colonies are normalised to WT colonies ( $p=0.03$ , **Fig 4.3-12C**). CD15 MFI was higher in *DNMT3A R882* mutant HSC/MPP derived colonies than WT colonies in 2 individuals ( $p=0.0001$  in AML2 and  $p=0.02$  in AML4), without reaching global significance across all samples (**Fig 4.3-12D**).

To further explore this altered differentiation pattern within the myeloid branch, granulocytic cells (CD15+) from either Gran or mixed Mono-Gran colonies were further analysed for expression of CD66b, a marker of more mature neutrophils, in 4 samples (ARCH4, AML2, AML3, AML4, **Fig 4.3-13A**). Fewer colonies containing only immature neutrophils (CD15+/CD66b-) were observed in the *DNMT3A R882* mutant HSC/MPP derived colonies in 2 of 4 samples in which this surface marker was assessed (**Fig 4.3-13B**). In all 4 samples, the proportion of CD66b+ cells within the CD15+ population was higher in the *DNMT3A R882* mutant colonies compared with WT colonies, while the proportion of CD66b- cells was lower without reaching statistical significance (**Fig 4.3-13B**). 3 of 4 individuals for whom CD66b was assessed had significantly higher CD66b MFI in CD15+ cells from *DNMT3A*

R882 mutant colonies compared with WT colonies (Fig 4.3-13C). These findings corroborate those of the colony proportions shown in Fig 4.3-12A. Overall, these analyses indicate that *DNMT3A* R882 mutation in HSC/MPPs leads to more efficient differentiation into mature neutrophils but less efficient monocyte differentiation *in vitro*.



**Fig 4.3-13 Expression of CD66b in granulocyte colonies to compare neutrophil maturity between *DNMT3A R882* and WT colonies within the same individual.** (A) Example of gating strategy showing progression of immature (CD66b-) to mature (CD66b+) neutrophils in HSC/MPP derived colonies. (B) Comparison of the proportions of granulocyte (CD45+/CD14-/CD15+) colonies containing at least 30 cells of immature (CD66b-), mixed (CD66b+/CD66b-) or mature (CD66b+) neutrophils derived from *DNMT3A R882* vs WT HSC/MPPs within each individual. (C) Proportion of CD66b+ cells (left) and CD66b- cells (right) within *DNMT3A R882* HSC/MPP derived granulocyte colonies normalised to WT granulocyte colonies within each individual. (D) Comparison of CD66b median fluorescence index (MFI) of CD15+ granulocyte cells between *DNMT3A R882* and WT derived HSC/MPPs within each individual, significance represents Fisher's exact test in (B), one sample t-test in (C) and two-tailed t-test in (D). n=4 for B-D.

#### *4.3.8.2.3 DNMT3A R882 mutant HSC/MPPs exhibit altered myeloid differentiation dynamics compared with WT HSC/MPPs*

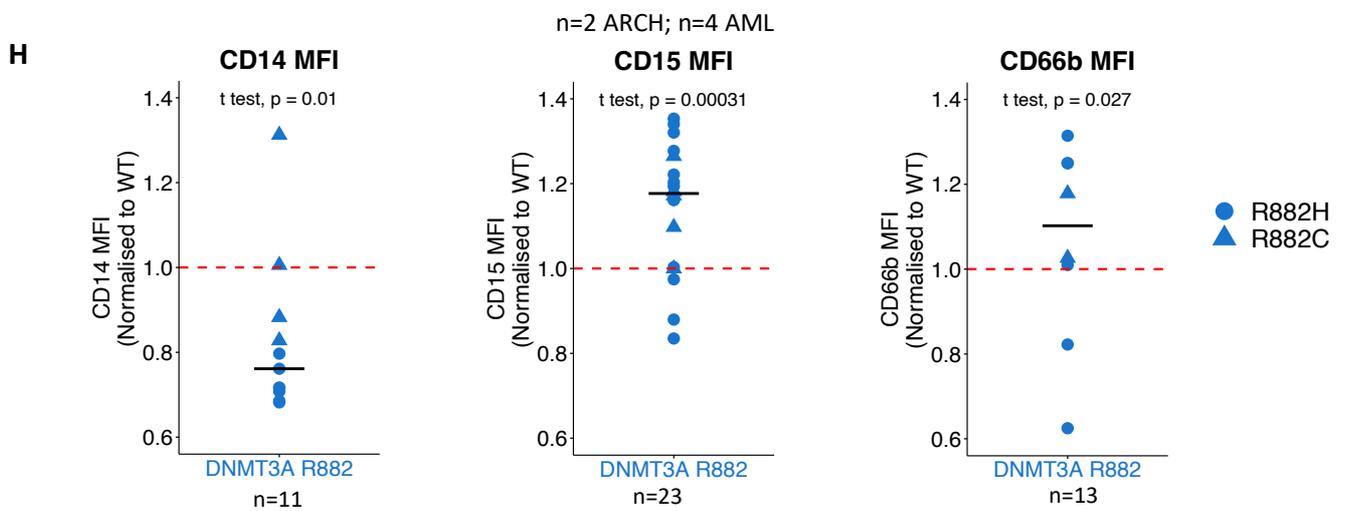
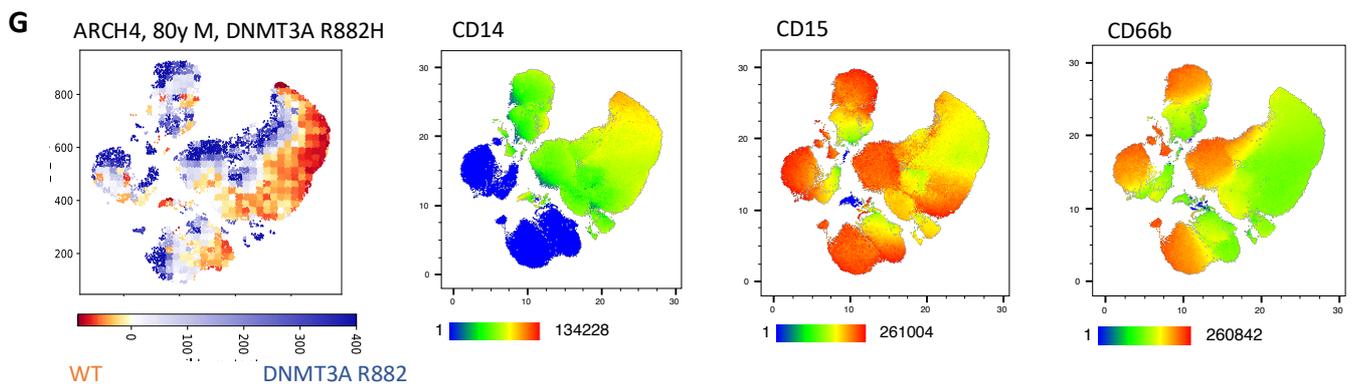
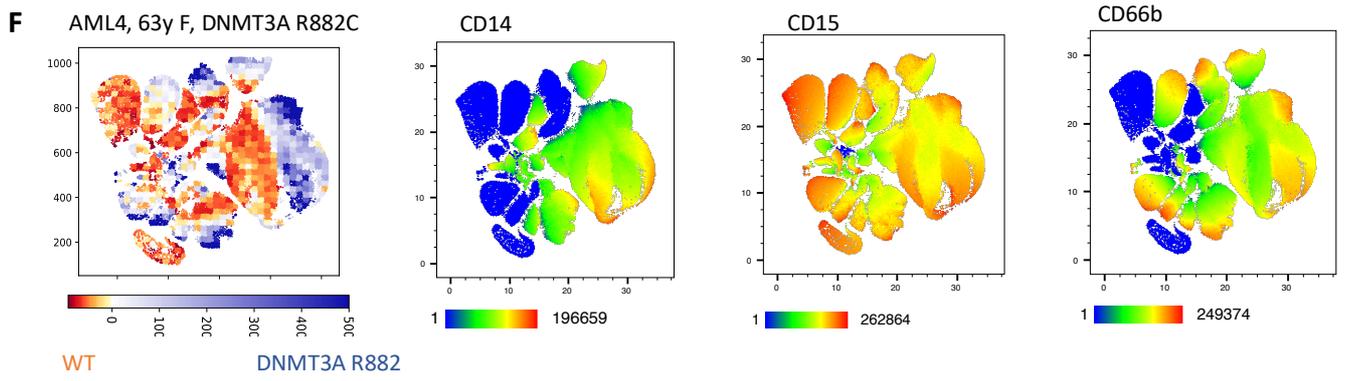
I then sought to validate the above findings using two bioinformatic analysis pipelines (FlowPAC and CytoTree), which assess distribution of cell surface marker expression across all mature cells pooled *in silico* from all HSC/MPP derived colonies within each individual sample. These unbiased tools were also used to understand differentiation dynamics of HSC/MPP differentiation by identifying the altered distribution of cell surface markers determined by flow cytometry across the *DNMT3A R882* mutant vs WT HSC/MPP landscape in an unbiased manner. This is a powerful unbiased approach that has been adapted for this indication based on previously available bioinformatic packages (CytoTree, R/Bioconductor<sup>281</sup>), or developed specifically for this purpose (FlowPAC, established by Daniel Hayler).

In FlowPAC, data extracted from flow cytometry based phenotyping of all mature cells derived from all single HSC/MPPs *in vitro* was pooled bioinformatically, analysing each individual ARCH or AML sample separately. Downsampling was performed so that 1000 cells were included for each colony. These data were used to perform dimensionality reduction and generate UMAPs based on the intensity of 7-8 different fluorochromes representing cell surface markers for mature haematopoietic cell types, and clustering was performed with the FlowSOM package (Fig 4.3-14A). Once these UMAPs were defined, the genotype of each cell within the UMAP was identified using data previously generated from targeted DNA sequencing of individual HSC/MPP derived colonies. Each UMAP was divided into 1600 equally sized square areas and the ratio of *DNMT3A R882* mutant:WT cells was calculated within each area of the UMAP. Individual fluorochrome intensities were also displayed on the UMAP to analyse each mature cell marker precisely and manually assign identities to each of the clusters (see Methods 2.7.2.1 for more detail on methodology).

First we qualitatively assessed which fluorochromes had higher MFIs in the areas of the UMAP where the *DNMT3A R882*:WT ratio was highest (blue) or lowest (red, left panels of Fig 4.3-14B-G). For all 6 individuals,

CD14 intensity was higher in areas of the UMAP where WT cells were more abundant than *DNMT3A R882* mutant cells (red areas; **Fig 4.3-14B-G**). This observation is consistent with the finding of fewer and less mature monocytes seen in *DNMT3A R882* mutant HSC/MPP derived colonies in the more conventional analysis of flow cytometry based phenotype above. CD15 expression was fairly widespread across the entire UMAP for all 6 individuals. In contrast, in the 4 samples in which CD66b was measured as a marker of neutrophil maturity, CD66b intense areas of the UMAP visually matched those with high *DNMT3A R882* mutant:WT ratio within all 4 individuals (**Fig 4.3-14D-G**).





**Fig 4.3-14 FlowPAC analysis showing association between CD66b, CD15, CD14 and *DNMT3A R882* mutant:WT cell ratio.** Pooled in silico bioinformatic analysis of all mature cells derived from HSC/MPPs within each individual separately, with each colony downsampled to 1000 cells (see Methods 2.7.2.1). (A) Representative UMAP from AML2 showing separation of cells into clusters based on intensity of cell surface marker expression. (B-G)) UMAP overlays showing distribution of *DNMT3A R882H:WT* ratio (left) and intensity of CD14, CD15 and CD66b fluorochromes of mature cell surface markers overlaid on to the same UMAP for each individual sample (right). (H) Expression of CD14, CD15 and CD66b MFI in *DNMT3A R882* cells normalised to WT cells within the same cluster, derived from all 6 individuals. Clusters selected show high expression of the relevant fluorochromes. n = number of clusters, across 6 samples for CD14 and CD15 and 4 samples for CD66b. P values represent one sample t-test. UMAPs generated by Daniel Hayler.

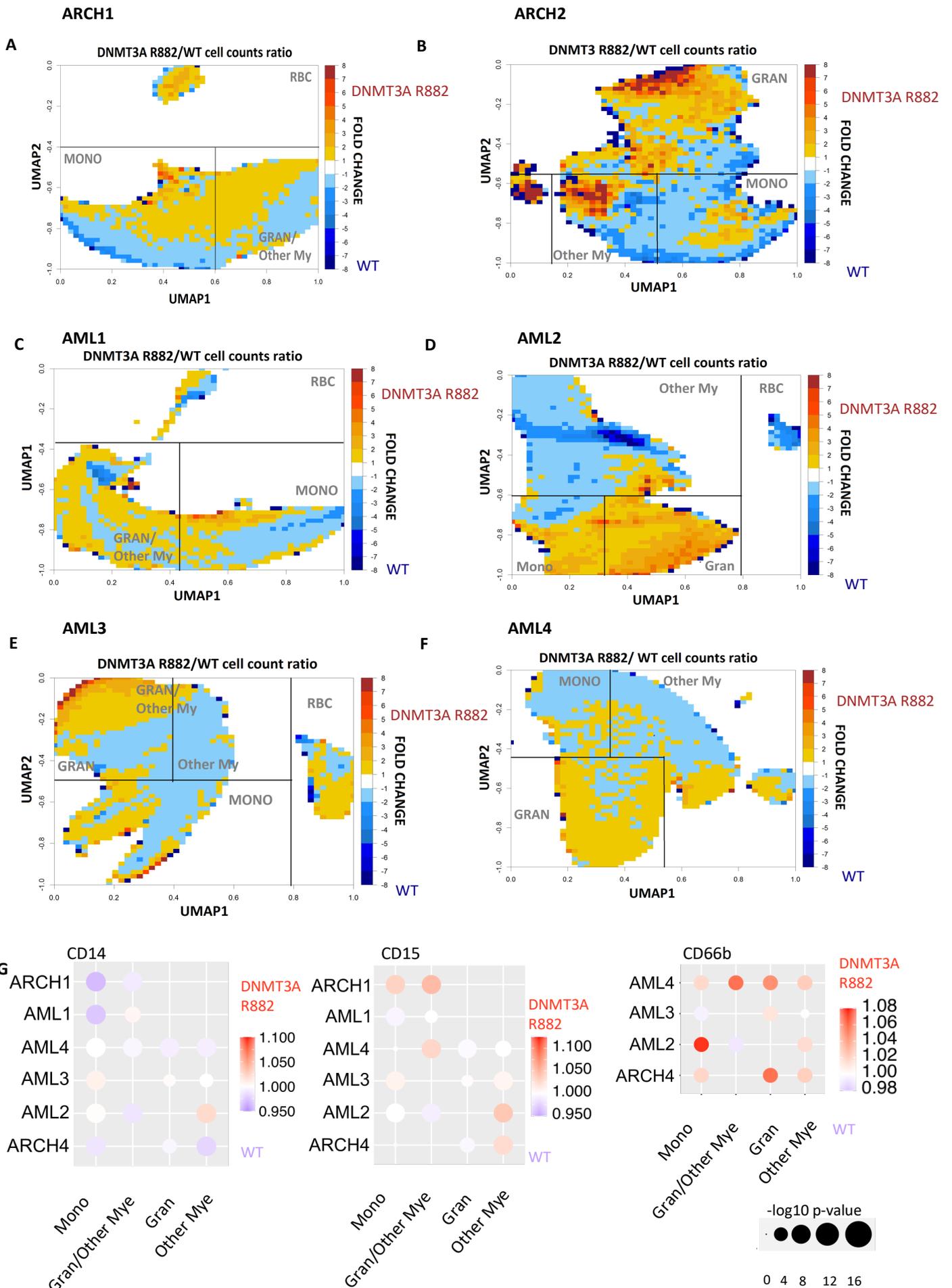
To measure this effect quantitatively, the MFI of the 3 fluorochromes of interest, CD14, CD15 and CD66b, was measured in selected clusters where these markers had high expression within each individual UMAP. The MFI of each fluorochrome within each relevant cluster in *DNMT3A R882* mutant HSC/MPP derived cells were normalised to MFI of the same fluorochrome in WT cells within the same cluster. The normalised expression of each fluorochrome within each relevant cluster across all 6 individuals shows a globally statistical difference. CD14 MFI is lower in *DNMT3A R882* mutant HSC/MPP derived cells (t test,  $p=0.01$ ), while CD15 and CD66b MFIs are higher (t-test,  $p=0.0003$  and  $p=0.02$  respectively, Fig 4.3-14H).

These results were reproduced using another unbiased bioinformatic approach based on Cytotree, a package available in R/Bioconductor. UMAPs were generated using fluorochrome intensity within each individual from all mature cells pooled from all HSC/MPP derived colonies as above but in this instance without downsampling. The UMAP was divided into 250 equally sized squares and the *DNMT3A R882* mutant:WT ratio within each square was calculated. To define which domain of the map corresponded to each mature cell type, colony assignment using conventional gating strategies as outlined above in 2.7.2.2, was overlaid on to these UMAPs, identifying 3-4 distinct differentiated domains in these maps. Qualitative visual assessment again showed that areas containing Gran (CD15hi or CD66bhi) cells largely correspond to areas of the UMAP that have a high *DNMT3A R882* mutant:WT ratio (red, Fig 4.3-15A-F).

Quantitative analysis using this method showed that within each individual, the overall expression of the fluorochromes of interest concurred with our findings using other methods above. CD14 MFI is significantly lower in *DNMT3A R882* cells than in WT cells in at least one domain for all 6 samples analysed. CD15 and CD66b MFI are significantly higher in *DNMT3A R882* mutant cells than in WT cells in at least one domain for all 6 samples analysed (Fig 4.3-15G).

Both in silico pooled approaches unbiasedly identified significantly lower expression of the monocyte marker CD14 and/or stronger expression of the mature neutrophil marker CD66b globally in mature cells derived from

*DNMT3A R882* HSC/MPPs than in those derived from WT HSC/MPPs. Hence, I conclude that *DNMT3A R882* mutation alters the differentiation dynamics of HSC/MPPs reducing mature monocyte production (CD14+) and increasing mature neutrophil (CD15+/CD66b+) generation *in vitro*.



**Fig 4.3-15 CytoTree analysis showing association between cell surface markers CD14, CD15 and CD66b and *DNMT3A R882* mutant:WT ratio.** Pooled in silico analysis of all mature HSC/MPP derived cells within each individual, without downsampling of colonies (see Methods 2.7.2.2). (A-F) UMAPs generated from each individual using CytoTree using fluorochrome intensity. *DNMT3A R882* mutant:WT genotype ratio of all cells within each area of the UMAP is overlaid and mature colony types shown as assigned by conventional gating strategies (G) Bubble plots with colours showing ratio of CD14, CD15 and CD66b MFIs in *DNMT3A R882* vs WT cells within each domain as listed on the x-axis and within each individual. Size of the bubbles show p values measured by Wilcoxon test. Mono = CD14+/CD15-, Gran/Other Mye = CD15+/CD66b-, Gran = CD15+/CD66b+, Other Mye = CD45+/CD14-/CD15-/CD66b-/CD56-/CD11b+. UMAPs generated by Dr Aleksandra Krzywon.

#### 4.3.8.2.4 *DNMT3A R882* mutant monocytes are transcriptionally less mature

As an independent and agnostic assessment of the effect of *DNMT3A R882* mutation on monocytic maturation of HSC/MPPs, I performed bulk RNAseq of colonies from the ARCH1 individual containing mature monocytes without any other mature cells (CD45+/CD11b+/CD14+/CD15-/CD56-/GlyA-). 33 HSC/MPP derived colonies were included in the analysis (see Methods 2.6.3) and clustered on a PCA plot based on gene expression of 605 Highly Variable Genes (Fig 4.3-16A). Cluster 3 was small, containing only 4 colonies with equal distribution between WT and *DNMT3A R882*. It was also marked by upregulation of genes and transcription factors associated with neutrophil progenitors such as *KLF5* (data not shown)<sup>18,19282</sup>, as observed on gene set enrichment analysis (GSEA) and differential expression analysis (DESeq) assessing genes differentially expressed between the clusters in the PCA.

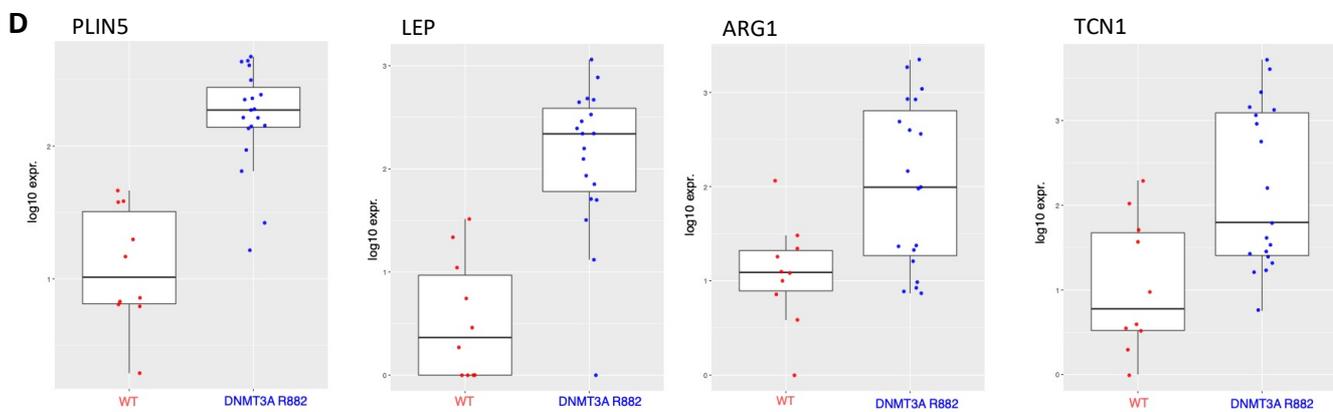
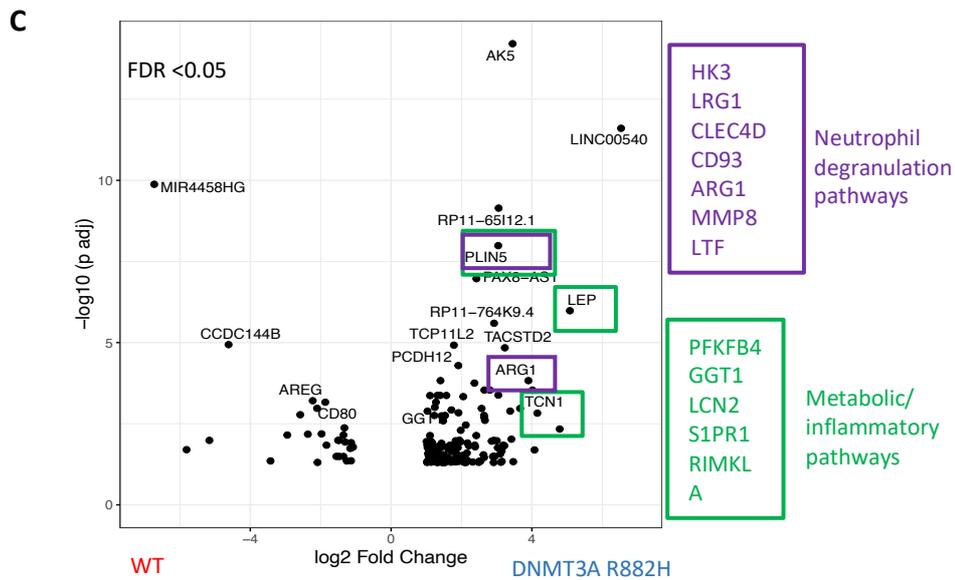
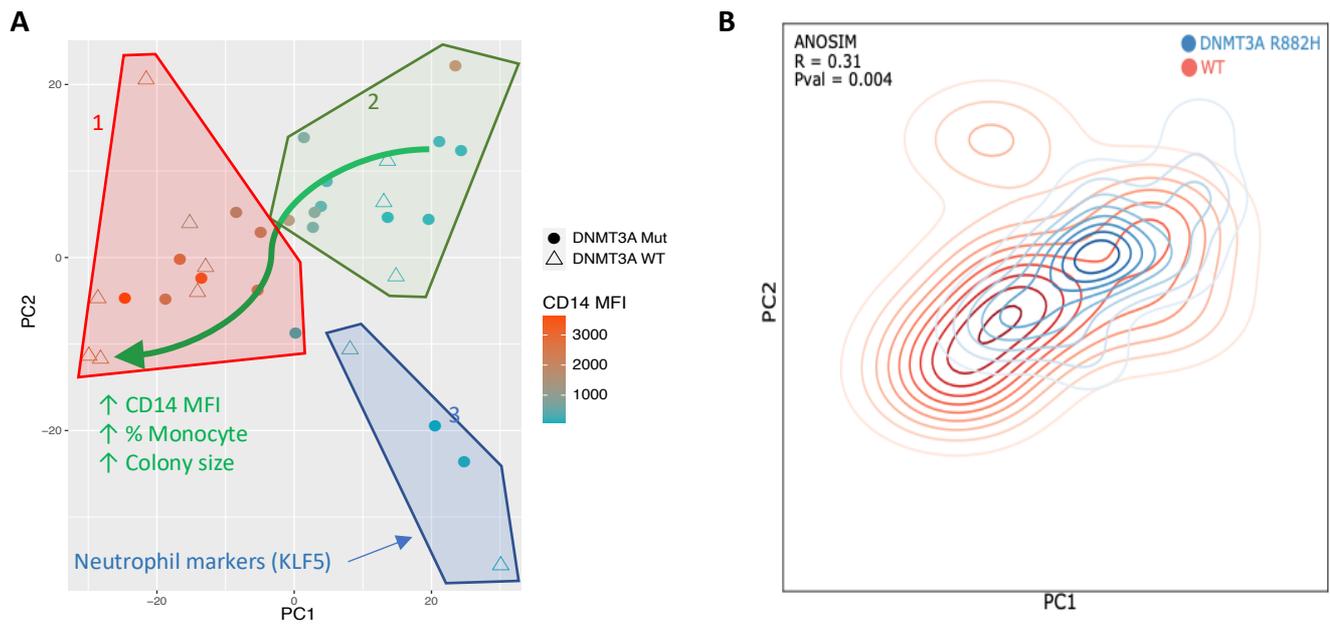
Progression of monocyte maturity based on percentage of all cells being CD14+ and CD14 MFI within each colony was identified as indicated in Fig 4.3-16A. Colony size also accompanied this progression. In an analysis of similarities (ANOSIM) test performed on the same PCA, a significant separation was observed between WT and *DNMT3A R882* mutant colonies, with the latter being significantly more represented on the less mature end of the monocyte trajectory, and with smaller colony size. This separation was quantitatively measured using the R statistic, which was 0.3, sufficiently different to 0.0 as measured by the p-value 0.004 to indicate significant separation of *DNMT3A R882* mutant and WT colonies in the PCA that did not occur by chance (Fig 4.3-16A & B).

To understand the genes and pathways that separated *DNMT3A R882* mutant and WT HSC/MPP derived monocytes, DESeq was performed between colonies derived from *DNMT3A R882* mutant and WT HSC/MPPs within clusters 1 and 2 in the PCA. 159 and 28 genes were respectively significantly (FDR<0.05) upregulated and downregulated in *DNMT3A R882* mutant monocyte colonies compared to WT. On gene set variation analysis (GSVA), monocyte colonies derived from *DNMT3A R882* mutant HSC/MPPs enriched for genesets involved in metabolic and inflammatory pathways including glycogen synthesis and metabolism (Pearson corr -0.8, p=1.46e-05 and Pearson corr -0.75, p=0.0003 respectively). Of these genesets, particular genes of interest include IL-4

and IL-13, which are associated with glycolysis, glutathione synthesis pathways and interleukin signalling. Furthermore, *PLIN5* was highly upregulated in *DNMT3A R882* mutant monocyte colonies, which is attributed to atherosclerotic/fatty acid pathways, promoting atherosclerosis<sup>283</sup>, and is associated with leptin (*LEP*), another gene that was upregulated in the *DNMT3A R882* monocyte colonies from the ARCH individual.

Several genesets identified as being enriched on GSVA in the *DNMT3A R882* mutant colonies compared with WT were visualised using the Reactome Gene Pathway Database. Most were attributed to neutrophil degranulation pathways (Pearson corr -0.63, p=0.002; examples of such genes within genesets include *HK3*, *LRG1*, *CLEC4D*, *CD93*, *ARG1*, *MMP8* and *LTF*; Fig 4.3-16C), Using the same approach, a number of genes were attributed to metabolic and inflammatory pathways, including *PFKFB4*, *GGT1*, *LCN2*, *S1PR1*, *LEP*, *TCN1* and *RIMKLA*. Differential expression of 4 representative genes from these two pathways are shown in the boxplots in Fig 4.3-16D.

Taken together, transcriptional analysis of monocytes derived from single HSC/MPPs in 1 ARCH individual substantiates previous findings that *DNMT3A R882* mutation results in less efficient monocyte differentiation. Further, genes involved in inflammation and metabolic pathways are significantly differentially expressed in *DNMT3A R882* mutant HSC/MPPs compared to WT. Further studies are required to investigate whether these differences may provide some functional understanding of how ARCH may be associated with cardiometabolic disease.



**Fig 4.3-16 : Transcriptional analysis of monocyte colonies derived from HSC/MPPs from ARCH1.** (A) PCA derived from bulk RNAseq data of 33 HSC/MPP derived colonies from ARCH1 containing only mature monocytes without granulocytes and erythroid cells (CD45+/CD11b+/CD56-/CD15-/CD14+/GlyA-). All cells within the colonies were sequenced. Clusters based on gene expression of 605 highly variable genes. (B) Analysis of similarities test to assess separation of *DNMT3A R882* and WT colonies as mapped on the PCA including clusters 1 & 2. R statistic >0 and p=0.004 indicates significant separation of *DNMT3A R882* and WT colonies in the adjoining PCA. (C) Volcano plot showing differential expression analysis comparing genes with significantly altered expression between *DNMT3A R882* and WT colonies as measured DESeq within clusters 1 & 2 only. Relevant genes identified by gene set and pathway enrichment analysis are highlighted in the boxed lists. (D) Differential expression of selected genes between *DNMT3A R882* and WT colonies included in the transcriptional analysis as measured by log10 of the expression of each gene. N= 33, box plots represent median and interquartile range.

## 4.4 Conclusions and future work

Using conventional phenotyping and targeted genotyping data from single HSC/MPP derived colonies from individuals with *DNMT3A R882* mutations, I have shown that proliferation and differentiation capacity of HSC/MPPs is not affected by this mutation. It confers no broad selective advantage *in vivo* compared with WT HSC/MPPs as measured by the correlation between the mutation burden (VAF) in PB MNCs and HSC/MPP derived colonies. HSC/MPPs with *DNMT3A R882* mutation do not have an overall bias towards myeloid/erythroid differentiation. Within the myeloid branch, there is a propensity towards neutrophil differentiation and away from mature monocyte differentiation in *DNMT3A R882* mutant HSC/MPPs.

To further elucidate a functional understanding of how myeloid differentiation is altered by *DNMT3A R882* mutation at the HSC level, monocytes and neutrophils derived from single HSC/MPPs from the samples in this cohort were sorted for single cell RNA sequencing on a number of unsuccessful attempts. A total of 3108 single monocytes and neutrophils were sorted from 499 colonies from 5 individuals. Once it was evident that 3108 sorted cells did not contain sufficient nucleic acid for RNAseq, further library preparation was abandoned in view of associated costs and minimal yield. Future work will need to optimise SMART-Seq approaches for myeloid cells.

Using PB MNCs from 3 separate individuals, our induced pluripotent stem cell (iPS) facility were able to successfully produce a total of 27 cell lines, including *DNMT3A R882* mutant and isogenic WT controls within each individual. I successfully expanded several cell lines. These will be used to further study the functional and metabolic effects of *DNMT3A R882* mutation on mature monocyte and neutrophil function, and further explore the effects of *DNMT3A R882* mutation on lymphoid differentiation.

# Chapter 5 Discussion

Haematopoiesis represents one of the most complex mammalian systems and attempts at understanding its careful orchestration has provided much insight into haematological disease states. In particular, studies looking at epigenetic modifications can unveil mechanisms by which heterogeneity across the HSC pool and indeed the haematopoietic tree diverges between health and disease<sup>47,284</sup>. Methyltransferases are a set of genes that encode for enzymes that play an important role in HSC function and span across health and disease. Clonal expansions of HSCs carrying mutations in methyltransferase genes are linked to clonal haematopoiesis and malignancy, as well as non-haematological and non-malignant disorders. Hence it is imperative to study the role of epigenetic mechanisms and in particular methyltransferases in the context of health and disease. Some frequently mutated genes that are of epidemiological and clinical significance include *DNMT3A*, *TET2*, *ASXL1* and *PRMT5*. In this thesis, two of these genes have been studied separately in either healthy or clonally expanded human HSCs.

## 5.1 PRMT5

Analysis of scRNAseq data from LT-HSCs during their exit from quiescence provided a unique data set from which to identify genes that have significantly altered expression during this critical period of HSC transition from quiescence into the cell cycle. Differential expression was evident in splicing genes and methyltransferases. The latter are of particular interest due to their role in stem cell biology and tumorigenesis, as well as the availability of pharmacologic inhibitors to allow their *in vitro* manipulation. Further, they can be readily applied to pre-clinical studies *in vivo*<sup>203,204</sup>.

I selected to further interrogate *PRMT5*, and the effects of pharmacological inhibition of this gene, because:

- a) It was significantly upregulated during quiescence exit of LT-HSCs
- b) It has an established association with haematological and non-haematological malignancies

The aims were to explore the effects of *PRMT5* inhibition on hHSC differentiation and the related potential myelotoxicity, which will allow better understanding of how to integrate this pharmacological approach into clinical practice. *PRMT5* inhibitors are currently being investigated in pre-clinical and clinical studies as potential targeted treatment for haematological and non-haematological malignancies, as outlined below. My data suggest that there is likely to be cytotoxicity related to reduced HSC proliferation from *PRMT5* inhibition as measured by reduced cell count following *in vitro* inhibition of *PRMT5* in hHSPCs.

### **5.1.1 Clinical implications**

Genes upregulated upon HSC activation may be subject to inhibition using small molecule inhibitors and/or modulation using shRNA or CRISPR technology, which may result in maintaining a state of quiescence in these

stem cells or decreased activation and differentiation. Self-renewal has been recognised as essential for tumour cell maintenance, suggesting that this process has therapeutic potential<sup>285</sup>. *PRMT5* has been implicated across solid<sup>202</sup> and haematological malignancies<sup>201,204</sup> in conjunction with other genes and pathways including *mTOR*<sup>193</sup>, *MYC*<sup>286</sup> and *MTAP*<sup>287</sup>. This property could also potentially be clinically manipulated in haematological malignancy, as *PRMT5* inhibition could maintain LSCs in quiescence, allowing relapsed/refractory patients to achieve or prolong remission status as a bridge to definitive treatment such as HSCT or CAR-T therapy. Further studies in the leukemic context are required before further implications can be inferred. This strategy has been used in Philadelphia chromosome positive leukaemia with great success<sup>288</sup>.

Based on the results of this study, I hypothesise that *PRMT5* overexpression would promote HSC differentiation, while downregulation leads to global reduction in cell proliferation. *PRMT5* is overexpressed in a number of haematological and solid malignancies and some groups have demonstrated its function in LSC survival and role as an oncogene<sup>198,199</sup>, representing an important therapeutic target. However, inhibition of this gene may result in fatal aplasia *in vivo* as observed in *PRMT5* deleted mutant mice<sup>218</sup>, further highlighting the importance of these dose dependent results of *PRMT5* inhibition in human cord blood samples as pre-clinical data to inform clinical studies evaluating the role of *PRMT5* inhibition as part of treatment or remission maintenance strategies in malignant conditions.

Jin et al.<sup>199</sup> demonstrated a positive feedback loop between *PRMT5* and *BCR-ABL* in CML cells, and an overexpression of *PRMT5* in human CML LSCs. *PRMT5* silencing by shRNA or pharmacological inhibition with the small molecule inhibitor PJ-68 in LSCs from CML patients led to decreased survival and colony formation of these cells. These results were verified by decreased self-renewal capacity of transplanted CML LSCs in *PRMT5* knockdown mice and prolonged survival of these mice. Hence *PRMT5* may provide a novel therapeutic target in LSCs, particularly in imatinib-resistant CML.

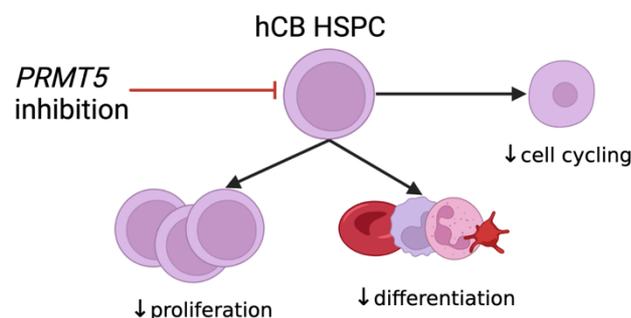
Kaushik et al.<sup>201</sup> have identified the therapeutic potential of *PRMT5* inhibition in *MLL*-rearranged AML. Their study showed that genetic silencing of *PRMT5* failed to initiate leukaemia in *MLL-AF9* expressing mice for over 120 days, while control and heterozygous *PRMT5* deleted mice died of AML within 90 days. Pharmacological inhibition of *PRMT5* with the same small molecule inhibitor used in my study, EPZ015666, led to cell cycle arrest and increased apoptosis in AML cells *in vitro*, with increased survival and delayed disease progression noted *in vivo* without a significant toxicity profile at the dose of 150mg/kg twice daily given to mice transplanted with *MLL*-AML cells. Similarly, anti-tumour activity was observed in mantle cell lymphoma (MCL) mouse models and cell death in MCL cell lines when treated with EPZ015666<sup>204</sup>. Taken together with the results of my study showing a dose dependent response in hHSCs to *PRMT5* inhibition, we can surmise that there may exist a therapeutic window for treatment with EPZ015666, at which LSCs respond to *PRMT5* inhibition with minimal HSC toxicity.

Alinari et al.<sup>225</sup> have shown that *PRMT5* is a critical driver in EBV-driven B-cell transformation, which can lead to lymphoma. Its expression in EBV-driven B cell lymphoma is limited to EBV-transformed B cells, and not seen in normal B cells, validating it as an ideal therapeutic target in this disease. Selective inhibition of *PRMT5* using the

small molecule inhibitor CMP5 has shown promising results, blocking EBV-driven B cell transformation, leading to cell death with relative sparing of normal resting B cells. I observed the same dose dependent decrease in cell growth *in vitro* with all 3 inhibitors of *PRMT5* in this study.

### **5.1.2 *PRMT5* inhibition induced decreased proliferation and differentiation block**

Pharmacological inhibition of *PRMT5* with 3 independent small molecule inhibitors results in decreased colony formation and cell proliferation, and without evidence of apoptosis using 1 inhibitor in hCB HSPCs *in vitro*. Further, there is a relative increase in the population of primitive CD34+ cells and decreased cell cycle activity, suggesting a shift towards less actively proliferating cells with pharmacological *PRMT5* inhibition (Fig 5.1-1). This reduced proliferation and relative preservation of CD34+ cells is dose dependent when hHSCs are inhibited by EPZ015666. Since the expression of this gene was upregulated with quiescence exit of LT-HSCs, these results are consistent with the hypothesis that inhibition of this gene should result in decreased cell proliferation and differentiation. These effects of pharmacological inhibition of *PRMT5* in hHSPCs, and the dose dependent nature of response has important implications for these small molecule inhibitors as a potential therapeutic agents in malignancies and for HSPC expansion, particularly in the context of its observed role in multiple cancer cell lines<sup>203</sup>.



**Fig 5.1-1 Graphical illustration of the effects of pharmacological *PRMT5* inhibition on human cord blood (hCB) haematopoietic stem and progenitor cells (HSPCs). *PRMT5* inhibition using 3 independent small molecule inhibitors results in decreased proliferation, reduced differentiation and reduced cell cycling in a dose dependent manner.**

Reduced HSPC proliferation as a result of reduced *PRMT5* activity has been reported in mouse models and now replicated in human HSPCs. In adult mice *Prmt5* KO results in fatal BM aplasia; 2 weeks following loss of *PRMT5* by poly(I:C) injection, mice develop severe pancytopenia, and BM cellularity reduces by 50% and 95% on days 7 and 15 respectively, suggesting that this gene is essential for sustaining normal adult haematopoiesis<sup>218</sup>. This effect is also observed in hHSPCs *in vitro* with almost complete paucity of colony formation and cell proliferation in cells treated with the *PRMT5* inhibitor after 14-days in MC CFC assays.

Effects of reduced *PRMT5* activity on HSPC differentiation are also broadly similar between mouse and human. Mature erythroid and myeloid cells were significantly reduced after 7 days in BM and spleens of *Prmt5*-deleted mice<sup>218</sup>, just as they were in hCB CD34+ cells in this study. Liu et al<sup>218</sup> also suggest that erythroid and megakaryocyte differentiation is more greatly affected than myeloid with *PRMT5* loss, although restricted erythroid differentiation in our assays limit assessment of this outcome in hHSPCs. While our assay does not support lymphoid differentiation, B and T cell development was differently affected than other lineages in Liu et al.'s study, with no change in frequency and proliferation of common lymphoid progenitors (CLPs) after 7 days in the BM of *PRMT5* deleted mice, and an increase in mature B cells.

Although there is overall dose dependent reduced proliferation in all hCB HSPCs treated with *PRMT5* inhibitors, there is relative preservation of the CD34+ compartment as a proportion of all cells. This effect has been observed in mice, with a transient increase in the stem-cell enriched cells (Lin<sup>-</sup>Sca<sup>+</sup>c-Kit<sup>+</sup>) and phenotypic LT-HSCs (CD48<sup>-</sup>Cd150<sup>+</sup>) after 7 days despite an overall 50% reduction in BM cellularity<sup>193,218</sup>. This was not sustained and by day 15 the number and frequency of all stem and progenitor cells was dramatically reduced<sup>218</sup>. In our hHSPCs, the relative increase in CD34+ cells with increasing *Prmt5* inhibition was maintained even after 14 days *in vitro* and was observed as early as 72 hours.

*PRMT5* was originally identified as JAK-binding protein 1<sup>289</sup>, and has been studied in myeloproliferative neoplasms, which are stem cell disorders thought to be driven by activating tyrosine kinase mutations. Oncogenic mutations within *JAK2* tyrosine kinase enhance its interaction with *PRMT5*<sup>195</sup>. Liu et al<sup>195</sup> used short hairpin RNA (shRNA) to knockdown *PRMT5* expression in hCB CD34+ cells achieving 60-70% knockdown. Downregulation of *PRMT5* led to increased colony formation and overexpression resulted in significant reduction in colony formation as well as erythroid differentiation. There was a mild reduction in CD11b+ cells. These results are in contrast to those of my study, where the effects of *PRMT5* inhibition using a small molecule inhibitor showed significant reduction in colony formation and cell proliferation in all mature blood cells including erythroid, myeloid and granulocytic/monocytic cells. This was consistently shown in a reproducible dose dependent manner across HSCs, progenitors and CD34+ hCB cells using 3 independent small molecule inhibitors. A possible explanation for these contrasting results may be the differential role of *PRMT5* in mouse and human stem cells. However, since my study was performed on hHSCs, it is possible that the *PRMT5* inhibitor does not reflect a true gene knockdown model, which can be more reliably reproduced using gene modulation with CRISPR technology and/or shRNA knockdown.

### **5.1.3 Mechanistic insights**

I have shown that the reduced proliferation and differentiation of HSPCs in response to *PRMT5* inhibition cannot be attributed to apoptosis. The same outcome was observed by Liu et al<sup>218</sup>, who reported the absence of a molecular signature consistent with apoptosis when comparing differentially expressed genes between control and *Prmt5*-null HSPCs. Although Liu et al propose that apoptosed cells may have been cleared too rapidly *in vivo* to be detected, the distinct lack of apoptosis was noted at early (24 hours), mid (72 hours) and late (7 days) time

points in my data, indicating this is not the explanation for reduced proliferation of HSPCs. However, Tan et al<sup>193</sup> observed increased HSC apoptosis in response to p53 activation when *Prmt5* activity was reduced in conditional KO mouse models and when EML cell lines were treated with the *PRMT5* inhibitor EPZ015666. This group also showed that reduced *PRMT5* activity through KO or inhibition leads to altered splicing of DNA repair genes and oxidative DNA damage, triggering p53-induced apoptosis. This was not consistent with my findings, nor with the mouse data previously discussed, hence p53 activation is the more likely explanation for these findings by Tan et al<sup>193</sup>.

Consistent with my findings, Liu et al. showed a dramatic decrease in the percentage of Lin<sup>-</sup> cells in the S phase after *Prmt5* deletion using *in vivo* BrdU assays, although the effect was different between stem and progenitor cells, with more stem cells and fewer progenitors observed in the S phase<sup>218</sup>. As the BrdU assay was performed on HSPCs in this study, the differential effect on cell cycle activity between stem and progenitor cells is not elucidated. These results, together with the results of my study suggest that there may be a differentiation block at the progenitor level with *PRMT5* inhibition. The increased BrdU and more positive Ki67 staining of *Prmt5*-null HSCs as shown by Liu et al. suggests there is a disturbed quiescent state in these cells.

Reduced cell cycling is indicative of less actively proliferating cells, which are usually more quiescent. They have a higher repopulation capacity *in vivo* with more robust and durable transplants. Assessing engraftment capacity and durability of engraftment in mice of *PRMT5* inhibitor treated HSPCs versus control HSPCs would help understand the implications of the findings outlined in this thesis.

Liu et al<sup>218</sup> also showed that *PRMT5* has a role in the regulation of cytokine signalling and cell surface expression of cytokine receptors, as shown by a significant reduction in FLT3, IL-6 and IL-3 receptor expression in *PRMT5* null HSPCs. This function of *PRMT5* deletion may be partly responsible for the decreased proliferation of HSPCs and studying hHSPCs under stress and inflammatory conditions such as with inflammatory cytokines (IL-6 and IL-3) with and without *PRMT5* inhibitor treatment may provide further insights into this as a possible mechanism of the reduced proliferation phenotype reproduced by 3 independent selective *PRMT5* small molecule inhibitors.

### ***5.1.4 Validation of phenotype by genetic modulation in hHSPCs***

I attempted genetic modulation of *PRMT5* in hCB CD34<sup>+</sup> cells and LT-HSCs with *PRMT5* LV overexpression (LVOE). Although overexpression of *PRMT5* using this approach did not reveal any statistically significant phenotypes in my study, the trend from the results of single biological replicate was consistent with what would be expected given the previous pharmacological inhibitor results. The lack of significance and hence reliability of this data is likely related to the low transduction in all assays after 3 days in culture. These studies warrant repeating after optimisation of the LV vectors.

## 5.2 Future studies on normal HSCs and progenitors

My research thus far has examined results of our group's scRNAseq data identifying genes expressed at each of the stages of transition from HSC quiescence to activation. Two of the significantly differentially expressed genes from quiescence to activation were studied using small molecule inhibitors to assess their functional role in the haematopoietic stem cell (HSC) proliferation and differentiation. Future examination of these selected genes, including *PRMT5*, should be done using genetic modification with LVOE, shRNA and CRISPR/Cas9. Other genes with significantly altered expression on HSC quiescence exit including *MAFF*, *DUSP1* and *GEMIN6* should also be examined for any effects they might have on HSC proliferation and function.

## 5.3 DNMT3A mutations in Clonal Haematopoiesis

It is now accepted that changes in haematopoiesis related to aging in humans can be explained by altered clonal dynamics such that a limited pool of clonally expanded HSC/MPPs predominate blood production in older individuals, with a steep increase in oligoclonality occurring after 70 years of age<sup>82</sup>. This oligoclonality correlates with changes in blood production commonly observed with aging including myeloid skewing<sup>291</sup>, cytopenias<sup>292</sup> and haematological malignancies<sup>88,293</sup>.

The incidence of myeloid versus lymphoid malignancies reverses in older individuals compared to youth<sup>294,295</sup>, with myeloid leukaemia being more common in elderly patients compared with younger adults, in whom lymphoid malignancies predominate. This observation is consistent with the myeloid skewing in normal blood production seen with aging<sup>296,297</sup>. Certain haematological conditions also mirror this myeloid skewing, including CLL<sup>278</sup>, beta-thalassemia and ET<sup>45</sup>. Further, I have elicited the same phenotype in patients with AML compared with age-matched healthy controls at any age, including a 38 year old male in this cohort. I have also demonstrated relative oligoclonality in the phylogenetic tree from the HSCs isolated from a healthy 63 year old female compared with an age matched control without CH. This is evidence that haematological disorders, particularly malignant conditions, likely emulate the phenotype observed in haematopoiesis with increasing age in humans. Further evidence to support this hypothesis is the finding that telomere length, reduction of which is a hallmark of aging, is shortened in AML<sup>298</sup>, CLL<sup>299</sup> and CML among other haematological malignancies<sup>300</sup>. Together, these observations suggest that haematological malignancies hasten the changes in blood production that occur with aging, or rather aging is an inevitable progression towards haematological disease.

The most common drivers of CH occur in genes that regulate epigenetic modification, particularly methylation. *DNMT3A* is responsible for DNA methylation, while *TET2* and *ASXL1* regulate DNA hydroxymethylation and histone methylation respectively. Methyltransferase genes recurrently drive CH and increase the risk of haematological malignancy. However, there are some questions relating to the mechanism by which these recurrently expressed genetic mutations increase risk of disease that I would like to address using published literature and my own data as evidence:

1. How and why do these recurrent mutations drive CH, outcompeting other mutations? I use published literature to address this below.
2. What effect do they have on human HSC/MPP function, including proliferation and differentiation? Functional data in my work provides insights into this.
3. How are the differentiation phenotypes induced by these mutations related to the clinical phenotypes associated with CH? Using what is known about CH related conditions, some inferences can be made, and some information can be gleaned from my RNAseq data, while further questions relating to this are yet to be addressed.

I have studied the inherent effects of *DNMT3A R882* mutation on haematopoiesis at the HSC/MPP functional level, which complements our current knowledge of this mutation's effects on mature cell function<sup>301</sup>, and our understanding of the transcriptional priming of HSPCs related to *DNMT3A* mutation<sup>274,275</sup>.

Since individuals with CH have normal blood counts, estimating the true prevalence of CH in the general population is challenging, and studying the effects of CH mutations on haematopoiesis in humans is even more challenging. As there are currently no consensus guidelines on the monitoring and management of patients with CH, there is no indication for screening such individuals at present. As a result, most studies in CH involve patients with abnormal blood counts or haematological disease such as MPN<sup>239</sup> and MM<sup>274</sup>, and to this end I studied patients with AML and *DNMT3A R882* mutations. To truly delineate pre-malignant, healthy HSC/MPPs harbouring CH mutations, single cell studies are critical. In this study I have used a robust method using 3 independent strategies to isolate healthy human HSC/MPPs from PB and BM and genotyped their progeny to study the functional effects of *DNMT3A R882* mutation versus WT at the HSC/MPP level.

To date, no comparisons have been made directly between *DNMT3A R882H* and *R882C* mutations regarding self-renewal, proliferation and differentiation capacity either functionally or transcriptionally. Transcriptional studies have combined these two mutations as *R882*<sup>274,275</sup>, and mechanistic studies have only investigated *R882H*<sup>302,303</sup>, hence there is a lack of understanding of whether these two mutations behave differently. Data from 3 *R882H* and 3 *R882C* individuals in my data suggest no difference in any of the phenotypes described in my report, though expanding this cohort is necessary to further investigate this effect.

### ***5.3.1 Early acquisition in life and the haematopoietic hierarchy***

*DNMT3A* is the most common mutation seen in CH<sup>88,93</sup> and *DNMT3A R882*, the most common *DNMT3A* mutation, is a dominant negative mutation, which determines hypomethylation<sup>304</sup>. It is acquired early in life<sup>82,239</sup>, long before a clinical phenotype is present, much like the *JAK2-V617F* mutation precedes the diagnosis of MPN<sup>305</sup>. I have further demonstrated that even in the presence of other subsequently acquired mutations such as *TET2*, *JAK2* and *ASXL1*, *DNMT3A R882* can lead to oligoclonality of blood production compared to age matched

peers who do not harbour this mutation in their HSC/MPPs, whilst maintaining a normal blood count phenotype. Acquisition early in life, even *in utero*, appears to be a consistent feature of this gene, yet it does not feature in childhood conditions, and earlier acquisition correlates with increased risk of haematological disease<sup>239</sup>. It always requires a cooperating second mutation such as *NPM1* or *FLT3* to cause malignancy yet can lead to cardiovascular and metabolic disease independent of other mutations<sup>95</sup>.

*DNMT3A* mutation affects HSC/MPP subtypes equally, as suggested by the equal distribution of cell surface markers for HSC/MPPs (CD90, CD34 and CD49f) between *DNMT3A R882* mutant and *DNMT3A* WT cells during single cell sorting by FACS. *Dnmt3a* expression has been shown to be highly enriched in the most primitive LT-HSCs compared to progenitors and differentiated cells in mice<sup>306</sup>. *Dnmt3a* mutation also leads to significant expansion of the HSC pool in secondary mouse recipients of *Dnmt3a-null* HSCs compared WT HSCs, while maintaining equal representation in PB<sup>306</sup>. Taken together, this is evidence that this mutation is acquired in the most primitive cells and penetrates to all blood cell subtypes from the HSC pool to mature blood cells. Of note, in two of our individuals, ARCH1 and AML1, the expression of CD49f, a cell surface marker for the most primitive stem cells<sup>10</sup> (LT-HSCs) was higher in *DNMT3A R882* mutant HSCs than in WT. Given that *DNMT3A R882* mutation is observed in all mature blood lineages in ARCH individuals with normal blood counts<sup>114</sup>, we can conclude that *DNMT3A* mutation does indeed occur at the stem cell level and that *DNMT3A R882* mutant HSC/MPPs are capable of producing cells within all major blood lineages. Limitations of my study mainly lie in the small numbers and studying other patients in this transition between ARCH and malignancy would help to address this.

### **5.3.2 Equal contribution to erythroid/myeloid lineages**

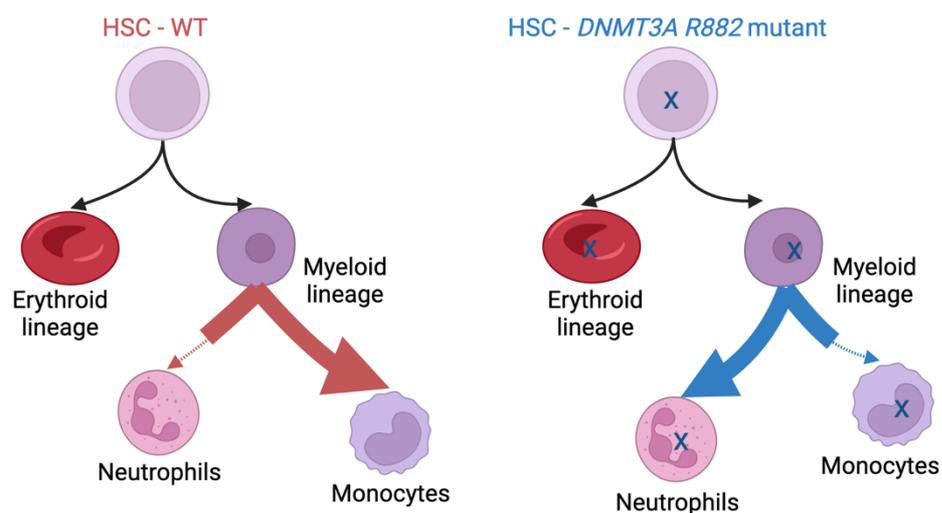
I have demonstrated a lack of any erythroid/myeloid bias due to *DNMT3A R882* mutation in HSC/MPPs in my cohort of 6 individuals. Our unbiased analysis pipelines (FlowPAC and CytoTree) also confirmed this finding using pooled analysis of all mature cells arising from single HSC/MPPs. Given the aforementioned ability of *DNMT3A R882* mutant HSC/MPPs to produce all mature blood lineages, our functional observations are not surprising. Some mouse and more recently human data has been published on transcriptional changes cause by *DNMT3A R882* mutation in HSC/MPPs<sup>274,275</sup>. *Dnmt3a KO* mice show transcriptional skewing towards erythroid and away from certain monocytic progenitor subtypes<sup>275</sup>. Other DNA methyltransferases including *Dnmt1* in mouse HSCs have been shown to confer significant differentiation bias away from lymphoid and towards myeloerythroid progeny<sup>180</sup>. Impaired erythroid differentiation and myeloid bias in BM of *Dnmt3a R878H* mutant mice has also been previously reported<sup>248</sup>. Although this difference was not observed between *DNMT3A R882* mutant and WT HSCs in our *in vitro* assays, the myeloid skewing seen overall in the AML patients in comparison to age matched healthy controls in my cohort suggests that myeloid skewing might be a disease phenomenon rather than being related to *DNMT3A R882* mutation *per se*, and this is supported by a similar phenotype seen in other haematological conditions in the absence of *DNMT3A* mutation<sup>45,278</sup>.

More recently, acknowledging the challenges of correlating genotypes and phenotypes within the heterogeneous human HSPC pool, other groups have adopted a single-cell sequencing approach to capture

genotype, and transcriptional data in progenitors from individuals with CH<sup>274</sup>. Nam et al<sup>274</sup>. used this methodology to demonstrate a myeloid bias in *DNMT3A* mutant progenitors and an expansion of megakaryocytic-erythroid primed progenitors as a result of *DNMT3A* mutation. While transcriptional priming at the progenitor level is a helpful indicator of the effects of CH, our data directly measures the differentiation capacity of *DNMT3A* mutant HSCs compared with WT HSCs *in vitro*, showing that myeloid-erythroid differentiation is not altered by *DNMT3A* mutation at the HSC level. The media in our experiments does not promote lymphoid differentiation, thus limiting direct comparisons between the transcriptional priming described in Nam et al. and actual differentiation capacity *in vitro*.

### 5.3.3 Novel monocyte/granulocyte phenotype

Importantly, the propensity of *DNMT3A R882* mutant HSC/MPPs to produce more mature neutrophils and less mature monocytes has not been previously shown (Fig 5.3-1). This is a consistent finding across our 6 independent samples from ARCH and AML individuals, across age and gender groups and independent of *R882H* or *R882C* mutation within the *DNMT3A* gene. This phenotype was demonstrated using conventional flow cytometry gating and analysis of CD14/CD15 proportions and MFI. Further, 2 independent bioinformatic analysis pipelines also confirmed this finding. These provide an unbiased examination of all the mature cells produced by all the HSC/MPPs isolated from an individual, mapping the cell surface marker expression to the genotype of each individual cell, hence are a novel application of this technique to correlate genotype and phenotype information from haematopoietic cells. There has been some indication of altered monocyte differentiation due to *DNMT3A* mutation from the analysis of transcriptional data from mouse HSCs<sup>275</sup>, but no other identification of altered monocyte/neutrophil differentiation has been reported in the literature in HSCs due to this mutation.



**Fig 5.3-1 Graphical illustration of the main effects of *DNMT3A R882* mutation on human hematopoietic stem cell and multipotent progenitor (HSC/MPP) differentiation. *DNMT3A R882* mutant HSC/MPPs have no erythroid/myeloid lineage bias compared with wild type (WT) HSC/MPPs. However, *DNMT3A R882* mutant HSC/MPPs exhibit a propensity towards neutrophil differentiation**

and away from monocyte differentiation compared with WT HSC/MPPs. X represents *DNMT3A R882* mutation.

We show that *DNMT3A R882* mutation in the absence of clinically evident disease has effects on the stem cell pool that are consistent with early changes that could lead to the overt phenotype we see in haematological malignancy. This phenotype that we have identified may lay the foundations for the clinical manifestations of ARCH, in particular the monocyte differentiation block seen in AML. Neutrophilia may also be associated with the inflammatory conditions linked with ARCH such as atherosclerosis, osteoporosis and more recently Chronic Obstructive Pulmonary Disease (COPD).

Examining the United Kingdom (UK) biobank data from ARCH patients, an association was observed between *DNMT3A* mutations and elevated neutrophil count in healthy individuals with *DNMT3A R882* mutation with VAF >2% in PB. This association was statistically significant for individuals with all *DNMT3A* mutations but not the *R882* subset alone due to limited numbers (personal communication, Prof George Vassiliou). We intend to further interrogate this association by expanding our cohort to other databases outside the UK. A clinically detectable association may provide a PB count marker that could be followed in ARCH individuals as a surrogate for *DNMT3A* mutation and provide a basis for translational studies investigating the functional effects of *DNMT3A* on haematopoietic differentiation, using my data as pre-clinical evidence.

### ***5.3.4 No selective advantage during in vivo differentiation***

The mutation burden of *DNMT3A* as measured by VAF in HSC/MPP derived colonies in our cohort of samples was the same as that in PB MNCs in the same individuals. This finding suggests that *DNMT3A R882* mutation in HSC/MPPs does not confer a selective advantage nor a differentiation block during *in vivo* differentiation. If there was a selective advantage for *DNMT3A R882* mutant HSC/MPPs during differentiation, the VAF in MNCs would be higher than in the HSC pool. However, VAF is equal in MNCs and HSC pool, despite VAF in the MNCs being artificially higher due to the presence of blasts in the AML samples. This finding holds true even when ARCH and AML samples are analysed separately, indicating the robustness of this finding. Since the progeny of an individual HSC can be assumed to carry the same mutation burden as the HSC itself, VAF in individual colonies is a reasonable approximation for VAF in the HSC pool. Positive selection of *DNMT3A* in HSCs, giving selective advantage to *Dnmt3a* mutant HSC/MPPs during serial transplantation has been observed by a number of groups in mouse models<sup>248,307</sup> and humans<sup>254</sup>, however these findings relate to increased self-renewal in *Dnmt3a* mutant HSC/MPPs. I have shown that despite the published evidence for increased self-renewal conferred by *Dnmt3a* mutation in mouse HSCs, *DNMT3A* mutation in human HSC/MPPs maintains equipoise in terms of differentiation potential without a selective advantage or differentiation block.

## 5.4 ARCH associated disease

Since CH clones can remain stable at relatively low levels for decades<sup>239</sup>, some HSC-extrinsic factors must be contributing to their clonal expansion (inflammation, infection, autoimmunity, chemo/drug/toxin exposure). These factors may provide a selective pressure that is conducive for expansion of cells harbouring ARCH mutations. Our finding of a neutrophil phenotype during differentiation due to *DNMT3A* mutation raises the question: is this intrinsic phenotype imparted by *DNMT3A R882* mutation on HSC/MPP differentiation part of a broader pro-inflammatory mechanism by which *DNMT3A* mutations lead to the clinical phenotypes observed in atherosclerosis and cardiovascular disease?

Inflammation biology in the context of cardiovascular disease has mainly focused on protein mediators (cytokines and chemokines) and small molecules (prostaglandins and reactive oxygen or nitrogen species). More recently cellular mediators of inflammation such as leukocytes have been studied<sup>106</sup>. A clinical response with anti-inflammatory treatment has been observed in patients with atherosclerosis, suggesting inflammation is an important mechanism in this disease<sup>308,309</sup>. Further, macrophages and monocytes have been shown to be key mediators in these disease processes<sup>310,311</sup>. Monocytes from patients with heart failure (HF) carrying *DNMT3A* mutations demonstrate significant upregulation of pro-inflammatory genes compared with HF patients without *DNMT3A* mutations<sup>312</sup>. Clonal expansion of *Dnmt3a*<sup>-/-</sup> HSCs is also driven by inflammation in the context of chronic mycobacterial infection and extrinsic recombinant interferon gamma<sup>313</sup>.

My data link these processes of inflammation, monocytes and CH together in the setting of a healthy individual with CH, without clinically apparent cardiovascular or haematological disease. Monocytes derived from *DNMT3A* mutant HSC/MPPs are transcriptionally associated with inflammatory cytokines and metabolic pathways, with genes including *LEP*, *PLIN5* and *TCN1* being significantly upregulated in *DNMT3A R882* mutant blood cells compared to WT.

Eradication of HSCs harbouring ARCH mutations could have much broader clinical implications than just prevention of haematological malignancy or disease relapse. The association with other inflammatory diseases could result in wider ranging public health impacts. My findings suggest that *DNMT3A* in inflammatory ARCH related conditions is more than a simple association and targeting this mutation and downstream inflammatory pathways could potentially lead to modification of disease risk and improved overall prognosis in ARCH patients. The practical challenge lies in identifying which patients are more likely to develop cardiovascular disease and hence more sophisticated prognostic strategies are required. Further, knowledge of therapeutic targets on LSCs distinct from HSCs can minimise the myelotoxicity commonly seen in clinical practice today, which is a major limiting factor in adequate treatment of leukaemia. The long-term clinical applications of the findings from this research may range from direct therapeutic strategies for haematological and non-haematological disease to supportive treatments for patients undergoing toxic therapies for these conditions.

## 5.5 Future studies in clonal haematopoiesis

To better understand the effects of ARCH mutations on haematopoiesis, future cohorts should investigate more patients who are in the transition from health to disease, as ARCH1 in our cohort. Of course, there is an obvious challenge to this approach, in that these individuals cannot be easily identified. Blood samples need to be prospectively collected and retrospectively analysed when the individual is identified as having developed an abnormal clinical phenotype e.g., leukaemia or MPN. This analysis will not only provide insights into the functional role of ARCH mutations in the absence of disease but in their likely pathogenic forms but can also provide epidemiological data on the prevalence of ARCH mutations in healthy individuals and risk factors for developing haematological or other diseases besides the presence of an ARCH mutation. Likely risk factors would include higher VAF or mutation burden, specific subtypes of mutations e.g., *R882* versus *non-R882*, timing of acquisition of mutations and the tissues affected by these somatic mutations. It is also important to identify which features of ARCH lead to non-haem disease as this carries the most significant public health risk, especially in the elderly population. It would be reasonable to hypothesise that *DNMT3A* mutation carriers who subsequently develop malignancy exhibit the differentiation phenotype described in my work, well in advance of actually developing clinically apparent disease.

Evaluation of the effect of other ARCH related mutations on HSC differentiation dynamics is warranted for genes with known association with a malignant phenotype, particularly *TET2*, *ASXL1* and splicing mutations (*SF3B1* and *SRSF2*).

Finally, this study could be extended to patients who have a known expanded driver clone in blood secondary to other toxic insults such as cytotoxic chemotherapy and smoking, which could provide insights into the mechanisms by which these clonal mutations lead to malignancy of the haematological and pulmonary systems respectively. Certain mutations and mutational signatures following cytotoxic treatment have been identified in paediatric patients, though a causal link not established to date<sup>314,315</sup>. These findings taken together with mine warrant further investigation of these mutations and signatures at a functional level to understand their role in the development of secondary haematological malignancies following cytotoxic therapy.

iPS cells (iPSC) have been derived from CD34<sup>-</sup> cells from 3 independent samples from our cohort, 1 ARCH and 2 AML. 2 of these 3 samples have a *DNMT3A* R882 mutant clone and an isogenic control, which I have expanded and frozen as working stock. Given the relative abundance of iPSC cells compared to primary human samples, they can be used to answer mechanistic questions arising from my work to date. It is anticipated that another PhD student in our group will use these iPSC cell lines to investigate the mechanisms by which *DNMT3A* R882 HSC/MPPs favour differentiation towards mature neutrophils and away from mature monocytes. Further, these lines can be used to identify any differences in the rate of differentiation towards mature cell lineages between *DNMT3A* R882 and WT iPSC cell lines from the same individual. Metabolic and functional assays examining differences in the functional phenotype of mature monocytes and neutrophils is also planned. While the functional examination of mature cells derived from iPSCs can be related back to mature cells derived from

primary HSC/MPP differentiation, the main limitations of iPSCs is the inability to directly correlate HSC function as these cells do not reliably recapitulate HSC/MPP function.

The laboratory has access to at least two mouse models of *Dnmt3a* KO available, which can be used to validate findings of altered monocyte and neutrophil differentiation patterns observed in my cohort. There are of course inherent difficulties with interpreting data from mouse models, as they do not adequately recapitulate the aging process and exposure to inflammation/infection and other stressful stimuli in humans across a life span. However, they can be used for validation and answering specific questions about altered differentiation patterns observed in human samples.

Summarising my results together with the published literature to date, we can conclude that *DNMT3A R882* mutation intrinsically alters myeloid differentiation from HSCs through to differentiated cells. I have detected a novel anomalous differentiation phenotype within the myeloid branch imparted by *DNMT3A* mutations. The main challenge now lies in extending beyond association and identifying a causal link between common CH mutations (*DNMT3A*, *TET2*, *ASXL1*) and haematological and non-haematological diseases.

## References

1. Ramalho-Santos, M. & Willenbring, H. On the Origin of the Term “Stem Cell”. *Cell Stem Cell* **1**, 35–38 (2007).
2. Maximow, A. The Lymphocyte as a stem cell common to different blood elements in embryonic development and during the post-fetal life of mammals. *Orig. Ger. Folia Haematol. Cell Ther Transpl.* **8**, 125–134 (1909).
3. Till, J. E. & McCulloch, E. A. Hemopoietic stem cell differentiation. *Biochim. Biophys. Acta BBA - Rev. Cancer* **605**, 431–459 (1980).
4. Johnson, G. R. & Metcalf, D. Pure and mixed erythroid colony formation in vitro stimulated by spleen conditioned medium with no detectable erythropoietin. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 3879–3882 (1977).
5. Till, J. E. & McCulloch, E. A. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. 1961. *Radiat. Res.* **175**, 145–149 (2011).
6. Cronkite, E. P. Hemopoietic stem cells: An analytic review of hemopoiesis. *Pathobiol. Annu.* **5**, 35–69 (1975).
7. Wang, J. C. Y., Doedens, M. & Dick, J. E. Primitive Human Hematopoietic Cells Are Enriched in Cord Blood Compared With Adult Bone Marrow or Mobilized Peripheral Blood as Measured by the Quantitative In Vivo SCID-Repopulating Cell Assay. *Blood* **89**, 3919–3924 (1997).
8. Spangrude, G. J., Heimfeld, S. & Weissman, I. L. Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62 (1988).
9. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: a human perspective. *Cell Stem Cell* **10**, 120–136 (2012).
10. Notta, F. *et al.* Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218–221 (2011).
11. Majeti, R., Park, C. Y. & Weissman, I. L. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell* **1**, 635–645 (2007).
12. Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* **11**, 585–593 (2010).
13. Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138–152 (2011).
14. Bonner, W. A., Hulett, H. R., Sweet, R. G. & Herzenberg, L. A. Fluorescence activated cell sorting. *Rev. Sci. Instrum.* **43**, 404–409 (1972).
15. Miller, R. G. & Price, G. B. Cell separation and surface markers. *Clin. Haematol.* **8**, 421–434 (1979).
16. Price, G. B., McCutcheon, M. J., Taylor, W. B. & Miller, R. G. Measurement of cytoplasmic fluorescence depolarization of single cells in a flow system. *J. Histochem. Cytochem. Off. J. Histochem. Soc.* **25**, 597–600 (1977).
17. Ye, F., Huang, W. & Guo, G. Studying hematopoiesis using single-cell technologies. *J. Hematol. Oncol. J Hematol Oncol* **10**, (2017).
18. Gundry, M. C. *et al.* Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. *Cell Rep.* **17**, 1453–1461 (2016).
19. Tothova, Z. *et al.* Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell Stem Cell* **21**, 547-555.e8 (2017).
20. Christen, F. *et al.* Modeling clonal hematopoiesis in umbilical cord blood cells by CRISPR/Cas9. *Leukemia* **1–9** (2021) doi:10.1038/s41375-021-01469-x.
21. Richardson, S. *et al.* Modelling the in-utero initiation of ETV6-RUNX1 in childhood acute lymphoblastic leukaemia using human pluripotent stem cells. *The Lancet* **387**, S86 (2016).
22. Adolfsson, J. *et al.* Upregulation of Flt3 Expression within the Bone Marrow Lin–Sca1+c-kit+ Stem Cell Compartment Is Accompanied by Loss of Self-Renewal Capacity. *Immunity* **15**, 659–669 (2001).
23. Sanjuan-Pla, A. *et al.* Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
24. Yamamoto, R. *et al.* Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* **154**, 1112–1126 (2013).
25. Laurenti, E. & Göttgens, B. From haematopoietic stem cells to complex differentiation landscapes. *Nature* **553**, 418 (2018).
26. Pellin, D. *et al.* A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nat. Commun.* **10**, 2395 (2019).

27. Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
28. Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* **167**, 1296–1309.e10 (2016).
29. Rufer, N. *et al.* Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. *J. Exp. Med.* **190**, 157–167 (1999).
30. Siminovitch, L., Mcculloch, E. A. & Till, J. E. THE DISTRIBUTION OF COLONY-FORMING CELLS AMONG SPLEEN COLONIES. *J. Cell. Comp. Physiol.* **62**, 327–336 (1963).
31. Paul, F. *et al.* Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**, 1663–1677 (2015).
32. Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20-31 (2016).
33. Chen, L. *et al.* Transcriptional diversity during lineage commitment of human blood progenitors. *Science* **345**, 1251033 (2014).
34. Haas, S., Trumpp, A. & Milsom, M. D. Causes and Consequences of Hematopoietic Stem Cell Heterogeneity. *Cell Stem Cell* **22**, 627–638 (2018).
35. Laurenti, E. *et al.* The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. *Nat. Immunol.* **14**, 756–763 (2013).
36. Laurenti, E. *et al.* CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells. *Cell Stem Cell* **16**, 302–313 (2015).
37. Bhatia, M., Wang, J. C. Y., Kapp, U., Bonnet, D. & Dick, J. E. Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice. *Proc. Natl. Acad. Sci.* **94**, 5320–5325 (1997).
38. Conneally, E., Cashman, J., Petzer, A. & Eaves, C. Expansion in vitro of transplantable human cord blood stem cells demonstrated using a quantitative assay of their lympho-myeloid repopulating activity in nonobese diabetic–scid/scid mice. *Proc. Natl. Acad. Sci.* **94**, 9836–9841 (1997).
39. Kent, D. G. *et al.* Prospective isolation and molecular characterization of hematopoietic stem cells with durable self-renewal potential. *Blood* **113**, 6342–6350 (2009).
40. 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).
41. Pietras, E. M. *et al.* Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. *Cell Stem Cell* **17**, 35–46 (2015).
42. 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**, 102–116 (2013).
43. Dykstra, B. *et al.* Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* **1**, 218–229 (2007).
44. 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).
45. Mende, N. *et al.* Unique molecular and functional features of extramedullary hematopoietic stem and progenitor cell reservoirs in humans. *Blood* [blood.2021013450](https://doi.org/10.1182/blood.2021013450) (2022) doi:10.1182/blood.2021013450.
46. Challen, G. A., Boles, N. C., Chambers, S. M. & Goodell, M. A. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. *Cell Stem Cell* **6**, 265–278 (2010).
47. Yu, V. W. C. *et al.* Epigenetic Memory Underlies Cell-Autonomous Heterogeneous Behavior of Hematopoietic Stem Cells. *Cell* **167**, 1310–1322.e17 (2016).
48. Dor, F. J. M. F. *et al.* Primitive hematopoietic cell populations reside in the spleen: Studies in the pig, baboon, and human. *Exp. Hematol.* **34**, 1573–1582 (2006).
49. Moignard, V. *et al.* Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* **15**, 363–372 (2013).
50. Belluschi, S. *et al.* Myelo-lymphoid lineage restriction occurs in the human haematopoietic stem cell compartment before lymphoid-primed multipotent progenitors. *Nat. Commun.* **9**, (2018).
51. Psaila, B. *et al.* Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways. *Genome Biol.* **17**, 83 (2016).
52. Cabezas-Wallscheid, N. *et al.* Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. *Cell* **169**, 807–823.e19 (2017).
53. Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).

54. Vedi, A., Santoro, A., Dunant, C. F., Dick, J. E. & Laurenti, E. Molecular landscapes of human hematopoietic stem cells in health and leukemia. *Ann. N. Y. Acad. Sci.* **1370**, 5–14 (2016).
55. Cheung, T. H. & Rando, T. A. Molecular regulation of stem cell quiescence. *Nat. Rev. Mol. Cell Biol.* **14**, 329–340 (2013).
56. Cheung, T. H. & Rando, T. A. Molecular regulation of stem cell quiescence. *Nat. Rev. Mol. Cell Biol.* **14**, 329 (2013).
57. Bernitz, J. M., Kim, H. S., MacArthur, B., Sieburg, H. & Moore, K. Hematopoietic Stem Cells Count and Remember Self-Renewal Divisions. *Cell* **167**, 1296–1309.e10 (2016).
58. Novershtern, N. *et al.* Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
59. Nakamura-Ishizu, A., Takizawa, H. & Suda, T. The analysis, roles and regulation of quiescence in hematopoietic stem cells. *Development* **141**, 4656–4666 (2014).
60. Xie, S. Z. *et al.* Sphingolipid Modulation Activates Proteostasis Programs to Govern Human Hematopoietic Stem Cell Self-Renewal. *Cell Stem Cell* **25**, 639–653.e7 (2019).
61. Ho, T. T. *et al.* Autophagy maintains the metabolism and function of young and old stem cells. *Nature advance online publication*, (2017).
62. Boulais, P. E. & Frenette, P. S. Making sense of hematopoietic stem cell niches. *Blood* **125**, 2621–2629 (2015).
63. Cheng, T. *et al.* Hematopoietic Stem Cell Quiescence Maintained by p21cip1/waf1. *Science* **287**, 1804–1808 (2000).
64. Arai, F. *et al.* Tie2/Angiopoietin-1 Signaling Regulates Hematopoietic Stem Cell Quiescence in the Bone Marrow Niche. *Cell* **118**, 149–161 (2004).
65. Ladha, M. H., Lee, K. Y., Upton, T. M., Reed, M. F. & Ewen, M. E. Regulation of Exit from Quiescence by p27 and Cyclin D1-CDK4. *Mol. Cell. Biol.* **18**, 6605–6615 (1998).
66. Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev.* **18**, 2747–2763 (2004).
67. Min, I. M. *et al.* The Transcription Factor EGR1 Controls Both the Proliferation and Localization of Hematopoietic Stem Cells. *Cell Stem Cell* **2**, 380–391 (2008).
68. Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
69. Viatour, P. *et al.* Hematopoietic Stem Cell Quiescence Is Maintained by Compound Contributions of the Retinoblastoma Gene Family. *Cell Stem Cell* **3**, 416–428 (2008).
70. Zhang, Y. W. *et al.* Hyaluronic acid-GPRC5C signalling promotes dormancy in haematopoietic stem cells. *Nat. Cell Biol.* (2022) doi:10.1038/s41556-022-00931-x.
71. Yoshihara, H. *et al.* Thrombopoietin/MPL Signaling Regulates Hematopoietic Stem Cell Quiescence and Interaction with the Osteoblastic Niche. *Cell Stem Cell* **1**, 685–697 (2007).
72. Grenier, J. M. P., Testut, C., Fauriat, C., Mancini, S. J. C. & Aurrand-Lions, M. Adhesion Molecules Involved in Stem Cell Niche Retention During Normal Haematopoiesis and in Acute Myeloid Leukaemia. *Front. Immunol.* **12**, (2021).
73. Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834 (2010).
74. Wagner, W. *et al.* Adhesion of hematopoietic progenitor cells to human mesenchymal stem cells as a model for cell-cell interaction. *Exp. Hematol.* **35**, 314–325 (2007).
75. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
76. Morales-Mantilla, D. E. & King, K. Y. The Role of Interferon-Gamma in Hematopoietic Stem Cell Development, Homeostasis, and Disease. *Curr. Stem Cell Rep.* **4**, 264–271 (2018).
77. Essers, M. A. G. *et al.* IFN $\alpha$  activates dormant haematopoietic stem cells in vivo. *Nature* **458**, 904–908 (2009).
78. Mack, R., Zhang, L., Breslin S, P. & Zhang, J. The Fetal-to-Adult Hematopoietic Stem Cell Transition and its Role in Childhood Hematopoietic Malignancies. *Stem Cell Rev. Rep.* **17**, 2059–2080 (2021).
79. Lee, Y., Decker, M., Lee, H. & Ding, L. Extrinsic regulation of hematopoietic stem cells in development, homeostasis and diseases. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, 10.1002/wdev.279 (2017).
80. Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* **44**, 1179–1181 (2012).
81. Busque, L. *et al.* Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood* **88**, 59–65 (1996).

82. Mitchell, E. *et al.* Clonal dynamics of haematopoiesis across the human lifespan. *Nature* **606**, 343–350 (2022).
83. Morita, K. *et al.* Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat. Commun.* **11**, 5327 (2020).
84. Williams, N. *et al.* Phylogenetic reconstruction of myeloproliferative neoplasm reveals very early origins and lifelong evolution. *bioRxiv* 2020.11.09.374710 (2020) doi:10.1101/2020.11.09.374710.
85. Coombs, C. C. *et al.* Therapy-Related Clonal Hematopoiesis in Patients with Non-hematologic Cancers Is Common and Associated with Adverse Clinical Outcomes. *Cell Stem Cell* **21**, 374-382.e4 (2017).
86. Gibson, C. J. *et al.* Clonal Hematopoiesis Associated With Adverse Outcomes After Autologous Stem-Cell Transplantation for Lymphoma. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **35**, 1598–1605 (2017).
87. Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
88. Jaiswal, S. *et al.* Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014).
89. Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**, 1472–1478 (2014).
90. Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9–16 (2015).
91. Fabre, M. A. *et al.* *The longitudinal dynamics and natural history of clonal haematopoiesis.* 2021.08.12.455048 <https://www.biorxiv.org/content/10.1101/2021.08.12.455048v1> (2021) doi:10.1101/2021.08.12.455048.
92. Yang, L., Rau, R. & Goodell, M. A. DNMT3A in haematological malignancies. *Nat. Rev. Cancer* **15**, 152–165 (2015).
93. Acuna-Hidalgo, R. *et al.* Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *Am. J. Hum. Genet.* **101**, 50–64 (2017).
94. Waterhouse, M. *et al.* Genome-wide profiling in AML patients relapsing after allogeneic hematopoietic cell transplantation. *Biol. Blood Marrow Transplant. J. Am. Soc. Blood Marrow Transplant.* **17**, 1450-1459.e1 (2011).
95. Jaiswal, S. *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. <https://doi.org/10.1056/NEJMoa1701719> <https://www.nejm.org/doi/10.1056/NEJMoa1701719> (2017) doi:10.1056/NEJMoa1701719.
96. Jaiswal, S. Clonal hematopoiesis and non-hematologic disorders. *Blood* (2020) doi:10.1182/blood.2019000989.
97. Meisel, M. *et al.* Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. *Nature* **557**, 580–584 (2018).
98. Miller, P. G. *et al.* Association of clonal hematopoiesis with chronic obstructive pulmonary disease. *Blood* **139**, 357–368 (2022).
99. Kim, P. G. *et al.* Dnmt3a-mutated clonal hematopoiesis promotes osteoporosis. *J. Exp. Med.* **218**, e20211872 (2021).
100. Steensma, D. P. Clinical consequences of clonal hematopoiesis of indeterminate potential. *Blood Adv.* **2**, 3404–3410 (2018).
101. Robertson, K. D. *et al.* The human DNA methyltransferases (DNMTs) 1, 3a and 3b: coordinate mRNA expression in normal tissues and overexpression in tumors. *Nucleic Acids Res.* **27**, 2291–2298 (1999).
102. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
103. Ramsahoye, B. H. *et al.* Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5237–5242 (2000).
104. He, Y. & Ecker, J. R. Non-CG Methylation in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **16**, 55–77 (2015).
105. Zhang, Z.-M. *et al.* Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature* **554**, 387–391 (2018).
106. Libby, P., Nahrendorf, M. & Swirski, F. K. Leukocytes Link Local and Systemic Inflammation in Ischemic Cardiovascular Disease: An Expanded ‘Cardiovascular Continuum’. *J. Am. Coll. Cardiol.* **67**, 1091–1103 (2016).
107. Gerhardt, T. & Ley, K. Monocyte trafficking across the vessel wall. *Cardiovasc. Res.* **107**, 321–330 (2015).
108. Yang, J., Zhang, L., Yu, C., Yang, X.-F. & Wang, H. Monocyte and macrophage differentiation: circulation inflammatory monocyte as biomarker for inflammatory diseases. *Biomark. Res.* **2**, 1 (2014).

109. Fuster, J. J. *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**, 842–847 (2017).
110. Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 868–874 (2002).
111. Olefsky, J. M. & Glass, C. K. Macrophages, inflammation, and insulin resistance. *Annu. Rev. Physiol.* **72**, 219–246 (2010).
112. Heuser, M., Thol, F. & Ganser, A. Clonal Hematopoiesis of Indeterminate Potential. *Dtsch. Arzteblatt Int.* **113**, 317–322 (2016).
113. Arends, C. M. *et al.* Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. *Leukemia* **32**, 1908–1919 (2018).
114. Buscarlet, M. *et al.* Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. *Blood* **132**, 277–280 (2018).
115. McCulloch, E. A., Howatson, A. F., Buick, R. N., Minden, M. D. & Izaguirre, C. A. Acute myeloblastic leukemia considered as a clonal hemopathy. *Blood Cells* **5**, 261–282 (1979).
116. Fialkow, P. J. *et al.* Clonal Development, Stem-Cell Differentiation, and Clinical Remissions in Acute Nonlymphocytic Leukemia. *N. Engl. J. Med.* **317**, 468–473 (1987).
117. Vogelstein, B., Fearon, E. R., Hamilton, S. R. & Feinberg, A. P. Use of restriction fragment length polymorphisms to determine the clonal origin of human tumors. *Science* **227**, 642–645 (1985).
118. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645–648 (1994).
119. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
120. Kreso, A. & Dick, J. E. Evolution of the cancer stem cell model. *Cell Stem Cell* **14**, 275–291 (2014).
121. Bartels, S. *et al.* SRSF2 and U2AF1 mutations in primary myelofibrosis are associated with JAK2 and MPL but not calreticulin mutation and may independently reoccur after allogeneic stem cell transplantation. *Leukemia* **29**, 253–255 (2015).
122. Larsson, C. A., Cote, G. & Quintás-Cardama, A. The Changing Mutational Landscape of Acute Myeloid Leukemia and Myelodysplastic Syndrome. *Mol. Cancer Res. MCR* **11**, 815–827 (2013).
123. Bejar, R. Splicing Factor Mutations in Cancer. *Adv. Exp. Med. Biol.* **907**, 215–228 (2016).
124. Graubert, T. A. *et al.* Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat. Genet.* **44**, 53–57 (2012).
125. Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat. Med.* **3**, 730–737 (1997).
126. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).
127. Dick, J. E. Stem cell concepts renew cancer research. *Blood* **112**, 4793–4807 (2008).
128. Jordan, C. T., Guzman, M. L. & Noble, M. Cancer Stem Cells. *N. Engl. J. Med.* **355**, 1253–1261 (2006).
129. Corces-Zimmerman, M. R., Hong, W.-J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 2548–2553 (2014).
130. Martelli, M. P. *et al.* CD34+ cells from AML with mutated NPM1 harbor cytoplasmic mutated nucleophosmin and generate leukemia in immunocompromised mice. *Blood* **116**, 3907–3922 (2010).
131. Taussig, D. C. *et al.* Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. *Blood* **112**, 568–575 (2008).
132. Sarry, J.-E. *et al.* Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2R $\gamma$ c-deficient mice. *J. Clin. Invest.* **121**, 384–395 (2011).
133. Eppert, K. *et al.* Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat. Med.* **17**, 1086–1093 (2011).
134. Gentles, A. J., Plevritis, S. K., Majeti, R. & Alizadeh, A. A. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA* **304**, 2706–2715 (2010).
135. Ng, S. W. K. *et al.* A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature* **540**, 433–437 (2016).
136. Rongvaux, A. *et al.* Human hemato-lymphoid system mice: current use and future potential for medicine. *Annu. Rev. Immunol.* **31**, 635–674 (2013).
137. Goyama, S., Wunderlich, M. & Mulloy, J. C. Xenograft models for normal and malignant stem cells. *Blood* **125**, 2630–2640 (2015).
138. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23 (2012).

139. Neumann, M. *et al.* Whole-exome sequencing in adult ETP-ALL reveals a high rate of DNMT3A mutations. *Blood* **121**, 4749–4752 (2013).
140. 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).
141. Odejide, O. *et al.* A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood* **123**, 1293–1296 (2014).
142. Asmar, F. *et al.* Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. *Haematologica* **98**, 1912–1920 (2013).
143. Couronné, L., Bastard, C. & Bernard, O. A. TET2 and DNMT3A mutations in human T-cell lymphoma. *N. Engl. J. Med.* **366**, 95–96 (2012).
144. Shlush, L. I. *et al.* Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature advance online publication*, (2017).
145. O'Brien, S. G. *et al.* Imatinib Compared with Interferon and Low-Dose Cytarabine for Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia. *N. Engl. J. Med.* **348**, 994–1004 (2003).
146. Longo, D. L. Imatinib Changed Everything. *N. Engl. J. Med.* **376**, 982–983 (2017).
147. Residual blood products from NHSBT platelet donation for use in assays. *Health Research Authority* <https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/residual-blood-products-from-nhsbt-platelet-donation-for-use-in-assays/>.
148. Sarma, N. J., Takeda, A. & Yaseen, N. R. Colony Forming Cell (CFC) Assay for Human Hematopoietic Cells. *J. Vis. Exp. JoVE* (2010) doi:10.3791/2195.
149. Hotchin, J. E. Use of Methyl Cellulose Gel as a Substitute for Agar in Tissue-Culture Overlays. *Nature* **175**, 352 (1955).
150. Warner, J. K. *et al.* Direct evidence for cooperating genetic events in the leukemic transformation of normal human hematopoietic cells. *Leukemia* **19**, 1794–1805 (2005).
151. Tuval, A. *et al.* Pseudo-mutant p53 as a targetable phenotype of DNMT3A-mutated pre-leukemia. 2021.05.30.446347 <https://www.biorxiv.org/content/10.1101/2021.05.30.446347v1> (2021) doi:10.1101/2021.05.30.446347.
152. Acuna-Hidalgo, R. *et al.* Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *Am. J. Hum. Genet.* **101**, 50–64 (2017).
153. Hiatt, J. B., Pritchard, C. C., Salipante, S. J., O'Roak, B. J. & Shendure, J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res.* **23**, 843–854 (2013).
154. Boyle, E. A., O'Roak, B. J., Martin, B. K., Kumar, A. & Shendure, J. MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinforma. Oxf. Engl.* **30**, 2670–2672 (2014).
155. Bushnell, B., Rood, J. & Singer, E. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* **12**, e0185056 (2017).
156. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
157. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* (2013).
158. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
159. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
160. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
161. Rimmer, A. *et al.* Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).
162. Biezuner, T. *et al.* An improved molecular inversion probe based targeted sequencing approach for low variant allele frequency. *NAR Genomics Bioinforma.* **4**, lqab125 (2022).
163. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
164. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* **29**, 15–21 (2013).
165. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31**, 166–169 (2015).
166. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).
167. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44 (2008).

168. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
169. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
170. Hänzelmann, S., Castelo, R. & Guinney, J. GSEA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* **14**, 1–15 (2013).
171. Ogino, S., Gulley, M. L., den Dunnen, J. T. & Wilson, R. B. Standard Mutation Nomenclature in Molecular Diagnostics. *J. Mol. Diagn. JMD* **9**, 1–6 (2007).
172. Boriack-Sjodin, P. A. & Swinger, K. K. Protein Methyltransferases: A Distinct, Diverse, and Dynamic Family of Enzymes. *Biochemistry* **55**, 1557–1569 (2016).
173. Yang, Y. & Bedford, M. T. Protein arginine methyltransferases and cancer. *Nat. Rev. Cancer* **13**, 37–50 (2013).
174. Jaenisch, R. DNA methylation and imprinting: why bother? *Trends Genet. TIG* **13**, 323–329 (1997).
175. Lei, H. *et al.* De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195–3205 (1996).
176. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
177. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247–257 (1999).
178. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells Devoted Mol. Cell. Mech.* **11**, 805–814 (2006).
179. Peters, S. L. *et al.* Essential role for Dnmt1 in the prevention and maintenance of MYC-induced T-cell lymphomas. *Mol. Cell. Biol.* **33**, 4321–4333 (2013).
180. Bröske, A.-M. *et al.* DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. *Nat. Genet.* **41**, 1207–1215 (2009).
181. Trowbridge, J. J., Snow, J. W., Kim, J. & Orkin, S. H. DNA Methyltransferase 1 Is Essential for and Uniquely Regulates Hematopoietic Stem and Progenitor Cells. *Cell Stem Cell* **5**, 442–449 (2009).
182. Deng, X. *et al.* Arginine methylation mediated by the Arabidopsis homolog of PRMT5 is essential for proper pre-mRNA splicing. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19114–19119 (2010).
183. Sanchez, S. E. *et al.* A methyl transferase links the circadian clock to the regulation of alternative splicing. *Nature* **468**, 112–116 (2010).
184. Gonsalvez, G. B., Rajendra, T. K., Tian, L. & Matera, A. G. The Sm-protein methyltransferase, dart5, is essential for germ-cell specification and maintenance. *Curr. Biol. CB* **16**, 1077–1089 (2006).
185. Bezzi, M. *et al.* Regulation of constitutive and alternative splicing by PRMT5 reveals a role for Mdm4 pre-mRNA in sensing defects in the spliceosomal machinery. *Genes Dev.* **27**, 1903–1916 (2013).
186. Liu, F. *et al.* Arginine methyltransferase PRMT5 is essential for sustaining normal adult hematopoiesis. *J. Clin. Invest.* **125**, 3532–3544 (2015).
187. Vougiouklakis, T., Nakamura, Y. & Saloura, V. Critical roles of protein methyltransferases and demethylases in the regulation of embryonic stem cell fate. *Epigenetics* **0**, 1–13 (2017).
188. Wu, Q. *et al.* CARM1 Is Required in Embryonic Stem Cells to Maintain Pluripotency and Resist Differentiation. *STEM CELLS* **27**, 2637–2645 (2009).
189. Tee, W.-W. *et al.* Prmt5 is essential for early mouse development and acts in the cytoplasm to maintain ES cell pluripotency. *Genes Dev.* **24**, 2772–2777 (2010).
190. Lee, S.-H. *et al.* A feedback loop comprising PRMT7 and miR-24-2 interplays with Oct4, Nanog, Klf4 and c-Myc to regulate stemness. *Nucleic Acids Res.* **44**, 10603–10618 (2016).
191. Yu, Z., Chen, T., Hébert, J., Li, E. & Richard, S. A Mouse PRMT1 Null Allele Defines an Essential Role for Arginine Methylation in Genome Maintenance and Cell Proliferation. *Mol. Cell. Biol.* **29**, 2982–2996 (2009).
192. Gkoutela, S., Li, Z., Chin, C. J., Lee, S. A. & Clark, A. T. PRMT5 is Required for Human Embryonic Stem Cell Proliferation But Not Pluripotency. *Stem Cell Rev. Rep.* **10**, 230–239 (2014).
193. Tan, D. Q. *et al.* PRMT5 Modulates Splicing for Genome Integrity and Preserves Proteostasis of Hematopoietic Stem Cells. *Cell Rep.* **26**, 2316–2328.e6 (2019).
194. Tan, D. Q. *et al.* PRMT5 Modulates Splicing for Genome Integrity and Preserves Proteostasis of Hematopoietic Stem Cells. *Cell Rep.* **26**, 2316–2328.e6 (2019).
195. Liu, F. *et al.* JAK2V617F-mediated phosphorylation of PRMT5 downregulates its methyltransferase activity and promotes myeloproliferation. *Cancer Cell* **19**, 283–294 (2011).

196. Fabbrizio, E. *et al.* Negative regulation of transcription by the type II arginine methyltransferase PRMT5. *EMBO Rep.* **3**, 641–645 (2002).
197. Durant, S. T., Cho, E. C. & Thangue, N. B. L. p53 methylation—the Arg-ument is clear. *Cell Cycle* **8**, 801–802 (2009).
198. Wang, L., Pal, S. & Sif, S. Protein Arginine Methyltransferase 5 Suppresses the Transcription of the RB Family of Tumor Suppressors in Leukemia and Lymphoma Cells. *Mol. Cell. Biol.* **28**, 6262–6277 (2008).
199. Jin, Y. *et al.* Targeting methyltransferase PRMT5 eliminates leukemia stem cells in chronic myelogenous leukemia. *J. Clin. Invest.* **126**, 3961–3980 (2016).
200. Shia, W.-J. *et al.* PRMT1 interacts with AML1-ETO to promote its transcriptional activation and progenitor cell proliferative potential. *Blood* **119**, 4953–4962 (2012).
201. Kaushik, S. *et al.* Genetic deletion or small-molecule inhibition of the arginine methyltransferase PRMT5 exhibit anti-tumoral activity in mouse models of *MLL*-rearranged AML. *Leukemia* (2017) doi:10.1038/leu.2017.206.
202. Braun, C. J. *et al.* Coordinated Splicing of Regulatory Detained Introns within Oncogenic Transcripts Creates an Exploitable Vulnerability in Malignant Glioma. *Cancer Cell* **32**, 411-426.e11 (2017).
203. Gerhart, S. V. *et al.* Activation of the p53-MDM4 regulatory axis defines the anti-tumour response to PRMT5 inhibition through its role in regulating cellular splicing. *Sci. Rep.* **8**, 9711 (2018).
204. Chan-Penebre, E. *et al.* A selective inhibitor of PRMT5 with *in vivo* and *in vitro* potency in MCL models. *Nat. Chem. Biol.* **11**, 432 (2015).
205. Chen, K., Dai, X. & Wu, J. Alternative splicing: An important mechanism in stem cell biology. *World J. Stem Cells* **7**, 1–10 (2015).
206. Boyer, L. A. *et al.* Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell* **122**, 947–956 (2005).
207. Mayshar, Y. *et al.* Fibroblast Growth Factor 4 and Its Novel Splice Isoform Have Opposing Effects on the Maintenance of Human Embryonic Stem Cell Self-Renewal. *STEM CELLS* **26**, 767–774 (2008).
208. Cesana, M. *et al.* A CLK3-HMGA2 alternative splicing axis impacts human hematopoietic stem cell molecular identity throughout development. *Cell Stem Cell* **22**, 575-588.e7 (2018).
209. Gao, Y., Vasic, R. & Halene, S. Role of alternative splicing in hematopoietic stem cells during development. *Stem Cell Investig.* **5**, 26 (2018).
210. Zahid, M. F., Patnaik, M. M., Gangat, N., Hashmi, S. K. & Rizzieri, D. A. Insight into the molecular pathophysiology of myelodysplastic syndromes: targets for novel therapy. *Eur. J. Haematol.* **97**, 313–320 (2016).
211. Saez, B., Walter, M. J. & Graubert, T. A. Splicing factor gene mutations in hematologic malignancies. *Blood* blood-2016-10-692400 (2016) doi:10.1182/blood-2016-10-692400.
212. Fabre, M. A. *et al.* Concordance for clonal hematopoiesis is limited in elderly twins. *Blood* **135**, 269–273 (2020).
213. Mitchell, E. *et al.* *Clonal dynamics of haematopoiesis across the human lifespan*. 2021.08.16.456475 <https://www.biorxiv.org/content/10.1101/2021.08.16.456475v1> (2021) doi:10.1101/2021.08.16.456475.
214. Walter, M. J. Abstract IA11: Spliceosome gene mutations in MDS: Biology and potential therapeutic strategies. *Clin. Cancer Res.* **23**, IA11–IA11 (2017).
215. ElHady, A. K., Abdel-Halim, M., Abadi, A. H. & Engel, M. Development of Selective Clk1 and -4 Inhibitors for Cellular Depletion of Cancer-Relevant Proteins. *J. Med. Chem.* **60**, 5377–5391 (2017).
216. Muraki, M. *et al.* Manipulation of Alternative Splicing by a Newly Developed Inhibitor of Clks. *J. Biol. Chem.* **279**, 24246–24254 (2004).
217. Sun, Q.-Z. *et al.* Discovery of Potent and Selective Inhibitors of Cdc2-Like Kinase 1 (CLK1) as a New Class of Autophagy Inducers. *J. Med. Chem.* **60**, 6337–6352 (2017).
218. Liu, F. *et al.* Arginine methyltransferase PRMT5 is essential for sustaining normal adult hematopoiesis. <https://www.jci.org/articles/view/81749/pdf> (2015) doi:10.1172/JCI81749.
219. Ma, Y., Dostie, J., Dreyfuss, G. & Van Duynes, G. D. The Gemin6-Gemin7 Heterodimer from the Survival of Motor Neurons Complex Has an Sm Protein-like Structure. *Structure* **13**, 883–892 (2005).
220. Atila-Gokcumen, G. E. *et al.* Dividing cells regulate their lipid composition and localization. *Cell* **156**, 428–439 (2014).
221. van Galen, P. *et al.* The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. *Nature* **510**, 268–272 (2014).
222. Araki, S. *et al.* Inhibitors of CLK Protein Kinases Suppress Cell Growth and Induce Apoptosis by Modulating Pre-mRNA Splicing. *PLoS ONE* **10**, (2015).

223. Kannan, M. B., Solovieva, V. & Blank, V. The small MAF transcription factors MAFF, MAFG and MAFK: Current knowledge and perspectives. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1823**, 1841–1846 (2012).
224. Hoppstädter, J. & Ammit, A. J. Role of Dual-Specificity Phosphatase 1 in Glucocorticoid-Driven Anti-inflammatory Responses. *Front. Immunol.* **10**, (2019).
225. Alinari, L. *et al.* Selective inhibition of protein arginine methyltransferase 5 blocks initiation and maintenance of B-cell transformation. *Blood* **125**, 2530–2543 (2015).
226. Muraki, M. *et al.* Manipulation of Alternative Splicing by a Newly Developed Inhibitor of Clks. *J. Biol. Chem.* **279**, 24246–24254 (2004).
227. Dvinge, H. & Bradley, R. K. Widespread intron retention diversifies most cancer transcriptomes. *Genome Med.* **7**, 45 (2015).
228. Miller, P. G. *et al.* Association of Clonal Hematopoiesis with Chronic Obstructive Pulmonary Disease. *Blood* (2021) doi:10.1182/blood.2021013531.
229. Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).
230. Ley, T. J. *et al.* DNMT3A Mutations in Acute Myeloid Leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
231. Delhommeau, F. *et al.* Mutation in TET2 in Myeloid Cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
232. Gelsi-Boyer, V. *et al.* Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. *Br. J. Haematol.* **145**, 788–800 (2009).
233. Cerchione, C. *et al.* IDH1/IDH2 Inhibition in Acute Myeloid Leukemia. *Front. Oncol.* **11**, (2021).
234. Abelson, S. *et al.* Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* **559**, 400–404 (2018).
235. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
236. Ho, P. A. *et al.* Leukemic mutations in the methylation-associated genes DNMT3A and IDH2 are rare events in pediatric AML: a report from the Children’s Oncology Group. *Pediatr. Blood Cancer* **57**, 204–209 (2011).
237. Hollink, I. H. I. M. *et al.* Low frequency of DNMT3A mutations in pediatric AML, and the identification of the OCI-AML3 cell line as an *in vitro* model. *Leukemia* **26**, 371–373 (2012).
238. Shiba, N. *et al.* DNMT3A mutations are rare in childhood acute myeloid leukaemia, myelodysplastic syndromes and juvenile myelomonocytic leukaemia. *Br. J. Haematol.* **156**, 413–414 (2012).
239. Williams, N. *et al.* Life histories of myeloproliferative neoplasms inferred from phylogenies. *Nature* 1–7 (2022) doi:10.1038/s41586-021-04312-6.
240. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
241. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
242. Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219–220 (1998).
243. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
244. Jeong, M. *et al.* Loss of Dnmt3a Immortalizes Hematopoietic Stem Cells In Vivo. *Cell Rep.* **23**, 1–10 (2018).
245. Thol, F. *et al.* Incidence and Prognostic Influence of DNMT3A Mutations in Acute Myeloid Leukemia. *J. Clin. Oncol.* **29**, 2889–2896 (2011).
246. Yan, X. J. *et al.* Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* **43**, 309–15 (2011).
247. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
248. Guryanova, O. A. *et al.* DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nat. Med.* **22**, 1488–1495 (2016).
249. Cancer Genome Atlas Research Network *et al.* Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
250. Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424–33 (2010).
251. Marcucci, G. *et al.* Age-Related Prognostic Impact of Different Types of DNMT3A Mutations in Adults With Primary Cytogenetically Normal Acute Myeloid Leukemia. *J. Clin. Oncol.* **30**, 742–750 (2012).
252. Renneville, A. *et al.* Prognostic significance of DNA methyltransferase 3A mutations in cytogenetically normal acute myeloid leukemia: a study by the Acute Leukemia French Association. *Leukemia* **26**, 1247 (2012).
253. Roller, A. *et al.* Landmark analysis of DNMT3A mutations in hematological malignancies. *Leukemia* **27**, 1573 (2013).

254. Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328–333 (2014).
255. Haferlach, T. *et al.* Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* **28**, 241–247 (2014).
256. Stegelmann, F. *et al.* DNMT3A mutations in myeloproliferative neoplasms. *Leukemia* **25**, 1217–1219 (2011).
257. Sakata-Yanagimoto, M. *et al.* Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat. Genet.* **46**, 171–175 (2014).
258. Grossmann, V. *et al.* The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. *Genes. Chromosomes Cancer* **52**, 410–422 (2013).
259. Kar, S. A. *et al.* Spliceosomal gene mutations are frequent events in the diverse mutational spectrum of chronic myelomonocytic leukemia but largely absent in juvenile myelomonocytic leukemia. *Haematologica* **98**, 107–113 (2013).
260. Walter, M. J. *et al.* Recurrent DNMT3A Mutations in Patients with Myelodysplastic Syndromes. *Leukemia* **25**, 1153–1158 (2011).
261. Palomero, T. *et al.* Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat. Genet.* **46**, 166–170 (2014).
262. Young, A. L., Tong, R. S., Birmann, B. M. & Druley, T. E. Clonal hematopoiesis and risk of acute myeloid leukemia. *Haematologica* **104**, 2410–2417 (2019).
263. Ley, T. J. *et al.* DNMT3A Mutations in Acute Myeloid Leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).
264. Gaidzik, V. I. *et al.* Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSG). *Blood* **121**, 4769–4777 (2013).
265. Ribeiro, A. F. T. *et al.* Mutant DNMT3A: a marker of poor prognosis in acute myeloid leukemia. *Blood* **119**, 5824–5831 (2012).
266. Patel, J. P. *et al.* Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *N. Engl. J. Med.* **366**, 1079–1089 (2012).
267. Patel, J. P. *et al.* Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. *N. Engl. J. Med.* **366**, 1079–1089 (2012).
268. Sehgal, A. R. *et al.* DNMT3A Mutational Status Affects the Results of Dose-Escalated Induction Therapy in Acute Myelogenous Leukemia. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **21**, 1614–1620 (2015).
269. Traina, F. *et al.* Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. *Leukemia* **28**, 78–87 (2014).
270. Ding, L. *et al.* Clonal evolution in relapsed acute myeloid leukemia revealed by whole genome sequencing. *Nature* **481**, 506–510 (2012).
271. Welch, J. S. *et al.* The Origin and Evolution of Mutations in Acute Myeloid Leukemia. *Cell* **150**, 264–278 (2012).
272. Rodriguez-Meira, A., O’Sullivan, J., Rahman, H. & Mead, A. J. TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *STAR Protoc.* **1**, 100125 (2020).
273. Nam, A. S. *et al.* Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* **571**, 355–360 (2019).
274. Nam, A. S. *et al.* Single-cell multi-omics of human clonal hematopoiesis reveals that DNMT3A R882 mutations perturb early progenitor states through selective hypomethylation. 2022.01.14.476225 Preprint at <https://doi.org/10.1101/2022.01.14.476225> (2022).
275. Izzo, F. *et al.* DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.* **52**, 378–387 (2020).
276. Lee-Six, H. *et al.* Population dynamics of normal human blood inferred from somatic mutations. *Nature* **561**, 473–478 (2018).
277. Grinfeld, J. *et al.* Classification and Personalized Prognosis in Myeloproliferative Neoplasms. *N. Engl. J. Med.* **379**, 1416–1430 (2018).
278. Santoro, A. *et al.* Chronic lymphocytic leukemia increases the pool of peripheral blood hematopoietic stem cells and skews differentiation. *Blood Adv.* **4**, 6310–6314 (2020).
279. Naeim, F., Nagesh Rao, P., Song, S. X. & Phan, R. T. Chapter 2 - Principles of Immunophenotyping. in *Atlas of Hematopathology (Second Edition)* (eds. Naeim, F., Nagesh Rao, P., Song, S. X. & Phan, R. T.) 29–56 (Academic Press, 2018). doi:10.1016/B978-0-12-809843-1.00002-4.
280. Triana, S. *et al.* Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nat. Immunol.* **22**, 1577–1589 (2021).

281. Dai, Y. CytoTree: A Toolkit for Flow And Mass Cytometry Data. (2022) doi:10.18129/B9.bioc.CytoTree.
282. Diakiw, S. M. *et al.* The granulocyte-associated transcription factor Krüppel-like factor 5 is silenced by hypermethylation in acute myeloid leukemia. *Leuk. Res.* **36**, 110–116 (2012).
283. Gallardo-Montejano, V. I. *et al.* Perilipin 5 links mitochondrial uncoupled respiration in brown fat to healthy white fat remodeling and systemic glucose tolerance. *Nat. Commun.* **12**, 3320 (2021).
284. Rodrigues, C. P., Shvedunova, M. & Akhtar, A. Epigenetic Regulators as the Gatekeepers of Hematopoiesis. *Trends Genet.* **37**, 125–142 (2021).
285. Lessard, J., Faubert, A. & Sauvageau, G. Genetic programs regulating HSC specification, maintenance and expansion. *Oncogene* **23**, 7199–7209 (2004).
286. Koh, C. M. *et al.* MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. *Nature* **523**, 96–100 (2015).
287. Kryukov, G. V. *et al.* MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. *Science* **351**, 1214–1218 (2016).
288. Slayton, W. B. *et al.* Dasatinib Plus Intensive Chemotherapy in Children, Adolescents, and Young Adults With Philadelphia Chromosome-Positive Acute Lymphoblastic Leukemia: Results of Children’s Oncology Group Trial AALL0622. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **36**, 2306–2314 (2018).
289. Pollack, B. P. *et al.* The Human Homologue of the Yeast Proteins Skb1 and Hsl7p Interacts with Jak Kinases and Contains Protein Methyltransferase Activity. *J. Biol. Chem.* **274**, 31531–31542 (1999).
290. Wilson, A. *et al.* Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. *Cell* **135**, 1118–1129 (2008).
291. Florian, M. C. *et al.* Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation. *Cell Stem Cell* **10**, 520–530 (2012).
292. Guralnik, J. M., Eisenstaedt, R. S., Ferrucci, L., Klein, H. G. & Woodman, R. C. Prevalence of anemia in persons 65 years and older in the United States: evidence for a high rate of unexplained anemia. *Blood* **104**, 2263–2268 (2004).
293. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. Engl. J. Med.* **371**, 2477–2487 (2014).
294. Juliusson, G. *et al.* Age and acute myeloid leukemia: real world data on decision to treat and outcomes from the Swedish Acute Leukemia Registry. *Blood* **113**, 4179–4187 (2009).
295. Wu, Y. *et al.* Global, regional, and national childhood cancer burden, 1990–2019: An analysis based on the Global Burden of Disease Study 2019. *J. Adv. Res.* (2022) doi:10.1016/j.jare.2022.06.001.
296. Lee, J., Yoon, S. R., Choi, I. & Jung, H. Causes and Mechanisms of Hematopoietic Stem Cell Aging. *Int. J. Mol. Sci.* **20**, 1272 (2019).
297. Liang, Y., Van Zant, G. & Szilvassy, S. J. Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells. *Blood* **106**, 1479–1487 (2005).
298. Lansdorp, P. M. Maintenance of telomere length in AML. *Blood Adv.* **1**, 2467–2472 (2017).
299. Jebaraj, B. M. C. & Stilgenbauer, S. Telomere Dysfunction in Chronic Lymphocytic Leukemia. *Front. Oncol.* **10**, 612665 (2021).
300. Nogueira, B. M. D., Machado, C. B., Montenegro, R. C., Moraes, M. E. A. D. & Moreira-Nunes, C. A. Telomere Length and Hematological Disorders: A Review. *In Vivo* **34**, 3093–3101 (2020).
301. Leoni, C. *et al.* Dnmt3a restrains mast cell inflammatory responses. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1490–E1499 (2017).
302. Dai, Y.-J. *et al.* Conditional knockin of Dnmt3a R878H initiates acute myeloid leukemia with mTOR pathway involvement. *Proc. Natl. Acad. Sci.* **114**, 5237–5242 (2017).
303. Anteneh, H., Fang, J. & Song, J. Structural basis for impairment of DNA methylation by the DNMT3A R882H mutation. *Nat. Commun.* **11**, 2294 (2020).
304. Brunetti, L., Gundry, M. C. & Goodell, M. A. DNMT3A in Leukemia. *Cold Spring Harb. Perspect. Med.* **7**, a030320 (2017).
305. Van Egeren, D. *et al.* Reconstructing the Lineage Histories and Differentiation Trajectories of Individual Cancer Cells in Myeloproliferative Neoplasms. *Cell Stem Cell* **28**, 514–523.e9 (2021).
306. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat. Genet.* **44**, 23–31 (2011).
307. Loberg, M. *et al.* Mutation in DNA Methyltransferase DNMT3A Confers Enhanced Self-Renewal Capacity Onto Multipotent Progenitor Cells and Predisposes to Acute Myeloid Leukemia (AML). *Blood* **132**, 2569 (2018).
308. Ridker, P. M. *et al.* Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. *N. Engl. J. Med.* **377**, 1119–1131 (2017).

309. Libby, P. *et al.* Atherosclerosis. *Nat. Rev. Dis. Primer* **5**, 56 (2019).
310. Ghattas, A., Griffiths, H. R., Devitt, A., Lip, G. Y. H. & Shantsila, E. Monocytes in coronary artery disease and atherosclerosis: where are we now? *J. Am. Coll. Cardiol.* **62**, 1541–1551 (2013).
311. Tabas, I. & Lichtman, A. H. Monocyte-Macrophages and T Cells in Atherosclerosis. *Immunity* **47**, 621–634 (2017).
312. Abplanalp, W. T. *et al.* Clonal Hematopoiesis-Driver DNMT3A Mutations Alter Immune Cells in Heart Failure. *Circ. Res.* **128**, 216–228 (2021).
313. Hormaechea-Agulla, D. *et al.* Chronic infection drives Dnmt3a-loss-of-function clonal hematopoiesis via IFN $\gamma$  signaling. *Cell Stem Cell* **28**, 1428-1442.e6 (2021).
314. Spitzer, B. *et al.* Bone Marrow Surveillance of Pediatric Cancer Survivors Identifies Clones that Predict Therapy-Related Leukemia. *Clin. Cancer Res.* **28**, 1614–1627 (2022).
315. Coorens, T. H. *et al.* Clonal hematopoiesis and therapy-related myeloid neoplasms following neuroblastoma treatment. *Blood* (2021) doi:10.1182/blood.2020010150.