The Role of DNA Methylation in Stem Cell Ageing

Oluwatosin O. Taiwo

Institute of Healthy Ageing

and

UCL Cancer Institute

University College London

This thesis is submitted for the degree of Doctor of Philosophy

Declaration

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

Work Carried Out In Collaboration

MeDIP-seq and RNA-seq Data Analysis: The core bioinformatics analyses of these aspects of the project were performed by bioinformaticians, Dr Gareth Wilson and Dr Warren Emmett, respectively. Project specific direction regarding the analysis of these data sets was provided by me and data mining was solely performed by me.

Publications: Aspects of this thesis were published in collaboration with others as detailed below;

Publications

1. **Methylome analysis using MeDIP-seq with low DNA concentrations.** Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S, Butcher LM. *Nat Protoc. 2012 Mar 8;7(4):617-36.*
2. **Methylomic analysis identifies the involvement of migration and adhesion genes in the ageing of primary haematopoietic stem cells.** Taiwo O, Wilson GA, Emmett W, Morris T, Bonnet D, Adejumo T, Pearce DJ and Beck S. *Genome Biology,* **Under Review.**

Oluwatosin O. Taiwo

Acknowledgement

I would like to express a deep gratitude to my supervisors, Dr Daniel Pearce, who offered me the opportunity to undertake this program and allowed me the freedom to develop as a scientist, and Prof. Stephan Beck, who introduced me to the world of epigenetics and whose ingenuity and excellent mentorship, has equipped me for the future.

Similarly, I am appreciative of Dr Edward Grimble, my college mentor, who taught me to believe in myself and Prof. Dallas Swallow, my undergrad mentor, who taught me to think like a scholar. I have indeed been fortunate to have encountered inspiring mentors whose guidance have impacted positively on me.

I am extremely grateful to my project bioinformaticians, Dr Gareth Wilson and Dr Warren Emmett. I could not have completed this study without their help.

Additionally, I would like to thank Dr Lee Butcher, for teaching me the MeDIP technique, and whose enthusiasm for research methods ultimately rubbed off on me. I also thank Dr Andrew Feber for constructive comments on my thesis.

I fully appreciate all members of the Medical genomics group, who provided a fantastic working environment, especially my fellow ‘weekenders’ Paul Guilhamon and Sabrina Stewart, who made weekend toiling fun.

I thank the MRC for providing me with the studentship, and IDEAL (EU FP7), for the funding to complete my project. I also thank the Nature Publishing Group for their permission to include the published MeDIP-seq article, which was originally published in *Nature Protocols*.

A great deal of my gratitude goes to my parents, siblings, and partner, who have been my biggest supporters and have cheered me on. Sorry for being a bore and thanks for understanding.

Above all, I give glory to God for seeing me through to completion.

Abbreviations

5-Aza'dC	5-Aza'deoxyCytidine
5HmC	5-HydroxyMethyl Cytosine
5mC	5-methylCytosine
aDMR	Age DMR
aDMG	Age Differentially Methylated Gene
AHR	Aryl Hydrocarbon Receptor
B6	C57BL/6 Mouse
B6-HSC	B6-derived HSC
BM	Bone Marrow
BML	Bone Marrow Lymphocytes
BS-seq	Bisulfite Sequencing
CAFC	Cobble-stone Area Forming Cell
cDMR	Cancer DMR
CGI	CpG Island
CpG	Cytosine-Guanine dinucleotide
ctDMR	Cell-type specific DMR
D2	DBA/2 Mouse
D2-HSC	D2-derived HSC
DE-aDMG	Differentially Expressed-aDMG
DMEM	Dulbecco`s Modified Eagle`s Medium
DMG	Differentially Methylated Gene
DMR	Differentially Methylated Region
DNAm	DNA Methylation
DNMT	DNA methyltransferase

ECM	Extracellular Cell Matrix
ES/ESC	Embryonic Stem Cell
EtBr	Ethidium bromide
FACS	Fluorescence Activated Cell Sorting
FCS	Foetal Calf Serum
FDR	False Discovery Rate
HBSS	Hank's Buffered Salt Solution
HATs	Histone Acetyltransferases
HDACs	Histone Deacetylases
HgG	Human Gamma Globulins
HMTs	Histone Methyltransferases
HPC	Haematopoietic Progenitor Cell
HSC	Haematopoietic Stem Cell
HSPC	Haematopoietic Stem and Progenitor Cell
IPA	Ingenuity Pathway Analysis
IPSCs	Induced Pluripotent Stem cell
KLS	c-Kit positive (K), Lineage negative (L) and Sca-1 positive (S)
LSP	Lower Side Population
LTC-IC	Long-Term Culture Initiating Cell
LT-HSC	Long-Term Haematopoietic Stem Cell
Ly-bi	Lymphoid Biased HSC
MBD	Methyl-CpG Binding Domain
MBP	mCG binding protein
MDR	Multi-Drug Resistance
MeDIP	Methylated DNA Immunoprecipitation
MeDIP-seq	MeDIP-sequencing
mPCR	MeDIP PCR

MPP	Multi-Potent Progenitor
MSP	Methylation Specific PCR
My-bi	Myeloid Biased HSC
NBMC	Nucleated Bone Marrow Cells
NGS	Next Generation Sequencing
PBS	Phosphate Buffered Saline
PCGT	Polycomb-group-proteins target gene
PCR	Polymerase Chain Reaction
PI	Propidium Iodide
qPCR	Quantitative PCR (Real Time only)
qRT-PCR	quantitative reverse-transcription PCR
RRBS	Reduced Representation Bisulfite Sequencing
RT	Room Temperature
SFEM	Serum-Free Expansion Media
SP	Side Population
SSC	Somatic Stem Cell
ST-HSC	Short Term reconstituting – HSC
STIFA	SCF + TPO + IGF-II + FGF-I + ANGPTL3
tDMR	Tissue DMR
TF	Transcription Factor
TFBS	TF Binding Site
TSS	Transcription Start Site
WBM	Whole Bone Marrow
WGA	Whole Genome Amplified/Amplification
β-Me	β-mercaptoethanol

Abstract

Ageing is a major factor contributing to human morbidity and disease, including cancer. To study the possible involvement of epigenetic changes in ageing, murine haematopoiesis was used as a model system. The key cells determining ageing in this system are thought to be lower side population (LSP) cells of the bone marrow, which are enriched for long-term reconstituting haematopoietic stem cells (LT-HSCs). In this thesis, rare primary LT-HSCs from young, middle-aged and old mice were isolated and phenotyped. A protocol, termed Nano-MeDIP-seq was developed for methylome analyses on such rare cells. DNA and RNA from these cells were then subjected to comprehensive methylome (MeDIP-seq) and transcriptome (RNA-seq) analysis.

Age-related changes in the LT-HSC methylome and transcriptome were observed in this study, many at genes associated with migration/adhesion and not previously implicated in ageing. These changes also include directional (young to old) global loss of DNA methylation of approximately 5%, 111 significantly (FDR < 0.2) age differentially methylated regions (aDMR) and more than one thousand significantly (FDR < 0.05) differentially expressed transcripts. Ingenuity pathway analysis (IPA) identified significant ($p << 0.0001$) age-related decline in B-Cell development and significant ($p << 0.0001$) alterations in pathways and functions associated with cell movement.

A number of genes were identified to be significantly age differentially methylated and differentially expressed. One aDMR, associated with the Serum deprivation protein response (*Sdpr*) gene, was functionally validated to demonstrate a negative correlation between promoter methylation and differential expression.

The findings in this thesis support a model involving an epigenetic dysregulation of the genes that control the interaction between LT-HSCs and their regulatory niche, during physiological ageing.

Table of Contents

LIST OF TABLES

Chapter 1

Introduction

1.1 **Theories of Ageing**

There are many theories of ageing, most of which are modelled under the evolutionary theory of ageing. Common ageing theories include the disposable soma theory of ageing, mutation accumulation theory of ageing and the theory of antagonistic pleiotropy [1]. The disposable soma theory of ageing was formulated by Thomas Kirkwood [2]. It suggests that an organism has to balance the efforts put in maintaining somatic cells with that of reproduction. This theory is likely based on Weissman's germ-plasm theory [3]. This theory suggests that organisms which segregate germ-line and somatic cells, will dedicate resources to the maintenance of the germ-line to ensure reproduction, while the soma (somatic cells) will be left to accumulate mutations overtime, leading to organismal ageing [3]. Likewise, the mutation accumulation theory is based on Peter Medawar's observation that the forces of natural selection decline with age [4]. Natural selection is an active force early in an organism's life, acting to eradicate deleterious mutations and selecting for genes that are beneficial earlier in life [4]. The forces of natural selection in a young organism act to maximize the chances of reproducing and to ensure that favourable genes are passed on to the next generation. However, later acting gene mutations will escape selection and an accumulation of such mutations is thought to contribute to the ageing process [4]. The antagonistic pleiotropy theory of ageing as put forward by G. C Williams [5], proposes that some genes that are beneficial earlier in life are harmful later in life. Such genes are selected for earlier in life, and since the forces of natural selection are less powerful with age, these genes persist [5]. An example in humans is the p53 tumour suppressor gene, which prevents excessive cell proliferation that could lead to cancer [6]. However, constitutive p53 expression in stem cells could result in disrupted cellular homeostasis and accelerated ageing [6].

Several other theories have been put forward in an attempt to explain the ageing process; however, a consistent feature in multicellular organisms is the progressive decline in organ

function and reduction in their regenerative potential. Somatic stem cells (SSC) are multipotent cells that are present in most tissues and responsible for replacing lost and/or damaged cells, thus acting to maintain organ integrity throughout an organism's lifespan.

1.2 **Somatic Stem Cells**

Somatic stem cells (SSC) are long lived cells and they are likely to suffer damage at some point in the course of their existence. Damage to somatic stem cells may be especially catastrophic, due to their ability to self-renew as well as differentiate into more specialized cell types. For example, haematopoietic stem cells (HSCs) are SSCs that generate all blood cell types. Damage to such a cell will be propagated to all subsequent cells arising from this single stem cell, eventually leading to haematological malignancies. It is therefore essential that damage be avoided in such cells. To this end, SSCs have evolved sophisticated damage avoidance mechanisms. Such mechanisms include the expression of drug transporters on their cell surfaces, which help efflux toxins and prevent potential DNA damage [7] . Additionally, SSCs are more likely to undergo apoptosis than attempt a fault prone repair [7,8]. This ensures that DNA replication errors are not retained in this cell population. However, despite these mechanisms, there is evidence that stem cell activity declines with age. The damage-induced stem cell theory of ageing predicts a decline in HSC numbers with age [9]. This theory was thought to explain the observed decline in HSC activity with age; however, it is now known that the number of phenotypically defined HSCs increases with age [9].

1.3 **Haematopoietic stem cells**

Haematopoietic stem cells (HSCs) are self-renewing, multipotent cells that reside in the bone marrow (BM) and are capable of differentiating into all specialized cells of the blood lineage (Fig. 1). The HSC population in the mouse bone marrow is a heterogeneous mix of long-term reconstituting HSCs (LT-HSCs) and short-term reconstituting HSCs (ST-HSCs). These cells, together with multipotent progenitors (MPP), are collectively referred to as KLS

cells or haematopoietic stem and progenitor cells (HSPCs) (Fig.1.1). LT-HSCs are a rare subset of bone marrow cells and are estimated as less than 0.01% in young mice [9]. These cells are largely non-cycling (quiescent) and most remain in the G0 phase of the cell cycle during steady state haematopoiesis [10].

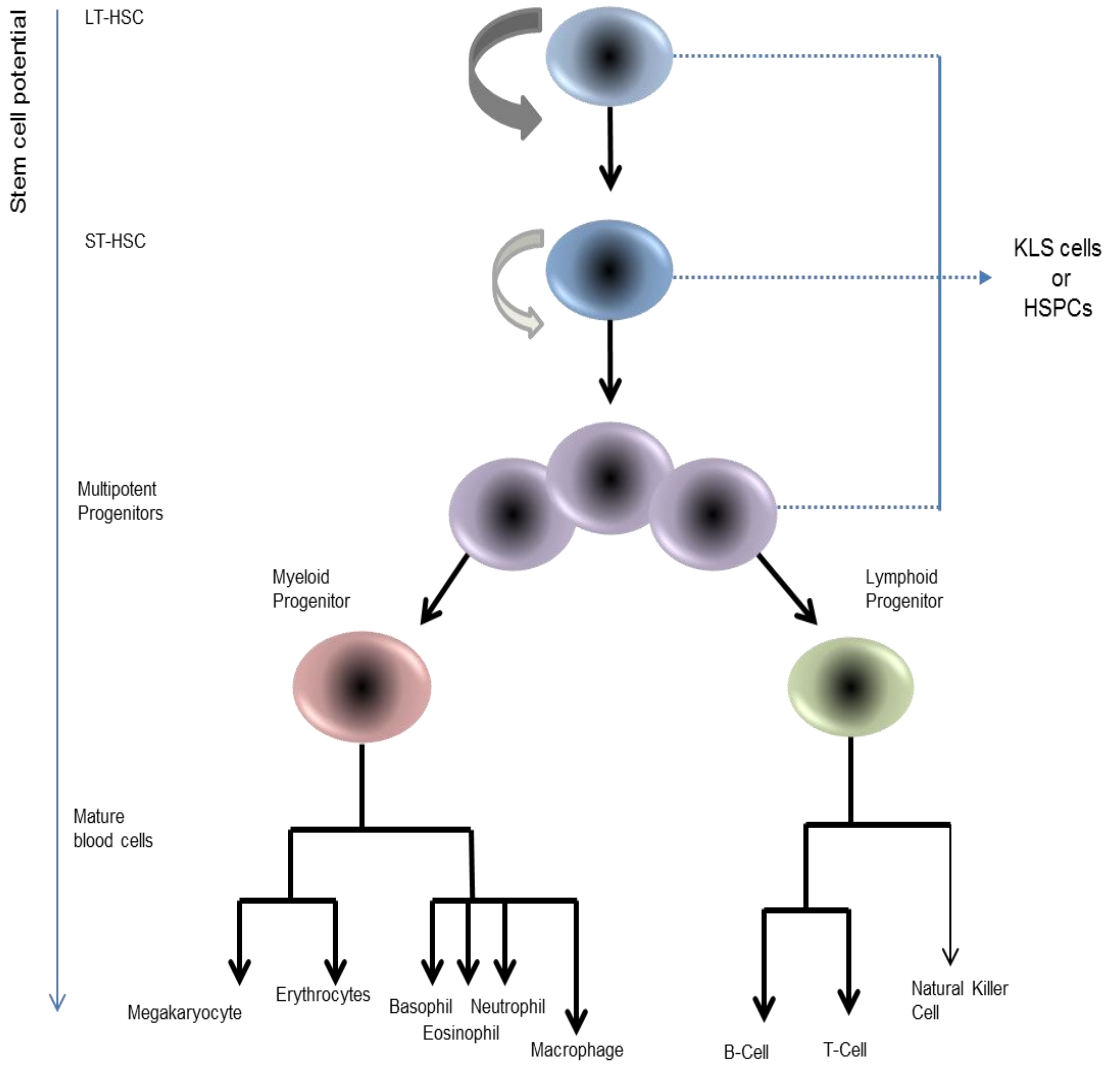


Figure 1-1: Hierarchy of cells in the mouse bone marrow.
The most primitive cells in this hierarchy are the long-term reconstituting HSCs (LT-HSCs). ST-HSCs are short-term reconstituting HSCs. Curly arrow denotes self-renewal. KLS cells are c-Kit positive, Sca-1 positive and lineage negative cells. HSPC stands for haematopoietic stem and progenitor cells. A cell above each arrow is capable of differentiating into the cells below it, forming a succession of increasingly committed downstream progenitors (not shown), through to mature short lived blood cells.

Canonical HSC properties include; expression of the c-kit receptor (K), the absence of lineage markers (Lin- or L) and the presence of the stem cell antigen-1 (Sca-1 or S) [11]. However, as mentioned earlier, KLS cells consist of both haematopoietic stem and progenitor cells. Additional markers of adult mouse HSCs, such as; the absence of CD34 [12,13], high and negative expression of Slam markers, CD150 and CD48 respectively [14,15], and Hoechst dye exclusion [16], are therefore required to enrich for LT-HSCs. Antibody or Hoechst labelled HSCs are typically isolated by fluorescence activated cell sorting (FACS).

Hoechst dye exclusion is a cheap and easy method for isolating LT-HSCs [16]. This method involves the active efflux of the Hoechst 33342 DNA binding dye by the ABCG2 multidrug resistant (MDR) transmembrane transporter, which is expressed on cell surfaces [17]. ABCG2 is exclusively expressed by stem cells and is highly expressed on HSCs [17]. As a result, HSCs retain the least Hoechst staining compared to other haematopoietic cells and are distinctively separated from other cells when excited by ultraviolet (UV) lasers, during FACS analysis. HSCs identified in this manner are referred to as side population (SP) cells (Fig. 1.7) and the SP, especially the lower SP, fraction of the mouse bone marrow has been shown to contain a cell population that is highly enriched for phenotypically and functionally determined LT-HSCs [7,16,18,19]. The availability of several well established markers, necessary for the identification and isolation of a homogeneous LT-HSC population, makes HSCs a convenient model for the study of SSC during mammalian ageing.

1.4 **Assays for studying HSCs**

Regardless of the availability of reliable markers to identify and isolate HSCs, functional assays are still required to identify *bona fide* HSCs. Such assays include *in vitro* and *in vivo* assays that examine the ability of putative HSCs to differentiate into mature progenitors, over a relatively long period of time. The latter property measures self-renewal, which is the key difference between true HSCs and MPPs (Fig. 1.1). Assays, such as the cobble-stone

area forming cell (CAFC) and the long-term culture initializing cell (LTC-IC) assays, can be used to confirm HSC activity *in vitro*.

1.4.1 **CAFC Assay**

CAFC involves the *in vitro* culture of putative HSCs on supportive stromal feeder cells, for a minimum of 28 days, followed by the visual confirmation (using a light microscope) of the presence of 'cobble-stone areas' (phase dark cells) [20,21]. It is assumed that mature progenitor cells would have been exhausted by 28 days; therefore, the observation of colonies after this period is indicative of the primitiveness of the original test cells. A drawback of this assay is that it is a retrospective measure of primitiveness and validated test cells cannot be recovered. Additionally, self-renewal cannot be measured directly. The CAFC assay is therefore unable to differentiate between LT- and ST- HSCs or MPPs, and is thus, only useful for the detection of HSPC activity.

1.4.2 **LTC-IC**

LTC-IC [22] is similar to the CAFC assay in that they both involve the *in vitro* culture of test cells on stromal feeder cells; they however differ in their end-points. Following the initial culture of test cells for approximately 4 weeks, LTC-IC cells are harvested and seeded on semi-solid (methyl-cellulose) media, to determine the presence of colony forming cells (CFC). The rationale behind this is that only primitive cells will differentiate, over prolonged culture periods, into progenitor cells (CFCs), which are then able to form colonies of mature blood cells when transferred to a suitable secondary media. Similar to CAFC above, LTC-IC and CFCs do not provide a measure of LT-HSC activity [23].

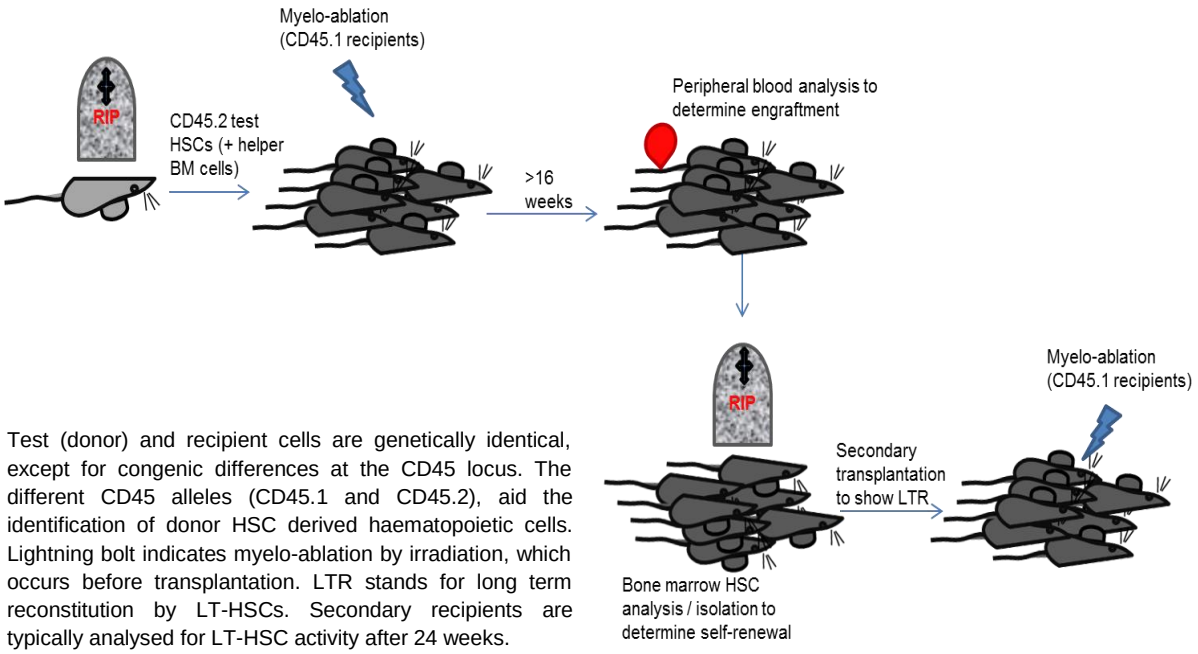
Other *in vitro* assays include the short-term *ex vivo* (<12 days) maintenance of HSCs in serum-free liquid medium and growth factors. *Ex vivo* assays are typically used to expand HSCs numbers in vitro, while maintaining them in an undifferentiated state. This is achieved by the addition of several growth factors such as Angiopoietin-like proteins, Fibroblast

growth factor 1 (FGF-1), Insulin-like growth factor 2 (IGF-2), Thrombopoietin and Stem cell factor (SCF) [24,25]. The advantage as well as disadvantage of this assay is its ability to maintain HSCs in an undifferentiated state. This is useful for the expansion of HSC numbers for various therapeutic purposes, however, the fact that expanded HSCs cannot differentiate in this system could be a disadvantage. This is because although cell surface markers may remain unchanged, one cannot be sure of the functionality of *ex vivo* expanded HSCs, which can only be ascertained by their ability to differentiate into mature blood cells and permanently reconstitute haematopoiesis.

1.4.3 **Transplantation assays**

In vivo reconstitution assays are the gold standard for determining HSC activity. These assays involve the transplantation of test (donor) HSCs into host mice, followed by the assessment of its engraftment and long-term reconstituting abilities [26]. Many variations of this assay exist; these include the use of myelo-ablated or non-ablated hosts, as well as a variation in the requirements for serial transplantation. Transplantation of HSCs into genetically different hosts, such as; across mouse strains or xenografts of human donor HSCs into mice, can also be performed. This is typically achieved by transplantation into immunocompromised hosts e.g. SCID mice. Nevertheless, the end-point assessment for HSC functionality remains the same, and must provide clear evidence of its ability to self-renew and reconstitute haematopoiesis. A typical HSC transplantation assay is the competitive repopulating unit (CRU). CRU assays (Fig. 1.2) [26] involve the transplantation of HSCs into myelo-ablated hosts, such that true HSCs will permanently reconstitute haematopoiesis in such hosts. Serial transplantation of HSCs into secondary hosts is often used to further ascertain self-renewal in LT-HSCs.

Figure 1-2: Competitive repopulating unit (CRU) assay.



Test (donor) and recipient cells are genetically identical, except for congenic differences at the CD45 locus. The different CD45 alleles (CD45.1 and CD45.2), aid the identification of donor HSC derived haematopoietic cells. Lightning bolt indicates myelo-ablation by irradiation, which occurs before transplantation. LTR stands for long term reconstitution by LT-HSCs. Secondary recipients are typically analysed for LT-HSC activity after 24 weeks.

1.5 **The Haematopoietic Stem Cell Niche**

The concept of a HSC niche was first proposed in 1978 by Schofield [27]. The niche as referred to here, is a regulatory micro-environment where HSCs reside, and are supported by specialised growth-factor secreting cells, and various other cytoskeletal and cell-adhesion molecules [28]. HSCs reside in various niches during development (Fig.1.3 [29-31]) before migrating to their final location, which is the bone marrow. This migration is thought to occur in embryos after approximately 18 days (E18) post conception (dpc) [29] and adult HSCs subsequently remain in the bone marrow throughout life. An exception to this is observed during pathological extramedullary haematopoiesis, where HSCs leave the bone marrow for other sites such as the spleen or liver, to proliferate and differentiate [32].

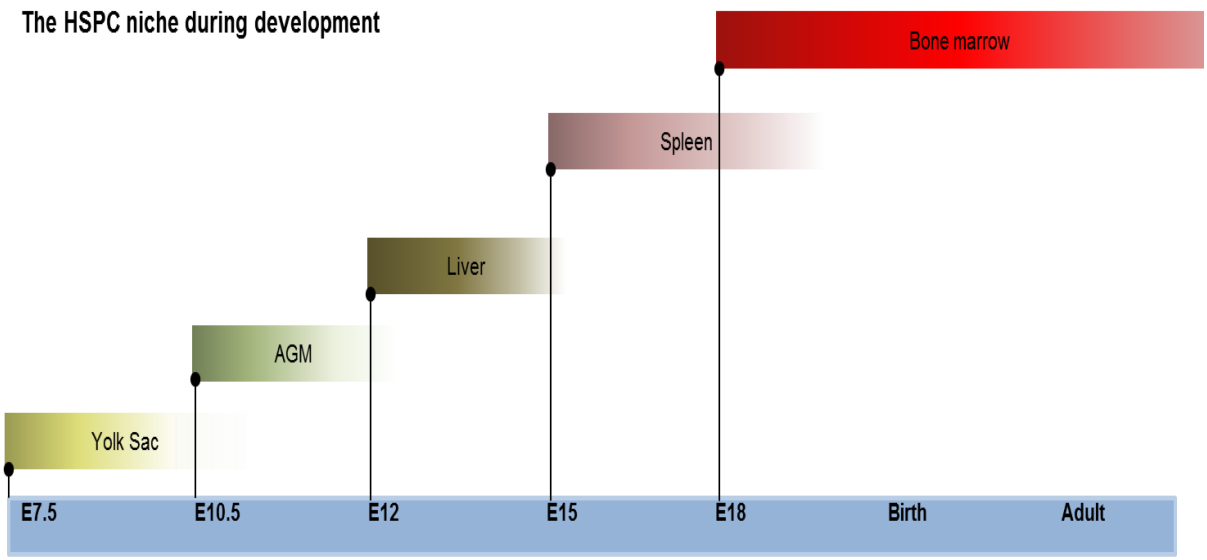


Figure 1-3: HSPCs reside in, and migrates to, various niches during ontogeny.
Primitive haematopoiesis begins in the yolk sac, while definitive haematopoiesis is first observed in the AGM of day 10.5 embryos (E10.5). The bone marrow is the final niche in adult haematopoiesis. AGM, Aorta-gonad mesonephros.

This thesis focuses on adult murine HSCs, which are resident in the bone marrow. The bone marrow is thought to consist of two anatomically distinct niches i.e. the endosteal and the vascular niche (Fig. 1.4 [33,34]). The endosteal region of the trabecular bone (endosteal niche) is thought to be involved in LT-HSC regulation and the maintenance of quiescence during steady state [34]. Osteoblasts are the main regulatory cells in this niche and they secrete key HSC factors such as osteopontin, angiopoietin, and stromal derived factor-1 (SDF-1, also known as CXCL12) [35-37]. Proliferation and differentiation into more specialized blood cells are thought to occur in a second niche, the vascular niche of the perivascular area [38], and resident cells such as megakaryocytes, mesenchymal stem cells and CXCL12 abundant reticular (CAR) cells, also produce angiopoietin and SDF-1. This overlap in regulatory factors and the observation that HSCs are able to interact simultaneously with both niches, without a change in their functional and cycling properties [14], makes the concept of distinct HSC niches untenable, and is currently a subject of high debate in HSC biology [39]. Although the exact mechanism by which HSCs are regulated in their niche is currently unclear, the bone marrow has been proven to be important in HSC maintenance and regulation [33-35,40].

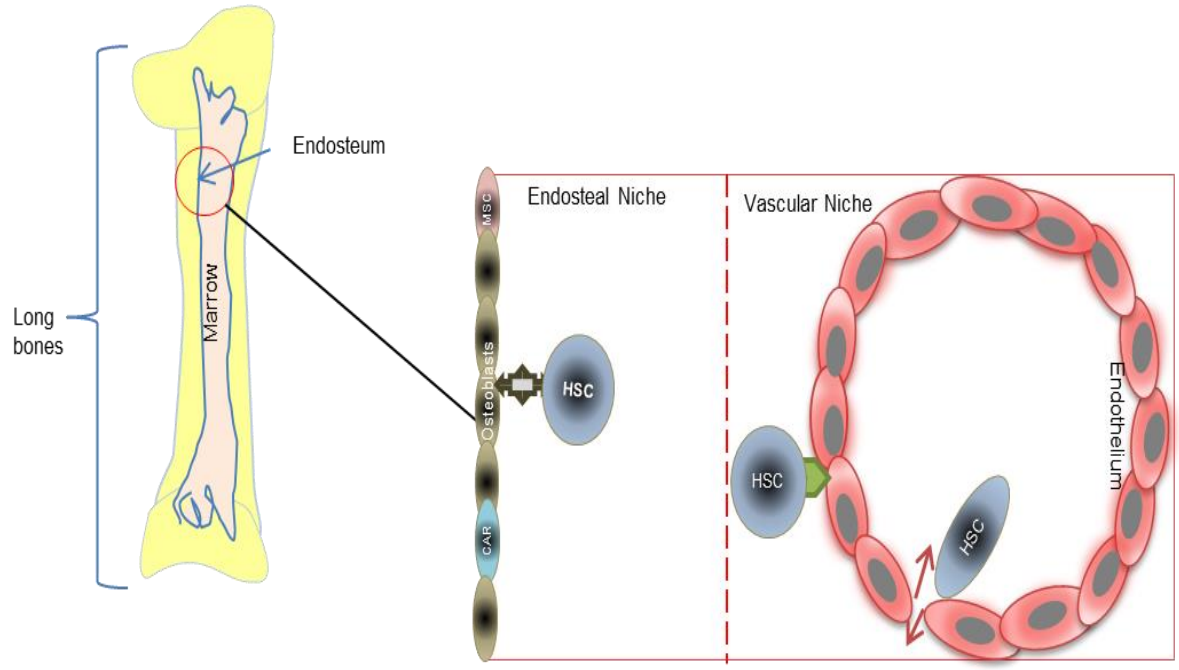


Figure 1-4: The endosteal and vascular niche of the mouse bone marrow.
Osteoblasts of the endosteal niche and sinusoid endothelium of the vascular niche both contribute in the maintenance of HSCs in the bone marrow. Other specialised cells such as CAR cells and MSCs are also present in both niches and these, as well as the main cells of each niche, secrete growth factors that regulate HSCs. It is currently unclear whether these niches are truly distinct or if they form parts of a single HSC niche. CAR- CXCL12 abundant reticular; MSC – Mesenchymal stem cell; HSC - Haematopoietic stem cell

1.6 **HSC Regulation**

The two key defining properties of HSCs are the ability to differentiate into all mature cells of the blood lineage and to self-renew for a relatively long period of time. The latter property is typically indexed by persistent reconstituting ability, following two to three rounds of serial transplantation (Section 1.4.3). As detailed above, the niche plays an important role in regulating these properties, but many other factors have also been shown to be involved in HSC regulation. Majority of LT-HSCs are quiescent at steady state [41], however, it is frequently necessary for these cells to proliferate in response to systemic signals, for the replenishment of all blood cell types. Therefore, a balance between HSC quiescence and proliferation is required to maintain homeostasis and prevent exhaustion. This is achieved by the action of transcription factors (TF) such as Hoxb4, Pbx1, JunB and Bmi-1. Hoxb4 positively regulates HSC self-renewal and cells overexpressing this TF exhibited an enhanced expansion, during *ex vivo* culture [42]. Hoxb4 also appears to collaborate with p21

in the regulation of HSC self-renewal [43]. p21 is a cyclin-dependent kinase inhibitor (CKI) that governs HSC quiescence, by regulating its entry into the cell cycle [44]. Overexpression of Hoxb4 in p21 mutant HSCs resulted in a significant increase in self-renewal, compared to wild-type (wt) HSCs or HSCs overexpressing Hoxb4 alone [43]. Conversely, Pbx1 regulates HSC quiescence and Pbx1 deficient LT-HSCs lost their self-renewal activity due to excessive proliferation, concomitant with exhaustion, and displayed ST-HSC functionality and a gene expression profile akin to that of MPPs [45]. Similarly, Bmi-1 mutants were unable to self-renew, suggesting a role for Bmi-1, in the control of HSC self-renewal [46]. Other factors such as cytokines and signalling pathways such as the Tgf-β, Wnt and Notch pathways have all been shown to be important regulators of HSC properties [19,47,48].

1.7 **Epigenetic regulation of HSCs**

Epigenetics, as defined by Adrian Bird, is "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" [49]. HSCs are required to maintain a tight balance between self-renewal and differentiation into mature cells of vastly different gene expression profiles [50]. As a result, a degree of plasticity in gene expression is necessary to allow fluid changes from quiescence, to proliferation and to differentiation. Epigenetic mechanisms are important in phenotypic plasticity [51], and are thus likely to be involved in the regulation of HSCs. Indeed, intricate epigenetic modifications at the promoters of lineage restriction genes, show a distinct pattern between HSCs and mature blood cells [52]. Epigenetics encompass the modification of histone tails, DNA methylation and gene regulation by non-coding RNA (ncRNA).

1.7.1 **Histone Modification**

Histones are proteins that are involved in the high order packaging of DNA into chromatin, within cells. The main proteins involved in this process are the core histones; H2A, H2B, H3 and H4, and the H1 linker histone. Key post translational histone modifications include acetylation by histone acetyltransferases (HATs) and methylation by histone

methyltransferases (HMTs). The recognition of these modifications by other (non-histone) proteins, results in dynamic changes in chromatin structure i.e. from transcriptionally inactive heterochromatin to active euchromatin and vice versa (Fig. 1.6) [53]. Gene expression is therefore regulated by histone modifications, which influence the accessibility of polymerases and TFs to DNA. Several studies have highlighted the importance of histone modifications in the determination of HSC fates [52,54,55]. HSCs exhibit a low level expression of lineage specific genes [56,57]. This is thought to be important in priming them towards lineage commitment, and the presence of bivalent (activating and repressive) histone marks on such genes [54], suggests that histone modifications are important in the regulation of lineage commitment in HSCs.

1.7.2 **DNA Methylation**

DNA methylation (DNAm) is an important epigenetic modification in the regulation of gene expression, and involves a stable but reversible addition of a methyl group to the carbon-5 position of cytosines, without a change in the underlying DNA sequence [58]. This process occurs *de novo*, through the action of DNA methyltransferase 3 (Dnmt3a and b), or as a maintenance mechanism in hemi-methylated DNA by DNA methyltransferase 1 (Dnmt1). In mammalian genomes, DNAm exists as variations of 5-methyl cytosine (5mC). 5mC DNA modifications are found mostly at Cytosine-Guanine (CpG) dinucleotides (mCG) and are recognised by protein complexes with methyl-CpG binding domains (MBD), which often act with other chromatin remodelling proteins, to modify gene expression [58,59].

mCG is enriched in the mammalian genome and its role in the regulation of gene expression is well appreciated. Genomic regions of high CpG density are known as CpG islands (CGI) and these are present in the promoter region of approximately 56% human genes [60]. Promoter CGIs are typically hypomethylated and DNAm of this region is associated with gene repression [60] (Fig. 1.5). In contrast, independent DNAm within gene

bodies maintain genome stability by preventing nonsense transcription from cryptic start sites and by suppressing intragenic repeat and mobile elements [60,61].

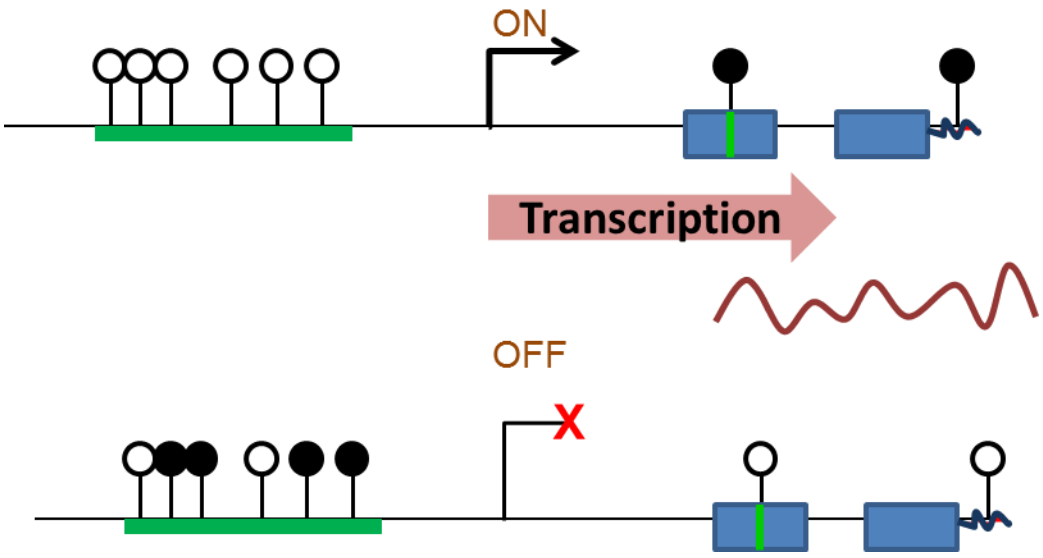


Figure 1-5: Gene regulation by DNA methylation.
Top; promoter (green bar) hypomethylation (open circle lollipops) allows gene expression i.e. transcription of DNA to protein coding mRNA (red squiggle), while gene body (blue bars) hypermethylation (closed circle lollipops) maintains genome stability, by repressing repeat elements (blue squiggle) and transcription from cryptic start site (green line). Bottom; transcription is repressed (red cross) in the case of promoter hypermethylation and gene body hypomethylation likely results in genome instability. Changes in promoter DNAm can modify gene expression, independently of those within the gene body.

As mentioned previously, there are other variations of 5mC. These include non CpG (mC) and 5HmC methylation. mC is structurally identical to mCG but occurs in the absence of a guanine nucleotide and is asymmetrical with regards to its presence on DNA strands [62]. In contrast to mCG which is equally represented in most mammalian cells, mC is only enriched in ESCs [63,64] and perhaps in SSCs. 5HmC, is another form of mammalian DNAm and involves the conversion of 5mC to 5HmC, in a 2-oxoglutarate (2OG) and Fe (II) dependent manner [65]. This conversion is catalysed by Tet proteins [66] and Tet1, Tet2 and Tet3 have all been shown to act as mCG dioxygenases [67]. Although the exact function of 5HmC is currently unknown, its presence on ES cells gene promoters coincides with increased gene expression [68]. Additionally, the downregulation of Tet1, as well as a decrease and increase in promoter 5HmC and 5mC respectively, is observed during ES cell differentiation [68].

These findings are consistent with a dynamic relationship between 5HmC and 5mC, as well as a possible mechanism for active demethylation, through the action of Tet proteins.

Although, DNAm and histone modifications may act in isolation to modulate gene expression, recent evidence indicates that these mechanisms may cooperate during gene regulation (Fig. 1.6) [59,69]. This is seen in the repression of p53BP2, as a result of the interaction between MBD1 and SETDB1, in the formation and maintenance of heterochromatin. MBD1 is a mCG binding protein (MBP) while SETDB1 is a histone methyltransferase (HMT) and MBD1 bound to mCG, recruits SETDB1 into a repressive complex that mediates the methylation of H3K9, resulting in gene repression [59].

As well as modulating gene expression by promoter methylation or through interactions with other epigenetic factors, DNAm has various functional consequences that indicate a possible role in the regulation of HSCs. The following sections detail the aspects of DNAm that 1) are relevant to this study, and 2) provide evidence for the possible involvement of DNAm in HSC ageing.

1.7.2.1 ***DNAm During Development***

The first epigenetic regulation of the mammalian genome occurs very shortly after fertilization. Reprogramming of the zygote epigenome occurs before any DNA replication, and involves rapid demethylation of the paternal genome [70,71]. The maternal genome also undergoes passive demethylation following multiples rounds of DNA replication [70,71]. While demethylation of the paternal genome is clearly an active mechanism, the factors involved in this event are currently unknown. However, DNA replication-independent accumulation of 5HmC, and Tet3 knockdown experiments, in one-cell embryos suggest that this process involves the direct conversion of 5mC to 5HmC by Tet3 [72]. Remethylation of the zygote genome begins in E6.5 mouse epiblasts, during implantation, by Dnmt3b catalysed *de novo* methylation of gene promoters [73]. Genes involved in pluripotency are targeted for methylation in early embryos, possibly as a mechanism to limit the differentiation

capacity of epiblast cells [73]. Reprogramming of the epigenome during embryogenesis is important in the maintenance of imprinting and in preventing the transmission of deleterious DNAm and/or epimutations, which could lead to diseases or shortened lifespan [74,75].

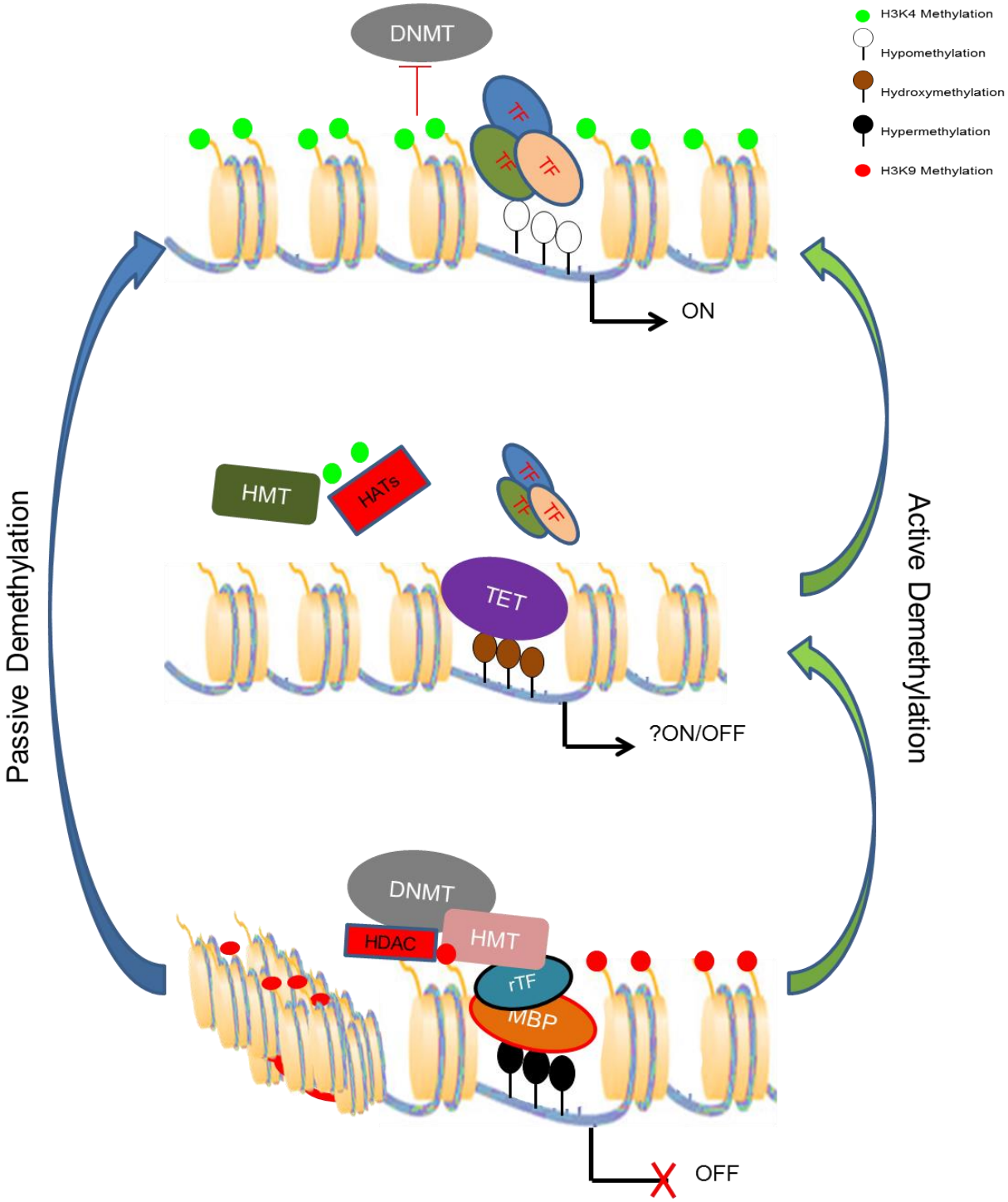


Figure 1-6: Simplified model of gene regulation.
Top; Hypomethylation of promoter CGIs and histone modifications, such as acetylation and H3K4 methylation, inhibit DNMT binding and result in an open chromatin (euchromatin). This allows the binding of transcription factors (TF) to DNA, thus allowing gene expression. Middle; putative intermediate step between gene expression and repression by Tet mediated DNA demethylation. HMT - histone methyltransferases, HAT – Histone acetyltransferase. Bottom; heterochromatin formation and gene repression, following histone deacetylation by histone deacetylases (HDAC) and promoter hypermethylation by DNMTs. mCG binding protein (MBP) binds to methylated DNA and recruits co-repressors (rTF) and repressive HMTs, e.g. SETDB1. This results in a stabilisation of heterochromatin and gene silencing [59,67,69,74].

1.7.2.2 ***DNAm and Lineage Specification***

The link between cell-type specific gene expression and DNA methylation was first demonstrated for the MASPIN gene, where its promoter was found to be unmethylated only in the cells that expressed the gene [76]. The advent of next generation methylome analysis techniques (Section. 1.11.2) has aided further elucidation of this mechanism, and it is now widely accepted that this phenomenon extends beyond single gene promoters. For example, analysis of several CGIs in human blood, brain, muscle and spleen, revealed multiple inter- and intragenic CGIs that were differentially methylated between these tissues [77]. These tissue-differentially methylated regions (tDMRs) were found to be enriched for developmentally relevant genes such as Pax and Hox genes [77]. Similarly, a study by the same group identified several cell-type specific DMRs (ctDMR), between human $CD4^+$ blood and brain cells [78]. ctDMRs were enriched at intragenic CGIs and although, many of these were not associated with promoters (orphan CGIs), DNAm at gene-associated intragenic CGIs was negatively correlated with gene expression. This indicates a functional role for intragenic CGIs in the regulation of cell specific gene expression. These findings are in slight contrast to related studies, which found tDMRs to be enriched at CGI shores [79-81], which are regions of relatively low CG density that are found approximately 2 Kb upstream of CGIs. CGI shore-tDMRs overlapped with cancer differentially methylated regions (cDMRs) and are evolutionary conserved between mice and humans [80]. Likewise, induced pluripotent stem cells (IPSCs) and the mature fibroblasts from which they were derived, were differentially methylated at CGI shores. These DMRs, termed rDMRs, showed considerable overlap with CGI shores-tDMRs [81]. These suggest that CGI shores are important regulatory regions in the maintenance of cell identity. Conversely, IPSCs generated from distinct cells retain certain DNAm marks that are reminiscent of the cells from which they originated [79]. Irrespective of the precise location of the DMRs highlighted above, these studies demonstrate that DNAm is important in high level cell lineage and tissue specification, as these marks define cellular identity and persist even after genome-wide reprogramming.

They also suggest that DNAm is involved in the choice between lineage specification and pluripotency and a disruption of this mechanism is likely involved in carcinogenesis.

1.7.2.3 ***DNAm in Cancer***

The link between global DNA hypomethylation and cancer was made several years ago [82] and there has since been a remarkable improvement in our understanding of the dynamic involvement of DNAm in oncogenic transformation. Indeed, several instances of global and gene specific changes in DNAm have been observed in almost all cancer types [83,84]. Global loss of DNAm results in chromosome instability (Section 1.7.2), and increased tumour incidence, as demonstrated in mice carrying hypomorphic alleles of Dnmt1 [85,86]. Likewise, hypermethylation at the promoters of several tumour suppressor genes, and their subsequent downregulation, is known be involved in carcinogenesis [83]. Additionally, overexpression and silencing of specific genes in stem cells, through aberrant DNAm, has been proposed as an initial step in neoplastic transformation [84].

1.7.2.4 ***DNAm in Stem cells***

The involvement of DNAm in embryonic stem cell (ESC) maintenance and differentiation is well appreciated [58,67,87]; but, until recently, its relevance in SSCs has been relatively understudied. Since the commencement of this study, several studies have highlighted the role of DNAm in HSPCs regulation. Such studies have demonstrated the importance of DNAm in maintaining HSC homeostasis and Dnmt mutant mice show defects in HSC self-renewal and differentiation [88-90]. Similarly, global hypomethylation in mice with reduced *Dnmt1* activity resulted in phenotypically defined LT-HSC expansion, but markedly impaired self-renewal *in vivo*, as well as a skewed lineage potential towards myeloerythropoiesis, at the expense of lymphopoiesis [88]. Lineage skewing here was as a result of markedly altered gene expression in HSCs, to one more similar to that of myeloerythroid progenitors. These changes in gene expression include the downregulation of several genes that are involved in self-renewal, and an upregulation of myeloid specification genes. Conversely,

Dnmt3a mutant HSCs upregulate multipotency genes and showed enhanced *in vivo* self-renewal at the expense of differentiation into mature blood cells [90].

The studies highlighted above (Section. 1.7.2), provide evidence for the role of DNA methylation in mammalian gene regulation and in the maintenance of cellular identity and integrity. Several of these studies also suggest an involvement of DNAm in HSC maintenance and differentiation, and in support for the latter, significant changes in DNAm were identified during haematopoietic lineage specification [91]. In the next sections, I discuss the ageing of HSCs and the possible mechanisms that are involved in this process.

1.8 **Haematopoietic Stem Cell Ageing**

HSCs from old mice show a reduction in the ability to efficiently reconstitute haematopoiesis in lethally irradiated hosts [7,9,92-94]. This functional defect is accompanied by an age-dependent increase in the number of phenotypically defined HSCs (Fig. 1.7) and a skewed differentiation potential towards myelopoiesis, at the expense of lymphopoiesis [93,95]. Additionally, aged HSCs show reduced ability to home to the bone marrow [7,96,97], which would affect their regulatory interaction with the niche. These reports are consistent with a decline in immune functions and an increase in the incidence of myelo-proliferative cancers with age [98].

Functional changes in HSCs during ageing have been well characterised, but the exact mechanism directing this change remains elusive. Nevertheless, various mechanisms have been postulated and these are summarised in the next section.

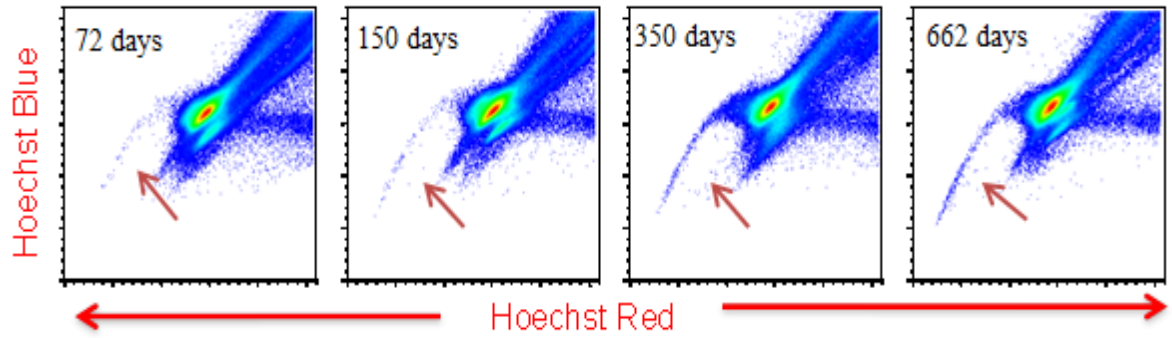


Figure 1-7: FACS plot of mouse side population (SP) cells.
These cells (arrowed) increase with age [7]. Figure was provided by Dr Daniel Pearce.

1.9 Theories of Haematopoietic Stem Cell Ageing

1.9.1 Intrinsic Ageing of HSCs

Gene expression analyses have provided evidence for an intrinsic change in HSCs during ageing [7,94,99]. A study by Rossi et.al [94] suggested that lineage skewing in aged HSCs is a result of upregulated myeloid specification transcripts and a corresponding downregulation of lymphoid transcripts during ageing [94]. The cell intrinsic nature of this change was further demonstrated by transplantation assays, which showed that exposure of HSCs from young mice to an aged niche, was not sufficient to recapitulate the defects observed in HSCs during ageing [94]. The cell intrinsic theory of HSC ageing was confirmed by another study, which observed an NF-kB mediated dysregulation of HSC gene expression during ageing [99]. Likewise, HSC activity in chimeric mice provided *in vivo* evidence for the intrinsic nature of HSC ageing [92]. Briefly, cellular chimeras were generated from two different mouse strains, C57BL/6 and DBA/2. The former is a long lived strain, living up to three years, while the latter is shorter lived and survives for up to two years. Haematopoiesis in young chimeras was evenly supported by both C57BL/6 and DBA/2 derived HSCs, however this was not the case in aged chimeras. Coinciding with the end of their natural lifespan, DBA/2 derived HSCs (D2-HSCs) appeared to have senesced in two year old chimeras,

leaving haematopoiesis to be sustained by C57BL/6 derived HSCs (B6-HSCs). Serial transplantation of 'senescent' D2-HSCs, into young chimeras, enforced their continued contribution to haematopoiesis, albeit for a few months [100]. Findings from this study support an intrinsic model of HSC ageing, as D2-HSCs appeared to age earlier than B6-HSCs, despite being exposed to identical niches. This study also hints at a role for the niche in HSC ageing, as 'senesced' D2-HSCs were temporarily reactivated upon exposure to a younger, and unoccupied, niche. As senescence involves the irreversible cessation of all replicative activity, it is unlikely that D2-HSCs in two year old chimeras were senescent, as they maintained the ability to proliferate in certain conditions. It is also possible that transplanted D2-HSCs were in fact ST-HSCs, which would explain their ability to proliferate only for a short period of time after transplantation. This is however unlikely, as these cells continued to be reactivated, after at least two rounds of serial transplantation [100]. Taken together, these findings suggest that intrinsic programmes cause HSCs to become deeply quiescent during ageing and certain niche conditions or indeed the process of transplantation and homing to a new unoccupied niche could restore or alter the functionality of aged HSCs.

1.9.2 **Clonal Compositional change of HSCs during ageing**

It has recently been shown that the mouse LT-HSC compartment consists of lineage biased subpopulations, which show a difference in their ability to differentiate into either the myeloid or the lymphoid lineage [101]. These subpopulations, termed myeloid biased (My-bi), balanced (Bala) and lymphoid biased (Ly-bi) HSCs, can be isolated based on their expression of the CD150 antigen [19,102] and their differential ability to efflux the Hoechst dye during SP analysis [19]. Lower SP and $CD150^{hi}$ LT-HSCs are thought to be enriched for My-bi clones while the upper SP and $CD150^{low}$ clones are enriched for lymphoid biased LT-HSCs [19,102]. The lineage bias shown by these cells is stable with age, such that specific subpopulations isolated from mice at any age point, consistently displays the same propensity towards a particular cell lineage [19]. This is particularly important during HSC

ageing, as it is the My-bi enriched lower SP fraction of LT-HSCs that increase with age [7,19]. This suggests that HSCs change in composition with age and the increase in myelopoiesis observed during ageing, is a consequence of this. My-bi clones have been shown to be more quiescent and longer lived that Ly-bi clones [19]. Additionally, My-bi clones show a greater ability to self-renew and generate Ly-bi clones, than the converse [102], suggesting that My-bi cells may be the most primitive HSC subsets. This ability to self-renew and persist for longer periods of time, than their Ly-bi counterpart, probably accounts for their accumulation with age and subsequent enrichment in old mice.

1.10 **Molecular Mechanisms of Haematopoietic Stem Cell Ageing**

Functional and expression studies of HSCs during ageing have identified a catalogue of genes, as well as HSC properties, that are altered with age [7,9,94,103]. Differentially expressed genes range from those that are involved in cell signalling pathways [94] and polarity [104], to those implicated in DNA damage, cell cycle and cancer [105-107]. Although these findings represent a significant improvement in our understanding of HSC ageing, the exact mechanism driving these age-related changes remains unknown. Stem cells are long lived and are therefore likely to accumulate DNA damage in the course of their lifespan [108]. The self-renewing properties of somatic stem cells also make them ideal targets for oncogenic transformation [84] and it is possible that age related alterations in HSCs are the initial steps to neoplasia (Fig. 1.8). Indeed, Chronic myeloid leukaemia (CML) is a cancer that becomes prevalent with age and has been shown to arise from oncogenic transformation of HSCs in mice [109]. Likewise, age-dependent promoter hypermethylation of genes that promote stem cell differentiation, suggest that aberrant DNAm may alter stem cell potentials during ageing [110]. Although the preferential production of myeloid cells in old age is now thought to be a result of clonal compositional changes in HSCs during ageing [19,102], it remains unclear why these clones increase in quantity or why there is a loss in their functionality with age. As discussed above, DNAm is important in HSCs regulation and its aberration is implicated in carcinogenesis [82,110]. Age is one of the greatest risk factors

in carcinogenesis, and cancers such as CML and other myelodysplastic syndromes (MDS), which have a high incidence in the older population, are currently being treated with DNA methylation altering agents such as 5-Aza'deoxyCytidine (5-Aza'dC) [111].

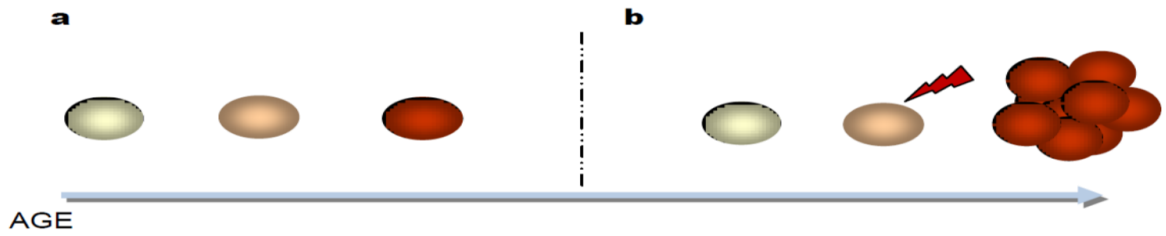


Figure 1-8: Ageing is a major risk factor in carcinogenesis.
a) Inevitable age related HSC changes could leave these cells vulnerable to neoplastic transformation; b) an additional hit (lightning bolt) on ageing HSCs could trigger malignant transformations, leading to age associated leukaemias.

1.10.1 **DNA Methylation and Cellular Ageing**

The otherwise identical genomes of monozygotic twins are differentially methylated with age [112], and similar to changes in DNAm during carcinogenesis, human $CD4^+$ blood cells gain DNAm at CGI promoters and become globally hypomethylated during ageing [113]. Likewise, multiple mouse tissues become differentially methylated with age [114]. An underlying problem of old age is a decline in the body's ability to respond adequately to damage; a problem that is attributable to the loss of phenotypic plasticity in SSCs. In the event of damage, normal SSCs respond to signals, which instruct them to differentiate and replace lost cells [115]. HSCs are quiescent at steady state [41], but proliferate rapidly to reconstitute haematopoiesis in myeloablated hosts. The transcriptome of quiescent cells has been shown to be distinct from that of proliferating cells [116]. Such changes in gene expression, and thus the switch from quiescence to proliferation (or vice versa), are likely underlined by epigenetic modifications. Indeed, recent studies of HSPC differentiation have identified several changes in DNAm, during lineage commitment [91,117].

Additionally, various studies have shown a dysregulation of DNAm in SSCs during ageing [110,117]. Teschendorff et.al found that human polycomb-group-proteins target genes (PcGTs), which are stably repressed in stem cells, also become hypermethylated with age in mature blood cells [110]. Similarly, analysis of human HSPCs identified several PcGTs to be *de novo* methylated at gene promoters, during HSPC ageing [117]. Additionally, several myeloid specification genes become hypomethylated in aged HSPCs [117]. These findings indicate an involvement of DNAm in the ageing of HSCs, and suggest that myeloid bias in aged HSCs is a result of DNAm-dependent upregulation of myeloid specification genes. It is also likely that promoter hypermethylation of PcGTs leads to enhanced self-renewal, at the expense of differentiation, culminating in the HSC expansion that is observed in aged mice.

In order to further elucidate the mechanisms of HSC ageing, I investigated the dynamics of DNAm, during mouse LT-HSCs ageing. In the next sections, I describe the techniques used to study DNAm in this thesis and conclude this Chapter with an overview of my investigations and findings.

1.11 **DNA Methylation Profiling**

Common techniques for DNAm analysis are bisulfite conversion based or enrichment based [118]. Each of these methods have multiple strengths and drawbacks, however, the most appropriate technique for any study is ultimately dependent on the type of information required, and the feasibility of applying the selected method to the samples being investigated. Nevertheless, the initial consideration, when deciding on a suitable method for DNAm analysis, is the scale at which the investigation is required. This can be separated into two categories; locus-specific and global DNAm (methylome) analysis.

1.11.1 **Locus-Specific Analysis of DNAm**

This approach requires *a priori* knowledge of the regions that will be analysed, and is commonly used to determine the DNAm status of candidate regions. Most locus-specific

(LS) DNAm techniques are bisulfite conversion and PCR based. Bisulfite conversion involves the chemical treatment of DNA with sodium bisulfite, which results in the deamination of unmethylated cytosine bases to uracil (converted to thymidine, following PCR amplification) [119]. This modification is methylation sensitive and is such that methylated cytosines remain unconverted and distinguishable from unmethylated cytosines, which will be converted to thymidine [119]. Depending on the information required, multiple end-point assays can be used to determine LS methylation (Fig.1.9).

1.11.1.1 ***Methylation specific PCR (MSP)***

MSP [120] is a sensitive method that can be used to determine the presence of DNAm at a particular locus when the amplified sequence data is not required. In MSP, PCR primers that are specific to either the methylated (MSP_m) or unmethylated sequence (MSP_u) are used in separate PCR reactions. Methylation is quantified by gel electrophoresis (or quantitative PCR, in the case of MethylLight [121]) and is based on the ability of MSP primers to generate products (Fig. 1.9). This method allows the sensitive detection of fully methylated and unmethylated regions, but does not take partial methylation into consideration. Another drawback is the labour intensive process of optimal primer design.

1.11.1.2 ***Combined Bisulfite Restriction Analysis (COBRA)***

This method is similar to MSP; in that it is only suitable for when the underling sequence data is not required. However, in contrast to MSP, PCR primers are not methylation specific but are designed to promote equal amplification of both methylated and unmethylated DNA. COBRA [122] involves the sequence-specific restriction enzyme (RE) digestion of PCR-amplified BS-treated DNA. BS conversion will result in a change in DNA sequence, and RE DNA digestion will be prevented. Therefore, only sequences that are resistant to BS conversion, i.e. methylated DNA, will remain unconverted, and thus cleaved. Digestion products are subsequently investigated by gel electrophoresis, and DNAm status is determined by the presence and intensity of digested and/or undigested fragments (Fig. 1.9).

A key limitation of this method is its dependence on restriction enzyme recognition sites, which may not be present in all regions of interest.

1.11.1.3 ***Locus Specific Bisulfite Sequencing (LS-BS-seq)***

Bisulfite sequencing [119] is considered the gold standard for DNAm detection. It involves the sequencing of PCR amplified BS-converted DNA, which provides a single nucleotide resolution of DNAm. There are two main variations of LS-BS-seq, allele-specific BS-seq [119] and BS- Pyrosequencing (BS-Pyroseq) [123]. Both of these methods provide a quantitative measure of DNAm levels at single cytosines nucleotides. A key advantage of BS-Pyroseq, over allele-specific BS-seq, is that cloning of PCR products is not required to determine absolute methylation. However, a major disadvantage is that a maximum of only 50 bp can be sequenced in a single reaction. Both methods require the design of efficient primers that, specifically amplify the desired region, and does this at equal efficiency in both methylated and unmethylated sequences. Due to the redundant nature of converted DNA, primer design is probably the most challenging step in bisulfite conversion based methods.

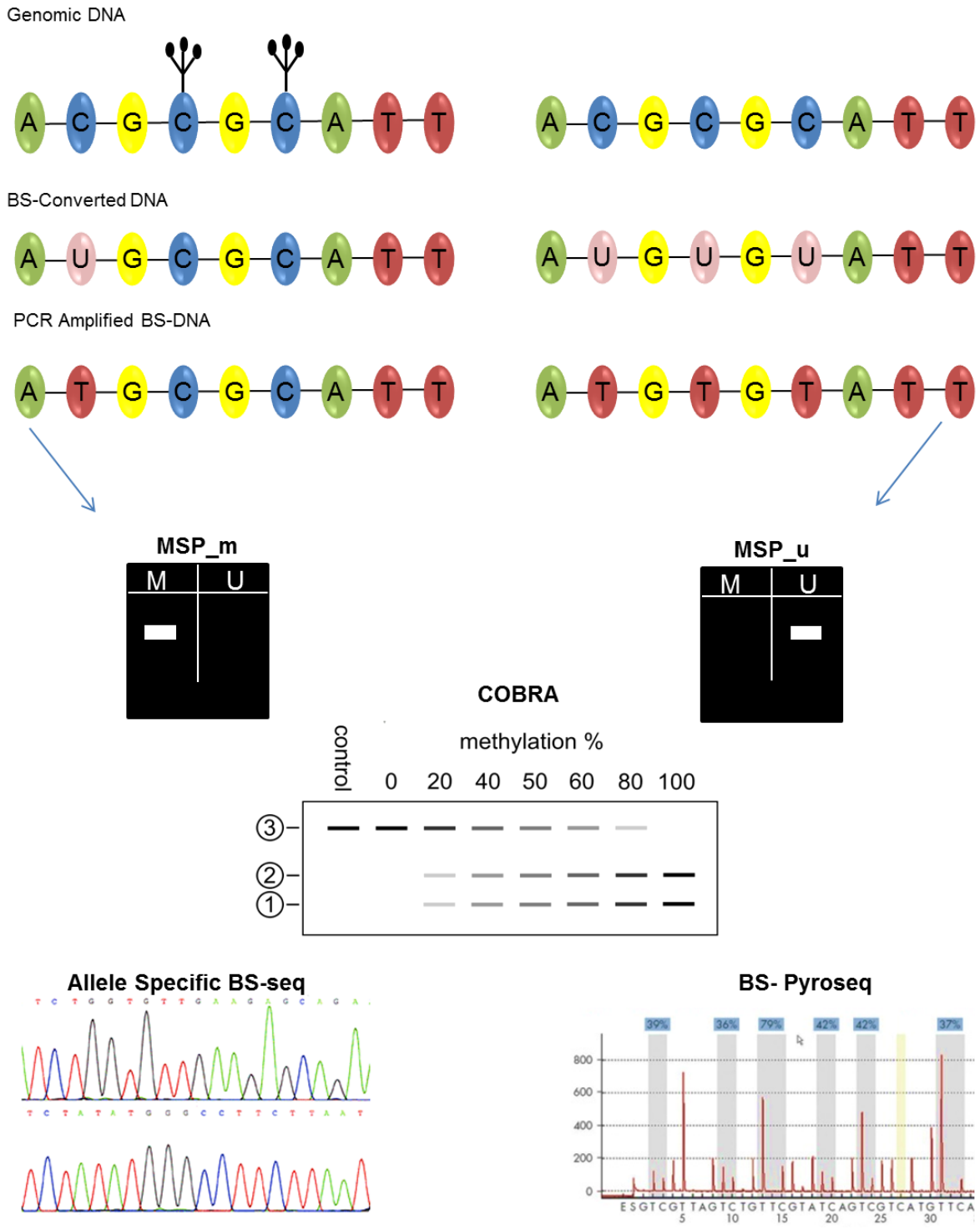


Figure 1-9: Bisulfite conversion (BS) of DNA and various quantification methods.
DNA on the left contains methylated cytosines (C with black stalks), which are resistant to BS conversion, while those on the right are unmethylated and thus BS converted. MSP; Methylated DNA will be amplified using MSP_m PCR primers, while unmethylated DNA will not. The reverse is the case, if MSP_u PCR primers are used. (MSP_m, methylation specific PCR for methylated DNA, MSP_u, methylation specific PCR for unmethylated DNA). COBRA; involves restriction enzyme (RE) digestion of DNA, following BS conversion and PCR amplification. Here, restriction enzyme cleavage will only occur in methylated and thus, unconverted sequences (lanes 1 and 2), while unmethylated sequences will remain uncut (lane 3). Row '3' depicts undigested products, while '1' and '2' depict digested fragments. Varying concentration of DNA will appear in either lane 3 (unmethylated) or lane 1 and 2 (methylated), depending on DNAm level. Allele Specific BS-seq; PCR amplified BS converted DNA is cloned and sequenced (Sanger sequencing). Absolute methylation level is then quantified by appropriate software, which assesses the resulting electropherogram for C/T mismatches. BS-Pyroseq; Absolute DNAm is determined directly, by pyrosequencing, following BS conversion and PCR, without the need for prior cloning.

1.11.2 **Global DNA Methylation Analysis**

DNAm analysis at this scale is commonly referred to as methylome analysis and will be described as such throughout this thesis. As mentioned earlier, common DNAm analyses are either bisulfite conversion based or enrichment based [118]. Commonly used global bisulfite conversion based methods are either quantified by next generation sequencing (NGS) as with whole-genome bisulfite sequencing (MethylC-seq or BS-seq) [62] and reduced representation bisulfite sequencing (RRBS) [124], or by hybridization to microarrays, the most frequently used being the Illumina Infinium Beadchip arrays (Currently 450K). Sequencing based methods are generally preferred to microarrays, as they do not require prior selection of regions of interest and are more informative with regards to repeat regions [118,125]. Additionally, MethylC-seq is more informative than RRBS, as the latter is biased for CpG rich regions [118] and is therefore not a true global analysis technique. Indeed, MethylC-seq is the gold standard method for methylome analysis as it provides a single-nucleotide resolution of the methylome; however, its application on large genomes remains prohibitively expensive. Other methods involve the enrichment of methylated portions of the genome, by antibody precipitation e.g. Methylated DNA immunoprecipitation (MeDIP) [126,127] or by precipitation with MBPs, as in the case of methyl-binding domain (MBD) assays [128,129]. These enrichment based methods are similar in terms of cost, technique, and principles [125], however, MBD assays are unable to detect non-CG methylation i.e. mC, as MBPs are only known to bind methylated DNA in the context of CG dinucleotides (mCG).

To carry out the necessary investigations for this thesis, an economic and sensitive method for mouse methylome analysis was required. Methylome analysis was favoured because several genes are thought to be involved in HSC ageing, and due to limited time and resources (HSCs are rare bone marrow cells), it was not feasible to investigate each of these individually. Likewise, due to limited funds, whole genome BS-seq was beyond the scope of this study. RRBS and MBD assays are biased towards CGIs and mCGs

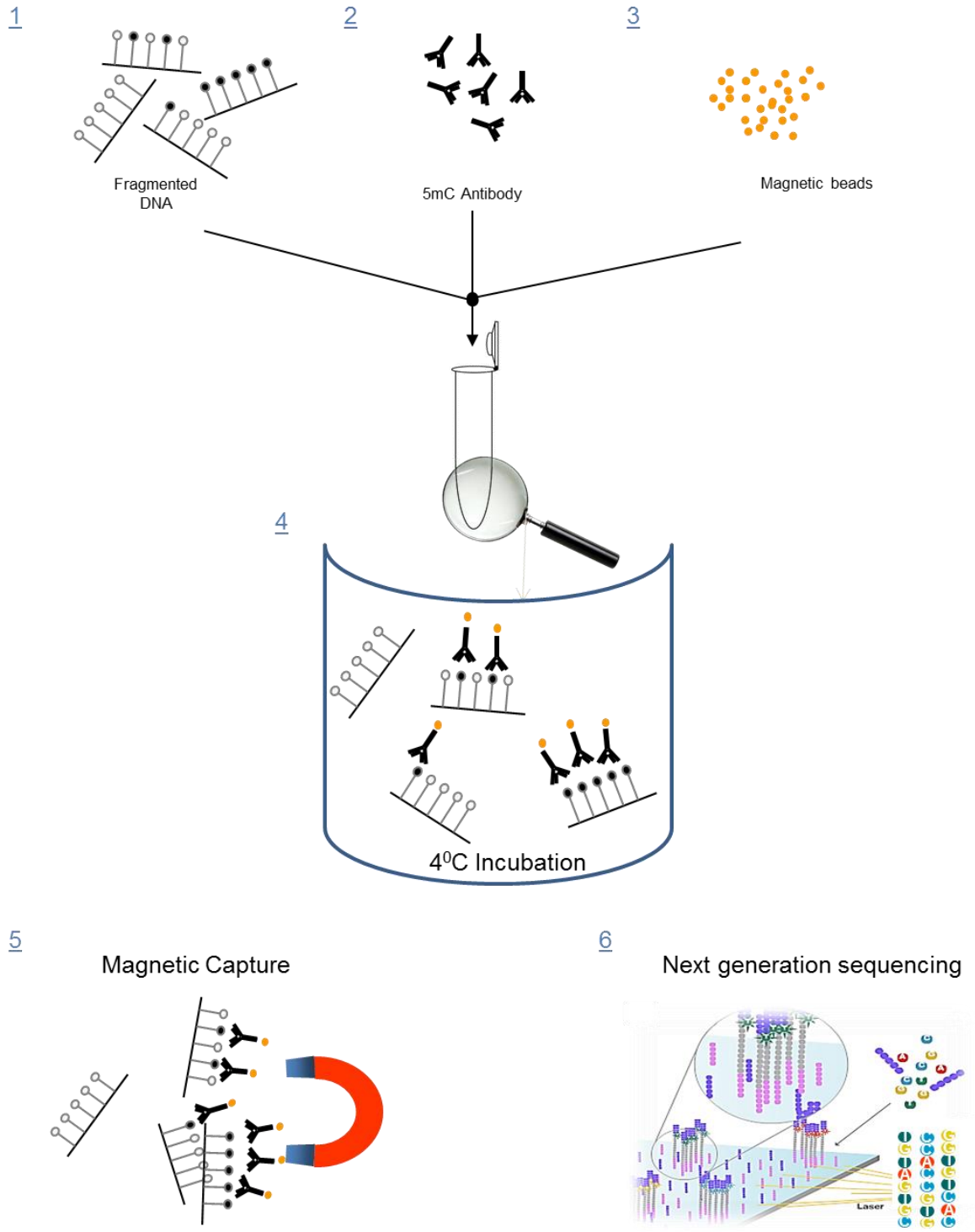
respectively [118] and there were no suitable hybridization arrays for mouse methylation studies. In aggregate, MeDIP-sequencing (MeDIP-seq) [127] was the most cost effective method for an unbiased analysis of the mouse methylome. I therefore used this method for methylome analysis of HSC ageing.

1.12 **MeDIP-Seq**

MeDIP-seq [126,127] involves the use of antibodies, directed against 5mC, to enrich for methylated regions of the genome, followed by next generation sequencing of enriched DNA (Fig. 1.10). 5mC antibodies recognize DNAm in CpG and non-CpG contexts, thus offering an unbiased analysis of the methylome. MeDIP-Seq provides a 100–250 bp resolution of the methylome and although this is low compared to BS-seq (1 bp), this resolution is sufficient for most applications, as DNAm of neighbouring CpGs show a high correlation for up to 1000 bp [130]. Therefore, MeDIP-Seq strikes a good balance between resolution, coverage and cost. Indeed, MeDIP-Seq was used to generate the first unbiased methylome of any mammalian genome [127] and has been shown to yield accurate and comparable results to the gold standard method (MethylC-seq) [118].

The original MeDIP-Seq protocol requires a starting genomic DNA concentration of ≥ 5 µg. However, HSCs are a rare cell population in mouse bone marrow (0.01–0.2%) and the maximum amount of DNA available for my investigation was ≤ 300 ng per sample. It was therefore necessary to improve the sensitivity of the MeDIP-seq assay. Additionally, a recently developed automated version of MeDIP-Seq, termed AutoMeDIP-Seq, was used in this thesis [131]. AutoMeDIP allows high throughput methylome analysis and is able to generate precise and accurate results [131], by minimizing human error during reactions.

Figure 1-10: Overview of the MeDIP-seq reaction.
DNA is fragmented (1) and incubated with monoclonal antibody that is specific to 5mC (2 and 4). Antibody-bound immunoprecipitated, methylated DNA (mDNA) is then incubated with magnetic beads (3 and 4), which bind to the 5mC antibodies and is used to purify mDNA by magnetic capture (5). mDNA is further purified, whole genome amplified (not shown) and subjected to next generation sequencing on Illumina sequencing platforms (6).



1.13 **Aims and Hypothesis**

This thesis aims to investigate the dynamics of DNA methylation during haematopoietic stem cell ageing. Mouse HSCs were used as a model to study mammalian HSC ageing because, they are easily isolated and obtainable and are the most characterised of all mammalian HSCs.

I hypothesise that changes in DNA methylation occur during HSC ageing and may underlie the functional changes that are observed during this process.

1.13.1 **Overview of Thesis**

Hoechst exclusion and side population analysis of the mouse bone marrow is currently the most cost effective method for the isolation of rare LT-HSCs [132]. I selected and used this method to obtain LT-HSCs and demonstrated the validity of its use in Chapter 3 of this thesis. As LT-HSCs are a rare cell population, it was necessary to improve the sensitivity of the MeDIP-seq method for use with limited starting material. I successfully optimized the MeDIP-seq method for a little as 50ng starting DNA concentration (Chapter 3). The optimized method is known as Nano-MeDIP-seq and this part of my thesis was published in a peer reviewed journal (Appendix) [133].

I used Nano-MeDIP-seq to generate and compare the methylome of mouse HSCs across three age time points; Young (8–12 weeks), Middle aged (~12 month), Old (22-24 months). Based on current evidence of the mechanisms by which DNAm alters cell and tissue dynamics during lineage specification, ageing and carcinogenesis (Section. 1.7 and 1.10); I hypothesised that a global loss of DNAm will accompany HSC ageing, and several genes that are involved in HSC regulation will be silenced by promoter hypermethylation. This hypothesis was partially supported, and I show in Chapter 4 that HSC ageing is indeed accompanied by global hypomethylation.

Previous studies have either investigated the ageing of heterogeneous HSPCs, or inadvertently introduced a sampling bias into their studies, by not selectively isolating the main population that change with age (My-bi HSCs). It is therefore likely that currently available gene expression data are not truly reflective of the changes that occur during HSC ageing. To address this, I purified a homogenous population of lower side population cells, which are the cells that specifically increase with age, and performed gene expression (transcriptome) analysis on these (Chapter 5). Analysis of my transcriptome data identified novel and key biological processes that elucidate the mechanism of HSC ageing.

I investigated the functional significance of aDMRs in HSC ageing, by integrated analysis of my methylome data and transcriptome data and functional validation of a candidate aDMR (Chapter 6). This analysis confirmed the role of DNAm in gene regulation and provided new insights into the dynamics of DNAm and gene expression during HSC ageing.

Finally, I discussed my main findings and their potential implications (Chapter 7) and concluded by identifying areas of my thesis that could be expanded upon, to further elucidate the mechanisms of HSC ageing.

Chapter 2

Materials and Methods

2.1 **Samples**

Wild type young (8-12 weeks), middle aged (~12 month) and old (22-24 months) C57BL/6 mice were sacrificed using an approved ('Schedule 1') method. A 5 month old C57BL/6 Mouse was used for the trial sequencing run. Mice were originally from the Jackson immunoresearch laboratories (Bar Harbour, ME) and some of the old and middle-aged mice were sourced from Harlan Laboratories (Holland). All the animals used in this thesis were bred as well as maintained in a specific pathogen-free animal facility.

2.2 **Nucleated Bone Marrow Cells Isolation**

Whole bone marrow cells were isolated from mouse femur, tibia and in some cases, hip bones, as follows;

1. Bone marrow was isolated from bones by flushing with RT 1x HBSS (Invitrogen) and 10mM Hepes solution (Sigma Aldrich) using a 27G syringe
2. Isolated bone marrow was chilled on ice for 5 min
3. Erythrocytes were lysed by adding 3 volumes of cold ammonium chloride (Stem Cell Technologies, Vancouver, Canada) to bone marrow suspension, and incubating on ice for 30 sec. 1 ml FCS was then added to marrow suspension to stop the lysis reaction
4. Cells were pelleted by centrifugation at 300g for 5–7 min at 4^0C
5. Cell supernatant was removed by aspiration and resuspended in cold 1 ml DMEM (Invitrogen) with 2% FCS (Invitrogen) and 10 mM Hepes solution; henceforth referred to as DMEM+
6. Debris were removed by filtering suspension through a 40 µM nylon cell strainer
7. 1–10 µl of isolated nucleated bone marrow cells (NBMCs) was stained with Trypan Blue solution and counted with a haemocytometer, under 10x magnification, using an inverted microscope (Leica).

2.3 **Fluorescent Assisted Cell Sorting (FACS)**

2.3.1 **Side Population Cells Phenotyping and Isolation**

1. Filtered NBMCs were pre-warmed to 37^0C in a closed, circulating water bath; henceforth referred to as water bath
2. Pre-warmed NBMCs were transferred to tissue culture flasks (vented cap, Corning) and resuspended at $1x10^6$ cells per ml of 37^0C DMEM+
3. Cell suspension was briefly incubated in a 37^0C water bath for 5 min
4. Hoechst 33342 (Sigma Aldrich) DNA binding dye; henceforth referred to as Hoechst, was added to cell suspension at 5 µg per ml and incubated for 90 min at 37^0C, with constant mixing at 30 min intervals
5. After 90 min, cells were transferred to 50 ml falcon tubes (BD Falcon) and pelleted by centrifugation, at 300g for 5–7 min at 4^0C
6. Supernatant was removed by aspiration and cell preparation varied depending on experimental end points, as described below.

2.3.1.1 ***SP Sort***

1. Cell pellets were resuspended in 1 ml cold DMEM+ and filtered using a 40 µM nylon cell strainer (BD plastics)
2. Filtered cells were resuspended in DMEM+ and 2 µg per ml Propidium iodide (PI, Sigma Aldrich) at approximately $1x10^7$ cells/ml
3. Cells were sorted on the Moflo cell sorter (Beckman Coulter)
4. Dead cells were excluded based on positive PI staining and cells with the least amount of Hoechst staining and an appropriate scatter profile, were isolated as lower SP (LSP) cells
5. LSP cells were sorted directly into 1.5 ml Eppendorfs containing either 1x PBS (Invitrogen) for pooled sorts, RLT cell lysis buffer (Qiagen) for single RNA extraction or Stemspan Serum-Free Expansion Media (SFEM, Stem Cell Technologies) for *ex-*

vivo culture. Cells sorted into PBS were pelleted and resuspended in up to 350 µl RLT cell lysis buffer (Qiagen).

2.3.1.2 ***SP Phenotyping***

Here, Hoechst stained NBMCs were analysed for LT-HSC phenotype by SP; KLS; CD150 staining. Antibody labelling was performed in a dark room, using the antibodies listed in Table 2.1 as follows;

1. Approximately $1x10^7$ cells were washed with cold 1x PBS, 2% FCS (Invitrogen) and 10mM Hepes solution, henceforth referred to as PBS+
2. Cells were pelleted by centrifugation, at 300g and at 4^0C for 5–7 min. Supernatant was removed by aspiration
3. Cell pellets were cooled on ice for 2 min and resuspended in up to $3x10^6$ cells/ 50 µl human gamma globulins (HgG)
4. A master mix of appropriate antibodies was prepared in the case of multiple samples, and transferred into the bottom of a 5 ml polypropylene snap top tubes (FACS tube) by pipetting
5. 1–$5x10^6$ cells-HgG suspension per sample was added to antibodies, and incubated in the dark at 4^0C for 30 min
6. Approximately $1x10^6$ cells-HgG suspension was incubated separately with individual antibodies as single staining (compensation) controls.
7. After 30 min, cells were washed with 20 volumes of 1x PBS+ and pelleted at 4^0C by centrifugation at 300g

   (For biotinylated antibodies; after 30 min of primary antibodies incubation, cells were washed and pelleted as described above and resuspended in streptavidin conjugated fluorochromes and PBS+. Cells were then incubated at 4^0C for 30 min, after which they were washed according to step 7).

8. Cell supernatant was removed by aspiration and cells were resuspended in 1 ml PBS+ and 2 µg per ml PI
9. Hoechst and antibodies stained cells were analysed on the LSRII (BD Biosciences).

Table 2-1: Phenotyping antibodies

Antibody (antimouse)	**Fluorochrome**	**Source**
CD150	Phycoerythrin (PE)	Biolegend
Sca-1	PE - Cyanine 7 (PE/Cy7)	Biolegend
CD 117 (c-Kit)	Allophycocyanine - Cy7 (APC/Cy7)	eBioscience
Lineage antibody cocktail	Biotin	BD Bioscience
Streptavidin	Percp	BD Bioscience
CD45.2	Fitc	Biolegend
CD 48	PE – Cy5	Biolegend

2.3.1.3 ***Post Sort SP Phenotying***

Approximately 1000–10000 freshly isolated or *ex vivo* cultured LSPs in a background of $1x10^5$ human KG1a cells, were analysed for LT-HSC enrichment by KS; CD150; CD48 antibody staining (Table. 2.1) staining, as described in Section 2.3.1.2. However, here cells were resuspended in 500 µl PBS+ and 1:1000 Dapi (Sigma Aldrich) and analysed on the Cyan ADP (Beckman Coulter). Mouse cells were differentiated from background human cells by positive staining for the anti-mouse CD45.2 antibody.

2.4 **Cell Culture**

2.4.1 **KG-1a Cell lines**

KG-1a, human acute myeloid leukaemia cell line, was cultured in 10% FCS RPMI cell culture medium and maintained at approximately $8x10^5$ cells/ml under 5% CO_2 in a 37^0C incubator;

2.4.2 **Ex vivo Culture of Primary HSCs**

1. SP cells were sorted into 1.5 ml falcon tubes containing 300 µl serum free enrichment media (SFEM)
2. Sorted cells were resuspended at 1000–2000 cells per 200 µl SFEM
3. 200 µl cell suspensions were transferred to a well in a round bottom 96 well plate (Corning) and cultured at 37^0C under 5% CO_2
4. Cells were cultured *ex vivo* for 4–5 days in the presence of 2 µg Heparin and STIFA cytokine cocktail (Table. 2.2).

Table 2-2 : *Ex vivo* cytokines (STIFA)

Cytokines (STIFA)	**Concentration/200 µl**	**Source**
recombinant mouse SCF	3 ng	Peprotech
recombinant mouse TPO	6 ng	Peprotech
recombinant mouse IGF-II	6 ng	R&D
recombinant human FGF-I	3 ng	R&D
recombinant mouse ANGPTL3	30 ng	R&D

*2.4.2.1 **Azacytidine Treatment of HSCs***

SP cells were cultured *ex vivo* as described above (Section 2.4.2). Cells were treated with two doses of 0–2 µM 5-Aza'deoxyCytidine (5-Aza'dC). The first dose was applied approximately 18 h after culture was initiated. Cells were maintained for a further 30 h after which 50% of media was replaced with fresh 100 µl SFEM, 1 µg Heparin and 33.3% STIFA cytokine cocktail. 50% of the initial 5-Aza'dC dose was added to cells 18 h after media replacement and cells were maintained in culture for a further 30 h. Treated SP cells were cultured *ex vivo* for a total of 4 days.

2.5 **Dual DNA/RNA Extraction**

For pooled sorts, DNA and RNA were simultaneously isolated from up to $8x10^4$ LSP cells.

1. Cells were pelleted by centrifugation at 300g for 5–7 min at 4^0C, and the supernatant was removed by gentle pipette aspiration
2. Cells were resuspended and lysed in 75 µl RLT plus lysis buffer (Qiagen) and 1% β-Me (Sigma Aldrich)
3. Up to four sets of 75 µl cell lysates were pooled for a total of $8x10^4$ cells
4. Cell lysate volume was adjusted to 350 µl with RLT plus lysis buffer and 1% β-Me
5. For single sorts, up to $1.5x10^4$ LSP cells were sorted directly into 350 µl RLT plus lysis buffer with 1% β-Me
6. DNA and RNA were extracted using the AllPrep DNA/RNA Micro Kit (Qiagen), according to manufacturer's instruction
7. Extracted DNA was eluted in approximately 85 µl or 35 µl EB buffer for pooled sorts and single sorts respectively, while RNA was eluted in 15 µl RNAse free water.

Cells in lysis buffer and extracted RNA were maintained at -80^0C until used. Extracted DNA was maintained at -20^0C for long term storage.

2.6 **Post Reaction DNA Purification**

DNA was purified from reactions using Agencourt Ampure XP (Beckman Coulter). Ampure XP is a Solid Phase Reversible Immobilisation (SPRI), paramagnetic-bead-based technology, which allows the highly efficient purification and recovery of DNA fragments from reactions. Ampure XP was used to purify DNA and cDNA from reactions according to manufacturer's instruction.

1. DNA or cDNA was mixed with 1.8x RT beads and incubated at RT for 5–15 min to allow beads to bind to DNA
2. Beads were separated from supernatant by magnetic capture

3. Immobilised beads were washed twice with 70% ethanol
4. Beads were left to air dry at room temperature for at least 5 min
5. Beads were resuspended with EB buffer (Qiagen) and left for 1 min at room temperature
6. Beads were separated from supernatant by magnetic capture
7. The supernatant, which contains purified DNA was collected and stored at -20^{0}C or used in downstream applications as appropriate.

In addition to using Ampure for reaction purification, I also adapted this protocol for the retention or exclusion of specific DNA fragment sizes, by varying the ratio of beads to DNA between 1.8x and 0.8x during purification (Section 3.4.2).

2.7 **Nucleic Acid Quantification**

2.7.1 **NanoDrop Spectrometer**

Purified high concentration (≥10 ng/ul) DNA and RNA samples were quantified by spectrometry, using the NanoDrop spectrometer (Thermo SCIENTIFIC). 1 µl of samples was typically loaded on the NanoDrop for quantification. The ratio of absorbance at 260nm and 280nm was used to asses sample quality. DNA samples with a ratio of 1.8 were generally accepted as good quality, while a ratio of 2.0 was accepted for RNA samples.

2.7.2 **Bioanalyzer Microfludics Platform**

The Bioanalyzer (Agilent) is a microfluidics based quantification assay. This was used to size and quantify samples of low concentrations and/or for which highly accurate size determination was required. The Bioanalyzer was also used to determine RNA Integrity Numbers (RIN) of RNA samples. 1 µl of samples were typically quantified using Bioanalyzer DNA High Sensitivity assay and chips or the Bioanalyzer RNA Pico assay and chips, all by Agilent and used according to manufacturer's instructions.

2.7.3 **Qubit Fluorometer**

Qubit (Life technologies) was used to quantify low concentration single and double stranded DNA and RNA samples, for which highly accurate and precise quantification was essential. 1 µl of samples were typically quantified using the Qubit DsDNA High Sensitivity assay, Qubit SsDNA assay or Qubit RNA assay, according to manufacturer's instructions.

2.8 **MeDIP-seq**

160–300 ng mouse genomic DNA was analysed for genome wide methylation, using an optimised version of the original MeDIP-seq protocol, termed Nano-MeDIP-seq (Section 3.4). The MeDIP-seq method included multiple steps, which are outlined below.

2.8.1 **DNA Fragmentation**

Genomic DNA was fragmented by sonication on the Bioruptor (Diagenode).

1. Sonicator was pre-chilled by filling water bath with crushed ice and leaving for 15 min
2. DNA was placed in a 1.5 ml tube and sonicated on 'high' for six, 15 min, cycles, each consisting of 30 sec on/off periods, to generate a fragment size range of 100–400 bp and fragment peak size of approximately 200 bp
3. DNA fragment size was validated on the Agilent Bioanalyzer 2100, using DNA high sensitivity chips and reagents, according to manufacturer's instructions.

2.8.2 **Library Preparation (Part 1)**

Library preparation was carried out using the NEBNext™ DNA Sample Prep Master Mix Set 1 Kit (New England BioLabs - NEB), according to manufacturer's instruction.

2.8.2.1 ***End Repair***

1. End Repair reaction was prepared on ice as follows:

Table 2-3: End repair mix

Component	Volume
NEBNext End Repair Enzyme mix	5 µl
NEBNext End Repair Reaction Buffer (10X)	10 µl
Fragmented DNA	85 µl
Total Volume	**100 µl**

2. The reaction mix was incubated at 20^0C for 30 min
3. End repaired DNA was purified using Ampure beads and eluted in 37 µl EB buffer (Qiagen).

*2.8.2.2 **dA- Tailing***

1. dA-Tailing reaction was prepared on ice as follows:

Table 2-4: dA-Tailing mix

Component	Volume
Klenow Fragment (3' → 5' exo)	3 µl
NEBNext dA-tailing Reaction Buffer (10X)	5 µl
Sterile H_2O	5ul
Purified End-Repaired, blunt-ended DNA	37 µl
Total Volume	**50 µl**

2. The reaction was incubated in a thermocycler at 37^0C for 1 h
3. dA tailed DNA was purified using Ampure beads and eluted in 25 µl EB buffer (Qiagen).

*2.8.2.3 **Adapter Ligation***

1. Adaptor ligation reaction was prepared on ice as follows:

Table 2-5: Adapter ligation mix

Component	Volume
Quick T4 DNA Ligase	5 µl
Quick Ligation Reaction Buffer (5x)	10 µl
DNA Adaptors (up to 5 µM)	10 µl
End Repaired, Blunt, dA- Tailed DNA	25 µl
Total Volume	**50 µl**

2. The reaction was incubated in a thermocycler at 18^0C for 2 h
3. Adaptor ligated DNA was purified using Ampure beads and eluted in 57 µl EB buffer.

2.8.3 **Methylated DNA Immunoprecipitation**

MeDIP was automated using the SX-8G IP-Star (Diagenode), according to manufacturer's instruction. Reagents (Diagenode) were prepared and dispensed in the appropriate wells of 0.2 ml 12-well strip tubes and loaded on the IP-Star as described below.

1. 150 µl 1x MagBuffer was prepared per sample by adding 120 µl of sterile H_2O to 30 µl 5x MagBuffer
2. A minimum of 16 µl of diluted antibody was prepared for up to six IP reactions, by adding 15 µl of sterile H_2O to 1 µl (2 µg) of 5-mC monoclonal mouse IgG1 antibody (Clone. b MeDIP-grade, Diagenode)
3. Antibody mix was prepared on ice as follows:

Table 2-6: Antibody mix

Antibody Mix	Volume/IP
Diluted antibody	2.40µl
MagBuffer A (5X)	0.60 µl
MagBuffer C	2.00 µl
Total Volume	**5.00 µl**

4. Incubation (IC) mix was prepared on ice as follows:

Table 2-7: Incubation mix

IC Mix	Volume/IP
MagBuffer A (5X)	24 µl
MagBuffer B	6 µl
Methylation control DNA	3 µl
Purified Adapter Ligated DNA	57 µl
Total Volume	**90 µl**

   a. IC mix was incubated at 100^0C for 10 min
   b. Samples were immediately placed on ice and incubated for 10 min.
   c. 75 µl of incubation mix was used in the immunoprecipitation (IP) reaction, while the remaining 15 µl was stored at -20^0C for later use as the negative IP control, henceforth referred to as Input DNA

5. IP mix was prepared on ice as follows:

Table 2-8: Immunoprecipitation mix

IP Mix	Volume/IP
Antibody Mix (step 3)	5 µl
1x MagBuffer A (step 1)	20 µl
IC Mix (step 4)	75 µl
Total Volume	**100 µl**

6. Reagents were dispensed in appropriate wells of a 12-well strip 0.2 ml tube, as listed in Table 2.9, and loaded onto the IP-Star

Table 2-9: Auto-MeDIP reagents

Well number	Reagent	Volume
1	DNA isolation Buffer (DIB)	92.5 µl
2	N/A	N/A

3	Magnetic (Dyna)beads (Dynabeads M-280 Sheep anti-mouse IgG)	10 µl
4	1x Magbuffer A	50 µl
5	1x Magbuffer A	50 µl
6	N/A	N/A
7	IP mix (See step 5 above)	100 µl
8	MagWash Buffer - 1	100 µl
9	MagWash Buffer - 1	100 µl
10	MagWash Buffer - 1	100 µl
11	MagWash Buffer - 2	100 µl
12	DNA isolation Buffer (DIB)	100 µl

7. IP was performed for 15 h and MeDIP DNA was eluted into DIB in well 12 of the 12-well strip tube
8. 7.5 µl of Input DNA was manually dispensed into DIB in well 1 of the 12-well strip tube
9. 1 µl of 20 mg/ml PCR grade proteinase K (Roche) was manually added to MeDIP and Input reactions, to digest antibody and release antibody-bound methylated DNA
10. The proteinase K reaction was incubated at 55^0C for 15 min, then to 100^0C for 15 min to stop the reaction
11. MeDIP DNA was separated from Dynabeads by magnetic capture
12. 9 µl of MeDIP and Input DNA was analysed by quantitative polymerase chain reaction (qPCR) to assess the recovery of spiked-in methylation control DNA and specificity of the MeDIP reaction (QC1; Section 2.8.3.1)
13. The remaining Input and MeDIP DNA were purified using Ampure beads and eluted in 15–30 µl EB buffer. Purified DNA was stored at -20^0C until used.

2.8.3.1 ***Quality Control (1) – QC1***

Here, the methylation control DNA that was included in the IC mix; a combination of methylated and unmethylated lambda fragments and henceforth referred to as Lambda spike, was analysed. Efficiency of the MeDIP reaction was measured by qPCR, using nested primers directed against either the methylated or the unmethylated lambda fragment.

Lambda spike and primers (Table. A-1) were obtained from Dr L.M Butcher and were constructed as previously described [131].

1. The qPCR assay was performed in triplicates for both MeDIP and Input DNA and prepared on ice as follows:

Table 2-10: qPCR reaction mix

Component	***Volume/Reaction**
10 µM Primer Pair	0.625 µl
Sterile H_2O	4.375 µl
2x SYBR qPCR master mix (Eurogentec)	6.25 µl
MeDIP or Input DNA	1.25 µl
Total Volume	**12.5 µl**

*A reaction master mix was prepared for a minimum of three reactions, as stated amounts, such as 0.625 µl, are difficult to pipette accurately.

2. qPCR was carried out on a real time PCR machine (Applied Biosystems) using the following program:

Table 2-11: qPCR program

STEP	**TEMP**	**TIME**	**CYCLES**
Activation	95^0C	5 min	1
Denature	95^0C	15 s	40 cycles
Anneal/Extend	60^0C	1 min	
Dissociation (melting curve)			

3. Recovery and specificity was calculated as follows:

    a. *Recovery* - As the quantity of Input DNA is only 10% of that of MeDIP, the Ct value obtained for Input was adjusted prior to calculating Recovery, using the formula;

$$\text{Adjusted Input Ct} = \text{Input Ct} - \log [2AE(10)]$$

Where AE is the amplification efficiency expressed as a number, e.g. 100% AE = 1 and 10 represents the dilution factor of Input to MeDIP (i.e. x10)

Recovery was then calculated as;

$$\text{Recovery (\%)} = 2AE^{(\text{Adjusted Input Ct} - \text{MeDIP Ct})} \times 100$$

b. *Specificity* was calculated as;

$$\text{Specificity} = 1 - \left\{ \frac{\text{Unmethylated DNA recovery}}{\text{Methylated DNA recovery}} \right\}$$

A MeDIP experiment was considered to be successful if specificity was ≥ 90% and recovery of unmethylated DNA was ≤ 1%.

2.8.4 **Library Preparation (Part 2)**

MeDIP reactions, which had passed QC1, were used in the second part of library preparation. This involves adapter-mediated PCR amplification of MeDIP and Input DNA. As the lambda DNA spike was not adapter ligated, only the sample DNA was amplified and sequenced here.

1. The PCR reaction was set up on ice as follows:

Table 2-12: Adapter-mediated PCR reaction mix

Component	Volume/reaction
PCR. PE. 1.0 (10 µM stock)	5 µl
PCR.PE.2.0 (10 µM stock)	5 µl
MeDIP or Input DNA	15 µl

Phusion™ or Kapa™ High-Fidelity PCR Master Mix, 2x	25 µl
Total Volume	**50 µl**

2. PCR reaction was incubated on a thermocycler, under the following conditions:

Table 2-13: Adapter-mediated PCR program

STEP	TEMP	TIME	CYCLES
Initial Denaturation	98^{0}C	30 sec	1
Denaturation	98^{0}C	10 sec	
Annealing	65^{0}C	30 sec	12-15
Extension	72^{0}C	30 sec	
Final Extension	72^{0}C	5 min	1
	4^{0}C	Hold	

3. PCR products were purified with Ampure beads and eluted in 18 µl EB buffer. Purified DNA was stored at -20^{0}C until used
4. 1.5 µl of purified DNA was quantified on the Nanodrop spectrophotometer
5. 0.5 µl of purified DNA was diluted up to 1 µl with sterile water and analysed on the bioanalyzer.

2.8.5 **Library Size Selection**

PCR amplified DNA was size selected to obtain uniform DNA sequencing libraries, with the required insert size. An insert size of 180–230bp was created for all MeDIP samples in this project.

1. 2% Tris Acetate EDTA (TAE, Thistle Scientific), Low melting point (LMP) agarose (VWR) gels were prepared as described below:
    i. 2 g of LMP agarose gel was dissolved in 100 ml 1xTAE (for small gels)

ii. Agarose mix was heated in a microwave oven at 800watts for 2 min
iii. Gel was cooled to ~30^{0}C
iv. Ethiduim bromide (EtBr) was added to agarose to obtain a final concentration of 0.5 µl/ml
v. Gel was transferred to a casting tray with 3 mm-wide gel combs and left to set on an even surface, at room temperature.

2. Sample DNA was mixed with 6x loading dye (Fermentas) and loaded onto wells within the agarose gel
3. Sample DNA was flanked by two wells containing 1 µg of 50bp DNA ladder (NEB). A gap of at least 2cm was retained between sample DNA and DNA ladder wells, or between multiple DNA samples
4. Gel electrophoresis was performed for 100 min at 100 volts, in TAE buffer containing EtBr (0.5µg/ml)
5. After electrophoresis, gel was visualised on a UV transilluminator
6. A strip of aluminium foil was placed underneath the gel to prevent cross linking of DNA
7. A library size of 300–350bp was excised with a clean scalpel. This provided an insert size of 180–230bp, which is the library size without the 120bp adapter/primer sequences
8. For contingency purposes, two 50bp size-range gel bands above and below the 300–350bp library were also excised
9. Samples were purified from agarose with the Minelute Gel DNA extraction kit (Qiagen), according to a modified version of manufacturer's instructions (Section 3.4.6)
10. Library DNA was eluted in 12 µl Elution buffer (Qiagen).
11. 1 µl of library DNA was quantified and size verified on the Bioanalyzer and DNA high sensitivity kit (Agilent).

12. 1 µl of library DNA was quantified with the Qubit fluorometer and Qubit high sensitivity dsDNA quantification kit (Invitrogen)
13. 2 ng of library DNA was diluted up to 9 µl with sterile H_2O and analysed by qPCR to validate the enrichment of known methylated genomic regions as detailed in Section 2.8.5.1

2.8.5.1 ***Quality Control (2) – QC2***

The aim of QC2 was to verify that the gel extracted library remained enriched for methylated DNA, and depleted for unmethylated fragments, following PCR. In contrast to QC1, QC2 involves the qPCR of endogenous DNA fragments, as the lambda DNA spike used in QC1 was not PCR amplified and therefore exists as negligible amounts in the library. Two mouse regions (Table. A-1) of known methylation status and of the similar CpG density (8%) were used as methylated and unmethylated controls. The methylation status of these regions was previously ascertained within the 'Medical Genomics' group. qPCR assays were performed in triplicates as described in Section 2.8.3.1. The comparative Ct method was used to calculate fold enrichment (FE) as follows:

$$\text{Fold enrichment ratio} = \frac{2AE^{(Ct_{input\_meth} - Ct_{MeDIP\_meth})}}{2AE^{(Ct_{input\_unmeth} - Ct_{MeDIP\_unmeth})}} \quad \textbf{\textit{Or}} \textit{ if no Input DNA} = 2AE^{(\mathbf{Ct}_{MeDIP\_unmeth} - \mathbf{Ct}_{MeDIP\_meth})}$$

Fold enrichment was occasionally calculated without the use of Input DNA. This is because the primers used for QC2 in this study were verified to have similar amplification efficiency (Appendix-Figure 1). Fold enrichment without Input DNA was calculated as shown above.

2.8.6 **MeDIP - Sequencing**

MeDIP libraries with a minimum fold enrichment of 100 for endogenous methylated regions were subjected to next generation sequencing.

Sequencing was carried out at a core facilities centre on the Genome Analyzer IIx or the Hi-SEQ platform (Illumina), by a specialist technician, according to manufacturer's instructions.

2.9 **Bisulfite Pyro-sequencing**

2.9.1 **Assay Design**

Forward, reverse and sequencing primers were designed for each pyro-sequencing assay, using the PyroMark assay design software v2.0 (Qiagen) as follows;

1. DNA sequence of target region was obtained from the UCSC genome browser (July 2007, NCBI Build 37), using the 'get DNA' function. Target DNA was flanked at either side with at least 50 bp extra, to allow flexibility during assay design
2. UCSC sequence was *in silico* bisulfite converted by the assay design software.
3. Multiple primer sets were designed for each assay
4. Top scoring primer sets were manually inspected and selection parameters were adjusted as required
5. PCR Primers were selected for further analysis, based on the following criteria:
    - Primer length of 20–32 nucleotides
    - High test CpG coverage
    - Amplicon length of less than 200 nucleotides
    - Melting temperature (Tm) of between 50^0C and 60^0C
    - Maximum Tm difference of 3^0C
    - No self-complementarity or secondary structures
    - %GC of 25–50
    - No alternate annealing sites on DNA template. When this was unavoidable, a maximum of two sites were permitted, provided annealing was very weak and outside of target DNA region.
6. Sequencing Primers were selected for further analysis, based on the following criteria
    - Primer length of 15–25
    - No alternate annealing sites on PCR amplicon

- No self-complementarity or secondary structures

2.9.2 **Assay Validation**

Test PCR reactions were performed for each PCR primer to confirm that the primer specifically amplified target DNA, and no primer dimers were generated during the PCR reaction. Primers not meeting these criteria were discarded or redesigned where possible. HPLC purified assay primers were obtained from Sigma Aldrich and one PCR primer, usually the reverse, was 5' biotinylated.

Bisulfite converted *in vitro* methylated, and whole genome amplified (WGA) DNA, were pyro-sequenced to validate the sensitivity of selected assays.

2.9.2.1 ***In vitro Methylation***

1. *In vitro* methylation reactions were set up on ice as follows

Table 2-14: *In vitro* methylation reaction mix

Component	**Volume/reaction**
10× NEBuffer 2	2 µl
S-adenosylmethionine (1.6 mM)	2 µl
SssI methyltranseferase (4 U/µl)	1 µl
Sterile H_2O	10 µl
DNA (3 µg)	5 µl
Total Volume/Reaction	**20 µl**

2. Reaction was incubated in a thermocycler and incubated for 2 h at 37^0C, followed by 20 min at 65°C
3. Reaction was purified using Ampure beads, and eluted in 60 µl EB buffer
4. *In vitro* methylation of DNA was verified by digestion with methylation sensitive restriction (HPAII, HpyCH4IV and BSTVI) enzymes. Whole genome amplified (WGA)

DNA was also digested in parallel, and used as the unmethylated control. Digestion reaction was prepared on ice as follows:

Table 2-15: Methylation-sensitive restriction enzyme digestion mix

Component	Volume/reaction
10× NEBuffer 1	4 µl
HPAII (10 U/µl)	1 µl
*HpyCH4*IV (10 U/µl)	1 µl
DNA (200 ng) + Sterile H_2O	Up to 34 µl
Total Volume	**40 µl**

5. Digestion reaction was incubated in a thermocycler for 1 h at 37^0C
6. 1 µl BSTVI (10 U/µl) restriction enzyme was added to the digestion incubation, at the end of step 5. (A separate reaction was required for this digestion as BSTVI is active at 60^0C, while HPAII and HpyCH4IV at 37^0C)
7. Reaction was incubated for 1 h at 60^0C
8. DNA fragment size was verified on 2% agarose gel (prepared as described in Section 2.8.5, but using standard agarose)
9. Gel electrophoresis was performed for 1 h at 120 volts

*2.9.2.2 **Whole genome Amplification***

1. 10 ng of DNA was fragmented and whole genome amplified, using the GenomePlex WGA 2 kit (Sigma Aldrich), according to manufacturer's instructions
2. WGA DNA was purified using Ampure beads, and eluted in 60 µl EB buffer
3. WGA reaction was validated by quantifying DNA, and by restriction enzyme digestion as described above (Section 2.9.2.1)

2.9.3 **Bisulfite Conversion**

1. 50 ng DNA or $1x10^4$ cells per sample were bisulfite converted, and purified using the EpiTect Plus (LyseAll) Bisulfite Kit (Qiagen), according to manufacturer's instructions
2. Conversion efficiency was determined by qPCR (Section 2.8.3.1)
3. Bisulfite converted DNA was maintained at -20^0C until used.

2.9.4 **PCR Amplification of Bisulfite Converted DNA**

Target regions were PCR amplified from bisulfite converted DNA, using the primers described in Section 2.9.1 and the PyroMark PCR kit, according to manufacturer's instructions, using the following cycling conditions:

Table 2-16: Bisulfite conversation PCR reaction

STEP	TEMP	TIME	CYCLES
Initial Denaturation	95^0C	15 sec	1
Denaturation	94^0C	30 sec	
Annealing	56^0C	30 sec	45
Extension	72^0C	30 sec	
Final Extension	72^0C	10 min	1
4^0C		Hold	

PCR products were verified by gel electrophoresis on standard agarose gels, as previously described (Section 2.9.2.1)

2.9.5 **Pyro-sequencing**

Pyro-sequencing of bisulfite converted PCR amplified DNA was performed on the PyroMark Q96 MD Pyro-sequencer, using PyroMark Gold Q96 reagents, according to manufacturer's instructions. Briefly;

1. Biotinylated PCR products were bound to Streptavidin beads

2. Bead-bound PCR products were washed with freshly prepared 70% ethanol
3. Bead-bound PCR products were denatured with 0.2 M NaOH
4. Bead-bound PCR products were washed with 1x Wash buffer (Qiagen)
5. PCR products were released into 96 well PSQ plates (Qiagen), containing the appropriate sequencing primer (0.3 µM) (Section. 2.9.1)
6. Sequencing primer was annealed to PCR products by heating the PSQ plate to 80^0C for 2 min
7. PCR products were cooled to room temperature and pyro-sequenced.

2.10 **RNA-seq**

RNA-sequencing libraries were prepared with the Tru-seq RNA sample preparation kit (Illumina), according to manufacturer's instructions. Enzymatic reactions were purified using 1.8–0.8x Ampure beads as previously described.

Briefly;

1. mRNA was isolated from 50–100 ng total RNA by oligo(dT) magnetic-bead enrichment
2. Purified mRNA was fragmented and converted to double stranded cDNA
3. cDNA was end repaired, A-tailed and adapter ligated
4. Adapter-ligated cDNA was whole genome amplified in a thermocycler, under the following cycling conditions:

Table 2-17: RNA-seq PCR reaction

STEP	TEMP	TIME	CYCLES
Initial Denaturation	98^0C	30 sec	1
Denaturation	98^0C	10 sec	
Annealing	65^0C	30 sec	12
Extension	72^0C	30 sec	

Final Extension	72^0C	5 min	1
	4^0C		Hold

5. 1 µl PCR amplified sequencing library was assessed on the Bioanalyzer, to verify fragment size
6. Sequencing libraries were quantified, using the Qubit fluorometer and dsDNA assays
7. cDNA libraries were subjected to next generation sequencing, at a core facilities centre, on the Genome Analyzer IIx or a Hi-SEQ platform (Illumina), by a specialist technician, according to manufacturer's instructions.

2.11 **Candidate Gene Expression Analysis**

The expression of single genes were analysed for quantitative reverse transcription-PCR (qRT-PCR), on a real time PCR machine (Applied Biosystems)

2.11.1 **cDNA Synthesis**

0.5–1x10^4 cells per sample were converted into single stranded cDNA, using the SuperScript III CellsDirect cDNA Synthesis kit (Invitrogen), according to manufacturer's instructions. Briefly;

1. Cells were resuspended in lysis buffer to lyse cells. 1 µl RNaseOUT (40 U/µl, Invitrogen) and 0.5 µl SUPERase.In (20 U/µl, Ambion) RNase inhibitors were included, per 10 µl, to prevent RNA degradation by RNases during cell lysis. (The addition of SUPERase.In is a deviation from the original protocol)
2. Cell lysate was treated with DNase I to digest any contaminating DNA
3. cDNA was synthesised from mRNA using oligo(dT) primers and SuperScript III reverse transcriptase
4. 2 units of RNase H were added per 30 µl reactions, to degrade any remaining template RNA.

2.11.2 **RT-qPCR**

The equivalent of 250–500 cells was used for each RT-qPCR reaction. TaqMan gene expression assays, which span exon boundaries, were used in this study. A stably expressed housekeeping gene, β-Actin, was used as reference control for all gene expression assays. Each reaction was performed in two to three technical replicates.

1. RT-qPCR reactions were prepared on ice as follows:

Table 2-18: RT-qPCR reaction

Component	**Volume/reaction**
20x TaqMan Assay	1 µl
2x TaqMan Universal Master Mix II	10 µl
cDNA	
Sterile H_2O	Up to 9 µl
Total Volume	**20 µl**

2. RT-qPCR reaction was performed on a real time PCR machine (Applied Biosystems) as follows:

Table 2-19: RT-qPCR program

STEP	**TEMP**	**TIME**	**CYCLES**
Activation	95^0C	10 min	1
Denature	95^0C	15 s	
Anneal/Extend	60^0C	1 min	50 cycles

2.11.3 **RT-qPCR Data Analysis**

The comparative C_T ($\Delta\Delta C_T$) method was used to calculate relative gene expression in this study as follows;

1. The Ct value for target gene was normalised to that of β-Actin (reference). This was calculated individually for both test and control samples as follows:

$$\Delta C_{T\,(test)} = C_{T(test\_target)} - C_{T(test\_ref)}$$

$$\Delta C_{T\,(control)} = C_{T(control\_target)} - C_{T(control\_ref)}$$

2. Normalised C_T values (ΔC_T) were compared to determine differences in gene expression as follows:

$$\Delta\Delta C_T = \Delta C_{T\,(test)} - \Delta C_{T\,(control)}$$

3. Fold change (FC) was calculated from $\Delta\Delta C_T$ as follows:

$$FC = 2^{-\Delta\Delta C_T}$$

2.12 **Data Analysis**

2.12.1 **MeDIP-seq Data Analysis**

MeDIP-seq data was analysed by Dr Gareth A. Wilson of UCL Cancer Institute, Medical Genomics group, using the MeDUSA [134] and MEDIPS [135] computational tools as follows:

1. Two separate FASTQ files, each corresponding to either the forward or the reverse DNA sequence reads were generated from raw MeDIP-seq data
2. FastQC v0.9.4 (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/) was used to perform quality control of the reads
3. BWA [136] v0.5.8 was used to align MeDIP-seq reads to the mouse reference genome (mouse genome build 37) using default settings
4. Aligned reads were filtered to retain only the reads that mapped as a true pair
5. MeDUSA v1.0.0 was used to discard read pairs in which neither read scored a BWA alignment score ≥ 10, and to discard all but one read from groups of non-unique

reads (i.e. reads that aligned to the exact same start and stop position on the same chromosome)

6. The MEDIPS package was used to perform MeDIP-seq specific quality control and to generate genome-wide methylation scores
7. Genome-wide correlation between and within sample sets were calculated using QCSeqs from the Useq package [137]. A 500 bp sliding (250 bp increments) window, consisting of at least 5 reads, was used to calculate sample correlations.
8. Regions containing a minimum of 10 reads (summed from a comparison set) were assessed for differentially methylated regions (DMRs), using the Bioconductor package DESeq v2.10 [138]
9. DMRs were placed in their relevant Ensembl (Build 62) annotated features, using custom perl scripts and bedtools [139] v2.10.2

2.12.2 **RNA-seq Data Analysis**

RNA-seq data was analysed by Dr Warren Emmett of UCL Genomics, using the RSEM [140] and DESeq [138] computational tools as described below:

1. BOWTIE [141] was used to align cDNA sequence reads to the mouse reference transcriptome library (mm9 genome build)
2. Aligned files were used to estimate transcript counts and transcript abundance
3. DESeq v2.10 was used to normalise count data according to library size, and to estimate the variance between observed data and the calculated mean
4. Differential expression of transcripts between samples was assessed by DESeq

2.12.3 **Gene Enrichment Analysis**

2.12.3.1 ***DMR Enrichment Analysis***

GREAT; Genomic Regions Enrichment of Annotations Tool, was used to predict the biological functions of DMRs, by looking for enriched annotations that were associated with the genomic regions.

1. The genomic coordinates of DMRs were used as input for GREAT v2.0.2 (http://great.stanford.edu/public/html/index.php)
2. The GREAT association rule was adjusted such that, each gene in the genome was assigned a proximal regulatory domain of at least 5 kb upstream, and 1 kb downstream of the TSS, and a distal regulatory domain of up to 100 kb in both directions
3. Each DMR was associated with all the regulatory domains (described in Step 2) it overlapped, as well as with their putative target genes. GREAT then performed enrichment analysis of functional annotations that were associated with DMRs.

2.12.3.2 ***Differentially Expressed Gene (DEG) Enrichment Analysis***

Ingenuity pathway analysis (IPA, https://analysis.ingenuity.com/pa/) was used to determine functional enrichment of differentially expressed genes;

1. Ensembl transcript ID of differentially expressed genes were input onto IPA
2. Annotated experimental and prediction based evidence from mammalian species, within the Ingenuity knowledge base, was used to determine functional associations within data sets.

Chapter 3

Methylome analysis of Long-Term Haematopoietic Stem Cells

3.1 **Introduction**

One aim of this thesis was to analyse the methylome of phenotypically defined long-term haematopoietic stem cells (LT-HSCs) with age. To do this, I isolated LT-HSCs using the side population (SP) method [16]. The SP compartment contains all the primitive cells of the mouse bone marrow (BM) [16] and these cells can be isolated based on their unique ability to effectively efflux the Hoechst DNA binding dye. SP is a cost effective method for isolating LT-HSCs as this method does not require the use of additional markers [13]. Commonly used markers for defining HSCs are summarised in Table 3.1

Table 3-1: Common HSC phenotyping markers.

Phenotype	HSC Property	Pros	Cons
c-Kit^{+}, Sca-1^{+}, Lin $^{-/low}$ (KLS)	LT-HSCs; Short-Term HSCs (ST-HSCs) and Multi-potent progenitors (MPPs) [142]	Well established and requires the use of fewer markers. Contains a higher frequency of primitive cells	Heterogeneous mix of cells
Lower Side Population (LSP)	LT-HSCs [13]	Cheap and identifies HSCs based on functional (ABC/G2) activity	Requires careful expertise as under-stained non HSCs will also appear as SP cells (False positives).
KLS; CD48^{-}; CD150^{+}	LT-HSC [14]	Identifies a homogeneous population of LT-HSCs. LT-HSCs can be sub fractionated based on CD150 expression	Costly, requires many markers and extensive set-up. Prone to false negatives (e.g. under-stained cells). LT-HSCs are heterogeneous for CD150 expression [15]
KLS; CD34^{-}; Flk2^{-}	LT-HSC [12,143]	Identifies a homogeneous population of LT-HSCs	Costly, requires many markers and extensive set-up. Prone to false negatives (e.g. under-stained cells).

LT-HSCs are a rare population of cells, accounting for 0.01–0.2% of the mouse bone marrow [7]; as a result, a methylome analysis method with low DNA input requirements was needed to complete this study. As discussed in Chapter one, MeDIP-seq was the most

suitable method for methylome analysis, as it offers a relatively unbiased view of the methylome. However, at the time of this study, MeDIP-seq required at least 5 µg starting DNA concentrations. It was therefore necessary to optimise the sensitivity of this method for the analysis of rare samples such as LT-HSCs.

In this Chapter, I first describe the successful isolation of C57BL/6 (b6) mouse LT-HSCs at three age time points; Young, 8–12 weeks; Middle-aged (Mid), ~12 Months; Old, 22–24 Months. I then outline the development of a more sensitive version of the MeDIP-seq method, termed Nano-MeDIP-seq, and demonstrate its application for the methylome analysis of rare LT-HSCs, using 160–300 ng starting DNA.

3.2 **Study Design**

Prior to starting the full-scale ageing project, I performed experiments to estimate the number of animals that would be required to complete this study, as well as the overall feasibility. To this end, I sorted and isolated DNA from bone marrow (BM) lymphocytes (BML) that were obtained from one adult C57BL/6 mouse. BMLs were sorted based on their cell profile (Fig. 3.1), which is similar to that of LT-HSCs, as determined by their forward scatter (FSC, correlates with cell volume/size) and side scatter (SSC, correlates with cell complexity e.g. granularity).

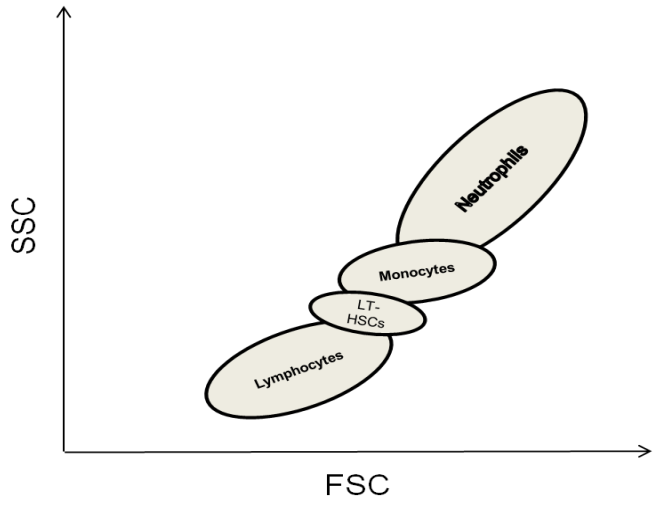


Figure 3-1: Diagrammatic representation of the scatter profile of nucleated bone marrow cells.

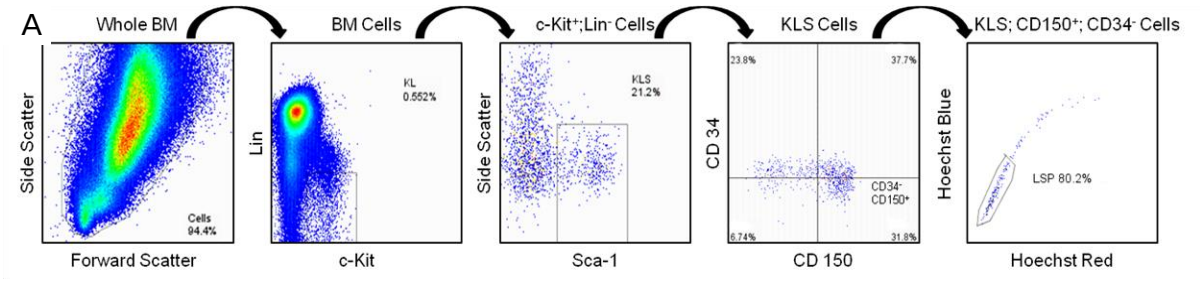
I extracted approximately 400 ng of DNA from $1x10^5$ FACS purified BMLs. Based on this, I estimated that ~$7.5x10^4$ FACS purified (B6 mouse) LT-HSCs would yield ~300 ng of DNA. Considering the frequency of LT-HSCs in the B6 mouse bone marrow (0.01–0.2% of 3–$5x10^7$), 15–30 young B6 mice were estimated to provide 300 ng of LT-HSC DNA. This was expected to be approximately 5–10 times lower for middle-aged and old B6 mice, based on the known accumulation of HSCs with age [95,144].

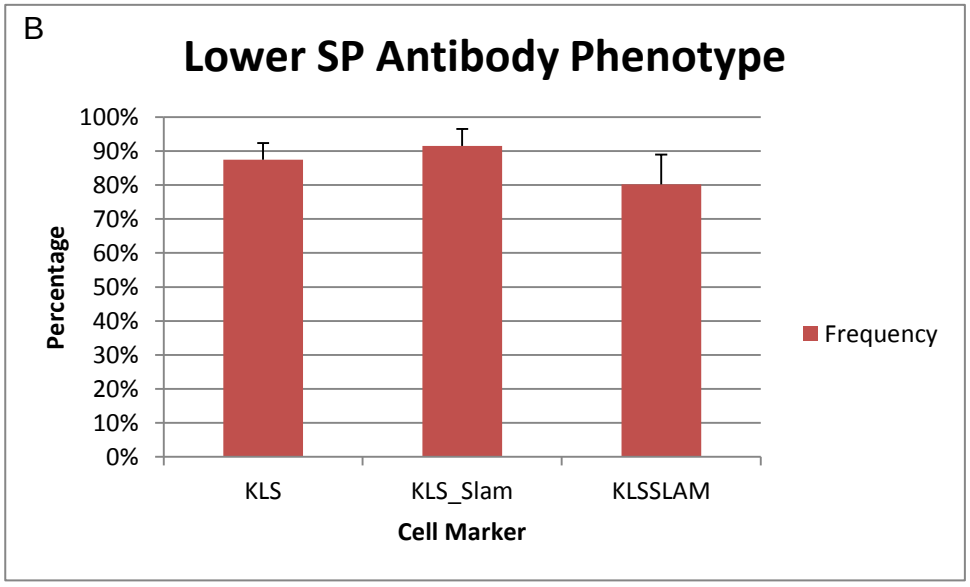
3.3 **Isolation and Validation of Ageing LT-HSCs**

The SP method used to isolate LT-HSCs in this study was first described by Margaret Goodell, in 1996 [16], for the isolation of HSC from mouse bone marrow. This method has since been widely used to purify stem cells from various other tissues as well as from other organisms [145,146]. I isolated the fraction of SP cells with the least amount of Hoechst staining (lower SP cells (LSPs)), as these cells have previously been shown to sustain long term haematopoiesis [13] and can therefore be referred to as LT-HSCs. Additionally, the LSP fraction of SP cells is the primary HSC population that undergoes age-related changes [7,19]. It was therefore appropriate that this population was selectively studied for DNA methylation and gene expression changes during ageing.

Before deciding to sort the LT-HSCs solely on their SP phenotype, I carried out a number of experiments to validate my staining and sorting techniques. Firstly, I investigated the SP phenotype of KLS; $CD150^+$; $CD34^-$ defined LT-HSCs (Table. 3.1) and found these cells to be enriched (80.2%) within the LSP region (Fig. 3.2a). I also confirmed the LT-HSC phenotype of LSPs (Fig. 3.2b) and found LSPs to be enriched for phenotypically defined LT-HSCs. This validated the use of the Hoechst exclusion (without the use of additional markers) for the identification and isolation of LT-HSCs in this thesis.

Figure 3-2: Side population phenotype of KLS, CD34$^+$ and CD150$^+$ defined HSCs.





A) FACS dot plot of total BM cells, which were stained with fluorescent antibodies to LT-HSC antigens as indicated on the x and y axis. Antibody stained cells were analysed on the LSRII flow cytometer. Each plot shows the phenotype of cells gated on the previous (left) plot, as indicated by the curly arrow. The final plot shows the Hoechst profile of antibody defined LT-HSCs. Data shown is from 24 months old mice. B) The lower SP (LSP) fraction of Hoechst stained bone marrow cells were analysed for LT-HSC cell surface markers. KLS antibodies are as previously described (measures HSPCs), KLS_Slam shows the frequency of KLS cells that are CD150+, while KLS_Slam shows the overall frequency of cells that are KLS and CD150+. Plot shows that LSP is enriched for phenotypically defined LT-HSCs (KLSSLAM). Data shown is from 24 months old mice. Two biological replicates are shown. Error bar is Standard Deviation.

As a result of this observation, I isolated phenotypically defined LT-HSCs by Hoechst staining alone throughout this thesis. However, to maintain stringency during the cell sorting, I also gated LSP cells on their FCS – SSC profile. Figure 3.3 shows a typical LSP cells/LT-HSC sort schematic. The phenotype shown in this figure is representative of the LSP cells used in this study, and demonstrates that these cells were a homogeneous population and not a mix of under-stained bone marrow cells.

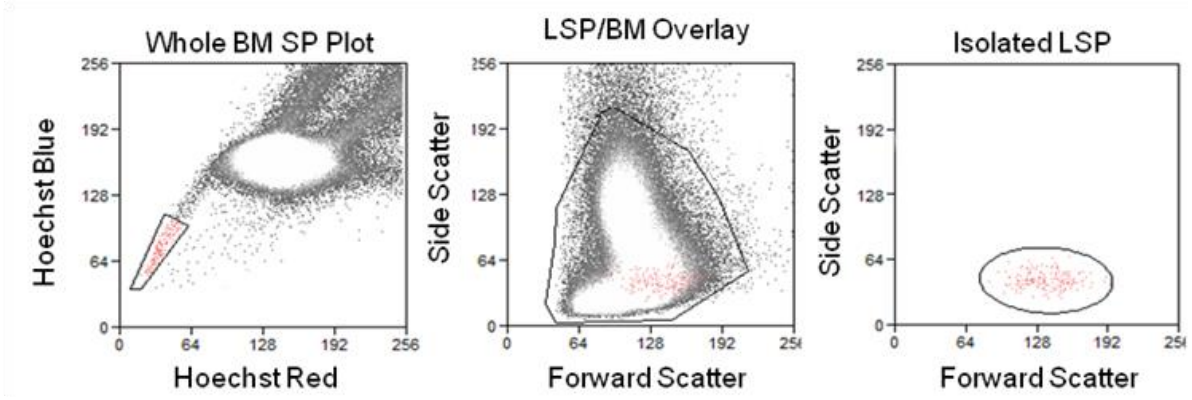


Figure 3-3: LSP sort.
FACS dot plot of a typical LSP sort, Hoechst stained whole bone marrow cells are shown on the left. Only the LSP cells that fall within the oval gate shown above (far right) were collected in this study

As a final validation of the sorting techniques in this study, I analysed the purity of sorted LSP cells, by investigating the expression of LT-HSC markers on these cells (Fig. 3.4). I confirmed that the LSP cells isolated in this study were Sca-1$^+$ and c-Kit$^+$. I also confirmed that the sorted LSP cells were enriched for CD150 positive cells and negative for the expression of CD48 as previously described [14].

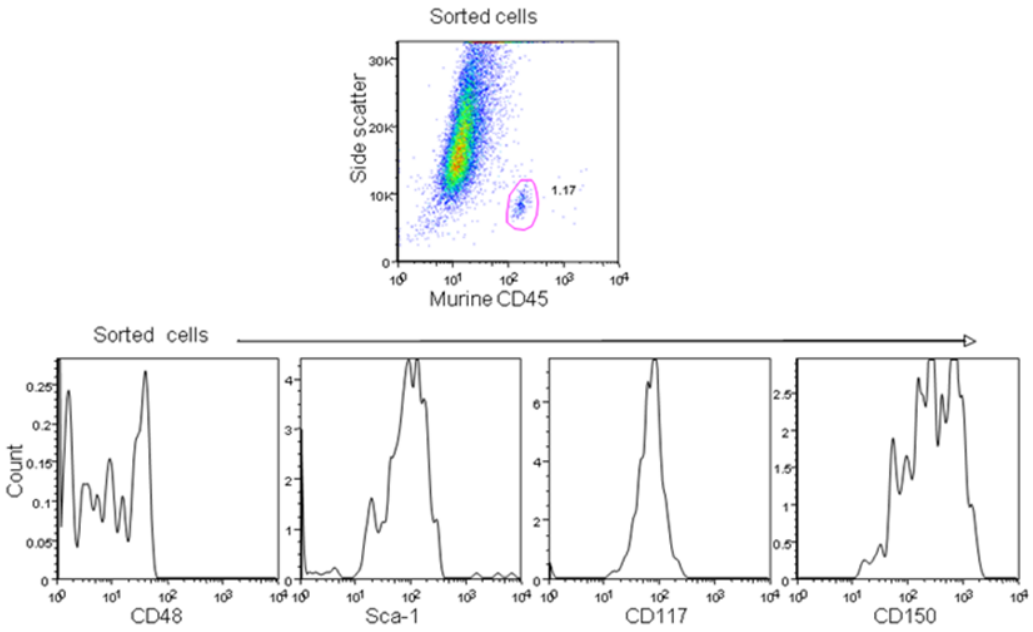


Figure 3-4: Antibody phenotyping of sorted lower side population cells.
FACS plot of sorted LSP cells, ~ 1x10^5 human KG-1a cells were used as carrier for 1–2x10^3 LSP cells, which were stained with fluorescent antibodies as indicated on the x axis. Mouse CD45.2 antibodies were used to distinguish mouse cells from human carrier cells (top plot)

Having validated the techniques used to identify LT-HSCs; I isolated phenotypically defined LT-HSCs from young, middle-age and old C57BL/6 mice and confirmed previous findings that these cells increase with age (Fig. 3.5).

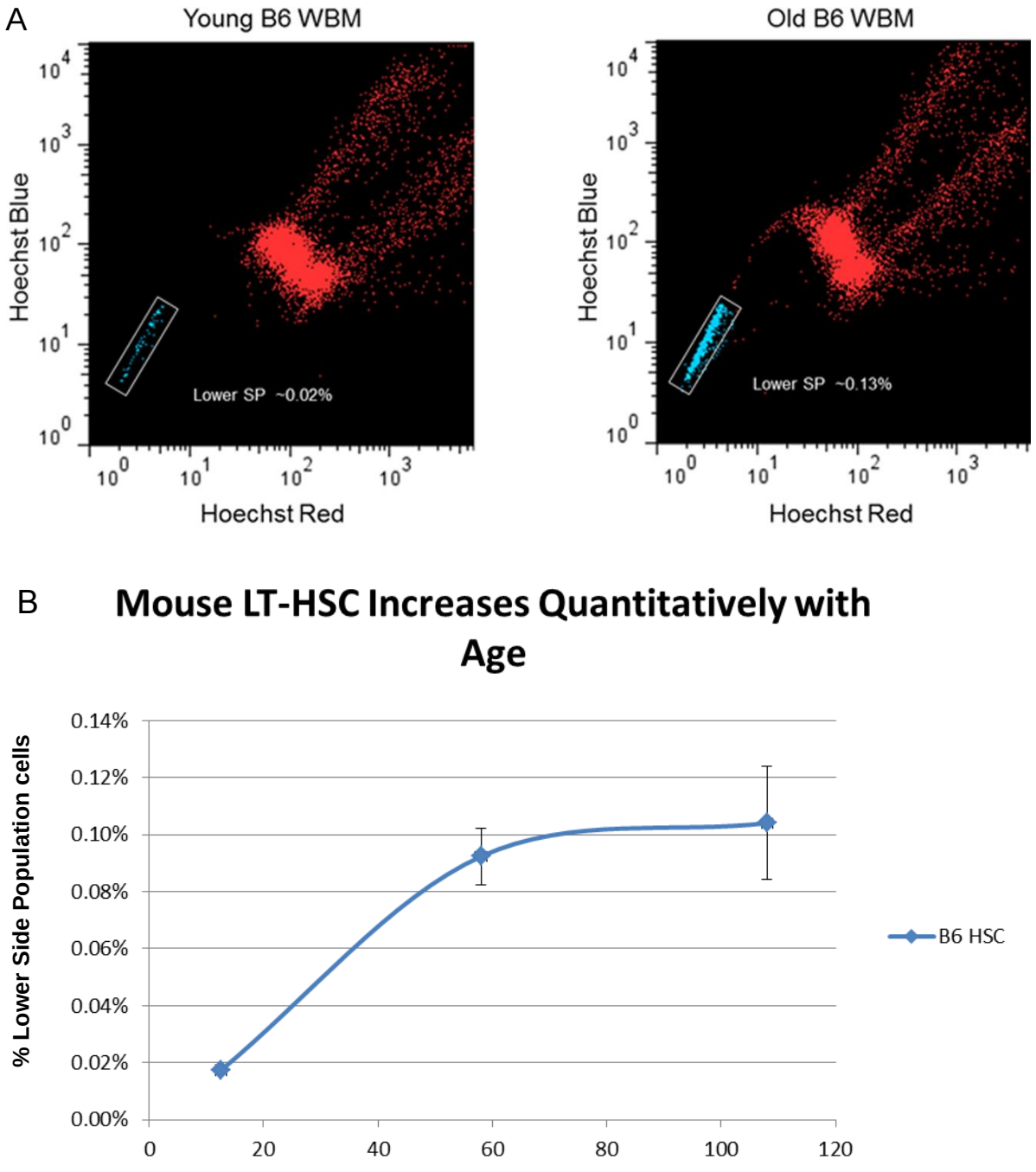


Figure 3-5: Age-related increase in side population cells.
Figure shows (A) a representative FACS dot profile of Hoechst stained bone marrow cells of young and old B6 mice as indicated. Lower SP cells are depicted as blue dots, and show an increase from young to old mice. Hoechst fluorescence, Red and Blue are shown the x and y axis respectively (B) The frequency of FACS sorted LSP cells from female b6 mice, at three independent time points. Error bar denotes standard error of the mean. 17, 11 and 10 biological replicates were used for Young, Mid and Old points respectively. Each biological replicate consists of 5 – 1 pools of mice. A pool of 5 mice was typically sorted for Young, while 3 – 1 were sorted for Mid and Old.

3.4 **Development of Nano-MeDIP-seq**

I optimised MeDIP-seq for low DNA concentrations to allow the analysis of LT-HSC methylomes. As the availability of LT-HSCs was limited in this study, I used DNA from FACS purified BML for the development of the Nano-MeDIP-seq protocol described below.

3.4.1 **DNA Fragmentation**

The methylation status of individual CpG sites is highly correlated over distances up to 1 kb across the genome [11]. Therefore, MeDIP libraries with an average sequence insert size of 200 bp were prepared (which represents a good balance between costs and resolution) and used throughout this study to analyse LT-HSC methylomes at 200 bp resolution. Randomly and evenly fragmented DNA is required as input into the MeDIP-seq reaction and DNA was fragmented by sonication in this study. I identified this step as a source of significant DNA loss, which was due to suboptimal sonication and fragmentation (Fig. 3.6a). I optimised the DNA sonication protocol such that majority of the fragmented DNA fell within the 200 bp library size range (Fig. 3.6b). The key optimisations were: 1) Longer sonication time of 90 min; 2) temperature was maintained below 10^0C throughout sonication. Additionally, I sonicated DNA in LoBind 1.5 ml tubes (Eppendorf), to minimize DNA loss due to non-specific DNA adhesion to tube walls during sonication.

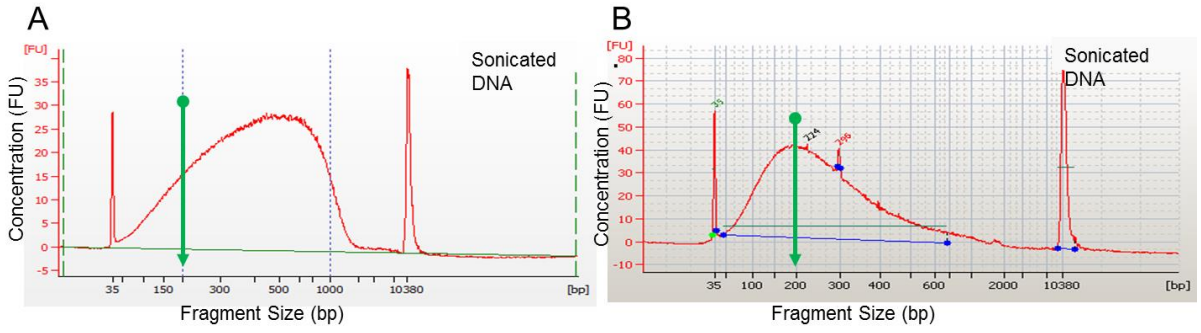
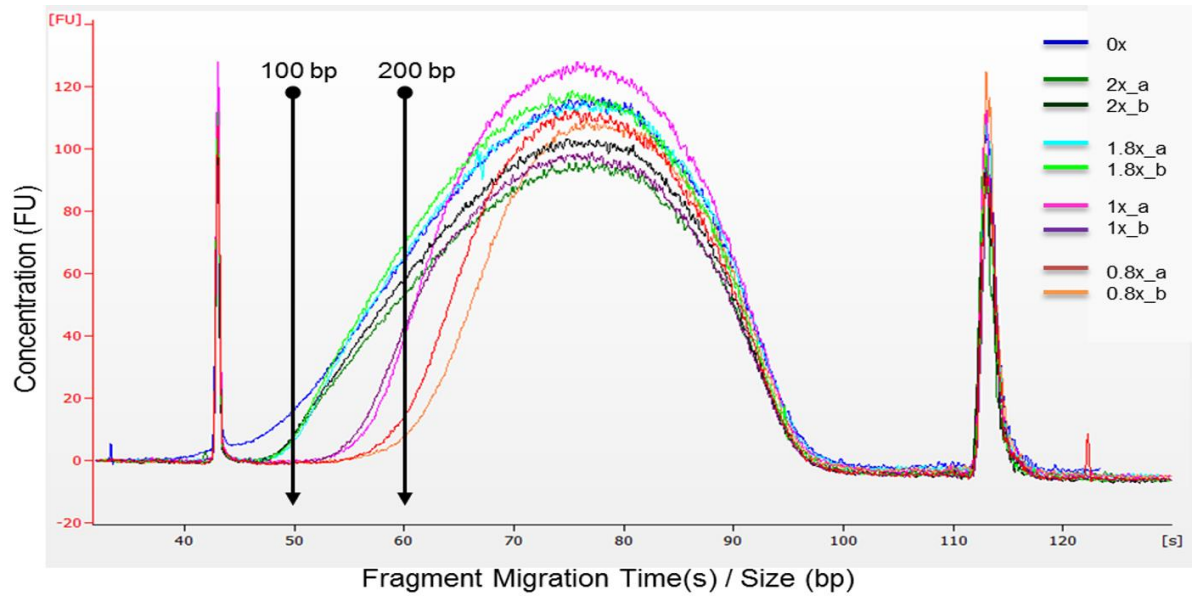


Figure 3-6: Electropherograph of sonicated DNA.
A) A standard sonication time of 30 min generates fragments that are mostly over the required 200bp mark (green arrow). A huge proportion of the DNA is therefore lost following library size selection. b) Optimised sonication, with a fragment peak size of 200bp. 35bp and 10380bp peaks represent the lower and upper DNA ladder peaks respectively. Electropherograms were generated as part of DNA analysis on the Bioanalyzer Microfluidics Platform. The graph is a plot of the DNA fragment size (bp) against fluorescence units (FU). DNA concentration is proportional to FU.

3.4.2 **Reaction Purification**

There are at least five purification steps in the MeDIP-seq protocol. I estimated a minimum of 30% loss of DNA, following each standard (affinity column based) purification step, making reaction purifications a main source of DNA loss. This estimate was derived from pre and post- affinity column –purification, quantification of (previously) purified genomic DNA, performed during the course of my project. As a result, I replaced all affinity column based purification steps with Solid Phase Reversible Immobilisation (SPRI) bead purification [147] and observed a marked improvement in DNA recovery. These paramagnetic beads (Trade name: Ampure) are able to bind reversibly to DNA in the presence of polyethylene glycol (PEG) [147]. I recovered at least 80% of high quality DNA following standard Ampure clean–up steps. Additionally, Ampure was able to exclude contaminating oligonucleotides such as excess primer and adapter dimers, from clean–up reactions. The DNA fragment size range recovered after Ampure purification is dependent on the PEG concentration in the Ampure beads buffer [148]. I adjusted PEG concentrations during DNA purification, by varying the ratio of Ampure mix to reaction volume i.e. between the standard 1.8x to 0.8x (Fig. 3.7). I used lower PEG concentrations (< 1x Ampure/reaction ratio) for the purification of DNA from PCR and adapter ligation reactions. This selectively excluded interfering excess and dimerised oligonucleotides (oligos), which are typically less than 150 bp, from downstream reactions.

Figure 3-7: Ampure purification.



DNA Electropherogram shows the results of Ampure purification of 200 ng DNA, at the indicated ratios (legend; top right). Fragment sizes of 100bp and 200bp are marked by black arrow. DNA was purified in two replicates at each test ratio and these were compared to unpurified control (0x). The 35bp and 10380bp lower and upper DNA ladder peaks are shown on the far left and far right (tallest peak) respectively. Electropherograms were generated as part of DNA analysis on the Bioanalyzer Microfluidics Platform. The graph is a plot of the migration time (s) against fluorescence units (FU), which is a measure of size and concentration respectively.

3.4.3 **Library Preparation (Part 1)**

This part of the MeDIP-seq protocol involves DNA end repair, 3'-dA tailing and adapter ligation. DNA end repair is necessary to repair the sheared ends of sonicated DNA, while A-tailing provides the 3'dA overhang that is required for adapter ligation by complementary base pairing to adapter 5' thymidine. DNA-Adapter ligation provides a universal sequence that allows whole genome amplification from a single primer pair.

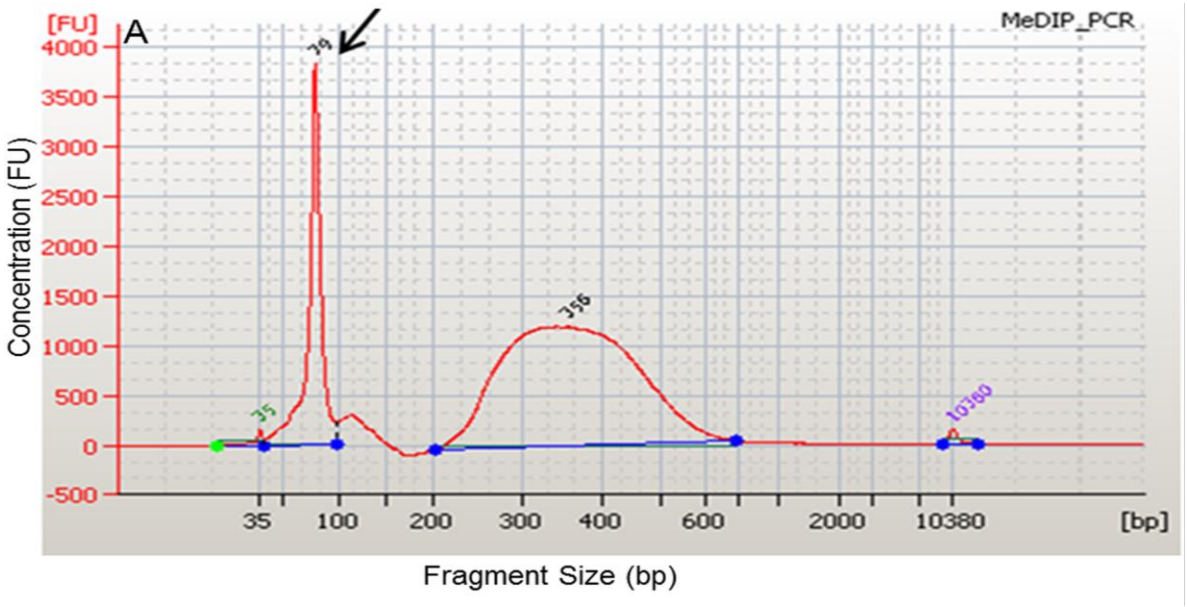
A key limitation in this part of the protocol was the inability to accurately determine the optimal adapter concentration that was required to ligate adapters to all DNA fragments in the reaction. This resulted in either unsuccessful ligation or a build-up of excess adapter dimers which inhibit subsequent PCR amplification and sequencing reactions. Generic library preparation protocols recommend a 10-fold molar excess of adapter sequences to DNA fragments ends. However, this recommendation is based on the starting DNA

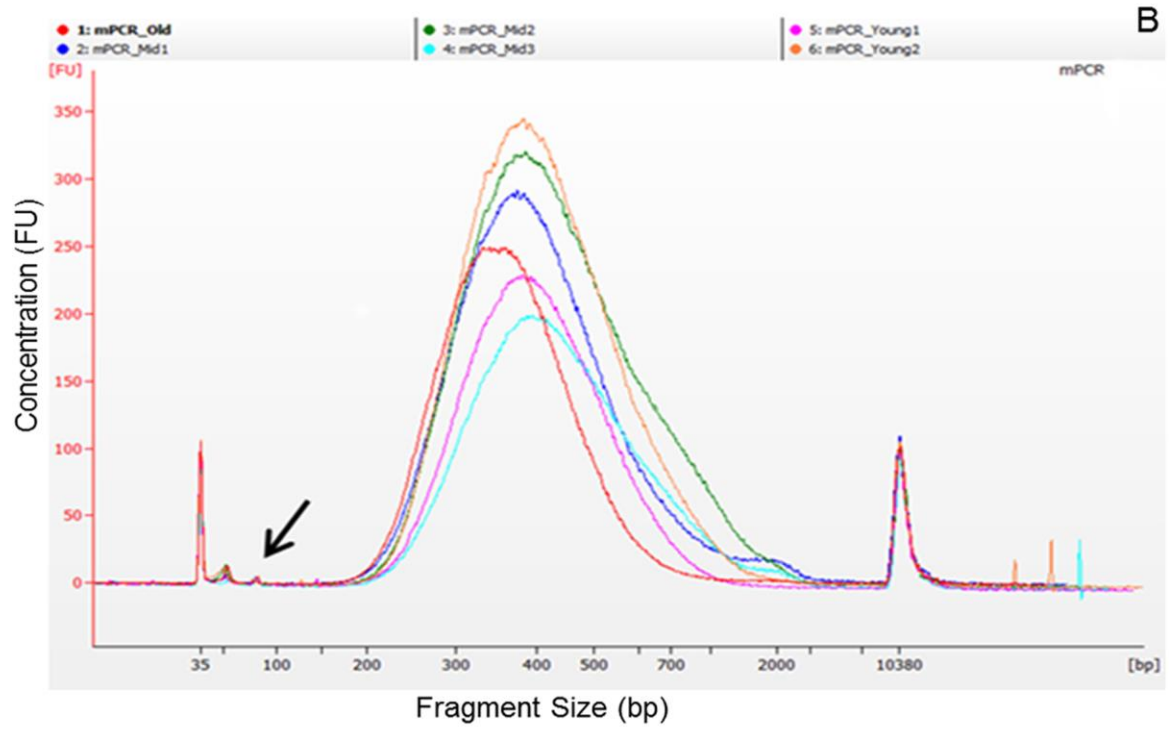
concentration and does not take into account the DNA loss in steps prior to adapter ligation, or the size into which the genomic DNA was fragmented, resulting in different molarities of ends. To circumvent this problem, I determined DNA molarity, using Bioanalyzer DNA high sensitivity reagents, following dA-tailing and prior to adapter ligation. This was used to calculate the optimal concentration of adapters that was required for the ligation reaction. Additionally, I purified adapter ligated DNA using low PEG concentration-Ampure clean-up. This helped to minimize the presence of adapter dimers in purified DNA (Fig. 3.8).

3.4.4 **MeDIP**

I performed MeDIP reactions as described in Section 2 but with minor modifications. For example, I adjusted reagent set-up to avoid the use of small volumes where possible. The Ab mix detailed in Step 3 of Section 2.8.3 originally required the addition of 0.30 µl diluted antibodies to the mix. 0.30 µl is an incredibly small volume and is in fact near invisible to the naked eye. As this could easily result in failure to add (sufficient) antibodies to the MeDIP reaction, I adjusted the concentration of stock antibodies, such that a more manageable volume was used, whilst maintaining the same reaction concentration.

Figure 3-8: Ampure purification efficiently removes unwanted oligos from sample DNA.





A) PCR Amplified MeDIP DNA was purified using affinity columns (Qiagen). The purified sample remained heavily contaminated with dimerised oligos (arrowed). B) Purification with Ampure beads consistently reduced the presence of contaminating oligos (arrowed). 35bp and 10380bp peaks represent the lower and upper DNA ladder peaks respectively. Electropherograms were generated as part of DNA analysis on the Bioanalyzer Microfluidics Platform. The graphs are plots of DNA fragment size (bp) against fluorescence units (FU).

3.4.5 **Library Preparation (Part 2)**

I typically obtained 2–10 ng purified immunoprecipitated (IP) DNA from MeDIP reactions, with ~7.5x10^4 cells. IP DNA is amplified in this part of the MeDIP-seq protocol, to increase DNA concentration for use in subsequent step. There is a caveat to PCR amplification; too few PCR cycles results in insufficient DNA for downstream use, while too many cycles lead to clonal amplification of small fragments present in the reaction. In order to determine the optimal conditions for amplifying MeDIP DNA (mDNA), I investigated the use of alternative, next generation DNA polymerases, and performed MeDIP PCR (mPCR) reactions under various conditions (Table. 3.2). 15 cycles of mPCR, using KAPA polymerase (Kapa Bioscience) resulted in the highest DNA yield and all subsequent mPCR reactions were performed under these conditions.

Table 3-2: Optimization of adapter-mediated PCR.

Condition	Pros	Cons
Standard 15 cycle PCR amplification of mDNA using the standard library prep polymerase (Phusion)	Optimal number of cycles for amplifying 10–20 ng mDNA	Insufficient yield from low concentration (< 20 ng) mDNA
Standard 15 cycle PCR amplification of mDNA using an alternative polymerase (Kapa)	Up to three times higher DNA yield than when using Phusion polymerase. Starting concentrations of as low as 3 ng mDNA was successfully amplified.	None observed.
'Dual mPCR' – mDNA is split into 2 equal parts and each part is amplified separately in a 12 cycle reaction, using Phusion polymerase	A higher DNA yield is obtained in most cases and it involves the use of lower PCR cycles which reduces the generation of clonal DNA fragments	mDNA is single stranded and thus difficult to quantify, so one cannot be sure of the exact concentration of mDNA used in PCR reactions. Using too little DNA can lead to complete failure of the PCR reaction. Dual mMeDIP PCR is not recommended where a low mDNA yield is expected, without prior extensive DNA quantification.
'Dual mPCR' – mDNA is split into 2 equal parts and each part is amplified separately in a 12 cycle reaction, using Kapa polymerase	A higher DNA yield than when using Phusion polymerase under the same condition.	Same as with Phusion polymerase under the same condition

3.4.6 **Library Size Selection**

PCR amplified mDNA (amDNA) is size selected to generate a uniform sequence library insert size of approximately 200bp. This is done by running the entire amDNA on agarose gels (Section 2.8.5) and excising a gel slice within the appropriate size range. DNA is then extracted from the gel slice with an affinity column based gel extraction kit (Qiagen). A significant amount (> 90%) of DNA was typically lost at this stage. As a result, I developed

an optimised gel extraction protocol. The key changes to the standard protocol are summarised in Table 3.3 below.

Table 3-3: Optimization of the Gel extraction protocol.

Optimisation	**Benefits**
Extracted DNA from ≤ 300 mg gel per column	Small gel slices dissolve more easily. It is important that gel is completely dissolved to fully release DNA from agarose and to prevent the clogging up of column membrane with undissolved gel fragments.
Dissolved gel slices in LoBind tubes	Prevents DNA loss by adhesion to reaction tubes
Added 3 M sodium acetate (pH 5.0) to buffer QG gel dissolution buffer (Qiagen)	Adsorption of DNA to column membrane is severely reduced at pH of > 7.0 and can result in total loss of DNA. Sodium Acetate lowers pH to facilitate DNA adsorption.
Eluted DNA in small volumes with 50^0C EB elution buffer (Qiagen)	Eluting in small volumes help concentrate the purified DNA. This can later be diluted if necessary. EB buffer has the optimal pH of 7.5–8, which is required to release DNA bound to column membrane. Distilled or purified H2O typically used in laboratories have a much lower pH and is therefore unsuitable.
Eluted in DNA with 50^0C EB elution buffer (Qiagen)	Elution with warm EB buffer increases permeability of column membranes which allows DNA to be released more efficiently

3.4.7 **Nano-MeDIP-seq**

I successfully prepared Nano-MeDIP-seq libraries from 50–300 ng starting DNA concentrations and sequenced one library generated from 300 ng starting DNA. The resulting sequence data was enriched for CpGs and covered the methylome at a sufficient depth for methylome analysis [133]. These results showed that Nano-MeDIP-seq can be used for methylome analysis from small amounts of DNA. I therefore used Nano-MeDIP-seq to analyse the methylome of LT-HSCs as detailed below.

3.5 **Nano-MeDIP-seq of LT-HSCs**

I extracted 160–300 ng of DNA from three biological replicates of young, middle-aged and old C57BL/6 mice, and carried out methylome analysis using Nano-MeDIP-seq. Each replicate included 3-30 pooled isogenic mice.

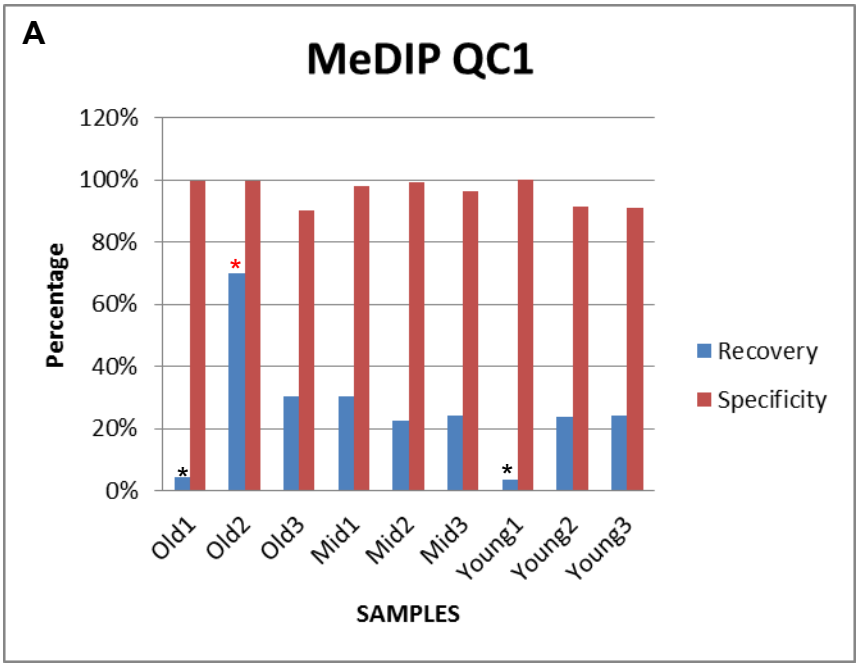
3.5.1 **Quality Control of LT-HSC Nano-MeDIP Libraries**

I confirmed the enrichment of Nano-MeDIP-seq libraries for methylated fragments by performing qPCR on all LT-HSC samples following MeDIP (QC1, Fig 3.9) and PCR amplification of MeDIP DNA (QC2). These were indexed by 'Recovery' (and 'Specificity') and 'Enrichment' respectively (Fig. 3.9). The spiked-in Lambda control fragments were assessed in QC1, while mouse DNA regions, of predetermined methylation status, were assessed in QC2. The full QC method is detailed in Section 2.8.3.1 and Section 2.8.5.1.

3.5.2 **LT-HSC Nano-MeDIP Sequence Data**

Nano-MeDIP libraries were subjected to next generation sequencing. One Young (Young 1) and two Old samples (Old 1 and Old 2) were sequenced on an Illumina Genome Analyzer (GAIIx). All other samples were processed on an Illumina Hi-seq sequencer, which is a newer and improved version of the GAIIx. This resulted in an average of 5.25 Gb raw paired-end reads per sample (S.D ±1.15 Gb) of which approximately 90% were successfully paired and aligned to the mouse genome (mm9 Genome Build 37). An average of 1.92 Gb (S.D ± 0.3 Gb) high quality ($q \geq 10$) uniquely mapping paired-end reads were obtained after filtering out clonal and low quality reads. Approximately 60% of all CpGs in the mouse genome were covered at least one fold (Fig. 3.10), and all MeDIP samples showed a clear CpG enrichment (Table. 3.4) when compared with Input control samples (Average MeDIP enrichment score = 2.73; S.D ± 0.12; n = 9, Average Input enrichment score of 1.11; S.D ± 0.03; n = 2). Additionally, there was a good correlation for all LT-HSC MeDIP samples (Table. A-6; average Pearson's correlation of 0.84, S.D ± 0.09). One replicate for the 'Old'

time point had a lower correlation of approximately 0.7 and was excluded from global analyses to maintain stringency.



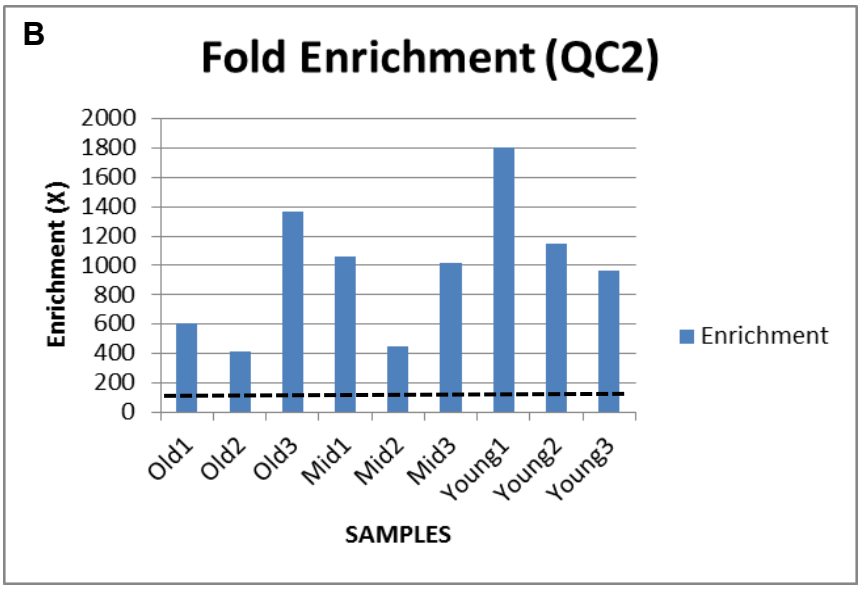
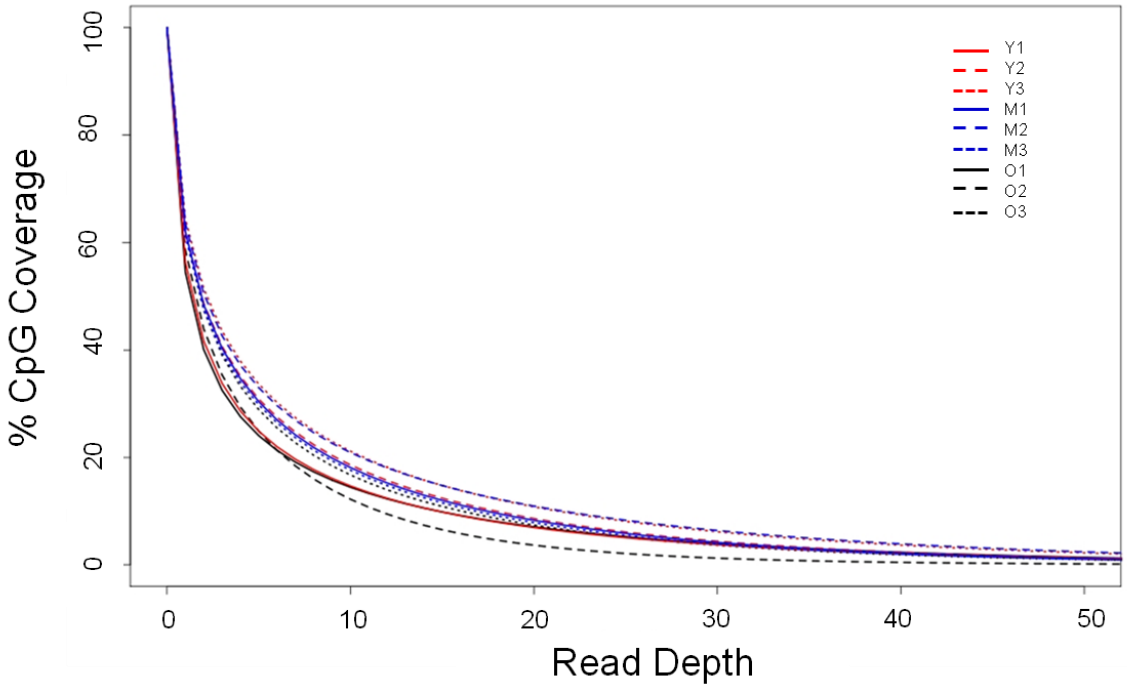


Figure 3-9: MeDIP QC.
A) Quality control after MeDIP. All samples show a highly specific recovery of methylated DNA. Recovery was assayed on spiked-in lambda DNA for all samples except those marked with *. A) All samples show a clear enrichment for methylated DNA following PCR amplification. Broken black line marks the minimum enrichment value (100) required for a successful experiment. This threshold was set based on previous observations within the Medical Genomic group. *a different lambda spike combination was included and assayed for this sample, *mouse genomic region were assayed for these samples and a low level recovery is typical for such regions. qPCR regions are detailed in Table. A-1. *Recovery* and *Specificity* were calculated as detailed in section 2.8.3.1 (QC1), while *Enrichment* was calculated as detailed in section 2.8.5.1 (QC2).

Table 3-4: CpG enchment of MeDIP and Input samples.
Enrichment scores were determined by the MEDIPS program. A score of >2.5 shows that MeDIP samples are enriched for CpGs, relative to input (Enrichment score of ~1).

Sample	CpG Enrichment Score
IP_Young_1	2.82
IP_Young_2	2.77
IP_Young_3	2.71
IP_Mid_1	2.65
IP_Mid_2	2.75
IP_Mid_3	2.7
IP_Old_1	2.98
IP_Old_2	2.57
IP_Old_3	2.65
Input_Young	1.13
Input_Old	1.09

Figure 3-10: CpG coverage of ageing LT-HSC MeDIP-seq libraries.



Average coverage at read depth (X) of 1, 5 and 10 was 60.5%, 30% and 17% respectively. (Standard deviation = 3.3%, 6% and 3% respectively). Coverage is consistent with other (high DNA concentration) MeDIP-seq studies [118,127,133,134].

3.6 **Conclusion**

I successfully isolated mouse phenotypically defined LT-HSCs across developmentally relevant time points and developed the Nano-MeDIP-seq method for DNAm analysis of rare samples. I applied Nano-MeDIP-seq for the generation of LT-HSC (ageing) methylomes and demonstrated the high quality of the data that was generated (Section 3.5.2; [133]).

3.7 **Discussion**

Until now, methylome analysis of LT-HSCs has been unattainable because of their rarity and the cost of isolating them. I overcame the latter by isolating LT-HSCs using the SP method. Hoechst exclusion is a bona-fide method for isolating LT-HSCs; however, its efficacy is easily affected by slight variations in technique. As a result, this method is often used in conjunction with multiple antibodies to augment purity [149] . I validated the SP method and confirmed that my staining techniques efficiently and consistently identified a pure population of LT-HSCs. This demonstrated that additional antibodies were not required and allowed me to isolate LT-HSCs from a large sample size at a reasonable cost.

Irrespective of the cost of isolating them, LT-HSCs remain a rare cell population in the bone marrow. This meant that an excessive number of animals would have been required to obtain sufficient cells for methylome analysis, using the methods that were available at the time of this study. I therefore developed the Nano-MeDIP-seq method for interrogating the methylomes from low numbers of cells and demonstrated its utility for the generation of ageing LT-HSC methylomes. The protocol developed here represents a higher than 10 fold improvement in sensitivity to the original MeDIP protocol and makes this method available to a wider range of studies. The full Nano-MeDIP-seq protocol and its development was published in the April 2012 issue of Nature Protocols [133] and can be found in the appendix of this thesis.

Chapter 4

DNA Methylation Analysis of Ageing LT-HSCs

4.1 **Introduction**

Altered DNA methylation (DNAm) is linked to profound functional changes, and has been implicated in normal development as well as in various diseases [51,80,150]. Mice deficient in enzymes that catalyse DNAm at CpG dinucleotides, DNA methyltransferases Dnmt1 and Dnmt3a, have abnormal HSC self-renewal and differentiation [88,90,151]. Dnmt1 mutant mice were unable to support lymphopoiesis and showed a quantitative increase in LT-HSCs, which rapidly became exhausted after multiple rounds of transplantation [88,151]. Similarly, Dnmt3a-null mice showed enhanced self-renewal at the expense of differentiation [90]. These findings suggest that DNAm is important for HSC regulation and in maintaining the balance between self-renewal and differentiation.

Aged mice show similar phenotypes to that of Dnmt mutants i.e. an increase in the number of phenotypically defined LT-HSCs and skewed differentiation towards myelopoiesis. It is therefore possible that alterations in DNAm contribute to age-related decline in LT-HSC functionality.

DNAm is a highly dynamic epigenetic mark which can fluctuate between cells of the same population [60]. This dynamism is a necessary feature of DNAm and is required for phenotypic plasticity [51]. In order to distinguish relevant age specific changes in DNAm from epigenetic noise, I generated methylomes for three biological replicates of Young, Mid and Old time points. Additionally, LT-HSCs were pooled from isogenic mice, further diluting out any random epigenetic signal that may exist within a particular sample.

In this Chapter, I investigated the role of differential DNAm in ageing LT-HSCs as follows;

1. I compared methylomes for three biological replicates of young, middle-aged and old C57BL/6 (b6) mice. From this, I identified age related changes in DNAm both on a global scale and at specific genomic regions.

2. I investigated the potential functional relevance of differentially methylated regions (DMRs) by assessing the regulatory features of the underlying loci.
3. I reconciled relevant age related changes in DNAm, with the ageing LT-HSC phenotype.

4.2 **Methylome Analysis of Ageing LT-HSCs**

I investigated the role of DNAm in LT-HSC ageing, by studying their methylomes at three progressive age time points. The methylome of Young LT-HSCs was set as the baseline and used as a reference for Mid and Old samples, such that any significant deviation from Young was assumed to be biologically relevant. Whilst I was primarily interested in absolute changes in DNAm with age, i.e. a change from Young to Old; Mid provided a progressive view of the methylome with age and was used to identify putative developmentally programed changes in DNAm.

Normalised sequence data for each sample was converted into wiggle format which I uploaded and viewed on the UCSC genome browser (mm9, NCBI Build 37). This provided a web interface where the mouse LT-HSC ageing methylome could be easily interrogated. I visualised the methylomes of LT-HSC from Young, Mid and Old samples (collectively referred to as ageing samples) and confirmed a good correlation between biological replicates (Fig. 4.1, Table. A-6). DMRs were also queried using this interface; this permitted the visual confirmation of all aDMRs, in biological replicates of each age time point.

Average methylation scores from all covered CpG sites in ageing samples were determined using MEDIPS, which converts enrichment into absolute DNAm values [135]. MEDIPS scores were then aggregated over all CpG sites for the three biological replicates, to generate a representative methylome for each sample. As mentioned in Chapter 3, one of the ‘Old’ samples (‘Old 2’) was excluded from global methylome analysis because of low (< 0.9) correlation compared to the other biological replicates.

Methylomes were depicted on a Circos plot (Fig. 4.2) and I visually assessed the intensity of the blue colour, which indicates global methylation levels (dark blue - high to pale blue/green - low) for each sample. I observed a reduction in the colour intensity of 'Old' methylomes compared to 'Young' methylomes, suggesting a global reduction in DNAm with age (Fig. 4.2).

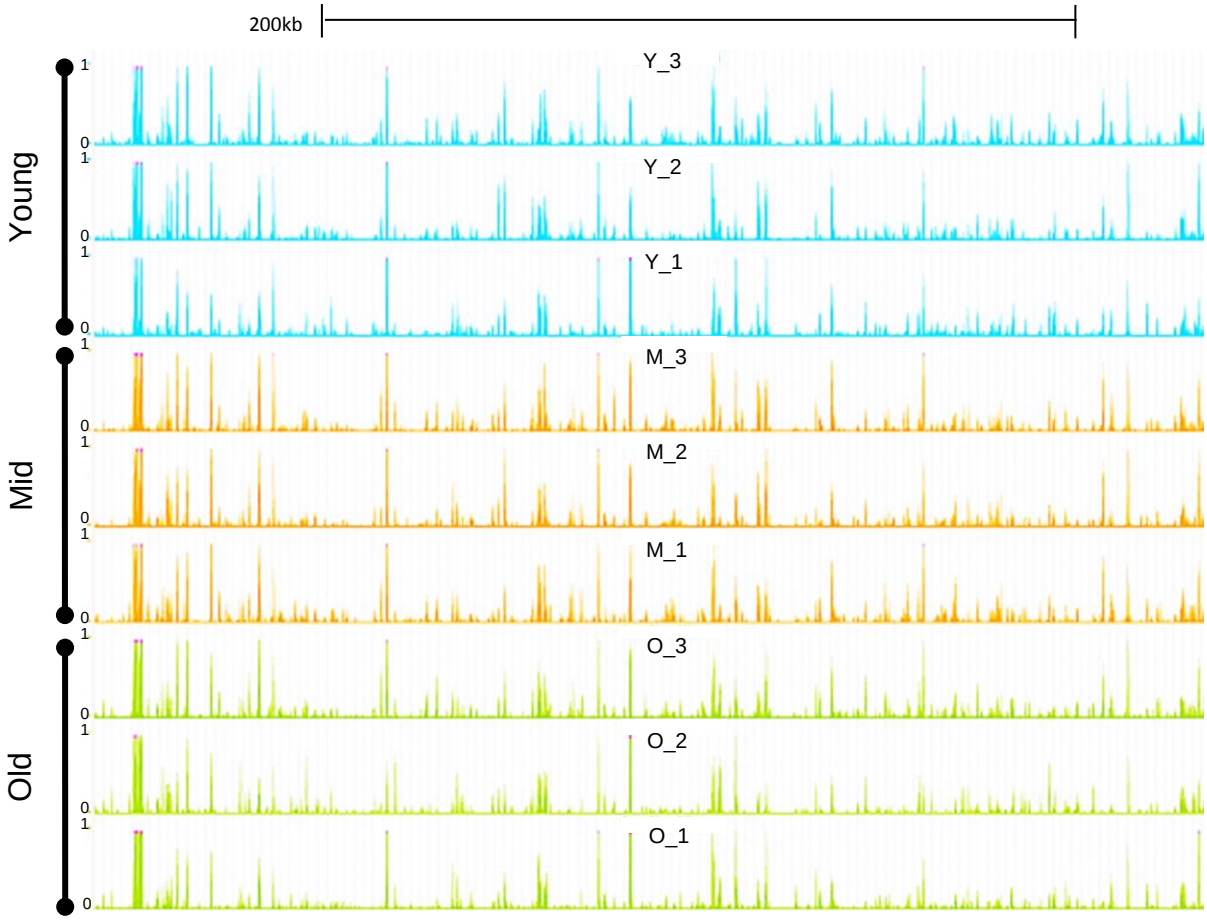


Figure 4-1: Screen shot of UCSC genome browser.
Each track shows the LT-HSC methylomes for ageing samples as indicated. A good correlation between sample replicates can be observed. Peak height represents DNAm levels; scale is from 0 to 1, with 0 being the lowest. Browser scale is 200kb

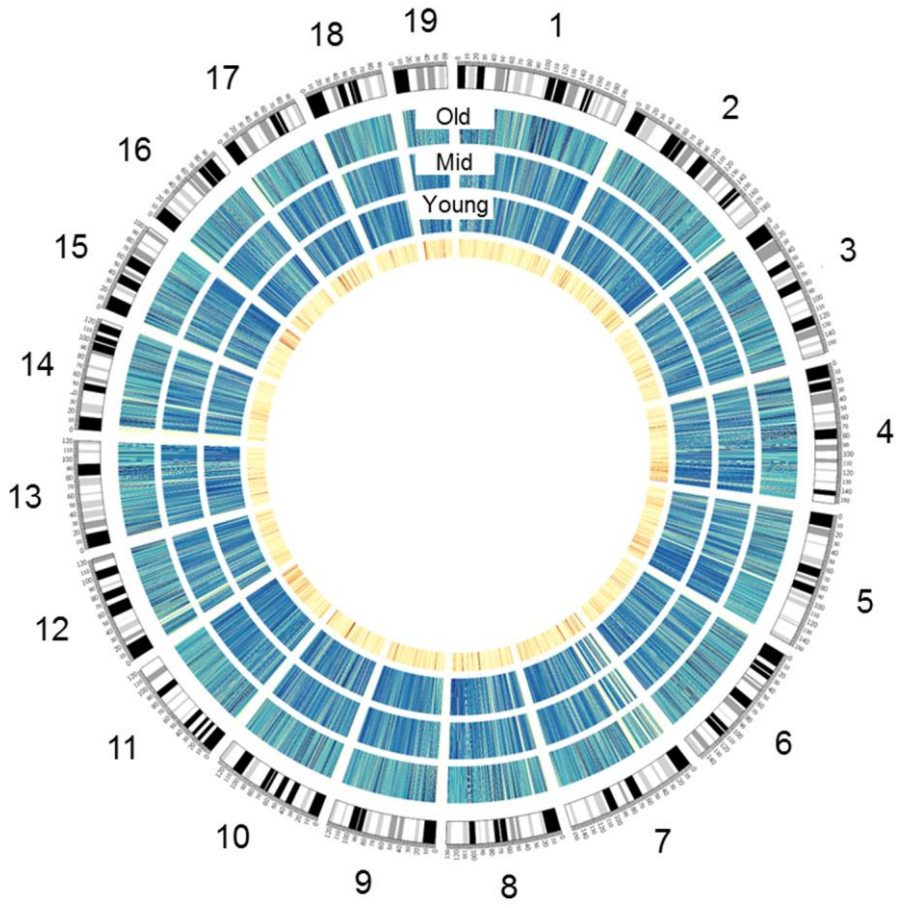


Figure 4-2: Circos plot of LT-HSC methylomes.
Outermost circle shows an ideogram of the mouse genome, dark bands are typically AT rich, heterochromatic regions, while light bands are typically GC rich euchromatic regions. The three blue circles show (from outer circle) Old, Mid and Young LT-HSC methylomes respectively (2 -3 biological replicates). Global methylation level is proportional to colour intensity i.e. dark blue for high methylated and light blue for low. The innermost circle depicts CpG density (dark orange, high and light orange low).

The observed differences in DNAm between the ageing methylomes were quantified using MEDIPS and a 5% ($p < 0.001$, Kolmogorov-Smirnov test) decrease in global DNAm was found in Old HSCs (Fig. 4.3a). This is in contrast to CpG islands (CGIs), which were globally hypermethylated in Old HSCs (Fig 4.3b). The gradual reduction in global DNAm described here is similar to that of a recent study, which found a progressive decrease in methylation levels in human $CD4^+$ T cells with age [152]. This suggests that global decline in DNAm levels may be a common feature of mammalian ageing.

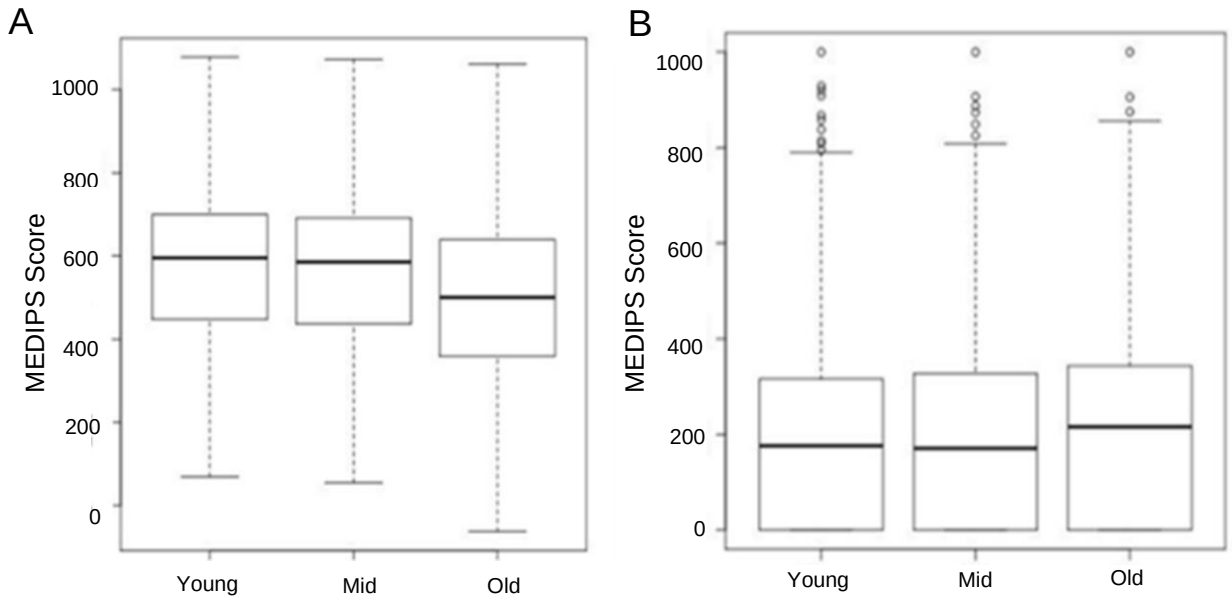


Figure 4-3: Global DNAm levels.
Box plots of A) global B) CGI DNA methylation levels for Young, Mid and Old LT-HSCs and. Global DNA methylation levels are expressed as MEDIPS scores.

4.3 **Age-specific Differentially Methylated Regions (aDMRs) in LT-HSCs**

aDMRs were determined using the MeDUSA pipeline [134] (v1.0.0), which compared normalised read counts for discreet regions between ageing samples as described in Chapter 2. The available tools for determining DMRs in this study were designed for samples where marked alterations in DNAm were expected (e.g. cancer studies). I reasoned that extreme changes in the DNAm of LT-HSCs will probably result in neoplastic transformation or cell death, none of which was observed in ageing LT-HSC samples. Moreover, I observed a high correlation between all LT-HSC methylomes, irrespective of age groups (Section 3.5.2, Section 4.2). It was therefore likely that subtle but significant changes in DNAm drive the ageing phenotype. In order to avoid excluding potential candidates, I selected a moderate FDR threshold of 0.2 for DMR calling in this study. This generated a list of 111 aDMRs, which were significantly differentially methylated between Young and Old samples. Of these aDMRs, 71 were hypermethylated (hyper-aDMRs) (Table. A-2) and 40 were hypomethylated (hypo-aDMR) (Table. A-3) in Old samples compared to Young. These aDMRs appeared to be randomly distributed across all mouse chromosomes (Fig. 4.4), with

the exception of chromosome 14 and 16 (where no DMRs were detected) and chromosome 5, where I found a hotspot that was significantly (p-value < 0.001) enriched for hypermethylated aDMRs. This hotspot corresponds to the location of at least 4 known members of the *Speer* gene family which includes 14 genes that encode putative glutamate-rich proteins of unknown function (Fig. 4.5). Speer genes are thought to arise from partial duplications of the *Dlg5* gene [153]. *Dlg5* is a member of the MAGUK super-family, which are scaffolding proteins that are important in cell adhesion and cell polarity [154].

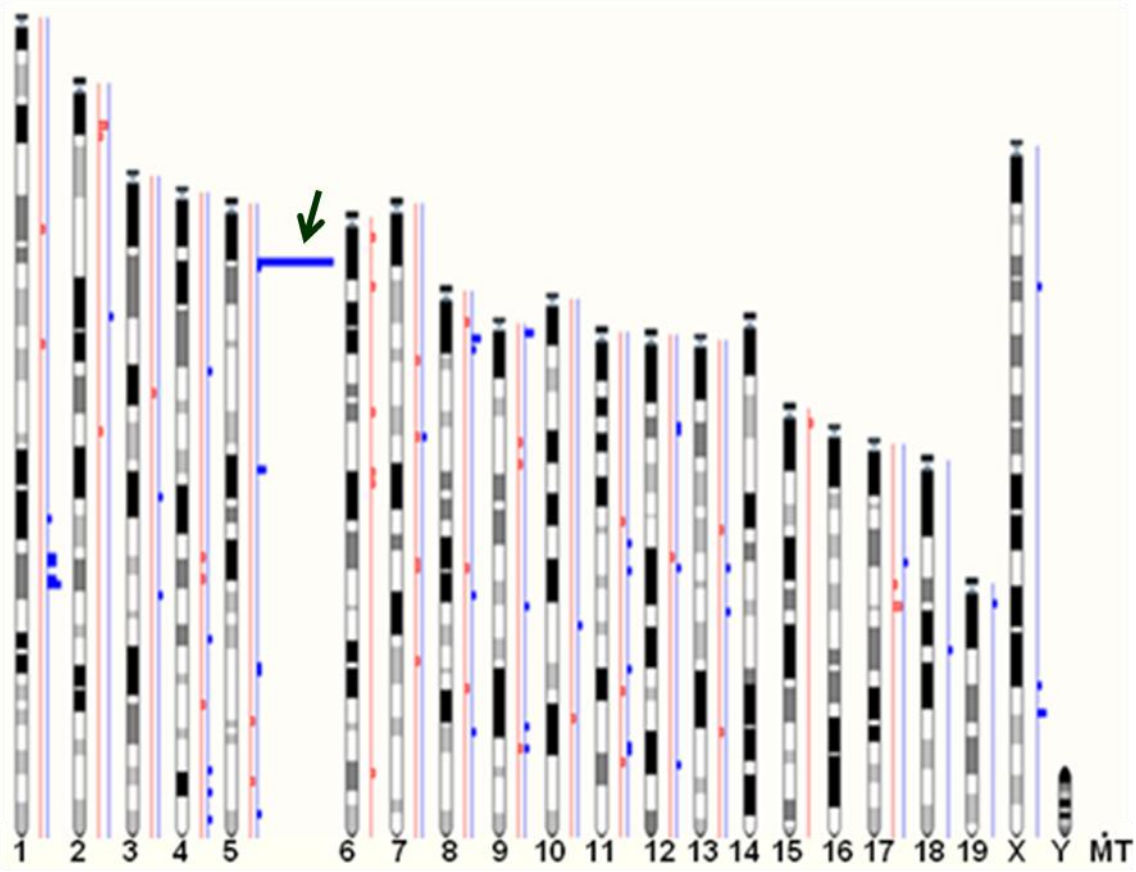


Figure 4-4: Mouse ideograph showing the location of aDMRs.
Hypo-aDMRs are depicted as red peaks while hyper-aDMRs are in blue and are enriched on a region in chromosome 5 (arrowed). Dark bands on mouse ideogram are AT rich, heterochromatic regions, while light bands are GC rich euchromatic regions.

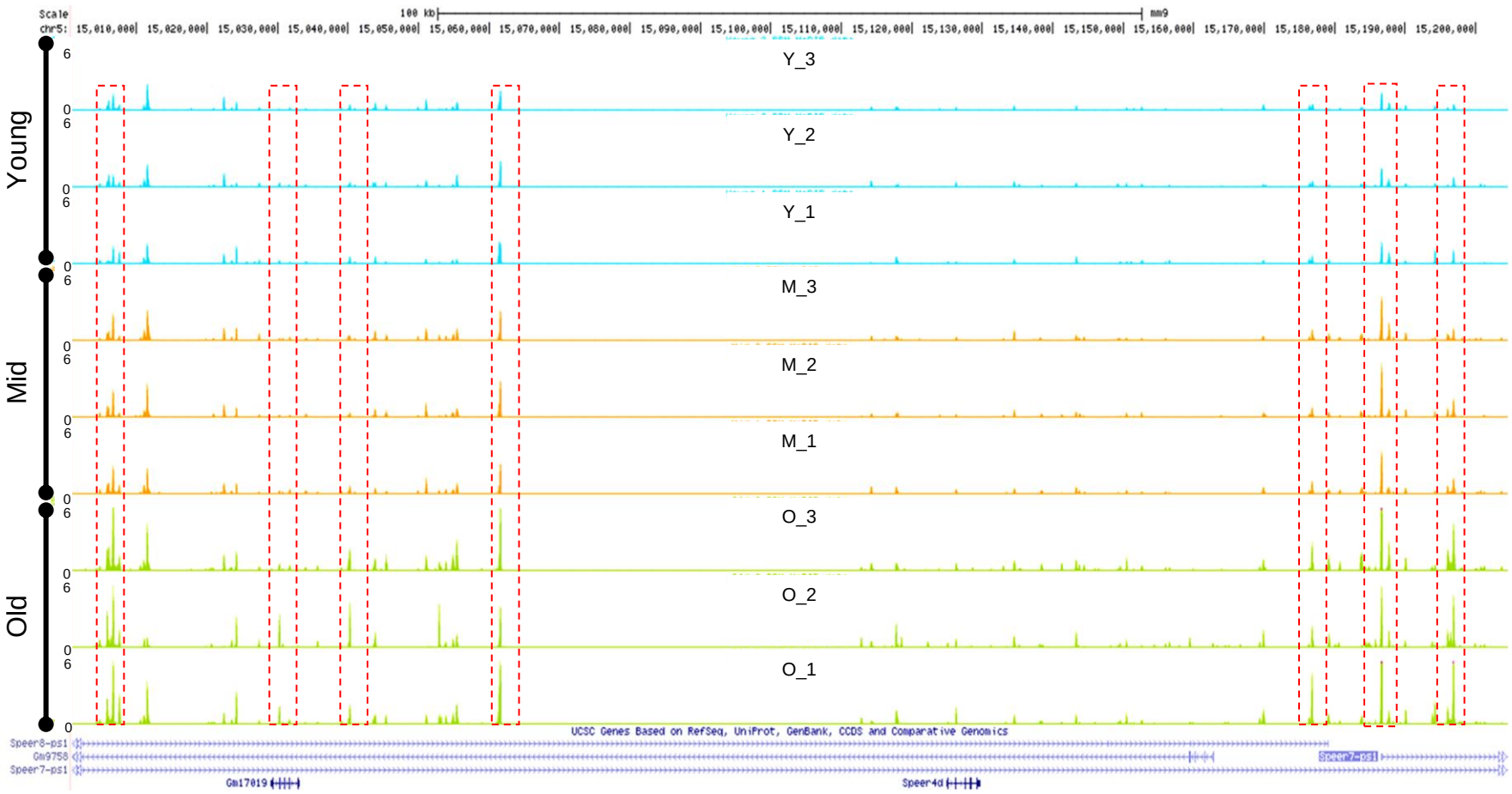


Figure 4-5: UCSC genome browser methylation tracks of ageing samples.
The screen shot shows aDMRs (red boxed) overlapping a group of *Speer* genes on mouse chromosome 5. Peak height represents DNAm levels; scale is from 0 to 6, with 0 being the lowest. Browser scale is 100kb

Using a previously determined cut-off of 100 Kb [155], 96% (107/111) of the identified aDMRs were associated with known genes; henceforth referred to as age-specific differentially methylated genes (aDMGs). I observed a progressive change in methylation with age for many aDMRs. Figure 4.6 shows a DNAm heatmap for 24 randomly selected aDMRs (from a total of 111 aDMRs), displaying progressive and directional DNAm change at during ageing. A heatmap of all 111 aDMRs is shown in Appendix-Figure 7.

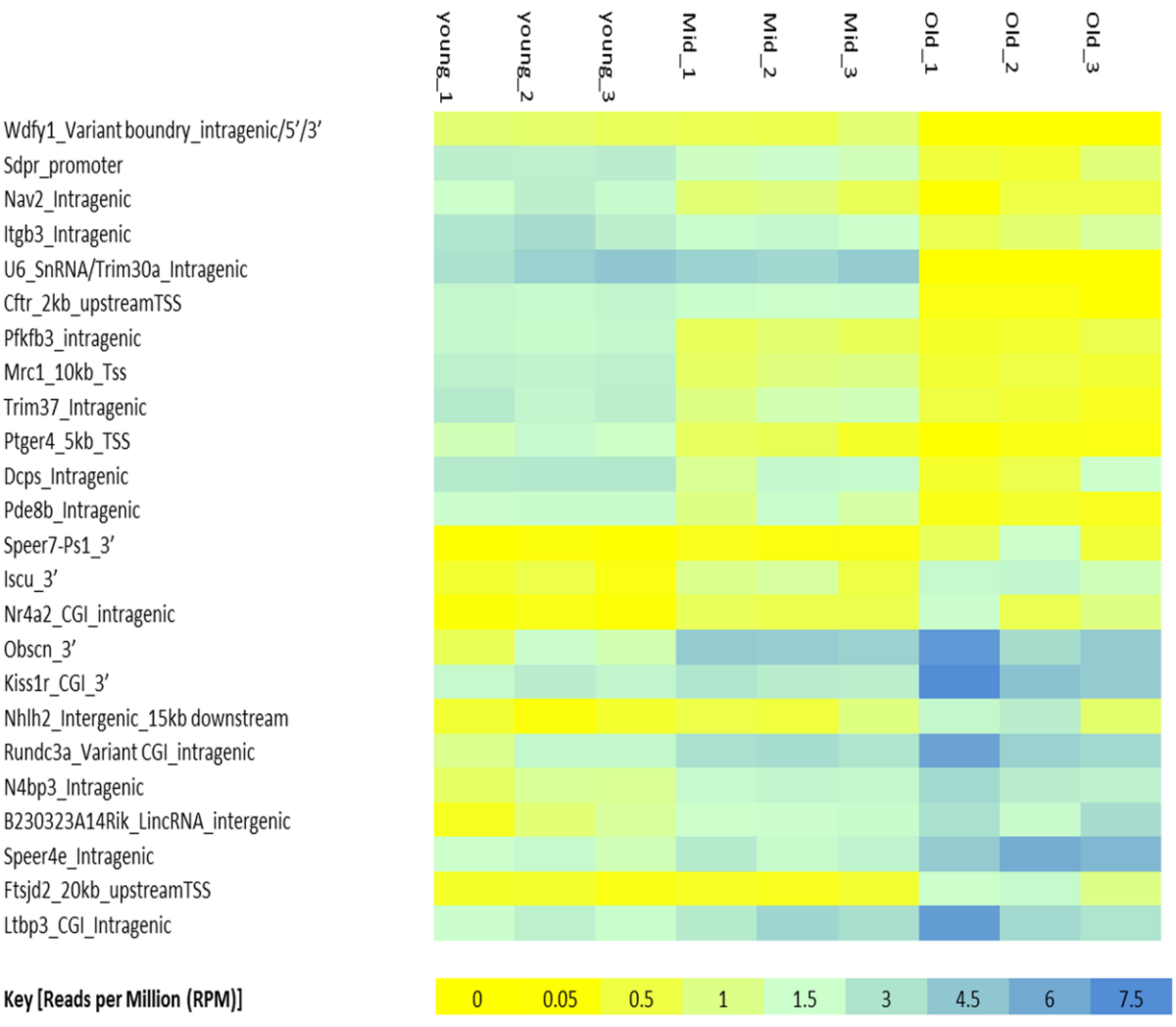


Figure 4-6: Methylation heatmap of selected DMRs in during ageing.
Blue represents hyper and yellow hypomethylation. Corresponding aDMGs are shown on the left.

I used GREAT v2.0.2 [156], a function and pathway enrichment analysis tool, to investigate the biological significance and association of aDMRs but found no significant enrichments. However GREAT, as most gene enrichment analysis tools, requires a large number of test data and the lack of significant enrichment in aDMRs could be as a result of low input data. Subsequently, aDMRs were categorised according to their annotated (Ensembl mm9, NCBIM37 genome build) genomic features. Hypermethylated aDMRs (in Old versus Young) were found to be significantly enriched in CGIs (Fishers Exact p-value = 2.33e-09), CGI shores (p-value = 4.30e-09) and exons (p-value = 6.18e-11), while hypo aDMRs were significantly enriched in exons (p-value = 2.14e-06), CGI shores (p-value = 8.39e-05), introns (p-value = 3.6e-03) and intergenic regions (p-value = 0.028). Further analysis using the Ensembl regulatory build (mm9, NCBIM37 genome build) revealed that CTCF binding sites were most enriched in the hyper aDMRs and c-MYB sites were most enriched in hypo aDMRs.

Many of the identified aDMRs were progressively differentially methylated with age (Fig. 4.6), and therefore potentially associated with ageing. To further elucidate the functional relevance of aDMRs, I investigated the functions of aDMGs and evaluated the regulatory functions of the aDMRs within these genes; this analysis was performed as follows:

1. I determined the biological functions of selected genes using annotations from scientific journals (PubMed: http://www.ncbi.nlm.nih.gov/pubmed) and from NCBI gene summaries (NCBI: http://www.ncbi.nlm.nih.gov/gene)
2. I queried the corresponding aDMRs on the UCSC genome browser as described for Figure 4.1, however here, the UCSC 'regulation' track was activated to show binding sites for UCSC default transcription factors (TFs) such as, CTCF, GATA1 and that of the TF co-activator, p300. I also utilised UCSC TF binding site (TFBS) tracks from a recently published article, which detailed TFBS in haematopoietic stem and progenitor cells [157].

3. From this analysis, I selected candidate genes (from a total of 107 aDMGs) that could be involved in the ageing of LT-HSCs and/or contained an aDMR that was enriched for TFBSs

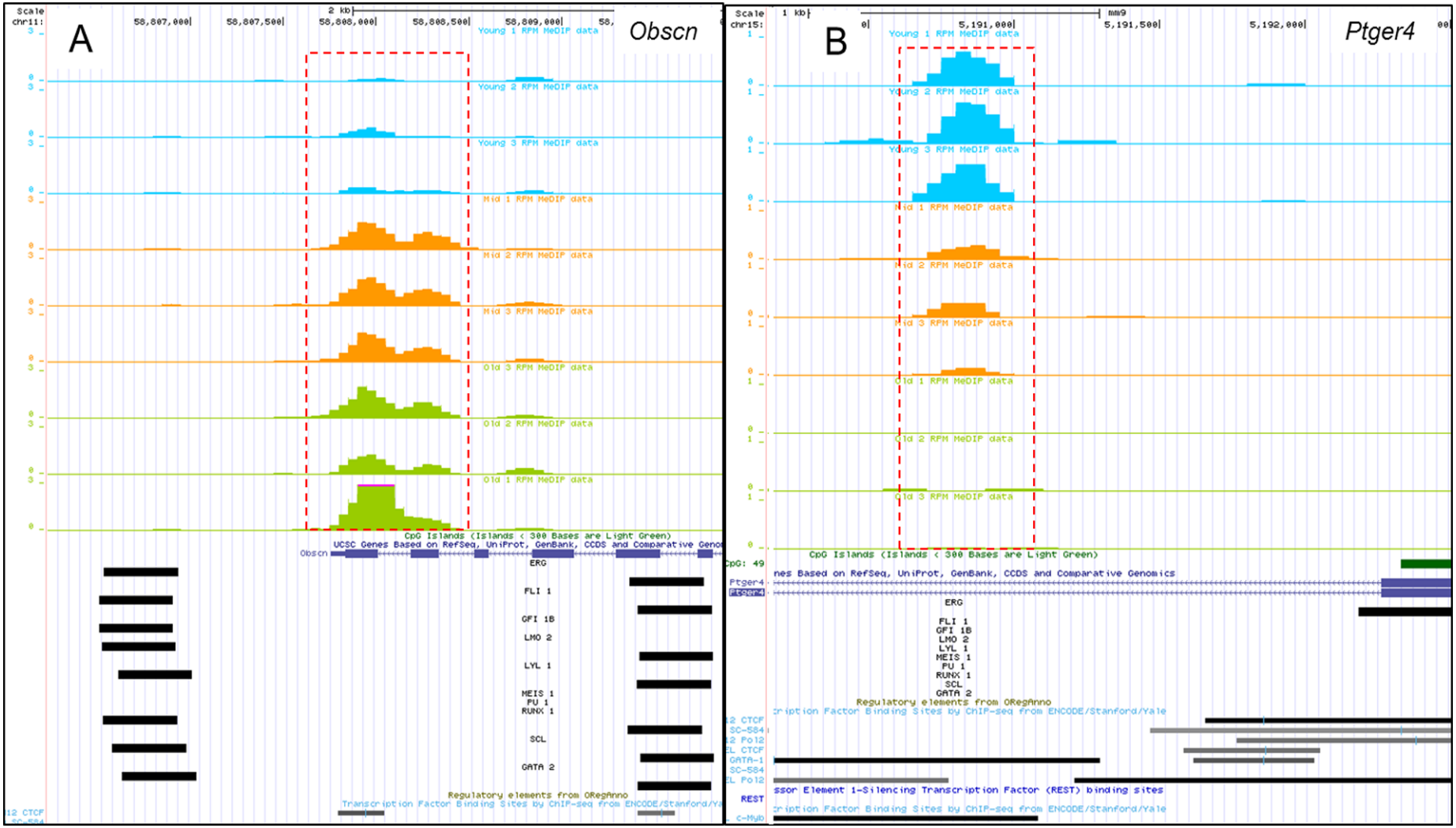
The selected candidate genes, with relevant biological functions, are summarised in Table 4.1. Some of these DMGs were also shown on UCSC genome browser tracks in Figure 4.7.

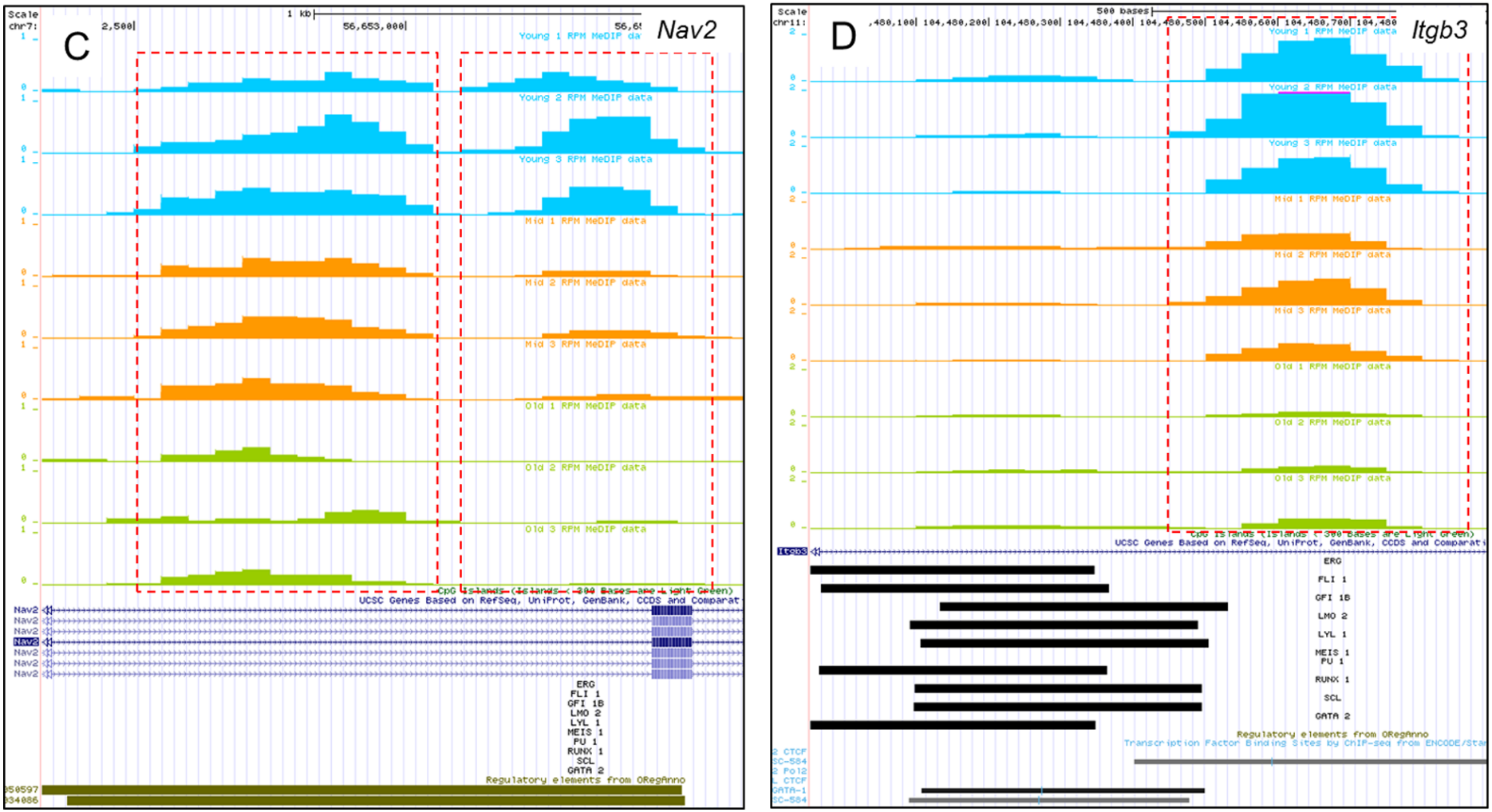
Table 4-1: Summary of candidate aDMGs

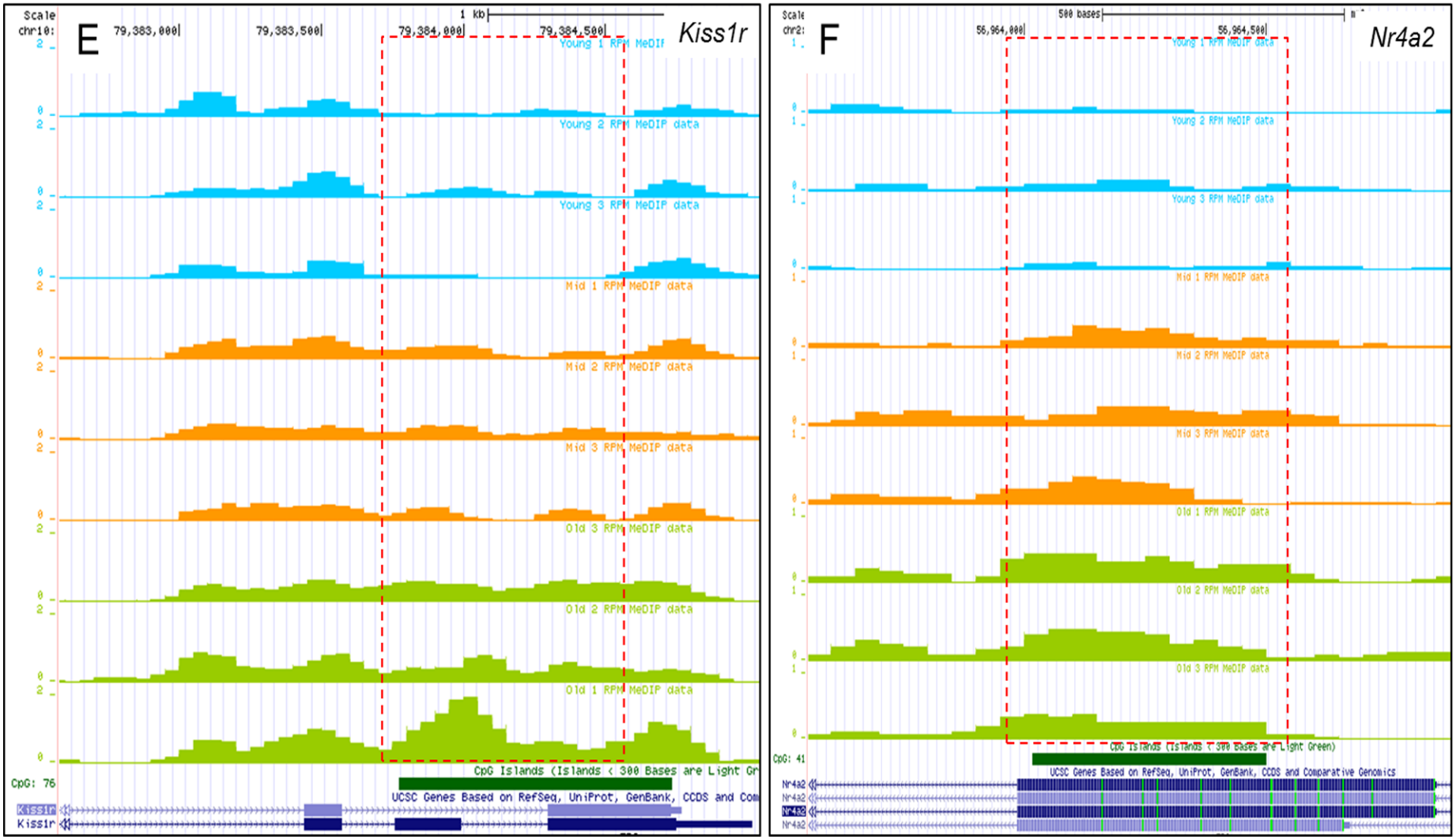
aDMG	Relevant Function	Source	DMR feature
Cftr	Similar function as ABC/G2 on LSP cells. *Cftr*, which is also known as *ABC/C7*, is a member of the ABC transporter super family which are involved in cell detoxification and multi-drug resistance (MDR)	NCBI Genes	Pronounced hypo-aDMR within 2kb of TSS
Cpne2	A calcium-dependent membrane-binding protein. Involved in various signalling pathways and membrane trafficking	PubMeD [158,159]	intronic hypo-aDMR at CTCF, p300 and Pol2 binding site (fig. 4.7G)
Cyfip2	Facilitates T-Cell adhesion. Overexpression increases fibronectin-mediated adhesion	[160]	3' hypo-aDMR at CTCF,p300 and Pol2 binding site
Dcps	scavenger decapping enzyme, role in mRNA decay and nuclear pre-mRNA splicing	[161]	Age progressive hypo-aDMR in gene intron
Hsf4	Activates heat shock response proteins under stress conditions. Involved in the negative regulation of DNA binding activity	NCBI Genes	Exon boundary hyper-aDMR
Itgb3	A cell surface protein which is important in cell signalling and cell adhesion	NCBI Genes	Binding site for several TFs within 300bp of intronic hypo-aDMR Fig 4.7D)
Kissr1	G-protein coupled receptor with a role in cancer cell growth and metastasis	PubMeD [162]	3' CGI hyper-aDMR (fig 4.7E)
Ltbp3	Mediates the targeting of latent TGF-beta into extracellular cell matrix	[163]	Exon boundary CGI hyper-aDMR

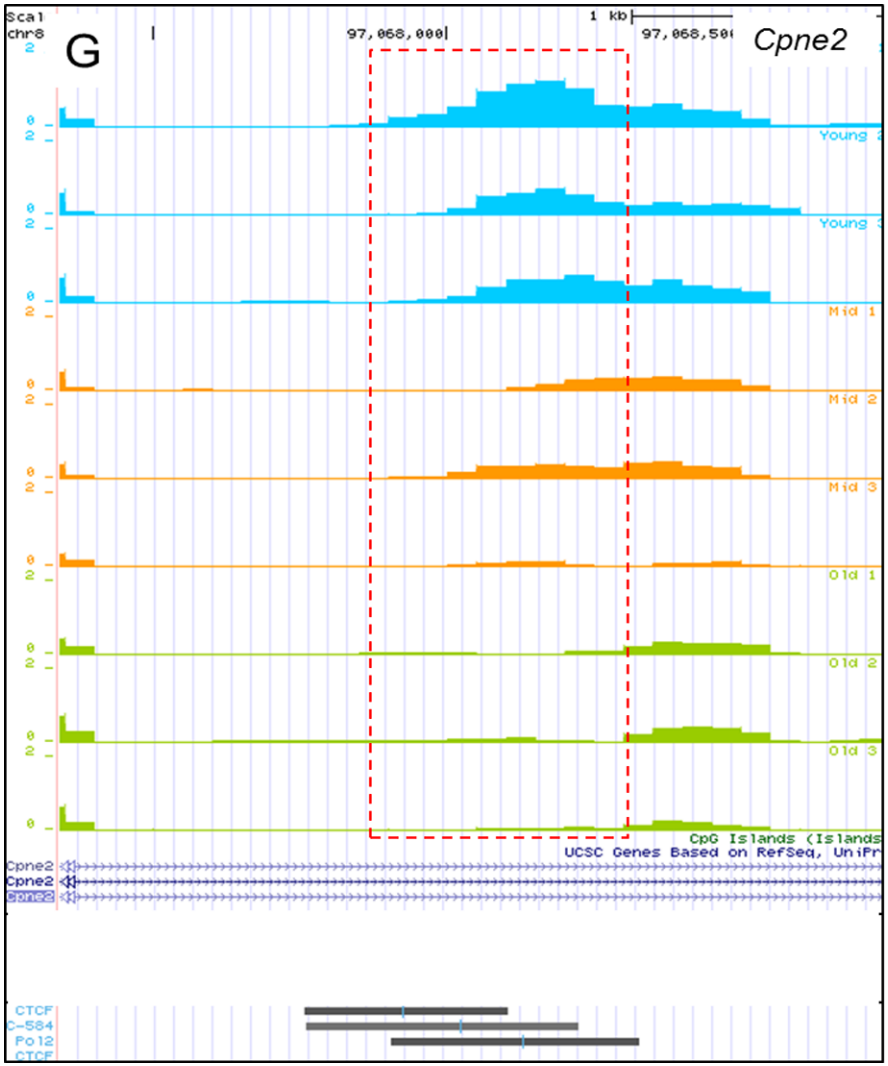
Mrc1	Mannose receptor with a role in immunity and inflammation	NCBI Genes	Pronounced hypo-aDMR at 3'UTR
Nav2	Involved in cellular growth and migration	NCBI Genes	Intronic hyper-aDMR as well as hypo-aDMR over UCSC annotated region (Fig 4.7C)
Nr4a2	Increases cell migration in response to stromal-derived factor-1a (SDF-1a)	PubMeD [164]	Hyper-aDMR over CGI and AUG codon rich exon (Fig 4.7F)
Obscn	Role in cellular assembly and loss of function mutations in this gene results in apoptotic resistance.	[165]	3' Hyper-aDMR overlapping CTCF binding site and flanked by several TFBS (Fig 4.7A)
Pde8	Regulates chemotaxis of activated lymphocytes	[166]	Hypo-aDMR in 8a and 8b
Ptger	G-protein coupled receptor for prostaglandin E2 that can activate T-cell signalling and enhance immature B-cell apoptosis	NCBI Genes	Marked intronic hypo-aDMR which overlaps c-MYB, Pol2 and GATA-1 binding sites and is within 1 kb of several other TFBS (Fig 4.7B)
Sdpr	Involved in caveolae formation and signalling. The expression of this phospholipid-binding protein increases in serum starved cells and during HSC ageing	[94,99,167,168]	Two hypo-aDMR around TSS and within 500bp of CTCF binding site (Fig 4.7H)
Trim30a	Negative regulator of NF-kappa B. Role in immunity and inflammation	[169]	Marked hypo-aDMR within intron which completely overlaps a snRNA

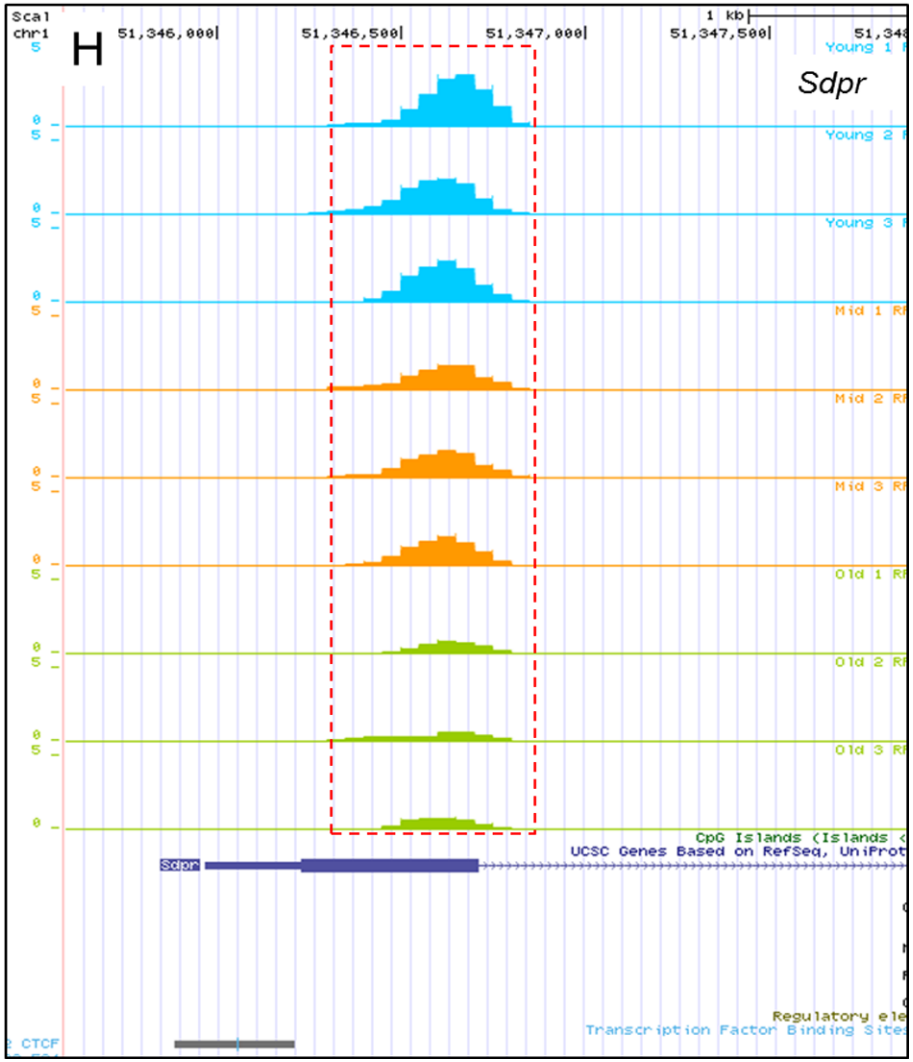
Figure 4.7: UCSC genome browser tracks of DMRs (red boxed) in selected genes as indicated.
Blue, orange and green peaks shown for all genes represent methylation in Young, Mid and Old samples respectively and peak height is proportional to methylation level. Solid black/grey bars in A, B, D, G and H represent TFBSs, green bars in B, E and F represent CGIs, and blue bars represent exons (solid) and introns (broken). Olive green bars shown for *Nav 2* (C) represent UCSC annotated regulatory elements while the broken green lines shown for Nr4a2 (F) represent methionine start codons.




C
Nav2
Young 1 RPM MeDIP data
Young 2 RPM MeDIP data
Young 3 RPM MeDIP data
Mid 1 RPM MeDIP data
Mid 2 RPM MeDIP data
Mid 3 RPM MeDIP data
Old 1 RPM MeDIP data
Old 2 RPM MeDIP data
Old 3 RPM MeDIP data
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparat
ERG
FLI 1
GFI 1B
LMO 2
LYL 1
MEIS 1
PU 1
RUNX 1
SCL
GATA 2
Regulatory elements from ORegAnno
D
Itgb3
Young 1 RPM MeDIP data
Young 2 RPM MeDIP data
Young 3 RPM MeDIP data
Mid 1 RPM MeDIP data
Mid 2 RPM MeDIP data
Mid 3 RPM MeDIP data
Old 1 RPM MeDIP data
Old 2 RPM MeDIP data
Old 3 RPM MeDIP data
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparat
ERG
FLI 1
GFI 1B
LMO 2
LYL 1
MEIS 1
PU 1
RUNX 1
SCL
GATA 2
Regulatory elements from ORegAnno


E
Kiss1r
F
Nr4a2
Young 1 RPM MeDIP data
Young 2 RPM MeDIP data
Young 3 RPM MeDIP data
Mid 1 RPM MeDIP data
Mid 2 RPM MeDIP data
Mid 3 RPM MeDIP data
Old 1 RPM MeDIP data
Old 2 RPM MeDIP data
Old 3 RPM MeDIP data
CpG Islands (Islands < 300 Bases are Light Green)
UCSC Genes Based on RefSeq, UniProt, GenBank, CCDS and Comparative Genomics


G
Cpne2
Young 1
Young 2
Young 3
Mid 1
Mid 2
Mid 3
Old 1
Old 2
Old 3
CpG Islands (Islands
UCSC Genes Based on RefSeq, UniPr
Cpne2
Cpne2
CTCF
Pol2
CTCF


H
Sdpr
Young 1
Young 2
Young 3
Mid 1
Mid 2
Mid 3
Old 1
Old 2
Old 3
CpG Islands (Islands
UCSC Genes Based on RefSeq, UniProt
Sdpr
CTCF
Regulatory ele
Transcription Factor Binding Site

4.4 **Conclusions**

I established that the ageing LT-HSC methylome becomes globally hypomethylated, and hypermethylated at some CGIs. I identified several aDMRs and their corresponding genes and highlighted many candidate genes that could account for age-dependent changes in LT-HSCs. These genes were selected as potential candidates for LT-HSC ageing based on their known or predicted functions and many of these appear to have a role in cell signalling and adhesion, which is important in haematopoiesis and HSC regulation. The availability of a focused list of relevant candidate genes represents a significant improvement in the understanding of HSC ageing.

4.5 **Discussion**

A key step in carcinogenesis is global hypomethylation and local hypermethylation at some CGIs [82,84,155]; it is therefore interesting that the same changes were observed during LT-HSC ageing. As age is a major risk factor for cancer, it is possible that these age-related changes in DNA methylation are a prerequisite for an oncogenic transformation in aged organisms. I identified several regions that were differentially methylated with age and found an enrichment of CTCF binding sites in regions that were hypermethylated in Old samples. CTCF is a methylation sensitive DNA binding transcription factor and boundary (insulator) element [170,171]. This TF binds DNA to insulate neighbouring genes from inappropriate (cross) regulation by cis-acting elements e.g. enhancers. CTCF has also been implicated in alternative splicing [172]. In this context, DNAm at CTCF binding sites promoted exon exclusion [172]. Hypermethylation of CTCF binding sites interferes with its binding to DNA resulting in aberrant gene expression, as observed in various diseases including imprinted diseases and cancer [170,173]. Similarly, hypermethylation of these sites during LT-HSC ageing could result in dysregulated gene expression, through the expression of normally silent genes or/and inappropriate activation of splice variants. Conversely, age-related

hypomethylated DMRs were enriched in binding sites for c-MYB, which is a key TF in haematopoiesis and HSC self-renewal [174,175]. Loss of methylation at these sites potentially enhances binding of c-MYB to DNA and perhaps has a wide-spread effect on the expression of several downstream targets. Age-related disruption of TF interaction with DNA, will most likely result in significant alterations in gene expression and may somewhat account for the loss of functionality that is observed in LT-HSCs, during ageing.

As well as global changes in the LT-HSC methylome, several genes were also found to be differentially methylated (Table 4.1; Table. A-2 and A-3). These include several genes with functions in cell signalling and adhesion, inflammation and immunity, cellular assembly, cell growth and metastasis. HSCs reside in the bone marrow where cell-cell interactions with specialized niche cells regulate their fate. These interactions take place through factors such as calcium ions, adhesion molecules and various soluble factors, and act to provide instructive signals to HSCs [176]. *Itgb3* is highly expressed on SP LT-HSCs [177] and was found to be hypomethylated in this study. This gene is a member of the Integrin family of adhesion molecules, which are involved in cell-cell/extracellular matrix (ECM) adhesion and HSC-niche interactions. These processes are crucial for HSC regulation [178]. Integrins mediate the adhesion process by binding to other ECM and cell surface molecules such as fibronectin [179]. Additionally, Itgb3 interacts with several proteins, such as matrix metalloproteinases (MMPs) and G proteins, within lipid rafts, to form so-called supramolecular complexes, which are potent signal transducers [180]. Other aDMGs that appear to be involved in the adhesion and cell signalling process include *Cyfip 2,* which is involved in fibronectin mediated cell-cell adhesion [160], *Sdpr*; a component of flask shaped lipid rafts (caveolae) [168], and *Ptger* and *Kiss1r,* which are both G-protein coupled receptors [162,181]. Additionally, genes involved in cell detoxification, inflammation and lymphocyte regulation were differentially methylated with age (Table. 4.1). DNAm associated dysregulation of genes that are involved in lymphopoiesis could account for the altered

differentiation capacity of aged HSCs, while alterations in inflammation genes could explain the proinflammatory environment and functional decline of the immune system with age.

Age-related changes in the DNAm of primitive haematopoietic cells have previously been described [117]. This study identified an age-dependent hypomethylation of differentiation-associated genes and reported several instances of *de novo* DNAm with age [117]. However, a heterogeneous haematopoietic progenitor cell (HPC) population was studied. Additionally, these HPCs were obtained from different anatomical locations i.e. umbilical cord blood for Young HPCs and mobilized peripheral blood HPCs for Old samples. The findings in this study were therefore likely affected by the different environment and that of mobilization. Such confounders were avoided in this thesis, as a more homogenous population of phenotypically defined LT-HSCs was studied for age-specific differential DNAm. Furthermore, the relevance of LSP cells in the study of HSC ageing was demonstrated in Chapter 3, i.e. this specific cell population is one that changes during ageing. The findings presented here are therefore likely to reflect a more accurate picture of HSC ageing and represent the first comprehensive account of differential DNAm in these cells during physiological ageing.

Chapter 5

Gene Expression Analysis of Ageing LT-HSCs

5.1 **Introduction**

The ability of HSCs to self-renew and differentiate into other blood cell types is tightly regulated by both intrinsic mechanisms and the micro-environment or niche in which they reside [33,35,40,46,182]. Several studies have reported a decline in HSC functionality and a differentiation bias towards myelopoiesis with age [7,9,93-95,99]. Such studies carried out gene expression profiling of purified HSCs and provided evidence for an intrinsic change in HSCs with age [7,94,99], which involved an up-regulation of myeloid specific transcripts [94]. It has since been shown that myeloid skewing with age is due to a compositional change in the ageing HSC pool [19,102]. As previous studies of HSC ageing performed gene expression profiling on the entire HSC pool, it is likely that the observed myeloid signature is a result of the overrepresentation of myeloid biased (My-bi) HSC clones in aged samples, rather than an intrinsic up-regulation of myeloid transcripts in balanced HSCs. Recent findings indicate that My-bi HSC clones are the most primitive subset of HSCs. They possess long-term reconstituting potential and show a high ABC/G2 pump activity that can be identified via Hoechst dye exclusion assays [16,19]. Indeed, cells with the least Hoechst dye retention, i.e. lower side population (LSP) cells, are enriched for My-bi clones as well as for LT-HSCs (Fig. 3.2, [19]). The LSP phenotype and differentiation potential is thought to be stable with age [19], however, I and others have found that it is the LSP fraction of HSCs in particular that accumulates with age (Fig. 3.5 [7,19]). Based on these notions, I concluded that the currently available data on HSC gene expression during ageing, most likely mirrors the change in composition of the HSC pool with age and is not a true reflection of the intrinsic changes that occur during LT-HSC ageing. In order to elucidate the mechanisms governing HSC ageing, I performed global gene expression analysis of LSP cells at the previously described ageing time points (Chapter 3) and investigated the changes that occur in gene expression during age. I analysed the transcriptomes of ageing HSCs by RNA-sequencing (RNA-seq), which in this case involves the sequencing of expressed (mRNA) transcripts on next-generation sequencing platforms. RNA-seq offers several advantages

over microarray hybridisation, which is another major gene expression analysis method [183]. These include 1) a high sensitivity for detecting very low expressed and rare transcripts, 2) a negligible background noise compared to the high background noise in microarray data and 3) the ability to detect novel transcripts and splice variants/isoforms. The latter advantage is due to the fact that RNA-seq involves the generation of cDNA sequences within a given sample which is then mapped back to the relevant genome, while hybridization techniques rely on a prior knowledge of existing transcripts, in order for these to be tilled on arrays, and subsequently detected in gene expression studies. In aggregate, RNA-seq is a sensitive method, which allows the accurate detection of changes in gene expression without *a priori* assumptions.

As this study concerns the ageing of an otherwise wild-type and homogeneous cell population, I hypothesised that HSC ageing will be accompanied by subtle but significant changes in gene expression, such as, the inappropriate activation of silent transcripts, splice variants and the generation of novel isoforms, all of which are only detectable by RNA-seq. As gene expression can vary within (allelic variation) and between (stochastic and population-specific variation) cells, I analysed the expression of three biological replicates for each time point, each consisting of HSCs pooled from several animals of a similar age. This ensured that biological noise was tapered and only significant and biologically relevant changes in gene expression were identified.

5.2 **RNA-sequencing of the Ageing LT-HSC Transcriptome**

I performed RNA-sequencing of mRNA from three biological replicates of HSCs from young (Young), middle-aged (Mid) and old (Old) animals using Illumina next-generation sequencing technology. One replicate each of Young and Old samples were sequenced on the Illumina GAIIx sequencing platform, while all other samples were sequenced on the Illumina Hi-seq platform, an improved model of the Illumina sequencers. Although there was a slight decrease in the correlation between samples sequenced on different platforms, compared to those on the same platforms, the correlation scores in either scenario was remarkably high (average Pearson's correlation score 0.94 versus 0.98 respectively). Based on this, I concluded that the use of data from the two different sequencing platforms was unlikely to influence the findings in this study. Moreover, I performed two technical replicates (one 36 bp and one 50 bp sequence read length) for all samples sequenced on the Hi-seq, and observed a perfect Pearson's correlation score of 1 for all samples (Table. 5.1), thus validating the reproducibility of the Hi-seq technology.

Finally, I observed a high correlation between all replicates and age time points (Table. 5.1), which suggests that gene expression of HSCs is relatively stable during ageing. This finding supports my hypothesis that gross alterations in gene expression are unlikely to govern HSC ageing (Section 5.1).

Table 5-1: Table of Pearson's correlation for RNA-seq libraries.
Correlation values highlighted in grey represent a sample's correlation to itself while those in red text represent correlation between technical replicates. 'YoungB6' and 'OldB6' were sequenced on the Illumina GAIIx while all others were sequenced on the Hi-seq platform. Correlation values range from 0–1 with 1 being the highest score.

Samples	YoungB6	YoungB6_2	YoungB6_2_50bp	YoungB6_3	YoungB6_3_50bp	MidB6_1	MidB6_1_50bp	MidB6_2	MidB6_2_50bp	MidB6_3	MidB6_3_50bp	OldB6	OldB6_2	OldB6_2_50bp	OldB6_3	OldB6_3_50bp
YoungB6	1.00	0.85	0.89	0.89	0.91	0.90	0.92	0.93	0.95	0.97	0.97	0.96	0.86	0.89	0.88	0.88
YoungB6_2	0.85	1.00	1.00	0.99	0.99	0.99	0.98	0.97	0.96	0.94	0.94	0.96	0.99	0.99	0.99	0.99
YoungB6_2_50bp	0.89	1.00	1.00	1.00	1.00	0.99	0.99	0.97	0.97	0.96	0.96	0.97	0.99	0.99	0.99	0.99
YoungB6_3	0.89	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.98	0.97	0.97	0.98	0.99	0.99	1.00	1.00
YoungB6_3_50bp	0.91	0.99	1.00	1.00	1.00	0.99	1.00	0.99	0.98	0.97	0.98	0.98	0.99	0.99	0.99	0.99
MidB6_1	0.90	0.99	0.99	1.00	0.99	1.00	1.00	0.99	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99
MidB6_1_50bp	0.92	0.98	0.99	0.99	1.00	1.00	1.00	0.99	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.99
MidB6_2	0.93	0.97	0.97	0.98	0.99	0.99	0.99	1.00	1.00	0.99	0.99	0.98	0.97	0.98	0.98	0.98
MidB6_2_50bp	0.95	0.96	0.97	0.98	0.98	0.98	0.99	1.00	1.00	1.00	1.00	0.99	0.96	0.97	0.97	0.97
MidB6_3	0.97	0.94	0.96	0.97	0.97	0.98	0.98	0.99	1.00	1.00	1.00	0.98	0.95	0.96	0.96	0.96
MidB6_3_50bp	0.97	0.94	0.96	0.97	0.98	0.98	0.98	0.99	1.00	1.00	1.00	0.98	0.95	0.96	0.96	0.96
OldB6	0.96	0.96	0.97	0.98	0.98	0.98	0.99	0.98	0.99	0.98	0.98	1.00	0.96	0.98	0.98	0.97
OldB6_2	0.86	0.99	0.99	0.99	0.99	0.99	0.99	0.97	0.96	0.95	0.95	0.96	1.00	1.00	1.00	1.00
OldB6_2_50bp	0.89	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.97	0.96	0.96	0.98	1.00	1.00	1.00	1.00
OldB6_3	0.88	0.99	0.99	1.00	0.99	0.99	0.99	0.98	0.97	0.96	0.96	0.98	1.00	1.00	1.00	1.00
OldB6_3_50bp	0.88	0.99	0.99	1.00	0.99	0.99	0.99	0.98	0.97	0.96	0.96	0.97	1.00	1.00	1.00	1.00

5.3 **Transcriptome Analysis of LT-HSC Ageing**

50,601 annotated ensembl transcripts were covered by the RNA-seq of ageing samples and these were analysed for gene expression and differential expression as described in materials and methods. Briefly, sequenced cDNA transcripts were mapped back to the mouse genome and each aligned transcript was used to determine relative transcript abundance, which was then compared across different age groups, to determine age-differential expression. Of these 50,601 transcripts, a total of 1290 were found to be significantly (FDR < 0.05) differentially expressed between Young and Old samples, with 430 genes downregulated and 860 genes upregulated in Old samples relative to Young samples (Y-O comparison). Additionally, 2074 transcripts were found to be differentially expressed between Young and Mid samples (Y-M comparison), while 1187 were differentially expressed between Mid and Old (M-O comparison). A count of differentially expressed genes for all comparison sets, in this study, is shown in Table 5.2.

67%, 59%, and 51% of differentially expressed transcripts were upregulated, in older animals of the aforementioned comparison sets, respectively. The observation that the majority of differentially expressed transcripts in HSCs were upregulated with age is consistent with an age-dependent global hypomethylation in HSCs (Chapter 4).

Table 5-2: Counts of differentially expressed genes

Comparison Set	Upregulated with age	Downregulated with age	Total changed with age
Young to Mid (Y-M)	**1215**	**859**	**2074**
Mid to Old (M-O)	**608**	**579**	**1187**
Young to Old (Y-O)	**860**	**430**	**1290**

I investigated the biological relevance of differentially expressed transcripts using the web based Ingenuity pathway analysis gene ontology tool (IPA). From this, I determined predicted biological functions that were altered during HSC ageing. Figure 5.1 provides a summary of the top ten biological functions that were up- and down regulated in HSCs at the three age points studied. This was further summarised in Figure 5.2, where it can be seen that the most functionally significant (p-value = 3.13E-14) age-dependent downregulation of transcripts occurred between young and middle-age, and involves genes that regulate proliferation of cells (arrowed), while the most functionally significant (p-value = 1.16E-14) transcript upregulation occurred between middle- and old age, affecting genes involved in cell movement (arrowed).

An important feature of foetal liver (FL) murine haematopoiesis is the presence of highly proliferative HSCs, which are necessary to rapidly establish haematopoiesis during ontology [41]. This highly proliferative state is observed in foetal and neonatal mice, where over 90% of all HSCs are actively cycling [184]. This is in contrast to haematopoiesis in adult mice (> 8 weeks) where over 90% of all HSCs are quiescent [41,184]. I determined that genes involved in the regulation of cell proliferation were the most significantly downregulated genes from Young (8–12 weeks) to Mid (~12 months) mice. This includes *Ikaros,* an important gene in the development of FL HSCs, whose overexpression in adult HSCs leads to diminished quiescence [25]. These observations are consistent with a switch from proliferative to quiescent HSCs during adult mouse haematopoiesis. Although this switch is thought to occur at approximately 4 weeks after birth [184], it is possible that the expression of genes involved in foetal HSC proliferation, persists beyond this period, and is only fully downregulated after puberty (> 12 weeks). I performed phenotyping experiments involving 6–7 week old adolescent mice and consistent with published observations [12,18], found that the majority of KLS defined HSCs from mice of this age did not possess the SP phenotype and were $CD34^{+}$ (Fig. 5.3), both of which are properties associated with foetal and actively cycling HSCs [12,34]. The SP phenotype, which is a marker of HSC quiescence [185], and

is thus highly enriched in adult LT-HSCs, appears to emerge in 8 week old mice (selected as Young in this thesis), but remains extremely rare compared to older mice (Fig. 3.5). Although only SP and thus bona fide adult LT-HSCs were isolated and compared in this thesis (Chapter 3), it may be that adult LT-HSCs from young mice retain the expression of several genes that were required for the highly proliferative states of FL HSCs, and these genes are only fully downregulated after puberty and by middle-age. The identification of such genes is important in our understanding of how HSC quiescence is achieved and maintained.

Following the downregulation of proliferation genes between Young and Mid, another highly significant change in gene expression appears to occur between Mid and Old. This involves the upregulation of genes involved in cell movement (Fig. 5.2). Cell movement is an important process in haematopoiesis and in the interaction of HSCs within specialised niches. I found that several genes involved in HSC-niche interactions, such as Selectins, Integrins, Metalloproteinases and Cell adhesion molecules (CAMs), were upregulated between Mid and Old samples. It is possible that by the upregulation of cell movement genes, HSCs from old mice alter their affinity for a certain niche, which causes them to migrate to an alternative niche, which subsequently results in a change in their development potential(s).

The Young - Old comparison dataset represents the ultimate change in gene expression during HSC ageing. I determined the overall age-dependent change in gene expression, by comparing HSCs from young- to that of old mice. I found that genes involved in the differentiation of blood cells (p-value = 3.42E-18) and cellular migration (p-value = 5.3E-12) were most significantly down- and upregulated respectively, during HSC ageing (Fig. 5.2). Additionally, IPA analysis of biological pathways that were differentially regulated in Old relative to Young samples revealed that the B-cell development (p-Value = 1.38E-07) and the glutamate degradation III (p-Value < 5E-03) pathways were most significantly down- and upregulated respectively, with age. While the relevance of the latter pathway in HSC ageing is currently unclear, age-dependent downregulation of the B-cell development pathway is

consistent with a reduction in the ability of aged HSCs to produce B-lymphocytes [94]. Furthermore, I found that the T-helper cell differentiation pathway was the second most significantly (p-Value = 5E-03) upregulated pathway during HSC ageing. Interestingly, the previously described reduction in B-lymphopoiesis in aged HSCs was reported to be accompanied by an increase in the production of T-cells (and myeloid cells) [94]. Taken together, these results suggest that age-related alterations in the differentiation potential of HSCs are a result of cell intrinsic changes in gene expression. This is consistent with the findings of several other studies of HSC ageing [7,94,99].

Finally, I performed IPA gene ontology analysis of all significantly age-dependent differentially expressed genes (Y-O). I detected a strong enrichment of genes involved in cell movement (p-value < 1E-10), 134 of which were upregulated and 85 downregulated (Table. A-4). These genes could be further subdivided into functional categories such as homing, migration, chemotaxis, and infiltration; both up- and downregulated, while other sub-categories like transmigration and extravasation were exclusively downregulated in HSCs from older animals. Additionally, Aryl hydrocarbon receptor (AHR), which regulates HSC migration through binding of its agonist, 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) [186,187], was identified as one of the most significantly (p < 1E-08) overrepresented upstream regulators of age-dependant differentially expressed genes. These findings implicate cell movement in HSC ageing.

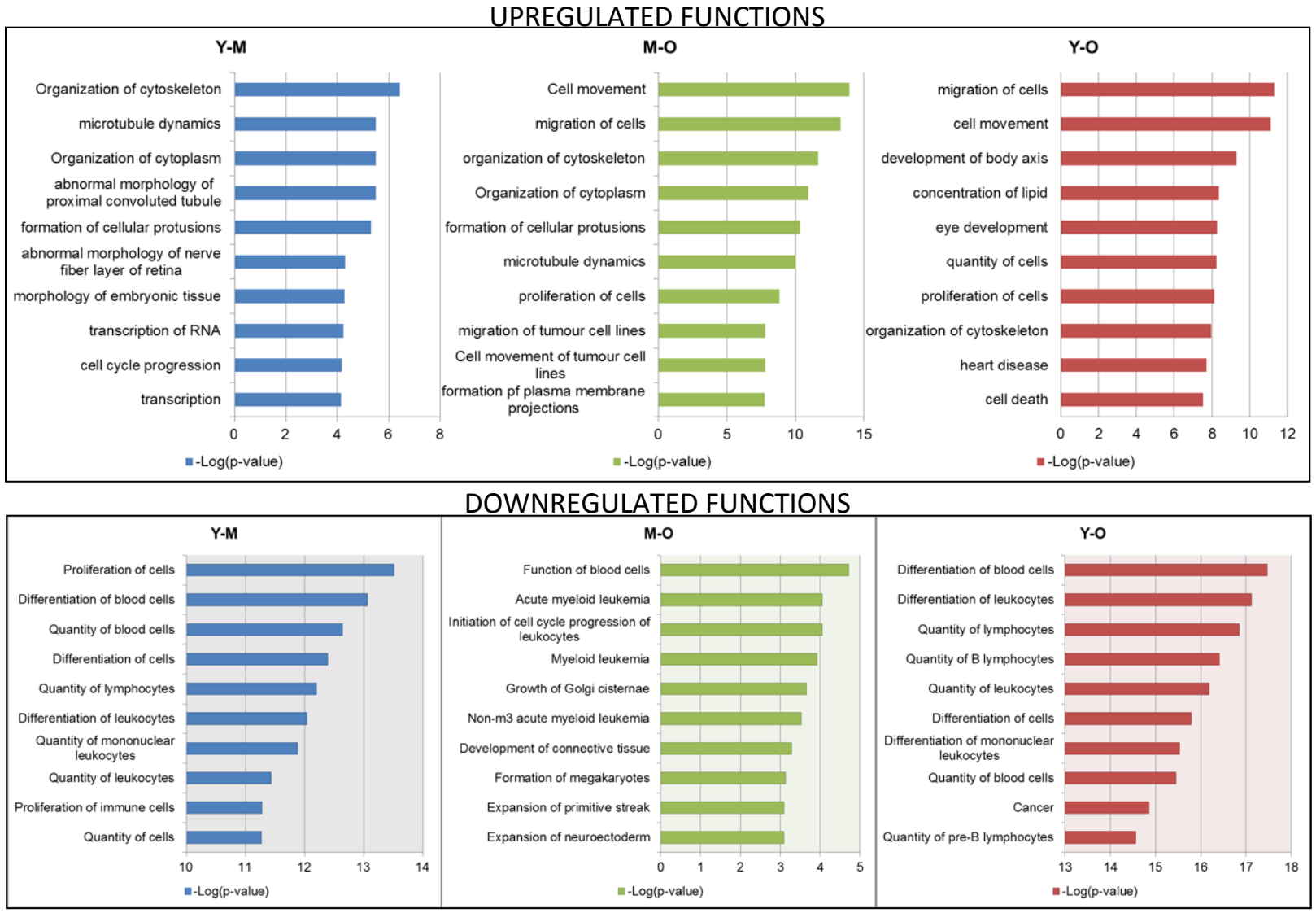


Figure 5-1: Top ten biological functions of age up- and down regulated genes at various time points.
Y, M and O are young, middle-aged and old respectively. Change shown is from the youngest to the oldest time point in each comparison set.

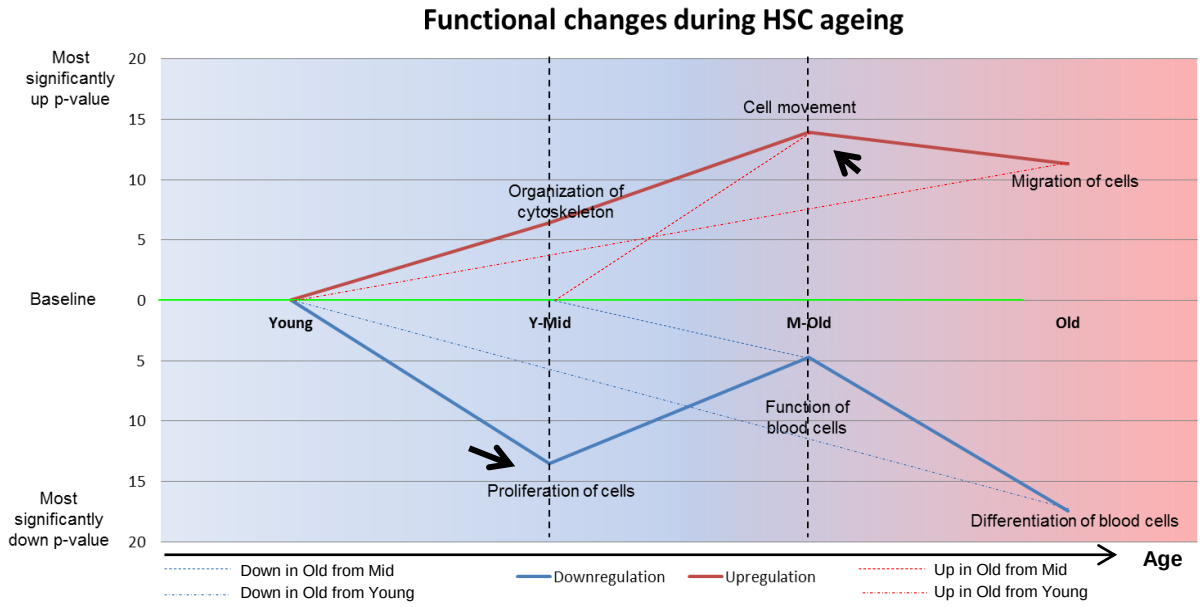


Figure 5-2: The dynamics of gene expression during HSC ageing.
Values shown on y axis is –Log(p-value). Cell Movement was the most significantly upregulated process during ageing, and this occurs between Mid and Old time points. Migration of cells appeared to be the overall most significantly upregulated process, from Young through Mid to Old. Proliferation of cells was the most significantly downregulated process during ageing, this occurs between Young and Mid time points. Differentiation of blood cells appeared to be the overall most significantly downregulated process from Young through Mid to Old.

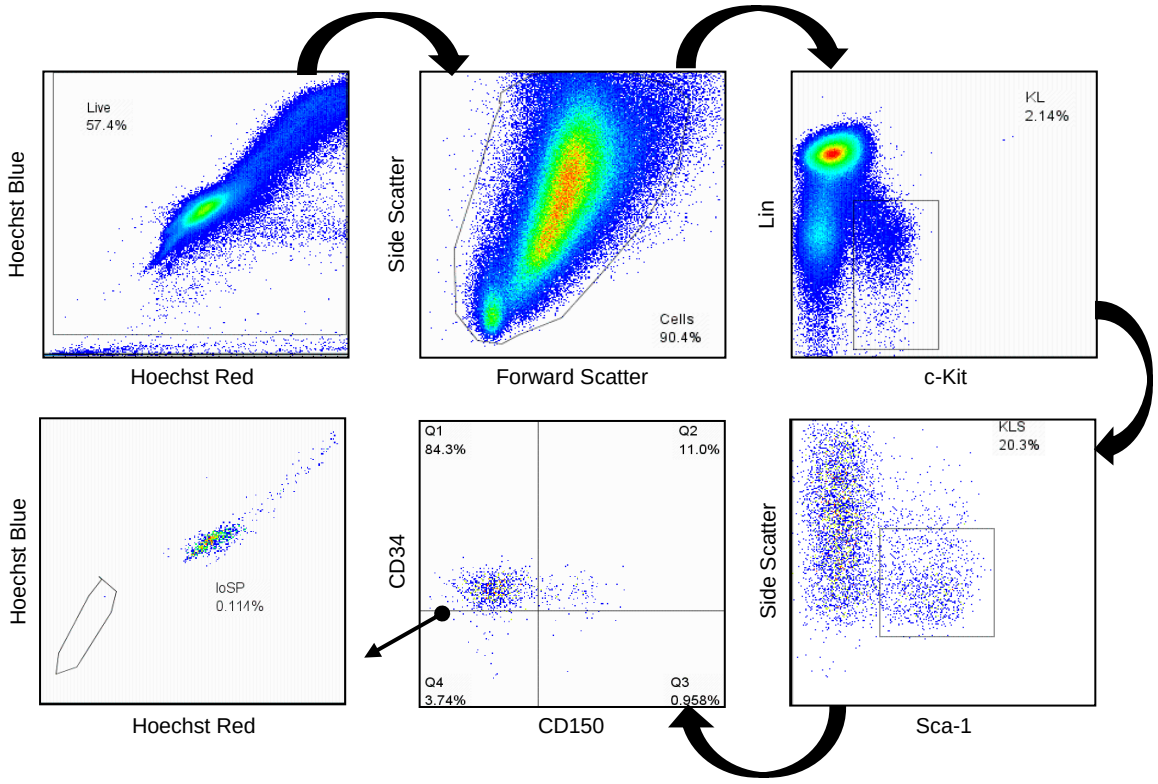


Figure 5-3: Antibody and SP phenotype of HSCs from 6-7 week old mice.
HSCs at this stage are reminiscent of foetal liver HSCs and are CD34+ and non-SP. Each plot shows the phenotype of cells gated on the previous plot as indicated by the curly arrow. The final plot shows the Hoechst profile of CD34+CD150- KLS (Q1) cells.

5.4 **Validation of RNA-seq by qRT-PCR**

To validate the RNA-seq data, I performed qRT-PCR on selected genes that were determined to be age-differentially expressed by RNA-seq. These candidates were selected based on their relevance to the aims of this thesis, as summarised below (Table. 5.3). The validation process also served as an independent confirmation of age differential expression, using independent biological replicates.

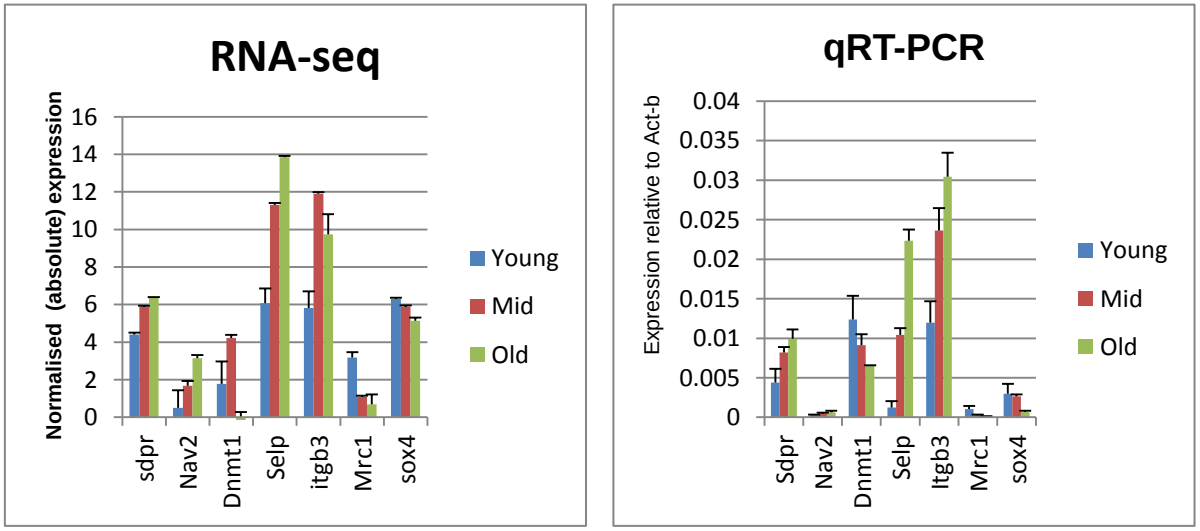
Table 5-3: Selected genes for RNA-seq validation by qRT-PCR.

Gene	Change with age	Summary
Dnmt1	Downregulated	Encodes the DNA methyltransferase enzyme which is responsible for the maintenance of DNA methylation marks in the genome. DNMT1 has been shown to be important in HSC regulation [151]
Itgb3	Upregulated	ITGB3 is a cell surface protein with a key role in cell signalling and cell adhesion (NCBI genes). This gene is hypomethylated in HSCs with age (Chapter 4)
Mrc1	Downregulated	Mannose receptor with a role in immunity and inflammation (NCBI genes). Hypomethylated in HSCs with age (Chapter 4)
Nav2	Upregulated	Involved in cellular growth and migration (NCBI genes) and both of these processes are differentially regulated in HSCs with age (Section 5.3). This gene is hyper- and hypomethylated in HSCs with age (Chapter 4)
Sdpr	Upregulated	Involved in caveolae formation and signalling. The expression of this phospholipid-binding protein increases in serum starved cells and during HSC ageing [94,99,167,168]. hypomethylated in HSCs with age (Chapter 4)
Selp	Upregulated	Has a key role in inflammation [188] and haematopoietic cell movement [189]. Previously identified as upregulated and important in HSC ageing [99]
Sox4	Downregulated	Encodes a transcription factor that is crucial for B lymphopoiesis [190]

I performed 'two-step' qRT-PCR on LT-HSCs that were obtained from three independent biological replicates as detailed in materials and method (Chapter 2). Briefly, FACS sorted LSP cells were lysed and converted directly into cDNA, which was analysed for gene expression on a qPCR machine using TaqMan assays for selected genes. qRT-PCR expression data was then compared to normalised expression values from RNA-seq. Figure 5.4 shows the expression of the genes detailed above as determined by RNA-seq (left) and by qRT-PCR (right). Overall, I observed a good agreement in the trend of gene expression between;

1. Genes - of the seven genes listed in Table 5.3, *Selp* and *Itgb3* were found by RNA-seq to show the highest expression in HSCs while *Nav2* and *Mrc1* had the lowest expression. This was confirmed by qRT-PCR.
2. Age groups for each gene – All seven genes were found to be differentially expressed with age by qRT-PCR and in the same direction as in RNA-seq. Additionally, Mid samples showed intermediate expression values for five out of the seven genes in both techniques.

Figure 5-4: Validation of RNA-seq by qRT-PCR



Plot show the gene expression of seven candidate genes in HSCs, from Young, Mid and Old mice. Gene expression was determined by RNA-seq and qRT-PCR. Data from the two technologies appear comparable. At least 3 biological replicates are shown for all samples. Error bar represents standard error of the mean.

5.5 **Comparison to Previous Studies of HSC Ageing**

As detailed in Section 5.1, there were two other similar data sets on adult HSC gene expression during ageing. These data sets were generated independently (Rossi et al and Chambers et al), and involved the expression profiling of HSCs from young and old C57BL/6 mice, as done in this study [94,99]. However, in contrast to the current study, a heterogeneous mix of HSCs was profiled and gene expression was analysed using microarray technology. Nevertheless, I was interested in comparing these data sets with my data, in order to;

1. Determine the degree of similarity between the gene expression of HSCs profiled by me and others.
2. Identify genes and biological processes that are commonly age-differentially regulated in all three data sets.

To perform this analysis, I obtained a list of genes that were differentially expressed between young and old mice in each of the three studies. These lists consisted of several genes, where more than one transcript was differentially expressed. I edited the lists such that each gene was represented only once (regardless of the number of differentially expressed transcripts. As a result, several 'duplicate' genes were removed and this prevented an artificial inflation of the degree of overlap between these studies. Subsequently, I compared the edited gene lists between studies and identified genes that were common to the three data sets. This comparison was performed separately for specified up- and downregulated genes, and is represented as a Venn diagram in Figure 5.5. All three studies including the current had more genes that were up- than downregulated, reiterating the significance of gene upregulation in HSC ageing. Given that the Rossi et al. [94] and Chambers et al. [99] data sets were generated using similar technology (microarray), and by purification strategies that supposedly yield very similar HSC populations (KLS;$CD34^-$;$Flk2^-$ and SP; KLS respectively), it was surprising to find a low overlap of differentially expressed genes

between these studies. In fact, the identified overlap is only a little higher than between the current study and the Chambers study (Fig. 5.5).

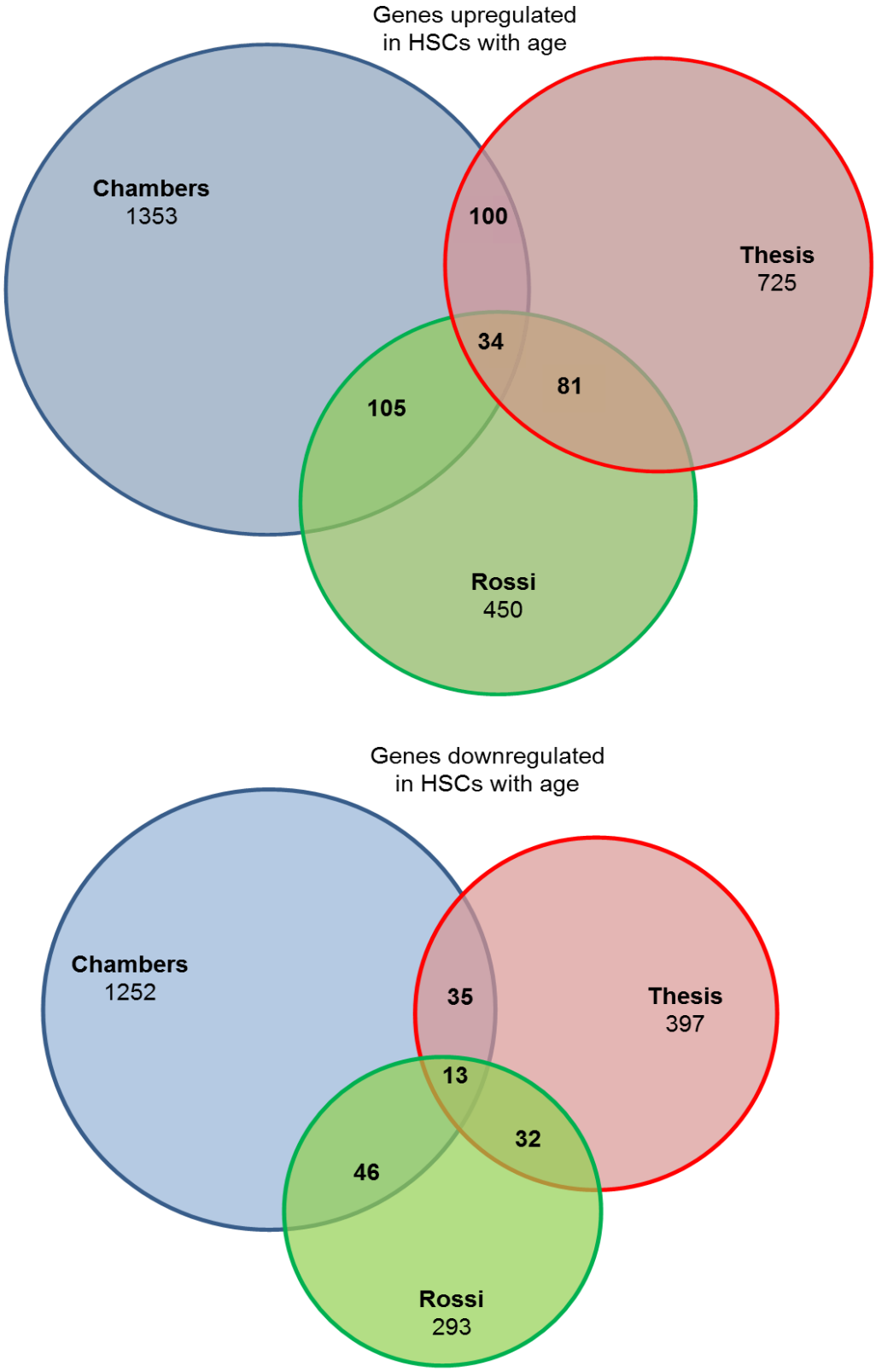
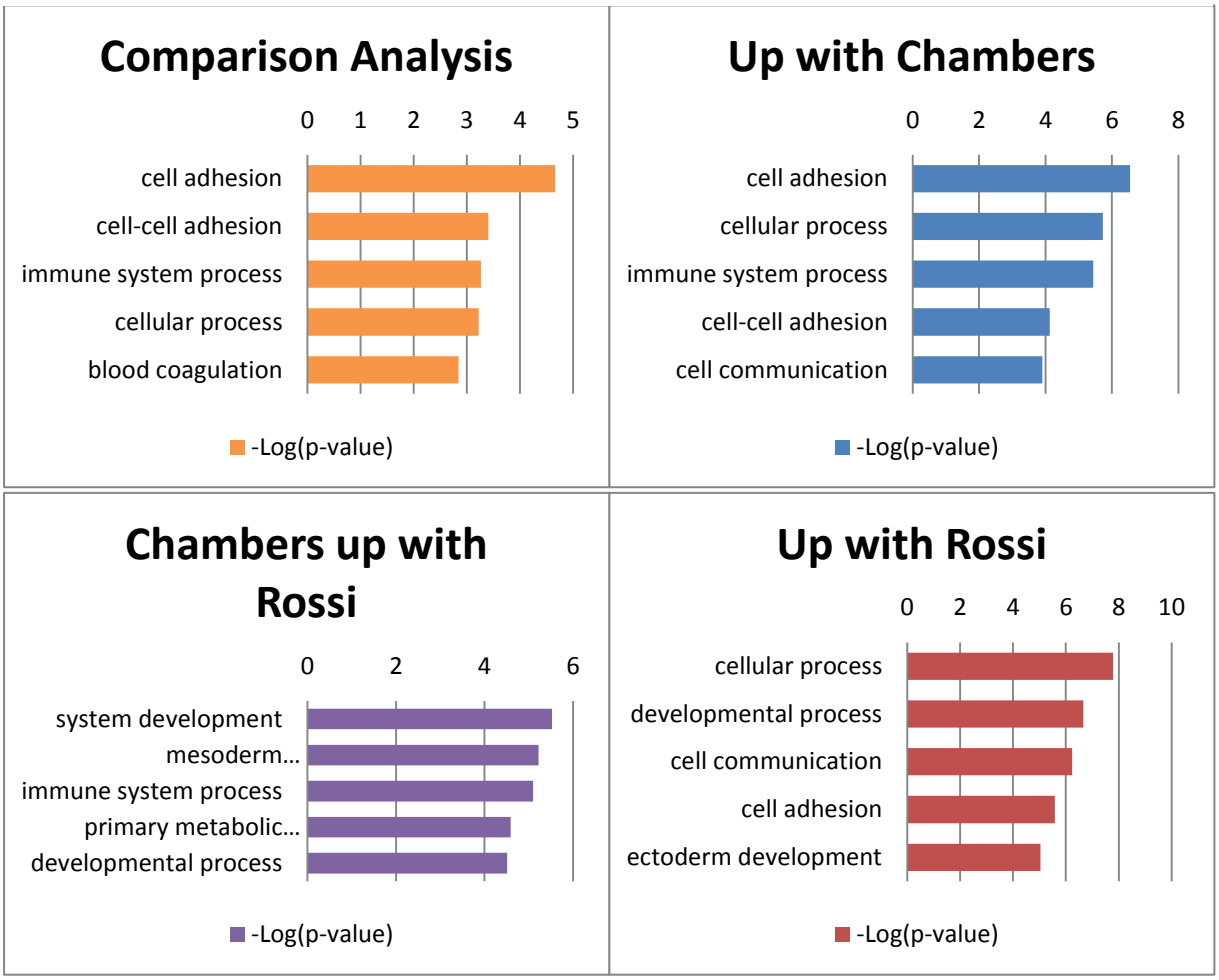


Figure 5-5: Comparative Analysis of the ageing HSC transcriptome.
The blue circle represents data from Chambers et al. [99]; the green represents that of Rossi et al. [94], while the red represent data from this study. Analysis was performed by identifying genes that were represented in common to the studies compared.

From the comparison analysis, I identified 100 age-upregulated genes that were shared between this study and that of Chambers et al., as well as 81 that were upregulated in common between the current study and Rossi et al. Of these, 34 genes were common to all three studies and I identified the top age upregulated biological processes by performing gene ontology (PANTHER) analysis of these genes ('Comparison Analysis'). I also investigated the biological processes common to this study and 1) Chambers et al. ('up with Chambers') and 2) Rossi et al. ('up with Rossi'). Finally, I determined the process common to the Chambers and Rossi studies ('Chambers up with Rossi'). The results of these analyses are displayed in Figure 5.6 and show that cell adhesion is the most significantly upregulated process that is common to all three data sets (Fig. 5.6, 'Comparison analysis'). However this was not true for the Chambers and Rossi comparison.

Figure 5-6: Biological processes that are common between multiple studies of HSC ageing.



Genes that were identified as significantly upregulated in common, during HSC ageing between the specified studies, were analysed for enriched biological processes, using the PANTHER program. Top 5 enriched processes are shown.

5.6 **Conclusion**

In this Chapter, I generated and analysed the transcriptomes of ageing LT-HSCs by RNA sequencing of cDNA obtained from LT-HSCs that were isolated from mice at three different age points. To my knowledge, these are the first RNA-seq data on ageing LT-HSCs. I observed a high correlation between all RNA-seq data sets, which represents a statistical indication of the high quality and reproducibility of the data generated in this study. Several age-differentially expressed genes were identified and I performed gene ontology analysis on these, revealing significant changes in key biological processes at specific stages during HSC ageing. This includes a downregulation of cell proliferation between young- and middle age as well as the upregulation of cell movement between middle- and old age. Overall i.e. up- and downregulated in Old compared to Young, I observed a highly significant change in the expression of several genes that are involved in cell movement. I also determined that the genes that were upregulated in common in this study and previous studies of HSC ageing are enriched in genes involved in cell adhesion. These findings indicate that several genes and key processes, involved in the interaction between HSCs and their bone marrow niche, are altered with age.

5.7 **Discussion**

I have identified several novel changes and provided insights into the dynamics of gene expression during HSCs ageing. Adult HSCs are known to be quiescent, with more than 90% of all HSCs in the bone marrow at the G0 stage of the cell cycle [41]. This is in stark contrast to FL haematopoiesis where HSCs are highly proliferative [29,41]. Although the above is a well-established fact in haematopoiesis, the exact stage and key regulators of the switch from foetal to adult haematopoiesis, is currently unclear. Previous studies have attempted to address this question by studying HSC cycling and have found that the switch from proliferative to quiescent HSCs does not occur until 3–4 weeks after birth [184]. However, I and others have found that several adult HSCs, of as old as 7 weeks, retain the FL HSC phenotype i.e. KLS $CD34^{+}$ [12,18]. I also showed that a major change in gene expression from young to middle-age, involved the downregulation of genes that were associated with cell proliferation. This suggests that although HSCs from pre-pubescent young mice are thought to be quiescent, the switch from proliferation to quiescence in HSCs is not instantaneous but involves the progressive downregulation of several genes that promote the proliferative phenotype. The identification of such genes is an important improvement in the understanding of HSC regulation and the mechanisms involved in their switch from a foetal/proliferative phenotype, to quiescence.

Another key difference between foetal and adult haematopoiesis is the niche in which HSCs reside, and adult haematopoiesis is preceded by the migration of HSCs from the foetal liver at approximately 15 days post-conception, to the spleen and finally the bone marrow, where they remain throughout adult life [29]. This transition is orchestrated by several adhesion molecules including selectins, integrins and chemokines, stromal cell-derived factor-1 (Sdf-1, also known as Cxcl12), which have crucial roles in the homing of HSCs to the bone marrow and their retention or lodgement once there [39,191-193]. The bone marrow provides a regulatory micro-environment for adult HSCs and this interaction is necessary for HSC self-renewal and quiescence [34]. It is therefore important that homeostatic conditions are

maintained within the bone marrow for normal haematopoiesis to occur. Cell movement and adhesion to specialized niches are clearly important processes in the regulation of HSCs during ontogeny; it was therefore interesting to find that these processes were significantly altered during ageing. It is tempting to explain these findings simply as another wave of HSC migration to a different niche i.e. away from an aged bone marrow; however, this is unlikely, as the bone marrow is the final known residence of adult HSCs [29], and although, extramedullary haematopoiesis is known to occur in adults, this is pathological [194]. Therefore, a more likely explanation could be the movement of HSC to a more supportive niche, within the bone marrow, during ageing. Indeed, It is thought that multiple niches exist within the bone marrow i.e. endosteal and vascular niche [14,33,182]. However, the exact nature of these niches and their exact physiological relationship with HSCs is highly debated [39]. Nevertheless, it is possible that intrinsic changes in aged HSCs cause them to move to a different niche within the bone marrow, which in turn alters their differentiation potential, resulting in increased quantity and myeloid bias, as seen in HSCs isolated from old mice. Interestingly, a study of haematopoietic stem and progenitor cell (HSPC) movement found that aged HSPCs showed increased motility and were found further away from the endosteum than their young counterparts [195]. These findings, and the change in the expression of cell movement and adhesion genes observed in this study, strongly suggest that altered cell movement is involved in HSC ageing.

Chapter 6

Methylomic Analysis of Ageing LT-HSCs

6.1 **Introduction**

I showed in Chapter 4 that the DNAm profile of HSCs is altered with age. I identified a 5% reduction in global DNAm levels and found several genes to be differentially methylated with age. As discussed in Chapter 1, DNAm is involved in gene regulation, and changes in DNAm result in altered gene expression, and compromised cell fate [51]. In line with previous findings [88,90,151], age-related changes in the DNAm of HSCs were accompanied by significant changes in gene expression (Chapter 5). DNAm regulates gene expression by various mechanisms, several of which are only beginning to be uncovered. Some of these mechanisms, such as epigenetic modification of promoter regions, are well established and known to have a major impact on gene expression, while the role of DNAm at other non-canonical regulatory regions, for instance, splice sites, CGI shores and non-coding intergenic regions, are just beginning to be elucidated [80,172,196]. The role of DNAm in gene regulation was highlighted in Chapter 1, and will not be discussed here; however, in order to understand the functional consequence of the changes in DNAm that were uncovered during HSC ageing, it was necessary to integrate the methylomes and transcriptomes of ageing HSCs. A key strength of this thesis is its study design; HSC DNAm and gene expression analysis were conducted on DNA and RNA from the same primary and functionally relevant cells. As a result, it was possible to assess the direct effects of aDMRs on gene expression. In this Chapter, I report the methylomics analysis of LT-HSC ageing, which was performed with the following aims;

1. Identify the relationship between global DNAm and gene expression in HSCs
2. Elucidate the dynamics of this relationship during ageing
3. Assess the effects of aDMRs on differential expression.

6.2 **Methylomic Analysis of Ageing LT-HSCs**

In order to understand the relationship between global DNAm and HSC gene expression, 50,601 annotated Ensembl transcripts, that were covered by the RNA-seq of HSCs in Chapter 5, were grouped according to four expression levels (0-25, 25-50, 50-75 and 75-100%). DNAm for these transcripts were then determined by the MEDIPS algorithm and aggregated over canonical gene features [197].

The analysis was done in two parts. Firstly, DNAm at putative promoter regions (-5kb upstream, transcription start site (TSS) and 1st exon) was analysed and pooled for each of the four gene expression levels, revealing a significant (Kruskal-Wallis, p-value < 0.001), step-wise decrease in promoter methylation with increasing gene expression, across all three age groups (Fig. 6.1). This indicates a negative correlation between promoter methylation and gene expression in LT-HSCs, and shows that this relationship is not altered during ageing.

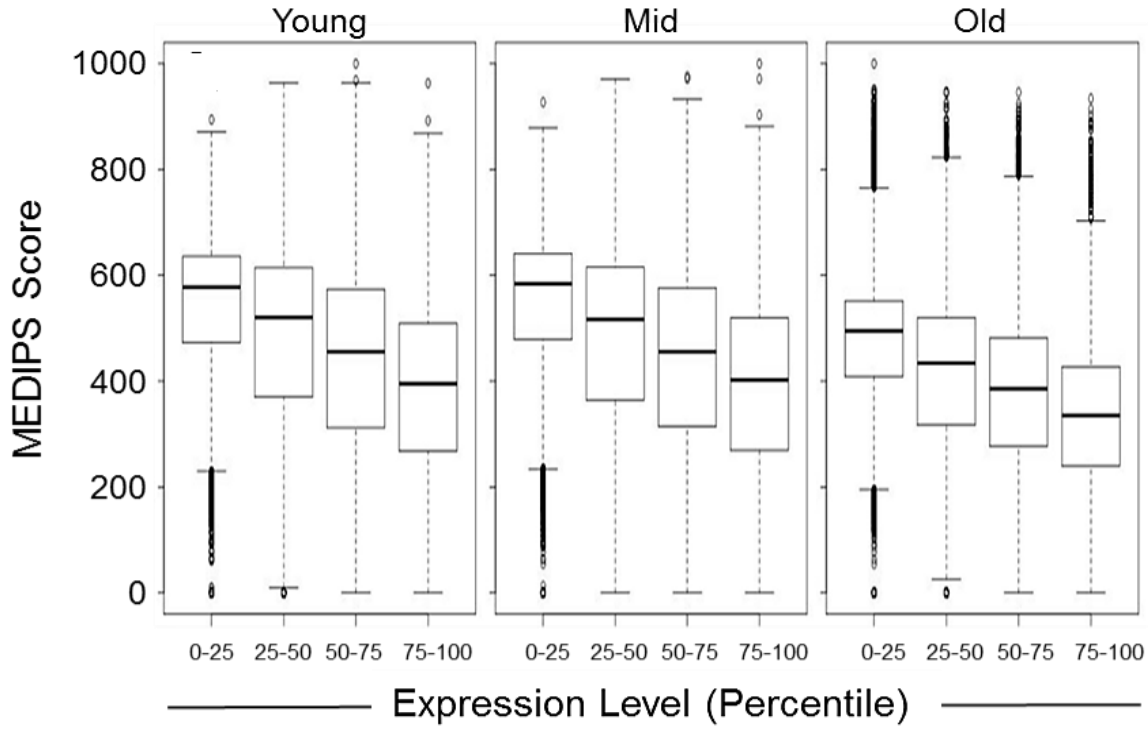


Figure 6-1: The relationship between DNA methylation and gene expression.
Box plot showing the negative correlation between DNAm and gene expression in HSCs, at different age points. MEDIPS score is proportional to DNAm levels.

As DNAm in non-promoter regions is also known to influence gene expression [60], it was important to assess the relationship between DNAm and global gene features, such as 3' and 5' UTRs, intragenic, intergenic, as well as introns and exons. To achieve this, HSC transcripts at two extreme mean expression levels (Low = 0-25% and High = 75-100%), were assessed for DNAm, over nine canonical gene features defined by Li et al [197]. Whereas, the negative correlation between promoter methylation and gene expression was maintained with age (Fig. 6.1), analysis of global gene features revealed significant (Kruskal-Wallis, p-value < 0.001) age-related alterations in the relationship between DNAm and HSC transcripts (Fig. 6.2). This analysis showed that while the DNAm profile of HSC transcripts were largely similar in Young and Mid HSCs; this profile was markedly altered in Old HSCs. The most prominent alteration shown in Figure 6.2 is the global decrease in DNAm that was identified in Chapter 4; however, closer examination of each gene features revealed a number of notable differences. For instance, the clear demarcation between the first exon, first intron and main exons, at positions 4–6 (arrowed), was lost in old HSCs. The flat-lining effect at these positions in Old samples is mostly attributable to loss of DNAm in the first intron (Fig. 6.2, position 5). Additionally, a loss in DNAm (ΔDNAm) at the 5–10 kb downstream positions (3'-untranslated region (UTR) can be observed. These findings show that loss of DNAm at gene bodies and intergenic regions occurs during HSC ageing and may be involved in this process. Likewise, the observation that DNAm at promoter regions was the only distinguishing feature between low and high expressed HSC genes, and the fact that this relationship is strictly maintained with age, confirms the importance of DNAm at this region and in HSC gene regulation.

6.3 **The Role of DNA methylation in Age Differential Gene Expression**

I examined the functional consequence of the aDMRs identified in Chapter 4, by investigating the correlation between age-dependant differential methylation and age-dependent differential expression. 13.5% of age differentially methylated genes (aDMG) were also found to be differentially expressed (Table. 6.1, Table 6.2). Hypo-aDMRs were

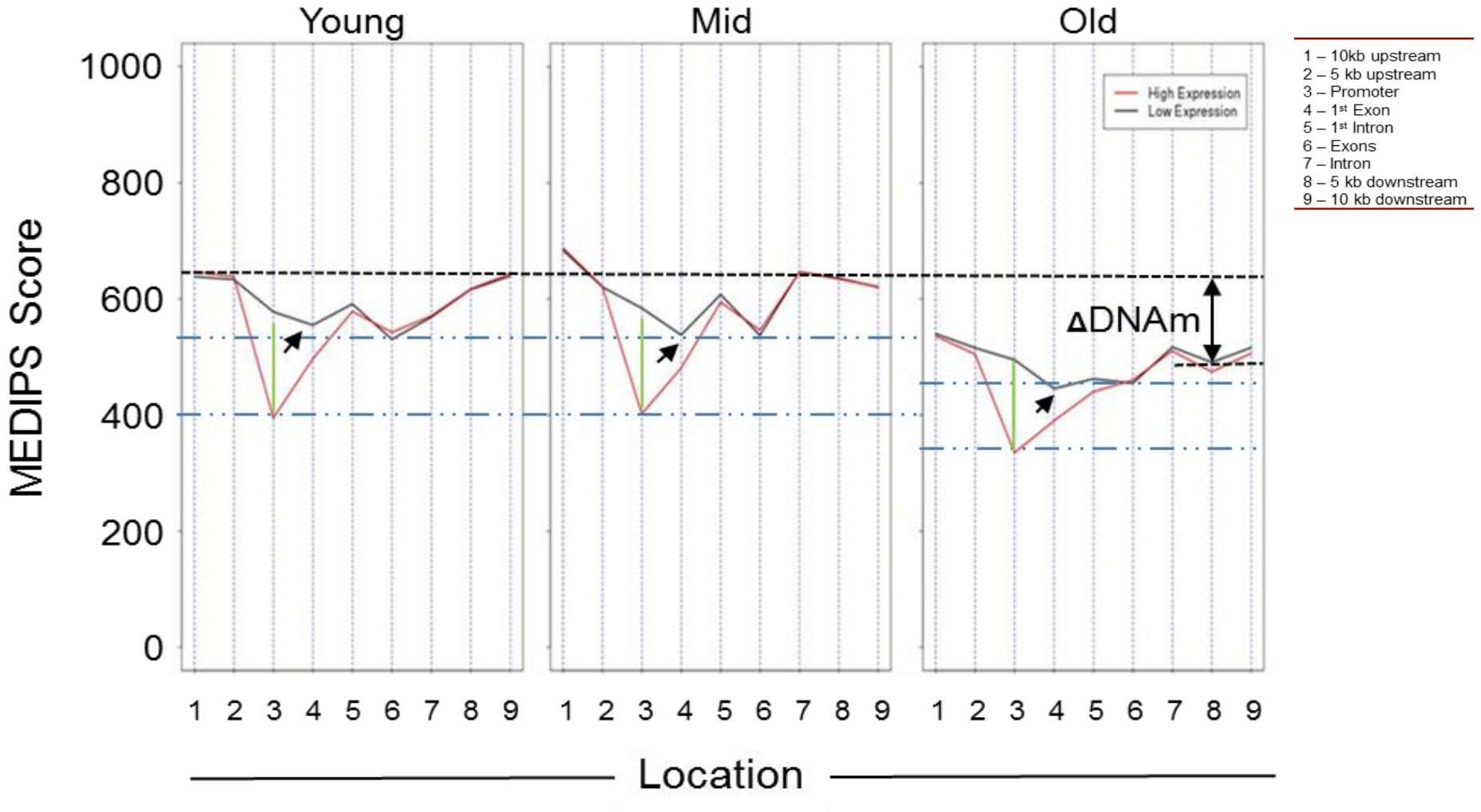


Figure 6-2: Genome-wide DNA methylation profiles of high and low expressed genes.
DNAm was aggregated and averaged for 9 canonical gene features for high (red line) and low (black line) expressed genes in Young, Mid and Old LT-HSCs. 3'UTR (locations 8 and 9) shows the most prominent drop in DNA methylation with age (ΔDNAm) while the negative correlation between DNAm levels and gene expression is maintained during ageing (green line). Loss of gene body methylation is indicated by single arrow heads.

.

more significantly associated with altered gene expression (p-value < 0.0001) than hyper-aDMRs (hyper-aDMR p-value = 0.04690). Additionally, 80% of differentially expressed aDMGs (DE-aDMGs) were upregulated, which is consistent with the global upregulation of genes identified in Chapter 5. These findings also suggest that loss of DNAm and increased gene expression are the predominant changes underlying ageing of HSCs.

Table 6-1: Summary of Methylomic Analysis Data

Analysis	Total Increased with age	Total Decreased with age	Total Observed with age
Age Differential Methylation	71	40	111
Age Differential Expression	860	430	1290
Age Differentially expressed aDMGs			14

Table shows a summary of the aDMR or age differentially expressed transcript counts from; methylation, gene expression and integrated (methylomics) analysis performed in this thesis. Change is from Young to Old.

Table 6-2: Differentially expressed aDMGs

aDMG	Log2 FC
Kiss1r	1.83
Nav2	2.21
Nav1	1.52
Rusc2	1.88
Hsf4	-5.55
Amotl2	1.33
Trim30a	-4.71
Mrc1	-2.46
Pfkfb3	1.13
Itgb3	2.33
Wdfy1	2.38
Nav2	2.21
Slco2a1	1.56
Sdpr	1.98
Pde8a	0.98

Blue Shading represents Hypermethylation and Yellow, Hypomethylation. Decreasing colour intensity depicts increasing FDR. Log2 FC is Log2 fold change in gene expression. Change is from Young to Old

6.4 **Mechanisms of HSC ageing**

The DE-aDMGs that were identified and listed in Table 6.2, were too few for gene ontology analysis, however, bearing in mind that cell movement was the most significantly altered process during HSC ageing (Chapter 5), I assessed the frequency of DE-aDMGs that were involved in cell movement and found that at least 40% of these genes are involved in cellular migration, which is a cell movement process [162,166,198-201]. These include *Itgb3, Nav1, Nav2, Amotl2, Kiss1r* and *Pde8a,* and suggest that age-related changes in DNAm are involved in the dysregulation of HSC cell movement genes with age.

6.4.1 **SDPR and HSC ageing**

As well as the cell movement genes highlighted above, *Sdpr* is another gene that was found to be differentially expressed and methylated with age. *Sdpr*, which is also known as *Cavin-2*, was originally identified as a gene whose expression was increased in serum-starved cells [167]; however, an independent role in caveolae formation and cell signalling has since emerged [168]. Caveolae are invaginated lipid rafts that provide a common platform for various signalling molecules, leading to enhanced cell signalling and potent signal transduction. Over-expression of *Sdpr* in cultured cells leads to caveolae elongation [168] and perhaps results in an increased surface area for interacting cell signalling molecules. Furthermore, previous studies of HSC ageing have identified a significant upregulation of *Sdpr* gene expression with age [94,99], suggesting that this gene is involved in HSC ageing.

I identified a progressive loss of promoter DNAm in the *Sdpr* gene with age (Fig. 6.3), and in agreement with findings from the integrated methylome/transcriptome analysis, described in Section 6.2; this reduction in promoter methylation was accompanied by a significant (FDR < 5E-09) age-dependent upregulation of the *Sdpr* gene. In addition, I found the Sdpr promoter aDMR to be conserved in several mammals and between other CAVIN family members (Fig. 6.3). These suggest that this aDMR is important for SDPR's involvement in caveolae biogenesis and may be involved in HSC ageing.

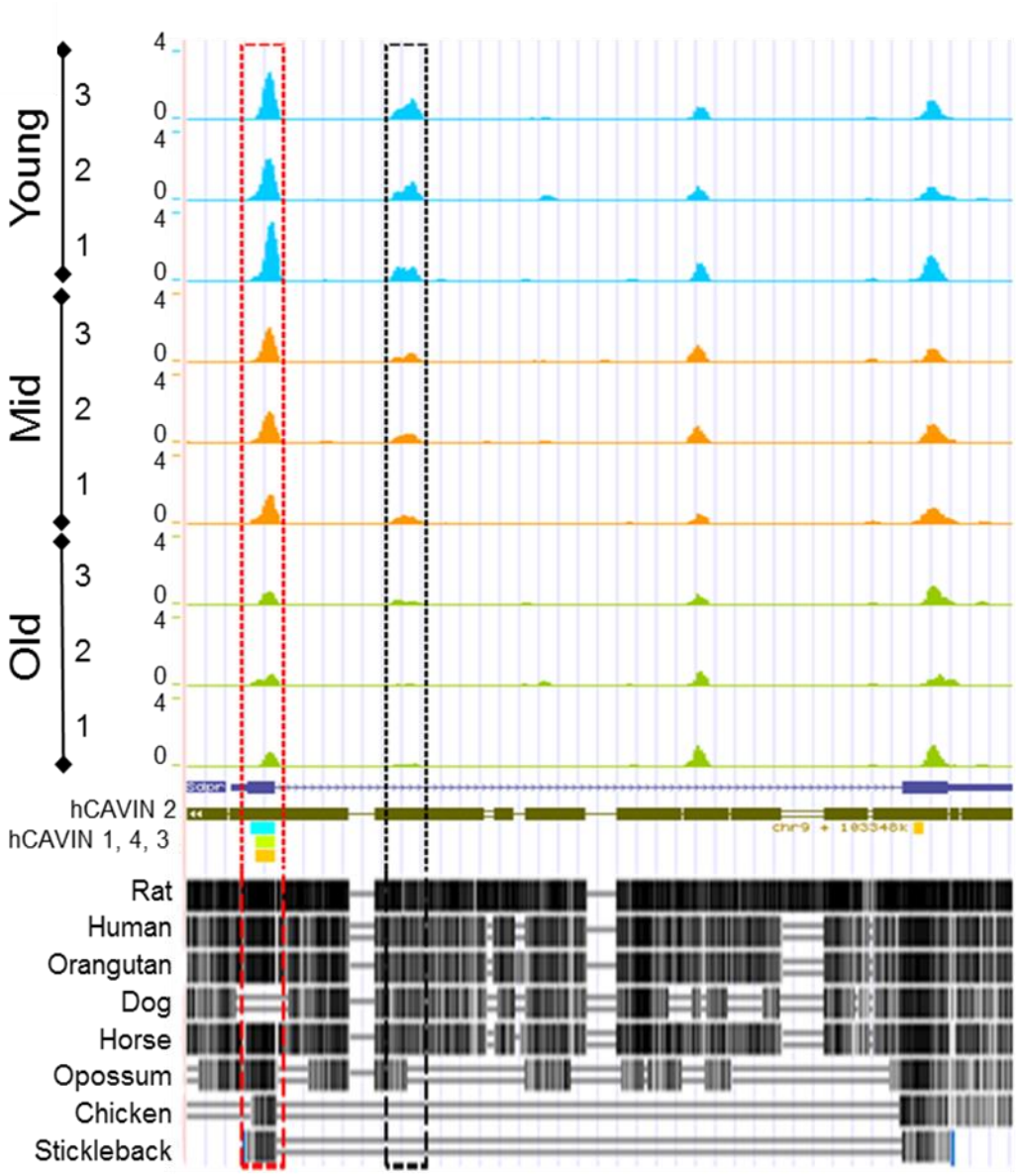


Figure 6-3: UCSC genome browser screenshot of MeDIP-seq profile of the Sdpr gene in ageing HSCs.
Tracks are colour-coded blue, orange and green for Young, Mid and Old samples, respectively. Three biological replicates, each consisting of pools of 5-30 mice, are shown for each age group. Peak height represents DNA methylation levels. The two aDMRs identified in the *Sdpr* gene are boxed in red and black. The red box depicts the aDMR that aligns with other members of the human CAVIN gene family and is highly conserved from human to fish (except dog) as indicated.

6.5 Validating the Role of DNAm in Age-dependent Differential Expression of the *Sdpr* Gene

I performed bisulfite pyro-sequencing of the *Sdpr* gene promoter aDMR (Fig. 6.3, red box), as a technical validation of MeDIP-seq, and to confirm the mechanism of age-dependent

upregulation of the *Sdpr* gene (upregulation of *Sdpr* was previously validated in Chapter 5, by qRT-PCR).

6.5.1 **Bisulfite Pyro-Sequencing of Low Amount of Cells**

Because of the harsh nature of the bisulfite conversion process, which causes DNA to degrade, large amounts of DNA are often required as input. This ensures that there is enough DNA left for PCR and pyro-sequencing reactions. The amount of LT-HSCs that were available for this thesis was limited; as a result, I used the Qiagen EpiTect Plus LyseAll Bisulfite Kit (EpiTect kit) for bisulfite conversion of DNA directly from cells. This helped to minimize DNA loss by eliminating the extra DNA extraction step. Additionally, the EpiTect kit is optimized to minimize degradation of DNA, during bisulfite conversion and thus allows the bisulfite conversion of low number of cells. As bisulfite conversion of DNA directly from cells was not a standardized or well established method, I validated its use by comparing its conversion efficiency with that of bisulfite conversion from equivalent amounts of DNA. To this end, I bisulfite converted $1x10^4$ cells and 50 ng of DNA in separate reactions, using the EpiTect kit, and compared conversion efficiencies by qPCR (materials and methods). This was performed for three technical replicates. As shown in Figure 6.4, bisulfite conversion efficiencies were remarkably high and comparable for both cells and DNA, thus validating the method of direct bisulfite conversion of DNA from cell lysates.

*6.5.1.1 **Bisulfite Pyro-sequencing Assay Validation***

Next, I designed (materials and methods) and validated the *Sdpr* promoter BS-Pyroseq assay. This involves the validation of BS-PCR primer sets (Table. A-1), to ensure that primers are able to amplify DNA of various methylation levels, with equal efficiency. There are multiple steps involved in this process. These include:

1. Construction and validation of *in vitro* methylated DNA that will be used to test BS-PCR primers.

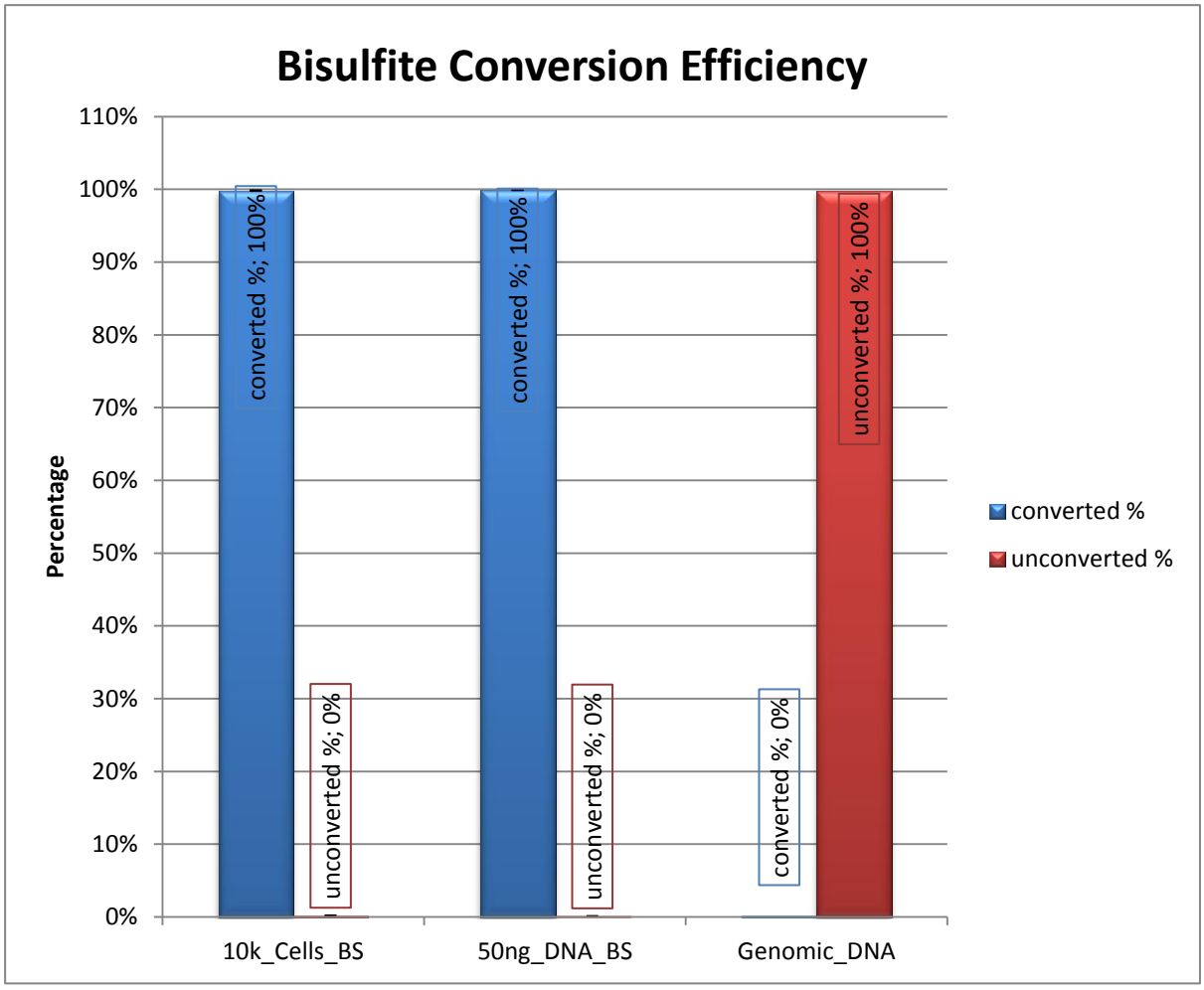


Figure 6-4: Validation of bisulfite conversion of DNA directly from cells lysates.
Plot shows that bisulfite conversion from cells (10k_cells_BS) is as efficient and reliable as from DNA (50ng_DNA_BS). Blue bars show the fraction of converted DNA over unconverted DNA (red bar). Genomic DNA that had not been treated with sodium bisulfite was used as control (Genomic DNA). Three replicates of 10000 cells and 50 ng DNA and one replicate of genomic DNA are shown. Error bars represent standard error of the mean.

2. PCR amplification of *in vitro* methylated DNA, over a range of methylation levels i.e.100%, 75%, 50%, 25% and 0%.
3. Pyro-sequencing of amplified products to ascertain assay linearity.

The Sdpr assay was tested by following the outlined steps. I *in vitro* methylated mouse genomic DNA as described in materials and methods, and verified that this was indeed methylated. Verification was achieved by digesting test DNA with methylation sensitive restriction enzymes (materials and methods). Figure 6.5 shows that the *in vitro* methylated DNA that was used to validate the *Sdpr* assay had been successfully methylated. This was

inferred from the presence of a single undigested DNA band for *in vitro* methylated DNA, compared to genomic DNA, which was digested and therefore appeared as a smear when electrophoresed.

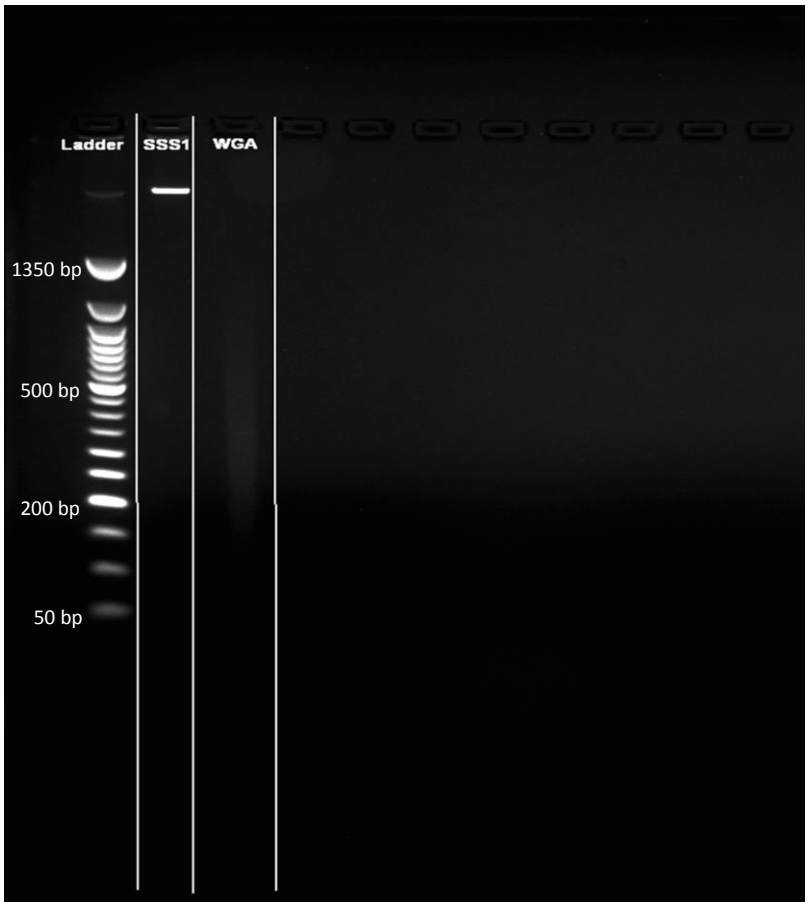


Figure 6-5: *In vitro* DNA methylation.
The 'SSS1' lane shows a single band of undigested high molecular weight DNA after treatment with methylation sensitive (MS) restriction enzymes. The 'WGA' lane contains DNA that has been completely demethylated (by whole genome amplification) and has therefore been digested by MS-restriction enzymes (smear). Far left lane shows a 50 bp size ladder.

Following validation of *in vitro* methylated DNA, I constructed 100%, 75%, 50%, 25% and 0% methylated DNA, by diluting the *in vitro* methylated with an appropriate amount of whole genome amplified, unmethylated DNA. These were bisulfite converted and used to test the *Sdpr* BS-PCR primers. Figure 6.6 shows that this primer set was able to specifically (no primer dimers or unwanted products) and consistently amplify the Sdpr promoter aDMR, regardless of methylation level.

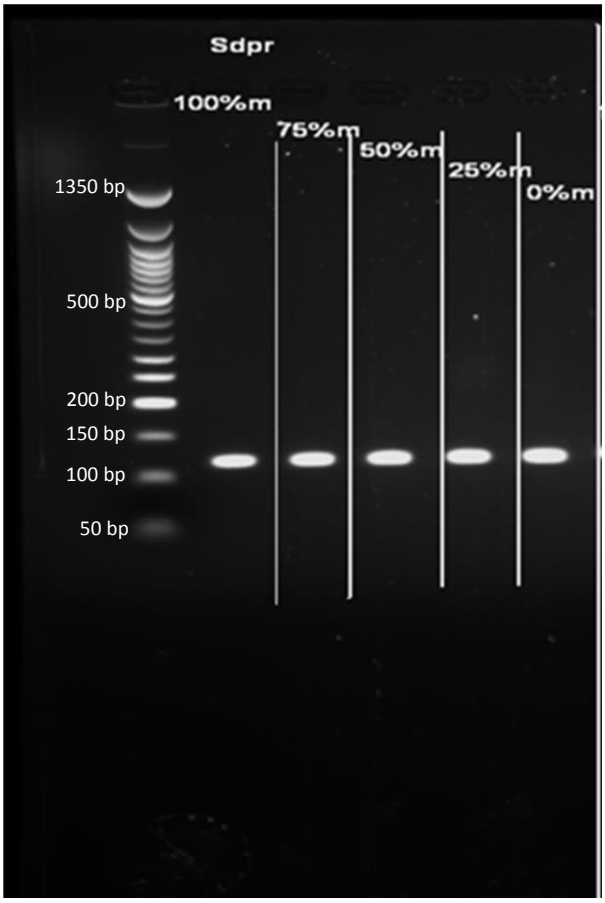


Figure 6-6: Validation of the Sdpr BS-PCR primers.
These primers show efficient amplification of DNA, irrespective of methylation status. Outer left lane shows the 50 bp ladder. The Sdpr amplicon is expected to be 118 bp and the bands seen in all sample wells, appear to be about this size.

Finally, I pyro-sequenced the amplified BS-converted DNA in order to test the linearity of the *Sdpr* assay, over the aforementioned methylation range. Figure 6.7 shows that the *Sdpr* assay is sensitive enough to discern the different levels of DNAm, and can therefore be used to validate the age-dependent *Sdpr* promoter hypomethylation that was detected by MeDIP-seq.

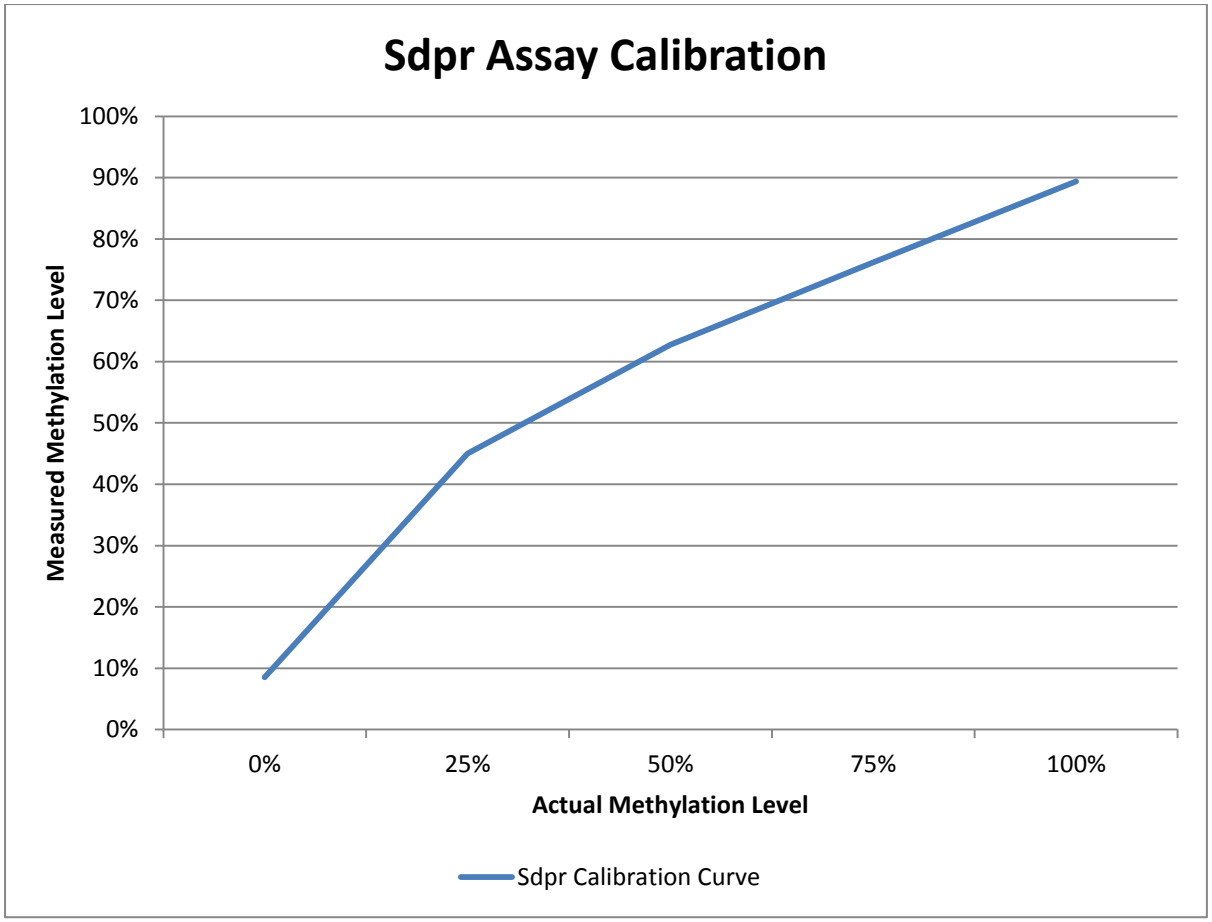


Figure 6-7: Validation of the Sdpr BS-Pyroseq assay.
Plot shows the range at which the Sdpr assay can detect DNAm levels. *In vitro* methylated and whole genome amplified (WGA) mouse DNA from WBM cells were used for this calibration. DNA of shown 'actual methylation levels' were constructed by appropriate mixing of 100% (*In vitro)* methylated DNA with 100% unmethylated (WGA) DNA, or by using undiluted versions of either, where appropriate.

6.5.2 **Validation of the Sdpr Promoter aDMR**

Following the establishment of the BS-Pyroseq method for use on low amount of cells, I bisulfite converted and pyro-sequenced four biological replicates of Mid and Old HSCs, as well as three biological replicates of Young HSCs. I confirmed by qPCR that the conversion efficiency was at least 99% for all samples (Appendix-Figure 2). These samples were independent of those used in MeDIP-seq reactions and were pyro-sequenced using the *Sdpr* assay (Table. A-1) described in Section 6.5.1.1.

Consistent with MeDIP-seq (Fig. 6.3), bisulfite pyro-sequencing of the *Sdpr* promoter revealed a progressive loss of DNAm at this region with age (Fig. 6.8). Because the *Sdpr*

assay was not 100% accurate in detecting absolute methylation values (Fig. 6.7), I normalised all measured values against the *Sdpr* calibration curve. This was necessary to correct for any measurement bias that may ensue. Nevertheless, similar trends of age-dependent *Sdpr* hypomethylation were observed, regardless of whether the measured or adjusted DNAm values were used (Fig. 6.8).

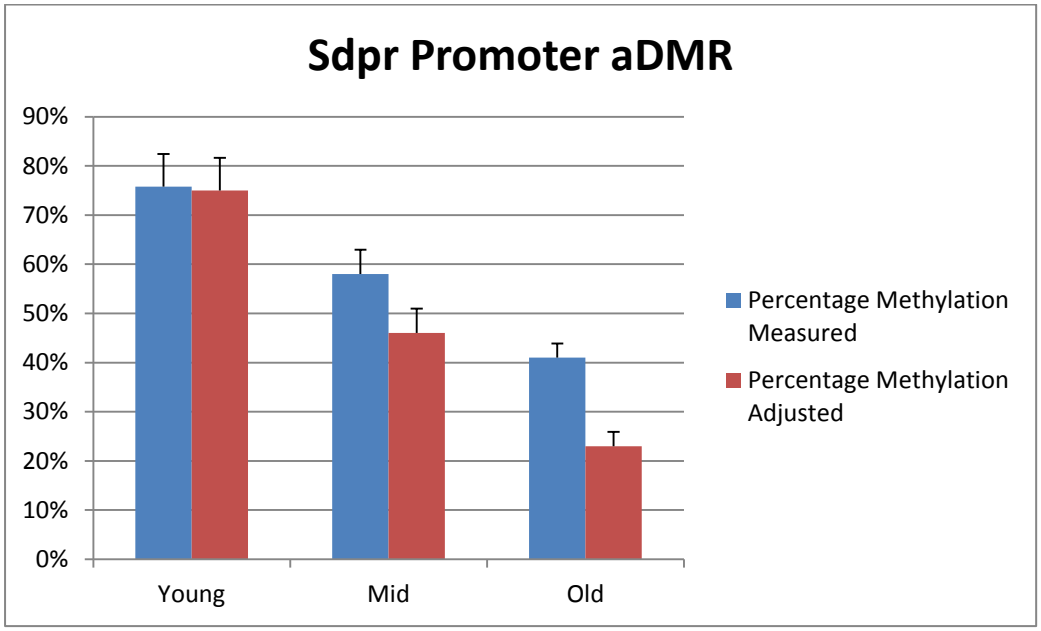


Figure 6-8: Validation of progressive Sdpr promoter hypomethylation with age, using BS-Pyro-seq.
Measured values (blue) and values normalized to the Sdpr calibration curve (red) are shown. Error bars represent standard error of the mean. 3 biological replicates of Young and 4 biological replicates of Mid and Old are shown.

6.5.3 **Validating the Link between DNAm and Gene Expression**

DNAm is known to regulate gene expression (Section 1.7.2), and I showed in Section 6.2 that promoter DNAm was negatively correlated with the gene expression of HSCs, irrespective of biological age. In this section, I sought to validate the link between promoter methylation and altered gene expression. As a result, I *ex vivo* cultured FACS purified LSPs, and treated them with 5-Aza'dC, to induce global demethylation. 5-Aza'dC is a methylation resistant analogue of cytosine, which can be incorporated into DNA to induce passive demethylation, following multiple rounds of cell division [202] (Fig. 6.9). To ensure that cultured LSPs divided sufficient times to allow 5-Aza'dC incorporation and demethylation, I sought to optimize the *ex vivo* expansion assay, to allow maximum self-renewal, without

inducing differentiation. I cultured LSPs *ex vivo* for five days and visually assessed cell growth, by culture well coverage, and primitiveness, by cell surface marker phenotyping, at days three and five. Three day *ex vivo* culture of LSPs resulted in insufficient amount of cells (Fig. 6.10a, left), while five day culture resulted in adequate expansion (Fig. 6.10a, right) but yielded less phenotypically defined HSCs.

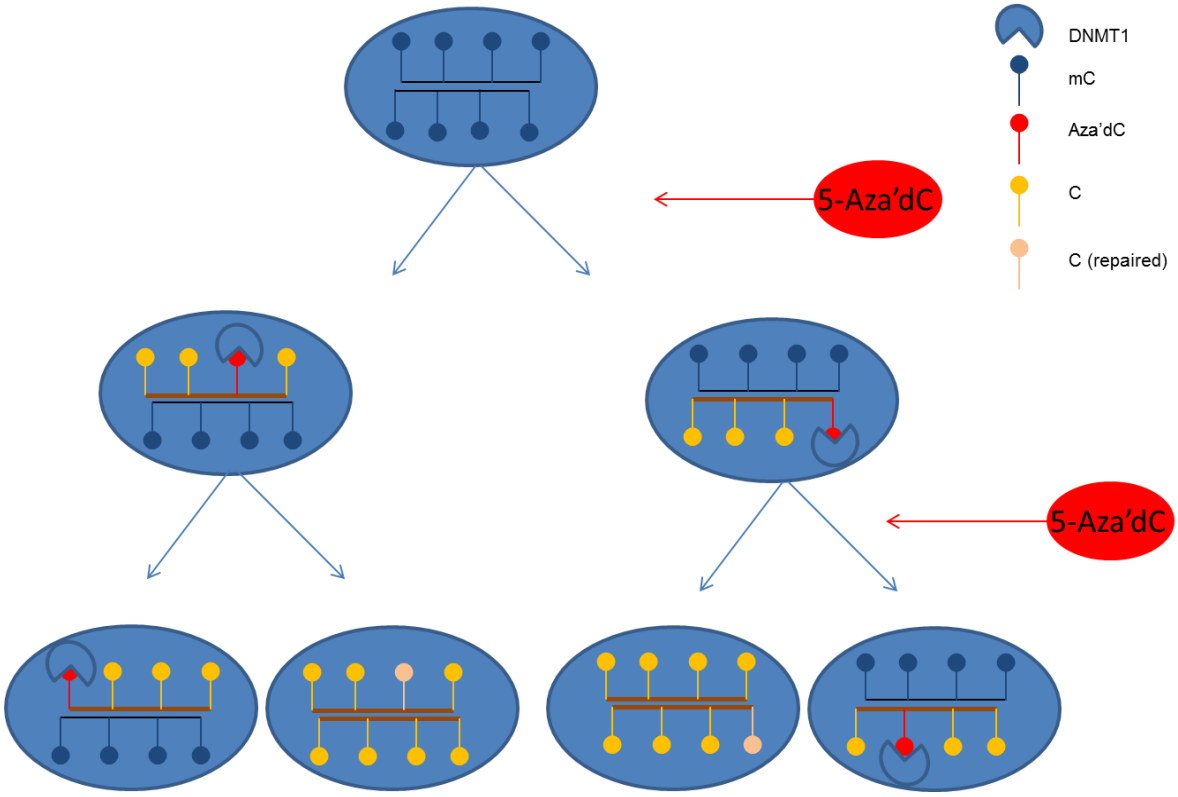


Figure 6-9: Ex vivo DNA demethylation by 5-Aza'dC treatment.
5-Aza'dC is incorporated during cell division and DNA replication. DNMT1 binds irreversibly to 5-Aza'dC and limits its availability for the maintenance of DNAm at hemi-methylated sites, leading to passive DNAm.

(Fig. 6.10b). Primitiveness was measured by the phenotyping of HSC cell surface antigens using fluorescent antibodies and flow cytometry. KLS antibodies were used to assess the presence of HSPCs, while this in combination with Slam markers (CD48 and CD150) were used to measure primitive HSCs i.e. LT-HSCs. Based on these criteria, I observed more primitive HSCs at day three, regardless of the marker combinations that were assessed. Additionally Old HSCs appeared to maintain their LT-HSC phenotype better than Young

HSCs, during *ex vivo* culture (Fig. 6.10b). Flow cytometry plots for the data displayed in Figure 6.10b are shown in Appendix-Figure 3.

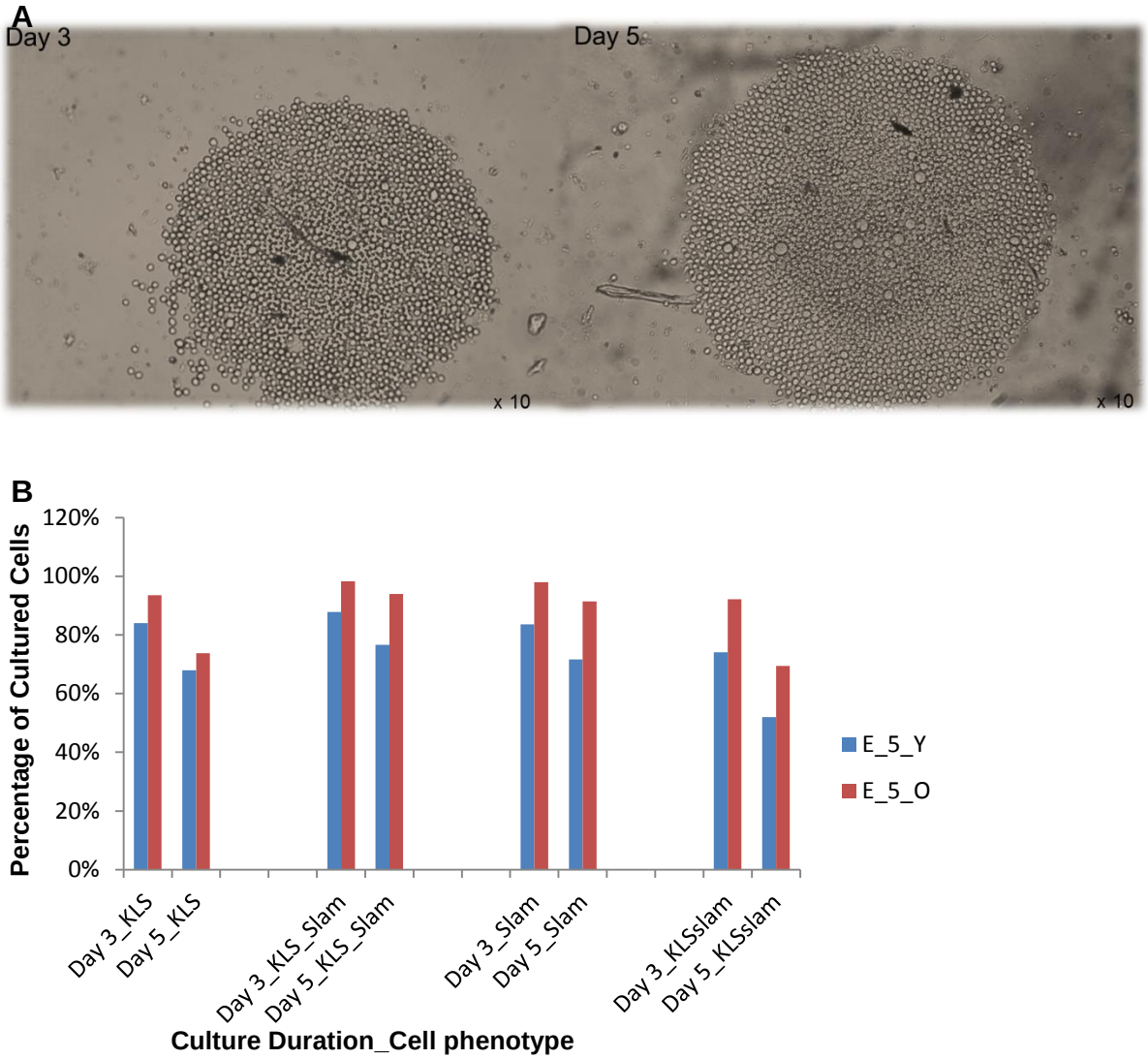


Figure 6-10: Ex vivo culture of HSCs.
A) Appearance of LSP cells after three or five days *ex vivo* culture. More cells can be seen in the culture well at day 5. B) Phenotype of *ex vivo* cultured Young (E_Y) and Old (E_O) LSP cells. KLS antibodies are as previously described and measures HSPCs, KLS_Slam shows the frequency of KLS cells that are CD48- and CD150+, Slam measures these CD markers without KLS. while KLS Slam shows the overall frequency of cells that are KLS. $CD48^-$ and $CD150^+$ i.e. LT-HSCs

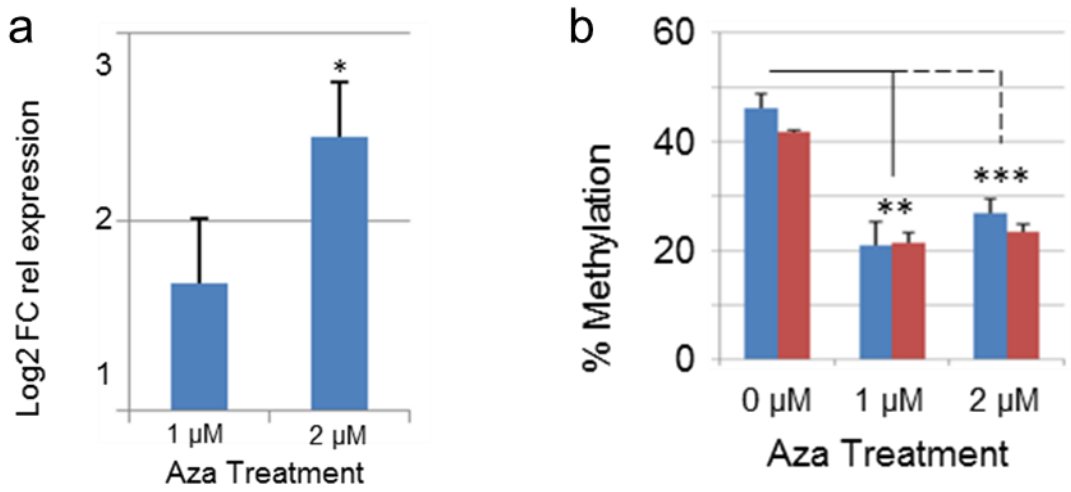
These results indicated that intermediate culture duration between three and five days was required for optimal phenotypically defined LT-HSCs expansion, I therefore *ex vivo* cultured FACS purified LSP cells for a maximum of four days. Subsequently, I treated LT-HSCs with a range of 5-Aza'dC concentrations and found that 1–2 µM was the optimum dose range

that was required to induce demethylation with minimal cytotoxicity in these cells. Morphological examination and antibody phenotying of treated cells showed that LT-HSC integrity had not been markedly compromised by the treatment (Appendix-Figure 4).

6.5.3.1 ***Sdpr Gene Regulation***

I assessed the expression of *Sdpr* in 5-Aza'dC treated LT-HSCs and observed a dose-dependent up-regulation of its gene expression (Fig. 6.11a, p-value = 0.028 for 2 µM treatment). Additionally, bisulfite pyro-sequencing of LT-HSCs revealed *Sdpr* promoter hypomethylation following 5-Aza'dC treatment (Fig. 6.12b, p-value < 0.005). These findings are consistent with the age-dependent difference in *Sdpr* methylation and expression observed in Young and Old LT-HSCs and suggests that age-related up-regulation of *Sdpr* is due to DNA hypomethylation.

Figure 6-11: Functional analysis of the Sdpr promoter aDMR.



Analysis was by ex-vivo culture of LT-HSCs in the presence of 5-Aza'dC (Aza) and (a) expression analysis using qRT-PCR. Data was obtained from independent experiments for a total of four biological replicates. Expression values shown are relative to *ex vivo* cultured, untreated control. (b) Validation of same Sdpr aDMR for promoter hypomethylation following ex-vivo culturing in the presence of 5-Aza'dC (Aza). Data is shown for two biological replicates (red and blue bars). 2-3 middle-aged mice were pooled per replicate and average values for two treatment replicates are shown for each biological replicate. *Sdpr* DNAm level for freshly isolated (uncultured) LSP cells from Mid mice is shown in Fig. 6.8. A lower methylation level is expected in the cultured equivalents, as these cells have undergone more cell divisions, and are therefore of a higher replicative age. FC is fold change. Error bars depict standard error of the mean (S.E.M) in all cases (*P < 0.05; **P < 0.005; ***P < 0.0005).

6.6 **Conclusion**

I performed the first known methylomics (integrated gene expression and DNAm) analysis of primary HSC ageing and began to investigate the regulation of gene expression by DNAm, during mammalian ageing. Consistent with several other studies [60,82,83], I found that DNAm, at gene promoters, is negatively correlated with gene expression in HSCs and showed that this relationship is stable with age. Age-related changes in DNAm at non-canonical regulatory gene features were also observed. Integrated analysis of aDMGs and age differential expression identified several cell movement genes such as *Itgb3* and *Nav2*, which appear to be age-differentially regulated by DNAm. Finally, I demonstrated a link between loss of promoter methylation and upregulation of *Sdpr* during HSC ageing. The findings in this thesis suggest that dynamic changes in DNA methylation constitute an epigenetic mechanism that possibly contributes to the decline of HSC functionality with age.

6.7 **Discussion**

Several studies have provided evidence for the intrinsic nature of HSC ageing, however, the mechanisms involved in these changes remained unclear. The integrated methylome and transcriptome analysis of purified HSCs that was described this thesis has provided novel insights into the underlying cause of HSC age-related dysregulation. A significant finding is that several genes involved in cell movement, migration and adhesion, are significantly altered in ageing LT-HSCs. Notable examples include up-regulated targets of the Aryl hydrocarbon receptor (AHR) ligand-dependent nuclear receptor (Table. A-5). Additionally, Integrin beta-3 (ITGB3), a key signalling and adhesion molecule, was upregulated by more than two fold in Old HSCs (Table. 6.2) and showed age-dependent hypomethylation at approximately 500bp downstream from at what appears to be a regulatory region (Fig. 4.7D), with binding sites for several HSC transcription factors [203]. *Sdpr* was also found to be differentially regulated with age and I demonstrated that age-dependent up-regulation of this gene is correlated with progressive hypomethylation of the *Sdpr* promoter region. While

age-related upregulation of *Sdpr* in HSCs has previously been observed [94,99], I showed that it is linked to loss of promoter methylation, using an *ex vivo* DNMT-1 inhibition assay for FACS purified HSCs. *Sdpr*, which is also known as *Cavin-2*, is involved in caveolae formation and cell signalling [168]. Interestingly, two out of the top five upregulated gene networks in this study involved cell-to-cell signalling and interaction. It is possible that age-dependent upregulation of *Sdpr* is involved in this signalling process. ITGB3 is a member of the integrin family and integrins are known to form supramolecular complexes, which direct potent signal transduction cascades and metalloproteinases-mediated cellular migration [180]. Lipid rafts, such as the caveolae, associate with supramolecular complexes and may regulate their formation. It is possible that the age-dependent upregulation of *Sdpr* has an effect on caveolae formation, which alters the dynamics of supramolecular complexes and their role in signal transduction and cell migration. As well as *Itgb3*, several members of the integrin family were also found to be significantly up-regulated during HSC ageing in this study.

The DNAm and gene expression data from this study provides a probable explanation for the functional decline in HSCs that has been reported during ageing [94,144]. Aberrant up-regulation of genes like *Sdpr* and *Itgb3* and their subsequent role in abnormal cell signalling, migration and adhesion could disrupt the ability of HSCs to interact appropriately with and/or be regulated by the bone marrow micro-environment. The ability of HSCs to respond appropriately to niche-associated signalling is crucial for processes like homing, migration and lodgement. Aged HSCs show reduced ability to home to the bone marrow [7,96,97], which could be due to reduced ability to respond to regulatory cues within the micro-environment. Additionally, *in vivo* imaging of mouse HSPCs revealed that aged HSPCs are found at distances further away from the endosteal niche, compared with younger HSPCs [195], suggesting a reduced affinity for that niche with age. Other age-related defects such as reduced lymphopoiesis and expansion of My-bi HSCs could be a result of sub-optimal

response to signals required for normal lymphopoiesis to occur. Indeed, *Dnmt1* mutant HSCs show global DNA hypomethylation and are unable to sustain lymphopoiesis [88].

Taken together, the data presented in this thesis support a model where a shift from homeostatic levels of DNAm in ageing LT-HSCs has a wide-spread effect on transcription. This could be a consequence of altered transcription factor binding to epigenetically modified target sites, culminating in abnormal interactions of HSCs with their niche, as a result of the dysregulation of key genes involved in HSC migration, homing, adhesion, and signalling. Although it is possible that other biological and epigenetic factors, such as telomere shortening, genetic mutations, histones- and RNA- modifications, are involved in HSC ageing, the observations in this study suggest that global and gene specific changes in DNAm, may contribute to the decline of HSC functionality with age. While it remains possible that the identified changes in DNAm are a consequence, rather that the cause of HSC ageing, and functional validation of these findings remains technically challenging, I was able to demonstrate proof of principle for the *Sdpr*-associated aDMR. Functional assays, i.e. gene knock-down assays, will be required to investigate the relevance of the genes implicated in this study, and to further improve our knowledge of the dynamics of HSC regulation during physiological ageing.

Chapter 7

Discussion and Outlook

7.1 **General Discussion**

The aim of this thesis was to investigate the involvement of DNAm in HSC ageing. This was done by whole genome methylation analysis and transcriptome analysis of mouse primary LT-HSCs. This study was undertaken at a time when methylome and transcriptome analyses of minute samples were not well established. As a result, a significant aspect of this thesis was focused on the development of suitable techniques for the analysis of the rare cells determining HSC ageing. To this end, I developed a protocol for the generation of methylomes from low input DNA [133] and performed methylome and transcriptome analysis on materials obtained from the same cells. The data generated in this study represents a valuable resource for the study of mammalian stem cell regulation and ageing. Additionally, I utilised and optimised various technically challenging assays to functionally validate the findings in this study. I anticipate that these assays will be useful to several other studies in related fields.

The findings in this thesis implicate migration and adhesion genes in the ageing of HSCs. Additionally, changes in the methylation and expression of *Sdpr*, which encodes a caveolae lipid raft protein, was identified and validated. In this chapter, I bring together the key findings of this thesis and discuss their implication in HSC regulation and ageing. I also provide recommendations for future development based on the work presented here.

7.2 **Methylome Analysis**

Nano-MeDIP-seq was developed and used to analyse ageing LT-HSC methylomes in this study. Based on multiple metrics, the data generated is of remarkably high quality (Chapter 3). However, Nano-MeDIP-seq was not compared directly to the original MeDIP method. It is therefore not currently possible to identify the drawbacks of performing MeDIP-seq with lower starting DNA concentrations. It is important that these limitations are known, in order to make informed decisions regarding suitable methylome analysis methods, when designing

future experiments. For example, the CpG enrichment scores observed in the data generated using Nano-MeDIP-seq appeared slightly lower, when compared to existing data from other studies within the 'Medical Genomics' group. A lower CpG enrichment score could be due to various reasons such as: 1) lower global methylation in HSCs compared to these other samples, which are typically cancer cells or tissues; 2) lower global methylation in the C57BL/6 mice compared to human samples, which are typically studied within the Medical Genomics group; 3) artificially lower CpG enrichment in Nano-MeDIP-seq data compared with the original MeDIP-seq method. In order to rule out the third scenario, it will be necessary for future studies to perform MeDIP-seq experiments with both methods, using the same samples. This will allow the direct comparison of methylomes generated by these methods, to determine any significant differences that may exist between them. Nevertheless, CpG coverage for all Nano-MeDIP-seq samples was comparable to that of MeDIP-seq [133], suggesting that Nano-MeDIP-seq is as efficient in detecting DNAm as the original MeDIP-seq method.

Third generation sequencing technologies such as those by Helicos Bioscience [204], Oxford Nanopore Technologies [205] and Pacific Bioscience [206], are capable of single molecule sequencing and are thought to be able to distinguish between cytosine and methyl-cytosine [205,207]. The advantages of these technologies are immediately apparent. Theoretically, only single molecules are required to generate an entire methylome. This eliminates the need for PCR, and removes the amplification bias that comes with it. Additionally the low input requirement of single molecule sequencing technologies allows the study of rare samples, without further optimizations. Although the Nano-MeDIP-seq method developed in this thesis was a significant breakthrough for unbiased methylome analysis, and makes the MeDIP-seq method available to studies where samples of interest are limiting, it is expected that third generation sequencing techniques will replace cumbersome bisulfite and enrichment based methods.

7.3 **MeDIP-seq Data Analysis**

The remarkable advancement in high throughput whole genome analysis technologies, which allow the generation of high resolution methylomes in a relatively short period of time, presents a significant challenge in terms of data analysis. Significant bioinformatics and statistical effort is required to analyse and interpret the enormous data set that is typically generated for methylome or transcriptome studies. As multiple tools and algorithms are required for data processing steps such as sequence alignment, filtering of clonal reads, data normalisation, read counts and differential read counts, it is often necessary to streamline the data analysis processes into discreet computational pipelines. Unfortunately, these packages are typically modelled around studies of gross abnormalities. This means that thresholds are often set too high to detect subtle deviations from the norm, such as those observed during HSC ageing. The MeDUSA pipeline [134] was used to generate HSC methylomes and to determine aDMRs in this study. Initial analysis with this pipeline yielded no differentially methylated regions, despite several obvious differences in sequence read counts between ageing samples. As a result, the threshold for DMR calling was lowered and rules within the pipeline were modified to allow the detection of subtle but significant changes during HSC ageing. This optimization yielded more DMRs, but also increased the incidence of false positive results (visual confirmation on UCSC genome browser). Subsequently, a more conservative version of the DESeq Bioconductor package (v2.10.2) was released and utilised. This version provided the highest level of true DMRs and the lowest incidence of false positives, as determined by visual confirmation of aDMRs on the UCSC genome browser. Nonetheless, the amount of aDMRs that were detected in this thesis remained relatively few, compared to most methylome analysis studies and this low amount of aDMRs appears at odds with the obvious decrease in global DNAm that was observed in aged HSCs. While the former could be attributed to the fact that a homogenous population of FACS purified cells was studied in this thesis, it is possible that the 5% difference that was observed in global DNAm is too small to be detected in discreet regions,

and several true aDMRs were missed, as a result of the low sensitivity of the available analysis tools. The MeDUSA pipeline and other similar analysis tools are constantly being further developed; it will be interesting to investigate the effects of these improvements on the data generated in this thesis.

The MeDIP-seq method can be used to detect other forms of DNAm, such as 5-hydroxymethylcytosine (5-HmC), however, this requires the use of 5-HmC specific antibodies and thus must be assayed separately to 5mC MeDIP reactions. Due to limited time and resources, this mark was not studied here; however, 5-HmC is an important feature of mammalian embryonic [67,68] and somatic stem cells [66]. The role of 5HmC DNAm in HSCs is currently unclear; however Tet2, a member of the Tet protein family, which catalyse the conversion of 5mC to 5HmC *in vivo,* has been shown to be important for haematopoiesis. Tet2 is highly expressed in bone marrow haematopoietic cells [208] and mutations in this gene are associated with various myeloproliferative malignancies [208]. Furthermore, Tet2 expression was upregulated during all-*trans* retinoic acid-induced granulocytic differentiation of the promyelocytic cell line, NB4 [208], suggesting that Tet2 is important in myelopoiesis and could therefore be involved HSC regulation. Future studies should be performed to ascertain the occurrence and relevance of this mark in HSCs.

Non-CG methylation is also thought to be an important feature of mammalian stem cells and the MeDIP-seq data generated in this thesis is able to detect these forms of 5mC. This data should be further analysed to determine the presence and involvement of non-CG methylated in HSC regulation and ageing.

7.4 **DMR Validation**

aDMRs were validated technically by BS-pyrosequencing and functionally by *ex vivo* HSC culture and 5-Aza'dC induced hypomethylation. These assays were very cumbersome and time consuming in terms of assay design and assay validation. BS-pyrosequencing in particular is technically challenging and is of very low throughput, as individual assays for a

maximum of 50 bp DMRs must be designed, and separate reactions are required for each assay. As a result, a significant amount of starting DNA concentration is required to validate multiple targets. Furthermore, regions refractory to PCR such as repeat and GC-rich regions are typically impossible to validate by this method. Consequently, it was only possible to validate a single aDMR (*Sdpr*) in this study.

Ex vivo demethylation of HSCs was performed, in preference to *in vivo* 5-Aza'dC treatments, in an effort to limit the confounding effects of 5-Aza'dC on other cell types. By isolating and selectively demethylating HSCs, I could be certain that subsequent changes are intrinsic to these cells. This assay was sufficient to confirm the link between *Sdpr* promoter hypomethylation and increased gene expression; however, it was unable to elucidate the relevance of changes in DNAm in HSC regulation and ageing. A more suitable endpoint for this *ex vivo* assay is the transplantation of 5-Aza'dC demethylated HSC (Aza-HSC). Transplantation of Aza-HSCs into myelo-ablated mice will provide insights into the effect of DNA hypomethylation on processes such as homing, migration and key stem cell functions such as self-renewal and differentiation. It must be noted that such experiments will yield very little information pertaining the effects of changes to specific genes. This is because 5-Aza'dC induces global demethylation, and several loci are expected to be hypomethylated, making it difficult to pinpoint the exact cause of any novel phenotypes.

The functional relevance of candidate genes such as *Sdpr* and *Itgb3*, in HSC regulation, is currently unknown. *In vivo* knock-down and overexpression assays that investigate the specific roles of the candidate aDMGs will further elucidate the mechanisms of HSC ageing. Likewise, novel technologies such as micro-droplet multiplexing PCR [209] and sequence-specific genome editing nucleases e.g. Transcription Activator-Like Effector Nucleases (TALENs) [210], will greatly improve the feasibility of functionally validating multiple candidate genes in future studies.

7.5 **Transcriptome analysis**

Transcriptome analysis by RNA-seq revealed several interesting changes in gene expression, during HSC ageing (Chapter 5). High quality data sets were generated from the RNA-seq analysis performed in this thesis and these data have the potential to be analysed for other changes in gene regulation, which may accompany HSC ageing. These changes include alternative splicing and variations in transcript isoforms at different age points. Although the gene expression data provided isoform specific expression for each age group, these were only used to determine differential gene expression, irrespective of the specific isoform. Detailed analyses that specifically identify changes in the representation of transcript isoforms, at specific age points, are required to identify isoform-specific regulation of HSC ageing. Likewise, due to time constrains, the possibility of alternative splicing during HSC ageing was not pursued. It is possible that certain genes, whose overall expression appeared unaltered with age, do in fact express different isoforms of distinct functions, at different stages during HSC ageing. For instance, a manual scan of selected candidate genes revealed that multiple isoforms of *Itgb3, Dnmt1* and *Selp* are differentially expressed at different age points. Some of these isoforms showed age dependent up or down regulation, while others showed age specific expression (Appendix-Figure 5). Methylome analysis of HSC ageing identified an overrepresentation of aDMRs at CTCF binding sites (Chapter 4). CTCF is a zinc-finger DNA binding transcription factor, which binds to target sites at alternatively spliced exons, to promote their inclusion during transcription [172]. Conversely, DNAm at these sites prevents CTCF binding, and allows exon exclusion [172]. The enrichment of aDMRs at CTCF binding sites in HSCs suggests that DNAm at these sites could result in differential splicing and isoform expression during ageing. Further analysis of the RNA-seq data should be performed to investigate these possibilities.

7.6 **SDPR and Cell Movement in HSC Ageing**

I showed for the first time that cell movement is the main process that is altered during HSC ageing. This was inferred by the overrepresentation of differentially expressed cell movement genes in ageing HSCs. I also identified *Sdpr* as a differentially methylated and differentially expressed candidate gene that could be important in HSC regulation and ageing. *Sdpr* was selected based on the observations in this thesis, its known functions, and previous reports from independent studies [94,99], which also identified this gene as being differentially expressed in ageing HSCs. As detailed in Section 6.4.1, *Sdpr* is involved in caveolae biogenesis and its overexpression leads to caveolae elongation [168]. Caveolae are flask shaped lipid rafts and these cholesterol rich structures are thought to be essential for the establishment of cell polarity during migration [211]. Cell polarity is necessary for directional cellular migration and loss of caveolae scaffold protein, Caveolin-1, results in the loss of cell polarity and impaired cellular migration [212,213]. SDPR and Caveolin-1 are functionally inter-dependent [168]; It is possible that age dependent hypomethylation of the *Sdpr* promoter, and its subsequent overexpression, is involved in the dysregulation of cell movement genes during HSC ageing. In support of this, several genes coding for lipid raft-associated cell movement molecules such as Integrins, G protein-coupled receptors, CD44 and MPPs [214], were found to be differentially expressed in ageing HSCs. The findings in this thesis are indicative of a role for *Sdpr* in the regulation of HSC cell movement during ageing. This is perhaps achieved by an increased surface area for lipid raft associated proteins to interact and generate signalling cascades, which alter cell polarity and cellular migration.

Increased activity of RhoGTPase Cdc42, which regulates cell polarity and cell-extracellular matrix adhesion, has recently been shown to result in HSC ageing and loss of polarity [104]. Additionally, loss of this gene leads to loss of HSC quiescence and altered HSC niche interactions [215]. This thesis implicates the involvement of several other cell movement

genes in HSC ageing. Future studies should investigate the link between Sdpr, Itgb3 and Cdc42 in HSC functionality with age.

7.7 **Epigenetic Regulation of HSCs**

This thesis focused on DNAm modifications during HSC ageing and identified several significant changes that accompanied HSC ageing. Although there was a modest overlap between aDMGs and differentially expressed genes, several changes in the ageing LT—HSC transcriptome could not be directly explained by DNAm. Similarly, differential methylation in several genes did not appear to have an effect on gene expression. It is possible that subtle changes in the methylome, which were not significant enough to be detected as aDMRs, have an indirect and widespread effect on gene expression. This effect could result from alterations in HSC transcription factors binding to DNA, as a result of changes in DNAm. In support of this, haematopoietic transcription factors binding was found to be negatively correlated with DNAm in HSCs [216].

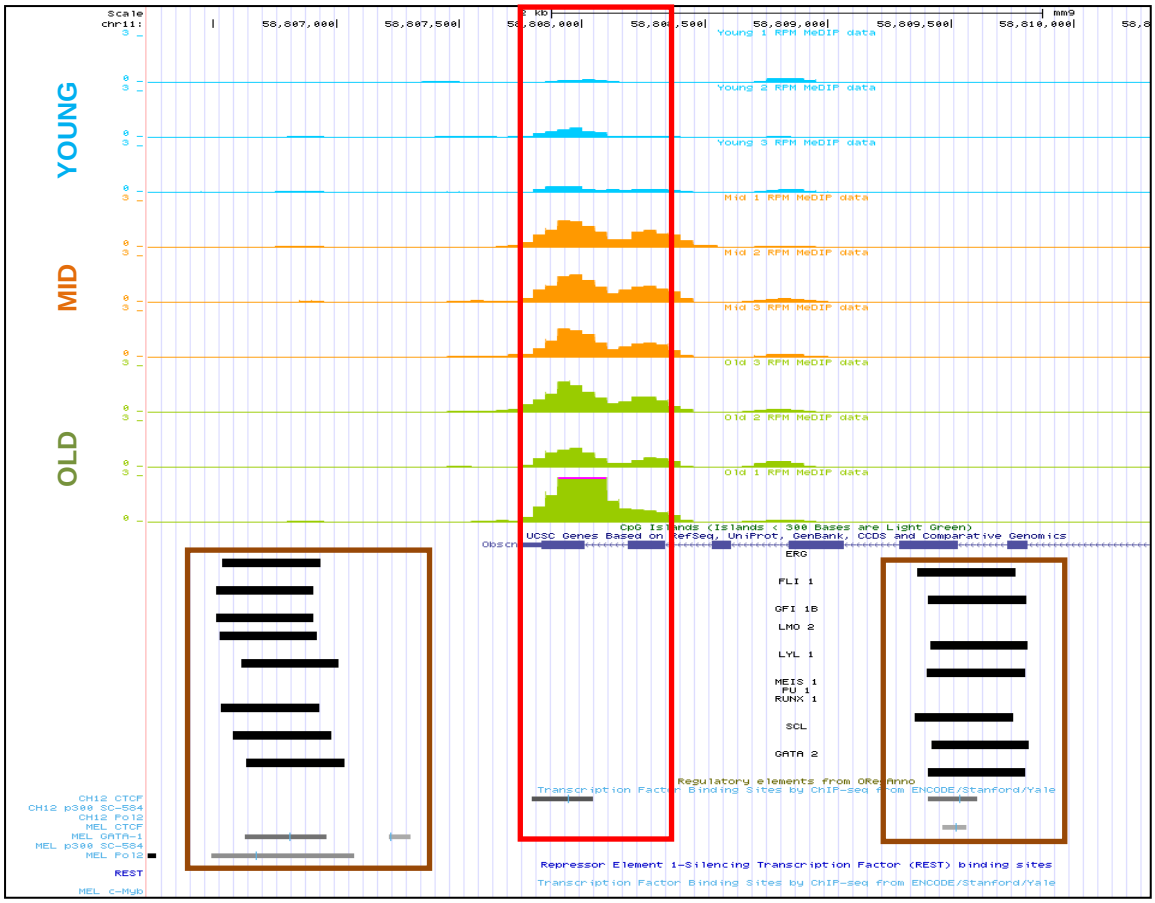


Figure 7-1: Hypermethylation at the 3' region of the *Obscn* gene in aged HSCs (red box).
This aDMR overlaps a CTCF binding site and is flanked by binding sites (brown box) for several HSC transcription factors.

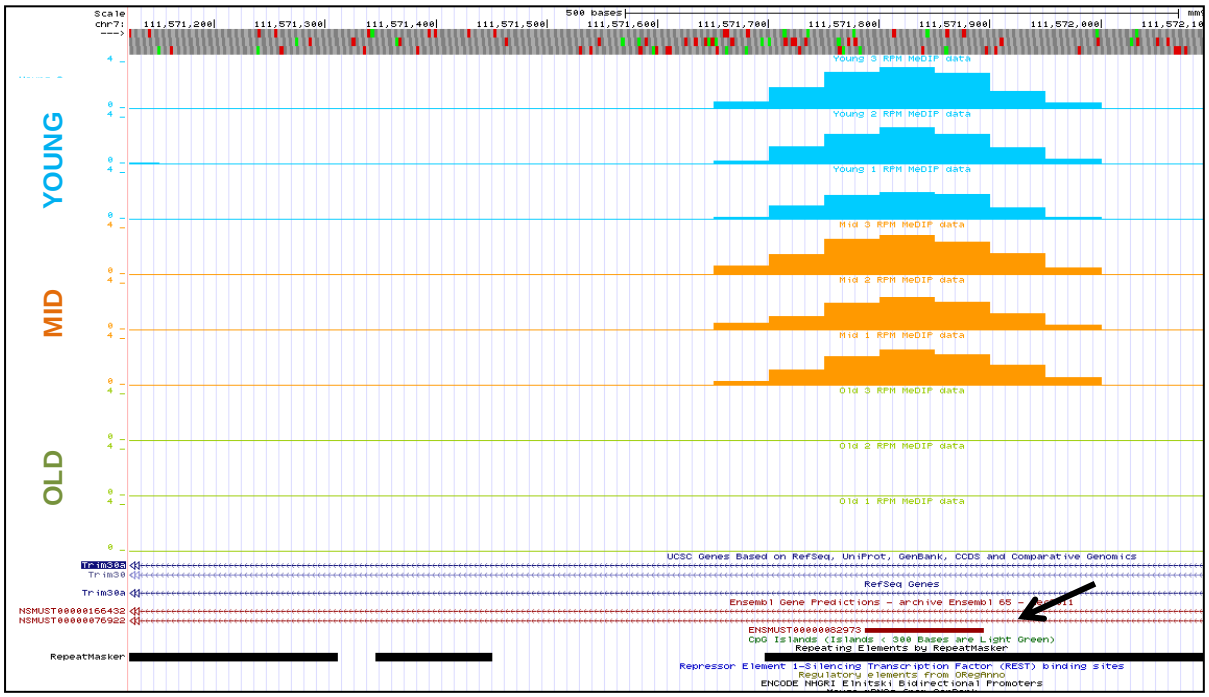
It was surprising to note that highly significant aDMRs that were associated with certain genes had no effect on their gene expression. For example, one of the most significant aDMR (FDR < 0.01) was found at the 3' end of the *Obscn* gene (Chapter 4, Fig. 7.1). Antisense microRNA (miRNA) can cause gene repression by binding to the 3'UTRs of mRNAs [217,218]. Age-dependent hypermethylation of the 3' end of the *Obscn* gene could interfere with the transcription of its 3'UTR, and thus with miRNA binding. While *Obscn* was not identified as upregulated with age in this study, miRNA has been suggested to induce gene repression without a change in mRNA level [218]. It is possible that age-dependent hypermethylation of genes like *Obscn* do indeed prevent gene repression, but this effect is only reflected in their protein abundance. Although, translation-only repression is a rare event [218], future studies should investigate the proteome of ageing HSCs, in order to obtain a more complete picture of HSC gene regulation, during ageing.

A hypermethylation hotspot was observed on chromosome 5, which corresponds to the location of several members of the *Speer* gene family (Chapter 4). Most of the *Speer* genes that were differentially methylated in this region were known pseudogenes. Pseudogenes were recently reported to modulate the expression of corresponding coding genes by competing for miRNA binding [219]. However, while *Speer* genes where not identified as differentially expressed in this thesis, it is possible that the hypermethylation of *Speer* pseudogenes have an effect on the expression of other coding genes.

The final example highlighted here is the age-related hypomethylation of *Trim30a*. This was the most significantly hypomethylated gene during HSC ageing and was also significantly down regulated with age (Table. 6.1, Fig. 7.2). This DNAm/gene expression correlation appears counter intuitive. However, the DMR overlaps a U6 non-coding small nuclear RNA (snRNA), and it is possible that the loss of DNAm in this region results in the activation of the snRNA, which perhaps causes the down regulation of Trim30a, through aberrant splicing. It will be interesting to investigate the effects of this snRNA on *Trim30a* regulation, and the impact this has on HSC functionality and ageing. Future studies could design synthetic forms of the ncRNA described here and transfect it into Young HSCs, in order to carry out this investigation.

In aggregate, DNAm is on its own an important epigenetic modification that is involved in the regulation of gene expression; however, it may act in collaboration with other forms of epigenetic mechanisms to modify gene expression. It is therefore important that these mechanisms are taken into consideration when designing future experiments.

Figure 7-2: aDMR in the *Trim30a* gene



This DMR overlaps a non-coding small nuclear RNA (arrowed)

7.8 **Conclusion**

I have generated and analysed methylomes and transcriptomes of phenotypically defined LT-HSCs, at key stages during physiological ageing. This thesis was, at the time of submission (Jan 2013), the most comprehensive study of HSC ageing (with regards to integrated methylome and transcription analysis). However, a very similar study was subsequently published, which performed and compared DNAm and gene expression in HSCs from foetal, young and old mice [220]. This study reported DNA hypermethylation in HSCs during ageing, which appears to be in contrast with the findings of this thesis. RRBS was the method used to quantify DNAm in the aforementioned study, and this method has been shown to be biased towards CGIs and promoter regions [118,125,221]. Additionally, its theoretical genome coverage is ~10% [125], while MeDIP-seq is able to determine truly global DNAm ([1]theoretical genome coverage of 100% [118,125,221]). As a result, the age-dependent hypermethylation that was reported in this newly published study, most likely reflects the fact that CG rich regions were compared during ageing. In support of this, I also observed an increase in DNAm at CGIs, during HSC ageing, while a significant loss ($p < 0.001$, Kolmogorov-Smirnov test) of DNAm was observed globally.

The methylomic analysis described in this thesis is particularly powerful, as DNA and RNA were concurrently isolated from the same functionally relevant cells. The findings in this thesis suggest that an epigenetic control of lipid raft associated proteins may be involved in the dysregulation of HSC cell movement during ageing. Further study of these lipid raft components is expected to improve our knowledge of the mechanisms directing HSC

[1]Genome coverage is not to be confused with global CG (methylome) coverage, which will be less than 100% in MeDIP-seq. This is because approximately 60-80% of CpGs are thought to be methylated in mammals [222] and MeDIP enriches for methylated portions of the genome. The methylome coverage in this thesis was ~60%.

ageing. The work presented in this thesis should be extended to other cell types and mammalian models, in order to gain insight into any universal mechanisms that underlie stem cell ageing.

7.9 **Outlook**

I found that locus specific alterations in the DNAm of aged HSCs potentially account for 1% of the gene expression changes that were observed in this thesis. While DNAm could also have an indirect effect on the gene expression of ageing HSCs (Section 7.5), it is likely that HSC ageing and age differential gene expression is multifactorial and regulated by several other mechanisms. As detailed in Chapter 1, the HSC niche is important in HSC regulation and interestingly, cell movement, which is a process that involves the dynamic interaction between HSC and their niche, was most significantly altered with age. It is therefore likely that age-related changes in the bone marrow will have a negative impact on the functionality of HSCs during ageing. Indeed fatty infiltration of the BM has been shown to occur during human ageing and BM fat and adipocyte contents are negatively correlated with HSC activity [223,224]. Increased cell movement in ageing HSCs could in part be a response to an altered BM, causing them to migrate to a more favourable niche. In support of this, active extramedullary haematopoiesis was observed in osteoblast deficient mice [225]. Furthermore, studies of muscle stem cells (satellite cells) suggest that circulating factors from young niches could rescue the functionality of aged stem cells [226]. However, transplantation of aged HSCs into a younger niche does not appear to rescue their functionality [94]. As a result, the role of the niche in regulating HSC ageing currently remains unclear. Nevertheless, it is possible that the process of transplantation somehow confounds these experiments; therefore, future studies should directly assess the effects of exposing aged HSCs to circulating factors from young mice. This could be achieved by heterochronic parabiotic pairing of young and old mice, where these pairs share the same circulation, or by *ex vivo* culture of aged HSCs on stroma that was derived from young mice. The former experiment was recently conducted with promising results [227], however, this article was subsequently retracted.

Polycomb group (PcG) proteins are transcription repressors and are important regulators of stem cell fate and gene expression [228]. A study of aged blood cells found PcG proteins

target genes (PcGTs) to be hypermethylated with age [110]. PcG proteins form two main complexes in mammals, which include polycomb repressive complex 1 and 2 (PRC 1 and PRC 2) [228]. Either one of the histone methyltransferases Ezh1 or Ezh2 are required in PRC2 complexes [229]. These proteins are important for the establishment of the repressive H3K27me mark [230] and Ezh1, which was downregulated in aged HSCs in this thesis, has recently been shown to be essential for HSC maintenance and the prevention of senescence [231]. Furthermore, *Runx1, Gata3, Myb, Cdc6 and Dntt*, which were differentially expressed in aged HSCs in this thesis, were identified as Ezh1 targets [231]. Age related downregulation of Ezh1 could in part be involved in the dysregulation of HSCs during ageing. Future studies should perform chromatin immunoprecipitation of PcG proteins such as Ezh1, to investigate their role in HSC ageing.

Young HSCs from short lived inbred strains such as the C3H/He, CBA/J and DBA/2 mice have a higher cycling activity than that of the longer lived C57BL/6 mice [232,233]. HSCs from these strains also show wide variations in cell frequency [234,235]. The genetic differences between these strains have been exploited by studies, which identified multiple quantitative trait loci (QTLs) and candidate genes that may underlie these differences [232,234,236,237]. This includes laxetin (*Lxn*), a gene whose expression is negatively correlated with the size of the HSC pool [237]. Acquisition of mutations in such genes could also influence the functionality of aged HSCs. Exome sequencing or resequencing of DNA from ageing HSCs could further elucidate the involvement of genetic factors in this process.

Reference List

Reference List

1. Ljubuncic P, Reznick AZ: **The evolutionary theories of aging revisited--a mini-review.** *Gerontology* 2009, **55:** 205-216.

2. Kirkwood TB: **Evolution of ageing.** *Mech Ageing Dev* 2002, **123:** 737-745.

3. Weismann A: **Essays Upon Heredity and Kindred Biological Problems.** Oxford, Clarendon Press; 1889.

4. Medawar PB: **An Unsolved Problem of Biology.** London: HK Lewis; 1959.

5. Williams GC: **Pleiotropy, natural selection, and the evolution of senescence.** *Evolution* 1957, 398-411.

6. Ungewitter E, Scrable H: **Antagonistic pleiotropy and p53.** *Mech Ageing Dev* 2009, **130:** 10-17.

7. Pearce DJ, Anjos-Afonso F, Ridler CM, Eddaoudi A, Bonnet D: **Age-dependent increase in side population distribution within hematopoiesis: implications for our understanding of the mechanism of aging.** *Stem Cells* 2007, **25:** 828-835.

8. Cairns J: **Somatic stem cells and the kinetics of mutagenesis and carcinogenesis.** *Proc Natl Acad Sci U S A* 2002, **99:** 10567-10570.

9. Morrison SJ, Wandycz AM, Akashi K, Globerson A, Weissman IL: **The aging of hematopoietic stem cells.** *Nat Med* 1996, **2:** 1011-1016.

10. Fleming WH, Alpern EJ, Uchida N, Ikuta K, Spangrude GJ, Weissman IL: **Functional heterogeneity is associated with the cell cycle status of murine hematopoietic stem cells.** *J Cell Biol* 1993, **122:** 897-902.

11. Spangrude GJ, Heimfeld S, Weissman IL: **Purification and characterization of mouse hematopoietic stem cells.** *Science* 1988, **241:** 58-62.

12. Matsuoka S, Ebihara Y, Xu M, Ishii T, Sugiyama D, Yoshino H *et al*.: **CD34 expression on long-term repopulating hematopoietic stem cells changes during developmental stages.** *Blood* 2001, **97:** 419-425.

13. Goodell MA, Rosenzweig M, Kim H, Marks DF, DeMaria M, Paradis G *et al*.: **Dye efflux studies suggest that hematopoietic stem cells expressing low or undetectable levels of CD34 antigen exist in multiple species.** *Nat Med* 1997, **3:** 1337-1345.

14. Kiel MJ, Yilmaz OH, Iwashita T, Yilmaz OH, Terhorst C, Morrison SJ: **SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells.** *Cell* 2005, **121:** 1109-1121.

15. Weksberg DC, Chambers SM, Boles NC, Goodell MA: **C.** *Blood* 2008, **111:** 2444-2451.

16. Goodell MA, Brose K, Paradis G, Conner AS, Mulligan RC: **Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo.** *J Exp Med* 1996, **183:** 1797-1806.

17. Zhou S, Schuetz JD, Bunting KD, Colapietro AM, Sampath J, Morris JJ *et al*.: **The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype.** *Nat Med* 2001, **7:** 1028-1034.

18. Pearce DJ, Ridler CM, Simpson C, Bonnet D: **Multiparameter analysis of murine bone marrow side population cells.** *Blood* 2004, **103:** 2541-2546.

19. Challen GA, Boles NC, Chambers SM, Goodell MA: **Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1.** *Cell Stem Cell* 2010, **6:** 265-278.

20. Ploemacher RE, van der Sluijs JP, Voerman JS, Brons NH: **An in vitro limiting-dilution assay of long-term repopulating hematopoietic stem cells in the mouse.** *Blood* 1989, **74:** 2755-2763.

21. Ploemacher RE, van der Sluijs JP, van Beurden CA, Baert MR, Chan PL: **Use of limiting-dilution type long-term marrow cultures in frequency analysis of marrow-repopulating and spleen colony-forming hematopoietic stem cells in the mouse.** *Blood* 1991, **78:** 2527-2533.

22. Sutherland HJ, Eaves CJ, Eaves AC, Dragowska W, Lansdorp PM: **Characterization and partial purification of human marrow cells capable of initiating long-term hematopoiesis in vitro.** *Blood* 1989, **74:** 1563-1570.

23. Trevisan M, Iscove NN: **Phenotypic analysis of murine long-term hemopoietic reconstituting cells quantitated competitively in vivo and comparison with more advanced colony-forming progeny.** *J Exp Med* 1995, **181:** 93-103.

24. Zhang CC, Lodish HF: **Murine hematopoietic stem cells change their surface phenotype during ex vivo expansion.** *Blood* 2005, **105:** 4314-4320.

25. Zheng J, Huynh H, Umikawa M, Silvany R, Zhang CC: **Angiopoietin-like protein 3 supports the activity of hematopoietic stem cells in the bone marrow niche.** *Blood* 2011, **117:** 470-479.

26. Harrison DE, Jordan CT, Zhong RK, Astle CM: **Primitive hemopoietic stem cells: direct assay of most productive populations by competitive repopulation with simple binomial, correlation and covariance calculations.** *Exp Hematol* 1993, **21:** 206-219.

27. Schofield R: **The relationship between the spleen colony-forming cell and the haemopoietic stem cell.** *Blood Cells* 1978, **4:** 7-25.

28. Haylock DN, Nilsson SK: **Stem cell regulation by the hematopoietic stem cell niche.** *Cell Cycle* 2005, **4:** 1353-1355.

29. Morrison SJ, Hemmati HD, Wandycz AM, Weissman IL: **The purification and characterization of fetal liver hematopoietic stem cells.** *Proc Natl Acad Sci U S A* 1995, **92:** 10302-10306.

30. de Bruijn MF, Speck NA, Peeters MC, Dzierzak E: **Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo.** *EMBO J* 2000, **19:** 2465-2474.

31. Muller AM, Medvinsky A, Strouboulis J, Grosveld F, Dzierzak E: **Development of hematopoietic stem cell activity in the mouse embryo.** *Immunity* 1994, **1:** 291-301.

32. Georgiades CS, Neyman EG, Francis IR, Sneider MB, Fishman EK: **Typical and atypical presentations of extramedullary hemopoiesis.** *AJR Am J Roentgenol* 2002, **179:** 1239-1243.

33. Zhang J, Niu C, Ye L, Huang H, He X, Tong WG *et al.*: **Identification of the haematopoietic stem cell niche and control of the niche size.** *Nature* 2003, **425:** 836-841.

34. Arai F, Suda T: **Maintenance of quiescent hematopoietic stem cells in the osteoblastic niche.** *Ann N Y Acad Sci* 2007, **1106:** 41-53.

35. Nilsson SK, Johnston HM, Whitty GA, Williams B, Webb RJ, Denhardt DT *et al*.: **Osteopontin, a key component of the hematopoietic stem cell niche and regulator of primitive hematopoietic progenitor cells.** *Blood* 2005, **106:** 1232-1239.

36. Petit I, Szyper-Kravitz M, Nagler A, Lahav M, Peled A, Habler L *et al*.: **G-CSF induces stem cell mobilization by decreasing bone marrow SDF-1 and up-regulating CXCR4.** *Nat Immunol* 2002, **3:** 687-694.

37. Arai F, Hirao A, Ohmura M, Sato H, Matsuoka S, Takubo K *et al*.: **Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche.** *Cell* 2004, **118:** 149-161.

38. Kopp HG, Avecilla ST, Hooper AT, Rafii S: **The bone marrow vascular niche: home of HSC differentiation and mobilization.** *Physiology (Bethesda)* 2005, **20:** 349-356.

39. Kiel MJ, Morrison SJ: **Uncertainty in the niches that maintain haematopoietic stem cells.** *Nat Rev Immunol* 2008, **8:** 290-301.

40. Arai F, Hirao A, Suda T: **Regulation of hematopoiesis and its interaction with stem cell niches.** *Int J Hematol* 2005, **82:** 371-376.

41. Pietras EM, Warr MR, Passegue E: **Cell cycle regulation in hematopoietic stem cells.** *J Cell Biol* 2011, **195:** 709-720.

42. Antonchuk J, Sauvageau G, Humphries RK: **HOXB4-induced expansion of adult hematopoietic stem cells ex vivo.** *Cell* 2002, **109:** 39-45.

43. Miyake N, Brun AC, Magnusson M, Miyake K, Scadden DT, Karlsson S: **HOXB4-induced self-renewal of hematopoietic stem cells is significantly enhanced by p21 deficiency.** *Stem Cells* 2006, **24:** 653-661.

44. Cheng T, Rodrigues N, Shen H, Yang Y, Dombkowski D, Sykes M *et al*.: **Hematopoietic stem cell quiescence maintained by p21cip1/waf1.** *Science* 2000, **287:** 1804-1808.

45. Ficara F, Murphy MJ, Lin M, Cleary ML: **Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence.** *Cell Stem Cell* 2008, **2:** 484-496.

46. Park IK, Qian D, Kiel M, Becker MW, Pihalja M, Weissman IL *et al*.: **Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells.** *Nature* 2003, **423:** 302-305.

47. Oh IH: **Microenvironmental targeting of Wnt/beta-catenin signals for hematopoietic stem cell regulation.** *Expert Opin Biol Ther* 2010.

48. Reya T, Duncan AW, Ailles L, Domen J, Scherer DC, Willert K *et al*.: **A role for Wnt signalling in self-renewal of haematopoietic stem cells.** *Nature* 2003, **423:** 409-414.

49. Bird A: **Perceptions of epigenetics.** *Nature* 2007, **447:** 396-398.

50. Miranda-Saavedra D, De S, Trotter MW, Teichmann SA, Gottgens B: **BloodExpress: a database of gene expression in mouse haematopoiesis.** *Nucleic Acids Res* 2009, **37:** D873-D879.

51. Feinberg AP: **Phenotypic plasticity and the epigenetics of human disease.** *Nature* 2007, **447:** 433-440.

52. Weishaupt H, Sigvardsson M, Attema JL: **Epigenetic chromatin states uniquely define the developmental plasticity of murine hematopoietic stem cells.** *Blood* 2010, **115:** 247-256.

53. Peterson CL, Laniel MA: **Histones and histone modifications.** *Curr Biol* 2004, **14:** R546-R551.

54. Attema JL, Papathanasiou P, Forsberg EC, Xu J, Smale ST, Weissman IL: **Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis.** *Proc Natl Acad Sci U S A* 2007, **104:** 12371-12376.

55. Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W *et al.*: **Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation.** *Cell Stem Cell* 2009, **4:** 80-93.

56. Miyamoto T, Iwasaki H, Reizis B, Ye M, Graf T, Weissman IL *et al.*: **Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment.** *Dev Cell* 2002, **3:** 137-147.

57. Pina C, Fugazza C, Tipping AJ, Brown J, Soneji S, Teles J *et al.*: **Inferring rules of lineage commitment in haematopoiesis.** *Nat Cell Biol* 2012, **14:** 287-294.

58. Goll MG, Bestor TH: **Eukaryotic cytosine methyltransferases.** *Annu Rev Biochem* 2005, **74:** 481-514.

59. Sarraf SA, Stancheva I: **Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly.** *Mol Cell* 2004, **15:** 595-605.

60. Suzuki MM, Bird A: **DNA methylation landscapes: provocative insights from epigenomics.** *Nat Rev Genet* 2008, **9:** 465-476.

61. Walsh CP, Chaillet JR, Bestor TH: **Transcription of IAP endogenous retroviruses is constrained by cytosine methylation.** *Nat Genet* 1998, **20:** 116-117.

62. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD *et al.*: **Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning.** *Nature* 2008, **452:** 215-219.

63. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J *et al.*: **Human DNA methylomes at base resolution show widespread epigenomic differences.** *Nature* 2009, **462:** 315-322.

64. Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird AP, Jaenisch R: **Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a.** *Proceedings of the National Academy of Sciences* 2000, **97:** 5237-5242.

65. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y *et al.*: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.** *Science* 2009, **324:** 930-935.

66. Kriaucionis S, Heintz N: **The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.** *Science* 2009, **324:** 929-930.

67. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y: **Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.** *Nature* 2010.

68. Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA *et al.*: **Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation.** *Nature* 2011, **473:** 398-402.

69. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T: **The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation.** *J Biol Chem* 2003, **278:** 4035-4040.

70. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T: **Demethylation of the zygotic paternal genome.** *Nature* 2000, **403:** 501-502.

71. Oswald J, Engemann S, Lane N, Mayer W, Olek A, Fundele R *et al.*: **Active demethylation of the paternal genome in the mouse zygote.** *Curr Biol* 2000, **10:** 475-478.

72. Wossidlo M, Nakamura T, Lepikhov K, Marques CJ, Zakhartchenko V, Boiani M *et al*.: **5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming.** *Nat Commun* 2011, **2:** 241.

73. Borgel J, Guibert S, Li Y, Chiba H, Schubeler D, Sasaki H *et al*.: **Targets and dynamics of promoter DNA methylation during early mouse development.** *Nat Genet* 2010, **42:** 1093-1100.

74. Reik W, Dean W, Walter J: **Epigenetic reprogramming in mammalian development.** *Science* 2001, **293:** 1089-1093.

75. Feng S, Jacobsen SE, Reik W: **Epigenetic reprogramming in plant and animal development.** *Science* 2010, **330:** 622-627.

76. Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H *et al*.: **Role for DNA methylation in the control of cell type specific maspin expression.** *Nat Genet* 2002, **31:** 175-179.

77. Illingworth R, Kerr A, Desousa D, Jorgensen H, Ellis P, Stalker J *et al*.: **A novel CpG island set identifies tissue-specific methylation at developmental gene loci.** *PLoS Biol* 2008, **6:** e22.

78. Deaton AM, Webb S, Kerr AR, Illingworth RS, Guy J, Andrews R *et al*.: **Cell type-specific DNA methylation at intragenic CpG islands in the immune system.** *Genome Res* 2011, **21:** 1074-1086.

79. Kim K, Doi A, Wen B, Ng K, Zhao R, Cahan P *et al*.: **Epigenetic memory in induced pluripotent stem cells.** *Nature* 2010.

80. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P *et al*.: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores.** *Nat Genet* 2009, **41:** 178-186.

81. Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R *et al*.: **Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts.** *Nat Genet* 2009, **41:** 1350-1353.

82. Feinberg AP, Vogelstein B: **Hypomethylation distinguishes genes of some human cancers from their normal counterparts.** *Nature* 1983, **301:** 89-92.

83. Feinberg AP, Tycko B: **The history of cancer epigenetics.** *Nat Rev Cancer* 2004, **4:** 143-153.

84. Feinberg AP, Ohlsson R, Henikoff S: **The epigenetic progenitor origin of human cancer.** *Nat Rev Genet* 2006, **7:** 21-33.

85. Eden A, Gaudet F, Waghmare A, Jaenisch R: **Chromosomal instability and tumors promoted by DNA hypomethylation.** *Science* 2003, **300:** 455.

86. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW *et al*.: **Induction of tumors in mice by genomic hypomethylation.** *Science* 2003, **300:** 489-492.

87. Bibikova M, Laurent LC, Ren B, Loring JF, Fan JB: **Unraveling epigenetic regulation in embryonic stem cells.** *Cell Stem Cell* 2008, **2:** 123-134.

88. Broske AM, Vockentanz L, Kharazi S, Huska MR, Mancini E, Scheller M *et al*.: **DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction.** *Nat Genet* 2009, **41:** 1207-1215.

89. Trowbridge JJ, Snow JW, Kim J, Orkin SH: **DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells.** *Cell Stem Cell* 2009, **5:** 442-449.

90. Challen GA, Sun D, Jeong M, Luo M, Jelinek J, Berg JS *et al*.: **Dnmt3a is essential for hematopoietic stem cell differentiation.** *Nat Genet* 2012, **44:** 23-31.

91. Ji H, Ehrlich LI, Seita J, Murakami P, Doi A, Lindau P *et al*.: **Comprehensive methylome map of lineage commitment from haematopoietic progenitors.** *Nature* 2010.

92. Van ZG, Holland BP, Eldridge PW, Chen JJ: **Genotype-restricted growth and aging patterns in hematopoietic stem cell populations of allophenic mice.** *J Exp Med* 1990, **171:** 1547-1565.

93. Sudo K, Ema H, Morita Y, Nakauchi H: **Age-associated characteristics of murine hematopoietic stem cells.** *J Exp Med* 2000, **192:** 1273-1280.

94. Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ *et al*.: **Cell intrinsic alterations underlie hematopoietic stem cell aging.** *Proc Natl Acad Sci U S A* 2005, **102:** 9194-9199.

95. Kim M, Moon HB, Spangrude GJ: **Major age-related changes of mouse hematopoietic stem/progenitor cells.** *Ann N Y Acad Sci* 2003, **996:** 195-208.

96. Liang Y, Van ZG, Szilvassy SJ: **Effects of aging on the homing and engraftment of murine hematopoietic stem and progenitor cells.** *Blood* 2005, **106:** 1479-1487.

97. Dykstra B, Olthof S, Schreuder J, Ritsema M, de HG: **Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells.** *J Exp Med* 2011, **208:** 2691-2703.

98. Silverman LR, Mufti GJ: **Methylation inhibitor therapy in the treatment of myelodysplastic syndrome.** *Nat Clin Pract Oncol* 2005, **2 Suppl 1:** S12-S23.

99. Chambers SM, Shaw CA, Gatza C, Fisk CJ, Donehower LA, Goodell MA: **Aging Hematopoietic Stem Cells Decline in Function and Exhibit Epigenetic Dysregulation.** *PLoS Biol* 2007, **5:** e201.

100. Van ZG, Scott-Micus K, Thompson BP, Fleischman RA, Perkins S: **Stem cell quiescence/activation is reversible by serial transplantation and is independent of stromal cell genotype in mouse aggregation chimeras.** *Exp Hematol* 1992, **20:** 470-475.

101. Sieburg HB, Cho RH, Dykstra B, Uchida N, Eaves CJ, Muller-Sieburg CE: **The hematopoietic stem compartment consists of a limited number of discrete stem cell subsets.** *Blood* 2006, **107:** 2311-2316.

102. Beerman I, Bhattacharya D, Zandi S, Sigvardsson M, Weissman IL, Bryder D *et al*.: **Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion.** *Proc Natl Acad Sci U S A* 2010, **107:** 5465-5470.

103. Muller-Sieburg C, Sieburg HB: **Stem cell aging: survival of the laziest?** *Cell Cycle* 2008, **7:** 3798-3804.

104. Florian MC, Dorr K, Niebel A, Daria D, Schrezenmeier H, Rojewski M *et al*.: **Cdc42 activity regulates hematopoietic stem cell aging and rejuvenation.** *Cell Stem Cell* 2012, **10:** 520-530.

105. Dumble M, Moore L, Chambers SM, Geiger H, Van ZG, Goodell MA *et al*.: **The impact of altered p53 dosage on hematopoietic stem cell dynamics during aging.** *Blood* 2007, **109:** 1736-1742.

106. Janzen V, Forkert R, Fleming HE, Saito Y, Waring MT, Dombkowski DM *et al*.: **Stem-cell ageing modified by the cyclin-dependent kinase inhibitor p16INK4a.** *Nature* 2006, **443:** 421-426.

107. Rossi DJ, Bryder D, Zahn JM, Ahlenius H, Sonu R, Wagers AJ *et al*.: **Cell intrinsic alterations underlie hematopoietic stem cell aging.** *Proc Natl Acad Sci U S A* 2005, **102:** 9194-9199.

108. Rossi DJ, Bryder D, Seita J, Nussenzweig A, Hoeijmakers J, Weissman IL: **Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age.** *Nature* 2007, **447:** 725-729.

109. Passegue E, Wagner EF, Weissman IL: **JunB deficiency leads to a myeloproliferative disorder arising from hematopoietic stem cells.** *Cell* 2004, **119:** 431-443.

110. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H *et al*.: **Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer.** *Genome Res* 2010, **20:** 440-446.

111. Issa JP, Garcia-Manero G, Giles FJ, Mannari R, Thomas D, Faderl S *et al*.: **Phase 1 study of low-dose prolonged exposure schedules of the hypomethylating agent 5-aza-2'-deoxycytidine (decitabine) in hematopoietic malignancies.** *Blood* 2004, **103:** 1635-1640.

112. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE *et al*.: **The relationship of DNA methylation with age, gender and genotype in twins and healthy controls.** *PLoS One* 2009, **4:** e6767.

113. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A *et al*.: **Distinct DNA methylomes of newborns and centenarians.** *Proc Natl Acad Sci U S A* 2012, **109:** 10522-10527.

114. Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J *et al*.: **Widespread and tissue specific age-related DNA methylation changes in mice.** *Genome Res* 2010, **20:** 332-340.

115. Renstrom J, Kroger M, Peschel C, Oostendorp RA: **How the niche regulates hematopoietic stem cells.** *Chem Biol Interact* 2010, **184:** 7-15.

116. Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R, Bahler J: **Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells.** *Cell* 2012, **151:** 671-683.

117. Bocker MT, Hellwig I, Breiling A, Eckstein V, Ho AD, Lyko F: **Genome-wide promoter DNA methylation dynamics of human hematopoietic progenitor cells during differentiation and aging.** *Blood* 2011, **117:** e182-e189.

118. Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C, Downey SL *et al*.: **Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications.** *Nat Biotechnol* 2010, **28:** 1097-1105.

119. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW *et al*.: **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** *Proc Natl Acad Sci U S A* 1992, **89:** 1827-1831.

120. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB: **Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands.** *Proc Natl Acad Sci U S A* 1996, **93:** 9821-9826.

121. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D *et al*.: **MethyLight: a high-throughput assay to measure DNA methylation.** *Nucleic Acids Res* 2000, **28:** E32.

122. Xiong Z, Laird PW: **COBRA: a sensitive and quantitative DNA methylation assay.** *Nucleic Acids Res* 1997, **25:** 2532-2534.

123. Colella S, Shen L, Baggerly KA, Issa JP, Krahe R: **Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites.** *Biotechniques* 2003, **35:** 146-150.

124. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R: **Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.** *Nucleic Acids Research* **33:** 5868-5877.

125. Beck S: **Taking the measure of the methylome.** *Nat Biotechnol* 2010, **28:** 1026-1028.

126. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL *et al*.: **Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells.** *Nat Genet* 2005, **37:** 853-862.

127. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E *et al*.: **A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis.** *Nat Biotechnol* 2008, **26:** 779-785.

128. Cross SH, Charlton JA, Nan X, Bird AP: **Purification of CpG islands using a methylated DNA binding column.** *Nat Genet* 1994, **6:** 236-244.

129. Serre D, Lee BH, Ting AH: **MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome.** *Nucleic Acids Res* 2010, **38:** 391-399.

130. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M *et al*.: **DNA methylation profiling of human chromosomes 6, 20 and 22.** *Nat Genet* 2006, **38:** 1378-1385.

131. Butcher LM, Beck S: **AutoMeDIP-seq: A high-throughput, whole genome, DNA methylation assay.** *Methods* 2010.

132. Goodell MA, McKinney-Freeman S, Camargo FD: **Isolation and characterization of side population cells.** *Methods Mol Biol* 2005, **290:** 343-352.

133. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D *et al*.: **Methylome analysis using MeDIP-seq with low DNA concentrations.** *Nat Protoc* 2012, **7:** 617-636.

134. Wilson GA, Dhami P, Feber A, CortÃ¡zar D, Suzuki Y, Schulz R *et al*.: **Resources for methylome analysis suitable for gene knockout studies of potential epigenome modifiers.** *GigaScience* 2012, **1**.

135. Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B, Lehrach H *et al*.: **Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage.** *Genome Res* 2010, **20:** 1441-1450.

136. Li H, Durbin R: **Fast and accurate short read alignment with BurrowsΓÇôWheeler transform.** *Bioinformatics* 2009, **25:** 1754-1760.

137. Nix DA, Courdy SJ, Boucher KM: **Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks.** *BMC Bioinformatics* 2008, **9:** 523.

138. Anders S, Huber W: **Differential expression analysis for sequence count data.** *Genome Biol* 2010, **11:** R106.

139. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.** *Bioinformatics* 2010, **26:** 841-842.

140. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.** *BMC Bioinformatics* 2011, **12:** 323.

141. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10:** R25.

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

143. Adolfsson J, Borge OJ, Bryder D, Theilgaard-Monch K, Astrand-Grundstrom I, Sitnicka E *et al*.: **Upregulation of Flt3 expression within the bone marrow Lin(-)Sca1(+)c-kit(+) stem cell compartment is accompanied by loss of self-renewal capacity.** *Immunity* 2001, **15:** 659-669.

144. Pearce DJ, Anjos-Afonso F, Ridler CM, Eddaoudi A, Bonnet D: **Age-dependent increase in side population distribution within hematopoiesis: implications for our understanding of the mechanism of aging.** *Stem Cells* 2007, **25:** 828-835.

145. Challen GA, Little MH: **A side order of stem cells: the SP phenotype.** *Stem Cells* 2006, **24:** 3-12.

146. Guo Y, Follo M, Geiger K, Lubbert M, Engelhardt M: **Side-population cells from different precursor compartments.** *J Hematother Stem Cell Res* 2003, **12:** 71-82.

147. DeAngelis MM, Wang DG, Hawkins TL: **Solid-phase reversible immobilization for the isolation of PCR products.** *Nucleic Acids Res* 1995, **23:** 4742-4743.

148. Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W *et al*.: **A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454.** *Genome Biol* 2010, **11:** R15.

149. Lin KK, Goodell MA: **Purification of hematopoietic stem cells using the side population.** *Methods Enzymol* 2006, **420:** 255-264.

150. Weber M: **[Profiles of DNA methylation in normal and cancer cells].** *Med Sci (Paris)* 2008, **24:** 731-734.

151. Trowbridge JJ, Snow JW, Kim J, Orkin SH: **DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells.** *Cell Stem Cell* 2009, **5:** 442-449.

152. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A *et al*.: **Distinct DNA methylomes of newborns and centenarians.** *Proc Natl Acad Sci U S A* 2012, **109:** 10522-10527.

153. Spiess AN, Walther N, Muller N, Balvers M, Hansis C, Ivell R: **SPEER--a new family of testis-specific genes from the mouse.** *Biol Reprod* 2003, **68:** 2044-2054.

154. te Velthuis AJ, Admiraal JF, Bagowski CP: **Molecular evolution of the MAGUK family in metazoan genomes.** *BMC Evol Biol* 2007, **7:** 129.

155. Feber A, Wilson GA, Zhang L, Presneau N, Idowu B, Down TA *et al*.: **Comparative methylome analysis of benign and malignant peripheral nerve sheath tumors.** *Genome Res* 2011, **21:** 515-524.

156. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB *et al*.: **GREAT improves functional interpretation of cis-regulatory regions.** *Nat Biotechnol* 2010, **28:** 495-501.

157. Hogart A, Lichtenberg J, Ajay SS, Anderson S, Margulies EH, Bodine DM: **Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites.** *Genome Res* 2012, **22:** 1407-1418.

158. Tomsig JL, Creutz CE: **Copines: a ubiquitous family of Ca(2+)-dependent phospholipid-binding proteins.** *Cell Mol Life Sci* 2002, **59:** 1467-1477.

159. Tomsig JL, Snyder SL, Creutz CE: **Identification of targets for calcium signaling through the copine family of proteins. Characterization of a coiled-coil copine-binding motif.** *J Biol Chem* 2003, **278:** 10048-10054.

160. Mayne M, Moffatt T, Kong H, McLaren PJ, Fowke KR, Becker KG *et al*.: **CYFIP2 is highly abundant in CD4+ cells from multiple sclerosis patients and is involved in T cell adhesion.** *Eur J Immunol* 2004, **34:** 1217-1227.

161. Liu H, Rodgers ND, Jiao X, Kiledjian M: **The scavenger mRNA decapping enzyme DcpS is a member of the HIT family of pyrophosphatases.** *EMBO J* 2002, **21:** 4699-4708.

162. Stafford LJ, Xia C, Ma W, Cai Y, Liu M: **Identification and characterization of mouse metastasis-suppressor KiSS1 and its G-protein-coupled receptor.** *Cancer Res* 2002, **62:** 5399-5404.

163. Koli K, Hyytiainen M, Ryynanen MJ, Keski-Oja J: **Sequential deposition of latent TGF-beta binding proteins (LTBPs) during formation of the extracellular matrix in human lung fibroblasts.** *Exp Cell Res* 2005, **310:** 370-382.

164. Maijenburg MW, Gilissen C, Melief SM, Kleijer M, Weijer K, Ten BA *et al*.: **Nuclear receptors Nur77 and Nurr1 modulate mesenchymal stromal cell migration.** *Stem Cells Dev* 2012, **21:** 228-238.

165. Perry NA, Shriver M, Mameza MG, Grabias B, Balzer E, Kontrogianni-Konstantopoulos A: **Loss of giant obscurins promotes breast epithelial cell survival through apoptotic resistance.** *FASEB J* 2012, **26:** 2764-2775.

166. Dong H, Osmanova V, Epstein PM, Brocke S: **Phosphodiesterase 8 (PDE8) regulates chemotaxis of activated lymphocytes.** *Biochem Biophys Res Commun* 2006, **345:** 713-719.

167. Gustincich S, Schneider C: **Serum deprivation response gene is induced by serum starvation but not by contact inhibition.** *Cell Growth Differ* 1993, **4:** 753-760.

168. Hansen CG, Bright NA, Howard G, Nichols BJ: **SDPR induces membrane curvature and functions in the formation of caveolae.** *Nat Cell Biol* 2009, **11:** 807-814.

169. Shi M, Deng W, Bi E, Mao K, Ji Y, Lin G *et al*.: **TRIM30 alpha negatively regulates TLR-mediated NF-kappa B activation by targeting TAB2 and TAB3 for degradation.** *Nat Immunol* 2008, **9:** 369-377.

170. Bell AC, Felsenfeld G: **Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene.** *Nature* 2000, **405:** 482-485.

171. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM: **CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus.** *Nature* 2000, **405:** 486-489.

172. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M *et al*.: **CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.** *Nature* 2011, **479:** 74-79.

173. Jelinic P, Shaw P: **Loss of imprinting and cancer.** *J Pathol* 2007, **211:** 261-268.

174. Lieu YK, Reddy EP: **Conditional c-myb knockout in adult hematopoietic stem cells leads to loss of self-renewal due to impaired proliferation and accelerated differentiation.** *Proc Natl Acad Sci U S A* 2009, **106:** 21689-21694.

175. Dasse E, Volpe G, Walton DS, Wilson N, Del PW, O'Neill LP *et al*.: **Distinct regulation of c-myb gene expression by HoxA9, Meis1 and Pbx proteins in normal hematopoietic progenitors and transformed myeloid cells.** *Blood Cancer J* 2012, **2:** e76.

176. Wang LD, Wagers AJ: **Dynamic niches in the origination and differentiation of haematopoietic stem cells.** *Nat Rev Mol Cell Biol* 2011, **12:** 643-655.

177. Umemoto T, Yamato M, Shiratsuchi Y, Terasawa M, Yang J, Nishida K *et al*.: **Expression of Integrin beta3 is correlated to the properties of quiescent hemopoietic stem cells possessing the side population phenotype.** *J Immunol* 2006, **177:** 7733-7739.

178. Wagers AJ, Allsopp RC, Weissman IL: **Changes in integrin expression are associated with altered homing properties of Lin(-/lo)Thy1.1(lo)Sca-1(+)c-kit(+) hematopoietic stem cells following**

mobilization by cyclophosphamide/granulocyte colony-stimulating factor. *Exp Hematol* 2002, **30:** 176-185.

179. Labat-Robert J: **Cell-Matrix interactions, the role of fibronectin and integrins. A survey.** *Pathol Biol (Paris)* 2012, **60:** 15-19.

180. Brown EJ: **Integrin-associated proteins.** *Curr Opin Cell Biol* 2002, **14:** 603-607.

181. Chandras C, Harris TE, Bernal AL, Abayasekara DR, Michael AE: **PTGER1 and PTGER2 receptors mediate regulation of progesterone synthesis and type 1 11beta-hydroxysteroid dehydrogenase activity by prostaglandin E2 in human granulosa lutein cells.** *J Endocrinol* 2007, **194:** 595-602.

182. Calvi LM, Adams GB, Weibrecht KW, Weber JM, Olson DP, Knight MC *et al*.: **Osteoblastic cells regulate the haematopoietic stem cell niche.** *Nature* 2003, **425:** 841-846.

183. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10:** 57-63.

184. Bowie MB, McKnight KD, Kent DG, McCaffrey L, Hoodless PA, Eaves CJ: **Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect.** *J Clin Invest* 2006, **116:** 2808-2816.

185. Hirao A, Arai F, Suda T: **Regulation of cell cycle in hematopoietic stem cells by the niche.** *Cell Cycle* 2004, **3:** 1481-1483.

186. Garrett RW, Gasiewicz TA: **The aryl hydrocarbon receptor agonist 2,3,7,8-tetrachlorodibenzo-p-dioxin alters the circadian rhythms, quiescence, and expression of clock genes in murine hematopoietic stem and progenitor cells.** *Mol Pharmacol* 2006, **69:** 2076-2083.

187. Casado FL, Singh KP, Gasiewicz TA: **Aryl hydrocarbon receptor activation in hematopoietic stem/progenitor cells alters cell function and pathway-specific gene modulation reflecting changes in cellular trafficking and migration.** *Mol Pharmacol* 2011, **80:** 673-682.

188. Zarbock A, Polanowska-Grabowska RK, Ley K: **Platelet-neutrophil-interactions: linking hemostasis and inflammation.** *Blood Rev* 2007, **21:** 99-111.

189. Tedder TF, Steeber DA, Chen A, Engel P: **The selectins: vascular adhesion molecules.** *FASEB J* 1995, **9:** 866-873.

190. Smith E, Sigvardsson M: **The roles of transcription factors in B lymphocyte commitment, development, and transformation.** *J Leukoc Biol* 2004, **75:** 973-981.

191. Christensen JL, Wright DE, Wagers AJ, Weissman IL: **Circulation and chemotaxis of fetal hematopoietic stem cells.** *PLoS Biol* 2004, **2:** E75.

192. Wright N, Hidalgo A, Rodriguez-Frade JM, Soriano SF, Mellado M, Parmo-Cabanas M *et al*.: **The chemokine stromal cell-derived factor-1 alpha modulates alpha 4 beta 7 integrin-mediated lymphocyte adhesion to mucosal addressin cell adhesion molecule-1 and fibronectin.** *J Immunol* 2002, **168:** 5268-5277.

193. Aiuti A, Webb IJ, Bleul C, Springer T, Gutierrez-Ramos JC: **The chemokine SDF-1 is a chemoattractant for human CD34+ hematopoietic progenitor cells and provides a new mechanism to explain the mobilization of CD34+ progenitors to peripheral blood.** *J Exp Med* 1997, **185:** 111-120.

194. Taniguchi H, Toyoshima T, Fukao K, Nakauchi H: **Presence of hematopoietic stem cells in the adult liver.** *Nat Med* 1996, **2:** 198-203.

195. Kohler A, Schmithorst V, Filippi MD, Ryan MA, Daria D, Gunzer M *et al*.: **Altered cellular dynamics and endosteal location of aged early hematopoietic progenitor cells revealed by time-lapse intravital imaging in long bones.** *Blood* 2009, **114:** 290-298.

196. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M *et al*.: **Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency.** *Nature* 2010, **463:** 1101-1105.

197. Li Y, Zhu J, Tian G, Li N, Li Q, Ye M *et al*.: **The DNA methylome of human peripheral blood mononuclear cells.** *PLoS Biol* 2010, **8:** e1000533.

198. Hayashi H, Sano H, Seo S, Kume T: **The Foxc2 transcription factor regulates angiogenesis via induction of integrin beta3 expression.** *J Biol Chem* 2008, **283:** 23791-23800.

199. Martinez-Lopez MJ, Alcantara S, Mascaro C, Perez-Branguli F, Ruiz-Lozano P, Maes T *et al*.: **Mouse neuron navigator 1, a novel microtubule-associated protein involved in neuronal migration.** *Mol Cell Neurosci* 2005, **28:** 599-612.

200. Muley PD, McNeill EM, Marzinke MA, Knobel KM, Barr MM, Clagett-Dame M: **The atRA-responsive gene neuron navigator 2 functions in neurite outgrowth and axonal elongation.** *Dev Neurobiol* 2008, **68:** 1441-1453.

201. Wang Y, Li Z, Xu P, Huang L, Tong J, Huang H *et al*.: **Angiomotin-like2 gene (amotl2) is required for migration and proliferation of endothelial cells during angiogenesis.** *J Biol Chem* 2011, **286:** 41095-41104.

202. Christman JK: **5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy.** *Oncogene* 2002, **21:** 5483-5495.

203. Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, Kaimakis P *et al*.: **Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators.** *Cell Stem Cell* 2010, **7:** 532-544.

204. Harris TD, Buzby PR, Babcock H, Beer E, Bowers J, Braslavsky I *et al*.: **Single-molecule DNA sequencing of a viral genome.** *Science* 2008, **320:** 106-109.

205. Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H: **Continuous base identification for single-molecule nanopore DNA sequencing.** *Nat Nano* 2009, **4:** 265-270.

206. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G *et al*.: **Real-Time DNA Sequencing from Single Polymerase Molecules.** *Science* 2009, **323:** 133-138.

207. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA *et al*.: **Direct detection of DNA methylation during single-molecule, real-time sequencing.** *Nat Meth* 2010, **7:** 461-465.

208. Langemeijer SM, Kuiper RP, Berends M, Knops R, Aslanyan MG, Massop M *et al*.: **Acquired mutations in TET2 are common in myelodysplastic syndromes.** *Nat Genet* 2009, **41:** 838-842.

209. Komori HK, LaMere SA, Torkamani A, Hart GT, Kotsopoulos S, Warner J *et al*.: **Application of microdroplet PCR for large-scale targeted bisulfite sequencing.** *Genome Res* 2011, **21:** 1738-1745.

210. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF *et al*.: **A TALE nuclease architecture for efficient genome editing.** *Nat Biotechnol* 2011, **29:** 143-148.

211. Manes S, Mira E, Gomez-Mouton C, Lacalle RA, Keller P, Labrador JP *et al*.: **Membrane raft microdomains mediate front-rear polarity in migrating cells.** *EMBO J* 1999, **18:** 6211-6220.

212. Grande-Garcia A, Echarri A, de RJ, Alderson NB, Waterman-Storer CM, Valdivielso JM *et al*.: **Caveolin-1 regulates cell polarization and directional migration through Src kinase and Rho GTPases.** *J Cell Biol* 2007, **177:** 683-694.

213. Grande-Garcia A, del Pozo MA: **Caveolin-1 in cell polarization and directional migration.** *Eur J Cell Biol* 2008, **87:** 641-647.

214. Murai T: **The role of lipid rafts in cancer cell adhesion and migration.** *Int J Cell Biol* 2012, **2012:** 763283.

215. Yang L, Wang L, Geiger H, Cancelas JA, Mo J, Zheng Y: **Rho GTPase Cdc42 coordinates hematopoietic stem cell quiescence and niche interaction in the bone marrow.** *Proc Natl Acad Sci U S A* 2007, **104:** 5091-5096.

216. Hogart A, Lichtenberg J, Ajay SS, Anderson S, Margulies EH, Bodine DM: **Genome-wide DNA methylation profiles in hematopoietic stem and progenitor cells reveal overrepresentation of ETS transcription factor binding sites.** *Genome Res* 2012, **22:** 1407-1418.

217. Parrish S, Fleenor J, Xu S, Mello C, Fire A: **Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference.** *Mol Cell* 2000, **6:** 1077-1087.

218. Baek D, Vill+®n J, Shin C, Camargo FD, Gygi SP, Bartel DP: **The impact of microRNAs on protein output.** *Nature* 2008, **455:** 64-71.

219. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP: **A coding-independent function of gene and pseudogene mRNAs regulates tumour biology.** *Nature* 2010, **465:** 1033-1038.

220. Beerman I, Bock C, Garrison BS, Smith ZD, Gu H, Meissner A *et al*.: **Proliferation-Dependent Alterations of the DNA Methylation Landscape Underlie Hematopoietic Stem Cell Aging.** *Cell Stem Cell* 2013.

221. Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H *et al*.: **Quantitative comparison of genome-wide DNA methylation mapping technologies.** *Nat Biotechnol* 2010, **28:** 1106-1114.

222. Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G *et al*.: **Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells.** *Nature* 2011, **471:** 68-73.

223. Naveiras O, Nardi V, Wenzel PL, Hauschka PV, Fahey F, Daley GQ: **Bone-marrow adipocytes as negative regulators of the haematopoietic microenvironment.** *Nature* 2009, **460:** 259-263.

224. Tuljapurkar SR, McGuire TR, Brusnahan SK, Jackson JD, Garvin KL, Kessinger MA *et al*.: **Changes in human bone marrow fat content associated with changes in hematopoietic stem cell numbers and cytokine levels with aging.** *Journal of Anatomy* 2011, **219:** 574-581.

225. Visnjic D, Kalajzic Z, Rowe DW, Katavic V, Lorenzo J, Aguila HL: **Hematopoiesis is severely altered in mice with an induced osteoblast deficiency.** *Blood* 2004, **103:** 3258-3264.

226. Conboy IM, Conboy MJ, Wagers AJ, Girma ER, Weissman IL, Rando TA: **Rejuvenation of aged progenitor cells by exposure to a young systemic environment.** *Nature* 2005, **433:** 760-764.

227. Mayack SR, Shadrach JL, Kim FS, Wagers AJ: **Systemic signals regulate ageing and rejuvenation of blood stem cell niches.** *Nature* 2010, **463:** 495-500.

228. Sauvageau M, Sauvageau G. Polycomb Group Proteins: Multi-Faceted Regulators of Somatic Stem Cells and Cancer. Cell Stem Cell 7[3], 299-313. 3-9-2010.

229. Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P *et al*.: **Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing.** *Science* 2002, **298:** 1039-1043.

230. Laible G, Wolf A, Dorn R, Reuter G, Nislow C, Lebersorger A *et al*.: **Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at S.cerevisiae telomeres.** *EMBO J* 1997, **16:** 3219-3232.

231. Hidalgo I, Herrera-Merchan A, Ligos J, Carramolino L, Nu+¦ez J, Martinez F *et al*.: **Ezh1 Is Required for Hematopoietic Stem Cell Maintenance and Prevents Senescence-like Cell Cycle Arrest.** *Cell Stem Cell* 2012, **11:** 649-662.

232. de HG, Nijhof W, Van ZG: **Mouse strain-dependent changes in frequency and proliferation of hematopoietic stem cells during aging: correlation between lifespan and cycling activity.** *Blood* 1997, **89:** 1543-1550.

233. de HG, Van ZG: **Genetic analysis of hemopoietic cell cycling in mice suggests its involvement in organismal life span.** *FASEB J* 1999, **13:** 707-713.

234. Muller-Sieburg CE, Riblet R: **Genetic control of the frequency of hematopoietic stem cells in mice: mapping of a candidate locus to chromosome 1.** *J Exp Med* 1996, **183:** 1141-1150.

235. Muller-Sieburg CE, Cho RH, Sieburg HB, Kupriyanov S, Riblet R: **Genetic control of hematopoietic stem cell frequency in mice is mostly cell autonomous.** *Blood* 2000, **95:** 2446-2448.

236. Jawad M, Cole C, Zanker A, Giotopoulos G, Fitch S, Talbot CJ *et al*.: **QTL analyses of lineage-negative mouse bone marrow cells labeled with Sca-1 and c-Kit.** *Mamm Genome* 2008, **19:** 190-198.

237. Liang Y, Jansen M, Aronow B, Geiger H, Van ZG: **The quantitative trait gene latexin influences the size of the hematopoietic stem cell population in mice.** *Nat Genet* 2007, **39:** 178-188.

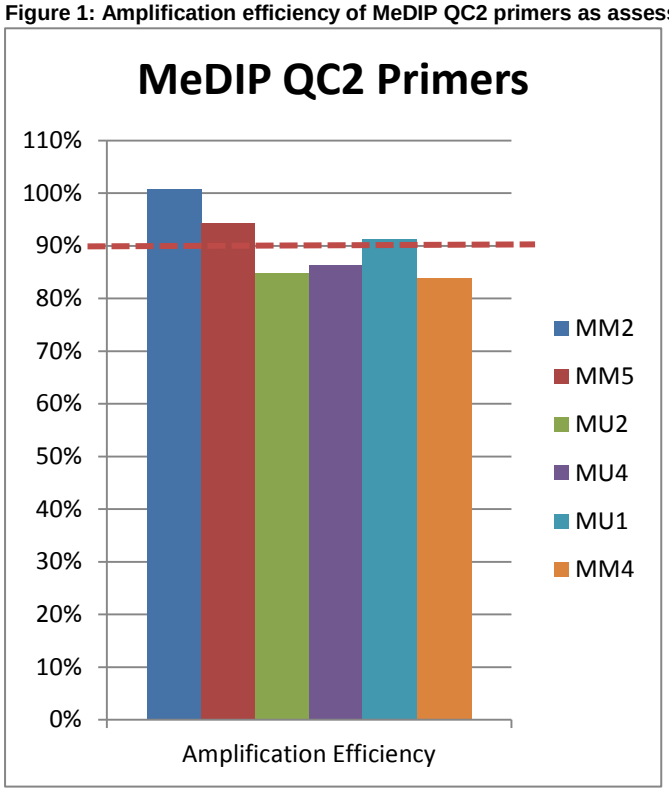
Appendix

Table A-1: Oligonucleotide Sequences

Sequence details of all oligonucleotides that were used in this thesis

MM5_F	Mouse MeDIP QC_qPCR	CATGGCCCACAAAGTAATAAAA
MM5_R		AACGACTTACAACGAGCTCAAA
MU1_F		GGCTAGAACTGACCAGACAGAC
MU1_R		ATCTGTAGCCAATCCTAGAGCA
15CpG_qPCR_F	Lambda MeDIP (spike) QC_qPCR	TATCACTGTTGATTCTCGC
15CpG_qPCR_R		GGTAAAGAGTTTGGATTAGG
20CpG_L_qPCR_F		GGTGAACTTCCGATAGTG
20CpG_L_qPCR_R		CAGTCATAGATGGTCGGT
20CpG_S_qPCR_F		GTTAGAGCCTGCATAACG
20CpG_S_qPCR_R		GAAAGAGCACTGGCTAAC
N4bp3_F	Mouse BS Conversion QC_qPCR	TGGGGTTAGTTTTTATGGAGGATTT
N4bp3_R		ACCTCCAACTACAACTATTTATAACT
Nasp_F		GGCTAGAACTGACCAGACAGAC
Nasp_R		ATCTGTAGCCAATCCTAGAGCA
Sdpr1_F	Mouse *Sdpr* BS pyro-seq primers	GGTTTTTATTAGTAATATAGTGAGTAAGTTG
Sdpr1_R		[Btn]CTCCAATCTCTTCACCTAAAC
Sdpr1_S		GTAAGTTGTTAGAGAAGTTT
PE.Adapter.1.0	Custom Illumina NGS Adapters (Sigma)	ACACTCTTTCCCTACACGACGCTCTTCCGATC*T
PE.Adapter.2.0		[Phos]GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG
PCR.primer.PE.1.0	Custom Illumina NGS PCR Primers (Sigma)	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T
PCR.primer.PE.2.0		CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T

Figure 1: Amplification efficiency of MeDIP QC2 primers as assessed by qPCR of serially diluted DNA



Mouse methylated 5 (MM5) and Mouse unmethylated 1 (MU1) primers were used for MeDIP QC2. These primers show equal amplification efficiencies

Age Differentially Methylated Regions between Young and Old Samples

Table A-2: Age Hypermethylated regions

FDR scores of ≥20, ≥ 13, ≥ 10 and ≥ 7 equals FDR of ≤ 1%, ≤ 5%, ≤ 10% and ≤ 20% respectively

DMR_ chr	DMR_start	DMR_stop	Length	CpG_ count	CpG_density	C_ count	C_density	FDR Score	Near_gene_ID_up_100000 _down_50000	Gene_coordinates	Strand	Associated Gene Name
9	3017681	3018217	537	13	4.841713	117	21.78771	126.3403	ENSMUSG00000074563	chr9:3017408-3019022	+	Gm10719
5	15195610	15197136	1527	10	1.309758	288	18.86051	75.43196	ENSMUSG00000089871	chr5:15186528-15220414	+	Speer7-ps1
5	15052419	15054269	1851	14	1.512696	349	18.85467	54.64918	ENSMUSG00000091897	chr5:15028950-15032998	-	Gm17019
5	15186252	15187052	801	12	2.996255	145	18.10237	47.01436	ENSMUSG00000089871	chr5:15186528-15220414	+	Speer7-ps1
5	14944958	14945908	951	11	2.313354	261	27.44479	37.3376	ENSMUSG00000091255	chr5:14933221-14938429	-	Speer4e
11	58807486	58808395	910	44	9.67033	221	24.28571	36.7908	ENSMUSG00000061462	chr11:58807758-58949904	-	Obscn
1	1.36E+08	1.36E+08	821	11	2.679659	238	28.98904	32.89294	ENSMUSG00000026458	chr1:136193360-136229505	-	Ppfia4
5	15005308	15007171	1864	15	1.609442	353	18.93777	29.10988	ENSMUSG00000072188	chr5:14974113-14978935	-	Gm10354
10	79383519	79384680	1162	82	14.1136	378	32.53012	28.42895	ENSMUSG00000035773	chr10:79379716-79385018	+	Kiss1r
5	14931969	14932505	537	3	1.117318	123	22.90503	28.02147	ENSMUSG00000091255	chr5:14933221-14938429	-	Speer4e
3	1.02E+08	1.02E+08	714	9	2.521008	230	32.21289	27.34599	ENSMUSG00000048540	chr3:101814068-101819415	+	Nhlh2
9	69589616	69590327	712	13	3.651685	148	20.78652	26.24206	ENSMUSG00000087375	chr9:69606547-69678652	+	B230323A14Rik
1	1.22E+08	1.22E+08	562	5	1.779359	134	23.84342	25.86844	ENSMUSG00000026389	chr1:122087334-122169282	-	Steap3
5	64421726	64422178	453	6	2.649007	93	20.5298	24.43467	ENSMUSG00000077209	chr5:64393592-64393723	-	SNORA17
5	14955913	14956552	640	0	0	117	18.28125	21.5311	ENSMUSG00000091255	chr5:14933221-14938429	-	Speer4e
5	15030146	15030726	581	7	2.409639	155	26.67814	21.23281	ENSMUSG00000091897	chr5:15028950-15032998	-	Gm17019
5	1.14E+08	1.14E+08	715	16	4.475524	241	33.70629	21.05837	ENSMUSG00000025825	chr5:114222757-114228297	+	Iscu
8	12133307	12134332	1026	0	0	555	54.09357	20.91275				
1	1.36E+08	1.36E+08	512	5	1.953125	137	26.75781	20.48887	ENSMUSG00000041559	chr1:135933831-135944854	+	Fmod
5	15176301	15177135	835	38	9.101796	259	31.01796	19.31867	ENSMUSG00000089871	chr5:15186528-15220414	+	Speer7-ps1
9	3020610	3021111	502	10	3.984064	109	21.71315	19.11427	ENSMUSG00000079719	chr9:3020155-3021593	+	Gm11167
13	66244325	66245141	817	21	5.140759	116	14.19829	17.61613	ENSMUSG00000074838	chr13:66222141-66222842	+	Gm16512
1	1.32E+08	1.32E+08	835	4	0.958084	164	19.64072	17.55861	ENSMUSG00000042581	chr1:131169879-132115855	+	Thsd7b
5	15039850	15040291	442	6	2.714932	131	29.63801	17.23883	ENSMUSG00000091897	chr5:15028950-15032998	-	Gm17019

19	5756493	5757280	788	44	11.16751	245	31.09137	16.27778	ENSMUSG00000024940	chr19:5740904-5758532	+	Ltbp3
1	1.37E+08	1.37E+08	501	6	2.39521	127	25.3493	15.67206	ENSMUSG00000066885	chr1:137007746-137008883	-	Ptprv
12	24781837	24782470	634	65	20.50473	225	35.48896	15.26373	ENSMUSG00000073158	chr12:24729566-24781833	-	9030624G23Rik
5	15178759	15179583	825	6	1.454545	177	21.45455	14.92996	ENSMUSG00000089871	chr5:15186528-15220414	+	Speer7-ps1
1	1.31E+08	1.31E+08	535	5	1.869159	90	16.82243	14.49169	ENSMUSG00000091877	chr1:130667176-130670998	-	Gm17699
5	14928379	14928962	584	4	1.369863	76	13.0137	14.08238	ENSMUSG00000091255	chr5:14933221-14938429	-	Speer4e
11	1.02E+08	1.02E+08	1036	58	11.19691	314	30.30888	14.00703	ENSMUSG00000006575	chr11:102254717-102263869	+	Rundc3a
4	1.52E+08	1.52E+08	781	5	1.28041	138	17.66965	13.79534	ENSMUSG00000028931	chr4:151764851-151851980	-	Kcnab2
11	51457270	51457950	681	28	8.223201	238	34.9486	12.1521	ENSMUSG00000001053	chr11:51456565-51464344	-	N4bp3
4	1.09E+08	1.09E+08	547	0	0	270	49.36015	12.13526	ENSMUSG00000086483	chr4:108615448-108644682	-	8030443G20Rik
7	56777668	56778256	589	23	7.809847	150	25.46689	12.06944	ENSMUSG00000052512	chr7:56501559-56865457	+	Nav2
2	56961684	56962234	551	13	4.718693	142	25.77132	11.66154	ENSMUSG00000026826	chr2:56959241-56976414	-	Nr4a2
3	78822359	78822885	527	0	0	182	34.5351	11.66154	ENSMUSG00000077487	chr3:78783844-78783973	+	SNORA17
17	29779939	29780556	618	5	1.618123	152	24.59547	11.588	ENSMUSG00000090083	chr17:29751735-29840304	+	Rnf8
5	15159419	15159918	500	2	0.8	134	26.8	11.44099	ENSMUSG00000069720	chr5:15158104-15162877	-	4930572O03Rik
5	1.48E+08	1.48E+08	818	12	2.933985	264	32.27384	11.38917	ENSMUSG00000053129	chr5:148000272-148002523	+	Gsx1
12	23094511	23094977	467	58	24.8394	168	35.9743	11.31543	ENSMUSG00000073164	chr12:22990656-23046943	+	2410018L13Rik
12	56639267	56640004	738	3	0.813008	284	38.48238	11.2481	ENSMUSG00000063129	chr12:56666226-56667883	+	Aldoart2
5	16973699	16974252	554	6	2.166065	147	26.5343	10.6545	ENSMUSG00000058643	chr5:16981916-16986754	+	Speer4f
1	1.37E+08	1.37E+08	598	11	3.67893	129	21.57191	10.38529	ENSMUSG00000009418	chr1:137331157-137481932	-	Nav1
4	43438124	43438886	763	31	8.125819	212	27.78506	10.27773	ENSMUSG00000035969	chr4:43394851-43439960	+	Rusc2
5	14972978	14973486	509	3	1.178782	108	21.21807	10.09287	ENSMUSG00000072188	chr5:14974113-14978935	-	Gm10354
8	1.08E+08	1.08E+08	536	34	12.68657	144	26.86567	10.05293	ENSMUSG00000033249	chr8:107793729-107799745	+	Hsf4
9	1.03E+08	1.03E+08	349	6	3.438395	61	17.47851	9.945994	ENSMUSG00000032531	chr9:102619002-102635748	+	Amotl2
X	34655763	34656411	649	0	0	56	8.628659	9.862435	ENSMUSG00000036572	chrX:34631673-34650335	-	Upf3b
9	97319946	97320447	502	6	2.390438	188	37.4502	9.862435	ENSMUSG00000077208	chr9:97339920-97340046	-	SNORA17
X	1.38E+08	1.38E+08	724	3	0.828729	101	13.95028	9.858895	ENSMUSG00000079418	chrX:137491446-137598813	+	Atg4a
5	64429438	64429915	478	11	4.60251	116	24.26778	9.719966	ENSMUSG00000077209	chr5:64393592-64393723	-	SNORA17
18	47313657	47314225	569	25	8.787346	112	19.68366	9.587079	ENSMUSG00000042705	chr18:47118515-47393938	+	Commd10

13	56143687	56144200	514	2	0.77821	62	12.06226	9.526354	ENSMUSG00000089379	chr13:56182824-56182883	+	AC163338.1
8	12131257	12131756	500	2	0.8	276	55.2	8.871338				
8	74213734	74214292	559	24	8.586762	161	28.80143	8.706437	ENSMUSG00000079019	chr8:74213113-74214474	+	Insl3
X	1.38E+08	1.38E+08	890	1	0.224719	125	14.04494	8.701147	ENSMUSG00000079418	chrX:137491446-137598813	+	Atg4a
1	1.32E+08	1.32E+08	536	4	1.492537	125	23.3209	8.449773	ENSMUSG00000042581	chr1:131169879-132115855	+	Thsd7b
4	1.4E+08	1.4E+08	512	2	0.78125	210	41.01563	7.95336	ENSMUSG00000086960	chr4:139947063-139952519	+	Gm13017
12	1.04E+08	1.04E+08	509	4	1.571709	207	40.66798	7.915029	ENSMUSG00000041669	chr12:104435118-104480360	-	Prima1
5	14905168	14905436	269	3	2.230483	24	8.921933	7.818868	ENSMUSG00000033219	chr5:14910122-14914899	-	Gm9758
11	82743854	82744714	861	127	29.50058	464	53.89082	7.818868	ENSMUSG00000086058	chr11:82738405-82746884	-	Gm12575
8	14534678	14535179	502	2	0.796813	175	34.86056	7.737277	ENSMUSG00000047495	chr8:14095865-14847684	+	Dlgap2
4	1.45E+08	1.45E+08	498	59	23.69478	189	37.95181	7.65151	ENSMUSG00000052912	chr4:145054107-145057323	+	Smarca5-ps
X	1.31E+08	1.31E+08	508	1	0.393701	32	6.299213	7.620456	ENSMUSG00000000223	chrX:130939090-130991112	+	Drp2
1	1.37E+08	1.37E+08	459	3	1.30719	120	26.14379	7.346188	ENSMUSG00000048096	chr1:137221384-137264642	+	Lmod1
5	1.13E+08	1.13E+08	502	6	2.390438	163	32.47012	7.215984	ENSMUSG00000058153	chr5:112848180-113006205	-	Sez6l
11	1.01E+08	1.01E+08	504	25	9.920635	153	30.35714	7.208128	ENSMUSG00000045007	chr11:101017221-101023101	+	Tubg2
5	15169336	15169837	502	3	1.195219	152	30.27888	7.17555	ENSMUSG00000069720	chr5:15158104-15162877	-	4930572O03Rik
5	14995989	14996496	508	0	0	128	25.19685	7.118167	ENSMUSG00000072188	chr5:14974113-14978935	-	Gm10354
1	1.31E+08	1.31E+08	497	3	1.207243	99	19.91952	7.022499	ENSMUSG00000091877	chr1:130667176-130670998	-	Gm17699

Table A-3: Age Hypomethylated regions

FDR scores of ≥20, ≥ 13, ≥ 10 and ≥ 7 equals FDR of ≤ 1%, ≤ 5%, ≤ 10% and ≤ 20% respectively

DMR_ chr	DMR_start	DMR_stop	Length	CpG_ count	CpG_density	C_ count	C_density	FDR Score	Near_gene_ID_up_100000 _down_50000	Gene_coords	Strand	Associated Gene Name
7	1.12E+08	1.12E+08	218	9	8.256881	38	17.43119	158.8895	ENSMUSG00000064907	chr7:111571788-111571894	+	U6
6	18118918	18119201	284	9	6.338028	68	23.94366	40.91621	ENSMUSG00000041301	chr6:18120687-18272768	+	Cftr
6	62964837	62965100	264	7	5.30303	48	18.18182	35.64826				
4	1.24E+08	1.24E+08	849	15	3.533569	204	24.02827	35.59651	ENSMUSG00000028907	chr4:124355404-124370844	-	Utp11l
2	14263017	14263967	951	18	3.785489	237	24.92114	25.83163	ENSMUSG00000036949	chr2:14309943-14416602	+	Slc39a12
11	87029807	87030541	735	18	4.897959	182	24.7619	24.4732	ENSMUSG00000018548	chr11:86940579-87034185	+	Trim37
9	30122131	30122752	622	12	3.858521	176	28.29582	24.24417	ENSMUSG00000064713	chr9:30099453-30099559	-	U6
15	5190435	5191234	800	11	2.75	160	20	23.12683	ENSMUSG00000039942	chr15:5156661-5194187	-	Ptger4
17	34355295	34356229	935	18	3.850267	293	31.3369	20.46584	ENSMUSG00000081512	chr17:34348257-34351404	-	Gm15821
9	34968957	34969885	929	12	2.583423	185	19.91389	20.2128	ENSMUSG00000032040	chr9:34931993-34983646	-	Dcps
17	39983239	39984285	1047	85	16.23687	299	28.55778	19.11427	ENSMUSG00000088272	chr17:39984071-39984165	+	mmu-mir-2133-2
8	67597647	67598343	697	14	4.017217	146	20.94692	17.9525	ENSMUSG00000074302	chr8:67598932-67601379	+	Gm10663
13	95903505	95904385	881	10	2.270148	207	23.49603	15.94277	ENSMUSG00000021684	chr13:95794409-96020291	-	Pde8b
2	11278297	11278847	551	6	2.177858	136	24.6824	14.26289	ENSMUSG00000080905	chr2:11274141-11274396	+	Gm13295
2	11434396	11435180	785	12	3.057325	174	22.16561	13.86083	ENSMUSG00000026773	chr2:11393057-11423728	-	Pfkfb3
2	84651064	84651726	663	24	7.239819	167	25.18854	13.18236	ENSMUSG00000027077	chr2:84651333-84662809	-	Smtnl1
11	1.04E+08	1.04E+08	919	14	3.04679	237	25.7889	13.11545	ENSMUSG00000020689	chr11:104469370-104531790	+	Itgb3
6	65487659	65488061	403	6	2.977667	88	21.83623	11.79247	ENSMUSG00000044162	chr6:65540392-65584034	+	Tnip3
1	79719216	79719541	326	9	5.521472	56	17.17791	11.66154	ENSMUSG00000073643	chr1:79698837-79772718	-	Wdfy1
17	39980211	39981034	824	66	16.01942	254	30.82524	11.66154	ENSMUSG00000076004	chr17:39981081-39981190	+	Mir715
11	46122379	46122890	512	11	4.296875	148	28.90625	11.65005	ENSMUSG00000020340	chr11:46007357-46125852	-	Cyfip2
7	56652863	56653697	835	20	4.790419	221	26.46707	11.566	ENSMUSG00000052512	chr7:56501559-56865457	+	Nav2
7	87918952	87919462	511	12	4.696673	130	25.44031	10.98418	ENSMUSG00000030536	chr7:87856469-87948217	-	Iqgap1
8	97067698	97068372	675	14	4.148148	221	32.74074	10.82393	ENSMUSG00000034361	chr8:97056890-97094429	+	Cpne2
10	1.03E+08	1.03E+08	501	8	3.193613	81	16.16766	10.67865	ENSMUSG00000090638	chr10:102492113-102492978	-	Gm17028
3	53157405	53158119	715	10	2.797203	158	22.0979	10.26196	ENSMUSG00000085174	chr3:53256541-53261719	+	Gm16206

8	8968710	8969450	741	23	6.207827	205	27.66532	10.21662				
9	1.03E+08	1.03E+08	502	5	1.992032	126	25.0996	10.14504	ENSMUSG00000032548	chr9:102910817-102990179	+	Slco2a1
5	1.39E+08	1.39E+08	500	5	2	137	27.4	9.859645	ENSMUSG00000090903	chr5:139461462-139461950	+	Gm17247
12	54823522	54824052	531	9	3.389831	112	21.09228	8.932791	ENSMUSG00000021010	chr12:54349664-55173162	+	Npas3
1	51348728	51349162	435	6	2.758621	80	18.3908	8.922945	ENSMUSG00000045954	chr1:51345970-51359803	+	Sdpr
4	88787486	88788071	586	13	4.43686	137	23.37884	8.591598	ENSMUSG00000062937	chr4:88783026-88826985	+	Mtap
7	38982135	38982636	502	3	1.195219	118	23.50598	8.585563	ENSMUSG00000054676	chr7:38968236-38982582	+	1600014C10Rik
5	1.25E+08	1.25E+08	504	6	2.380952	116	23.01587	8.297465	ENSMUSG00000029401	chr5:124913274-124928375	-	Rilpl2
4	93852965	93853469	505	9	3.564356	93	18.41584	8.214479	ENSMUSG00000085931	chr4:93756267-94092279	-	Gm12648
6	6141710	6142250	541	8	2.957486	152	28.09612	7.907489	ENSMUSG00000015112	chr6:5991219-6167173	-	Slc25a13
6	1.35E+08	1.35E+08	348	10	5.747126	92	26.43678	7.90583	ENSMUSG00000042992	chr6:134589645-134661202	+	Loh12cr1
6	47335738	47336239	502	8	3.187251	122	24.30279	7.371706	ENSMUSG00000029686	chr6:47403397-47476138	+	Cul1
13	46031365	46031859	495	9	3.636364	101	20.40404	7.369538	ENSMUSG00000046876	chr13:45650262-46060345	-	Atxn1
7	88359509	88360005	497	11	4.426559	99	19.91952	7.230497	ENSMUSG00000025584	chr7:88358482-88479419	+	Pde8a

Table A-4: Age differentially expressed cell movement genes.

Process	Upregulated	Downregulated
Cellular Movement	AGTR1	ABR
	ALCAM	ADAM15
	AMICA1	ANGPTL4
	ANKRD28	ANXA2
	APBB2	ARHGAP24
	APOB	CACNA1E
	AR	CAMK1D
	ARFGEF1	CAPN1
	ARID5B	CAPN2
	ARNT	CAST
	BCR	CAT
	BMP4	CCR7
	BMPR1A	CCR9
	C4B (includes others)	CD19
	C5AR1	CD34
	CD28	CD44 (includes EG:100330801)
	CD38	CD86
	CDCP1	CDH5
	CDKN2C	CDK1
	CHRM3	CFH
	CHRNA7	CTGF
	CLASP2	E2F2
	CLU	EBF1
	CMKLR1	EGFL7
	COL18A1	EPB41L5
	CXCR5	EPHB2
	CYSLTR2	EPS8
	DCBLD2	EVL
	DCN	FABP4
	DDR1	FLT1
	DNM2	FLT4
	EFNA1	FOXO3
	EPB41L5	FSCN1
	F10	FUT7
	FAP	GFRA1
	FBLIM1	IL15 (includes EG:16168)
	FBLN5	IL15RA
	FCER2	IL33
	FGFR2	IL6R
	FHL2	ITGA9
	FOSL2	ITGAE
	FOXC2	ITGB5
	FYB	LCK
	GATA3	LEF1
	GNAO1	LGALS1
	GPR183	LSP1 (includes EG:16985)
	GRB7	LTB
	HNRNPA2B1	LY6D
	HOXB9	MCF2
	HRH1	MMP2
	HRH2	MYB
	HTATIP2	NES

HTRA1	NOV
ID1	P2RY2
IL10	PBK
IL18BP	PF4
IL33	PIP5K1A
ISL1	PLEC
ITGA3	POSTN
ITGA6	POU2AF1
ITGAE	PPARG
ITGAV	PPBP
ITGB3	PRKCB
JAG1	PTPRO
KANK1	RAPGEF2
KISS1R	RELN
LAMA5	RGS4
LIF	ROBO4
LMO7	RUNX2
LOX (includes EG:16948)	S100A10
LRP6	SDC1 (includes EG:20969)
LSP1 (includes EG:16985)	SELE (includes EG:20339)
LTBP2	SEMA3A
MAGI2	SLC9A3R2
MAP1B	SPARC
MAPK8IP3	ST3GAL5
MCF2	ST6GAL1
MINK1	STAB1
MMP14	STAT1
MTA1	STMN1
MUC1	TLR9
NAV1	TUBB2B
NBL1 (includes EG:17965)	VIM
NCK1	WNT11
NEO1	ZYX
NR1H4	
NTN4	
PARK2	
PDCD1	
PDE4D	
PDGFD	
PENK	
PLAT	
PLCL1	
PLEC	
PLXNB2	
PMP22	
PPAP2A	
PRKCZ	
PRNP	
PTGS2	
PTK2B (includes EG:19229)	
PTP4A3	
PTPRK	
PTPRU	
RAMP2	

	RASSF5
	ROBO3
	ROR2
	RORC
	SELP
	SEMA3E
	SEMA6D
	SERPING1
	SFRP1
	SKIL
	SORBS3
	SPP1 (includes EG:20750)
	SPRY4
	STAT3
	SYNM
	TAC1
	TACR1
	TGM2
	THBD
	THBS1
	THRB
	TNK2
	TRPM4
	VLDLR
	VWF
	WASF3
	WWTR1
	ZBTB16

Figure 2: Bisulfite conversion efficiency

Plot of BS conversion efficiency in Old (O_1, 2, 3 and 4), Mid (M_1, 2, 3 and 4) and Young (Y_1, 3 and 4) samples. Genomic DNA was also assessed as an untreated control.

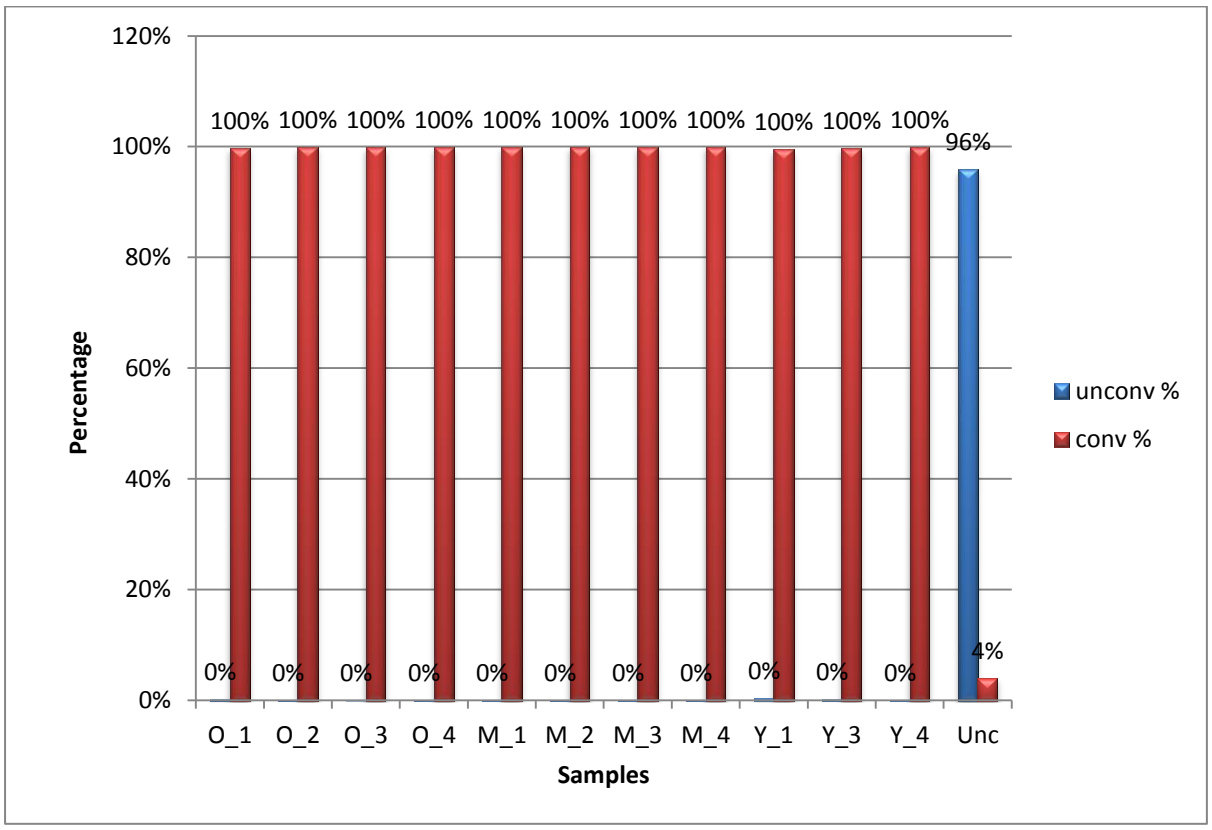
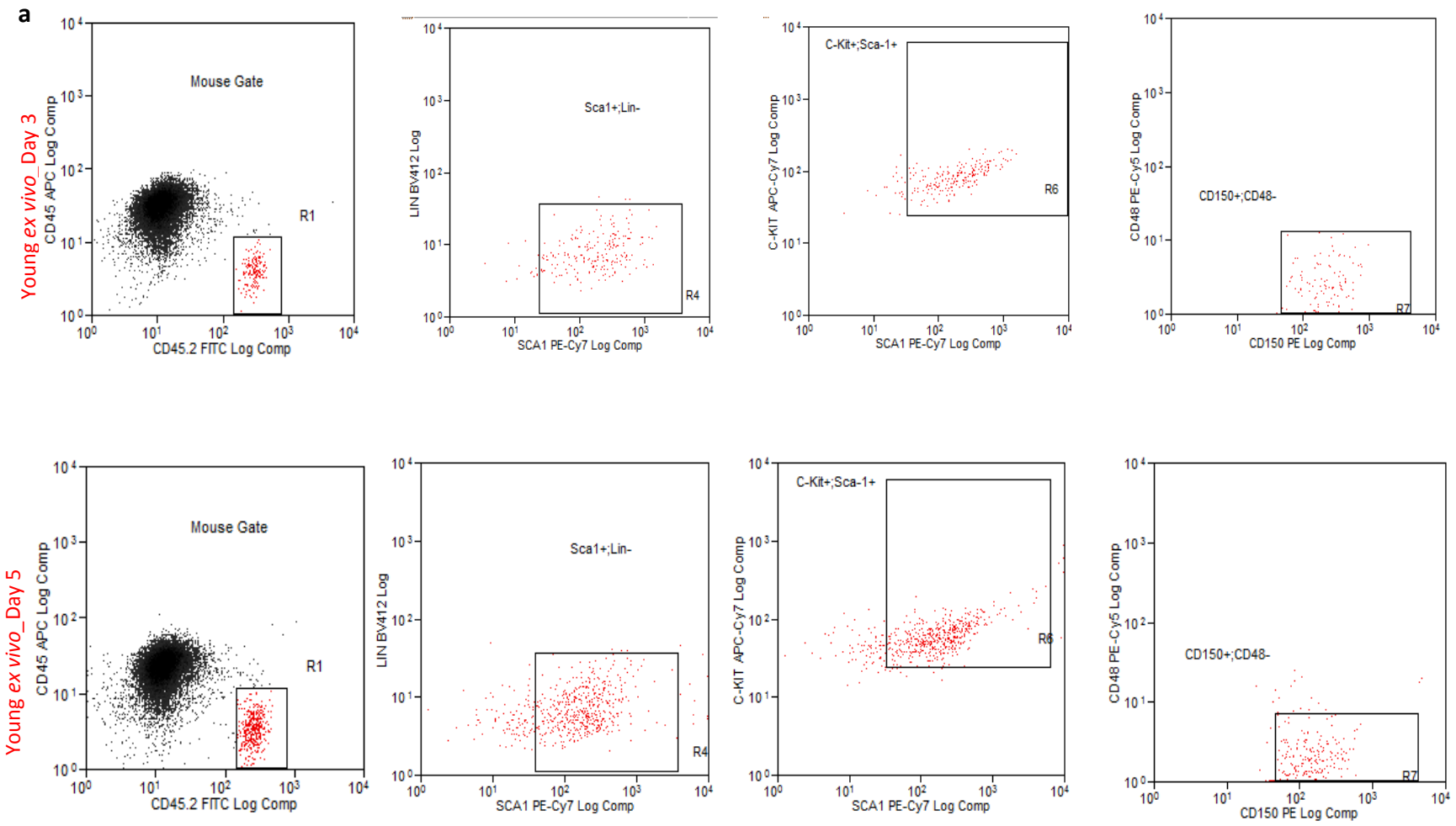


Figure 3: Antibody phenotype of *ex vivo* cultured LT-HSCs
LT-HSC shown are from (a) Young mice after three (top plot) and five (bottom plot) days in culture; (b) Old mice after three (top plot) and five (bottom plot) days in culture.



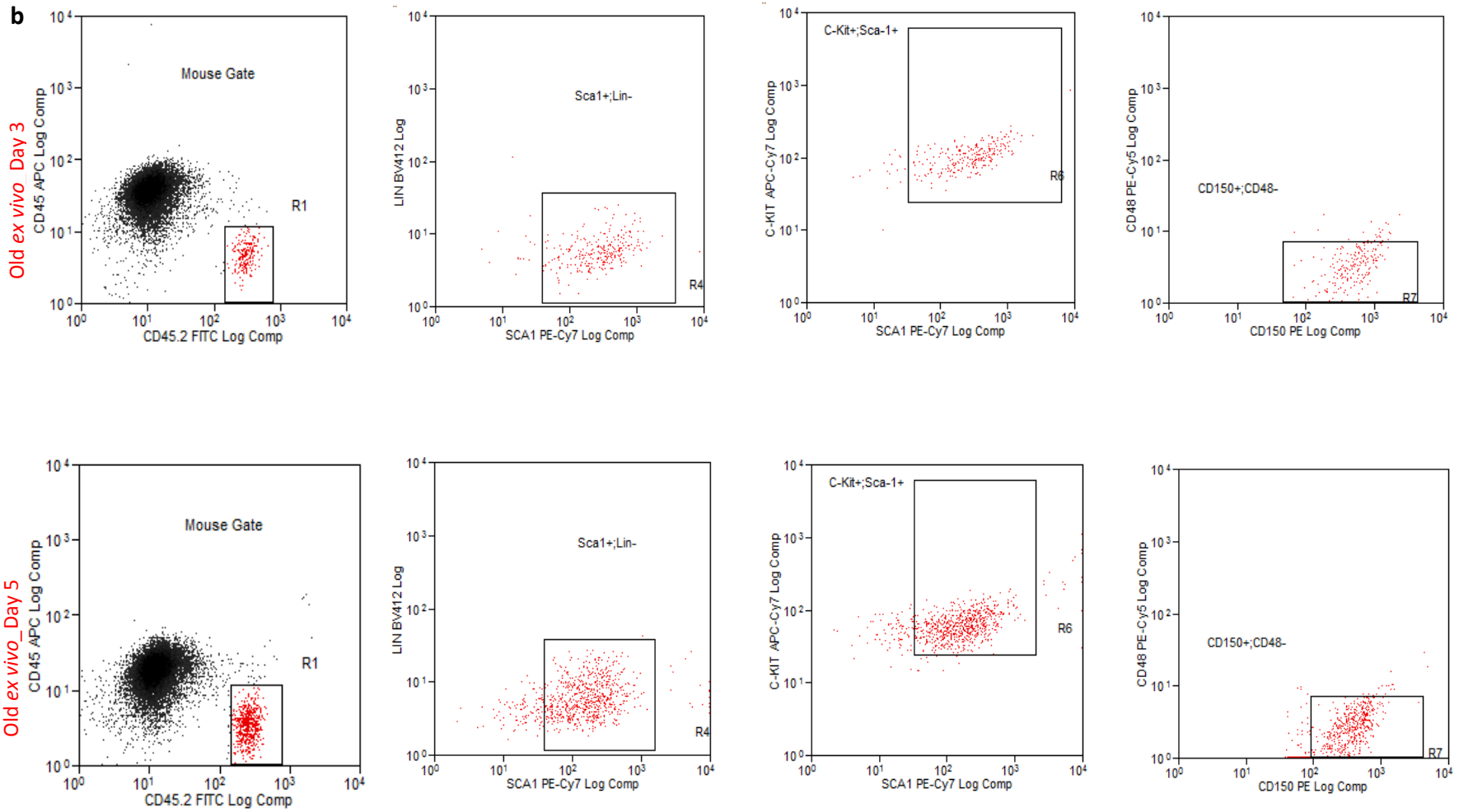

b
Old ex vivo Day 3
Mouse Gate
CD45 APC Log Comp
CD45.2 FITC Log Comp
R1
Sca1+;Lin-
LIN BV412 Log
SCA1 PE-Cy7 Log Comp
R4
C-Kit+;Sca-1+
C-KIT APC-Cy7 Log Comp
SCA1 PE-Cy7 Log Comp
R6
CD150+;CD48-
CD48 PE-Cy5 Log Comp
CD150 PE Log Comp
R7
Old ex vivo_Day 5
Mouse Gate
CD45 APC Log Comp
CD45.2 FITC Log Comp
R1
Sca1+;Lin-
LIN BV412 Log
SCA1 PE-Cy7 Log Comp
R4
C-Kit+;Sca-1+
C-KIT APC-Cy7 Log Comp
SCA1 PE-Cy7 Log Comp
R6
CD150+;CD48-
CD48 PE-Cy5 Log Comp
CD150 PE Log Comp
R7

Figure 4: Assessment of 5-Aza'dC (Aza) treated ex vivo cultured LT-HSCs.
a) Aza treated cells were growth retarded, however their morphology appeared consistent with untreated cells and b) the phenotype of treated cells (1 µM) appeared consistent with primitive LT-HSCs. Cells shown in c-Kit, Sca-1, CD48 and CD150 plots were gated side scatter (R1) and the murine CD45.2 marker (R20)

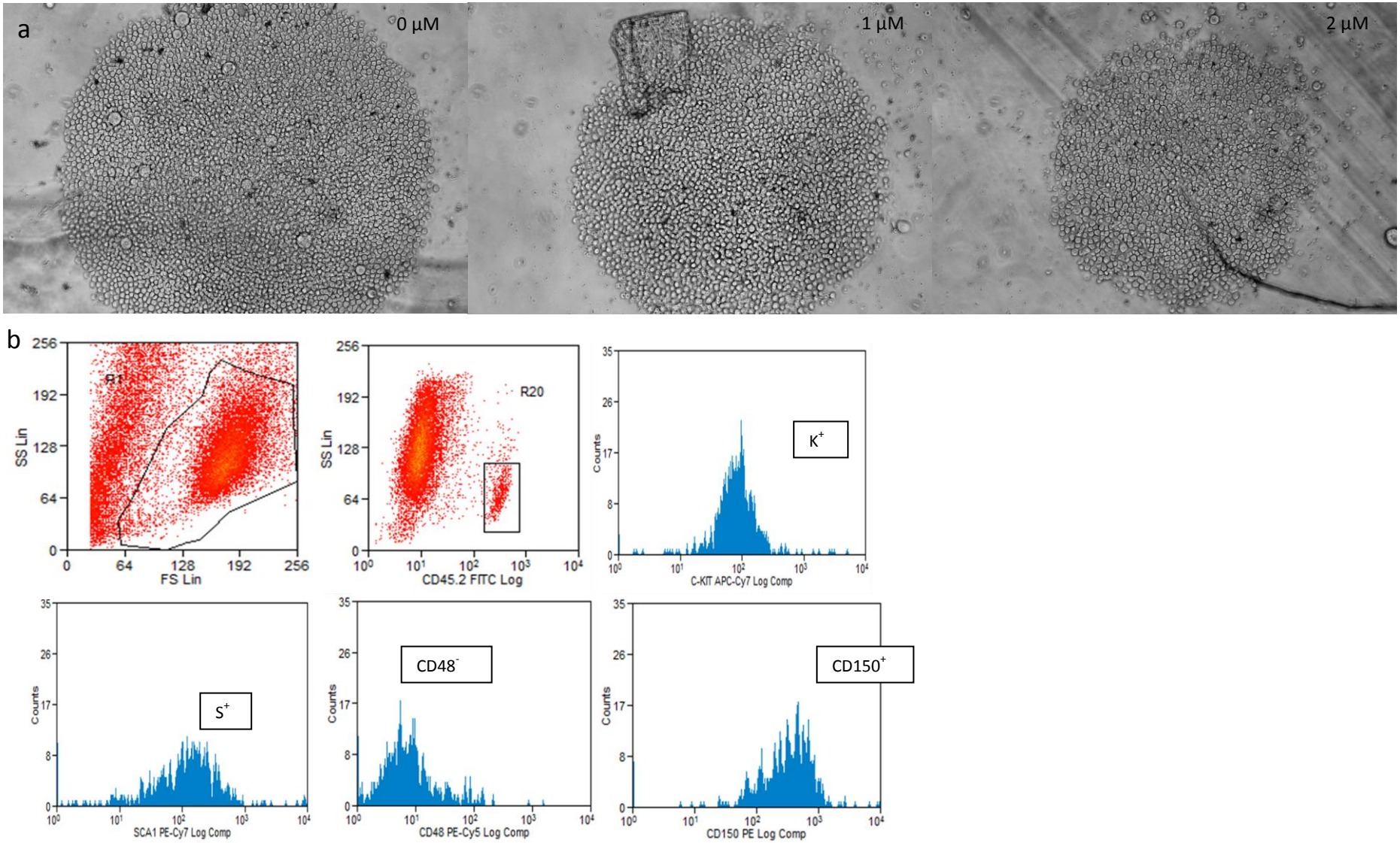


Table A-5: AHR target genes that were age differentially expressed

Ensembl ID	Genes
ENSMUST00000039559	THBS1
ENSMUST00000025875	SLC1A1
ENSMUST00000140739	SLC16A5
ENSMUST00000031594	SDSL
ENSMUST00000035065	PTGS2
ENSMUST00000033941	PLAT
ENSMUST00000164721	NRG4
ENSMUST00000034214	MT2A
ENSMUST00000034215	MT1E
ENSMUST00000002073	LTBP2
ENSMUST00000028735	JAG1
ENSMUST00000155249	ITGA6
ENSMUST00000003818	INSIG2
ENSMUST00000016673	IL10
ENSMUST00000169036	HECTD2
ENSMUST00000117872	FGFR2
ENSMUST00000021603	FBLN5
ENSMUST00000027632	EPB41L5
ENSMUST00000105287	DCN
ENSMUST00000031668	COL1A2
ENSMUST00000044565	COL16A1
ENSMUST00000039267	CHD7
ENSMUST00000063531	CDKN2C
ENSMUST00000097981	CDKN2B
ENSMUST00000050918	CAMK2N1
ENSMUST00000149051	ARNT
ENSMUST00000130229	ACOX1

Figure 5: Multiple transcript isoforms appear to be differentially regulated with age.

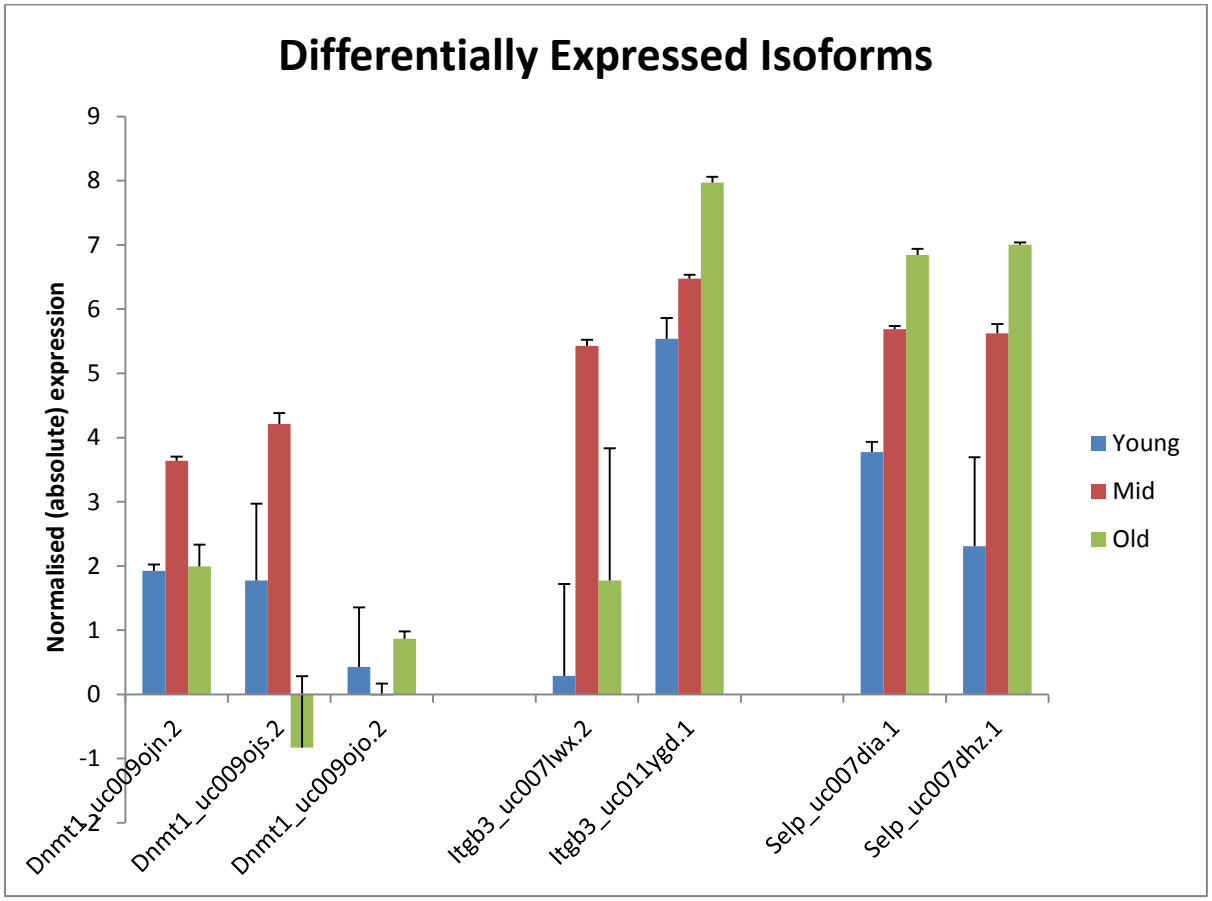


Figure 6: Heatmap of all 111 age differentially methylated regions.

Blue represents regions that are hypermethylated in Old age, while Yellow represents those that are hypomethylated.

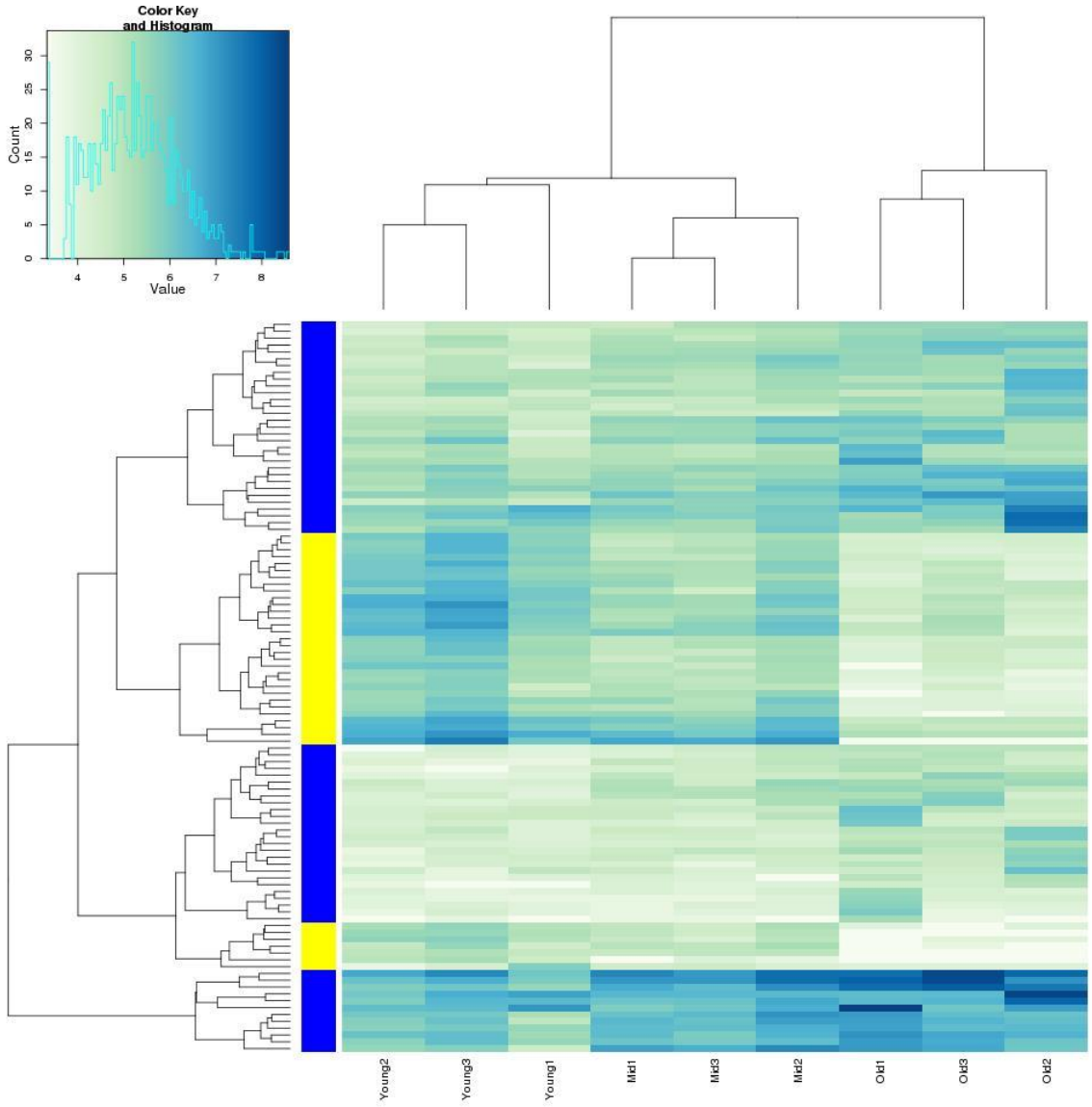


Table A-6: Pearson correlation of MeDIP-seq samples.
Values are from 0 to 1, with 1 being the highest.

MeDIP-seq

Samples	Y1	Y2	Y3	M1	M2	M3	O1	O2	O3
Y1		0.81	0.83	0.83	0.84	0.82	0.91	0.74	0.82
Y2	0.81		0.94	0.92	0.93	0.92	0.83	0.66	0.92
Y3	0.83	0.94		0.93	0.94	0.93	0.85	0.66	0.92
M1	0.83	0.92	0.93		0.93	0.92	0.85	0.67	0.92
M2	0.84	0.93	0.94	0.93		0.93	0.86	0.66	0.93
M3	0.82	0.92	0.93	0.92	0.93		0.84	0.68	0.92
O1	0.91	0.83	0.85	0.85	0.86	0.84		0.73	0.84
O2	0.74	0.66	0.66	0.67	0.66	0.68	0.73		0.68
O3	0.82	0.92	0.92	0.92	0.93	0.92	0.84	0.68	

Table Key

MeDIP_seq Samples	Platform
Y1, O1, O2	GAIIx
Others	Hi-seq
Correlation codes_MeDIP-seq	**Comparison Platform**
Black	GAIIx
Red	Hi-seq
Green	Cross Platforms
Blue fill	Self-correlation