Age-associated changes in promoter CpG island methylation and their potential role in cancer development

Thesis submitted for the degree of Doctor of Philosophy

Hannah Elizabeth Gautrey

Institute for Ageing and Health
Newcastle University

September 2012

Abstract

Changes in DNA methylation patterns are a hallmark of both cancer and ageing, and may underlie susceptibility to developing age-related diseases such as cancer. To uncover the impact that such variation has on health and ageing, we assessed DNA methylation patterns in the peripheral blood leukocytes (PBL) of 480 participants in the Newcastle 85+ study to determine the levels, and degree of inter-individual variation, of DNA methylation in the promoter region of a panel of genes by Pyrosequencing.

We found considerable inter-individual variation in promoter CpG methylation in several genes and a remarkable similarity to leukaemic patterns of aberrant methylation. This included specific methylation of the same sets of genes, strong correlations between methylation of the genes (in a pattern reminiscent of the CpG island methylator phenotype observed in cancer) and the presence of densely methylated alleles in highly methylated individuals, identical to patterns observed in cancer cells. This suggests that ageing and cancer related methylation may be closely linked.

Further analysis of PBL DNA methylation levels in Newcastle 85+ study participants with a previous history of cancer (n=113) versus cancer-free individuals (n=113) found significantly higher methylation in those with a cancer history (10.71% vs. 10.21%, p=0.04). Further, a separate group of 72 individuals diagnosed with cancer within the 3 year duration of the study had similarly increased methylation levels (10.96% vs. 10.36%, p=0.03), suggesting that pre-existing methylation in normal cells may increase risk of cancer and may be evident prior to clinically detectable disease. In addition, individuals with increased DNA methylation were more likely to be categorised as frail (as defined by Fried) than those with lower DNA methylation measures, indicating that disrupted methylation patterns are associated with detrimental effects on healthy ageing.

Subsequently, a GWAS analysis found that SNPs in two genes, *DSCAM* and *DSCAML1*, appeared to be associated with determining DNA methylation levels in the Newcastle 85+ study participants.

Acknowledgements

Firstly, I must thank my two wonderful supervisors, Dr Gordon Strathdee and Professor John Mathers. Without your invaluable guidance, advice and discussion, this thesis would not be much longer than this page. I am also grateful to the BBSRC for providing my studentship.

Thanks to all members of the lab particularly Sanne, Shabnam and Stefano and everyone from the old crucible lab (gone, but not forgotten...), every single one of you helped to make my time in Newcastle so amazing on both the good days and the bad.

I must also graciously thank those individuals who are much more experienced than me in certain areas including Dr Thahira Rahmen for help with high-throughput sequencing, Dr Simon Cockell for bioinformatics analysis, Andrew Kingston and Mohammad Yadegarfar for their assistance with regression models, Dr Joanna Collerton for performing the frailty analysis, Dr Kristin Ayers and Dr Heather Cordell for performing the GWAS analysis, Dr Angela Pyle for determining mitochondrial haplogroups, and Dr Jill McKay for teaching me the wonders of Pyrosequencing what feels like a lifetime ago! Additional thanks go to everyone involved in the Newcastle 85+ study, especially the participants without whom my research would not have been possible.

Thank you to all of my friends and family, both in Newcastle and beyond, for their support during this time and putting up with me even when you don't have a clue what I'm talking about. Particular thanks to my grandparents for everything you've done to help me over the years, including providing the laptop that has worked it's socks off for the past couple of months so I could write this thesis! Special thanks to Emma, Lou, Roisin and Sian as your years of friendship extend way beyond the duration of this thesis.

I simply can't put into words how much I value the encouragement of my parents, Mick and Rosemary. You've been there for me every step of the way and I couldn't ask for a better role model in life than the two of you.

Finally to Scott, you picked up my chin when it was down, always made me laugh and encouraged me to "just keep going". I couldn't have done it without you. I love you $\infty^{(\infty)}$

Table of Contents

List of Tables

List of Figures

List of Abbreviations

5hmC	5'hydroxymethylcytosine
A	adenine
AD	Alzheimer's disease
ALL	acute lymphoblastic leukaemia
ALX	adaptor in lymphocytes of unknown function X/hematopoietic SH2 domain containing
AML	acute myeloid leukaemia
APC	adenomatous polyposis coli
APOE	apolipoprotein E
APP	amyloid beta (A4) precursor protein
APS	adenosine 5' phosphosulphate
ASM	allele specific methylation
ATF2	activating transcription factor 2
ATP	adenosine 5' triphosphate
Bio	biotin
BLAST	Basic Local Alignment Search Tool
B-NHL	B-cell non-Hodgkin lymphomas
bp	base pair
BRCA1	breast cancer 1, early onset
C	cytosine
CACNA1B	calcium channel, voltage-dependent, N type, alpha 1B subunit
CALCA	calcitonin-related polypeptide alpha
CDH1	E-cadherin
CDH13	H-cadherin
CDKN2A	cyclin-dependent kinase inhibitor 2A
CELF4	CUGBP, Elav-like family member 4
CH_3	methyl group
CIMP	CpG island methylator phenotype
c-jun	jun proto-oncogene
CLL	chronic lymphoblastic leukaemia
CML	chronic myeloid leukaemia
c-myc	v-myc myelocytomatosis viral oncogene homolog (avian)
COBRA	combined bisulfite restriction analysis

COMT	catechol-O-methyltransferase
CpA	cytosine – phosphate – adenine
CpC	cytosine – phosphate – cytosine
CpG	cytosine – phosphate – guanine
CpT	cytosine – phosphate – thymine
CREB	cAMP response element-binding
CRC	colorectal cancer
CTCF	CCCTC-binding factor (zinc finger protein)
CV	cardiovascular
CVD	cardiovascular disease
CYGB	cytoglobin
ddH_2O	double deionised water
DIP2B	DIP2 disco-interacting protein 2 homolog B (Drosophila)
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	deoxyribonucleotide triphosphate
DSCAM	Down syndrome cell adhesion molecule
DSCAML1	Down syndrome cell adhesion molecule like 1
dsDNA	double stranded deoxyribonucleic acid
DZ	dizygotic
EBF4	early B-cell factor 4
EDERADD	EDAR-associated death domain
ENCODE	encyclopedia of DNA elements
EPHA10	ephrin type-A receptor 10
ERN2	endoplasmic reticulum to nucleus signalling 2
ESC	embryonic stem cell
ESR1	estrogen receptor 1 (alpha)
EWAS	epigenome wide association study
F2RL3	coagulation factor II (thrombin) receptor-like 3
FBS	fetal bovine serum
FFS	Fried frailty status
FOXA1	forkhead box A1
FZD9	frizzled family receptor 9
G	guanine

GAIIx	Genome Analyser IIx
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GFP	green fluorescent protein
GNASAS	GNAS antisense RNA 1 (non-protein coding)
GP	general practitioner
GPR15	G protein-coupled receptor 15
GR	glucocorticoid receptor
GST	glutathoione-S-transferase
GWAS	genome wide association study
H1/2A/2B/3/4	Histone 1/2A/2B/3/4
H19	imprinted maternally expressed transcript (non-protein coding)
HAND2	heart and neural crest derivatives expressed 2
HAP1	huntingtin-associated protein 1
HapMap	International Haplotype Map Project
HAT	histone acetyltransferase
HCC	hepatocellular carcinoma
HCK	hemopoietic cell kinase
HDAC	histone deacetylase
HDM	histone demethylase
HLXB9	homeobox HB9 also known as motor neuron and pancreas homeobox 1
HMT	histone methyltransferase
HOXA4	homeobox A4
HOXA5	homeobox A5
HOXB6	homeobox B6
HOXD4	homeobox D4
IAP	intracisternal A-particle
ICF	immunodeficiency, centromeric region instability and facial anomalies
IDH1	isocitrate dehydrogenase 1 (NADP+), soluble
IDH2	socitrate dehydrogenase 2 (NADP+), mitochondrial
IGF1/2	insulin-like growth factor 1/2
IL-6	interleukin 6 (interferon, beta 2)
INS	insulin
iPSC	induced pluripotent stem cell
IVM	in vitro methylated
K	Lysine

kb	kilobase
LB	lysogeny broth
LD	linkage disequilibrium
LEP	leptin
LINE	long interspersed nuclear element
MAT1A	methionine adenosyltransferase I, alpha
MBD	methyl binding domain
MBD1/2/3	methyl binding domain containing protein 1/2/3
MBP	methyl binding domain containing protein
MCAM	methylated CpG island amplification and microarray
me/me2/me3	methylation/demethylation/trimethylation
MeCP2	methyl CpG binding protein 2 (Rett syndrome)
MeDIP	methylated DNA immunoprecipitation
MGMT	O-6-methylguanine-DNA methyltransferase
miRNA	micro ribonucleic acid
MLH1	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)
MMR	mismatch repair
MMSE	Mini Mental State Examination
mRNA	messenger ribonucleic acid
MS	methionine synthase
mtDNA	mitochondrial DNA
MTHFR	5,10-methylene tetrahydrofolate reductase
MYOD	myogenic differentiation
MZ	monozygotic
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
ncRNA	non coding ribonucleic acid
NESP55	neuroendocrine secretory protein 55
NK	natural killer
NKX2-5	NK2 homeobox 5
NPTX2	neuronal pentraxin 2
NS	not significant
ONS	Office for National Statistics
ORF	open reading frame
p15	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)

p16	cyclin-dependent kinase inhibitor 2A
p53	tumor protein p53
PBL	peripheral blood leukocytes
PBMC	peripheral blood mononuclear cells
PcG	polycomb group
PCNA	proliferating cell nuclear antigen
PCR	polymerase chain reaction
PDCD5	programmed cell death 5
PEG3	paternally expressed 3
PenStrep	Penicillin and Dihydrostreptomycin
PPi	pyrophosphate
PR	progesterone receptor
PRC	polycomb repressor complex
qPCR	quantitative polymerase chain reaction
RA	rheumatoid arthritis
RARβ2	retinoic acid receptor, beta
RASSF1A	ras association (RalGDS/AF-6) domain family member 1
RFI	Rockwood frailty index
RLGS	restriction landmark genome scanning
RNA	ribonucleic acid
ROS	reactive oxygen species
RPMI	Roswell Park Memorial Institute medium
SAH	S-adenosyl homocysteine
SAM	S-adenosyl methionine
SD	standard deviation
SH2D4A	SH2 domain containing protein 4A
SINE	short interspersed nuclear element
sir2	silent information regulator 2
SiRNA	small interfering ribonucleic acid
Sirt1/4/6	sirtuin 1/4/6
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
Sp1	specificity protein 1
T	thymine
T2D	type II diabetes

TAE	Tris base, acetic acid and EDTA
TCL1B	T-cell leukemia/lymphoma 1B
tDMR	tissue specific differentially methylated region
TE	tris, EDTA
TET	ten-eleven-translocation enzyme
TOM1L1	target of myb1 (chicken)-like 1
TpG	thymine – phosphate – guanine
TSAd	T cell–specific adapter
TSS	transcriptional start site
TUSC3	tumour suppressor candidate 3
TWIST2	twist homolog 2 (Drosophila)
UDG	uracil DNA glycosylase
UHRF1	ubiquitin-like, containing PHD and RING finger domains, 1,
VEGF	vascular endothelial growth factor
VHL	von Hippel-Lindau
WHO	world health organisation
WT	wild type
XAF1	XIAP associated factor 1

Chapter 1 – Introduction

1. Epigenetics

Epigenetics refers to the heritable DNA modifications and gene expression changes which occur independently of any alterations to the underlying DNA sequence (Bird, 2007). The term heritable in this definition refers to the transmission through mitotic division, but not necessarily heritable through the germline. Driven by advances in technology, the interest in studying gene regulation by epigenetics mechanisms such as DNA methylation, histone modifications and non-coding RNA has grown exponentially over recent years.

DNA methylation marks are the mostly widely studied of the epigenetic modifications and are characterised by the addition of a methyl group (CH_3) to the carbon at position 5 of cytosine nucleotides, forming 5'methyl cytosine (Figure 1.1). This typically occurs when cytosine is followed by guanine in the nucleotide sequence, referred to as CpG sites.

The primary role of DNA methylation is to restrict gene expression. Methylation of CpG sites surrounding the transcriptional start site of a gene is typically, but not exclusively, associated with silenced expression, whereas the promoter region of actively transcribed genes is usually free of methylation. However, this is not always the case, and it is clear that the interaction between DNA methylation and histone modifications are important for determining gene activity levels which will be discussed more thoroughly in subsequent sections (illustrated in Figure 1.2).

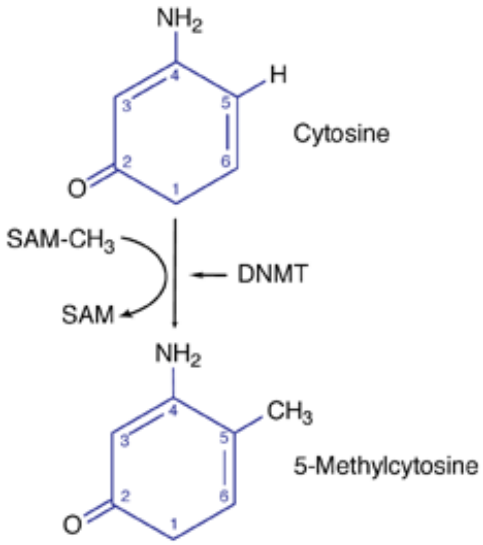


Figure 1.1: Methylation of cytosine forming 5'methylcytosine

S-adenosylmethionine (SAM) acts as a methyl donor for the methylation reaction at the carbon 5 position of cytosine, catalysed by DNA methyltransferase enzymes (DNMTs). Figure from (Strathdee and Brown, 2002).

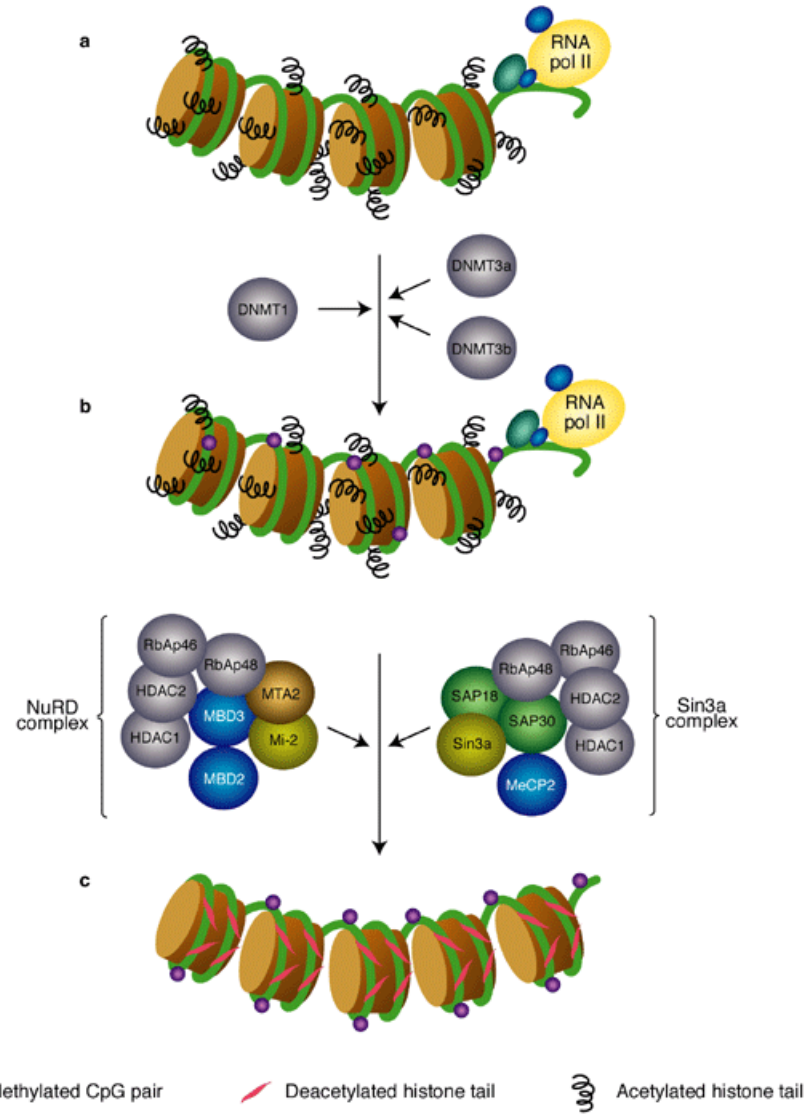


Figure 1.2: Interaction between DNA methylation and histone modification for transcriptional repression

(a) Chromatin in its active form is associated with acetylated histone tails and bound by RNA polymerase II. (b) Methylation at CpG sites is catalysed by DNA methyltransferase (DNMT) and (c) 5' methylcytosine is targeted by methyl binding proteins including MeCP2, MBD2 and MBD3, and large complexes (such as NuRD and Sin3a) containing histone deacetylases (HDAC) and other chromatin remodelling factors are recruited, leading to altered chromatin structure and deacetylation of histone tails. Figure from (Strathdee and Brown, 2002).

1.1 DNA methylation

As described earlier, DNA methylation occurs almost exclusively symmetrically at CpG dinucleotides in the mammalian somatic genome, as the CpG sequence is a signal for the DNA methyltransferase enzymes which catalyse the addition of the methyl group to cytosine nucleotides (McKay *et al.*, 2004).

1.1.1 Non CpG methylation

Although the primary site of DNA methylation is at CpG sites, in certain instances methylation can occur in a Non-CpG context. This can often be non-symmetrical and appears to be largely restricted to embryonic stem cells (ESC) or induced pluripotent stem cells (iPSC), and is very rare in differentiated or somatic cells (Lister *et al.*, 2009). A comparison of ESCs and somatic cells found that levels of CpG methylation are mostly similar between ESCs and somatic cells, however CpA methylation levels were around 15% of the total methylated cytosines, CpT around 3% and CpC approximately 1% in ESCs, which was far more frequent than in somatic cells (Ramsahoye *et al.*, 2000). The presence of non-CpG methylation appears to be independent of DNMT1, since its knockdown in ESCs had no effect on non-CpG methylation levels; however it may be that DNMT3A and DNMT3B are able to compensate for a lack of DNMT1. The authors were able to induce small amounts of CpG, CpA and CpT in Drosophila (which has no endogenous methylation) by expressing DNMT3a, suggesting a role for this particular DNA methyltransferase enzyme in directing non-CpG methylation. This was corroborated by another study which found partial knockdown of DNMT3a in ESC line reduced levels of CpA, CpT and CpC methylation, with little effect on CpG methylation (Ziller *et al.*, 2011). Non-CpG methylation does not appear to be a stable modification, with a large degree of variation in amount of CpA methylation between different ESC lines and the same cell lines at different passage numbers suggesting it is not well conserved (Ziller *et al.*, 2011). The function of non-CpG methylation is unclear; however the fact that it is lost upon differentiation, but recurs after reprogramming to form iPSCs suggests that it may be important for controlling expression of genes associated with differentiation, in order to sustain pluripotency.

1.1.2 DNA methylation patterns across the genome

The incidence of CpG sites is vastly underrepresented in the genome. Cytosine is the most unstable of the four nucleotide bases, susceptible to a deamination reaction forming uracil (Duncan and Miller, 1980). The uracil DNA glycosylase (UDG) enzymes function to cleave uracil bases from the DNA strand, triggering base excision repair (BER) at that site and subsequent replacement with cytosine (Nilsen *et al.*, 2001). However, failure of repair mechanisms can lead to the base pairing of adenine with uracil, and replacement with thymine (Duncan and Miller, 1980). In addition, methylated cytosine is subjected to an increased rate of spontaneous deamination compared with cytosine (Shen *et al.*, 1994) and the direct deamination product of 5'methylcytosine is thymine, which is more difficult to recognise and repair compared to uracil this nucleotide is not foreign to DNA. This results in a number of cytosine to thymine mutations, replacing CpG sites with TpG sites (Coulondre *et al.*, 1978). The consequence of this is that CpG sites are typically infrequent throughout the bulk of DNA.

Mutations involving methylated cytosine are estimated to be responsible for as much as 30% of all mutations occurring in both germline and somatic cells (Shen *et al.*, 1994). For example, *p53* is frequently mutated in a multitude of cancer types, and the frequency of G:C→A:T transitions at methylated CpG sites suggests that deamination of 5'methylcytosine may be responsible for a large proportion of these. In fact, around 50% of germline mutations observed in Li-Fraumeni syndrome correspond to mutation of CpG sites (Greenblatt *et al.*, 1994).

Regions of the genome exist where the frequency of CpG dinucleotides is close to the expected level. These are known as CpG islands, and are often located at the 5' end of annotated genes, in the promoter region or first exon (Saxonov *et al.*, 2006) The functional consequences of DNA methylation, both within and outside CpG islands, are greatly dependent on the genomic location of the modification, which will be discussed more thoroughly in subsequent sections.

1.1.3 CpG Island methylation

A CpG island was originally defined as being larger than 200bp in length, having a CG content greater than 50% and an observed/expected CG ratio of more than 0.6 (Gardiner-Garden and Frommer, 1987). In recent years though, more stringent criteria for defining CpG islands have generally been utilised. One such definition was proposed by Takai and Jones (2002) and stated a CpG island should have a minimum length of 500bp, CG content greater than 55% and observed/expected CG ratio of 0.65 (Takai and Jones, 2002). Numerous other threshold values have been used to predict the number of CpG islands in the genome, which leads to variation in the number reported (Illingworth and Bird, 2009). Repetitive regions of the genome, such as Alu elements, often have a high CG content and observed/expected CG ratio but a more stringent threshold for defining CpG islands has led to the exclusion of many repetitive regions from CpG island analysis (Takai and Jones, 2002).

Studies have suggested that approximately 70% of human gene promoters have a high CpG content (Saxonov *et al.*, 2006), and approximately 50-55% of genes have a defined CpG island crossing the transcriptional start site (Wang and Leung, 2004; Illingworth *et al.*, 2008).

Promoter associated CpG islands are typically free of methylation regardless of expression status (Weber *et al.*, 2007), suggesting that mechanisms exist to prevent aberrant methylation of these regions. The unmethylated status of gene promoters may partly explain why such a large proportion of CpG islands are maintained in these regions, as cytosine which is unmethylated are not subjected to as high a rate of deamination and mutagenesis as methylated cytosine, so maintain higher observed/expected CG ratios (Saxonov *et al.*, 2006).

Housekeeping genes are constitutively active, have a methylation free promoter region and are frequently associated with Sp1 elements, suggesting a possible role for Sp1 in maintaining active transcription and unmethylated promoters (Holler *et al.*, 1988). Subsequently, it has been shown that Sp1 does appear to protect CpG island associated sites from acquiring methylation and may play an important role in establishing correct methylation patterns in early development (Brandeis *et al.*, 1994). There are examples of genes such as E-cadherin (*CDH1*) and von Hippel-Lindau (*VHL*) which contain a

normally unmethylated CpG island surrounded by heavily methylated Alu rich repeat sequences. At the border separating the methylated/unmethylated region there is enrichment for Sp1 binding sites, again implicating a role for Sp1 in controlling against spreading methylation, although protection could be overridden by overexpressing DNMTs (Graff *et al.*, 1997).

Disruption of gene expression levels can induce methylation changes and affect long term expression of target genes. For example, siRNA knockdown of *ESR1* led to decreased ESR1 binding, and the initial recruitment of histone deacetylases (HDACs) and Polycomb repressors. This was later followed by DNA methylation and silencing of multiple downstream targets, including the progesterone receptor (PR). Re-expression of *ESR1* could not trigger the re-expression of PR and subsequent demethylation of the PR promoter was required. This suggests that maintaining active transcription may be required to retain an unmethylated promoter and expression of a gene. Additionally, DNA methylation changes induced by altered expression of regulatory gene can lead to permanent silencing of target sequences (Leu *et al.*, 2004).

Increased levels of the active chromatin mark, dimethylated lysine 4 of histone H3 (H3K4me2), have been described at active, unmethylated promoters. Inactive promoters with H3K4me2 marks are suggested to remain unmethylated, whilst inactive and methylated promoters are not enriched for H3K4me2, suggesting the presence of specific histone modifications may also be involved in protecting against the acquisition of aberrant methylation (Weber *et al.*, 2007).

Enrichment for genomic CpG islands followed by sequencing found that approximately 50% of mammalian CpG islands are at so called “orphan promoters” which are not associated with annotated genes, but share some features of promoters, such as overlapping H3K4Me2/3, a mark of active chromatin. These “orphan promoters” are much more likely to be methylated than the CpG islands associated with annotated gene promoters (Illingworth *et al.*, 2010).

Only a small number of genes have extensive methylation of their promoter regions, predominantly imprinted genes or those subjected to X chromosome inactivation (Weber *et al.*, 2007). Genome wide microarray analysis of CpG islands in human blood have revealed 4% of approximately 5,500 autosomal genes to be densely methylated.

Such methylation was associated with silenced expression in a subset of genes selected for further investigation, and re-expression could be induced by DNA demethylation treatment or DNMT knockout. This implies that DNA methylation, and not other regulatory factors, was responsible for controlling expression at these genes (Shen *et al.*, 2007).

Methylation of CpG islands in exons, introns or between genes is much more frequent than within promoters. A study found that of the 6-8% of CpG islands which were methylated in several human tissues including blood, brain, muscle and spleen, the vast majority were intergenic or intragenic rather than promoter associated (Illingworth *et al.*, 2008). Additional data found that although the majority of these non-promoter CpG islands are methylated (~71% in at least one tissue), a proportion of those which are unmethylated demonstrate binding of RNA polymerase II and show enrichment for the H3K4me3 active chromatin mark, suggesting either that transcription may be originating from some of these sites or that they may function as transcriptional regulators (Rakyan *et al.*, 2008).

Investigating the features of genes which acquire methylation at CpG islands has suggested that the underlying DNA sequence, predicted DNA structure and an enrichment for particular sequence patterns (namely CACC/GGTG) could all predict the susceptibility of the CpG island to becoming methylated (Bock *et al.*, 2006). Other studies have suggested that overall CpG content could also predispose to gains of methylation; genes with an intermediate CpG content, or 'weak CpG islands' appear more likely to be methylated, than those with traditional CpG islands (Weber *et al.*, 2007).

1.1.4 Non-CpG Island methylation

The majority (approximately 80-89%) of CpG dinucleotides outside of CpG islands, in the bulk of DNA, are methylated (Eckhardt *et al.*, 2006; Gibney and Nolan, 2010; Gronniger *et al.*, 2010), but the purpose of such methylation is less clearly understood than CpG island methylation.

The presence of DNA methylation in the intragenic region of a gene has been linked to transcriptional silencing. This was demonstrated by inserting a transgene of the human

p16 promoter which was driving expression of green fluorescent protein (GFP) into murine erythroleukaemia cells. The transgene promoter region was unmethylated whilst the downstream gene body region was methylated. Reduced GFP expression was observed when compared to a completely unmethylated control, and a corresponding lack of association with RNA polymerase II, absence of H3K4me2/3, the active chromatin marks at the methylated region and the formation of tightly packaged chromatin; all of which is incompatible with transcriptional elongation (Lorincz *et al.*, 2004). Conversely, DNA methylation in the gene body region of loci in human B cells has been correlated with active expression, rather than repression (Rauch *et al.*, 2009). A specific example is the active form of the X allele in females which demonstrate a greater proportion of gene body methylation after X inactivation has occurred (Hellman and Chess, 2007). However, there is no evidence that such DNA methylation present in gene bodies is directly influencing transcription.

Other studies have linked non-CpG methylation to determining tissue specificity. Analysis of CpG shores, located several kilobases (kb) from CpG islands, revealed an association between differential methylation of these sites and tissue specific differences in gene expression. The hypermethylated CpG shores of silenced genes within specific tissues showed upregulated expression with either 5'aza-2'deoxycytidine treatment, or after knockout of both *DNMT1* and *DNMT3B*, suggesting methylation of these regions is regulating expression, at least in part. Additionally, CpG shores showed greater extent of aberrant DNA methylation in cancer tissue, than CpG dinucleotides within an island or gene promoter, and therefore may be responsible for some of the dysfunction gene expression associated with cancer, but this remains to be verified (Irizarry *et al.*, 2009).

1.1.5 DNA methylation at repetitive regions

A large proportion of the human genome is represented by two main types of repetitive DNA sequences, which are usually densely methylated and silenced. Tandem repeats, are highly repetitive simple nucleotide repeats including satellite DNA, minisatellites and microsatellites; usually associated with centromeric or heterochromatic regions. Interspersed repeats are typically transposable elements, able to induce recombination events in the genome, comprising of either short interspersed nuclear elements (SINEs) or long interspersed nuclear elements (LINEs). Most of these are artefacts and are no longer able to induce transpositions even if not silenced by methylation (Weisenberger

et al., 2005). However, the requirement for silencing of these elements is notable in some rare cases of human cancer where the re-activation of transposable elements disrupts important genes, such as the *Myc* oncogene in breast cancer (Morse *et al.*, 1988) and the *APC* tumour suppressor in a colorectal tumour (Miki *et al.*, 1992), both of which were disrupted by the insertion of a mobile LINE-1 element. Hypomethylation of LINE-1transposable elements is frequently observed in tumours compared to normal cells, such as hepatocellular carcinoma, but this was not associated with increased expression (Lin *et al.*, 2001). Conversely in chronic myeloid leukaemia (CML), loss of LINE-1 methylation was related to activation of transcription, and poor prognosis (Roman-Gomez *et al.*, 2005). Despite this, there is no direct evidence for demethylation of transposable elements resulting in their activation.

1.1.6 DNA Methyltransferases

Three major functional DNA methyltransferase enzymes have been described; DNMT1, DNMT3A and DNMT3B. These enzymes are characterised by the presence of an N-terminal domain, responsible for protein localisation to the nucleus and interactions with other proteins, and the C-terminal active domain responsible for DNA association and active catalytic function (Jurkowska *et al.*, 2011).

1.1.6.1 DNMT1: Maintenance methylation

In somatic cells, the faithful inheritance of DNA methylation in successive cell divisions depends mainly on the mitosis-linked enzyme DNA methyltransferase 1 (DNMT1), first identified in 1988 (Bestor *et al.*, 1988). DNMT1 has a high preference for the hemimethylated CpG sites which are present on the newly synthesised DNA strand after replication, as opposed to completely unmethylated DNA (Goll and Bestor, 2005). This is likely to be mediated via UHFR1 protein which is comprised of domains able to bind both hemimethylated DNA and DNMT1 (Bostick *et al.*, 2007). DNMT1 is localised with newly replicated DNA at the replication fork during S phase (Leonhardt *et al.*, 1992) and is possibly recruited there via a direct interaction with part of the replication machinery, PCNA (Schermelleh *et al.*, 2007). The symmetry of CpG dinucleotides in a complementary sequence ensures that the pattern of methylation can be retained after DNA replication and the gene expression patterns of the progeny cells reflect those of the parental cell (Wigler *et al.*, 1981; Stein *et al.*, 1982). However, the

ability of DNMT1 to authentically copy methylation patterns is imperfect under certain conditions, such as in CpG dense repetitive regions and the de novo methyltransferases (discussed below) are also required for maintaining methylation (Leonhardt *et al.*, 1992).

Maintaining methylation is essential for development in mice; knockout of *Dnmt1* causes abnormalities in growth, severely reduced levels of DNA methylation and embryonic lethality at mid gestation (Li *et al.*, 1992). Initial experiments of *DNMT1* knockout in a human colorectal cancer cell line (HCT116) resulted in a global reduction of methylation level by approximately 20%. However, these cells were able to continue to survive and proliferate, and the p16 tumour suppressor gene promoter remained densely hypermethylated and silenced (Rhee *et al.*, 2000). Subsequent investigations revealed that this method failed to produce incomplete knockouts and instead led to the formation of hypomorphs, resulting in only partial loss of function. A truncated protein is produced which can still function as a methyltransferase, albeit less efficiently, despite no longer being able to associate with the replication fork, explaining the limited phenotypic effects observed in the previous study (Egger *et al.*, 2006). When complete inactivation of the catalytic domains of *DNMT1* was induced in HTC116 cell lines, the presence of hemimethylated DNA and a small reduction in global DNA methylation was observed. In addition, the majority of cells stopped proliferating. Of those which did proliferate, there was a high frequency of chromosomal abnormalities, which triggered a DNA damage response and these cells ultimately underwent apoptosis. This suggests that DNMT1 is essential for proliferation and survival in human cells (Chen *et al.*, 2007).

1.1.6.2 DNMT3A and DNMT3B: De novo methylation

There are two types of methyltransferases which are responsible for the establishment of de novo methylation in vivo, DNMT3A and DNMT3B. Whilst these enzymes are primarily responsible for de novo DNA methylation, they also demonstrate some affinity for hemimethylated sites. Both *Dnmt3a* and *Dnmt3b* are expressed at high levels in undifferentiated cells and embryos in the mouse, which decreases upon differentiation and in adult cells, suggesting these enzymes may function to completely re-establish genome wide methylation patterns in the early stages of development (Okano *et al.*, 1998). Both proteins are essential; complete inactivation of *Dnmt3a* and

Dnmt3b in mouse embryonic stem cells and embryos resulted in failure to establish de novo methylation after the genome wide demethylation which occurs in very early development. Knockout of *Dnmt3a* resulted in apparently healthy offspring, which soon exhibited stunted growth and death within 4 weeks of birth. Knockout of *Dnmt3b* was lethal and failed to generate any live offspring, with embryos demonstrating numerous defects in development. Double *Dnmt3a* and *Dnmt3b* knockouts had a more severe phenotype which was lethal extremely early on in development. This suggests some overlapping functions of the two methyltransferases in early development, but not at later stages (Okano *et al.*, 1999). The genomic targets for DNA methylation by these enzymes appear to be slightly different. For example Dnmt3b, but not Dnmt3a, knockouts demonstrated clearly reduced methylation of satellite repeat sequences, suggesting such sequences are a unique target of Dnmt3b (Okano *et al.*, 1999).

In humans, mutations of DNMT3B affecting the catalytically active part of the enzyme are responsible for an extremely rare disease termed Immunodeficiency, Centromeric region instability and Facial anomalies syndrome (ICF syndrome). ICF syndrome is characterised by chromosome breakage and reduced levels of DNA methylation particularly at heterochromatin surrounding the centromeres. Patients with this syndrome demonstrate severely impaired immunity due to reduced levels of B-cells, T-cells and serum immunoglobulins, and typically die in early childhood due to susceptibility to infection (Ehrlich, 2003). Additionally, mutations in *DNMT3A* are frequently observed in acute myeloid leukaemia (AML) patients and linked with poor prognosis. A recent study demonstrated disrupted methylation processivity in these patients, providing a possible explanation for some of the highly aberrant methylation patterns observed in AML (Holz-Schietinger *et al.*, 2012).

1.1.6.3 DNMT3L

DNMT3L protein is highly related to the de novo methylating enzymes although not a functional methyltransferase itself. DNMT3L is unable to bind DNA but has been shown to increase the activity of mouse Dnmt3a and Dnmt3b by approximately 1.5-3 fold (Suetake *et al.*, 2004). Crystallography imaging suggests DNMT3L interacts with chromatin structure through the binding of unmethylated H3K4. However, this interaction could be completely prevented when H3K4 tails were trimethylated, a mark of active chromatin, proposing a possible mechanism by which actively expressed genes

are protected from de novo methylation (Ooi *et al.*, 2007). DNMT3L also binds the catalytic domain of DNMT3a (Jia *et al.*, 2007) In addition, DNMT3L has been shown to improve the processivity of DNMT3A. Experiments show that approximately 5-8 consecutive CpG sites are methylated by DNMT3A before a significant proportion of the enzyme becomes unbound, whereas the presence of DNMT3L increases the number of sites progressively methylated, proposing an important role to ensure spreading of methylation within a particular region (Holz-Schietinger and Reich, 2010). DNMT3L function is also important for establishment of imprinting patterns. Disturbance of *Dnmt3l* expression in mouse embryonic stem cells is lethal to female homozygotes part way through gestation and subsequent analysis of embryos revealed a loss of maternal methylation and transcriptional control at imprinted loci. Homozygous males fail to produce sperm (Bourc'his *et al.*, 2001).

1.1.6.4 DNMT2

Despite similarity to other DNMTs, knockout of *Dnmt2* has no effect on DNA methylation in mice, and human DNMT2 has been shown not to methylate DNA, rather acting to methylate a form of transfer RNA (Goll *et al.*, 2006).

1.1.7 DNA methylation during development

After the zygote is formed at fertilisation, there is a genome wide demethylation of DNA, removing parental methylation patterns and allowing the cell totipotency. At implantation, the blastocyst is subjected to a round of de novo methylation, dependent on DNMT3A and DNMT3B, to establish correct methylation patterns which are maintained in the somatic cells. The formation of the germ cells require an additional round of methylation loss and de novo methylation in order to establish correct marks of imprinting dependant upon parental origin (Jurkowska *et al.*, 2011).

1.1.8 DNA methylation and gene expression

1.1.8.1 Relationship between DNA methylation levels and transcription

Unmethylated gene promoters are associated with gene transcription whereas the presence of dense hypermethylation of the region surrounding the transcription start site of genes is strongly concomitant with transcriptional silencing (Eckhardt *et al.*, 2006;

Bell *et al.*, 2011). DNA methylation was determined to be responsible for a large proportion of gene silencing by experiments with DNMT inhibitors such as 5'azacytidine, leading to demethylation and re-expression of genes (Christman, 2002). Not all unmethylated genes are actively transcribed in every cell (Weber *et al.*, 2007), and expression likely depends on the binding of additional transcriptional activators. For example, unmethylated gene promoters of active genes are more frequently associated with RNA polymerase II binding close to proposed transcriptional start sites (Weber *et al.*, 2007). The effect of partial methylation at gene promoters on gene trasncription is unclear and is likely to be position dependent. Silencing of the *XAF1* gene has been found in both gastric and colon cancer, with DNA methylation at specific CpG sites in the promoter demonstrating differing ability to repress expression (Zou *et al.*, 2006). In addition, the differential methylation status of a single CpG site in the *IL-6* promoter region between rheumatoid arthritis (RA) patients and controls was related to expression levels of the protein (Nile *et al.*, 2008). However, the methylation differences at these sites have not been proven to be causative for the observed discrepancies in gene expression.

1.1.8.2 Mechanisms of DNA methylation based gene silencing

1.1.8.2.1 Transcription factor binding

One manner in which DNA methylation is thought to repress gene expression is by preventing the direct physical interaction between transcription factors and other activators of transcription with such promoters. Methylation of CpG sites within the target sequences of protein binding sites is thought prevent the recognition by the appropriate proteins. Specific examples include methylation at the imprinting centre on the paternally inherited allele of *IGF2*, which prevents binding of the CTCF insulator protein. CTCF is therefore unable to block initiation of transcription at the *IGF2* promoter driven by upstream enhancers and the gene is expressed (Bell and Felsenfeld, 2000). Additionally, a second study found that DNA methylation in the promoter of both the mouse and human insulin gene (specifically at the cAMP responsive element) repressed transcription and preventing binding of ATF2 and CREB proteins, which are thought be involved in promotion of insulin transcription (Kuroda *et al.*, 2009). However, only a relatively small number of transcription factors are unable to bind

their target sequence due to the presence of methylation (Chen *et al.*, 2011), and thus alternative methods for transcriptional silencing must exist.

Transcriptional control at the single gene level may be partly achieved by the ability of certain transcription factors to associate with gene promoters and form a direct interaction with the de novo DNA methyltransferase enzymes Dnmt3a and Dnmt3b, recruiting them for site specific methylation at particular loci. In the following study, interaction between Dnmt3a and the c-myc transcription factor could trigger the methylation of CpG sites in c-myc boxes located in downstream target genes, such as *CDKN2A*. In total, 42 transcription factors interacting with both Dnmt3a and Dnmt3b were described, proposing a possible role for transcription factors in driving hypermethylation (Hervouet *et al.*, 2009). Once methylation has been established it is likely followed by the by the recruitment of either proteins which bind methylated DNA, or proteins which modify chromatin structure, in order to repress individual genes (Gibney and Nolan, 2010). The DNA methyltransferases have also been shown to associate with a variety of chromatin modifying enzymes, suggesting these enzymes may have a silencing function in addition to the catalysis of DNA methylation (Klose and Bird, 2006).

1.1.8.2.2 Interaction between DNA methylation and histone modification

DNA contained within cells is packaged around histone proteins to form chromatin. The tails of these proteins are subjected to a variety of post-translational modification, which govern the resulting chromatin structure. An alternative mechanism for control of gene transcription involves the interaction between both DNA methylation and these histones modifications via the recruitment of histone remodelling complexes which will be discussed later in further detail (Section 1.2.5).

1.1.9 Imprinted genes, tissue & gender specific methylation patterns

1.1.9.1 Imprinted genes

Imprinted genes demonstrate monoallelic expression due to DNA methylation and silencing of one copy of the gene, dependent on parental origin (Li *et al.*, 1993). Such imprinting is established upon formation of the germline cells and is critical for normal

development and growth. Several genetic diseases are caused by defects in imprinted regions including Prader-willi syndrome, Angelman syndrome and Beckwith-Wiederman syndrome (Butler, 2009). One example of an imprinted gene is *IGF2*, which regulates foetal growth. The alleles are differentially methylated such that the paternally derived allele is expressed, whilst expression is repressed from the maternal allele. This regulation appears to play a key role in controlling offspring size, as when a mutated copy of *Igf2* was paternally inherited in mice, the resulting offspring were dwarfs (DeChiara *et al.*, 1991).

1.1.9.2 Tissue specific methylation patterns

All cells types contain the same underlying DNA sequence, but have a multitude of diverse functions. It has already been determined by genome wide analysis that different tissues can have vastly diverse methylation patterns and it is hypothesised that tissue specific methylation (in addition to other epigenetics modifications) may be responsible for regulating tissue specific gene expression profiles. However, it is unclear the extent to which tissue specific patterns of DNA methylation are required for the establishment and maintenance of tissue specific gene expression.

In humans, only a small proportion of tissue specific methylation appears to occur in gene promoter regions, with rest located within gene bodies (Eckhardt *et al.*, 2006). Analysis of chromosome 1 revealed DNA methylation differences in several tissues including heart, liver, lung and separate brain regions which were conserved between individuals. Closer investigation demonstrated that the vast majority of the variability in DNA methylation was predominantly in regions outside of CpG islands (De Bustos *et al.*, 2009). Other studies have shown that certain genes can have differential methylation of their CpG islands, including those associated with the 5' end of the gene, in different tissues of the same individual, suggesting methylation may impact upon tissue specific expression (Ghosh *et al.*, 2010). Analysis of promoter associated CpG islands which were methylated in a tissue specific manner in blood, brain, muscle and spleen found an enrichment for genes involved in developmental processes and transcription factors such as the HOX gene family (Illingworth *et al.*, 2008). A comprehensive genome wide methylated DNA immunoprecipitation (MeDIP) combined with microarray analysis suggested that around 18% of all genomic regions could be defined as tissue specific differentially methylated regions (tDMRs) in a range of somatic tissues and sperm. Such

tDMRs were observed across all genomic regions, including promoters, but were most frequent at lower CpG density promoters. Hypomethylated DNA unique to sperm was responsible for 27% of all tDMRs. A comparison of tDMRs and gene expression levels in individual tissues found promoter tDMRs were negatively correlated with gene expression, implying a fundamental role for such methylation in determining tissue specific expression profiles. An association between higher methylation levels in gene bodies and upregulated expression was also observed although the significance of this is unclear. Finally, gene ontology analysis showed that tDMR promoters were over-represented by genes with a tissue specific function, adding evidence that methylation is imperative for defining tissue specific differences (Rakyan *et al.*, 2008).

1.1.9.3 Gender and Methylation

Numerous investigations have suggested that DNA differences in methylation exist between males and females in addition to those due to X inactivation. Global DNA methylation is usually estimated by measuring methylation levels at repetitive elements as a surrogate marker, since they contribute the greatest proportion to total methylation levels. The reported differences of global methylation levels between the sexes are often conflicting. Several investigations of healthy volunteers have found significantly increased methylation of LINE-1, or both LINE-1 and Alu in males (El-Maarri *et al.*, 2007; Zhang *et al.*, 2011); however, others reported that males have lower methylation of Alu than females (Zhu *et al.*, 2012). This inconsistency may arise due to the different techniques used for methylation analysis. At the level of single genes, a study of a healthy Singapore Chinese population found significantly higher methylation of *CALCA, MGMT* and *MTHFR* in males even though these genes had relatively little methylation in any individual (Sarter *et al.*, 2005). The imprinted genes *PEG3, NESP55* and *H19* also demonstrated a slight tendency towards higher methylation in males although this was not significant (El-Maarri *et al.*, 2007). A genome wide study of over 20,000 CpG sites in saliva samples revealed a total of 580 sites which were differentially methylated between males and females, of which approximately 53% were more highly methylated in females and 47% in males. The study also looked at 1,298 CpG sites in human blood and only 36 demonstrated gender differences, 21 sites were higher in females and the remaining were increased in males. The overlap of genes showing sex specific methylation in both tissues was minor (Liu *et al.*, 2010). This data

suggests that neither sex is particularly predisposed to increased methylation at a global level; however gender variation may be loci and tissue specific.

1.1.10 DNA demethylation

As previously described, the complete removal of DNA methylation marks is essential in germline cells, in order to remove and re-establish marks of imprinting. Additionally, the loss of methylation is a key feature of both cancer and ageing cells. The passive mechanism by which methylation marks are removed from DNA typically follows from the reduced activity of DNMT1, leading to failure to recognise the hemi-methylated CpG sites produced on newly synthesised DNA strands during progressive rounds of replication, causing cytosine nucleotides to remain unmethylated after replication. DNA methylation inhibitors such as 5'azacytidine typically induce passive demethylation as they become incorporated into the DNA, causing DNMT1 to become irreversibly bound and unable to function (Christman, 2002).

The presence of an active mechanism for DNA demethylation was inferred from the rapid demethylation which exists in early development, and although many speculated about the manner by which active demethylation of DNA was occurring; there was little firm evidence until relatively recently. Unlike the passive loss of DNA methylation, active demethylation is not reliant upon replication and requires enzymatic activity to remove methylated cytosine and replace with cytosine. The identification of an enzyme with demethylation activity remained elusive until the discovery of the Ten-Eleven-Translocation (TET) enzymes in 2009, which actively convert 5'methylcytosine into 5'hydroxymethylcytosine (5hmC) (Tahiliani *et al.*, 2009). The presence of 5hmC appears dependent on the prior existence of 5'methylcytosine as there is a complete absence of 5hmC in *Dnmt1*-/-, *Dnmt3a*-/-, *Dnmt3b*-/- triple knockout mouse ESCs (Ficz *et al.*, 2011). Furthermore, the TET enzymes are involved in the formation of other intermediates; 5'formylcytosine and 5'carboxylcytosine from methylated cytosine. This outlines a plausible pathway for the replacement of 5'methylcytosine by cytosine via a decarboxylation reaction, which as yet remains to be elucidated (Ito *et al.*, 2011). IDH1/2 mutations which lead to the inactivation of TET2 are often observed in myeloid malignancies alongside genome wide hypermethylation. A methylation array analysis of approximately 14,000 gene promoters revealed that methylation levels at promoters were significantly higher in IDH1/2 mutant acute myeloid leukaemia (AML) compared

to wild type AML patients, generating a hypothesis that higher levels of aberrant methylation accumulate in cells with dysfunctional TET, due to an ineffective mechanism for demethylation of DNA (Figueroa *et al.*, 2010).

A genome wide study in mouse ESCs found both high expression of TET1 and TET2 enzymes and high levels of 5hmC widely spread throughout the genome, although this was enriched at CpG islands and promoters and associated with increased transcription. Decreasing TET levels during differentiation or by TET1 and TET2 knockdown was associated with declining 5hmC at promoters of developmental associated promoters and accompanied by increasing 5'methylcytosine and loss of gene expression (Ficz *et al.*, 2011), suggesting a possible additional role for 5hmC in maintaining pluripotency. However, others have found that TET1 may repress transcription, by recruitment of a Sin3a histone deacetylase complex. Additionally, only ~10% of TET1 target genes change expression following TET1 knockdown and a greater proportion of these genes were up regulated, contradictory to previous findings (Williams *et al.*, 2011).

In addition to the demethylation produced by the oxidation of methylcytosine by the TET proteins, an alternative mechanism for demethylation is though to involve DNA repair. The growth arrest and DNA damage, or *Gadd45a*, gene was identified from expression screening for possible DNA demethylases in *Xenopus laevis*. In addition, active DNA demethylation was shown to be mediated by DNA repair processes via the interaction of *Gadd45a* with XPG, the endonuclease responsible for the 3' incision of damaged DNA during nucleotide excision repair (NER) (Barreto *et al.*, 2007). Since Gadd45a demonstrates no catalytic activity by itself, the mechanism behind its role in demethylation appears to be via DNA repair pathways. Several models have been hypothesised as to how Gadd45a and additional cofactors can mediate demethylation. The proposed mechanisms involve either the excision of methylated cytosines and repair (NER) or by the deamination of methylated cytosine followed by base excision repair (BER) (Niehrs and Schafer, 2012).

1.1.11 Trangenerational inheritance of DNA methylation

There is debate as to whether epigenetic transgenerational inheritance exists in mammals, whereby epigenetic memories can be transmitted to subsequent generations via the germline cells. It seems unlikely as most DNA methylation and histone

modifying marks are eradicated during the meiotic formation of the germ cells (Sasaki and Matsui, 2008), however the epigenetic state of some regions, such as the Intracisternal A-particle (IAP) at the agouti variable yellow (A^{vy}) and axin-fused ($Axin^{fu}$) loci in mice, can be partially inherited by offspring due to incomplete erasure of methylation marks at these loci (Morgan *et al.*, 1999; Rakyan *et al.*, 2003). There is little indication that such transgenerational inheritance exists in humans, and it remains difficult to determine as there is high degree of genetic variability in the population and it is now apparent that underlying genetic sequence is important for defining methylation patterns (Morgan and Whitelaw, 2008). If any transgenerational inheritance could be shown to exist in humans, then this could significantly impact upon DNA methylation levels at birth and the accumulation of DNA methylation during ageing.

1.2 Histone Modification

1.2.1 Histone Structure

In all mammalian cells, the DNA is compacted and packaged into chromatin, consisting of repeating components termed nucleosomes. Nucleosomes consist of a core of key histone proteins; an octamer of H2A, H2B, H3 and H4 to which 147bp of DNA is bound, with the linker histone H1 separating repeating nucleosomes. Histones are positively charged and so form a tight interaction with the negative charge of DNA phosphate groups producing dense chromatin (Cedar and Bergman, 2009). The N-terminal tails of these histone proteins are exposed in the area surrounding the nucleosomes and are subjected to a multitude of post translational modifications including acetylation, methylation, phosphorylation and ubiquitinylation (Grant, 2001). The presence of such modifications alters the structure and function of chromatin, subsequently leading to changes in gene expression. The many combinations and permutations caused by the different types of modification possible at various amino acid residues of different histones have generated a hypothesis that they may form a complex “histone code”, whereby a particular pattern of histone modifications is translated by recruitment of specific proteins in order to exert a downstream effect (Agalioti *et al.*, 2002; Lennartsson and Ekwall, 2009).

1.2.2 Histone acetyltransferase

The most frequently studied modification is the acetylation of lysine residues which is executed by histone acetyltransferase (HAT), a large family of enzymes with multiple catalytic subunits. HATs can form large multi-protein complexes with other subunits, enabling specific targeting to unique target sites (Lee and Workman, 2007). Presence of histone acetylation is well known for its disruptive effect on the conformation of tightly compacted chromatin (Garcia-Ramirez *et al.*, 1995) and is associated with active transcription, possibly by making the DNA within nucleosomes more accessible for transcriptional machinery binding. Nucleosomes with experimentally hyper-acetylated histone tails within nucleosomes, demonstrated an increased rate transcriptional elongation compared to nucleosomes with histones in their native state (Protacio *et al.*, 2000), suggesting the presence of acetylation may ease the recruitment, or passage, of proteins necessary for transcription along the DNA strands.

1.2.3 Histone deacetylase

Histone deacetylase (HDAC) form a large family of enzymes, which have the opposite function to HATs, acting to remove acetyl marks from lysine residues and encourage the reformation of dense chromatin. HDACs have a key role in transcriptional suppression and are vital for mammalian development and cellular differentiation. Disruption of *Hdac1* in mice is lethal at mid gestation and associated with aberrant up regulation of genes controlling cell cycle progression (Lagger *et al.*, 2002). Zupkovtiz et al (2006), also described a subset of genes which were abnormally regulated in mouse ESCs lacking *Hdac1* compared to wild type (WT) cells, the majority of which were up regulated. There was direct recruitment of HDAC1 to target gene promoters with repressed transcription in WT, and increased levels of acetylation at H3 and H4 of these same target genes in the absence of *Hdac1* (Zupkovitz *et al.*, 2006). This suggests a robust link between HDAC binding, decreased acetylation and reduced expression.

1.2.4 Methylation of histones

The affect of methylation at histone residues can depend on the position of the affected amino acid. Several are associated with transcriptional activation including trimethylation of lysine 4, or dimethylation of lysine 79 on H3 (H3K4Me3 and

H3K79Me2), which are both found at high levels at the promoter region of expressed genes (Schneider *et al.*, 2004; Schubeler *et al.*, 2004). Others are linked to gene silencing due to their association with repressed promoters including trimethylation of lysine 9 or 27 on H3 (H3K9Me3 and H3K27Me3) (Barski *et al.*, 2007). The methylation of lysine 36 on histone H3 (H3K36Me), is catalysed by the Set2 methyltransferase, which is linked to RNA polymerase II. Both di- and trimethylation of H3K36 are localised to the 3' end of the open reading frame (ORF) and these marks are recognised by part of the RpD3S HDAC complex, leading to deacetylation and localised compaction of chromatin structure at the ends of coding regions, in order to supress spurious initiation transcription from within the ORF (Carrozza *et al.*, 2005). These examples highlight the complex nature of histone modifications such that methylation of histone may promote, supress or refine gene expression.

1.2.5 Relationship between DNA methylation, histone modification and gene expression

As previously described, gene activity is thought to be determined, at least in part, by the interaction between both DNA methylation and histone modifications via their effect on chromatin structure. The presence of methylated DNA indirectly represses gene transcription by the binding of methyl binding domain containing proteins (MBPs) including MBD1, MBD2, MBD3 and MeCP2. These proteins specifically recognise and bind 5'methylcytosine. MBPs silence gene expression by their interaction with co-repressor proteins and recruitment of large complexes of HDACs and chromatin remodelling factors, to the DNA, facilitating the formation dense heterochromatin which is incompatible with transcription (Jones *et al.*, 1998) (Figure 1.2). Such multi-protein complexes include NuRD (Tyler and Kadonaga, 1999) and Sin3A (Nan *et al.*, 1998). The recruitment of MBPs is an important mechanism by which methylated genes are transcriptionally silenced, as demonstrated by cells lacking MeCP1 which are unable to effectively repress methylated genes (Boyes and Bird, 1991). In addition, the DNA methylation induced recruitment of chromatin modifiers and subsequent changes to chromatin structure may prevent the physical interaction of DNA and transcription machinery.

1.2.6 Histone modification in embryonic stem cells

In ESCs, key genes associated with development have been defined by the presence of bivalent chromatin, marked by both active (methylated H3K4) and repressive (methylated H3K27) histone modifications in the same region. Such genes are poised for either rapid activation or inactivation. In pluripotent cells these genes are typically expressed at low levels (Bernstein *et al.*, 2006). The existence of bivalent chromatin in ESCs is associated with targets of Polycomb group (PcG) proteins, which form Polycomb repressor complexes (PRCs). PRC2 catalyses the methylation of H3K27, which is then accompanied by the recruitment of PRC1 members. This leads to the formation of repressive chromatin (Cao *et al.*, 2002). However, this repression is not permanent; upon differentiation, bivalent promoters have the potential to be activated or silenced, usually retaining only one of the modifications depending on transcriptional status (Cui *et al.*, 2009). Demethylation of H3K27 by histone demethylase (HDM) is likely to be responsible for activation of lineage specific genes; whereas long term silencing is associated with loss of methylated H3K4 and the local recruitment of DNA methylation (Surface *et al.*, 2010).

1.3 Non-coding RNA

Non-coding RNA (ncRNA) does not form a functional protein product; however it has been demonstrated to play an important role in chromatin remodelling and regulation of gene expression.

1.3.1 Role of non-coding RNA (ncRNA)

One function for ncRNA is the X inactivation which occurs in every female mammalian cell. One copy of the X chromosome is randomly condensed into extremely dense heterochromatin via the X inactivation centre, which includes the non-coding RNA Xist. The inactive X is covered by Xist RNA which prevents binding of transcriptional machinery and permanent silencing is achieved by recruitment of chromatin modifying proteins including PRCs, followed by DNA methylation (Wutz and Gribnau, 2007). ncRNAs have also been implicated in the control of monoallelic expression at imprinted loci and silencing transposons (Costa, 2008). Additionally, ncRNA may regulate expression of the HOX genes, which are located within four clusters on separate

chromosomes. Analysis of the HOX clusters found 231 ncRNAs originating from these regions, which seem to play a role in long-range regulation of expression of genes at distance sites. For example, HOTAIR, a trans-acting ncRNA within the HOXC locus is able to interact with PRC2 and trigger H3K27 methylation at the HOXD locus to repress expression (Rinn *et al.*, 2007).

1.3.2 microRNA (miRNA)

Small single stranded molecules, approximately 22bp in length are termed microRNA (miRNA). It is estimated that approximately 30% of genes may be regulated by miRNAs, which are thought to exert their action via recognition and binding of mRNA sequences, but several alternative mechanisms underlying such action have been proposed. It is thought that the number of potential targets of any specific miRNA is large (Nilsen, 2007). The potential importance of miRNA based regulation is emphasised by a number of diseases associated with dysfunctional miRNA including the overexpression of miR-21 and a reduction of miR-125a in tumour cells, and altered miR-1 and miR-133 in cardiovascular disease (Lu *et al.*, 2008), although direct causation is yet to be determined.

1.4 Epigenetics and cancer

The epigenome of a cancer cell is highly abnormal, typically demonstrating reduced levels of DNA methylation overall compared to its normal counterpart due to loss of methylation in the bulk of the genome. This is accompanied by localised increases of DNA methylation, particularly at promoter CpG islands and accompanied by gene silencing, particularly at tumour suppressor genes (Ehrlich, 2002). Additionally, histone modifications are frequently altered in tumour cells, although this could simply be reflecting gene expression changes.

1.4.1 Altered histone modifications and cancer

Aberrations in the pattern of histone modifications have been described in cancer compared to normal cells. One particular genome wide study found a loss of H4 lysine 13 acetylation and trimethylation of H4 lysine 20 (H4K13ac and H4K20me3) in cancer cells which was associated with loss of DNA methylation at repetitive stretches of

DNA. These losses were observed in early stages of malignancy and accumulated with advancing disease in a mouse skin cancer model (Fraga *et al.*, 2005b). Silencing of specific genes is associated with increased H3K9 methylation and reduced methylation of the active mark H3K4 (Nguyen *et al.*, 2002), in addition to the trimethylation of H3K27 mediated by PcG proteins (Kondo *et al.*, 2008). However, these likely represent modifications which occur as a consequence of the altered expression state of genes after transformation, rather than being directly involved in the initiation of cancer associated expression changes. Overexpression of HDACs results in repression of tumour suppressors, likely due to the removal of activating acetylation marks, and has been frequently associated with poor prognosis in cancer (Kawai *et al.*, 2003; Song *et al.*, 2005; Hayashi *et al.*, 2010). In addition, the overexpression of several histone methyltransferase (HMT) enzymes responsible for addition of a variety of repressive marks have been noted in several types of cancer including leukaemia, breast and prostate and are responsible for the addition of a variety of repressive marks. Such changes are thought to play a key role in inactivation of key tumour suppressor genes (Sharma *et al.*, 2010).

Recent analyses involving whole exome sequencing of human cancer cells has revealed the presence of mutations within proteins which form part of histone remodelling complexes. These mutations have been associated with subsequent disruptions to histone modification patterns and chromatin structure with potential downstream effects on gene expression (You and Jones, 2012). A specific example is *EZH2*, a component of the polycomb repressor complex 2 (PRC2) which is responsible for the repressive methylation of H3K27. Overexpression of *EZH2* is often associated with invasive and metastatic disease in breast and prostate cancer, and acquired mutations within *EZH2* have been identified in lymphoid and myeloid leukaemia, where the resulting effect was increased function of the gene (Chase and Cross, 2011). This example highlights the potential overlapping role between genetic mutation and epigenetic alterations in cancer development.

The presence of frequent abnormalities in histone modification patterns in cancer cells has led to a great interest in the use of therapies for reversal, with the hope of treating the disease. Various inhibitory compounds have been subjected to clinical trials and elicited diverse consequences for tumour cells including apoptosis, stimulation of differentiation and growth arrest at the G1/S phase (Carew *et al.*, 2008). HDAC

inhibitors have been investigated in an attempt to prevent the abnormal loss of acetylation observed in cancer. Despite the exact mechanisms of action being unknown, the outcomes of clinical trials utilising HDAC inhibitors have demonstrated good success (Wagner *et al.*, 2010). HDAC inhibitors are able to effectively induce the re-expression of some genes silenced in cancer cells (Yang *et al.*, 2000), however, it appears that only a proportion of genes are affected by HDAC inhibitors in this manner and a number of loci are actually suppressed (Richon and O'Brien, 2002). In addition, since these enzymes can also act as deacetylases for other proteins, such as α-tubulin and p53, the effects of inhibition may not be mediated through altered histone modification (Glozak *et al.*, 2005). Inhibition of HMT enzymes has also been shown to effectively re-express genes silenced in AML, including *p15* (Lakshmikuttyamma *et al.*, 2010).

1.4.2 Hypomethylation and cancer

Hypomethylation affects a much larger proportion of the genome than hypermethylation but the functional consequence of such DNA methylation loss in cancer is much less clear than the role of hypermethylation. Sequences which are susceptible to hypomethylation in cancer are those which are normally highly methylated including retrotransposons, satellite DNA, imprinted genes and much more rarely, single copy genes (Ehrlich, 2002). It has been suggested that the hypomethylation of retrotransposable elements in cancer may be associated with their increased expression, initiating transposition events and genome instability, although there is limited evidence to support such a hypothesis (Daskalos *et al.*, 2009). A possible functional consequence of global hypomethylation in cancer is suggested by ICF syndrome (mutations in DNMT3B) which is accompanied by loss of DNA methylation at repetitive sequences and chromosome instability (Ehrlich, 2003), drawing similarities with the genetic instability which is a hallmark of a cancer cell (Rodriguez *et al.*, 2006). Gene specific DNA methylation loss has been described much more rarely, however there are some examples such as the hypomethylation of growth regulating genes in colorectal neoplasms compared to their normal counterparts (Goelz *et al.*, 1985); and although reduced methylation levels of the oncogenes *c-Myc* and *c-Jun* was associated with up regulated expression in mouse liver tumours, the effect of *c-Myc* hypomethylation in human colorectal cancer and hepatocellular malignancies is undetermined (Ehrlich, 2002).

1.4.3 Hypermethylation and cancer

Hypermethylation of DNA typically occurs at CpG islands. This has been much more widely studied in cancer and has clearly important functional consequences, as hypermethylation of promoter associated CpG islands leads to transcriptional inactivation of the linked gene. Many genes have been shown to be inactivated by this mechanism in cancer, including well established tumour suppressors and genes which regulate the DNA damage response and cell cycle control (Baylin, 2005). The epigenetic marking of these genes provides an alternative mechanism of silencing in addition to classical genetic mutations which are well described in cancer (Jones and Baylin, 2002). Indeed studies suggest that gene silencing by hypermethylation may be a more frequent event than inactivation of genes by DNA sequence mutation (Schuebel *et al.,* 2007). Germline mutations in several tumours suppressors including *MLH1* and *BRCA1* greatly increase the risk of developing a familial colorectal or breast cancer respectively (Bronner *et al.,* 1994; Lancaster *et al.,* 1997), and these genes are also frequent targets for inactivation by hypermethylation in sporadic tumours (Herman *et al.,* 1998; Esteller *et al.,* 2000a).

In addition to these single gene methylation events, the methylation of multiple genes in a tumour cell is a recognised phenomenon referred to as the CpG island methylator phenotype (CIMP), initially identified in colon cancer (Toyota *et al.,* 1999). This phenotype refers to the simultaneous methylation of multiple loci in a subset of tumours. Such tumours are liable to hypermethylate several important tumour suppressor genes concurrently and contain numerous transcriptionally inactivated genes. Whether there is a specific defect causing the concordant aberrant methylation of specific loci in these tumours is undetermined.

The suggestion that certain genes may be more susceptible to acquiring methylation has gained support from several recent papers describing a link between PcG target genes and DNA methylation. PcG proteins form a complex which acts as a transcriptional repressor by inducing suppressive histone modifications in undifferentiated cells (Klauke and de Haan, 2011). Results from genome wide methylation arrays have shown that genes which acquire cancer associated hypermethylation are more likely to be targets of PcG complex binding in stem cells (Ohm *et al.,* 2007). This finding suggests

that PcG complexes may have a specific role in the induction of aberrant methylation during cancer development, and may also alter the susceptibility to DNA methylation of a subset of genes to methylation in different cell types, as regulation of genes by the PRC2 can be cell type specific (Margueron and Reinberg, 2011).

1.5 Epigenetics and ageing

We are living in an ageing population. The world health organisation (WHO) estimates that between the year 2000 and 2050, the number of over 60s in the world will vastly increase from 600 million to 2 billion individuals. With the passage of time, molecular changes are accrued which are often detrimental to the normal activity of cells and organs, meaning the elderly often experience frailty, a term for the collection of several functional impairments (Fulop *et al.*, 2010). There is often a high degree of variability with regards to the degree of frailty experienced on an individual level (Collerton *et al.*, 2009). The role of altered epigenetics in ageing and in age-related disease is unclear, but could play a significant part in determining health status, or frailty, particularly in later life.

1.5.1 Histone modifications and ageing

The potential role of histone modifications in regulating lifespan was proposed by observations of decreased lifespan in *S. cerevisiae* when the HDAC, sir2 was deleted. Conversely, up regulation of sir2 dramatically extends the lifespan and replicative potential of the yeast. The activation of sir2 orthologs in other species such as *C. elegans* and *D. melanogaster* produces an equivalent lifespan effect, however the exact mechanisms responsible for this effect are unclear (Longo and Kennedy, 2006). In mammals, recent data has shown that the lifespan of male mice can be significantly extended by overexpression of *Sirt6*. Altered *Sirt6* expression triggered differential expression of many downstream genes and reduced the activating phosphorylation levels of members of the IGF1 signalling pathway (Kanfi *et al.*, 2012). Disruption of this pathway has previously been demonstrated to have a role in determining longevity (Holzenberger *et al.*, 2003). In various model organisms, multiple other enzymes including HMT and histone demethylases have been shown to affect levels of methylation at specific lysine residues, leading to aberrant gene expression and altered lifespan (Berdasco and Esteller, 2012). In mammals, more specific ageing associated

histone modifications have been uncovered, such as a genome wide increase in H4 lysine 20 trimethylation (H4K20me3) in old mouse tissues compared to young (Sarg *et al.*, 2002) and such changes have been implicated with functional decline. For example, aged mice have been shown to have decreased H4 lysine 12 acetylation levels in the hippocampus and an associated failure to express downstream gene targets associated with learning and memory (Peleg *et al.*, 2010).

The role for altered histone modifications in age-related diseases has been proposed. For example, altered histone acetylation patterns and an increase in phosphorylated H3 have been detected in Alzheimer's disease (Mastroeni *et al.*, 2011). However, these observations have not been shown to be causative in the disease process. Treatment of age-related diseases with epigenetic therapies has attracted attention. In cardiovascular disease (CVD), the use of HDAC inhibitors stimulates changes within cells of the arterial wall, such as reduced VEGF signalling and decreased macrophage adherence to the endothelium, leading to inhibition of atherosclerosis, but whether the mechanism of action occurs via alterations to histone proteins is unclear (Ordovas and Smith, 2010).

1.5.2 DNA methylation and ageing

The altered pattern of DNA methylation occurring throughout ageing has mostly been investigated by comparisons of young and old individuals, but despite this, the consequences of such methylation changes are not well understood.

1.5.2.1 Global DNA methylation and ageing

Global 5'methylcytosine content appears to be reduced in several organisms, including mouse, rat and humans, as a consequence of the ageing process suggesting genome wide hypomethylation is occurring (Vanyushin *et al.*, 1973; Wilson *et al.*, 1987; Fuke *et al.*, 2004). However, a measure of methylated cytosine content creates no assessment of the genomic location of methylation loss; therefore it is not possible to determine the consequences of the age related reduction in global methylation from these studies. Also, some studies have failed to find any effect of age on global DNA methylation levels. For example, an analysis of entire chromosomes involving various tissue types and ages suggested that there was no effect of age on global CpG methylation content (Eckhardt *et al.*, 2006). However, this study used pooled data from age and sex matched

samples meaning the effect of age on DNA methylation levels may be lost due to variation, with some individuals displaying more hypomethylation, whilst others may be more susceptible to gene specific hypermethylation

1.5.2.2 DNA methylation at repetitive regions and ageing

In addition, it has been reported that the epigenetic control of retrotransposons may be disrupted with age. Methylation of an IAP located in the mouse *m.nocturnin* gene was progressively lost throughout ageing, which was associated with its activation (Barbot *et al.*, 2002).

1.5.2.3 Genome wide DNA methylation and ageing

A comparison of DNA methylation patterns at the extreme ends of human lifespan by whole genome bisulfite sequencing, and methylation array containing 450,000 CpG sites, revealed that centenarian DNA was hypomethylated overall compared to newborn DNA samples. This decreased methylation was seen in gene bodies and intergenic regions, as well as at promoters, although these were mostly associated with tissue specific genes and did not contain a CpG island. (Heyn *et al.*, 2012).

In order to examine the frequency of age-related methylation changes in CpG islands, Tra et al. (2002) utilised a restriction landmark genome scanning (RLGS) approach to inspect around 2000 loci in T lymphocytes from newborns, adult (average 27 years of age) and elderly donors (65+ years of age). Their analysis suggested only 1% of loci were differentially methylated with age, and proposed that methylation is not significant affected by ageing, at least in T lymphocytes (Tra *et al.*, 2002). However, other analyses have disagreed with these findings. A methylated CpG island amplification and microarray (MCAM) analysis representing almost 9,000 genes in normal human colon samples from young (35 years) and older (68 years) individuals found increased methylation at 10% of loci, whereas 1% of loci has lower methylation in the older individuals. A similar analysis in young (3 month) and old (35 month) mice revealed and even higher frequency of changes, 21% of genes has higher methylation and 13% showed lower methylation in old compared to young mice (Maegawa *et al.*, 2010). This suggests that age-related methylation changes may affect a significant proportion of the genome in mammals. Such discrepancies indicate that the analysis of specific tissues

and the use of different techniques for detecting DNA methylation may impact upon the reported findings.

1.5.2.4 Tissue specific DNA methylation and ageing

In order to determine the tissue specificity of age-related methylation changes, Christensen et al performed an analysis of CpG sites in almost 800 genes, using normal samples from ten anatomical sites including bladder, blood, brain, cervix, head & neck, kidney, lung, pleura and small intestine. A comparison of methylation differences both between individuals, and across tissues within the same individual found that CpG sites within islands had methylation increases, whereas the trend was for decreased DNA methylation in non CpG island sites in all tissues; confirming the genome wide hypomethylation and gene specific hypermethylation associated with the ageing epigenome. Interestingly, this study observed that methylation levels of *DNMT3B* decreased as a function of age, proposing a mechanism by which the deregulation of proteins involved in the DNA methylation reaction could impact upon DNA methylation levels (Christensen *et al.*, 2009).

1.5.2.5 Gene specific DNA methylation and ageing

On the basis of the observed relationship between DNA methylation and age, a recent study published that the methylation levels of three CpG sites located within three gene promoters, *EDERADD*, *NPTX2* and *TOM1L1* in saliva samples, was sufficient to accurately determine a person age to within approximately 5 years. Suggesting that, at least at these loci, DNA methylation changes are occurring at a highly similar linear rate with ageing at a population level (Bocklandt *et al.*, 2011).

The consequences of age-associated methylation changes which are not cancer associated have been investigated by a recent epigenome wide association study (EWAS) of monozygotic twins (MZ) which identified a number of DMRs displaying hypermethylation in association with chronological age, and also DMRs which were linked to several age-related phenotypes such as cholesterol levels, lung function, maternal longevity and blood pressure. However, only one DMR was overlapping between these groups, suggesting that methylation changes accumulated during ageing are not contributing to age associated phenotypes (Bell *et al.*, 2012). Further

understanding of DNA methylation changes with age are clearly required to elucidate the functional consequences.

1.6 Environmental mediators of age-associated DNA methylation changes

Although 5'methylcytosine is a stable modification, it is not fixed and several factors are known to be significant for determining or altering DNA methylation patterns. Therefore, distinct exposures may be important regulators of age associated methylation changes.

1.6.1 Nutrition and DNA methylation

The critical methyl donor for the DNA methylation reaction is S-adenosyl methionine (SAM), which is derived from methionine and ATP. After the methyl (CH_3) group from SAM has been transferred to DNA, its unmethylated form, S-adenosyl homocysteine (SAH), is converted to homocysteine. Homocysteine can be converted back to methionine in a reaction which is dependent on 5'methyltetrahydrofolate as a cofactor alongside the methionine synthase (MS) enzyme and vitamin B12. The formation of active 5'methyltetrahydrofolate is dependent on a series of folate reduction reactions, also involving 5,10-methylene tetrahydrofolate reductase (MTHFR) (Scott, 1999). The efficiency of this pathway determines how effectively SAM can be produced. This means that levels of components of this cycle including folate and vitamin B12, as well as single nucleotide polymorphisms (SNPs) affecting the function of enzymes required for these reactions, can directly impact upon SAM formation and consequently, DNA methylation levels (Feil, 2006; Hazra *et al.*, 2010).

1.6.2 *In utero* exposures

It is likely that the genome is more vulnerable to altered methylation in utero. In order to establish cell lineages in the developing embryo, there is radical reprogramming of DNA methylation, facilitated by highly active DNMTs. Anything which influences the regulation of this dynamic process could result in DNA methylation arising at spurious locations, or else accumulating to abnormally high levels during the prenatal period.

1.6.2.1 Altered maternal nutrition during pregnancy

There are several examples of altered nutrition during pregnancy influencing DNA methylation levels, gene expression and phenotype of the offspring, particularly in animal models where diet can be easily controlled and manipulated.

1.6.2.1.1 The agouti mouse model

The effect of altering intake of methyl donors within the diet has been elegantly demonstrated by mouse models featuring the viable yellow agouti locus (A^{vy}). These mice are inbred; however coat colour is determined by DNA methylation of an intra-cisternal A particle (IAP) transposon. Agouti (brown) coloured mice have complete methylation of IAP and are healthy and lean, whereas viable yellow mice have completely unmethylated IAP transposons, aberrant agouti expression driven by a long terminal repeat region within the IAP, and are prone to obesity and hyperinsulinemia. Epigenetic mosaicism leads to a range of variable coat colours between these two extremes. Increasing methyl donor supply in the diet of pregnant dams by supplementation with folate, vitamin B12, choline and betaine increased DNA methylation at the agouti locus of their offspring leading to expression changes and altered coat colour towards agouti (Wolff *et al.*, 1998).

1.6.2.1.2 Methyl donors

Restricting folate, vitamin B12 and methionine supply in sheep peri-conceptually led to offspring which were fatter and heavier, had higher blood pressure and showed evidence of insulin resistance compared with offspring from mothers fed a control diet with adequate methyl donor supply and resulted in hypomethylation at several CpG sites across the genome (Sinclair *et al.*, 2007).

1.6.2.1.3 Under nutrition

In humans, maternal under nutrition can be experienced during times of famine, such as the Dutch hunger winter in 1944. At its worst, the average calorie intake of pregnant women fell to around 400-800 calories per day. Children *in utero* during this time were born smaller, with altered methylation patterns which persisted into adulthood

(Heijmans *et al.*, 2008). Additionally, they appear to have an increased incidence of symptoms of poor health in later life such as impaired glucose tolerance, coronary heart disease and obesity (Roseboom *et al.*, 2006). It has been hypothesised that these modifications may enhance survival of the foetus but could cause undesirable gene expression in the adult leading to disease, so called the "developmental origin of adult disease" hypothesis (Barker, 2004), but the mechanisms underlying these observation are unclear.

1.6.2.2 Environmental exposures (cigarette smoking)

Maternal smoking during pregnancy has been linked to adverse health effects including low birth weight and impaired lung function (Hofhuis *et al.*, 2003), and interestingly children prenatally exposed to smoking display global hypomethylation and increased methylation of specific gene promoters, similar to the aberrant methylation pattern of many diseases (Breton *et al.*, 2009). Again however, whether altered DNA methylation contributes towards these adverse health markers, or ensues as a consequence, is undetermined.

1.6.3 Postnatal Exposures

1.6.3.1 Postnatal nutrition and diet

The most widely studied postnatal dietary components known to influence methylation are those which modify the availability of methyl donors including folate, vitamin B12, betaine and choline. Food sources of these compounds include green vegetables, eggs and meat. Patients with hyperhomocysteinaemia demonstrate hypomethylation of DNA, due to the accumulation of high levels of homosysteine which inhibits the action of DNMTs. DNA methylation levels in these individuals could be corrected by supplementation with folate, presumably by increasing levels of the reduced folate derivatives which are essential for SAM generation and the subsequent methylation reaction (Ingrosso *et al.*, 2003). A study in rats demonstrated that feeding a high fat diet was associated with increased DNA methylation of the leptin promoter and reduced circulating leptin levels. Leptin is vital for appetite regulation (mice with loss of functional leptin are obese), suggesting that diet can lead to altered DNA methylation and regulation of functionally relevant genes (Milagro *et al.*, 2009).

1.6.3.2 Postnatal environmental exposures

1.6.3.2.1 Exposure to pollutants/alcohol

Recent exposure to environmental pollutants, particularly those particulates caused by traffic, has been linked to reduced global methylation measurements (Baccarelli *et al.*, 2009). The effect of numerous environmental exposures in a genome wide analysis of DNA methylation found several associations of interest. Higher methylation levels of CpG sites in the lung were observed in individuals in contact with asbestos compared to unexposed individuals. Additionally, differential methylation was observed in the peripheral blood of people who drank alcohol as opposed to those who abstain (Christensen *et al.*, 2009).

1.6.3.2.2 Exposure to cigarette smoke

Smoking has been frequently linked to aberrant methylation, including over 100 CpG sites (almost 10% of those analysed) in the lung. Also, in peripheral blood, increasing packyears was significantly related to methylation levels of two genes including *MLH1* which is known to be frequently hypermethylated in cancer cells (Christensen *et al.*, 2009) . The functional significance of these changes is yet to be established, but an additional study has shown that a number of tumour suppressor genes are targets for aberrant methylation in the lungs of current and former smokers. These same targets are similarly hypermethylated in lung tumour cells (Belinsky *et al.*, 2002). However, it is possible that smoking leads to early pre-neoplastic transformation of cells, inducing many molecular changes including altered DNA methylation as a consequence, even in cells with apparently normal morphology.

1.7 Parallels between methylation in ageing and cancer

It is now becoming apparent that similar epigenetic changes can be observed with advancing age to those seen in cancer in many tissues; however the consequences of this are not fully understood (Richardson, 2003).

1.7.1 Hypomethylation

Similar to the observation in cancer cells, methylation is characteristically lost at repetitive regions of the genome, such as Alu elements, which exhibit a linear decline in methylation levels with advancing age (Bollati *et al.*, 2009).

1.7.2 Hypermethylation

The presence of age related gene specific hypermethylation was first appreciated by researchers focusing on DNA methylation of the *ESR1* gene and its subsequent loss of expression in breast cancer cells and other types of cancer. They discovered a very high level of methylation in colorectal cancer samples, and surprisingly also in normal colorectal tissue samples with no associated tumours (Ottaviano *et al.*, 1994). Further analysis of *ESR1* showed that although normal colorectal mucosa from young individuals was almost completely unmethylated, the amount of methylation in a subset of cells increased progressively with advancing age. This could be correlated with a decrease of *ESR1* expression, implying that this age associated DNA methylation had a functional impact (Issa *et al.*, 1994).

Examination of colorectal tissue has since identified several genes which follow the same pattern as *ESR1*, being hypermethylated in cancer and progressively methylated with age including *TUSC3* and *MYOD*; In contrast, no such effect was seen in other genes, implying that the acquisition of methylation is partially loci dependent (Ahuja *et al.*, 1998). Interestingly, this study also found that some individuals may be more susceptible to acquiring methylation at multiple loci than others; a young individual with very high methylation and several older individuals with very low methylation showed this consistent pattern across several genes tested (Ahuja *et al.*, 1998). It is possible that these individuals' differences may be phenotypically relevant.

Genome wide DNA methylation studies have been utilised to replicate the initial findings, and previously reported genes, including *ESR1*, *MGMT*, and *RARB2*, also show age related methylation increases and hypermethylation in cancer and can be identified by these types of analysis, validating the observations of single gene studies (Christensen *et al.*, 2009).

The overlap between the types of genes susceptible to increased DNA methylation during ageing and the loci which are hypermethylated in cancer has been a key outcome from many genome wide analyses. Developmental genes which are silent in stem cells, characterised by the presence of bivalent chromatin domains in their promoter region and targeted for PcG protein binding, are highly enriched targets for preferential methylation in both ageing and cancer, proposing that the mechanisms by which methylation is occurring in ageing and cancer are similar, if not the same (Rakyan *et al.*, 2010; Teschendorff *et al.*, 2010).

The premise that DNA methylation changes during ageing may precede those observed in neoplastic cells has been explored. Analysis of specific genes such as *ESR1*, found that the same promoter associated CpG sites which were hypermethylated in colorectal tumours and polyps were also hypermethylated in a subset of cells in the normal colon and that this became more marked with increasing age, suggesting that ageing associated methylation changes may precede cancer onset (Issa *et al.*, 1994). Recent genome wide investigation of 27,000 CpG sites in normal cervical cells found substantial DNA methylation variability in women who developed cervical cancer after three years compared to those remaining disease free. Furthermore, these highly variable sites were overlapping with those displaying aberrant methylation during ageing, including Polycomb group targets. A total of 140 'risk' CpG sites were used to accurately predict neoplastic transformation in an independent sample set (Teschendorff *et al.*, 2012). These findings propose that DNA methylation changes are occurring prior to the appearance clinically detectable disease and could be useful for identifying at risk individuals.

Whilst the age related increased DNA methylation of specific genes in normal samples and hypermethylation of tumour suppressors in cancer tissues have frequently been observed in human samples, it remained unclear whether this is specific to long lived humans or is evident in other species. Methylated CpG island amplification and microarray analysis of old (35 month) compared to young (3 month) mouse intestine revealed numerous methylation increases (774 genes) and decreases (466 genes). There was partial conservation in age related methylation patterns in human and mouse microarrays. Several candidate genes which were reported to be hypermethylated in normal ageing colon or in colon cancer were also hypermethylated in the older mouse

samples, suggesting that age related methylation is not restricted to humans, but also affects mice despite their relatively short lifespan (Maegawa *et al.*, 2010).

1.8 Inter-individual variation in DNA methylation

Currently, little is known about how the accumulation, or progressive loss, of DNA methylation may vary on an individual level during ageing and whether this may contribute to susceptibility to diseases such as cancer, for which advancing age is a significant risk factor.

1.8.1 DNA methylation in twins

Comparisons of monozygotic (MZ) twins, who have the same DNA nucleotide sequence, as opposed to dizygotic (DZ) twins sharing approximately half the same genetic sequence, are useful for studying the effects of heritability and environment. Phenotypic or methylation differences in an MZ twin pair are likely to be due to the effects of the environment or a random process, rather than have a genetic basis. Microarray data suggests that DZ twins have significantly higher methylation differences than MZ twins, providing some evidence for a type of genetic control of DNA methylation at specific regions of the genome (Kaminsky *et al.*, 2009). Others have also confirmed this tendency for hereditability, observing a trend of more similar DNA methylation in MZ compared to DZ twin pairs (Schneider *et al.*, 2010).

The theory that aberrant DNA methylation is accumulated during a lifetime is illustrated by MZ twins pairs who display either no or exceedingly little differences at very early ages, whereas older twins often have much more divergent and variable DNA methylation profiles. This disparity presents as both global 5'methylcytosine content, and methylation of specific genes which could be linked with more gene expression differences in older compared to younger twin pairs. Twin pairs with the most divergent methylation profiles were often older, had spent less time in a shared environment and had more dissimilar health status (Fraga *et al.*, 2005a). This implies that DNA methylation mechanisms may become less controlled over time, or that differential exposures may influence the accumulation of DNA methylation over time. Identification of several loci differentially methylated in an age related manner has since been confirmed in both twins and singletons (Boks *et al.*, 2009). Combined longitudinal

(10 year follow up) and cross-sectional (age range 18-89 years) analysis was used to examine age-related methylation alterations at several disease associated promoters including *GNASAS, IGF2, INS* and *LEP* for MZ twin pairs. Similarly to findings described above, it was also found that variable DNA methylation patterns were more pronounced in older twins and that twin pairs become more divergent with advancing age. During the longitudinal study of older twins, a similar pattern of methylation changes were observed after 10 years to those seen when comparing young to elderly twins (Talens *et al.*, 2012). These studies suggest that the degree of DNA methylation variability at an individual level progressively increases throughout life in twins.

1.8.2 DNA methylation changes over time

A further investigation comparing global levels of DNA methylation in PBL DNA of healthy volunteers taken approximately 11 years apart found almost 30% of individuals had a greater than 10% loss or gain of methylation (Bjornsson *et al.*, 2008). Additionally, analysis of up to three generations of the same family revealed not only intra-individual variation after 16 years between sample collections, but also that those samples from within families were clustered together. This again implies that there may be a genetic component to maintaining methylation states, especially since the majority of family members tested were not living within a shared environment for the years between sampling (Bjornsson *et al.*, 2008).

However, analysis of *COMT*, a gene implicated in psychiatric disorders, found differential methylation in MZ twin pairs at age 5. As the twins were likely to have had largely shared environmental exposures at this young age, this suggests that environment, shared genetics or the passage of time was not a factor in producing the variable DNA methylation of these genes (Mill *et al.*, 2006). Bisulfite sequencing of 1,760 CpG sites in the human major histocompatibility complex uncovered highly variable DNA methylation in both MZ and DZ twins, implying little heritability in this region (Gervin *et al.*, 2011). Methylation of an imprinted locus, *IGF2/H19*, measured in blood samples from adolescents compared to middle aged MZ and DZ twins found DNA methylation of *IGF2* ranged from 43-62%, with much variation across the individual that did not appear to be related to age (Heijmans *et al.*, 2007). A longitudinal analysis of genes implicated in cardiovascular and metabolic disease found that although methylation levels were highly variable between individuals, they were

relatively stable after 11-20 years of follow up in blood samples and after 2-8 years in buccal swabs from individuals ranging from 14-62 years (Talens *et al.*, 2010). These examples highlight that individual variation is not only a feature advancing age, but can be present throughout life and the role of genetics in determining methylation status is unresolved.

1.9 Mechanisms for individual variation of DNA methylation

Whitelaw and Whitelaw (2006) outlined several hypotheses proposing why individual variation of DNA methylation may be observed, particularly at advanced ages (Whitelaw and Whitelaw, 2006).

1.9.1 Altered activity of DNA methyltransferases

Analysis of human fetal lung fibroblast cells found *DNMT1* gene expression, protein levels and enzymatic activity was reduced in aged cells, whereas *DNMT3A* and *DNMT3B* levels and activity was increased. This implies a direct role for DNMT's causing the aberrant methylation which is frequently observed in many tissues and species at increasing ages. Global DNA hypomethylation may be due to reduced maintenance of methylation, and gene specific hypermethylation possibly occurs due to increased activity of the de novo methylating enzymes. Activities of all DNMT's are also increased after neoplastic transformation which demonstrates a very similar pattern of aberrant methylation to that of age related DNA methylation patterns (Casillas *et al.*, 2003). In human liver samples, the expression levels of *DNMT3A* and *DNMT3B* were significantly higher in older compared to younger individuals and levels varied among individuals, implying that such differences could underlie inter-individual differences in DNA methylation (Xiao *et al.*, 2008). However, others have failed to find any association between *DNMT1* levels and age (Hammons *et al.*, 1999).

The biological effect of increased expression of the de novo methylating enzymes is illustrated by the overexpression of *Dnmt3a* and *Dnmt3b* in $Apc^{min/+}$ mice which are susceptible to colon tumours. Increased *Dnmt3b* was associated with an increased number of tumours, as well as higher levels of DNA methylation and silencing of several tumour suppressor genes. The gene targets displaying increased methylation induced by *Dnmt3b* overexpression were the same in normal, as well as the cancerous

tissue (Linhart *et al.*, 2007). This suggests that certain loci may be more susceptible to acquiring DNA methylation than others when DNMTs are abnormally expressed.

1.9.2 Genetic effects (SNPs)

Underlying genetic sequence has been proposed to affect methylation levels, demonstrated by MZ twin pairs demonstrating more epigenetic similarity than DZ twins (Kaminsky *et al.*, 2009; Schneider *et al.*, 2010). To examine this link further, genome wide DNA methylation analysis of over 22,000 CpG sites in lymphoblastoid cell lines from 77 individuals was integrated with data from over 3 million SNPs. Methylation levels at 180 CpG sites were associated with a cis-acting SNP within 5kb and at a significant proportion of these genes; methylation was negatively associated with expression (Bell *et al.*, 2011). A recent EWAS of twins calculated the average genome wide heritability of DNA methylation to be around 18%, still leaving a significant proportion of methylation changes which cannot be explained by underlying genetic sequence (Bell *et al.*, 2012).

1.9.3 Stochastic events linked to replication

Variation could also occur due to a lack of fidelity of the enzyme responsible for maintaining methylation patterns. An *in vitro* study in mice suggests that although *Dnmt1* has an initial 30-40 fold preference for hemimethylated DNA; it is diminished by the addition of fully methylated sequences. This suggests that as newly synthesised DNA becomes methylated in vivo, the function of *Dnmt1* could be detrimentally affected (Goyal *et al.*, 2006). Additionally, the de novo methylation rate of this enzyme could be as high as 5% per replication (Vilkaitis *et al.*, 2005). Examining the fidelity of methylation patterns at unmethylated CpG islands suggest an error rate of approximately 0.1% per cell replication at any CpG site. This rate is doubled at sites outside of the promoter region. Methylated CpG sites have a much lower error rate (Ushijima *et al.*, 2003).

1.9.4 Differential environmental exposures

Nutritional factors such as folate, vitamin B12, choline and betaine modify the availability of methyl donors so variable intakes of these nutrients can affect DNA

methylation levels during a person's lifetime (Feil, 2006). The phenotypic effect of this is can be observed by increasing methyl donor supply in the diet of pregnant dams by supplementation, which caused increased DNA methylation at the agouti locus of their offspring and altered coat colour (Wolff *et al.*, 1998). Further discussion of the effect of differential environmental exposures, such as pollutants and cigarette smoking, on DNA methylation is detailed in previous sections (see section 1.6.3.2).

1.9.5 Replication independent events

DNA methylation can also be altered in non-replicative tissues such as the brain. The licking and grooming behaviour of parent rats towards their offspring induced differences in DNA methylation and expression of the glucocorticoid receptor (GR) in their offspring. Such changes were correlated with deregulation of the hypothalamic-pituitary-adrenal stress response which persisted to adulthood (Weaver *et al.*, 2004). Human studies of suicides have linked increased DNA methylation, decreased expression of neuron specific *GR* and altered stress response with childhood abuse, compared to suicides with no history of abuse or non-suicidal death (McGowan *et al.*, 2009).

1.10 Summary

In summary, the role of DNA methylation is well characterised in certain circumstances, particularly in cancer cells where hypermethylation of gene promoter regions typically results in transcriptional silencing. It is apparent that similar methylation changes accumulate with ageing, although usually not at such dramatic levels, therefore the consequences of age-related methylation, particularly methylation increases at the promoter region are unknown. This accumulation of DNA methylation with advancing age can vary between individuals, and together with the fact that methylation marks are modifiable by diet and environmental exposures, make it an attractive target worthy of investigation for a possible role in determining individual differences age-related health or disease status.

1.11 Hypotheses

In addition to the highly abnormal DNA methylation patterns evident in malignant cells, it is becoming apparent that supposedly unmethylated gene promoters may also acquire methylation at CpG sites with progressing time. However, the absolute amount of methylation present in an elderly population at specific genes is not well understood. Secondly, it is unclear whether all loci or just a subset are susceptible to increasing levels of DNA methylation. Lastly, although altered methylation is attracting study for its potential role in disease predisposition, it needs to be established that levels of DNA methylation are variable in the population, as it is the presence of such variability which may underlie differences in disease susceptibility. If changes in DNA methylation levels are uniform between individuals during ageing, then it is unlikely to play a role in differential disease susceptibility in individuals of the same age.

Our hypothesis for this study was that individuals at age 85 would demonstrate highly variable levels of DNA methylation at promoter associated CpG islands and that certain individuals may be at more susceptible to acquiring gene specific DNA methylation across the genome (possibly due to genetic background or differential environmental exposures). We hypothesised that an individual predisposed to increased levels of DNA methylation during ageing would be at an increased risk of cancer, since key genes (including tumour suppressor genes) may be methylated and transcriptionally silenced. This theory could partly explain the strong link between increasing age and increasing cancer incidence.

1.12 Aims

In order to more clearly investigate the age related changes in promoter associated CpG island methylation, we had four specific aims which are detailed below. Firstly,

1. To measure the extent and inter-individual variability of gene specific promoter associated CpG island methylation in a normal elderly population.

The presence of aberrant methylation in cancer has been frequently described; however whether these changes occur as a consequence of neoplastic transformation, or are evident before cancer is initiated remains to be demonstrated. Since advancing age is a

significant risk factor for the majority of cancer, it is possible that age-related DNA methylation increases may be responsible for this increased risk. In order to investigate this further, our second aim is,

2. To determine if individuals with either a previous history or future diagnosis of cancer, where hypermethylation of numerous genes is a frequent observation; have increased levels of DNA methylation suggesting they may be more susceptible to cancer?

The presence, and potential role, of altered DNA methylation in other age-related diseases is not as well defined as in cancer, but since there is a wealth of clinical information accompanying participants in the Newcastle 85+ study, our third aim asks,

3. Are differences in DNA methylation levels associated with the presence of other clinical features in the Newcastle 85+ study participants?

Finally, others have recently reported a significant overlap between sets of genes which become hypermethylated in cancer and sets of genes that become methylated during ageing. Therefore our final aim will be,

4. To begin to assess whether identification of methylation of a gene in cancer implies susceptibility to age related methylation in the corresponding normal tissue.

1.13 Objectives

In order to demonstrate that levels of DNA are highly variable in the very elderly population, a panel of genes will be chosen to represent gene specific loci throughout the genome. Methylation levels at the promoter region of these genes will be measured by Pyrosequencing, a highly quantitative technique for DNA methylation measurements, in DNA isolated from peripheral blood leukocytes (PBL) of participants in the Newcastle 85+ study. All participants were aged 85 years at the time of sample collection.

Next, to establish whether elderly individuals with higher methylation levels in PBL DNA are at increased risk of cancer in the Newcastle 85+ study, the DNA methylation measurements of participants both with and without cancer will be compared. Both retrospective comparison of participants with a previous history of any cancer (excluding non melanoma skin cancer) prior to their recruitment on the study, and prospective analysis of participants diagnosed with cancer during the study (within three years of DNA sample donation) will be performed.

Subsequently, to identify a possible role for altered DNA methylation in other age-related diseases, DNA methylation measurements of participants in Newcastle 85+ study with cardiovascular disease, dementia, type II diabetes and stroke will be compared to individuals without the disease. This has the potential to identify whether differential DNA methylation may play a role in other age-related diseases.

Lastly, to determine whether methylation of a gene in cancer implies susceptibility to age related methylation in the corresponding normal tissue, novel cancer associated hypermethylated loci will be identified by genome wide methylated DNA enrichment of lymphoma cells lines combined with high throughput sequencing. Targets will be chosen at random for further confirmation of their methylated status in primary cancer samples (ALL). Finally, DNA methylation levels of these same genes will be measured in normal PBL DNA samples from the 85 year old population, to conclude whether increased methylation is evident.

Chapter 2 – Methodology

2.1 Newcastle 85+ Study samples for gene specific DNA methylation analysis

The Newcastle 85+ Study invited all eligible participants who were born in 1921, therefore aged 85 upon commencement of the study in 2006, and registered with a GP surgery based in Newcastle upon Tyne or Tyneside to take part in the study (Collerton *et al.*, 2007). In total, 1042 individuals were recruited including participants who were living in care homes, those with disabilities or cognitive impairments; representative of the characteristics of the population at that age (Martin-Ruiz *et al.*, 2011). Informed consent was provided by all participants or their carers if individuals lacked capacity to consent (Collerton *et al.*, 2009). Comprehensive GP record reviews were performed detailing disease and medication use. In addition, a more detailed assessment of health status was completed by a nurse visiting the home residence of each participant; including blood sampling, questionnaires, measurements and functional tests (Collerton *et al.*, 2009). These assessments were carried out at baseline (Phase I) and then again at 18 (Phase II) and 36 months (Phase III) after initial recruitment.

Blood samples were taken after an overnight fast where possible and were subjected to routine hematological tests including a full blood count using standard automated analysis systems, and measurement of nutritional markers including red cell folate, plasma vitamin B12 and total plasma homocysteine. More extensive details regarding these methods can be found elsewhere (Martin-Ruiz *et al.*, 2011). In addition, DNA was extracted from peripheral blood leukocytes (PBL) and peripheral blood mononuclear cells (PBMC) with the QiaAmp DNA Maxi kit (Qiagen Ltd, Crawley, UK) for future analyses (including this study of DNA methylation) (Collerton *et al.*, 2007; Den Elzen *et al.*, 2011). All measures were carried out by the Newcastle 85+ study team.

2.1.1 Chapter 3 Newcastle 85+ study participant samples (Pilot study)

DNA from peripheral blood leukocytes (PBL) samples were selected from a random sample of 50 participants at Phase I of the Newcastle 85+ Study and the corresponding peripheral blood mononuclear cell (PBMC) DNA samples for 37 of these individuals which were available for analysis at Phase II of the study.

2.1.2 Chapter 4 Newcastle 85+ study samples (Association with clinical outcomes)

2.1.2.1 Initial analysis

DNA from PBL was obtained at Phase I of the Newcastle 85+ study. For the initial analysis, DNA samples were obtained from 113 participants with a diagnosis of any cancer (excluding non-melanoma skin cancer) prior to recruitment onto the study (previous cancer). The control group for this analysis comprised of 113 participants randomly selected from all remaining participants. The total number of DNA samples analysed in the first analysis was 226, however only 220 of these were unique as six participants with a previous diagnosis of cancer had been included in the pilot study.

2.1.2.2 Validation analysis

For the validation analysis, the 72 individuals diagnosed with any cancer (excluding non-melanoma skin cancer) during the 3 year duration of the study (future cancer) were selected. However, 36 of these participants had been unknowingly included in the first analysis (as data on cancer development during the study had not been available at the time), meaning 36 individuals remained for analysis. In addition, all 24 current smokers, 36 dementia patients, and 57 participants diagnosed with Type II diabetes were selected to validate associations observed in the first analysis, however due to an overlap of participants in more than one group, for example individuals with both Type II diabetes and dementia, this represented 103 unique DNA samples. For this analysis, the control group was comprised of 71 participants randomly selected from all remaining individuals. In total, the number of DNA samples analysed in the validation analysis was therefore 210 and thus 436 individuals were analysed in Chapter 2.

In total, 480 Newcastle 85+ study DNA samples were tested for the complete analysis including the samples from the pilot study, and associations with cancer and other clinical outcomes (Table 2.1).

2.1.3 Chapter 5 Newcastle 85+ study participant samples (Examination of targets identified via genome wide methylation analysis)

In order to measure the DNA methylation levels of genes identified from a genome wide DNA methylation analysis of lymphoma cell lines in PBL DNA from the Newcastle 85+ study, DNA samples were largely selected from individuals in the pilot study, with the exception of 3 samples which removed and replaced with 3 samples from the larger 481 sample set, to ensure an equal proportion of these 50 samples fell within each quartile of the larger sample set, based on average methylation levels of the five variably methylated genes.

		Selection	Number of Unique Samples
Chapter 1	Pilot Study	Random	50
Chapter 2	First Analysis	Previous Cancer	107
		No Previous Cancer (Random)	113
Chapter 2	Validation Analysis	Control Group (Random)	71
		Future Cancer	31
		Type II Diabetes	50
		Dementia	27
		Current Smoker	18
		Dementia + Diabetes	4
		Dementia + Current Smoking	4
		Future Cancer + Diabetes	2
		Future Cancer + Current Smoking	1
		Future Cancer, Current Smoking + Diabetes	1
		Future Cancer, Diabetes + Dementia	1
		Total	**480**

Table 2.1: Selection of PBL DNA samples from Newcastle 85+ Study participants for gene specific DNA methylation analysis

Table shows the selection criteria used and the number of participants analysed for different stages of the study.

2.2 Pyrosequencing assay design

Gene sequences, including predicted transcriptional start sites, for primer design were obtained from the GenBank entry on NCBI and the region surrounding transcription start site (promoter and first exon) was examined for the presence of a CpG island. Gene sequences were altered to reflect their expected sequence pattern after bisulfite modification and input into PSQ assay design software (Qiagen, Crawley, UK). Forward, reverse and sequencing primers were designed according to the software criteria to amplify a region within the CpG island. Sequencing primers were designed to be positioned close to a minimum of three consecutive CpG sites, with the first site being within 10 base pairs of the primer sequence. The exception was the imprinted gene *IGF2*, for which the CpG sites examined were located in the differentially methylated region 0 (DMR0) upstream of exon 2, which is known to be important in control of imprinting (Murrell *et al.*, 2008). Either the forward or reverse primer was labelled with biotin, dependent on which strand the sequencing primer is complementary to, to allow specific separation of the strand which was to be sequenced during Pyrosequencing. Primers containing a CpG site were designed with a degenerate base at the position of the cytosine to allow amplification of both methylated and unmethylated sequences (Y = C/T nucleotide, R = A/G nucleotide on complementary strand). Gene loci primer sequences are listed, along with the number of CpG sites analysed and their location relative to the transcriptional start site (TSS) (Table 2.2).

Gene	Forward Primer	Reverse Primer	Sequencing Primer	CpG Sites	Distance from TSS
APOE	GGAGTTTTATAATTGGATAAGT	AAAATCCCAACTCTTTCT (Bio)	GGGATTTTTGAGTTTTATT	3	-24 → +123
EPHA10	ATTATTTATGGGAATTAATTATTAG	CCCACCAATTAAATATTC (Bio)	TATTTATGGGAATTAATTATTAGGTA	3	+28 → +143
ESR1	GGAGATTAGTATTTAAAGTTGGAGGT	CCTAAAAAAAAAACACAACCC (Bio)	TGGGATTGTATTTGTTTT	5	+92 → +294
HAND2	CCAAATTTTAATTATCTTATA (Bio)	GGATTTTTTAGTAAGATTT	ATTTTATTACCCAAAACTC	4	-29 → +185
HLXB9	TAGTTYGGGTTGTTTAGGAT	TACRCTACAACCCTCAAA (Bio)	GAAGTYGYGTTGAGTTTTAAG	4	-161 → +61
HOXA4	TACACTTCACAAATTAATAACCATAAACTC (Bio)	GTTGTTGTAGYGGTAGGTGTTG	AACCCAAATTCCCTCCCTT	4	+31 → +219
HOXA5	CCACAAATCAAACACACATATC (Bio)	GGTTGGTTGTATTTGGGT	CTACRACTACAATAACATAAATCT	5	+30 → +327
HOXD4	GAAATTAATGGTTATGAGT	CCCRAAACTAAAAATCTA (Bio)	ATTTGTAGGGYGGTTATTTAGGYGAG	5	+242 → +398
IGF2	GGGTTAAGGTAGTTTTTTTGGGAATG	AAATATAAAAACCTCCTCCACCTCC (Bio)	AGGGGGTTTATTTTTTTAGGAA	3	DMR0
MLH1	AGTTTTTTTTTTAGGAGTGAAGG	ATAAAACCCTATACCTAATCTATC (Bio)	GTAGTATTYGTGTTTAGTTT	6	-569 → -375
TUSC3	GAATAGGATGTTTTGTTAGTT (Bio)	TAACTAAACCCTCCCAAATA	CCAAATAACAACCACTTTAC	6	+25 → +116
p15	GGTTTGGGGTTTYTGTAGTGG	CRACCCCTTAACCCAACTAAAAAC (Bio)	GAGGAAGTAGTTTAGTTTAAAG	5	-207 → +41
RARB2	TATAGAGGAATTTAAAGTGTGGGTTG	AAAACTTAAAAACTCCCAACAACCCT (Bio)	GTTTTTAAGGTGAGAAATAG	4	-416 → -118
RASSF1A	GGGGGAGTTTGAGTTTATTGA	CTACCCCTTAACTACCCCTTCC (Bio)	GGGTYGTATTYGGTTGGAG	5	+136 → +434
TWIST2	AACAACTATTTAACAACCCAACCCCAAC	GGGYGAGTTGGAGTTTTTTTTATGG (Bio)	CAAAACCTTTCCAACAAC	4	-26 → +208

Table 2.2: The fifteen gene loci selected for DNA methylation analysis by Pyrosequencing in the Newcastle 85+ study participants

For each gene the forward, reverse and sequencing primer sequences are shown and the biotin labelled primer is indicated (Bio). Degenerate bases are used at the cytosine position of CpG sites (Y = C/T nucleotide, R = A/G nucleotide on reverse strand). The number of CpG sites analysed in the Pyrosequencing assay and the region amplified by forward and reverse primers relative to the transcriptional start site are also listed.

2.3 Sodium bisulfite modification of DNA

Sodium bisulfite modification of DNA is widely used as the first step in many procedures for DNA methylation analysis. It is a chemical based treatment which converts unmethylated cytosine to uracil (replaced with thymine after PCR amplification) by a deamination reaction, whereas methylated cytosine is protected from this conversion (Clark *et al.*, 2006).

A detailed diagram of the steps involved in sodium bisulfite conversion of cytosine to uracil is shown in Figure 2.1.

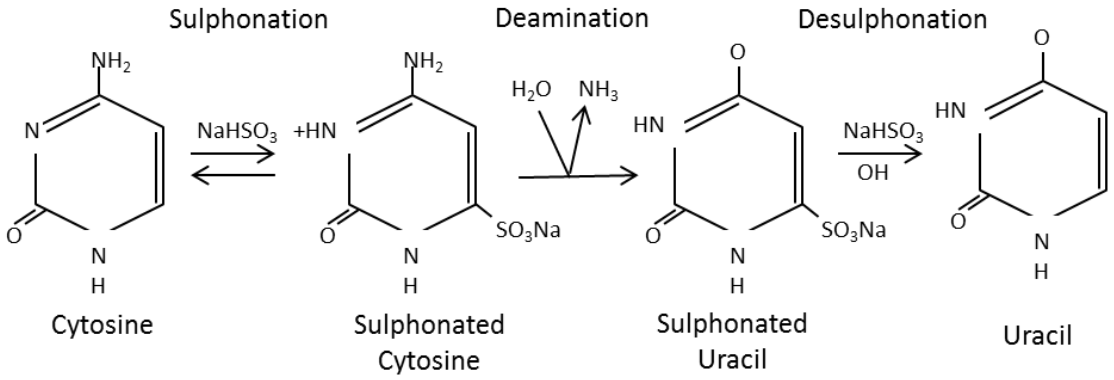


Figure 2.1: Diagram of sodium bisulfite conversion of cytosine to uracil

The degree of methylation is subsequently determined via the ratio of C:T nucleotides at a particular CpG site, the converted sequence of an unmethylated site being TG and a methylated site represented by CG (illustrated in Figure 2.2). The main purpose of this is to fix the DNA methylation pattern as a sequence change, since methyl groups are lost from DNA during PCR amplification. For all samples, 200ng of DNA was bisulfite modified using the Epigentek Methylamp™ One-Step DNA Modification Kit (Cambridge BioScience, Cambridge, UK) according to the manufacturer's protocol.

Briefly, DNA was mixed with sodium bisulfite (DNA modification reagent) and denatured by incubation of DNA at 99°C for 6 minutes, followed by incubation at 65°C for 90 minutes, for optimal deamination of cytosines on single stranded DNA with minimal degradation. DNA samples were then bound to a column with DNA capture reagent and washed with cleaning reagent and 90% ethanol to allow efficient removal of sodium bisulfite salts, followed by elution in 15µl of DNA elution solution. Modified DNA was stored at -20°C. A full protocol is detailed in the appendix (A1.1).

Figure 2.2: Nucleotide sequence change induced by treatment of DNA with sodium bisulfite and subsequent PCR amplification.
Methylated cytosine (green C) is protected from sequence change whereas unmethylated cytosine (red C) is susceptible to uracil conversion after sodium bisulfite treatment, which is replaced with thymine during PCR amplification

2.4 PCR amplification

All PCR reactions were carried out in a final volume of 25µl, using 1µl (14ng) of sodium bisulfite modified DNA for each PCR amplification with 0.2µl FastStart taq polymerase (Roche, Welwyn Garden City, UK), 2.5µl of manufacturers' buffer, 0.5µl of dNTPs, 1-4mM of $MgCl_2$ and 150ng of forward and reverse primer (Eurofins MWG Biotech, London, UK), with the exception of *ESR1* and *HOXA5* for which 75ng of each primer was used. PCR amplification was performed using either the G-Storm GS1 (Gene Technologies, Braintree, UK) or Thermo Hybaid PxE (Thermo Fisher) thermal cycler with the following conditions: one incubation at 95°C for 5 minutes, followed by 40 cycles of 95°C for 30 seconds, primer set specific annealing temperature of 51°C (*HAND2*), 52°C (*HOXD4*), 53°C (*EPHA10*), 55°C (*APOE, ESR1, HLXB9, TUSC3*), 58°C (*HOXA5, MLH1, APP, CYGB, ERN2, HAP1, SH2D4A*), 61°C (*p15, RARB2*) or 63°C (*HOXA4, IGF2, RASSF1A, TWIST2*) for 30 seconds and 72°C for 30 seconds followed by a final elongation step of 72°C for 5 minutes (Table 2.3). Forty cycles of PCR were performed to reduce the presence of unused biotin labelled primer. Amplification of all DNA samples was checked prior to Pyrosequencing analysis by gel

electrophoresis using 1.5% agarose (Alpha Laboratories, Hampshire, UK) gel in 1xTAE (Tris base, acetic acid and EDTA) (Appendix A1.2) and visualised with GelRed™ (Cambridge BioScience, Cambridge, UK) staining.

Stage	Temperature (°C)	Time (seconds)	Cycles
Hot Start	95	5	1
Denaturing	95	30	40
Annealing	Primer set specific	30	
Elongation	72	30	
Final Elongation	72	5	1
Hold	4	∞	

Table 2.3: Thermal cycler conditions for PCR reactions

2.5 Pyrosequencing

Quantification of DNA methylation levels was carried out by Pyrosequencing analysis using the PyroMark MD instrument (Qiagen, Crawley, UK). Pyrosequencing is based on sequencing by synthesis approach and is routinely used for genotyping, as well as DNA methylation analysis. This technique utilises a biotin labelled, single stranded DNA template to which a sequencing primer is annealed followed by synthesis of the complementary DNA strand. Firstly, dNTP's are added in turn and DNA polymerase catalyses the insertion of complementary nucleotides. Upon nucleotide incorporation, a pyrophosphate (PPi) is released in an equivalent proportion to the number of incorporated nucleotides. The PPi is converted to ATP by the ATP sufurylase enzyme in the presence of adenosine 5' phosphosulfate (APS). This ATP becomes a substrate for luciferase to convert luciferin to oxyleuciferin and generates light. Light is detected and presented as a peak in the data output file (pyrogram) and the intensity of the light emission and peak corresponds to the frequency of that particular nucleotide in the sequence, allowing the complementary DNA strand sequence to be determined. The apyrase enzyme degrades unincorporated nucleotides and ATP before the next nucleotide is added so that the light signal can be correctly attributed to a specific nucleotide (Ahmadian *et al.*, 2006) (Figure 2.3). For DNA methylation analysis, methylation status at a CpG site is determined by the presence of either a C

(methylated) or T (unmethylated) nucleotide in the sequence (or a G (methylated) or A (unmethylated) nucleotide if the complementary strand is sequenced), and the ratio of C:T allows quantification of percentage methylation at a specific CpG site. All Pyrosequencing assays attempt to incorporate non complimentary bases into the newly synthesised strand to confirm the specificity of the assay, and that only expected nucleotides are being integrated. In addition, a bisulfite control is performed at a non CpG cytosine base (which should always be converted to thymine after sodium bisulfite modification) by also adding cytosine nucleotides at this position. If any become incorporated into the synthesised strand, this indicates that sodium bisulfite modification of DNA was incomplete and methylation measurements maybe inflated and inaccurate. Samples with cytosine levels >4.5% detected at this position are automatically failed. Biotin-labelled Pyrosequencing PCR products were bound to streptavidin coated sepharose beads (GE Healthcare Life Sciences) and made into single stranded DNA templates by performing strand separation using a vacuum pump tool on the Pyromark Q96 Vacuum Workstation (Qiagen, Crawley, UK). The vacuum tool picks up the PCR product, bound to the beads which are then washed with 70% ethanol. Next, the vacuum tool is added to sodium hydroxide (NaOH) to denature the DNA, causing the biotin labelled strand to remain associated with the beads and vacuum, whilst the other strand is removed. After washing with a 10mnM tris acetate buffer (Qiagen, Crawley, UK), the vacuum is switched off and the biotin labelled single stranded DNA dissociates and is added to a Pyrosequencing plate containing 10pmol of gene specific sequencing primer (Eurofins MWG Biotech, London, UK), which becomes annealed to this template. Pyrosequencing reaction enzymes, substrates and nucleotides were obtained from the PyroGold Q96 Reagent kit (Qiagen, Crawley, UK). A full protocol can be found in the appendix (A1.3). Each DNA sample was measured in duplicate. For each gene, methylation was measured at multiple consecutive CpG sites (3-6), each of which are shaded grey on the resulting pyrogram (see example Figure 2.4), and the average amount of methylation pooled across all CpG sites was calculated. The PyroMark MD instrument scores the results of the methylation analysis by a colour coded system which is visible on the pyrogram output file. In the following example pyrograms, blue colouring represents a sample which passes necessary quality controls including a check for complete bisulfite conversation (Figure 2.4), yellow samples have warnings and need to be manually checked for quality (Figure 2.5) and red samples are judged to have failed the Pyrosequencing analysis so should be discarded and repeated (Figure 2.6).

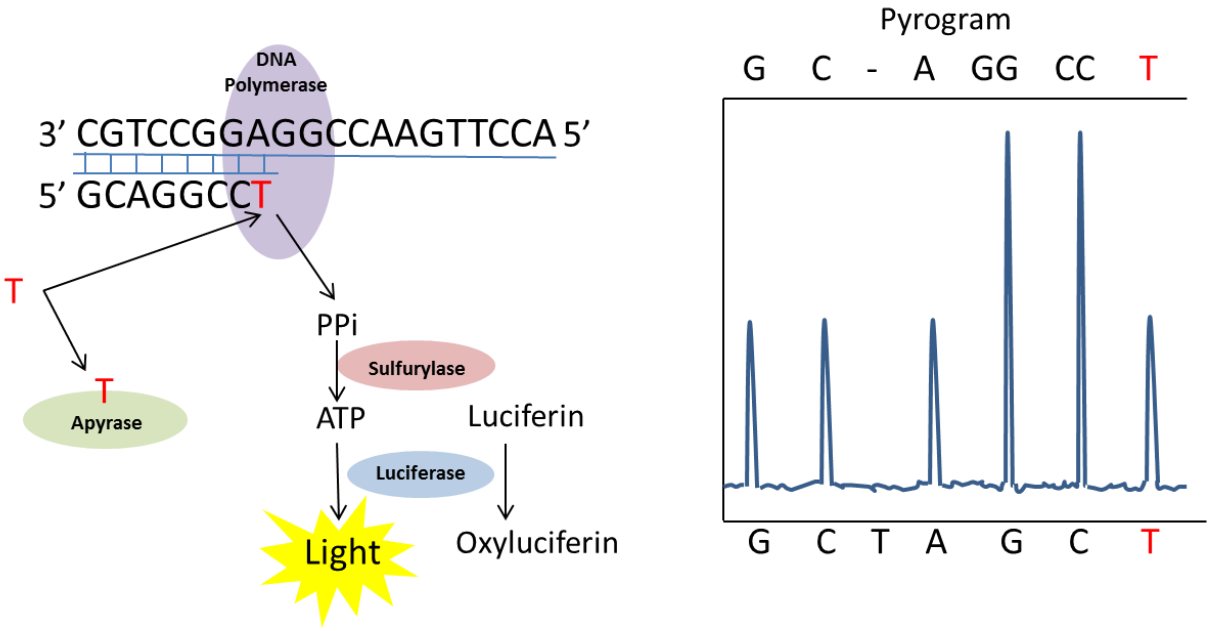


Figure 2.3: Pyrosequencing reactions and example output pyrogram

After the incorporation of a complementary nucleotide, a series of reactions trigger the emission of light. This light is detected and presented as a peak on the pyrogram output. Height of the peaks is directly proportional to the amount of incorporated nucleotide. Non complimentary bases do not generate light or a peak. From the pyrogram, the sequence of the complementary DNA strand can be determined.

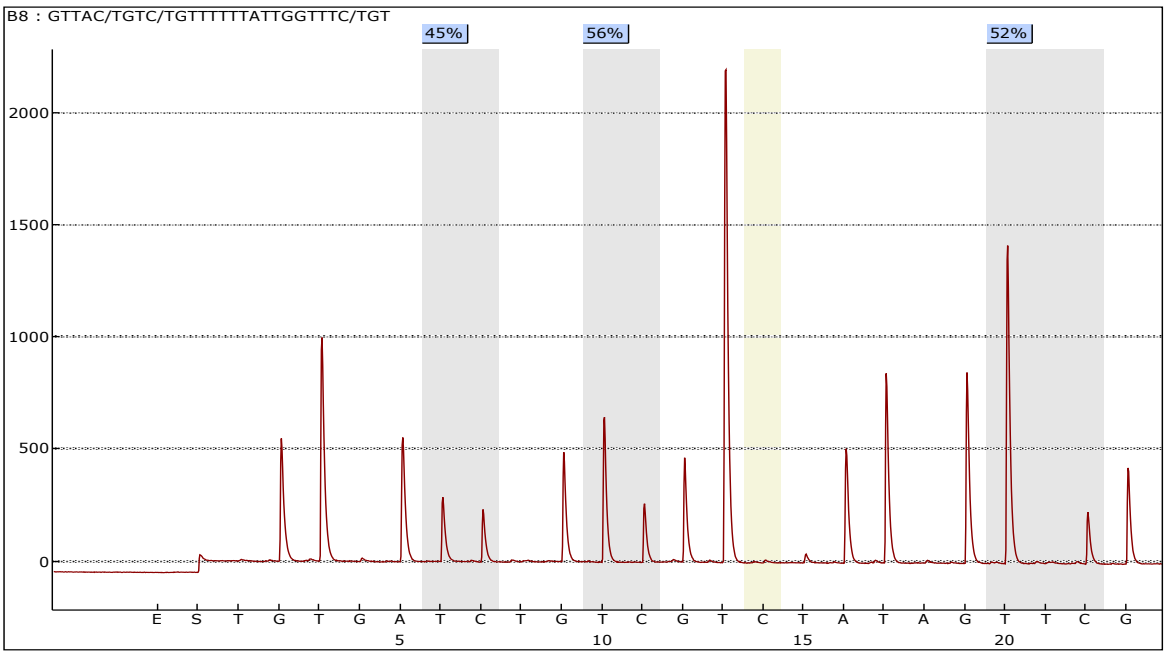


Figure 2.4: Pyrogram of a passed *IGF2* Pyrosequencing assay sample

The pyrogram features three CpG sites, and the level of DNA methylation (%) at each particular CpG site (shaded grey) is calculated by the ratio of cytosines and thymines incorporated into newly synthesised DNA strand. The blue colour of the methylation percentage indicates the sample has passed quality control checks, including the specific incorporation of expected nucleotides. The site at which complete bisulfite conversion of DNA is checked is shaded yellow.

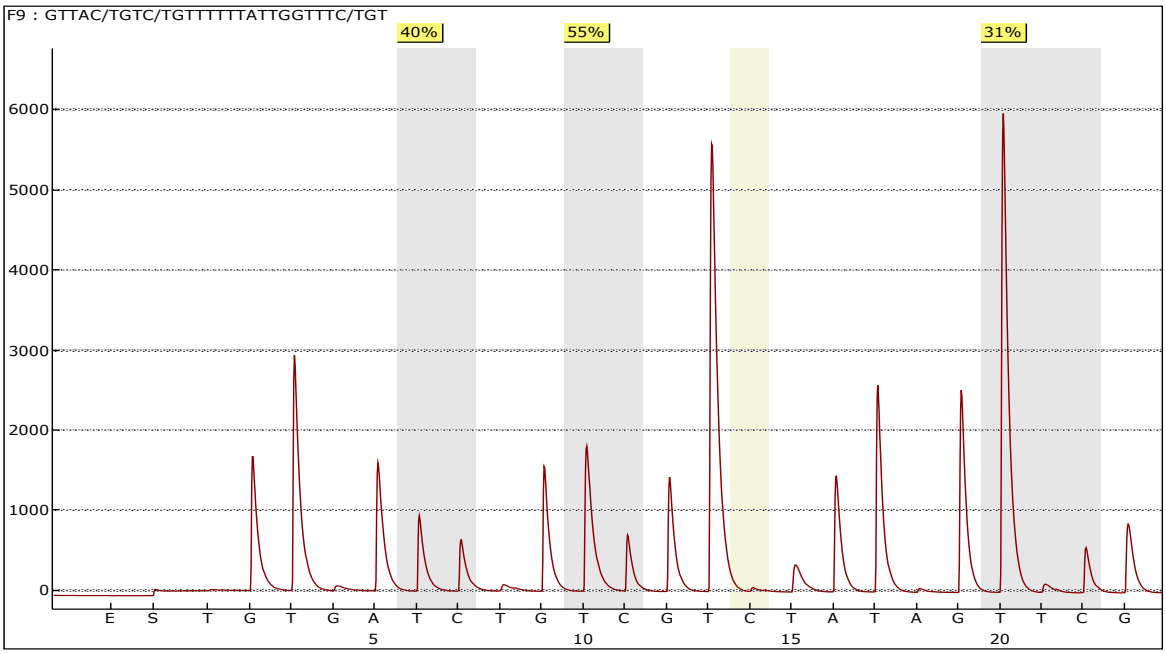


Figure 2.5: Pyrogram of an *IGF2* Pyrosequencing assay sample with warnings

The yellow colour of the methylation percentage calculations indicate that there may be a problem with some of the quality control checks performed on this sample. In this case, the warning was for wide peaks, which is caused by too much product in the reaction. Samples with warnings should be repeated to ensure methylation measurements are accurate.

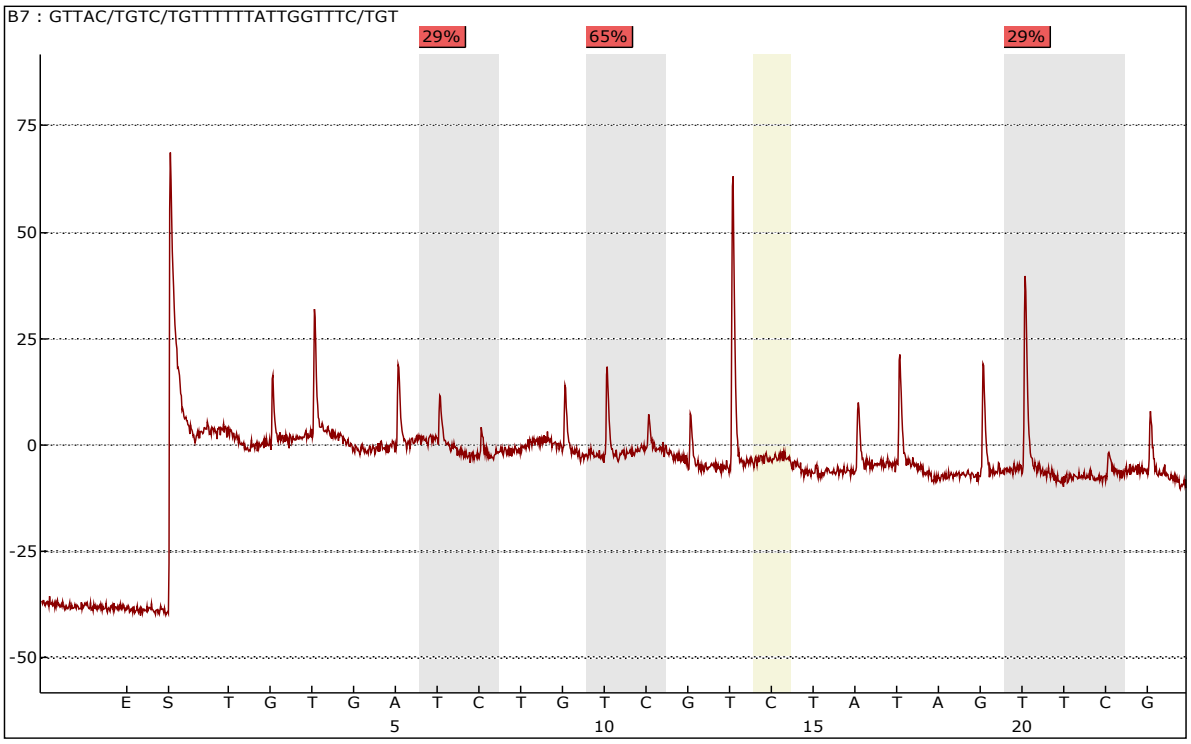


Figure 2.6: Program of a failed *IGF2* Pyrosequencing assay sample

Methylation measurements which are coloured red are deemed to have failed quality control checks. In this case, the heights of the peaks generated by nucleotide incorporation are very low. This is due to insufficient product in the analysis and these results are likely to be highly unreliable.

2.5.1 Pyrosequencing validation

Each Pyrosequencing assay was individually validated to ensure that the observed levels of methylation measured by the Pyrosequencer are accurately reflecting the true amount of DNA methylation. Measured methylation data from the Pyrosequencer is plotted against the expected methylation of known standards to give a validation curve. Standards of 0%, 5%, 10%, 15%, 20%, 30%, 40%, 50%, 60%, 70% and 100% methylation were created by diluting CpGenome™ Universal Methylated Control DNA (Millipore Ltd, Watford, UK) which represents a human *in vitro* 100% methylated sample, with DNA from peripheral blood lymphocytes from a young volunteer, which represents mostly unmethylated CpG islands. These known methylation DNA standards were subjected to bisulfite modification prior to Pyrosequencing.

2.6 Sodium bisulfite sequencing of single PCR amplified DNA molecules

In order to decipher methylation status at all allelic level, *de novo* sequencing of a single PCR amplified sodium bisulfite modified DNA molecule (which has a sequence equivalent to a single allele in the original DNA sample) was performed. This analysis has two advantages over Pyrosequencing; one that it allows determination of the methylation status of all CpG sites within the amplified region (a greater number of CpG sites than analysed by Pyrosequencing) and secondly that it allows determination of methylation is distributed amongst individual alleles. Firstly, the sample was subjected to PCR amplification and cloning of the PCR products into competent E. coli cells using the TA Cloning® Kit (Invitrogen Ltd, Paisley, UK). Briefly, 1µl of PCR product was added to 1µl 10x Ligation buffer, 2µl pCR®2.1 vector and 1µl T4 DNA Ligase for overnight incubation at 14°C. 2µl of the ligation product was added to 50µl of One Shot® aTOP10 Chemically Competent E. coli cells and incubated on ice for 30 minutes followed by transformation by heat shock treatment of 42°C for 30 seconds. Cells were briefly shaken (225 rpm) in 250µl S.O.C. medium at 37°C for 1 hour prior to plating of 80µl and 160µl of cells on LB agar containing kanamycin (50µg/ml) (appendix A1.4) for overnight incubation. Transformants were selected by culturing individual colonies overnight at 37°C in 5 ml of LB media with kanamycin (50µg/ml) (Appendix A1.4). Vector DNA was isolated and purified by lysis of bacterial cells and adsorption of DNA using QIAprep Spin Miniprep kit (Qiagen, Crawley, UK) according to the manufacturer's protocol and eluted in 50µl of elution buffer (Appendix A1.5). The presence of the PCR amplified region of interest was confirmed by EcoRI (Fermentas, York, UK) restriction enzyme digestion which cuts the vector either side of the incorporated PCR product. Ten microlitres of DNA was added to 0.5µl EcoRI enzyme, 2µl Buffer H and 2µl BSA (10x) in a final reaction volume of 20µl and incubated at 37°C for 2 hours. Samples were separated by gel electrophoresis using 2% agarose gel in 1xTAE and visualised with GelRed™ (Cambridge BioScience, Cambridge, UK) staining. The DNA concentration of vectors containing the insert was determined using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific) and all samples were sent to Source Bioscience, at a concentration of 100ng/µl, for Sanger sequencing (Source Bioscience LifeSciences plc, Nottingham, UK). The methylation status of all CpG sites within the amplicon was determined by the DNA sequence after bisulfite modification, TG indicating the cytosine was unmethylated, CG representing a methylated cytosine.

2.7 Samples for genome wide DNA methylation analysis

2.7.1 Cell lines

A total of four human lymphoma cell lines were used for genome wide methylation analysis; DB, HT, Karpas-422 and SUDHL8. All four are diffuse large B-cell lymphoma cell lines, a subtype of B-cell non-Hodgkin lymphoma (B-NHL).

2.7.2 Primary acute lymphoblastic leukaemia (ALL) samples

The methylation status of gene targets identified from the genome wide methylation analysis was confirmed in 18 primary adult ALL primary samples. Age and ALL immunophenotype associated with each sample can be found in Table 2.4.

2.8 Cell culture of lymphoma cell lines

A total of four lymphoma cell lines were cultured; DB, HT, Karpas-422 and SUDHL8. All cell lines were maintained in RPMI 1640 with 10% fetal bovine serum (FBS) and 1% PenStrep antibiotic in 95% air/5% CO_2 at 37°C. All cell lines were grown in suspension in T75 flasks containing 15ml RPMI media, with FBS and PenStrep at a density of 0.5-2x10^6 cells/ml and passaged every 3-4 days at a ratio of 4:1.

2.9 DNA extraction

DNA was extracted from 4-5 million pelleted cells using a phenol chloroform extraction, by adding 340µl of Reagent B (Nucleon – Appendix A1.6) to lyse cells followed by RNase treatment for degradation of RNA using 2.5µl RNase A (50µg/ml) solution and incubation at 37°C for 30 minutes. Then, 100µl of 5M Sodium Perchlorate was added and samples incubated at 37°C for 20 minutes followed by 65°C for 20 minutes, to prevent precipitation of proteins alongside DNA. A combination of 440µl phenol and 440µl chloroform was added to samples, mixed and centrifuged at full speed for 10 minutes, which results in separation into two distinct phases, a lower phase of phenol associated with proteins, and an upper aqueous phase containing DNA. The upper phase was removed into a new microcentrifuge tube and 880µl of chloroform added followed by centrifugation at full speed for 10 minutes to remove lipids and

ensure complete removal of the phenol phase. The supernatant was again transferred to a new microcentrifuge tube and DNA precipitated with 2 volumes of cold 100% ethanol, followed by centrifugation at full speed for 10 minutes. If no DNA precipitate was visible upon addition of the ethanol, samples were incubated overnight at -20°C prior to centrifugation. Finally, samples were washed with 70% ethanol and centrifuged at full speed for 5 minutes. The DNA pellet was dried briefly and resuspended in appropriate volume of TE buffer (Appendix A1.7). DNA concentration was tested using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific).

Sample ID	Age at Diagnosis	Immunophenotype
1628	48	T
3322	59	CALLA
3517	49	T
3684	46	PRE-PRE B
4508	49	CALLA
5031	71	CALLA
6095	70	B
6654	19	T
7431	23	PRE-PRE B
7769	15	CALLA
8849	19	CALLA
10140	20	CALLA
10226	40	PRE B
12032	63	B
12439	23	CALLA
12868	65	CALLA
12896	29	CALLA
15566	20	PRE B

Table 2.4: ALL sample characteristics

For each sample, age at diagnosis and immunophenotype, determined by detection of expressed cell surface markers is shown. B = B-cell ALL, CALLA = common acute lymphoblastic leukaemia antigen, PRE B = Precursor B-cell ALL, PRE-PRE B = Precursor B-cell ALL, T = T-cell ALL

2.10 Preparing DNA fragments for sequencing

2.10.1 Bioruptor® Plus sonication system

The Bioruptor® Plus sonication system (Diagenode SA, Liege, Belgium) allows up to six samples in sealed 1.5ml tubes to be sonicated simultaneously. This system utilises ultrasound energy to stress and fragment DNA whilst the tubes containing DNA samples are rotated in an iced water bath, allowing an equal distribution of energy. Conditions were optimised prior to analysis to determine cycle length, number of cycles and the amount and concentration of DNA required to reproducibly yield fragments of the required 300-500 base pairs in length. The 1.5ml microcentrifuge tubes required a volume of at least 100µl of DNA for successful sonication, and the optimal concentration was found to be 100ng/µl, therefore a total of 10µg of DNA was used for sonication, comprising of 2.5µg of DNA extracted from each of the four lymphoma cell lines DB, HT, Karpas-422 and SUDHL. Samples were vortexed vigorously for 5 seconds prior to sonication was this was found to improve the consistency of fragment size between different samples during testing. The sonication was performed on the high power setting and the programme consisted of 20 cycles of power on for 30 seconds, then power off for 30 seconds. 500ng of sonicated DNA was subjected gel electrophoresis using 1% agarose gel in 1xTAE, visualised with GelRed™ staining and size determined with 1 Kb Plus DNA Ladder (Invitrogen Ltd, Paisley, UK) to check DNA fragment distribution.

2.10.2 Covaris Instrument

The Covaris instrument (KBiosciences, Hoddesdon, UK) utilises Adaptive Focused Acoustics™ technology, using shorter wavelengths of energy which can be focused more precisely on the sample. It is more efficient than traditional sonication, where energy is often lost as heat and samples can be exposed to overheating. To further fragment DNA which was larger than the required 200-300bp size for Illumina GAIIx sequencing, a total of 6.7µg of DNA was input into the optimised Covaris instrument with the assistance of Dr Thahira Rahman, with the following settings:
Duty cycle 20% (percentage of time energy output is “on”)
Intensity 4.0 (acoustic power)
Cycles per burst 200 (number of pressure waves in a burst)

Time 60 seconds x2

Temperature 6°C

2.10.3 Fragment profile measurements

Sizing of DNA fragments was measured with the assistance of Dr Thahira Rahman after Bioruptor® Plus sonication, Covaris sonication and upon ligation of sequencing adapters on the 2100 Bioanalyzer (Agilent Technologies UK Ltd, Cheshire, UK). Reagents, including DNA dye and gel matrix, and DNA LabChip were obtained from the Agilent DNA 1000 Kit (Agilent Technologies UK Ltd, Cheshire, UK) for sizing of DNA in 25-2000bp range. The 2100 Bioanalyzer assay programme used was electrophoresis → dsDNA → DNA 1000. The sonicated DNA sample and reagents were added to appropriate wells on the DNA LabChip using the Chip Priming Station (Agilent Technologies UK Ltd, Cheshire, UK). Sizing accuracy was confirmed by the inclusion of a size ladder, and two internal markers at 15bp and 1500bp.

2.10.4 DNA purification

DNA was purified after fragmentation, during sequencing library preparation and after MethylCap enrichment of DNA to remove very short nucleotide fragments <40bp, primers, salts and enzymes using the QIAquick PCR Purification Kit (Qiagen, Crawley, UK) according to the manufacturer's protocol. DNA fragments are bound to the column membrane in high salt concentrations (pH<7.5) whilst contaminants are lost, and DNA is eluted once pH is increased by the addition of elution buffer (pH 8.5). The full protocol is listed in the appendix (A1.8).

2.11 Sequencing library preparation

Sequencing library preparation was performed with the assistance of Dr Thahira Rahman. The lymphoma cell line DNA sample was prepared for sequencing using the Paired-End DNA Sample Preperation Kit (Illumina) to add adapter sequences onto the ends of DNA fragments for 75bp paired end reads.

2.11.1 Perform end repair

After DNA fragmentation, the ends of DNA are not uniform and are comprised of both 3' and 5' overhangs and blunt ends. All fragments were made into blunt ends by filling in 5' overhangs and removing 3' overhangs by T4 DNA polymerase and Klenow enzyme. T4 PNK catalysed phosphorylation at the 5' blunt end of DNA allowing subsequent ligation. A total of 6.7µg DNA was incubated with 10µl DNA Ligase Buffer with 10mM ATP, 4µl of 10mM dNTP Mix, 5µl of T4 DNA Polymerase, 1µl Klenow Enzyme and 5µl T4 PNK in a total reaction volume of 100µl was incubated at 20°C for 45 minutes. Samples were purified with QIAquick PCR Purification kit (Qiagen, Crawley, UK) in 32µl of Qiagen EB (Appendix A1.8).

2.11.2 Adenylate 3' ends

The 3' end of the blunt DNA fragment was adenylated with a single 'A' nucleotide, which prevents fragments ligating to each other. The adapter sequences contain a complementary 'T' nucleotide for ligation with the DNA fragment. The 32µl DNA sample was incubated at 37°C for 45 minutes with 5µl Klenow buffer, 10µl mM dATP and 3µl Klenow Exo - in a total volume of 50µl. Samples were purified with QIAquick MinElute PCR Purification kit (Qiagen, Crawley, UK) according to manufacturer's protocol and eluted in 10µl Qiagen EB (Appendix A1.8). The protocol for the QIAquick MinElute PCR Purification kit is identical to the QIAquick PCR Purification kit but includes a specifically designed column which allows elution of DNA in very small volumes.

2.11.3 Ligate adapters

In this step, the sequencing adapters were ligated to the ends of the DNA fragments. Distinctive sequences were added to the 5' and 3' ends of each strand. The 10µl purified DNA sample was incubated at 20°C for 45 minutes with 25µl DNA Ligase Buffer 2X, 10µl PE adapter Oligo Mix and 5µl DNA Ligase in a total volume of 50µl.

2.11.4 Purify adapter ligated fragments

Purification of the adapter ligated DNA fragments was performed using Agencourt AMPure XP magnetic beads (Beckman Coulter). DNA and magnetic beads were incubated for 30 minutes at room temperature prior to separation of the DNA bound magnetic beads in a magnetic rack. Beads were washed with 70% ethanol and dried. DNA was eluted by adding 17µl of nuclease free H_2O followed by incubation at room temperature for 30 minutes before placing on a magnetic rack and removing the DNA containing supernatant.

2.12 Enrichment of methylated sequences

In order to recover DNA fragments containing methylated regions within total genomic DNA, the MethylCap kit (Diagenode SA, Liege, Belgium) was utilised to selectively bind fragments containing methylated CpG sites. This technique utilises a H6-GST-MBD fusion protein, comprising of a C-terminal methyl binding domain (MBD) isolated from the human methyl binding protein MeCP2, with Glutathoione-S-transferase (GST) and an N-terminal His6-tag, for the capture and isolation of DNA fragments that contain one or more methylated cytosines.

2.12.1 Capture of methylated DNA

Firstly, the capture reaction mix was prepared by adding 1.2µg of the adapter ligated DNA fragments at a concentration of 100ng/µl to 129.8µl of buffer B in a total volume of 141.8µl. Per capture reaction, 119µl of the capture reaction mix was mixed with 1µl H6-GST-MBD fusion protein and incubated on a rotating wheel at 40rpm at 4°C for 2 hours; forming a complex between the H6-GST-MBD fusion protein and methylated DNA. The remaining excess capture reaction mix of 22.8µl served as an INPUT control sample. Thirty microlitres of meDNA Capture beads were washed several times with Wash Buffer 1 prior to their addition to the capture reaction mix containing H6-GST-MBD fusion protein and methylated DNA and subsequently incubated on a rotating wheel at 40rpm at 4°C for 1 hour.

2.12.2 Bead washes

After the capture reaction, samples were placed in a Magnetic Rack (Diagenode SA, Liege, Belgium), and the buffer containing DNA which has not bound to magnetic beads was removed. The methylated DNA bound magnetic beads were washed several times by incubation with Wash Buffer 1 and then Wash Buffer 2, on a rotating wheel at 40rpm at 4°C for 5 minutes.

2.12.3 Elution of captured DNA

The MethylCap kit includes three elution buffers of progressively increasing salt concentration which enable differential fractionation of double-stranded DNA based on increasing CpG methylation density. To elute captured DNA, 150µl of low salt buffer was added to magnetic beads and incubated on a rotating wheel at 40rpm at 4°C for 10 minutes, then placed in magnetic rack. The supernatant solution contained methylated DNA fragments eluted by the low salt fraction. These steps were repeated with medium and high salt buffers and the INPUT DNA (no enrichment for methylated sequences) was eluted with the high salt buffer. DNA from all fractions was purified with QIAquick PCR Purification Kit (Qiagen, Crawley, UK) according to the manufacturer's protocol and eluted in 30µl ddH_2O (Appendix 1.8).

2.13 Quantitative PCR (qPCR)

Successful enrichment for methylated fractions of the genome was confirmed by measuring the amount of DNA template present in each of the salt eluted fractions at a known methylated (*HOXA5*) and unmethylated (*GAPDH*) gene by qPCR. Reactions were carried out in a final volume of 10µl with 5µl SYBR® Green JumpStart™ Taq ReadyMix (Sigma, Gillingham, UK), 37.5ng of forward and reverse primer and 0.5µl of DNA template and the reaction performed by the TaqMan 7900HT (Applied Biosystems, Warrington, UK) with the following cycling conditions; one cycle of 94^{o}C for 15min, followed by cycles of 94°C for 30sec, 57°C (*GAPDH*) or 64°C (*HOXA5*) for 30sec and 72°C for 30sec, with plate reads carried out at 86°C (*GAPDH*) or 91°C (*HOXA5*). Each sample was tested in triplicate. The primer sequences for *HOXA5* were forward primer 5'TCGGAAGCTGGGCGATGAG and reverse primer 5'GTGCACTAATAGGGGAGTTGGG and for *GAPDH* were forward primer

5’CTCAGGCCTCAAGACCTTG and reverse primer 5’GACAGTCAGCCGCATCTTC.

2.14 Illumina Genome Analyzer IIx sequencing

Paired end, high throughput sequencing of MBD enriched DNA fragments was performed on the Illumina Genome Analyzer IIx (GAIIx) sequencing platform using a sequencing by synthesis approach. Library prepared samples were placed on a flow cell where the single stranded fragments anneal to oligos on the surface via the complementary adapter ligated sequences. PCR was performed whereby double stranded bridge sequences were formed on the flow cell surface, generating millions of clusters containing clonally amplified fragments utilising the TruSeq PE ClusterKit V5 CS-GA (Illumina). Reverse strand fragments were removed, forward strand ends blocked and sequencing primers annealed to the forward strand template to begin sequencing of reverse strand (read 1). Clusters were simultaneously sequenced on the GAIIx platform. Paired end 75bp sequencing reads were performed on the Genome Analyzer IIx (GAIIx) platform using 2x TruSeq SBS Kit V5-GA 36 cycles (Illumina). Briefly, fluorescently labelled, reversibly terminated dNTPs and polymerase were added to the synthesis reaction. After single nucleotide incorporation, the base is identified by colour imaging and the fluorescent dye removed along with reversible terminator. This process continues for a further 74 cycles. After this, cluster generation was repeated utilising the paired end module of the GAIIx platform, to enable synthesis of the forward strand (read 2). Paired end sequencing allows greater confidence in identifying fragments, as both 5’ and 3’ ends of the fragment can be mapped back to the genome, as opposed to a single read, enabling a more precise alignment of reads.

2.15 Bioinformatics analysis

Analysis was performed by Dr Simon Cockell from the Newcastle University Bioinformatics support unit. Raw sequencing data from the Illumina GAIIx platform in FASTQ format was subjected to quality control analysis by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which removes problem sequencing reads. This analysis considers quality scores along the length of the 75bp read, nucleotide content and proportion of N (undetermined bases) in the sequencing reads. Filtered sequencing reads were mapped to the human reference genome using the

Burrows-Wheeler Aligner (BWA) (http://bio-bwa.sourceforge.net/) program, for rapid identification of reads. Regions of genome which were enriched by MethylCap analysis were determined by the peak calling algorithm, Model-based Analysis of ChIP-Seq (MACS) (http://liulab.dfci.harvard.edu/MACS/). The alignment files were used to identify regions which deviate from an expected Poisson distribution of aligned sequence reads, generating approximately 150,000 peaks. Peaks were filtered to exclude those in repetitive regions, as these are more likely to be sequenced and are usually highly methylated. Remaining peaks were intersected with a gene list to assign each peak to its closest gene. Only peaks which were within 300bp of, or overlapped, the transcriptional start site of a gene were retained, leaving 10,000 peaks. Peaks were assigned to their respective genes according to the above criteria, and those which matched to pseudo genes were removed. This left a total of 7,266 genes, of which approximately 1,500 were RNA genes (particularly ribosomal), of which a significant proportion are thought to be methylated in normal cells (Gagnon-Kugler *et al.*, 2009). Remaining peaks were assigned into the top or bottom 10% of all sequenced peaks based on their p-values, representing a confidence score of the observed peak, determined by MACS, generating a list of genes associated with the top and bottom 10% of the filtered peaks.

2.16 COBRA assay

Validation of the methylation status of gene sequences from high-throughput sequencing analysis was performed by the COmbined Bisulfite Restriction Analysis (COBRA) assay (Xiong and Laird, 1997). The genes for validation were chosen at random (using the random number generator function in Microsoft Excel) and primers were designed to produce equivalent amplification of unmethylated and methylated DNA, either by ensuring that the primer sites did not contain any potentially methylated sites or contained a "wobble" base at any CpG sites which were included (i.e. either Y instead of C or R instead of G at the CpG site) (Table 2.5). PCR amplification was performed as described in section 2.4, with primer specific annealing temperatures of 55°C (*CELF4*), 58°C (*APP, CACNA1B, CYGB, ERN2, HAP1, SH2D4A*), 61°C (*HCK, TCL1B*) or 63°C (*EBF4, PDCD5*) followed by digestion with restriction enzymes to discriminate between methylated and unmethylated sequences after sodium bisulfite modification. The identification of appropriate restriction enzymes was determined using NEBcutter V2.0 software (New England BioLabs, URL:

http://tools.neb.com/NEBcutter2/). The sequence of the DNA amplicon after bisulfite modification was input into the software, with CpG sites remaining as CG, which returns a list of restriction enzymes which will cut somewhere within the sequence. The restriction enzymes utilised for this analysis with their associated recognition sequences can be found in Figure 2.7. Enzymes with recognition sequences containing a cytosine and guanine can therefore be used to distinguish between methylated (CG – will cut) and unmethylated (TG – will not cut) sequences (demonstrated in Figure 2.8). For each COBRA assay, 11.5µl of gene specific PCR product was mixed with 0.5µl of either BsiEI, BstUI, HhaI, HinFI, HpyCH4IV, RsaI or TaqI (New England BioLabs, Hitchin, UK) (Table 2.6), 2µl of the appropriate enzyme buffer and 2µl BSA (10x) in a final volume of 20µl and incubated at 37°C (HhaI, HinFI, HpyCH4IV, RsaI), 60°C (BsiEI, BstUI) or 65°C (TaqI) for 2 hours. Digested PCR products were separated on 2–3% agarose gels, visualized by GelRed™ (Biotium) staining and size determined with 1 Kb Plus DNA Ladder (Invitrogen Ltd, Paisley, UK). Standards of known methylation status (100%, 66%, 33%, and 0%) were created by diluting *in vitro* methylated DNA (Millipore Ltd, Watford, UK) into DNA extracted from normal peripheral blood (typically unmethylated/low levels of methylation) to allow a comparison with tested samples and also for confirmation of efficient restriction digestion.

2.16.1 Development of Pyrosequencing assays corresponding to COBRA assays

A more sensitive technique than COBRA was required for testing the methylation status of genes identified by genome wide DNA methylation analysis in the Newcastle 85+ study samples, as it is not possible to determine small differences in methylation (which is expected in normal human samples) using COBRA. For this reason we developed Pyrosequencing assays corresponding to the same region as the COBRA primers (Table 2.7). Pyrosequencing assay design was carried out as described (see section 2.2).

Gene	Forward Primer	Reverse Primer	Distance from TSS	Size (bp)
	Genes associated with top 10% of sequencing peaks			
APP	TGGGGGTTAAAAAATGAGGTT	TCRTCCCCRTAAACTTAAATCATC	-499 → -280	219
CACNA1B	GGYGGTTAGAGGTTTTTTTGAGTTGAGAT	TAAATACCCTCRACTCCTCAAACTC	-563 → -279	284
CELF4	AGGTTGGATYGGTTTAAGTTTTTAGTTAG	ACRCCTCCCRAAAAACCATAC	-11 → +248	256
CYGB	GGATTTTTTTYGTTATTAGTTGGGTTAGAA	CTCTAAACRCCCAAACTCTAAAAAATT	-656 → -393	263
ERN2	GGGATTYGTTGGTGTAATGATAAAGAG	CCRCCCTTCTCATCACACT	-80 → +182	262
HAP1	TTYGGAGATGGAGAGGGTAGAAAA	TCRAAATCCTACTCTCTATCCAATACC	-243 → +171	414
SH2D4A	GTGGAAAGGTTTAGATTTGTAGATTTAAT	ACCCRAACRCAAAACAACTCTTATA	-190 → +119	309
	Genes associated with bottom 10% of sequencing peaks			
EBF4	GTTTTTTTYGAGGGAGYGTTTAAGAGG	CATCCAAAAACRCACTAAACCCAAAC	+600 → +864	264
HCK	GGGTTTGGGATTTTTTTTGTAGGA	TCCTACACTATCTAAATCCCTTCC	-108 → +129	238
PDCD5	ATAGAGGTTGGTYGAGTTGTAGGTT	AACCCRCTCCAAACTAAAAAACCC	+106 → +404	298
TCL1B	GGAAAGTTATAYGTGTGAGTTTAGAG	ATTACTACRCTCCTAAACCTACAC	+1 → +355	354

Table 2.5: COBRA assay primer sequences of genes associated with sequencing peaks from the genome wide DNA methylation analysis

The forward and reverse primer sequences for each amplicon are shown with the location of primers relative to the transcriptional start site and overall size of the PCR product amplicon (Y = C/T nucleotide, R = A/G nucleotide on reverse strand).

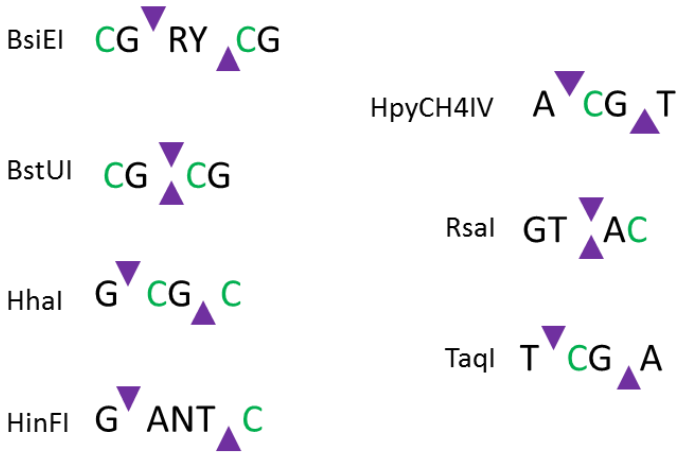


Figure 2.7: COBRA assay restriction enzyme recognition sequences and cut sites

The seven enzymes shown will cut only when the DNA sequence contained a methylated cytosine nucleotide (green C). The BsiEI, BstUI and HhaI enzymes will only cut when two consecutive cytosines are methylated. Cut sites are indicated by purple triangles. (N = any nucleotide, R = either A/G nucleotide, Y = either C/T nucleotide).

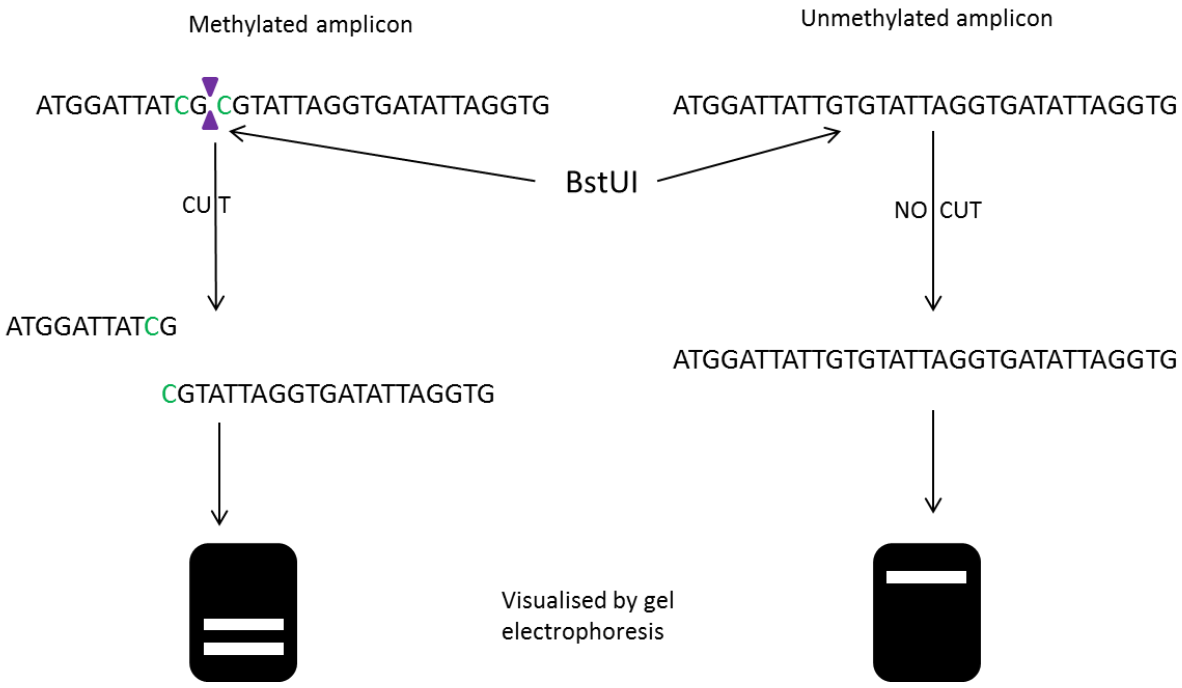


Figure 2.8: COBRA assay example

A sodium bisulfite modified and methylated amplicon sequence will be recognised and cut into smaller fragments by restriction enzymes, whilst unmethylated amplicons remain full length. DNA methylation status can then be determined by sizing of DNA fragments visualised by gel electrophoresis.

Gene	Size (bp)	Restriction Enzymes	Number of cuts	Enzyme cut position
	Genes associated with top 10% of sequencing peaks			
APP	219	HpyCH4IV	2	91/93, 142/144
		TaqI	1	113/115
CACNA1B	284	HinFI	3	119/122, 154/157, 269/272
		BstUI	2	101, 117
CELF4	256	BsiEI	1	126/124
		TaqI	1	203/205
CYGB	263	TaqI	3	114/116, 167/169, 192/194
		BstUI	1	103
ERN2	262	RsaI	2	130, 142
		BstUI	1	227
HAP1	414	HhaI	2	133/131, 286/284
		RsaI	1	383
SH2D4A	309	BstUI	5	235, 137, 222, 258, 271
		TaqI	2	151/153, 178/180
	Genes associated with bottom 10% of sequencing peaks			
EBF4	264	BstUI	3	33, 159, 161
		HinFI	3	50/53, 129/132, 220/223
HCK	238	BstUI	1	130
		HinFI	1	47/50
PDCD5	298	BstUI	3	69, 101, 208
		HhaI	1	101/99
TCL1B	354	BstUI	2	176, 279
		TaqI	1	58/60

Table 2.6: Restriction enzymes used in COBRA assays

For each assay, the size of the amplicon, the two restriction enzymes used, the number and position of cuts in the sequence are shown.

Gene	Forward Primer	Reverse Primer	Sequencing Primer	CpG Sites
APP	TGGGGGTTAAAAAATGAGGTT (bio)	Same as COBRA	CCCCRTAACTTAAATCA	3
CACNA1B	No suitable Pyrosequencing assay could be designed			
CELF4	No suitable Pyrosequencing assay could be designed			
CYGB	Same as COBRA	CTCTAAACRCCCAAACTCTAAAAAATT (bio)	GGTYGGAGGTGYGTTTAAG	4
ERN2	Same as COBRA	CCRCCCTTCTCATCACACT (bio)	GTGTTTGGTGGGGTGATT	3
HAP1	Same as COBRA	TCRAAATCCTACTCTCTATCCAATACC (bio)	ATTYGGGGATTTAGTAGTAT	3
SH2D4A	Same as COBRA	ACCCRAACRCAAAACAACTCTTATA (bio)	GGTTTTGGAGAGTTTTTAG	5

Table 2.7: Primers for Pyrosequencing analysis of genes identified from genome wide methylation analysis

For *APP, CYGB, ERN2, HAP1* and *SH2D4A* the sequence of the primer and reverse are indentical to those used for COBRA assays with the addition of a biotin label (bio) on either the forward or reverse primer. Sequencing primers were generated by the PSQ assay design software. The number of CpG sites measured by the assay is also shown. No suitable Pyrosequencing assays could be designed for either *CACNA1B* or *CELF4*

2.17 Data analysis and Statistics

2.17.1 Bonferroni correction

This is applied in order to compensate for the effects of multiple testing and comparisons, which increase the likelihood of discovering a significant effect purely by chance. The significance level of the p value is altered from a standard 0.05 level, to a more stringent level by the following formula:

$$\text{Bonferroni significance level} = \frac{0.05 \text{ significance}}{\text{Number of comparisons}}$$

2.17.2 Pearson Correlation coefficient

The strength of correlation between two data variables was measured by calculating the Pearson Correlation coefficient (r) and corresponding p value for determining significance of the relationship. The Pearson Correlation coefficient was calculated by Wessa.net statistical software version 1.1.23-r7 (URL http://www.wessa.net/) (Wessa, 2012).

2.17.3 Linear regression analysis

Linear regression analysis was performed in Excel 2010 (Microsoft Office) to investigate the relationship between two variables. The r^2 value of the regression indicates how well the regression line fits the actual data. An r^2 value of +1 or -1 represents a perfect linear positive or negative relationship between the variables respectively, whilst an r^2 value of 0 signifies no relationship.

2.17.4 Student's t-test

An Independent-Samples Student's t-test was performed using SPSS Statistics (version 19) for testing the hypothesised differences in DNA methylation levels between groups. The null hypothesis (no difference between groups) can be rejected if a significance level below 0.05 is reached.

2.17.5 Z-Score

The Z-score is an indication of how variable a measurement is compared to the population mean measurement, by determining how many standard deviations separate the measurement and the mean, calculated by the following formula:

$$\text{Z-Score} = \frac{(\text{Individual measurement} - \text{Population mean})}{\text{Population SD}}$$

2.17.6 Linear regression analysis model

2.17.6.1 DNA methylation levels and clinical outcomes in the Newcastle 85+ study

Construction of a step-down linear regression analysis model was performed with the assistance of Andrew Kingston (Newcastle 85+ study statistician), using SPSS Statistics (version 19). The methylation level of each gene, *HOXD4, TUSC3, TWIST2, HAND2, EPHA10,* the overall mean of these 5 genes and *HOXA5* were individually analysed as the dependent variable whilst other variables including lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking, history of cardiovascular events, type II diabetes, dementia, stroke and survival were built into the model as independent variables. Non-significant associations were removed (exclusion criteria $p>0.10$) and only the effects of significant associations on methylation levels were retained (retention criteria $p<0.05$).

2.17.6.2 DNA methylation levels and mitochondrial haplogroup in the Newcastle 85+ study

Construction of a step-down linear regression analysis model was performed with the assistance of Mohammad Yadegarfar (Newcastle 85+ study statistician) using SPSS Statistics (version 19). The methylation level of each gene, *HOXD4, TUSC3, TWIST2, HAND2, EPHA10* and the overall mean of these 5 genes were analysed as the dependent variables whilst lymphocyte percentage, gender and mitochondrial haplogroup were built into the model. Non-significant associations were removed (exclusion criteria $p>0.10$) and only the effects of significant associations on methylation levels were retained (retention criteria $p<0.05$).

Chapter 3 – Inter-individual variation of DNA methylation in an elderly population

3.1 Introduction

The extent to which epigenetic marks vary across the population and the impact that such variation has on health and ageing are not clear. Therefore, in order to begin understand the consequence of altered DNA methylation in ageing, we performed a pilot study using DNA extracted from peripheral blood leukocytes of 50 volunteers in the Newcastle 85+ study. Our aim was to quantify the levels of promoter DNA methylation in this population, and to investigate the extent of inter-individual variation in DNA methylation levels at a panel of genes, using a Pyrosequencing-based approach. The purpose of this pilot study was to determine whether there was evidence of significant inter-individual differences in the degree of methylation at specific loci across the genome. Such inter-individual variation may result in reciprocal changes in gene expression and, therefore, in cell function. This could ultimately play a role in determining inter-individual variation in biological ageing and in susceptibility to age related disease.

3.2 Selection of genes for gene specific methylation analysis

A total of 18 genes were selected to act as markers of promoter methylation levels throughout the genome, to enable us to examine both the extent and amount of variation in DNA methylation levels in PBL samples from a population of 85 year olds. This number of genes was chosen on pragmatic grounds since it was a realistic target given the time intensive nature of developing and validating gene specific Pyrosequencing assays.

Several genes were chosen because they had been the focus of previous or on-going study by our group and were known to be hypermethylated in haematological malignancies. These genes included *TWIST2* (Thathia *et al.*, 2012), *TUSC3* (Scholz *et al.*, 2005), *HOXA4* (Strathdee *et al.*, 2007a; Irving *et al.*, 2011), *HOXA5* (Strathdee *et al.*, 2007a), *HOXD4*, *HLXB9* (Ferguson *et al.*, 2011) and *HOXB6* (Strathdee *et al.*, 2007a).

Additional genes were selected from a literature search for genes displaying aberrant DNA methylation in any subtype of leukaemia with some evidence of methylation in normal individuals. These genes included *EPHA10* (Kuang *et al.*, 2010), *HAND2* (Tong *et al.*, 2010), *ESR1* (Issa *et al.*, 1996b), *IGF2, p15* (Wong *et al.*, 2000) and *FZD9* (Martin-Subero et al., 2009) .

Finally, genes were chosen from a literature search for genes which display age related DNA methylation increases in normal tissues. These genes included *APOE* (Wang *et al.*, 2008), *RARB2* (Brait *et al.*, 2009), *RASSF1A* (Dammann *et al.*, 2005), *MLH1* (Nakagawa *et al.*, 2001) and *NKX2-5* (Kwabi-Addo *et al.*, 2007).

3.3 Pyrosequencing assay design

We used Pyrosequencing to examine gene promoter methylation status. This is a highly quantitative method which enables the detection of low levels of, or small changes in, methylation levels with high accuracy. PSQ assay design software was used to design suitable forward, reverse and sequencing primers for each assay, based on the sequences which would be obtained following sodium bisulfite modification of DNA. Primer sets for the candidate genes were designed to amplify a region close to the transcriptional start site, corresponding to the promoter region or first exon of the gene, as this region is the most frequent site of aberrant hypermethylation in ageing (Richardson, 2003) and in cancer cells (van Vlodrop *et al.*, 2011).

Optimisation of PCR was performed to ensure good amplification of products for Pyrosequencing. Several primer sets had to be redesigned at this stage due to inadequate amplification including primers for *ESR1* and *IGF2*. Once a suitable PCR amplicon was obtained, a small number of samples were subjected to Pyrosequencing analysis to ensure correct binding of the sequencing primer and that the nucleotide peaks generated matched the expected sequence. We experienced issues at this stage caused by mispriming of the specific sequencing primer to the amplicon. This problem is due to the decreased sequence specificity caused by sodium bisulfite modification of DNA, which reduces the frequency of unique sequence combinations, and affected several assays including *HOXA4, TUSC3* and *RARB2*. This resulted in inconsistency in the position where the primer anneals on the amplicon, leading to erroneous nucleotide signal peaks which do not match the sequence adjacent to where the sequencing primer

was expected to anneal. These issues were resolved by redesigning new sequencing primers. Only assays which demonstrated strong PCR amplification and yielded the expected sequencing signals were taken forward for validation.

3.4 Validation of Pyrosequencing assays

Validation of each Pyrosequencing assay was carried out using a series of samples of known methylation status, created by diluting *in vitro* methylated DNA (where the methylation level approaches 100%) with peripheral blood lymphocytes DNA from a young volunteer (expected to have very low methylation), to confirm that the methylation values measured by the Pyrosequencer were as expected and that there was no preferential amplification of methylated or unmethylated DNA during PCR.

For each assay, measurements across all the CpG sites (3-6 depending on the assay) were used to calculate an average methylation measurement for the sample. Each sample was tested in duplicate to verify that the assay was reproducible.

Linear regression analysis was used to analyse the relationship between the methylation levels measured by the Pyrosequencer and the expected methylation of the known standards. The r^2 value of the regression indicates how well the regression line fits the actual data points. An r^2 value of 1 indicates a perfect fit. The regression lines for each assay illustrate a strong positive relationship between measured and expected methylation measurements, with all r^2 values close to 1 (Figure 3.1). This confirms that the Pyrosequencing assays provide an accurate measure of true methylation in each assay and that the assays could then be taken forward to assess methylation levels in the Newcastle 85+ study participants.

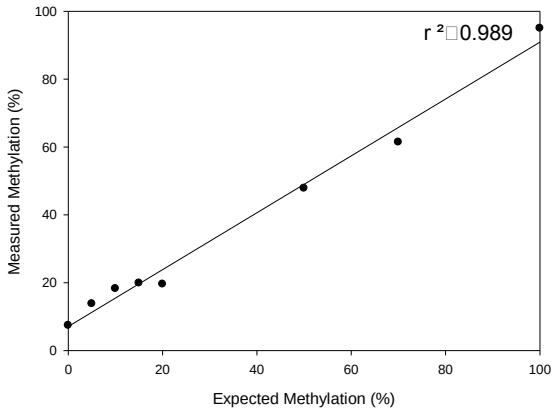
Initial validation was unsuccessful for several of these genes including *HOXA5, TUSC3, TWIST2, FZD9, HOXB6* and *NKX2-5,* as the DNA methylation levels measured by the Pyrosequencer were significantly different from expected values (>5% difference in methylation measurement between PCR repeats of the same sample). Sequencing primer redesign followed by repeated validation was successful for *HOXA5, TUSC3* and *TWIST2*. However, despite attempting alternative primers, we were unable to validate the assays for *FZD9, HOXB6* and *NKX2-5* and found that these assay were not reproducible i.e. repeat assays of the same sample gave variable results which

demonstrated that the assay was unlikely to be reliable for detection of small between-sample differences (from 0-20%) in DNA methylation.

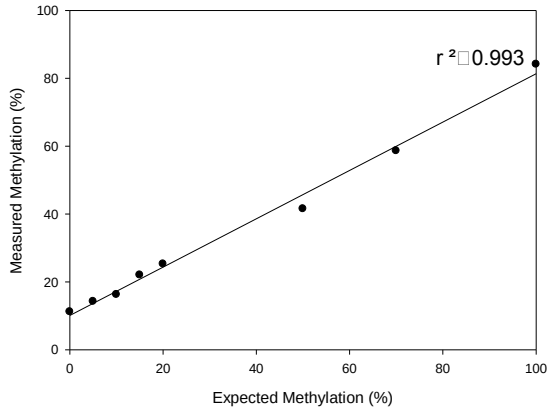
APOE, EPHA10, ESR1, HAND2, HLXB9, HOXD4, MLH1, TUSC3, p15, RARB2, RASSF1A and *TWIST2* are genes which were expected to be either unmethylated, or have very low levels of methylation in normal peripheral blood either due to our own observations, or findings reported by others (Issa *et al.*, 1996b; Cameron *et al.*, 1999; Zemlyakova *et al.*, 2003; Wang *et al.*, 2008; Avraham *et al.*, 2010; Cho *et al.*, 2010; Kuang *et al.*, 2010; Tong *et al.*, 2010; Ferguson *et al.*, 2011; Pineda *et al.*, 2012; Thathia *et al.*, 2012). Validation of Pyrosequencing assays was performed using known methylated controls of 0%, 5%, 10%, 15%, 20%, 50% 70% and 100% methylation, to cover the whole range of methylation levels from zero to 100% and enable quantification of any level of methylation. More controls were included at the lower end of the range, since age related methylation increases in genes with already low levels of methylation may be small.

The validation of these twelve genes demonstrated high reproducibility between repeats and a strongly linear relationship between expected and measured methylation across the whole range of methylation from 0 – 100%. The results also suggested that small differences in methylation (5%) can be successfully detected by these Pyrosequencing assays (Figure 3.1).

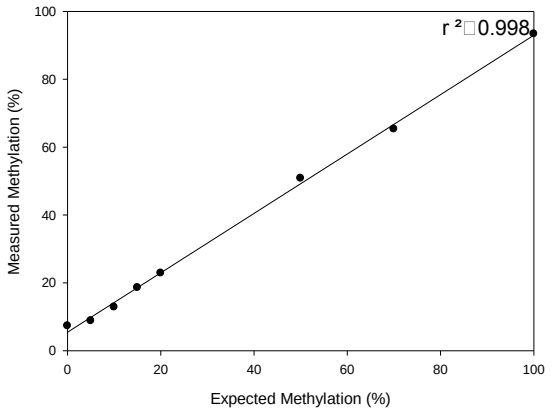
APOE Assay Validation: Expected and Measured Methylation



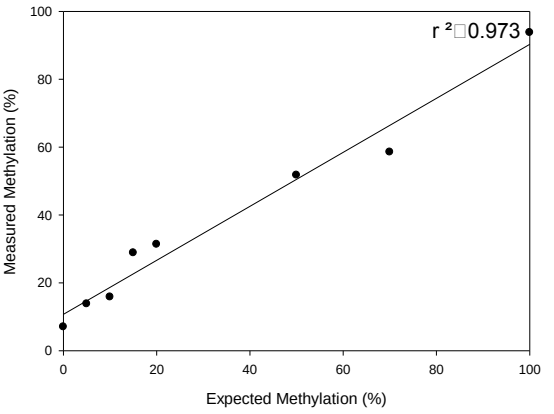
EPHA10 Assay Validation: Expected and Measured Methylation



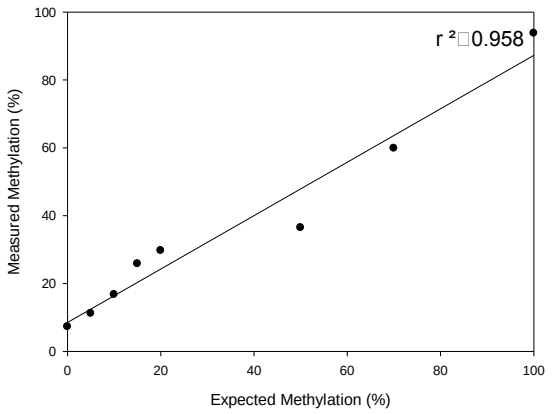
ESR1 Assay Validation: Expected and Measured Methylation



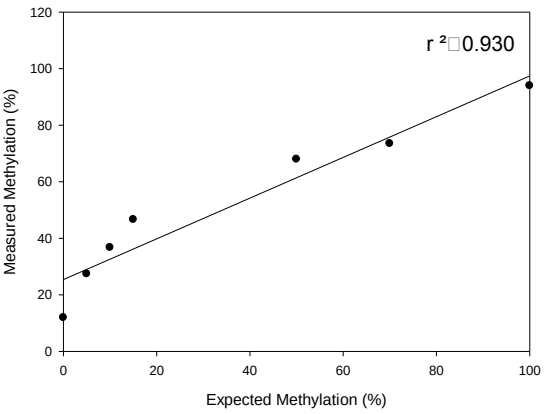
HAND2 Assay Validation: Expected and Measured Methylation



HLXB9 Assay Validation: Expected and Measured Methylation



HOXD4 Assay Validation: Expected and Measured Methylation



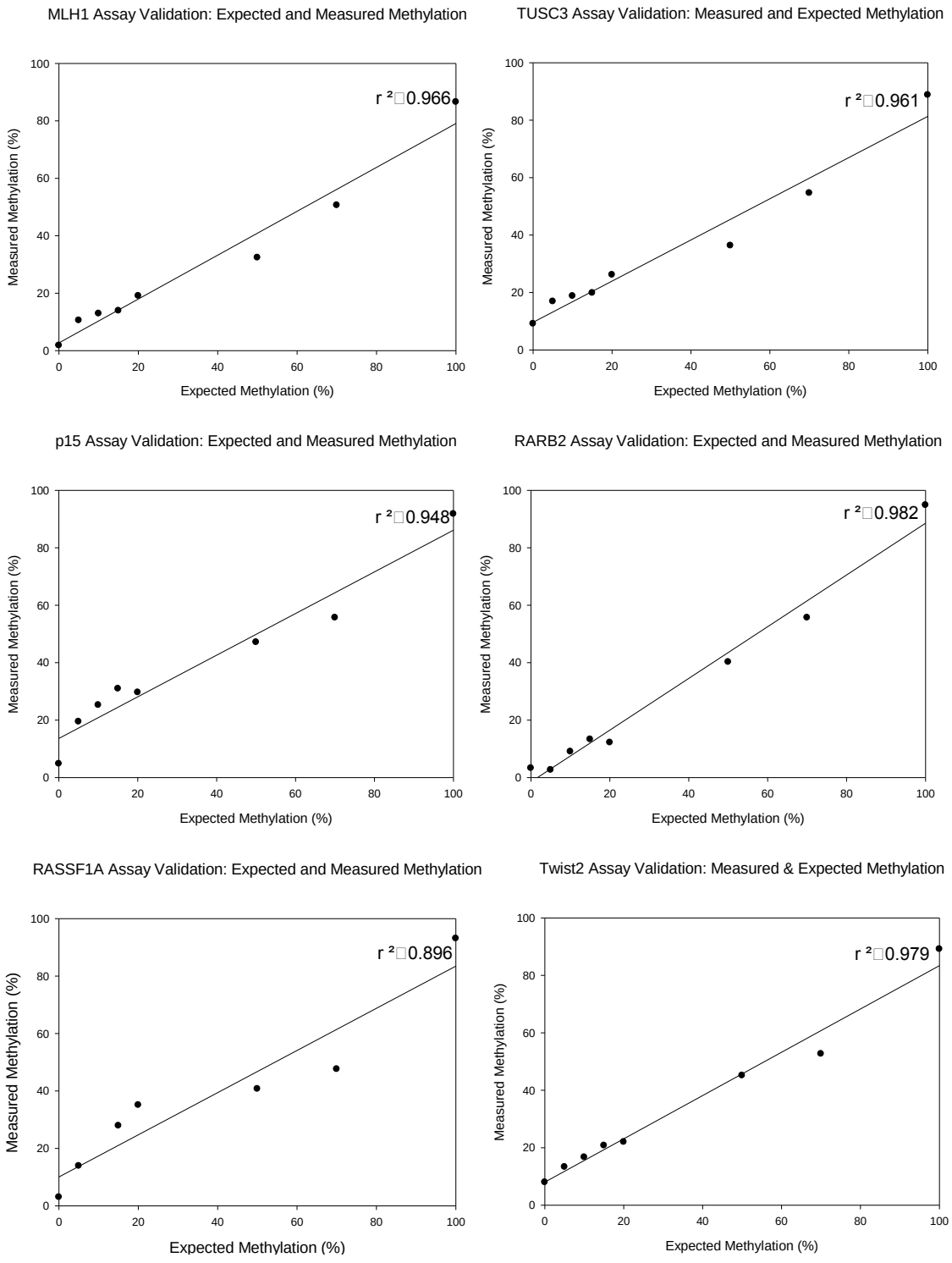


Figure 3.1: Validation of Pyrosequencing assays using for twelve loci with low levels of methylation in normal PBL DNA.

Expected DNA methylation controls were created by mixing normal PBL DNA with *in vitro* methylated DNA prior to PCR amplification. Linear regression analysis was performed by plotting the expected methylation of each sample against the methylation levels measured by the PyroMark MD instrument, with all assays demonstrating a robust positive relationship between these measurements (respective r^2 values shown on each graph), confirming that a range of methylation levels from 0-100% can be reliably

quantified. All samples were tested in duplicate and an average of 3-6 CpG sites calculated.

Due to differences in the intrinsic methylation status of certain genes which are not low/unmethylated in normal peripheral blood (including *HOXA4*, *HOXA5* and the imprinted gene *IGF2*), the validation curves for these assays were performed in a similar fashion, but with known standards primarily covering the range from 50-100%.

Normal *IGF2* expression is controlled by imprinting, specifically the maternal allele is densely methylated whilst the paternal allele remains free of methylation (Giannoukakis *et al.*, 1993; Issa *et al.*, 1996a). The peripheral blood leukocyte (PBL) sample used to create standards of known methylation represents the *in vivo* methylation status, which is around 50% in the case of *IGF2*. For validation of this gene the controls used were 50%, 65%, 70%, 75%, 80%, 85% and 100% methylation and the data in Figure 3.2A demonstrate successful validation of this assay.

We had expected that the promoter of *HOXA4* would be unmethylated (Strathdee *et al.*, 2004b), but when carrying out validation of this assay using 0%, 5%, 10%, 15%, 20%, 50% 70% and 100% controls, the measured methylation in normal PBL samples isolated from young volunteers was much higher than anticipated, at around 50%. To test whether this was due to an artefact of the Pyrosequencing measurement, cloned *HOXA4* mRNA sequence which had been inserted into a vector was sodium bisulfite modified. This represents unmethylated *HOXA4* sequence as the procedure of cloning requires a PCR step during which methylation marks are lost prior to sodium bisulfite modification. Subsequent Pyrosequencing analysis of this sequence revealed an average methylation of 3.5%, suggesting that the assay is able to accurately identify sequences which are unmethylated or exhibit low levels of DNA methylation, so it is likely that the high methylation observed in the PBL samples from the young volunteers is a true indication of high methylation levels *in vivo*. Due to the intrinsic methylation present in normal PBL samples, the methylation standards used represent expected methylation levels were 50%, 52.5%, 55%, 57.5%, 60%, 75%, 85% and 100% at the *HOXA4* locus and the measured methylation levels fit with this pattern (Figure 3.2B). In agreement with my finding, a recent publication reported approximately 60% methylation in *HOXA4* in DNA from normal white blood cells (Avraham *et al.*, 2010).

The promoter region of the *HOXA5* gene has around 50% in normal haematopoietic cell types (Strathdee *et al.*, 2007b). Therefore the known methylation standards used for validation of this gene were 50%, 65%, 70%, 75%, 80%, 85% and 100% (Figure 3.2C).

After Pyrosequencing validation, a panel of 15 genes in total remained for assessment of methylation levels in DNA from participants in the Newcastle 85+ study.

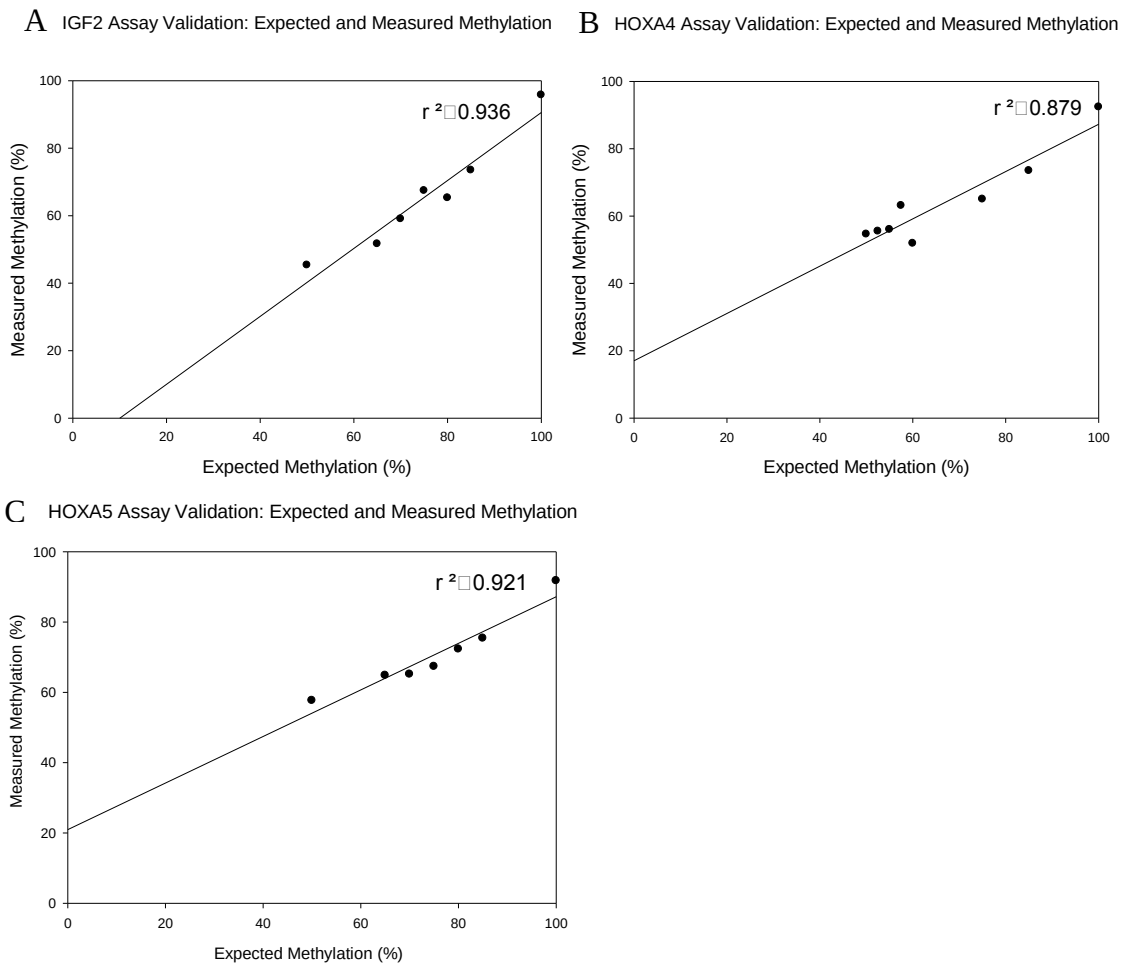


Figure 3.2: Validation of Pyrosequencing assays using for three loci which are partially methylated in normal PBL DNA.

The *IGF2* DMR (A) and the *HOXA4* (B) and *HOXA5* (C) promoters are typically around 50% methylated in normal peripheral blood. Expected methylation controls were created by mixing normal PBL DNA with in vitro methylated DNA prior to PCR amplification. Linear regression analysis was performed by plotting the expected methylation of each sample against the methylation levels measured by the PyroMark MD instrument, with these assays demonstrating a strong positive relationship between these measurements (respective r^2 values shown on each graph). All samples were tested in duplicate and an average of 3-6 CpG sites calculated.

3.5 Pyrosequencing results reveal the presence of genes with inter-individual variation

Having established the reliability of the Pyrosequencing assays for accurate quantification of DNA methylation levels, DNA samples from a subset of 50 Newcastle 85+ study participants were selected at random to perform a pilot study, examining DNA methylation status at the panel of 15 gene promoters in an 85 year old population.

For each sample, PCR was performed in duplicate before being subjected to Pyrosequencing analysis. Reproducibility of the Pyrosequencing measurement itself was tested by repeating the Pyrosequencing measurement on exactly the same sample (i.e. same PCR amplified product) which revealed highly consistent results for all assays, typically less than 1% difference between the two measures. As an example, the results of the *TWIST2* gene tested in the Newcastle 85+ population demonstrate very small differences; an average of 0.29 ± 0.29% [mean ± SD], difference in DNA methylation measurements between two separate Pyrosequencing runs. Likewise, two separate PCR reactions using the same sample template also exhibit high levels of reproducibility, typically less than 3% difference between the two PCR reactions with an average of 1.46 ± 1.1% difference in *TWIST2* DNA methylation levels between repeats (Table 3.1). More detailed data can be found in the appendix (A2.1 and A2.2).

	Duplicate analysis of the same PCR product	Analysis of duplicate PCR products
Number of samples	23	25
Range of methylation difference (%)	0-1.05	0.18-4.05
Mean difference (%)	0.29	1.46
Standard deviation (%)	0.29	1.10

Table 3.1: Reproducibility of DNA methylation measurements by Pyrosequencing after duplicate analysis of the same PCR product and duplicate PCR products for the *TWIST2* gene

Any sample which had >5% methylation difference between PCR repeats was repeated a third time and the two repeats which were within 5% of each other used to calculate the sample average. If methylation difference of >5% persisted after three repeats, the sample was discarded from the analysis due to low reproducibility.

Table 3.2 describes the DNA methylation levels measured at all fifteen gene promoters for all 50 participants. The two most variably methylated genes were *HOXA5* (measurements from 28-92%) and *HOXA4* (measurements from 13-64%), followed by the imprinted gene *IGF2* (measurements from 30-50%) which showed a much wider range of methylation measurements than any of the other loci (Figure 3.3). The extreme range of methylation observed for the two HOXA genes may be related to the highly variable methylation we have observed across the region of the HOXA cluster. There are multiple regions of 0, 50% or 100% methylation, which presumably is required for normal control of this gene cluster which may become less tightly controlled at an elderly age. We tested a small number of PBL DNA samples from young adults and found methylation levels at the *HOXA5* proximal promoter are close to 50% and show far lesser inter individual variation (52.5 ± 6.1% [mean ± SD]), which is clearly different from the measurements we obtained for the 85+ study participants (Table 3.3). The *HOXA4* promoter appears similarly more tightly controlled in the younger volunteers than the 85+ population, with the exception of one sample for which the DNA methylation levels were unusually low, which has a dramatic effect on the mean and standard deviation measurements given the low number of samples (47.7 ± 11.5% - Table 3.4). In addition, at the DMR of *IGF2,* although DNA methylation levels varied from 30-50% in this population of 85 year olds, we also observed similar variation at this *IGF2* locus when we tested a small number of PBL DNA samples from younger individuals (45 ± 4.7%) which suggests that the variation we observed in the 85 year olds participants is not unusual or age associated (Table 3.5). These three genes all represent loci with partially methylated promoters. The fact that methylation is already pre-existing may mean that increased variability is observed due to the fact that methylation can change in two directions, either increase or decrease; as opposed to genes with an unmethylated promoter region which can only show a methylation change in one direction (increase).

Gene	Methylation range (%)	Mean Methylation (%)	Standard Deviation (%)
HOXA5	28-92	61.5	15.1
HOXA4	13-64	41.5	11.8
HOXD4	7-35	14.8	6.0
IGF2	30-50	40.2	4.4
EPHA10	7-31	12.0	3.9
TUSC3	4-22	8.9	3.4
TWIST2	4-20	8.4	3.3
HAND2	2-18	7.3	2.8
HLXB9	1-15	3.3	2.1
APOE	6-13	8.4	1.9
p15	2-12	4.9	1.9
ESR1	4-11	5.5	1.6
RARB2	1-5	2.5	1.0
RASSF1A	1-4	2.5	0.7
MLH1	1-2	1.3	0.3

Table 3.2: Properties of fifteen genes analysed by Pyrosequencing in 50 PBL samples from the Newcastle 85+ study participants

For each gene, the minimum and maximum DNA methylation levels of all the samples is shown, alongside average DNA methylation calculated from all CpG sites in all samples, and the standard deviation as a measure of variation.

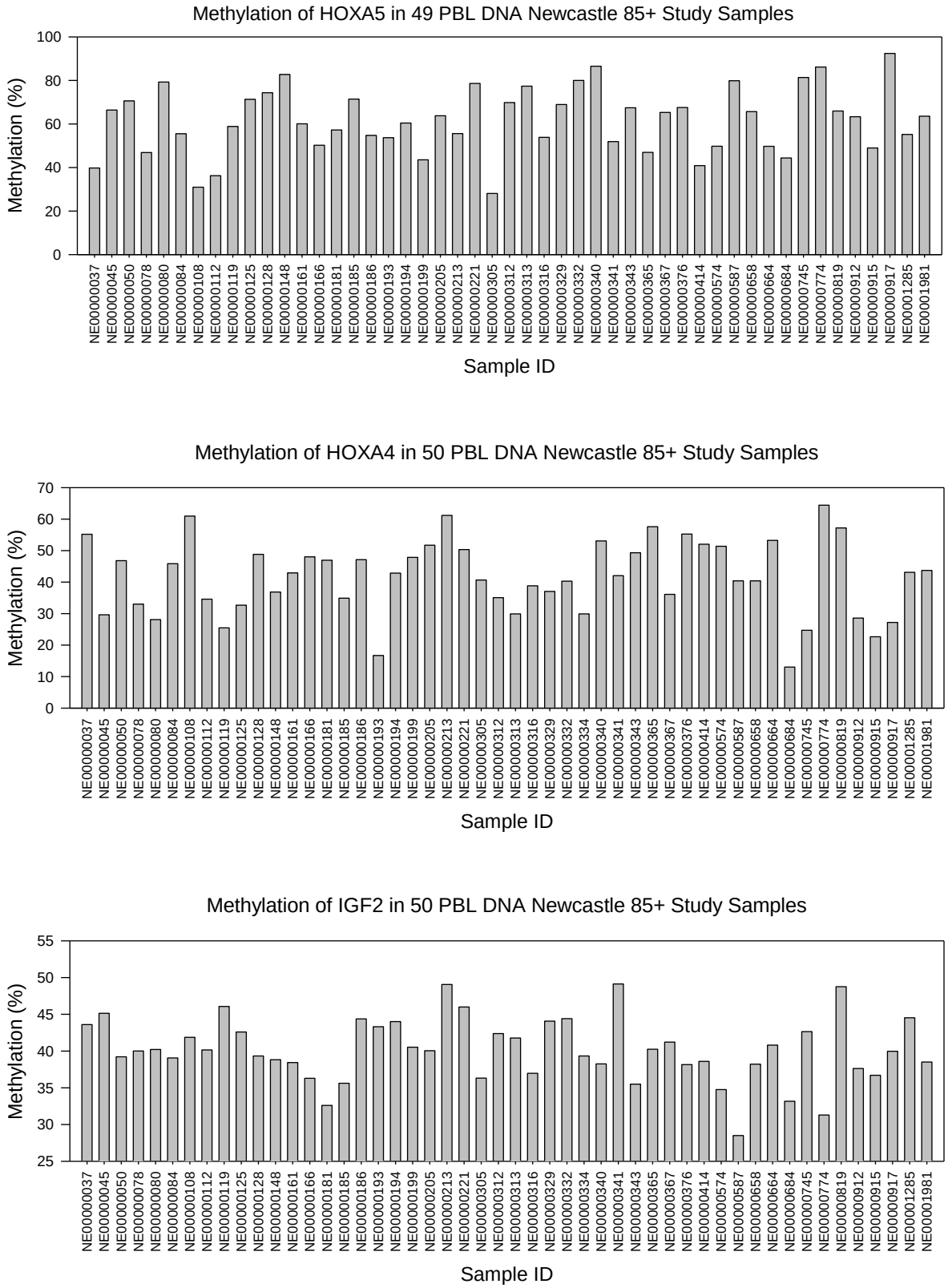


Figure 3.3: DNA methylation levels of *HOXA5, HOXA4* and *IGF2* in PBL samples from the Newcastle 85+ population

HOXA5, *HOXA4* and *IGF2* are the three most variably methylated genes measured by Pyrosequencing.

Sample	Mean Methylation (%)
PBL 1	50.8
PBL 2	52.3
PBL 3	49.6
PBL 4	47.3
PBL 5	48.5
PBL 6	55.8
PBL 7	66.1
PBL 8	50.0
Range	47-66
Mean	52.5
SD	6.1

Table 3.3: DNA methylation levels of *HOXA5* in PBL samples from eight young volunteers measured by Pyrosequencing

Sample	Mean Methylation (%)
PBL 1	43.4
PBL 3	53.6
PBL 4	59.4
PBL 5	53.6
PBL 6	24.1
PBL 7	49.8
PBL 8	50.2
Range	24-59
Mean	47.7
SD	11.5

Table 3.4: DNA methylation levels of *HOXA4* in PBL samples from seven young volunteers measured by Pyrosequencing

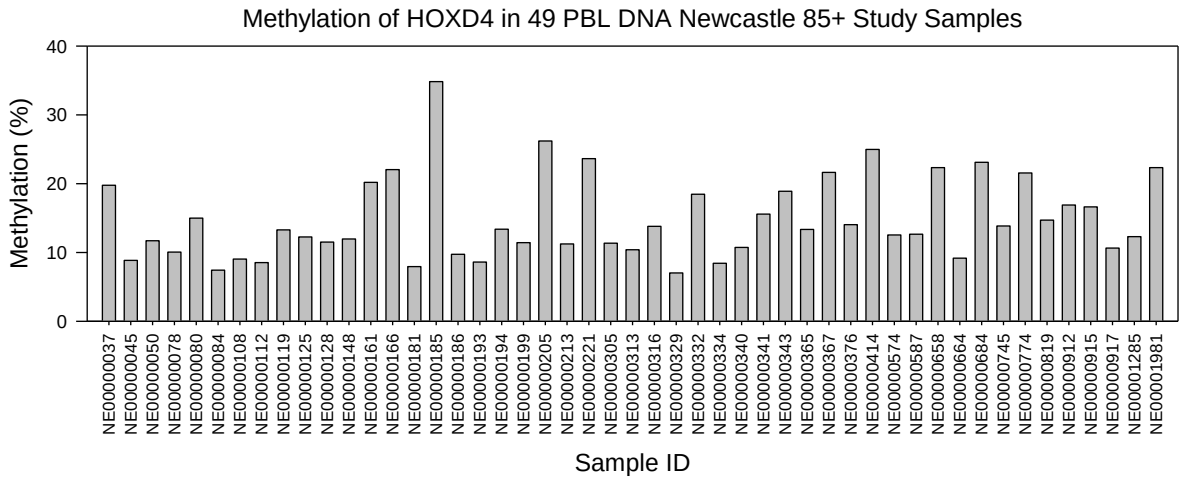
Sample	Mean Methylation (%)
PBL 2	46.2
PBL 3	45.3
PBL 4	36.1
PBL 5	49.4
PBL 6	42.8
PBL 7	50.0
PBL 8	45.3
Range	36-50
Mean	45.0
SD	4.7

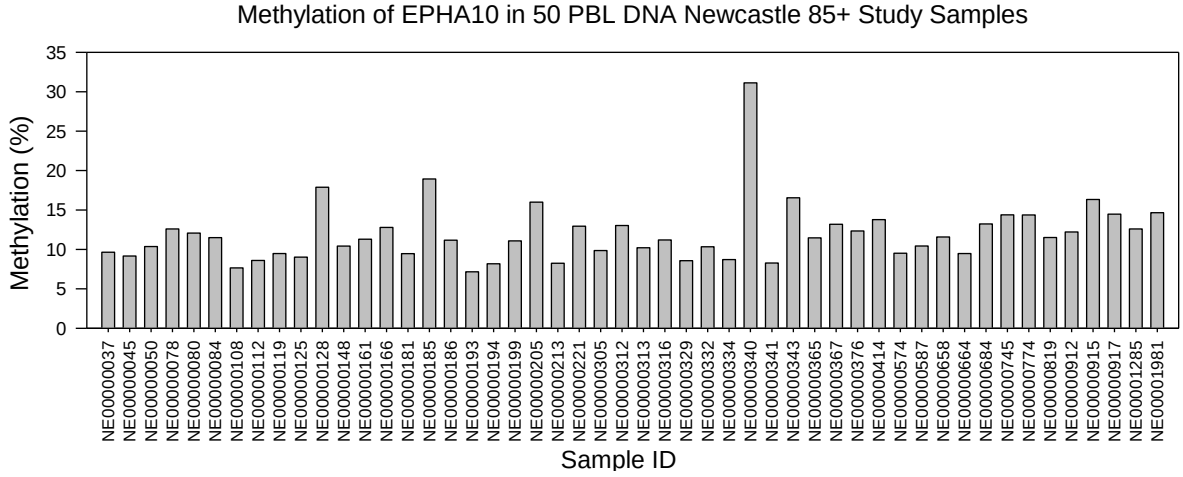
Table 3.5: DNA methylation levels of *IGF2* in PBL samples measured from seven young volunteers by Pyrosequencing

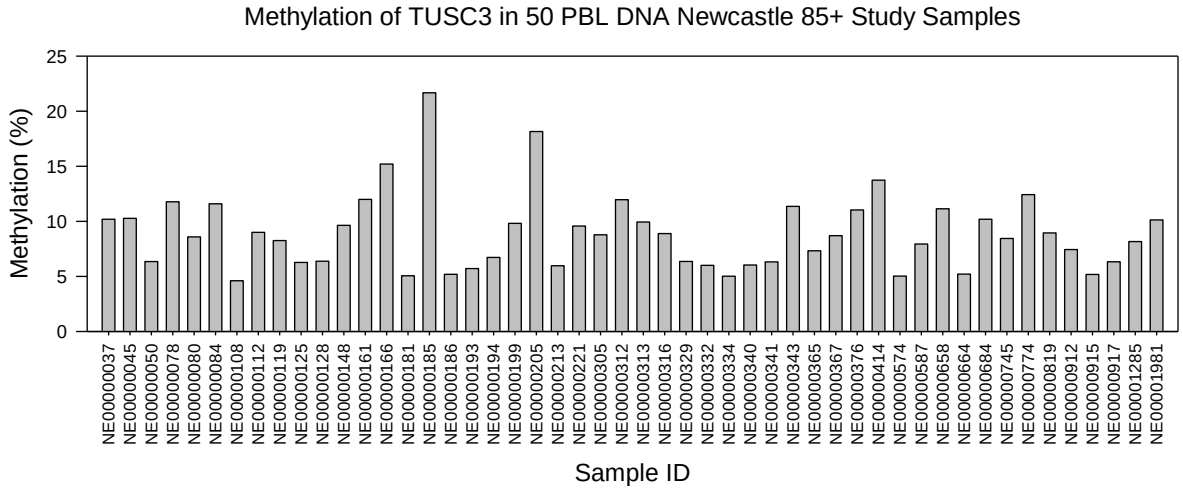
Of the remaining twelve genes with a theoretically unmethylated promoter region, five were noted to have both increased DNA methylation with a considerable amount of inter-individual variation (Figure 3.4). These genes were *HOXD4* (14.8 ± 6.0%), *EPHA10* (12.0 ± 3.9%), *TUSC3* (8.9 ± 3.4%), *TWIST2* (8.4 ± 3.3%) and *HAND2* (7.3 ± 2.8%).

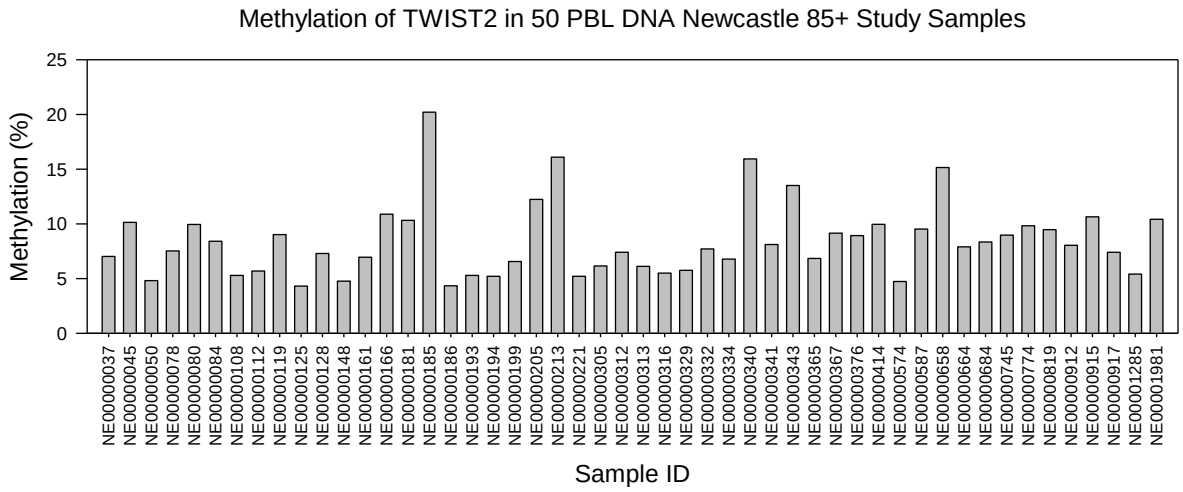
Other genes including *HLXB9* (3.3 ± 2.1%), *APOE* (8.4 ± 1.9%), *p15* (4.9 ± 1.9%) and *ESR1* (5.5 ± 1.6%) showed some evidence towards variability although these genes tended to show less inter-individual variation than the genes described above, and overall methylation levels were typically lower at these loci (Figure 3.5).

Finally, the remaining three genes: *RARB2* (2.5 ± 1%), *RASSF1A* (2.5 ± 0.67%) and *MLH1* (1.3 ± 0.3%), exhibited either very little methylation or low variation in any of the pilot study participants (Figure 3.6). The amount of methylation observed here is likely to be below the limit of detection for the Pyrosequencing assay so we could not be confident that any observation of inter-individual variation at these loci is accurate.


Methylation of HOXD4 in 49 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
10
20
30
40
NE00000037
NE00000045
NE00000050
NE00000078
NE00000080
NE00000084
NE00000108
NE00000112
NE00000119
NE00000125
NE00000128
NE00000148
NE00000161
NE00000166
NE00000181
NE00000185
NE00000186
NE00000193
NE00000194
NE00000199
NE00000205
NE00000213
NE00000221
NE00000305
NE00000313
NE00000316
NE00000329
NE00000332
NE00000334
NE00000340
NE00000341
NE00000343
NE00000365
NE00000367
NE00000376
NE00000414
NE00000574
NE00000587
NE00000658
NE00000664
NE00000684
NE00000745
NE00000774
NE00000819
NE00000912
NE00000915
NE00000917
NE00001285
NE00001981
Sample ID


Methylation of EPHA10 in 50 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
30
35
NE00000037
NE00000045
NE00000050
NE00000078
NE00000080
NE00000084
NE00000108
NE00000112
NE00000119
NE00000125
NE00000128
NE00000148
NE00000161
NE00000166
NE00000181
NE00000185
NE00000186
NE00000193
NE00000194
NE00000199
NE00000205
NE00000213
NE00000221
NE00000305
NE00000312
NE00000313
NE00000316
NE00000329
NE00000332
NE00000334
NE00000340
NE00000341
NE00000343
NE00000365
NE00000367
NE00000376
NE00000414
NE00000574
NE00000587
NE00000658
NE00000664
NE00000684
NE00000745
NE00000774
NE00000819
NE00000912
NE00000915
NE00000917
NE00001285
NE00001981
Sample ID


Methylation of TUSC3 in 50 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
NE00000037
NE00000045
NE00000050
NE00000078
NE00000080
NE00000084
NE00000108
NE00000112
NE00000119
NE00000125
NE00000128
NE00000148
NE00000161
NE00000166
NE00000181
NE00000185
NE00000186
NE00000193
NE00000194
NE00000199
NE00000205
NE00000213
NE00000221
NE00000305
NE00000312
NE00000313
NE00000316
NE00000329
NE00000332
NE00000334
NE00000340
NE00000341
NE00000343
NE00000365
NE00000367
NE00000376
NE00000414
NE00000574
NE00000587
NE00000658
NE00000664
NE00000684
NE00000745
NE00000774
NE00000819
NE00000912
NE00000915
NE00000917
NE00001285
NE00001981
Sample ID



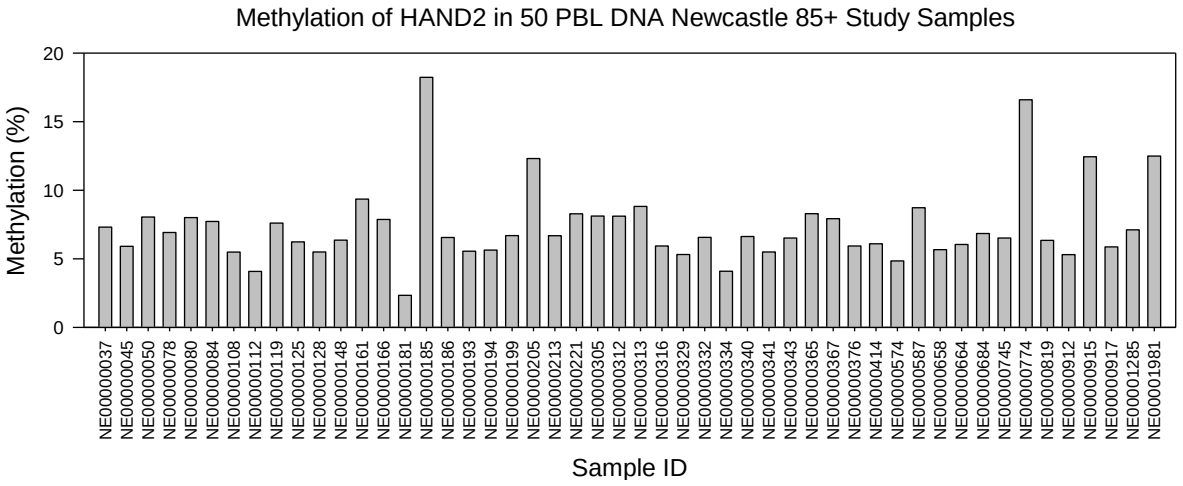
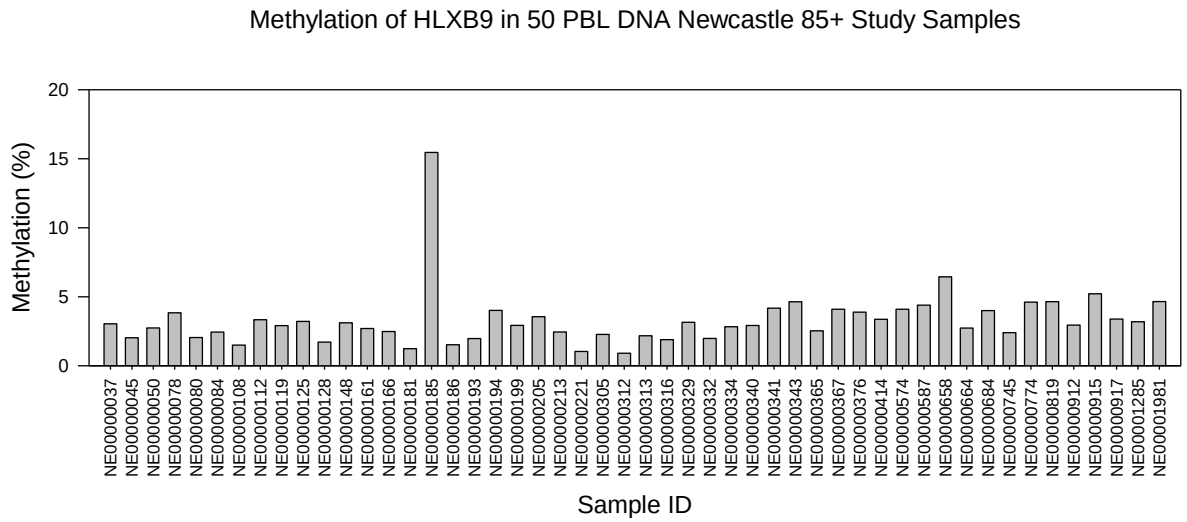
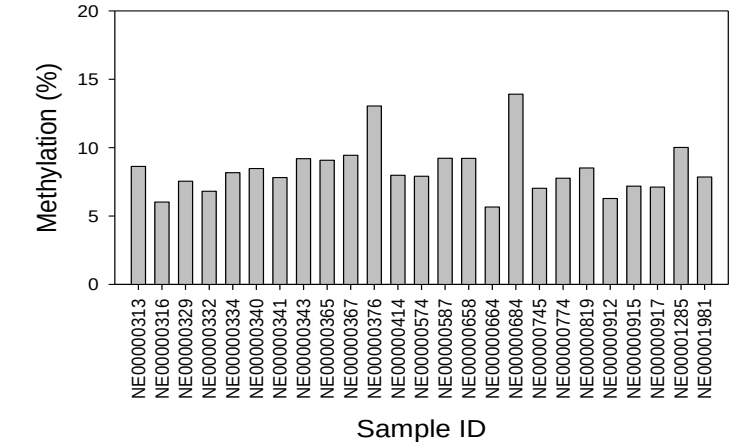
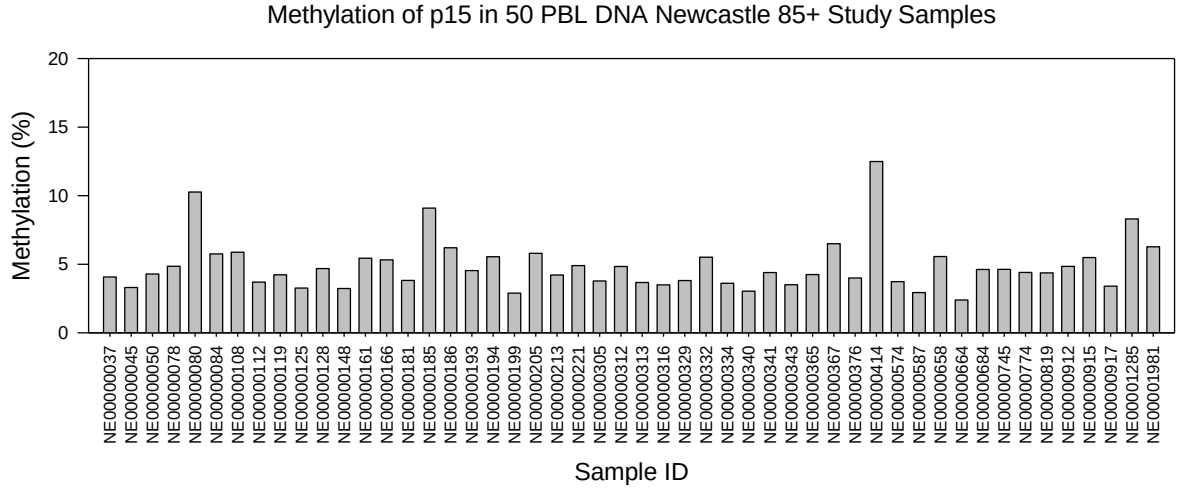


Figure 3.4: DNA methylation levels of *HOXD4, EPHA10, TUSC3, TWIST2* and *HAND*2 genes in PBL samples from the Newcastle 85+ study participants

These five genes display significant inter-individual variation in DNA methylation levels measured by Pyrosequencing. Lack of reproducible methylation measurements meant one sample was excluded from *HOXD4* result.


Methylation of HLXB9 in 50 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
NE00000037
NE00000045
NE00000050
NE00000078
NE00000080
NE00000084
NE00000108
NE00000112
NE00000119
NE00000125
NE00000128
NE00000148
NE00000161
NE00000166
NE00000181
NE00000185
NE00000186
NE00000193
NE00000194
NE00000199
NE00000205
NE00000213
NE00000221
NE00000305
NE00000312
NE00000313
NE00000316
NE00000329
NE00000332
NE00000334
NE00000340
NE00000341
NE00000343
NE00000365
NE00000367
NE00000376
NE00000414
NE00000574
NE00000587
NE00000658
NE00000664
NE00000684
NE00000745
NE00000774
NE00000819
NE00000912
NE00000915
NE00000917
NE00001285
NE00001981
Sample ID


Methylation of APOE in 25 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
NE00000313
NE00000316
NE00000329
NE00000332
NE00000334
NE00000340
NE00000341
NE00000343
NE00000365
NE00000367
NE00000376
NE00000414
NE00000574
NE00000587
NE00000658
NE00000664
NE00000684
NE00000745
NE00000774
NE00000819
NE00000912
NE00000915
NE00000917
NE00001285
NE00001981
Sample ID


Methylation of p15 in 50 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
NE00000037
NE00000045
NE00000050
NE00000078
NE00000080
NE00000084
NE00000108
NE00000112
NE00000119
NE00000125
NE00000128
NE00000148
NE00000161
NE00000166
NE00000181
NE00000185
NE00000186
NE00000193
NE00000194
NE00000199
NE00000205
NE00000213
NE00000221
NE00000305
NE00000312
NE00000313
NE00000316
NE00000329
NE00000332
NE00000334
NE00000340
NE00000341
NE00000343
NE00000365
NE00000367
NE00000376
NE00000414
NE00000574
NE00000587
NE00000658
NE00000664
NE00000684
NE00000745
NE00000774
NE00000819
NE00000912
NE00000915
NE00000917
NE00001285
NE00001981
Sample ID

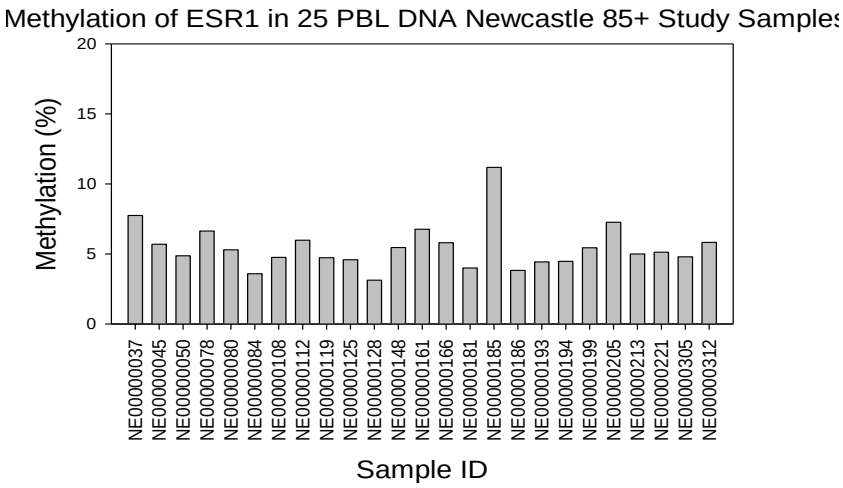


Figure 3.5: DNA methylation levels of *HLXB9, APOE, p15* and *ESR1* genes in PBL samples from the Newcastle 85+ study participants.

These four genes display limited inter-individual variability of methylation levels measured by Pyrosequencing. Pyrosequencing of the *APOE* and *ESR1* promoter was performed for only 25 Newcastle 85+ study participants.

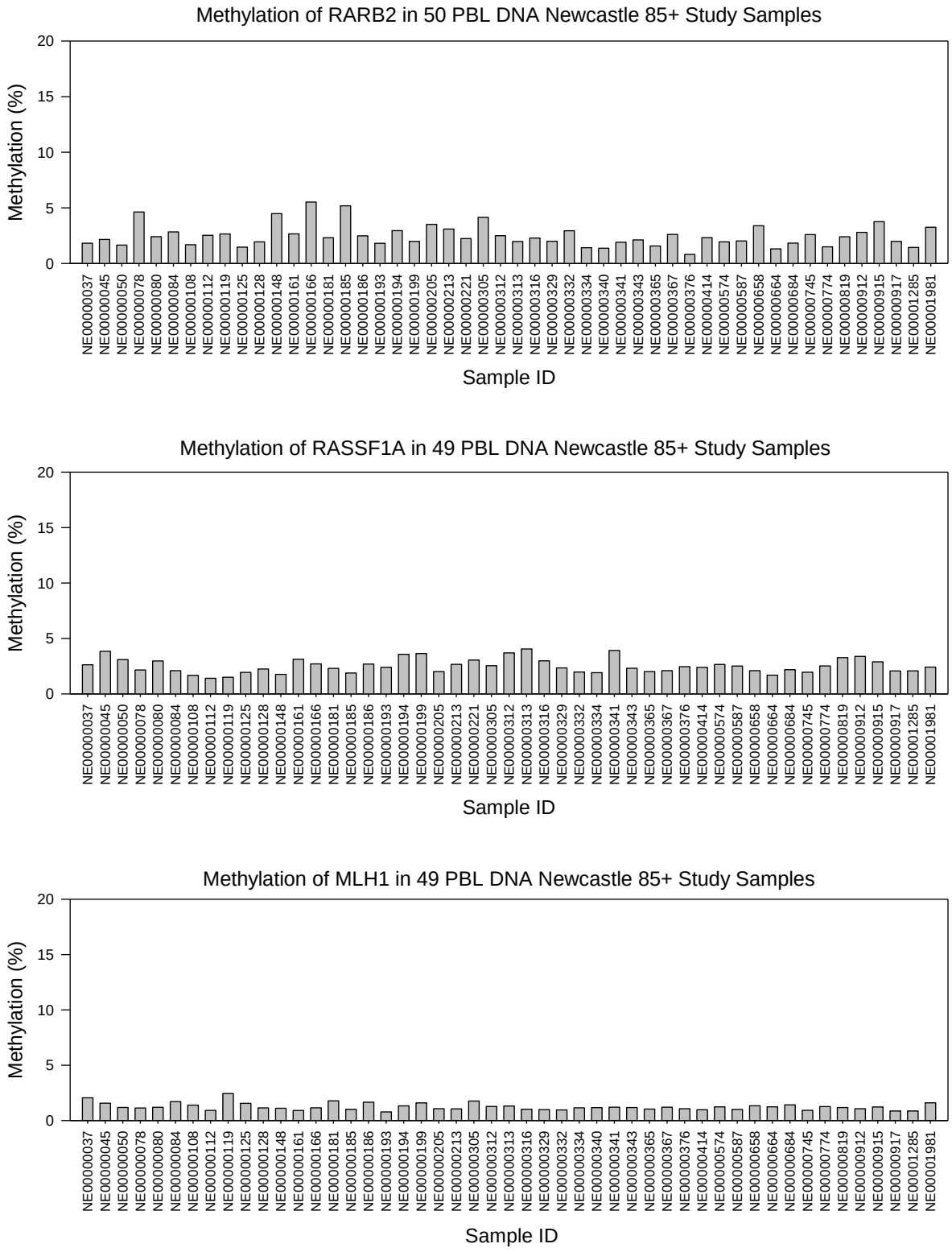


Figure 3.6: DNA methylation levels of *RARB2*, *RASSF1A* and *MLH1* genes in PBL samples from the Newcastle 85+ Study participants

These three genes exhibit both low levels of methylation and a lack of inter-individual variation, measured by Pyrosequencing. Lack of reproducible methylation measurements meant one sample was excluded from *RASSF1A* and *MLH1* result.

The results of the pilot study indicate that aberrant DNA methylation was fairly common at many of the loci examined and that substantial amounts of methylation were detected in some individuals at certain gene promoters which are expected to be free of methylation. In addition to this, we found DNA methylation was also highly variable between individuals of the same elderly age, within this population.

These observations imply that the propensity to aberrantly hypermethylate CpG islands, at least at certain loci, varies between individuals of the same age within the study group; and therefore may underlie, at least partially, the vast differences between individuals in their susceptibility to age related disease.

3.6 Features of variably methylated genes

3.6.1 DNA methylation levels between pairs of genes are strongly correlated

We performed pairwise correlations of DNA methylation levels for each of the twelve genes with evidence of variation for all the fifty individuals in the Newcastle 85+ study, using the Pearson correlation coefficient test. The genes were assessed in two groups; one composed of nine genes (*TWIST2*, *HOXD4*, *EPHA10*, *TUSC3*, *HAND2*, *p15*, *HLXB9*, *ESR1* and *APOE*) for which the promoter regions typically have low levels of DNA methylation. The other group contained those genes which are regulated by DNA methylation (*IGF2*, *HOXA5* and *HOXA4*); therefore the promoter regions of these genes are partially methylated in normal haematopoietic cells.

For the first group of nine genes, DNA methylation levels were strongly positively correlated within individuals with high significance at the majority of genes (Table 3.6), meaning that individuals with high levels of DNA methylation at one loci, are likely to have high levels of methylation at any other loci. This suggests that a subset of the Newcastle 85+ study participants exhibited high DNA methylation levels at most/all of the genes assessed. Consequently, when methylation was averaged across multiple genes, the overall levels of methylation also varied widely in this population, since individuals were likely to have high or low methylation at several genes simultaneously. The exceptions to this were *APOE* and *p15*. *APOE* was not significantly correlated with any other loci after correction for multiple comparisons. This may be due to the fact that all the other genes have evidence of hypermethylation in leukaemia, whereas *APOE*

does not, suggesting the acquisition of methylation for this gene in haematopoietic cells may be different to the other genes. Similarly, DNA methylation levels of *p15* were not significantly correlated with *TWIST2, EPHA10, HLXB9* or *ESR1* and only displayed a weakly significant association with *TUSC3* or *HAND2* methylation levels, suggesting that this gene does not show the same pattern of methylation acquisition as the other seven genes in this group. This may be partly due to the observation that *p15* methylation is more frequently found to be methylated in myeloid forms of leukaemia, whereas the other genes are associated with lymphoid leukaemia, demonstrating the differential DNA methylation status of specific genes in distinct haematopoietic cell types.

The three genes with partially methylated promoters, *HOXA4, HOXA5* and *IGF2*, which demonstrated significant inter-individual variation in DNA methylation, did not show any significant correlation either with each other, or any of the other variably methylated genes which demonstrated pairwise correlations with each other (Table 3.7).

These results show that not only are DNA methylation levels highly variable in the Newcastle 85+ population, also that individuals exhibiting high DNA methylation levels at one loci have significantly more chance of exhibiting high levels at other loci, implying that the propensity to aberrantly methylate CpG islands varies within the study group. This pattern is reminiscent of the "CpG Island methylator phenotype" a term first used to describe the methylation of multiple loci simultaneously in a subset of colon cancer cells. Such tumours may be liable to hypermethylate several important tumour suppressor genes concurrently and contain numerous transcriptionally inactivated genes (Toyota *et al.*, 1999). The results suggest that a similar acquisition of methylation at multiple loci may also be a feature of normal haematopoietic cells in an elderly population.

	TWIST2	*HOXD4*	*EPHA10*	*TUSC3*	*HAND2*	*p15*	*HLXB9*	*ESR1*
TWIST2								
HOXD4	r = 0.51 p = 0.0002							
EPHA10	r = 0.55 p = 0.00004	r = 0.36 p = 0.01						
TUSC3	r = 0.49 p = 0.0003	r = 0.72 p = 5.0E-09	r = 0.31 p = 0.03					
HAND2	r = 0.43 p = 0.001	r = 0.64 p = 7.0E-07	r = 0.38 p = 0.006	r = 0.64 p = 0.000002				
p15	r = 0.24 p = 0.09	r = 0.51 p = 0.0001	r = 0.18 p = 0.27	r = 0.39 p = 0.005	r = 0.31 p = 0.03			
HLXB9	r = 0.60 p = 0.00003	r = 0.60 p = 0.00006	r = 0.30 p = 0.04	r = 0.54 p = 0.00006	r = 0.59 p = 0.000007	r = 0.28 p = 0.05		
ESR1	r = 0.58 p = 0.002	r = 0.77 p = 7.7E-06	r = 0.44 p = 0.026	r = 0.81 p = 9.9E-07	r = 0.77 p = 5.4E-06	r = 0.35 p = 0.08	r = 0.78 p = 3.7E-06	
APOE	r = 0.16 p = 0.22	r = 0.33 p = 0.06	r = 0.20 p = 0.17	r = 0.33 p = 0.05	r = -0.09 p = 0.33	r = 0.09 p = 0.33	r = 0.25 p = 0.12	

Table 3.6: Pair-wise correlations of PBL DNA methylation levels between nine variably methylated loci from the Newcastle 85+ study

DNA methylation data for 50 participants (25 individuals for *ESR1* and *APOE*) was correlated at each pair of loci using Pearson test of correlation. The Pearson correlation coefficient (r) and corresponding p values are shown. The majority of correlations are highly significant. Black p value = not significant (NS), Orange p value = significant at <0.05 level, Red p value = significant at <0.0014 level after performing Bonferroni correction for multiple comparisons.

	IGF2	*HOXA4*	*HOXA5*
IGF2			
HOXA4	r = 0.02 p = 0.87		
HOXA5	r = -0.04 p = 0.79	r = -0.12 p = 0.42	
TWIST2	r = -0.19 p = 0.18	r = 0.12 p = 0.40	r = 0.21 p = 0.15
HOXD4	r = -0.22 p = 0.13	r = 0.07 p = 0.63	r = 0.10 p = 0.49
EPHA10	r = -0.29 p = 0.03	r = 0.09 p = 0.51	r = 0.39 p = 0.005
TUSC3	r = -0.24 p = 0.09	r = 0.05 p = 0.68	r = -0.02 p = 0.87
HAND2	r = -0.27 p = 0.05	r = 0.07 p = 0.59	r = 0.22 p = 0.13
p15	r = 0.03 p = 0.81	r = -0.007 p = 0.96	r = -0.09 p = 0.53
HLXB9	r = -0.26 p = 0.06	r = -0.06 p = 0.64	r = 0.09 p = 0.52
ESR1	r = -0.20 p = 0.33	r = -0.06 p = 0.75	r = 0.02 p = 0.91

Table 3.7: Pair-wise correlations of *HOXA4*, *HOXA5* and *IGF2* PBL DNA methylation levels with other variably methylated genes in the Newcastle 85+ Study

DNA methylation data for fifty participants (25 individuals for *ESR1*) was correlated at each pair of loci using the Pearson test of correlation. The Pearson correlation coefficient (r) and corresponding p values are shown. The results show mostly weak or no correlation Black p value = NS, Orange p value = significant at <0.05 level, Red p value = significant at <0.0019 level after performing Bonferroni correction for multiple comparisons.

3.6.2 Genes susceptible to age-related and cancer related DNA methylation acquisition are overlapping

For the 8 loci, *TWIST2, HOXD4, EPHA10, TUSC3, HAND2, HOXA4, HOXA5* and *IGF2,* which exhibit considerable susceptibility to inter-individual variation in DNA methylation levels, there is also evidence that these same genes become hypermethylated in haematological malignancies.

For the four genes which showed some evidence of variability *APOE, p15, HLXB9* and *ESR1,* there is some previous evidence suggesting abnormal methylation patterns of *p15, HLXB9* and *ESR1* in haematological malignancies, but none for the *APOE* gene.

Of the remaining non variably methylated genes, *RARB2, MLH1* and *RASSF1A* although there is copious evidence for either their hypermethylation status in solid tumours, or the presence of an age associated increase in methylation in normal tissues, there is limited evidence that they are subjected to aberrant methylation in leukaemia. The age related DNA methylation increases in these genes which have been observed in other studies may be tissue specific and not present in peripheral blood leukocytes.

Therefore these results suggest that the targets for both age related methylation and aberrant DNA methylation in haematological malignancies may be overlapping. This finding has been supported by the outcomes from other recent studies (Rakyan *et al.*, 2010; Teschendorff *et al.*, 2010), although the increased DNA methylation levels observed in this elderly population are less dramatic than those occurring in cancer cells, which typically have very high levels of methylation (>50%).

3.6.3 Bisulfite sequencing reveals the presence of densely methylated alleles

The results presented so far describe variable DNA methylation levels in an elderly population; however it is unclear from the experimental technique used, what the pattern of methylation is like at the allelic level. This is important with regards to how DNA methylation affects gene transcription and expression since DNA methylation functions at an allelic level. Pyrosequencing measures the average amount of methylation at a particular CpG site by detecting whether a methylated cytosine is present or not at a specific CpG site in one of the vast number of alleles and cells

represented within a particular DNA sample. Using this technique, it is not possible to determine whether consecutive CpG sites within an individual allele have the same methylation status (methylated or unmethylated across consecutive sites); or whether methylation is essentially random at any CpG site in any particular allele.

This is illustrated by the simplified scenario presented in Figure 3.7. This example demonstrates two different samples in which 5 consecutive CpG sites are represented across ten individual alleles. Using Pyrosequencing analysis, both samples will measure 30% methylation at each of the five CpG sites, but the distribution of methylation across the alleles is very different. In sample 1, 30% of the alleles have complete methylation of consecutive CpG sites, whilst the remaining 70% are free of methylation. Sample 2 contains only partially methylated alleles, with sporadically methylated CpG sites. These two scenarios would likely have a very different biological consequence. An allele which is completely methylated will almost certainly result in transcriptional silencing, such as in the case of imprinted genes (Li *et al.*, 1993), X inactivation (Kelkar and Deobagkar, 2010), cell type specific genes (De Smet *et al.*, 1999) and tumour suppressor genes in cancer (Baylin, 2005). However, the effect of partial methylation on gene expression is not clear (Costello *et al.*, 1996; Qin *et al.*, 2004). A study examining the *p15* promoter in leukaemia samples found that gene inactivation could be achieved by the presence of alleles with a methylation density of 30-40%, although samples comprised of alleles of lower methylation density than this still expressed *p15*, suggesting the existence of a methylation density threshold at which a gene is no longer expressed (Cameron *et al.*, 1999). Therefore it is more likely that the scenario in sample 1, with its densely methylated alleles, will produce a negative effect on gene transcription than sample 2, although it is possible that partial methylation could affect expression levels of the gene (Nile *et al.*, 2008).

In order to determine DNA methylation patterns at the allelic level in PBL DNA samples from the Newcastle 85+ study population, we examined the promoter region of two variably methylated genes, *TWIST2* and *HOXD4*, by bisulfite sequencing of multiple alleles obtained from an individual with either high or low DNA methylation levels at these genes (determined by Pyrosequencing). For each sample and gene, PCR products were cloned into a pCR®2.1 vector and transfected into competent *E. coli* cells. Each clone isolated from a growing colony contains a single PCR amplified sequence, which is representative of a single allele from the whole sample. *TWIST2*

methylation was analysed in one individual, DNA sample NE00000185 (20% methylation – high). *HOXD4* methylation was analysed in three individuals, DNA samples NE00000185 (35% methylation – high), NE00000205 (26% methylation – high) and NE00000334 (8% methylation – low).

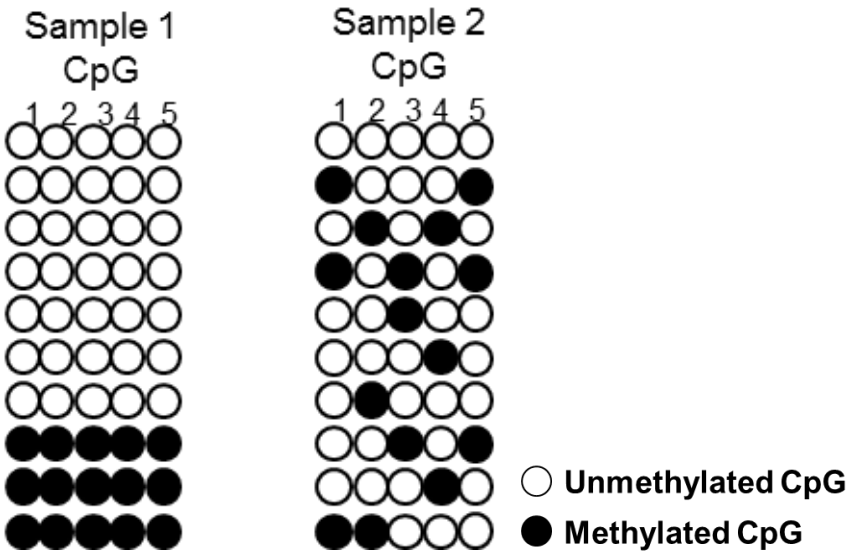


Figure 3.7: Allelic methylation patterns in two hypothetical samples
Each sample is represented by ten different alleles which comprise of various possibilities for methylation at consecutive CpG sites. Black dots represent methylated CpG sites and white dots unmethylated CpG sites. The average amount of methylation measured by Pyrosequencing of both sample 1 and 2 would be 30%. Sample 1 is comprised of 30% completely hypermethylated alleles, whereas for sample 2 only unmethylated or partially methylated alleles exist.

The sequencing results for the *TWIST2* locus suggest that the 20% methylation observed in participant NE00000185 by Pyrosequencing consists of mixture of alleles which are either completely unmethylated/have very low levels of DNA methylation, or have very dense methylation in the region (Figure 3.8).

The results for the *HOXD4* locus suggest that the participants with high levels of DNA methylation at this gene (NE00000185 & NE00000205) also contain a similar combination of highly methylated alleles alongside those which are unmethylated (Figure 3.9A and B). However the highly methylated alleles are not as densely methylated as those observed in the *TWIST2* region. There are fewer highly methylated alleles in participant NE00000205 than NE00000185, which was reflected by the lower

overall methylation DNA measurement of this individual by Pyrosequencing. In addition, no heavily methylated alleles were detectable in our analysis of participant NE00000334 and methylation was observed as a sporadic event at individual CpG sites (Figure 3.10).

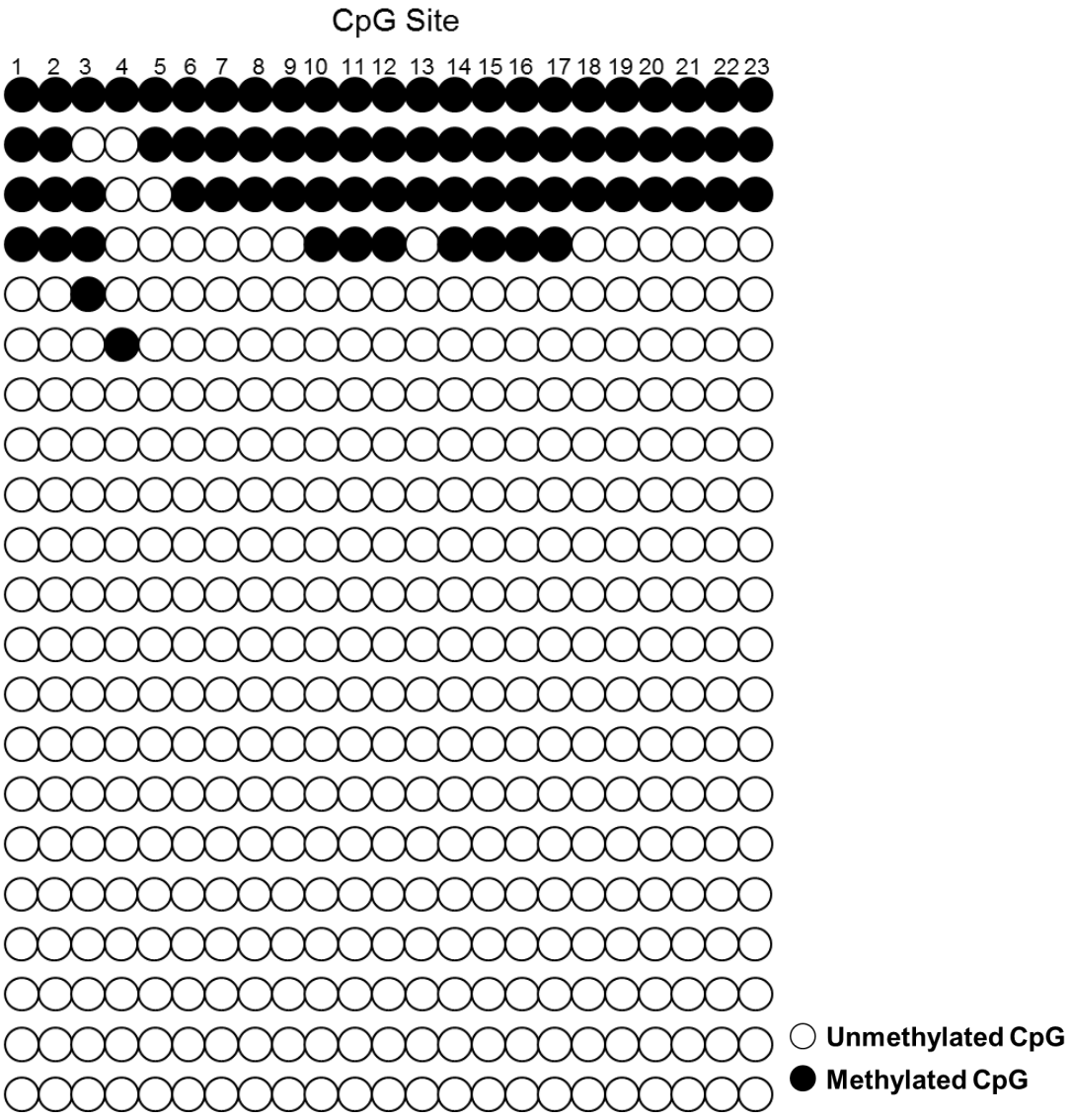


Figure 3.8: Bisulfite sequencing results at the *TWIST2* promoter of PBL DNA sample NE00000185 from the Newcastle 85+ study

Bisulfite sequencing results of 23 consecutive CpG sites within the *TWIST2* promoter represented by 21 alleles from sample NE00000185**.** Methylated CpG sites are represented by black circles an unmethylated CpG sites by a white circles. The sequencing results indicate a combination of either very densely methylated or completely unmethylated alleles

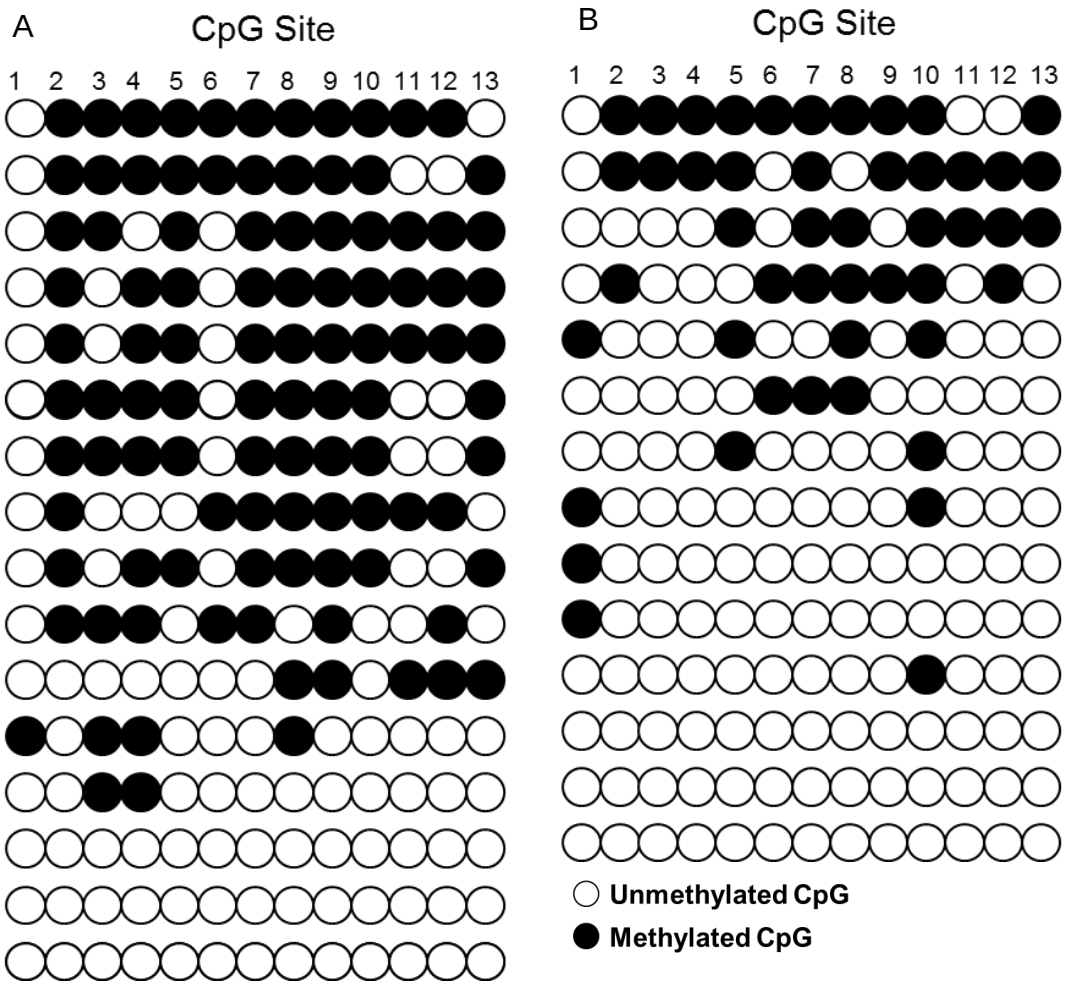


Figure 3.9: Bisulfite sequencing results at the *HOXD4* promoter of PBL DNA samples NE00000185 and NE00000205 from the Newcastle 85+ study

Bisulfite sequencing results of 13 consecutive CpG sites within the *HOXD4* promoter represented by (A) 16 alleles from DNA sample NE00000185. The sample was comprised of alleles with high levels of methylation (>50% CpGs methylated) and those with low levels of methylation. There are more densely methylated alleles in this sample than sample NE00000205; and (B) 14 alleles from DNA sample NE00000205**.** This sample again contained a mixture of alleles with either high or low levels of methylation but there are fewer highly methylated alleles than DNA sample NE00000185. Methylated CpG sites are represented by black circles and unmethylated CpG sites by white circles.

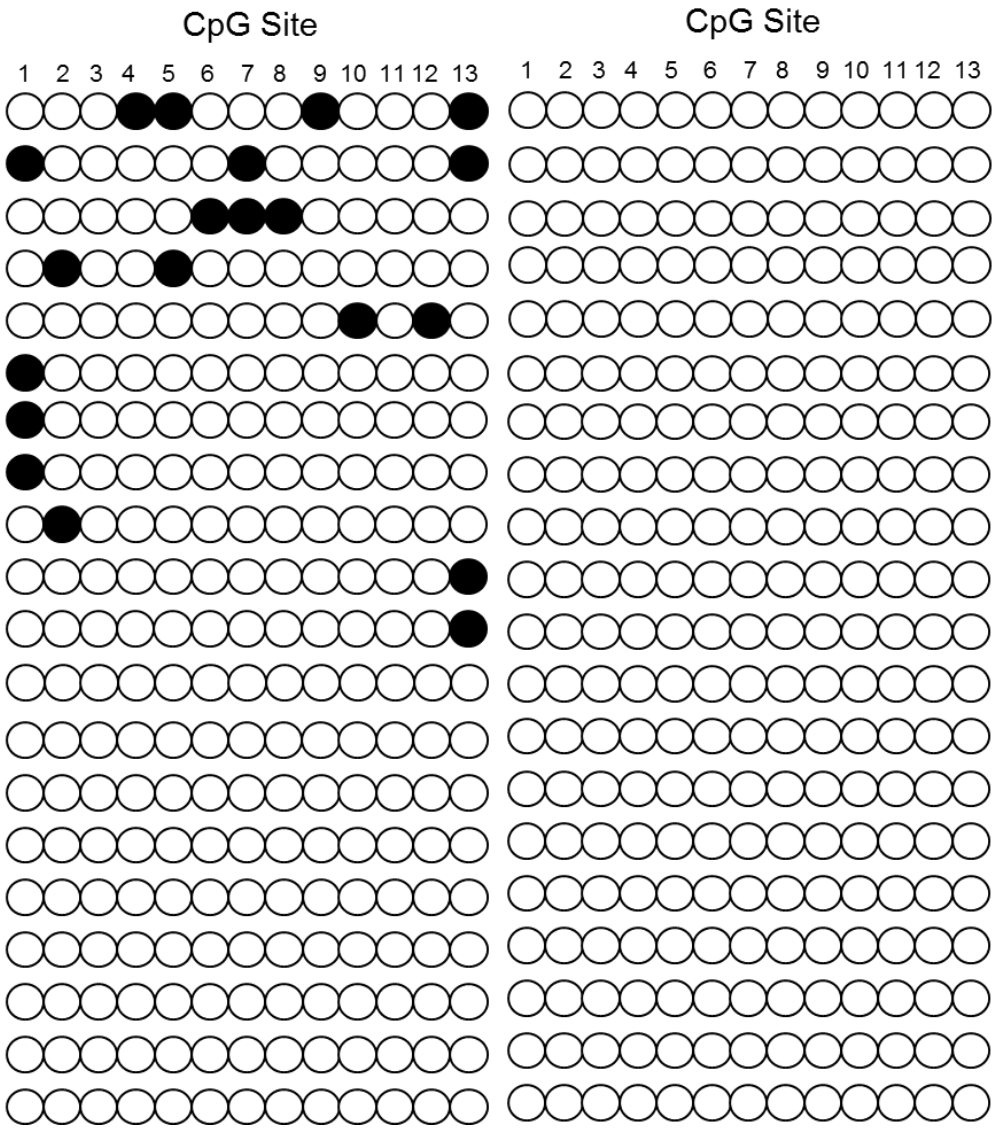


Figure 3.10: Bisulfite sequencing results at the *HOXD4* promoter of PBL DNA sample NE00000334 from the Newcastle 85+ study

Bisulfite sequencing results of 13 consecutive CpG sites within the *HOXD4* promoter represented by 40 alleles from DNA sample NE00000334. All alleles are either completely unmethylated, or contain very low levels of sporadic methylation. Methylated CpG sites are represented by black circles and unmethylated CpG sites by white circles.

Although the total number of clones sequenced in this analysis was fairly small, the results imply that significant numbers of alleles (10-20%) will be transcriptionally inactivated due to abnormal DNA hypermethylation in Newcastle 85+ study

participants with high methylation levels but these alleles are not present (or present at very low levels) in individuals with low methylation.

3.7 DNA methylation levels are correlated with proportion of lymphocytes

Recently published data has suggested that lymphocytes exhibit increased levels of methylation at differentially methylated regions compared with monocytes (Rakyan *et al.*, 2010). Many of the genes examined in this study have been shown to be associated specifically with lymphoid and not myeloid leukaemia; therefore it was important to ascertain whether the proportion of lymphocytes in the Newcastle 85+ study peripheral blood samples was related to the DNA methylation levels observed at the five variably methylated genes. The percentage of lymphocytes within the total white blood cell population was calculated using measurements from full blood counts (Chapter 2.1) and was correlated to the DNA methylation levels at the *TWIST2*, *HOXD4*, *EPHA10*, *TUSC3* and *HAND2* promoter region for each participant. The resulting correlations are to some extent consistent with the idea that lymphocytes may harbour higher levels of methylation at the 5 variable genes, seen by a positive correlation between percentage of lymphocytes in an individual and methylation levels at any gene. However, the correlations appear to be fairly weak, are not statistically significant at most of the loci, only at *HOXD4* (r=0.43, p=0.002) and *TUSC3* (r=0.47, p=0.005) (Table 3.8 and Figure 3.11). Therefore lymphocyte percentage appears to explain only a minority of the extensive variation seen in methylation levels in these 50 individuals within the Newcastle 85+ study population.

In addition, there is extensive data within the Newcastle 85+ study including flow cytometry characterisation of specific leukocyte subtypes within individual blood samples. Calculation of the Pearson correlation coefficient indicated no relationship between a participants average DNA methylation level and B-cell number in the pilot study Newcastle 85+ study samples (r=-0.12, p=0.41).

	TWIST2	*HOXD4*	*EPHA10*	*TUSC3*	*HAND2*	Average
Percentage of lymphocytes	r = 0.22 p = 0.13	r = 0.43 p = 0.002	r = 0.09 p = 0.53	r = 0.47 p = 0.005	r = 0.24 p = 0.10	r = 0.38 p = 0.006

Table 3.8: Correlation between the DNA methylation levels of *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* and the percentage of lymphocytes measured in the same individual.

The Pearson correlation coefficient (r) and corresponding p value was calculated for each locus. Black p value = NS, Red p value = significant at <0.008 level after performing Bonferroni correction for multiple comparisons.

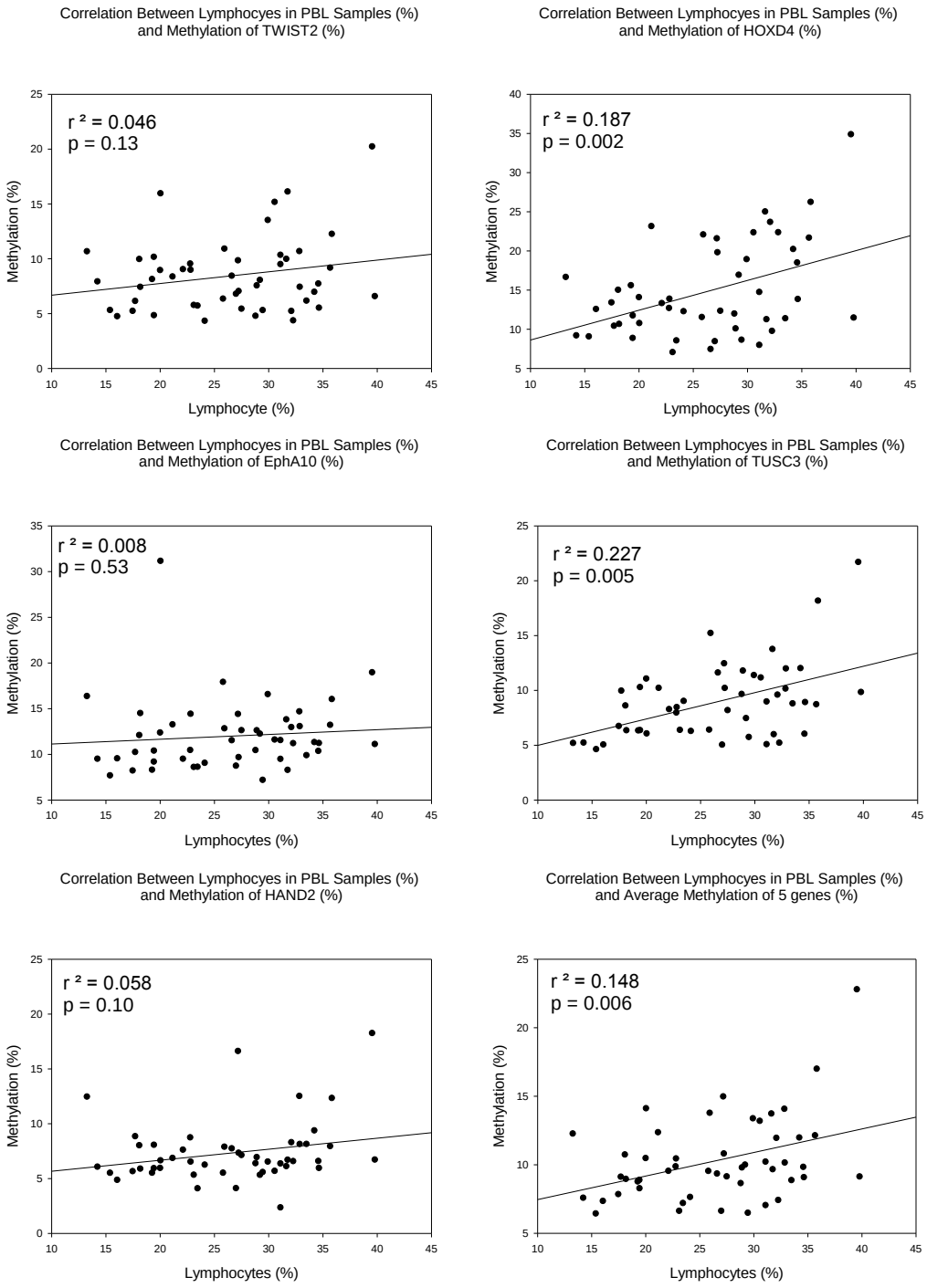


Figure 3.11: Relationship between the DNA methylation levels of *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* and the percentage of lymphocytes measured in the same individual in the Newcastle 85+ study

Methylation values of *TWIST2, HOXD4, EPHA10, TUSC3, HAND2* and the overall average DNA methylation level of the five genes were plotted against lymphocyte percentage calculated from full blood counts (Chapter 2.1). The linear regression line and r^2 value indicate a fairly weak relationship at most loci.

3.8 DNA methylation measurements are stable over 18 months

In order for DNA methylation to be considered a useful measurement or biomarker to be analysed in relation to health outcomes, it should be relatively stable without fluctuations over a short space of time. In order to test this, Pyrosequencing measurements of *TUSC3*, *TWIST2*, *HOXD4*, *HAND2* and *EPHA10* was performed for 37 of the 50 pilot study participants for which a DNA sample was available at phase II of the Newcastle 85+ study, 18 months after initial sample collection at phase I. The expectation was that DNA methylation measurements should have changed very little in such a short space of time in most individuals, as previous studies have shown stability of gene promoter methylation, even over a much longer period of 11-20 years (Talens *et al.*, 2010).

Unfortunately, although PBL DNA (i.e. total white blood cells) was analysed in phase I this was not available at phase II and DNA from peripheral blood mononuclear cells (PBMCs) was obtained instead. This led to some important differences in DNA methylation levels due to the differences in cell type composition. Typically a high percentage of cells in a PBL sample are neutrophils with the remainder being lymphocytes, monocytes, eosinophils and basophils, whereas PBMCs are comprised of the same cells but excluding neutrophils. The fact that the proportion of lymphocytes in the PBMC sample is increased compared to a PBL sample, in addition to recently published data suggesting lymphocytes exhibit increased methylation levels of differentially methylated regions, led us to the anticipate that DNA methylation measurements will be higher in samples consisting of just PBMCs compared to PBLs. For this reason, the methylation levels at phase I with phase II cannot be directly compared and instead the Pearson test of correlation was used to ascertain the relationship between methylation levels in the same individual at the two different phases. The results show that methylation levels at phase I are much lower than phase II as predicted based on the different cell compositions (Table 3.9), but the correlations between phase I and phase II measurements are strong and highly significant (Table 3.10 and Figure 3.12). This means that DNA methylation measurements are likely to be stable over time, as fluctuations in DNA methylation within an individual over time would result in a lack of significant relationship between DNA methylation at phase I and phase II across the whole population.

	Phase I (n=37)			Phase II (n=37)		
Gene	Methylation range (%)	Mean Methylation (%)	Standard Deviation (%)	Methylation range (%)	Mean Methylation (%)	Standard Deviation (%)
HOXD4	4-35	14.5	6.4	8-43	20.1	7.2
EPHA10	7-31	12.0	4.1	8-35	15.0	4.7
TUSC3	5-22	9.3	3.6	5-28	13.5	5.7
TWIST2	3-20	8.9	3.7	4-29	9.8	6.0
HAND2	2-18	7.4	2.8	4-22	11.0	3.9

Table 3.9: Comparison of DNA methylation properties at the *HOXD4, EPHA10, TUSC3, TWIST2* and *HAND2* promoter in the phase I PBL samples and Phase II PBMC samples from the Newcastle 85+ study participants

Methylation was measured in the 37 samples for which DNA from both phases was available. The mean, minimum and maximum methylation levels for each gene are shown alongside standard deviation as a measure of variation

	Phase I					
	TWIST2	*HOXD4*	*EPHA10*	TUSC3	*HAND2*	Average
Phase II	r = 0.48 p = 0.003	r = 0.62 p = 0.00006	r = 0.82 p = 8.2E-10	r = 0.83 p = 1.4E-10	r = 0.85 p = 3.7E-11	r = 0.76 p = 5.6E-08

Table 3.10: Correlation between DNA methylation levels at the, *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* promoter in Phase I PBL DNA sample and Phase II PBMC samples from the Newcastle 85+ study participants

The Pearson correlation coefficient (r) and corresponding p value is given for each locus. Correlations are strong, positive and highly significant, signifying a robust relationship. Red p value = significant at <0.008 level after performing Bonferroni correction for multiple comparisons.

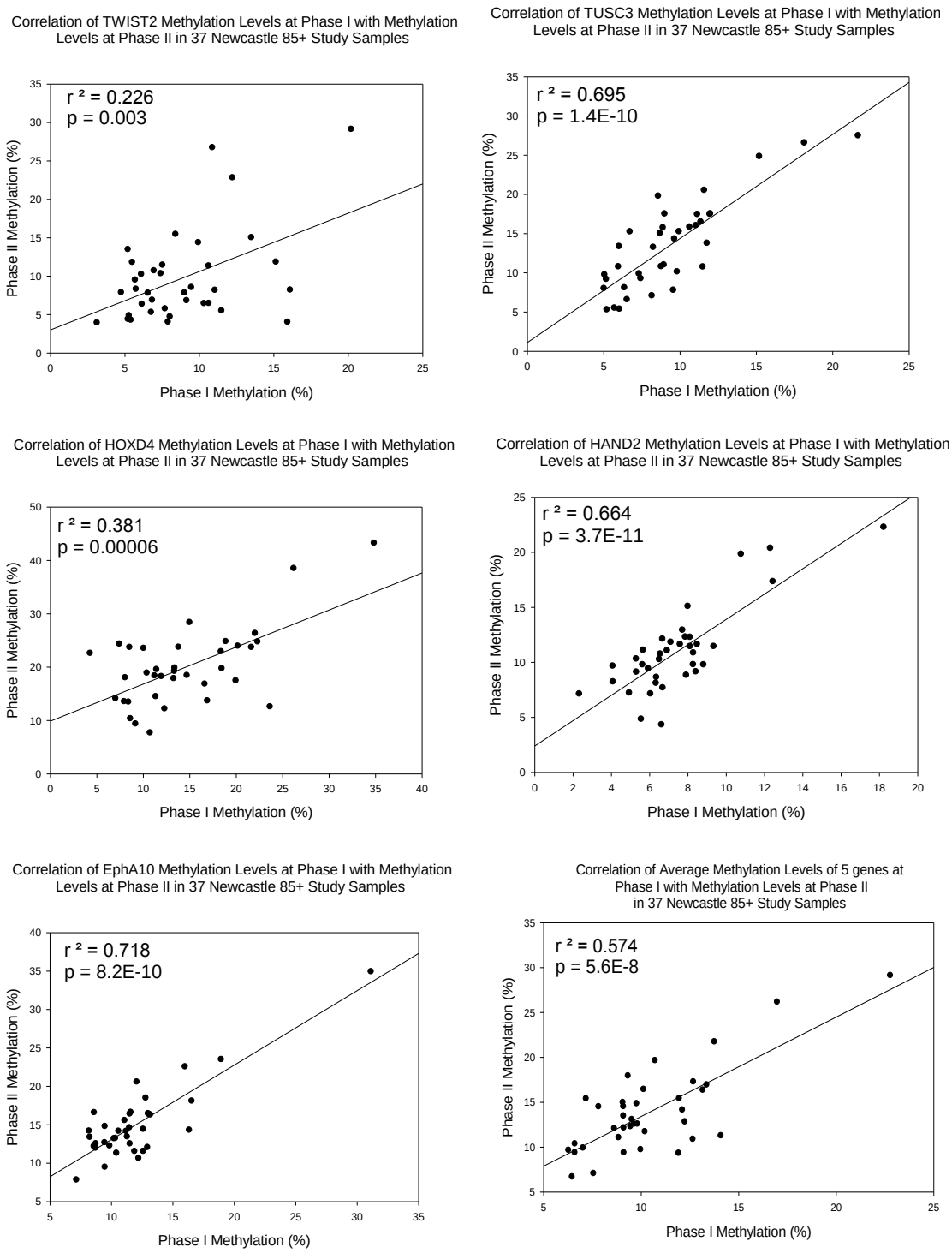


Figure 3.12: Correlation between DNA methylation levels in PBL DNA samples from Phase I and PBMC DNA samples from Phase II of the Newcastle 85+ study participants

DNA methylation levels of *TWIST2, TUSC3, HOXD4, HAND2, EPHA10* and the overall average methylation level of the five genes at Phase I and Phase II were plotted. The linear regression line and r^2 value indicate a strong positive relationship between the two phases at all loci.

3.9 Discussion

The purpose of the pilot study was to determine both the extent of methylation which is occurring in elderly individuals, and the possible presence of inter-individual variation in DNA methylation within this population as a potential mechanism to partly explain individual differences in disease susceptibility during ageing.

A highly quantitative technique, Pyrosequencing, was employed for measuring DNA methylation with a high degree of accuracy, which was ensured by validating each individual assay with control samples of known methylation status. These controls were created by mixing *in vitro* methylated DNA (near to 100% methylated across the genome) with normal PBL DNA from young volunteers (typically unmethylated at gene promoters). An alternative method to validate these assays involves use of a nested PCR to amplify a region slightly larger than the amplicon of interest. This nested PCR product will be free of methylation and can be subjected to *in vitro* methylation using SSS1 methylase enzyme. Mixing differing proportions of the unmethylated and methylated PCR product enables the creation of known methylation standards which range from 0-100%. Since three of the genes in this study, *HOXA4*, *HOXA5* and *IGF2*, are all partially methylated in normal cells, it could be argued that using a control which is also partially methylated is not ideal (as PBL DNA would be at these loci), as the assay has not demonstrated detection of methylation below the normal *in vivo* measurement (with the exception of *HOXA4*). A disadvantage of using a nested PCR approach however, is that the controls used to validate the assay (PCR product) will be different from the test sample (PBL DNA), whereas the method that we used ensures that the same kind of template is used for both validation and test samples (both PBL DNA). Additionally, assays may be vulnerable to contamination with amplified DNA from the first round of PCR due to the formation of large quantities of PCR product by nested PCR.

The results demonstrated that methylation levels at a particular set of loci can accumulate both high levels of DNA methylation, and are extremely variable between individuals of the same age. Included within the group of genes exhibiting considerable levels of methylation and inter-individual variation are those which are known to be expressed within the haematopoietic lineage and/or have a functional role in haematopoietic cells, so it is possible that increased DNA methylation level at these

genes could have a biological consequence. For example, *HOXA4* is expressed in haematopoietic progenitor cells, suggesting that it could play a role in control of differentiation. Also, *HOXA5* is a key regulator of myeloid differentiation, as shown by altered monocytic and erthyroid cell differentiation after *HOXA5* over-expression and inhibition experiments (Crooks *et al.*, 1999; Fuller *et al.*, 1999). In addition, *TWIST2* is involved in the negative regulation of myeloid lineage development and *TWIST2* deficient mice have increased mature myeloid cell type populations including macrophages, neutrophils and basophils compared with WT mice (Sharabi *et al.*, 2008).

Upon investigation of the properties of these fifteen loci, we found that a significant proportion of these loci including *TWIST2, TUSC3, HOXD4, EPHA10, HAND2, p15, HLXB9* and *ESR1* were susceptible to increased and variable methylation in haematopoietic cells. For these eight loci, there is evidence in the literature that they exhibit aberrant hypermethylation in haematological malignancies.

Previous studies have described hypermethylation of the *TWIST2* promoter in chronic lymphoblastic leukaemia (CLL) samples which is associated with gene silencing (Raval *et al.*, 2005). Bisulfite sequencing found minimal *TWIST2* methylation in normal blood cells or CD19+ B cells, however only two samples were examined. *TWIST2* methylation status has attracted further study within our lab and we have recently shown that the *TWIST2* promoter is frequently hypermethylated in both child and adult acute lymphoblastic leukaemia (ALL) which was associated with silenced expression. We also confirmed the methylation previously described in CLL samples, although at lower levels than ALL samples (Thathia *et al.*, 2012). The re-expression of *TWIST2* in ALL cell lines led to an inhibition of growth, induction of apoptosis as well as increasing susceptibility to chemotherapy, suggesting that methylation induced silencing of *TWIST2* may confer a survival advantage for leukemic cells *in vivo*.

DNA methylation at the *TUSC3* promoter was recently identified as being hypermethylated in cases of adult ALL, but not AML by microarray analysis (Scholz *et al.*, 2005). As far as I am aware, this remains the only published finding of *TUSC3* methylation in haematological malignancy, however our findings also indicate that hypermethylation of *TUSC3* is frequently seen in adult and childhood ALL, where approximately one third of patients have methylation levels in excess of 50%. In contrast methylation of TUSC3 appears rarely in myeloid leukaemia.

Although there is no published data describing the methylation status of the *HOXD4* promoter in hematopoietic malignancies, we have found hypermethylation of this region in various leukaemia subtypes. In addition, a study looking for germline mutations in the HOX cluster which may be causal in childhood ALL samples found only three mutations unique to the childhood ALL cases. Of these, two were the same missense mutation in the *HOXD4* locus which resulted in some loss of function (van Scherpenzeel Thim *et al.*, 2005). This suggests that disturbing the normal function of *HOXD4* may play a role in ALL development, therefore disruption of the gene by DNA methylation rather than mutation may also be functionally relevant.

Hypermethylation of the Eph/ephrin family gene promoters is frequent event in ALL, including *EPHA10,* with methylation levels exceeding 60% in most ALL samples. In addition, this study included ten normal blood samples for which methylation levels were highly variable, ranging from 7.3-27.1% (Kuang *et al.*, 2010).

Abnormal DNA methylation of the *HAND2* promoter was identified from a promoter methylation microarray of aberrantly methylated genes in CLL. Methylation density of the *HAND2* promoter was typically in the range of 10-49% with some samples measuring >50% methylation. The methylation levels of blood cells from healthy controls were in the range of 2-8%, although this is based on a relatively small sample size of 10 individuals (Tong *et al.*, 2010). Highly significant age related increases in *Hand2* methylation have also observed in mouse tissues such as small intestine, spleen, kidney and liver (Maegawa *et al.*, 2010).

Hypermethylation of the *p15* promoter is common in various types of leukaemia (Wong *et al.*, 2000; Chim *et al.*, 2001a; Chim *et al.*, 2001b) and is significantly associated with disease progression including reduced disease free survival, lower overall survival, lower likelihood of achieving remission in AML (Shimamoto *et al.*, 2005) and increased likelihood of relapse if *p15* remains methylated in complete remission of both AML and ALL (Agrawal *et al.*, 2007). Notably, p15 methylation has primarily been reported to be methylated in myeloid, but not lymphoid leukaemia.

Hypermethylation (>50%) of the homeobox gene *HLXB9* is a frequent event in adult and childhood ALL, although rarely appears in other leukaemia subtypes (Ferguson *et al.*, 2011).

ESR1 promoter hypermethylation has been described in a wide range of haematopoietic neoplasms including both childhood and adult ALL, adult AML, chronic myeloid leukaemia (CML) and lymphoma samples (Issa *et al.*, 1996b). Normal PBL samples examined as controls for this study of haematopoietic malignancies found fairly low levels of *ESR1* promoter methylation (1-4%) and no association with age; however this conclusion was based on a small sample size of 17 individuals aged 29-67 years which are considerably younger than the Newcastle 85+ study population. Methylation was measured by southern blot analysis which is a less sensitive technique than Pyrosequencing, meaning some of the *ESR1* promoter methylation in normal PBL samples may have remained undetected in previous analysis.

Furthermore, we observed that methylation levels at any one of these loci, was strongly and positively associated with methylation levels at other genes, which could imply that certain individuals are predisposed to higher levels of DNA methylation. The fact that multiple loci across the genome are affected hints towards a general deregulation of methylation mechanisms with ageing. A similar co-methylation of multiple genes is noted in the literature as a feature of a both leukemic (Melki *et al.*, 1999; Garcia-Manero *et al.*, 2002) and solid tumour cells (Bai *et al.*, 2004; Lee *et al.*, 2004), although since the levels of methylation tend to be very high in tumour cells, these studies looked solely for the frequency of hypermethylation at multiple genes, rather than comparing a quantitative level of methylation between different genes as we did in the pilot study.

For the other highly methylated loci, *HOXA4*, *HOXA5* and *IGF2*, which have high levels of inter-individual variation, there is similar evidence that they are hypermethylated in leukaemia. However, methylation levels at these loci are not correlated with any other gene, which likely is due to differences in how methylation is acquired and regulated at partially methylated compared to methylation free promoters.

High levels of *HOXA4* promoter methylation have been discovered in CLL (Strathdee *et al.*, 2006), and the presence of such methylation highlighted as a possible biomarker for prognosis in CLL patients (Irving *et al.*, 2011). Aberrant hypermethylation of *HOXA4* is a feature of many leukemia subtypes including ALL, CML, where it is strongly associated with disease progression and AML, where is may be linked to poor prognosis (Strathdee *et al.*, 2007a).

Very high levels of *HOXA5* promoter methylation (>80%) have been described in numerous type of leukaemia including ALL, AML CLL and CML (Strathdee *et al.*, 2007a; Kim *et al.*, 2010b) and again been linked to patient outcome.

Studies have identified extensive methylation of *IGF2* in leukemic cell lines and in a small number of primary samples examined (Issa *et al.*, 1996a). Other studies have identified abnormal regulation of this gene in AML and CML leading to biallelic expression, but whether this is the consequence of altered DNA methylation levels is unclear (Wu *et al.*, 1997; Randhawa *et al.*, 1998).

Conversely, for the loci which display low levels of methylation and a lack of variability including *MLH1*, *RASSF1A*, and *RARB2*; or a lack of correlation with methylation levels at other genes such as *APOE*, there is little or no evidence for their hypermethylation in leukaemia, despite these genes presenting age associated aberrant methylation patterns.

The *MLH1* gene is involved in DNA repair, specifically mismatch repair (MMR). Defects in the MMR pathway, due to MLH1 silencing by hypermethylation are associated with a microsatellite instability (MSI) phenotype in colorectal cancer (CRC) (Poynter *et al.*, 2008) and endometrial cancer (Simpkins *et al.*, 1999). However, a study examining *MLH1* status in colorectal tissue from disease free individuals found that absolute methylation measurements were extremely low, around 2% in CRC samples and even lower in controls (Al-Ghnaniem *et al.*, 2007). This suggests DNA methylation increases are small at this locus which agrees with our findings.

Hypermethylation of *RASSF1A* has been described in numerous types of cancer (detailed extensively in (Dammann *et al.*, 2005)) and partial methylation described in normal prostate tissue (Kwabi-Addo *et al.*, 2007) and kidney samples, which was related to increasing age of the individual (Peters *et al.*, 2007).

Similarly, aberrant methylation of *RARB2* is described in solid tumours of tissues such as the prostate (Kwabi-Addo *et al.*, 2007), and a positive correlation between methylation of *RARB2* and age was also described in this tissue (Kwabi-Addo *et al.*, 2007; Christensen *et al.*, 2009).

Currently, there is no evidence for the presence of *APOE* methylation in haematological malignancies, A study found differential DNA methylation of *APOE* in the brains of Alzheimer's disease (AD) patients compared to controls which was associated with advancing age, and the amount of methylation was variable within all individual studied. However, these findings were not observed in lymphocytes. (Wang *et al.*, 2008).

In addition to these outcomes, we found that DNA methylation of the *TWIST2* and *HOXD4* promoter at the allelic level is represented by a mixture of alleles containing either densely methylated or unmethylated CpG sites in normal haematopoietic cells in the Newcastle 85+ study population. This again draws similarities to a cancer-like cell, in which genes at the allelic level are often densely, rather than partially methylated (Corn *et al.*, 2000; Esteller *et al.*, 2000b; Varley *et al.*, 2009).

Taken together, these findings led to the development of the pre-existing methylation hypothesis (Figure 3.13) to explain how an individual's level of DNA methylation at loci which are susceptible to both hypermethylation and acquisition of methylation during ageing, may underlie risk of developing cancer.

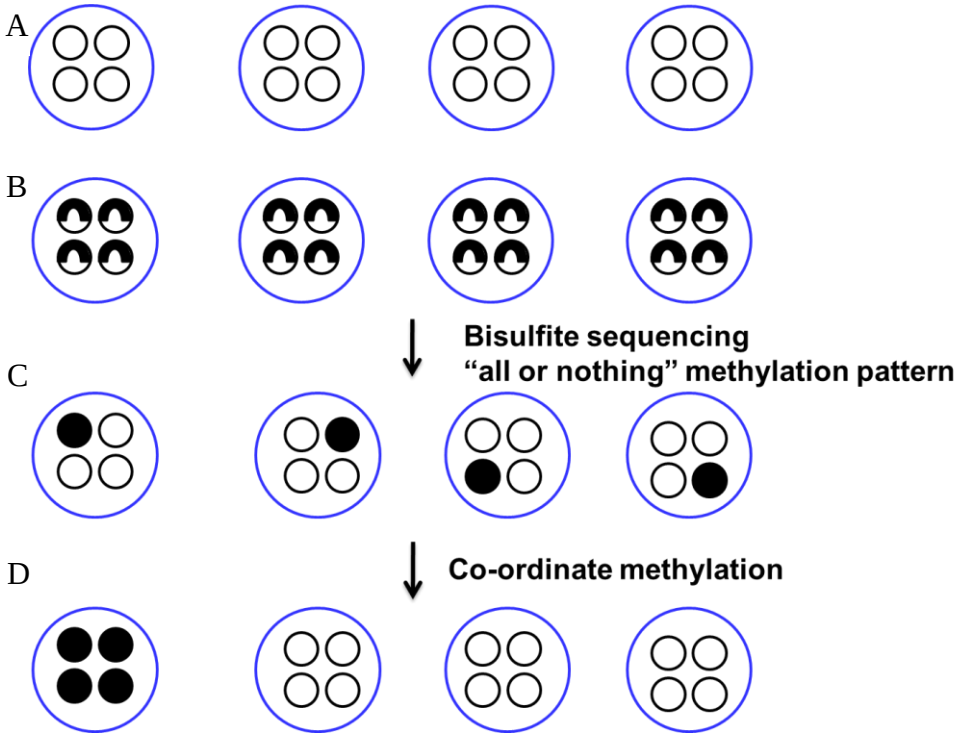


Figure 3.13: The pre-existing methylation hypothesis

Previously, it was believed that the majority genes in a normal stem cell were completely unmethylated. In the example shown (Figure 3.13A), four haematopoietic stem cells (circles, blue outline) are shown and four genes susceptible to aberrant methylation in cancer and ageing are represented within the cells (circles, black outline) which entirely lack methylation. It has since become apparent, through others and the observations presented here, that these types of loci are partially methylated to variable degrees during ageing. Figure 3.13B illustrates a possible pattern of partial methylation in the four stem cells whereby all four cancer-related genes are partially methylated (partial black colouring) in all cells a normal sample. However, we know from our bisulfite sequencing results of the *TWIST2* and *HOXD4* promoter region that partially methylated alleles are not common, rather the allele appears to be either densely methylated or completely unmethylated. In Figure 3.13C, each of the four stem cells have complete methylation of a one gene (black circle) which is different in each cells, with the others remaining unmethylated (white circles). Expression of the gene will likely be silenced in the cell where it is heavily methylated but the gene may be expressed within cells where it remains unmethylated. Furthermore, at present we know that co-ordinate methylation of numerous genes occurs at the level of an individual, and we can speculate that this may be occurring at the level of an individual stem cell. This is illustrated in Figure 3.13D, whereby all four loci, susceptible to both cancer and age associated methylation, are completely methylated in one cell (black circles), and unmethylated in others (white circles). This is identical to observations of tumour cells, in which numerous genes are hypermethylated and silenced. Consequently, if we hypothesise that the partial methylation measured in the PBL DNA of Newcastle 85+ study participants is represented by the existence of completely methylated cells such as those in Figure 3.13D, an individual with higher levels of methylation may likely have higher numbers of these abnormally methylated cells. Our hypothesis is therefore, that an individual, who is predisposed to acquiring abnormal methylation during ageing, may be more susceptible to cancer as they have an increased pool of normal cells with an abnormal methylation pattern from which cancer can originate via clonal expansion, without requiring any additional methylation changes.

The relationship between methylation levels and the percentage of lymphocytes was suspected due to the recent observation of increased methylation of lymphocytes compared to monocytes (Rakyan *et al.*, 2010). A study of haematopoietic differentiation in mice also revealed higher levels of global methylation in lymphocytes than myeloid

lineage cells (Ji *et al.*, 2010). Although we found some evidence of an association in our pilot study, it was weak and not significant at most of the loci, therefore is unlikely to be having a substantial impact on the DNA methylation levels measured in this pilot study.

A comparison between methylation levels of DNA samples taken 18 months apart at phase I and phase II of the Newcastle 85+ study was undertaken to confirm that DNA methylation measurements were relatively stable within an individual over a short space of time and not susceptible to day-to-day fluctuations. Fluxing levels would render DNA methylation an impractical measurement for relating to health outcomes. Our analysis was made more complicated by the different sources of DNA at the two phases, derived from PBLs in Phase I and PBMCs in Phase II, which affects overall levels of DNA methylation due to intrinsic methylation differences between cells types. In spite of this, methylation levels at phase I and phase II were strongly, positively correlated suggesting little variation of methylation over time, within the same individual.

In conclusion, these results demonstrate that in elderly individuals, increased DNA methylation at the promoter region of multiple loci is evident but does not accumulate to the same level throughout the population, and instead is highly variable. In addition, these age-related methylation changes share similarities with those seen in tumour cells including the presence of densely methylated alleles and co-methylation of genes indicative of a "CpG island methylator phenotype", suggesting age-associated changes may have a functional consequence.

Chapter 4 – Association between DNA methylation levels and clinical outcomes in the Newcastle 85+ Study

4.1 Introduction

For many diseases, advancing age is a significant risk factor. According to Cancer Research UK, approximately 63% of all cancer (excluding non-melanoma skin cancer) is diagnosed in individuals aged over 65 years, and despite representing a relatively small proportion of the population over a third of cancers are found in those aged over 75 (URL http://info.cancerresearchuk.org/cancerstats/incidence/age/).

Numerous factors affect a person's risk of developing a cancer. Germline mutations in genes such as *BRCA1* are associated with a vastly increased incidence of familial breast or ovarian cancer (Lancaster *et al.*, 1997), but such high penetrance germline defects account for only a small proportion of all cancer cases and are typically responsible for cancers at a younger age of onset than sporadic cancers (Brandt *et al.*, 2008). Studies examining the presence of SNPs within the genome have highlighted regions which may be important for determining cancer risk, including the C677T SNP in *MTHFR*, a gene involved in the metabolism of folate and the provision of substrates for DNA synthesis and for methylation of DNA and other cell macromolecules. The TT genotype encodes an enzyme with reduced activity compared with the CC genotype and appears to decrease risk of colorectal cancer (Levine *et al.*, 2010) and acute lymphoblastic leukaemia (Skibola *et al.*, 1999). However, this effect was not observed in all analyses of colorectal cancer (Plaschke *et al.*, 2003), or in myeloid leukemia (Wiemels *et al.*, 2001). In addition, specific environmental exposures are known to contribute to cancer risk. For example, the incidence of lung cancer is significantly higher in cigarette smokers compared with never smokers (Freedman *et al.*, 2008).

These genetic factors alone do not appear to account for all of the risk of cancer development. This is illustrated by MZ twins, who despite sharing an identical genetic background and typically very similar environments (at least early in life), often demonstrate discordance with regards to diseases, particularly cancer. A study of childhood cancer incidence in MZ twins described extremely low concordance of approximately 5% for leukaemia and 2% for solid tumours between pairs (Buckley *et al.*, 1996). A large study of both MZ and DZ twins found that although there was some

evidence of genetic susceptibility, many types of cancer displayed considerable discordance between twin pairs (Ahlbom *et al.*, 1997).

Although stochastic events can likely explain the majority of disease discordance, another proposed explanation underlying such differences is DNA methylation. Indeed differential methylation status of genes has been implicated in the discrepancy of disease incidence between monozygotic twins including schizophrenia (Petronis *et al.*, 2003), systemic lupus erythematosus (Javierre *et al.*, 2010) and psoriasis (Gervin *et al.*, 2012). In addition, a recent EWAS of MZ twins discordant for Type I diabetes identified a number of differentially methylated CpG sites in the affected twin, which were also present in singletons with the disease, even prior to disease onset, suggesting a potential causal role for DNA methylation in Type I diabetes (Rakyan *et al.*, 2011).

The results from the pilot study (Chapter 3) suggest strongly that DNA methylation levels are highly variable between individuals at age 85. The presence of such variation may be important when considering the dramatic differences in susceptibility to diseases such as cancer. In addition to the increased epigenetic variability between individuals with age, there are similarities between the DNA methylation patterns observed in peripheral blood leukocyte cells of elderly individuals and those in cancer cells. Such features include targeting of the same or an overlapping set of genes, the co-methylation of genes, suggestive of a "CpG island methylator phenotype" and the presence of densely hypermethylated alleles.

These observations led to the construction of a hypothesis that pre-existing methylation in normal cells may underlie an individual's susceptibility to disease, particularly cancer, since the consequence of altered methylation in cancer is well described. Specifically I hypothesised that those 85 year olds with the highest amounts of DNA methylation in the promoter regions of age- and cancer-related genes may be at an increased risk of developing cancer, as they contain an increased pool of abnormally densely methylated cells, which could potentially clonally expand. To test this hypothesis, methylation levels of the genes *HOXD4, TUSC3, TWIST2, HAND2, EPHA10* and *HOXA5* shown to be variable methylated in the pilot study, will be used as surrogate loci for methylation changes occurring at genes throughout the genome and DNA methylation levels of individuals with a cancer diagnosis will be compared to those who are cancer free.

In addition to cancer, the prevalence of cardiovascular disease, dementia, type II diabetes and stroke also rises dramatically with age and altered DNA methylation has been proposed to play a role in the aetiology of such diseases (Sharma *et al.*, 2008; Barrachina and Ferrer, 2009; Ling and Groop, 2009; Baccarelli *et al.*, 2010). However these studies primarily described associations between altered DNA methylation and disease status and did not demonstrate a direct biological (causal) role for the epigenetic changes. In addition, many of these analyses measured global DNA methylation levels and the functional consequences of such changes are not clearly understood

To uncover potential relationships between methylation levels and the incidence of other diseases including cardiovascular disease, type II diabetes, stroke, and dementia, methylation levels were compared between individuals with and without these diseases. In addition, relationships between potential key confounders such as gender or smoking behaviour and methylation were investigated.

4.2 Results – initial analysis

To test our hypothesis that pre-existing methylation levels in normal haematopoietic cells may underlie susceptibility to cancer within the Newcastle 85+ study population, PBL DNA samples from all 113 participants in the study who had a previous incidence of any cancer (excluding non-melanoma skin cancer) prior to their recruitment onto the study, were selected for DNA methylation analysis. A control group consisting of PBL DNA from 113 participants was selected at random from all remaining study participants.

4.2.1 Pyrosequencing Results – initial analysis

DNA methylation levels were measured at six of the same genes tested in the pilot study including *HOXD4, TUSC3, TWIST2, HAND2* and *EPHA10*, as these were the most variably methylated of the 'normally unmethylated' gene promoters, in addition to *HOXA5*, as this was the most extensively and variably methylated of all genes.

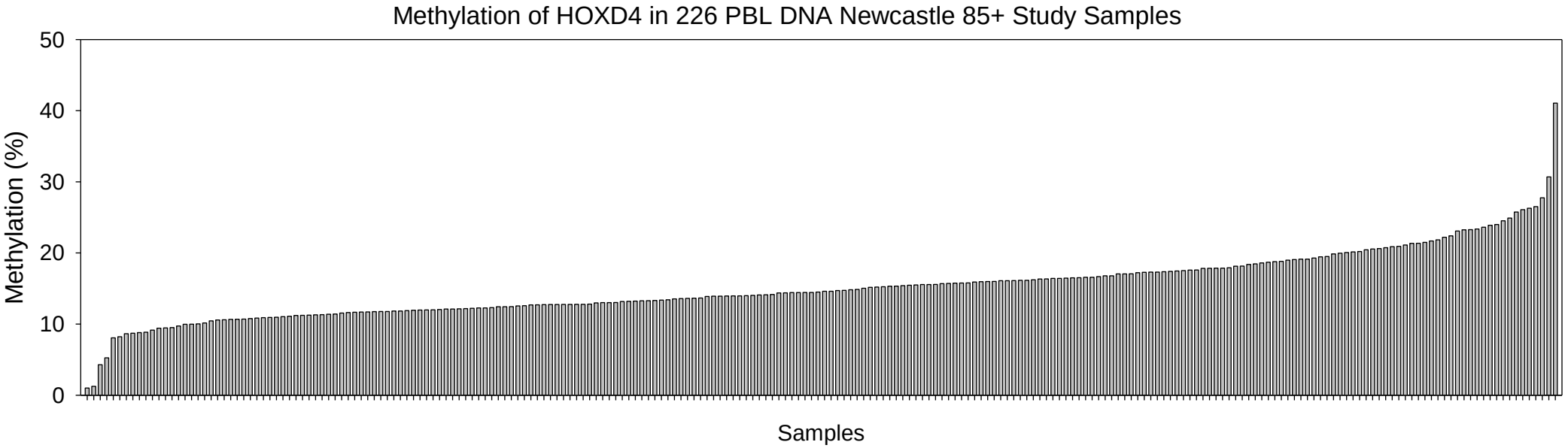
The results for the total sample set of 226 participants confirmed the highly variable nature of DNA methylation which was observed in the pilot study at *HOXD4* (15.3 ±

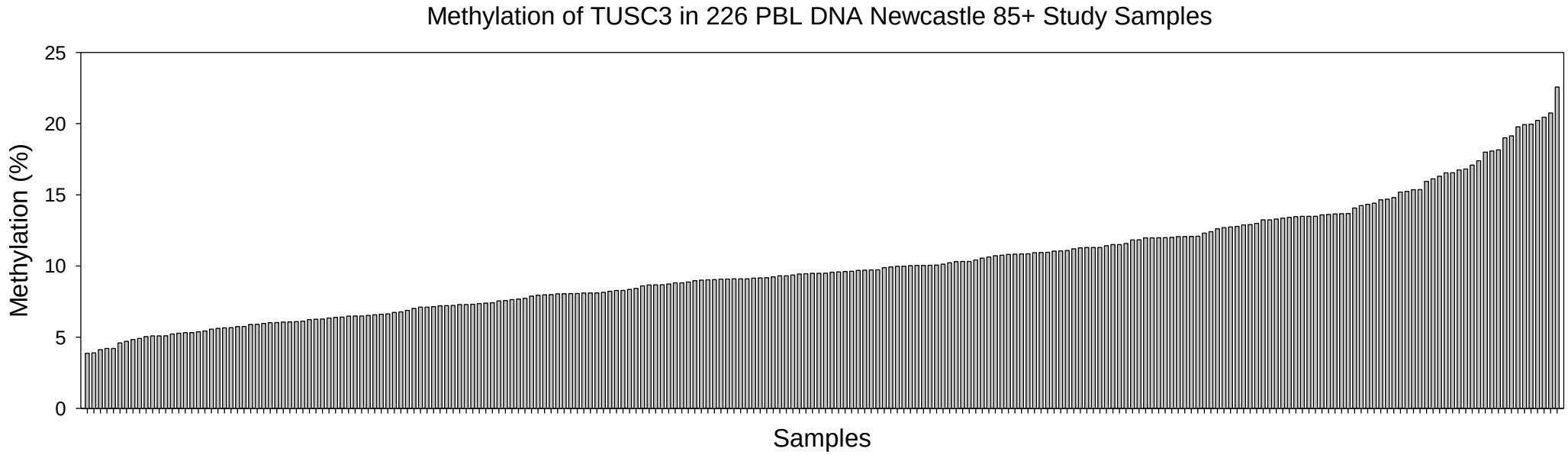
4.8% [mean ± SD]), *TUSC3* (10.1 ± 3.8%), *TWIST2* (6.8 ± 3.1%), *EPHA10* (11.7 ± 2.9%) and *HAND2* (8.4 ± 2.7%). DNA methylation levels at the *HOXA5* promoter (ranging from 31-91%) were measured in 113 participants (56 with no cancer and 57 with previous cancers) and demonstrated extremely variable measurements within the 85 year old population (Table 4.1 and Figure 4.1). Overall, these measurements were very similar to those which were observed in the pilot study at *HOXD4* (14.8 ± 6.0%), *TUSC3* (8.9 ± 3.4%), *EPHA10* (12.0 ± 3.9%), *HAND2* (7.3 ± 2.8%) and *HOXA5* (61.5 ± 15.1%) (Chapter 3.5). DNA methylation levels in this expanded set of participants were slightly lower for *TWIST2* than were observed in the pilot study (6.8 ± 3.1% and 8.4 ± 3.3% respectively).

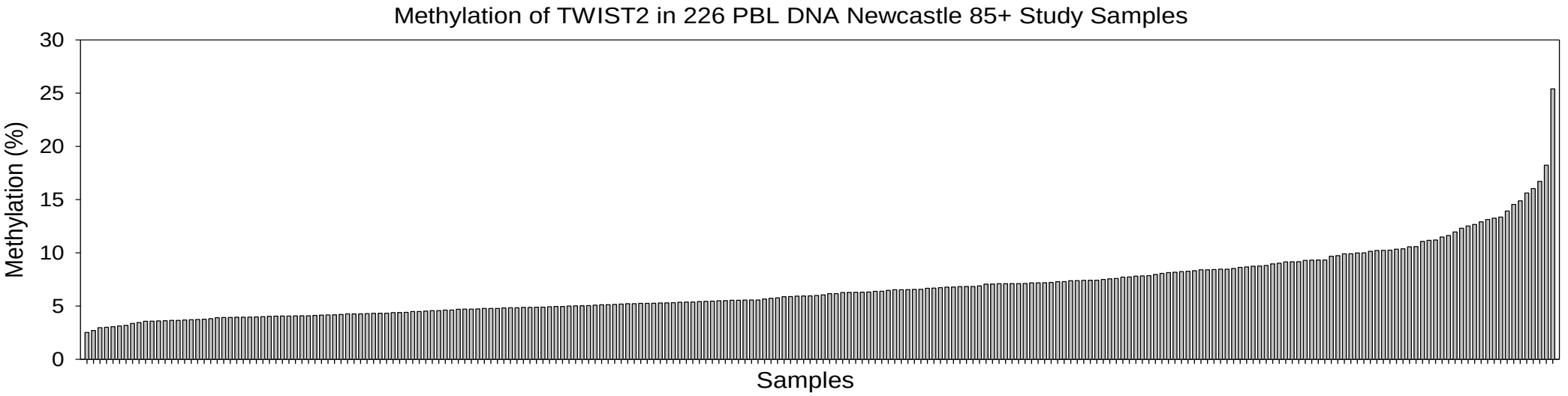
Gene	n	Methylation Range (%)	Mean Methylation (%)	Standard Deviation (%)
HOXA5	113	31-91	61.9	12.7
HOXD4	226	1-41	15.3	4.8
TUSC3	226	4-23	10.1	3.8
TWIST2	226	3-25	6.8	3.1
EPHA10	226	3-24	11.7	3.9
HAND2	226	3-27	8.4	2.7

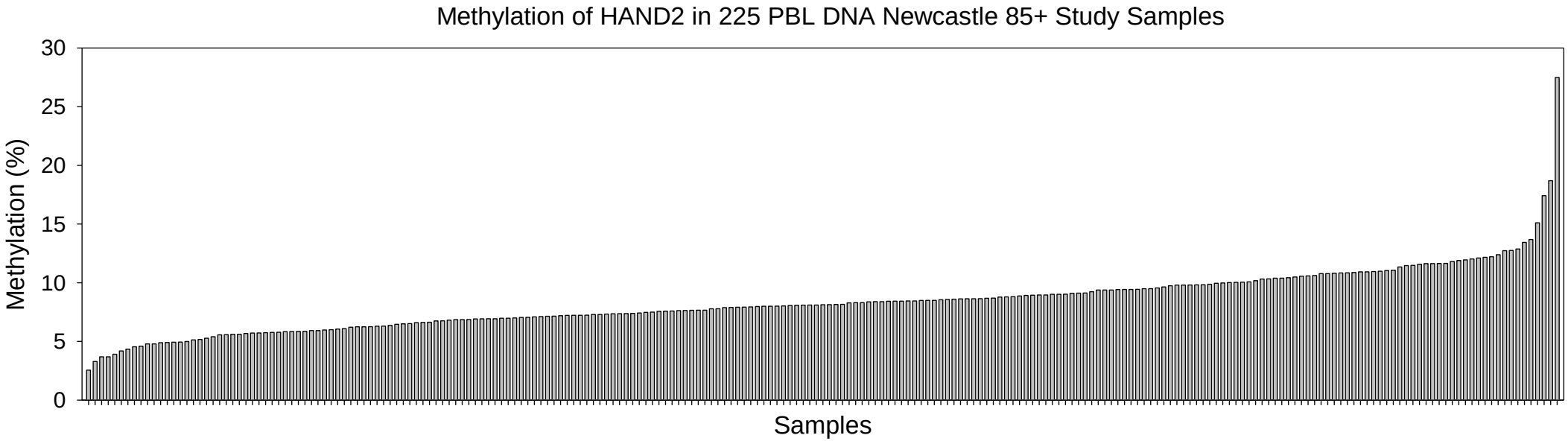
Table 4.1: Summary of methylation measurements in PBL DNA from 113 participants with a previous cancer history and 113 cancer-free participants from the Newcastle 85+ study

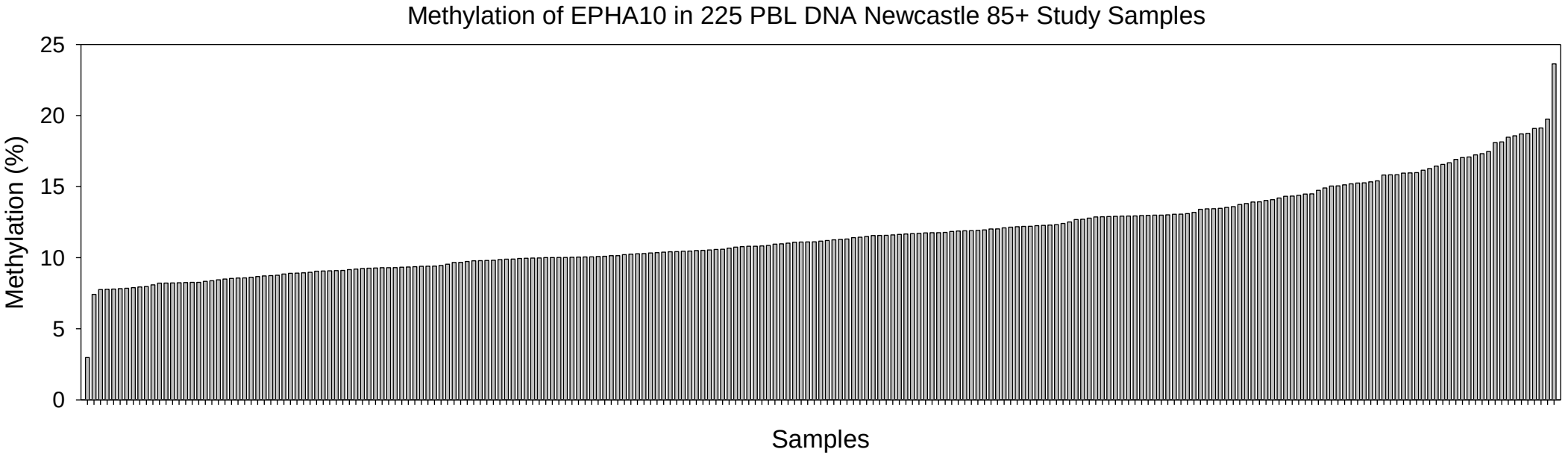
For each gene, the minimum and maximum methylation levels in the samples is shown, together with mean methylation calculated from 3-6 consecutive CpG sites for all participants and the standard deviation as a measure of variation.


Methylation of HOXD4 in 226 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
50
40
30
20
10
0
Samples


Methylation of TUSC3 in 226 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
25
20
15
10
5
0
Samples


Methylation of TWIST2 in 226 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
30
Samples


Methylation of HAND2 in 225 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
30
Samples


Methylation of EPHA10 in 225 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
Samples

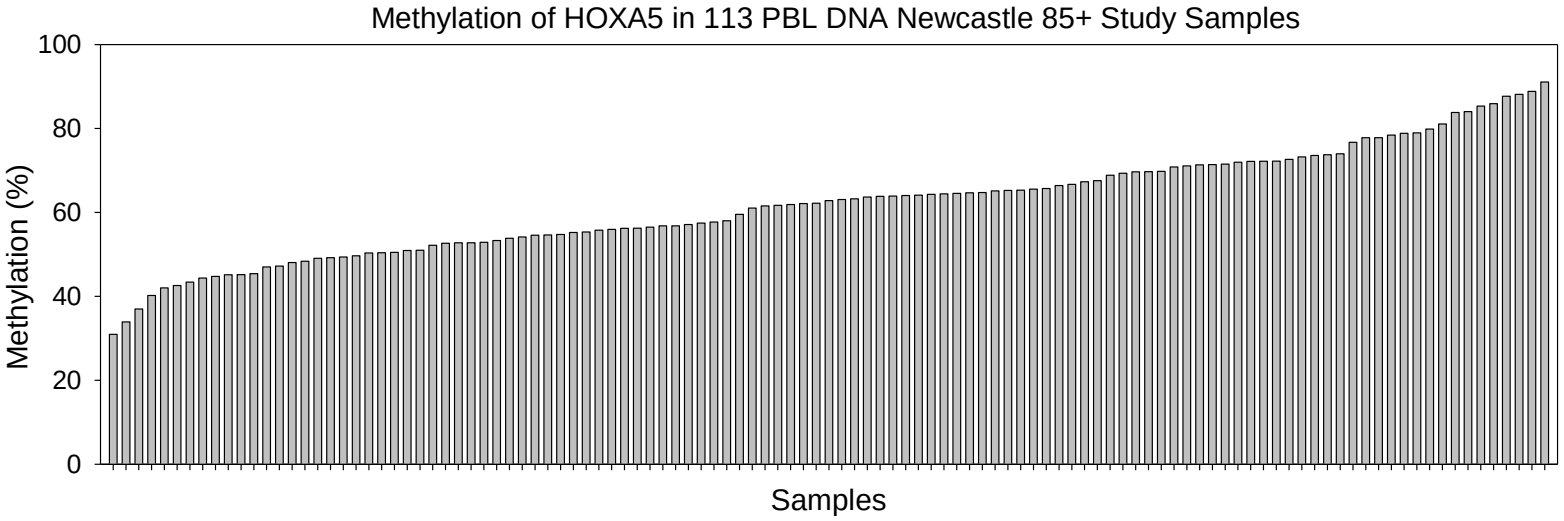

Methylation of HOXA5 in 113 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
20
40
60
80
100
Samples

Figure 4.1: Methylation levels in PBL DNA at *HOXD4, TUSC3, TWIST2, HAND2 EPHA10* and *HOXA5* promoters from 113 participants with a previous cancer history and 113 cancer-free participants from the Newcastle 85+ study

Samples are listed in ascending order of DNA methylation measurements at each particular gene. All genes exhibit highly variable DNA methylation levels between individuals. Lack of reproducible methylation measurements meant one sample was excluded from *EPHA10* and *HAND2* results. Only 113 samples were tested at the *HOXA5* locus.

4.2.2 DNA methylation levels between genes are strongly correlated

Next, in order to confirm the strong positive correlations observed between genes in the initial 50 participants, DNA methylation levels of *HOXD4, TUSC3, TWIST2, HAND2*, *EPHA10* and *HOXA5* measured in 226 participants were used to calculate the Pearson correlation coefficient and corresponding p value for each pairwise comparison.

Exceptionally strong, positive and highly significant relationships were observed between methylation levels at the *HOXD4, N33, TWIST2, HAND2* and *EPHA10* promoter regions (Table 4.2). These correlations were both stronger and more significant than the correlations observed in the pilot study (see Chapter 3.6.1). Also in agreement with the pilot study result, there was no association of *HOXA5* methylation levels with any other loci.

	TWIST2	*HOXD4*	*EPHA10*	*TUSC3*	*HAND2*
TWIST2					
HOXD4	r = 0.73 p < 1e-16				
EPHA10	r = 0.51 p = 2.2e-16	r = 0.59 p < 1e-16			
TUSC3	r = 0.73 p < 1e-16	r = 0.66 p < 1e-16	r = 0.60 p < 1e-16		
HAND2	r = 0.53 p = 0	r = 0.48 p = 5.7e-15	r = 0.36 p = 2.2e-08	r = 0.42 p = 5.3e-11	
HOXA5	r = 0.04 p = 0.62	r = -0.03 p = 0.72	r = -0.002 p = 0.98	r = -0.02 p = 0.77	r = 0.003 p = 0.97

Table 4.2: Pair-wise correlations of methylation of *TWIST2, HOXD4, EPHA10, TUSC3, HAND2* and *HOXA5* in DNA from 226 Newcastle 85+ Study participants

DNA methylation data for 226 participants was correlated at each pair of loci using Pearson test of correlation. The Pearson correlation coefficient (r) and corresponding p values are shown. DNA methylation levels at *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* are strongly positively correlated. Methylation of *HOXA5* demonstrates no relationship with methylation at the other loci investigated. Black p value = NS, Red p value = significant at <0.003 level after performing Bonferroni correction for multiple comparisons.

4.2.3 DNA methylation levels are highly correlated with proportion of lymphocytes

The results from the pilot study indicated a weak relationship between DNA methylation levels from PBL cells and the percentage of lymphocytes in the same individual (Chapter 3.7). Using the data from 226 participants, correlations between methylation levels at the *TWIST2, HOXD4, EPHA10, TUSC3, HAND2* and *HOXA5* promoter, and the proportion of lymphocytes in each individual was examined. This revealed very strong positive associations which were highly significant at *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* (Figure 4.2). In contrast, there was no relationship between *HOXA5* methylation levels and the percentage of lymphocytes.

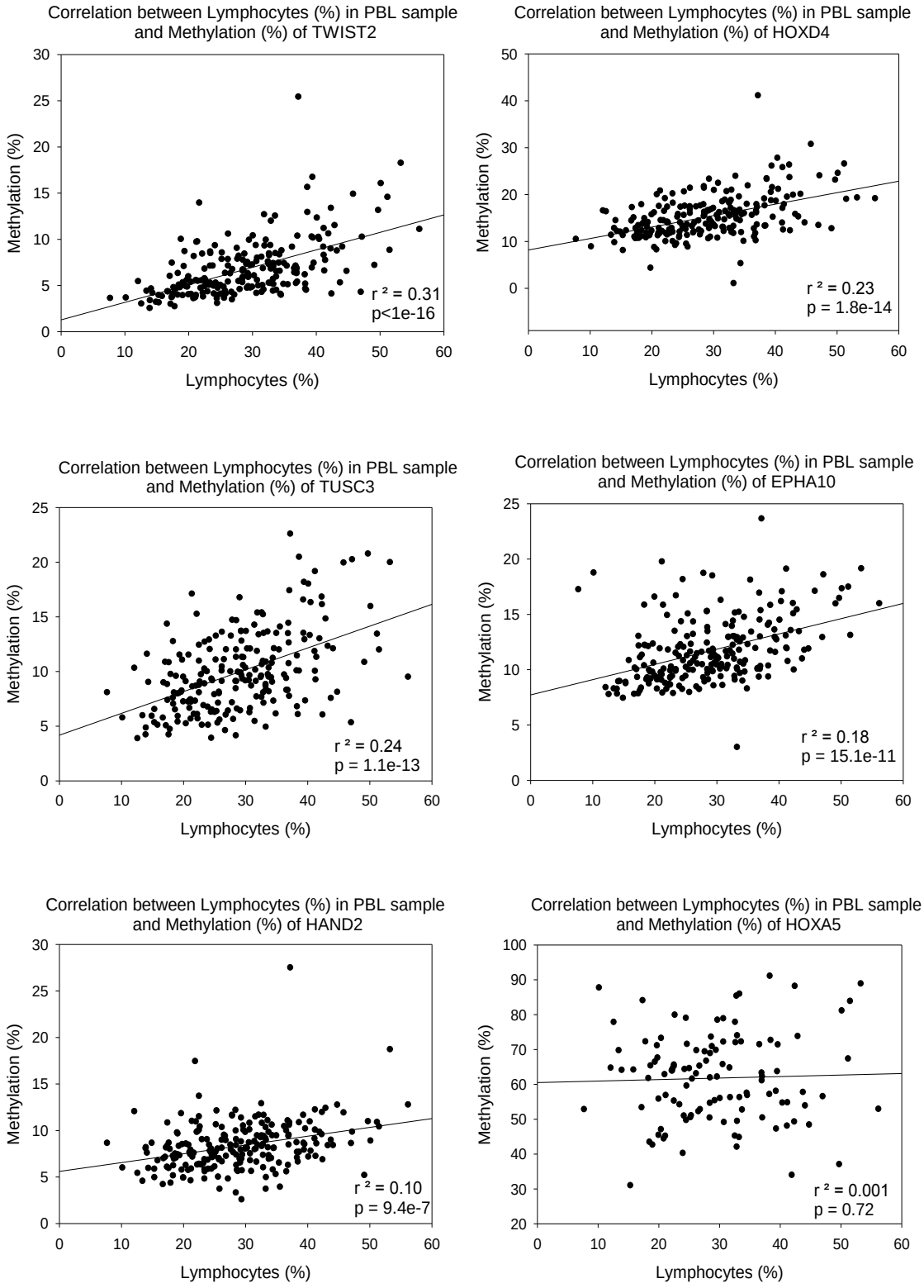


Figure 4.2: Correlations between DNA methylation levels and lymphocyte percentage in blood samples from Newcastle 85+ study participants.

DNA methylation values of *TWIST2*, *HOXD4*, *EPHA10*, *TUSC3*, *HAND2* and *HOXA5* were plotted against lymphocyte percentage calculated from full blood counts (Chapter 2.1). The linear regression line and r^2 value indicate a positive relationship for all genes except *HOXA5*, for which there is no association.

4.2.4 Correction of DNA methylation values to account for proportion of lymphocytes – initial analysis

These results indicate that the percentage of lymphocytes in any individual may significantly affect the levels of DNA methylation at these five loci. Furthermore, a recent analysis of DNA methylation levels in these same five genes in isolated haematopoietic cell populations demonstrated that lymphocytes, particularly B-cells, exhibit significantly higher levels of DNA methylation than other cells types in a PBL sample (example in Figure 4.3 and further examples in A3.1) (Unpublished data, performed by Sanne van Otterdijk). B-cells typically represent approximately 5-10% of total lymphocyte count and consequently their numbers will increase with increasing lymphocytes, and subsequently raise DNA methylation levels.

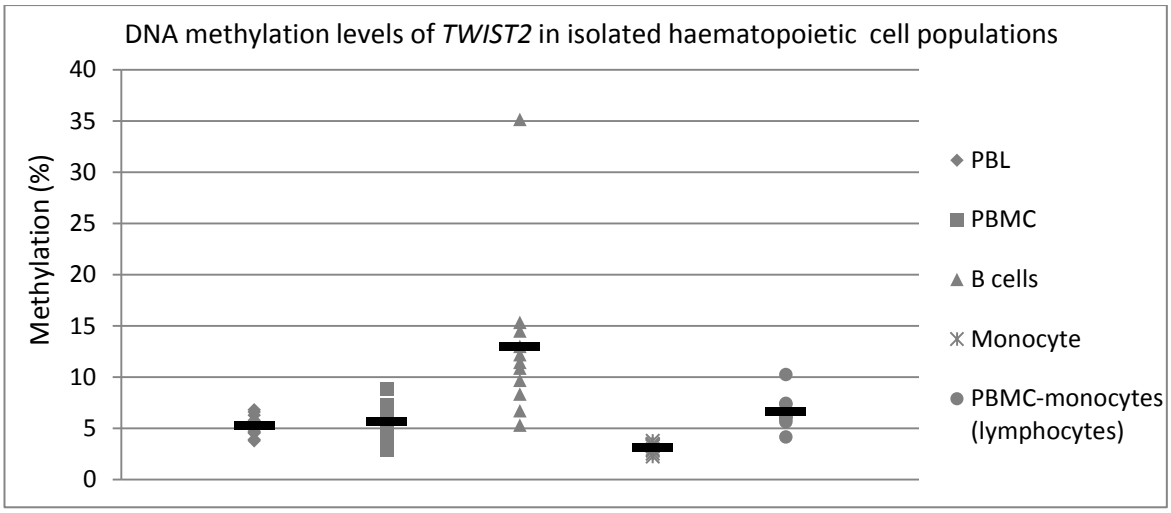


Figure 4.3: DNA methylation levels of *TWIST2* in isolated haematopoietic cell populations

DNA was isolated from PBLs (lymphocytes, neutrophils, and monocytes), PBMCs (lymphocytes and monocytes), B cells and monocytes and DNA methylation levels at the *TWIST2* promoter measured by Pyrosequencing. PBMC-monocytes represent total lymphocytes (T cells, B cells and NK cells). DNA methylation levels are lowest in monocytes and highest in B cells.

As data regarding the characterisation of specific leukocyte populations by flow cytometry was available for all Newcastle 85+ study participants, the association between DNA methylation levels and number of B-cells was investigated to decipher whether the proportion of B-cells within an individual blood sample was responsible for

the observed relationship between lymphocyte percentage and DNA methylation levels. Calculation of the Pearson correlation coefficient indicated no relationship between DNA methylation level and B-cell number in these Newcastle 85+ participants (r=0.07, p=0.12).

Therefore, the DNA methylation levels measured in the Newcastle 85+ study participants were corrected to account for the proportion of lymphocytes in the blood sample. This will ensure that differences between individuals are due to genuine differences in cellular levels of DNA methylation and are not secondary to differences in cell populations.

Firstly, in order to establish a single methylation measurement for each individual, the mean of the DNA methylation levels at *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* would typically be calculated. However, absolute methylation measurements for the five genes are quite different; for example the average amount of methylation in the total population at the *TWIST2* locus is 6.8%, whereas mean methylation of *HOXD4* is 15.3%, over 2 times greater. This would result in particular loci (such as *HOXD4*) having a greater influence on the overall methylation average than other loci. In order to prevent this, methylation measurements across all five loci were adjusted, such that each gene contributes an equal proportion to the overall mean. The average methylation level across five genes in all 226 participants was 10.5%; therefore, the absolute methylation levels of all genes were scaled such that the overall methylation percentage at each locus was 10.5%. The variation between individuals within the population will remain, as the methylation levels at each gene for each individual were corrected by multiplying by a factor, calculated by dividing the total population mean (10.47%) by the specific gene mean.

Next, linear regression analysis was used to examine the relationship between the overall methylation mean, calculated by averaging the adjusted methylation levels at *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* for each individual, and the proportion of lymphocytes in that individual (Figure 4.4). The r^2 value of 0.33 revealed a positive relationship between these two variables. The equation of the regression line is shown below:

$$y = 0.1811x + 5.1549$$

This can be used to determine the magnitude of the effect. If methylation is plotted on the y axis and lymphocytes along the x axis, the following equation states that the slope of the regression line is 0.1811, meaning that every 1% increase in lymphocytes (x), is responsible for a 0.1811% increase in methylation levels (y).

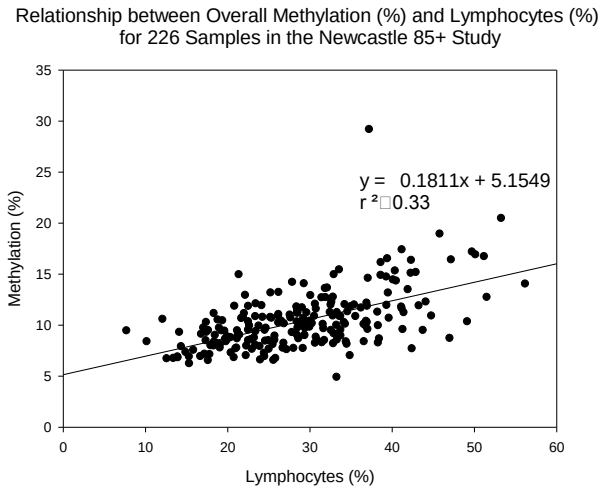


Figure 4.4: Relationship between overall mean DNA methylation levels and percentage of lymphocytes in the blood sample of 226 Newcastle 85+ study participants

Linear regression analysis was performed by plotting the overall mean methylation of each participant against the percentage of lymphocytes. A positive relationship is indicated by the r^2 value and the equation of the regression line is also shown.

In order to remove the influence of lymphocyte proportion on DNA methylation levels, measurements for all individuals were corrected to account for this effect by calculating by the following lymphocyte normalisation factor:

((Sample lymphocytes (%) – population mean lymphocytes (29.02%)) * DNA methylation increase determined by lymphocytes (0.1811%))

This individually calculated factor was then applied to methylation levels at all five genes, and either reduces or increases methylation levels according to whether that individual had a higher or lower than average proportion of lymphocytes. This should result in removing any influence of differing cell populations on methylation levels at

TWIST2, HOXD4, EPHA10, TUSC3 and *HAND2*. All analysis after this point was performed using the corrected methylation values. Changes to the mean, standard deviation and range of methylation measurements for each gene after this correction are shown in Table 4.3. The pairwise strong positive correlation between DNA methylation levels at these genes was still evident after adjusting for lymphocyte proportion indicating that this relationship is not driven by the relationship between increased DNA methylation levels and increased lymphocyte percentage (Table 4.4). No changes were made to the *HOXA5* locus measurements since this DNA methylation at this gene demonstrates no relationship with the proportion of lymphocytes in the same blood sample.

Gene	Methylation Range (%)	Mean Methylation (%)	Standard Deviation (%)
HOXD4	0-27	10.47	2.9
TUSC3	2-22	10.47	3.5
TWIST2	3-37	10.47	4.1
EPHA10	2-20	10.47	2.4
HAND2	3-33	10.47	3.2

Table 4.3: Properties of DNA methylation measurements after correction for proportion of lymphocytes in the blood sample of 226 Newcastle 85+ Study participants

For each gene, the corrected minimum and maximum methylation levels in the samples is shown, alongside the normalised average methylation level calculated as described above, and the standard deviation as a measure of variation.

	TWIST2	*HOXD4*	*EPHA10*	*TUSC3*
TWIST2				
HOXD4	r = 0.61 p < 1e-16			
EPHA10	r = 0.30 p = 5.2e-06	r = 0.49 p = 4.2e-15		
TUSC3	r = 0.64 p < 1e-16	r = 0.56 p < 1e-16	r = 0.47 p = 8.0e-14	
HAND2	r = 0.38 p = 2.7e-09	r = 0.41 p = 1.9e-10	r = 0.29 p = 7.8e-06	r = 0.31 p = 3.1e-06

Table 4.4: Pair-wise correlations of PBL DNA methylation levels at *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* after adjusting for proportion of lymphocytes in 226 Newcastle 85+ Study participants

The Pearson correlation coefficient (r) and corresponding p values are shown for 226 participants. DNA methylation levels at *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* are strongly positively correlated. Black p value = NS, Red p value = significant at <0.005 level after performing Bonferroni correction for multiple comparisons.

4.2.5 DNA methylation levels of Newcastle 85+ study participants with a previous history of cancer compared to those with no cancer

In order to determine whether there was a difference in DNA methylation levels between individuals either with no cancer, or a previous incidence of cancer, the difference in DNA methylation levels between these groups were compared using the Student's t-test. Alongside the 226 participants, the 50 pilot study participants were also included, In total, 254 Newcastle 85+ study participants were analysed, as 6 of the pilot study individuals had a previous cancer history and 16 participants originally selected for the cancer free control group had developed cancer during the duration of the study; therefore the results for these participants were removed.

The results revealed that methylation levels at the *HOXD4, EPHA10, TWIST2, TUSC3* and *HAND2* promoters, as well as the overall mean, were indeed higher in participants with a previous history of cancer compared to those with no cancer, in accordance with our hypothesised relationship (Table 4.5). There was a statistically significant difference

when comparing the overall mean, as well as at two individual loci *TUSC3* and *HOXD4*. At the *HOXA5* promoter, no significant differences were found (Table 4.5).

DNA methylation	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Mean no cancer (%)	141	**9.90**	10.61	**10.14**	10.04	10.37	**10.21**	62.42
Mean Previous cancer (%)	113	**10.78**	10.77	**10.79**	10.48	10.72	**10.71**	60.40
t-test p-value		**0.021***	0.376	**0.041***	0.122	0.158	**0.043***	0.188

Table 4.5: Comparison of PBL DNA methylation levels between individuals with a previous history of cancer compared to those with no cancer (excluding future cancers) in the Newcastle 85+ study

Student's t-test was used to analyse difference in methylation levels between the two groups. Methylation levels of *TUSC3*, *HOXD4* and the overall mean were significantly different between the two groups *($p < 0.05$).

4.2.6 Association between DNA methylation levels with gender, smoking behaviour and levels of folate, vitamin B12 and homocysteine in the Newcastle 85+ Study

The following associations were tested in the total Newcastle 85+ study sample set of 270 individuals, to undercover relationships between DNA methylation and gender, smoking and levels of dietary methyl donors. Several previous publications have suggested such factors may impact upon levels of DNA methylation (see Chapter 1).

4.2.6.1 DNA methylation and gender

Several studies have suggested that DNA methylation differences exist between males and females, however reports are often conflicting and are dependent on the type of assay used for measurements (see section 1.1.9.3). The Student t-test was utilised again to determine the existence if significant differences between the six loci in males and females. As expected due to the greater life expectancy of females, there were more women (n=157) than men (n=113) in the Newcastle 85+ study population. At all the

gene promoters measured, males had higher levels of methylation than females. This was a statistically significant difference at *TUSC3, HAND2, EPHA10* and the overall mean (Table 4.6). The same association was also observed at the *HOXA5* promoter (Table 4.6). These results indicate that males have significantly increased levels of gene specific DNA methylation in the Newcastle 85+ study population.

DNA methylation	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Males (%)	113	**10.97**	10.97	10.82	**10.81**	**11.28**	**10.96**	**64.96**
Females (%)	157	**9.92**	10.76	10.27	**10.02**	**10.08**	**10.21**	**59.79**
t-test p-value		**0.007***	0.35	0.072	**0.030***	**0.0001***	**0.008***	**0.008***

Table 4.6: Comparison of DNA methylation levels between males and females
Student's t-test was used to analyse difference in methylation levels between the sexes. Methylation levels of *TUSC3, HAND2, EPHA10*, the overall mean and *HOXA5* were significantly different between the two groups *($p < 0.05$).

4.2.6.2 DNA methylation and smoking behaviour

Smoking has been linked to aberrant DNA methylation patterns (see section 1.6.3.2.2); therefore we aimed to identify any differences in DNA methylation levels between individuals in the Newcastle 85+ study who were either never smokers, current smokers or former smokers.

DNA methylation levels were not significantly different in current smokers than either former or never smokers at most loci, including the overall mean; although at the *HOXD4* locus, never smokers had the highest level of DNA methylation (Table 4.7). The Student's t-test indicated a statistically significant difference in the *HAND2* methylation levels of former compared to never smokers (Table 4.8). Also, methylation levels of *HOXA5* were significantly higher in current compared to former smokers (Table 4.8).

Overall, the results indicate little evidence for any difference in DNA methylation levels between current, former or never smokers at most loci, although there was a significant difference at two of the loci, therefore the link between smoking and DNA methylation levels of the genes is not clear in the 85+ population.

DNA methylation	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Current Smokers (%)	18	10.74	12.03	10.68	10.64	10.97	11.01	**66.67**
Former Smokers (%)	156	10.37	10.47	10.35	**10.01**	10.64	10.37	**60.26**
Never Smokers (%)	95	10.30	11.23	10.72	**10.87**	10.42	10.70	62.75

Table 4.7: DNA Methylation levels in PBL samples from current, former and never smokers in 269 Newcastle 85+ study participants

DNA methylation levels of *TUSC3*, *TWIST2*, *HOXD4*, *HAND2*, *EPHA10*, the overall mean and *HOXA5* in 269 participants. Information regarding smoking was missing for one sample.

Student t-test	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Current v. Never	0.32	0.28	0.48	0.41	0.19	0.35	0.11
Current v. Former	0.33	0.06	0.32	0.19	0.32	0.18	**0.04***
Former v Never	0.44	0.08	0.18	**0.02***	0.27	0.17	0.14

Table 4.8: Student's t-test for differences between DNA methylation levels in PBL samples from current, former and never smokers in 269 Newcastle 85+ study participants

The Student's t-test was used to analyse difference in DNA methylation levels between current and never smokers, current and former smokers, and former compared to never smokers. There was a statistically significant difference in *HOXA5* methylation between current and former smokers, and *HAND2* methylation levels of former compared to never smokers *($p < 0.05$).

4.2.6.2 DNA methylation and levels of folate, vitamin B12 and homocysteine

Both folate and vitamin B12 are essential for DNA methylation, as they are involved in the synthesis of the methyl donor S-adenosyl methionine (SAM), from methionine. Homocysteine is formed after the transfer of a methyl group and is converted back to methionine through a series of reactions requiring both folate and vitamin B12. A build-up of homocysteine can indicate a lack of reconversion and a reduced level of SAM. Therefore, both folate and vitamin B12 levels have a positive impact on SAM levels, whilst homocysteine is negatively associated. Levels of homocysteine, vitamin B12 and red cell folate were measured in the Newcastle 85+ Study participants (Chapter 2.1). For each individual, a Z-score was calculated for these three measures combined by the following formula:

$$\text{Z-Score} = \frac{(\text{Individual measurement} - \text{Population mean})}{\text{Population SD}}$$

A combined Z-Score was calculated by the following formula, to reflect the conflicting relationship between levels of homocysteine, vitamin B12 and red cell folate:

Combined Z-Score = Z-Score (folate) + Z-score (vit B12) – Z-Score (homocysteine)

The correlation between this combined Z-Score and DNA methylation levels at the six loci was tested by calculating the Pearson correlation coefficient which revealed no evidence of any association between the measures (Table 4.9). Figure 4.5 illustrates the lack of relationship between the overall mean methylation and the combined homocysteine, vitamin B12 and red cell folate Z-score. This implies that the variable DNA methylation levels observed in the Newcastle 85+ study participants do not appear to be significantly affected by differences in levels of folate, vitamin B12 and homocysteine.

	TUSC3	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Combined Z – Score	r = -0.02 p = 0.72	r = 0.03 p =0.58	r = 0.01 p = 0.98	r = 0.09 p = 0.14	r = 0.02 p = 0.78	r = 0.02 p = 0.73	r = 0.01 p = 0.90

Table 4.9: Correlation between DNA methylation levels and combined homocysteine, vitamin B12 and red cell folate Z-score measure in blood samples from the Newcastle 85+ study participants (initial sample set)

The Pearson correlation coefficient (r) and corresponding p values are shown. There was no association between the two measures at any gene ($p<0.05$).

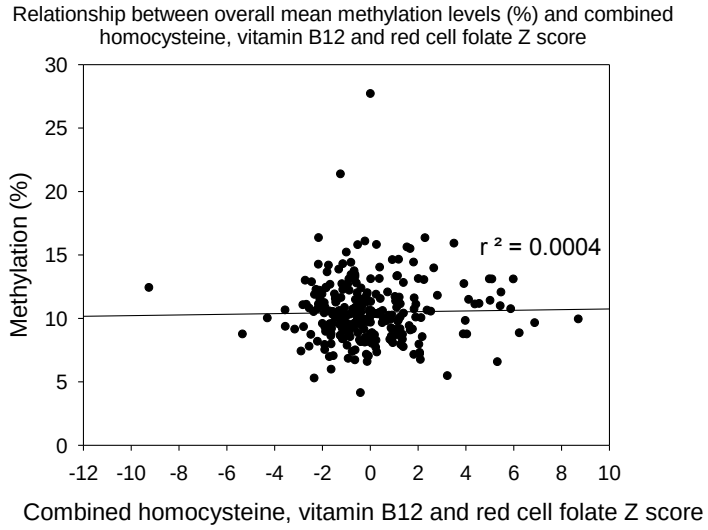


Figure 4.5: Relationship between overall mean DNA methylation levels and combined homocysteine, vitamin B12 and red cell folate Z-score in the Newcastle 85+ study participants (initial sample set)

There was no evidence for any association.

4.2.7 Association between DNA methylation levels and the presence of age-related clinical outcomes in the Newcastle 85+ study population

Next, possible relationships between DNA methylation levels and the presence of age related clinical outcomes were investigated. For comparison, the DNA methylation levels of participants with or without a particular clinical outcome were analysed for the overall mean of *TUSC3, TWIST2, HOXD4, HAND2* and *EPHA10*, since these loci appear to exhibit co-methylation whereas *HOXA5* was analysed separately as it does not fit with the similar pattern of methylation as the other five genes. The comparison of overall DNA methylation levels in individuals with cardiovascular disease (CVD), Type II diabetes (T2D), Dementia/Alzheimer's disease (AD), stroke or who had died within three years can be found in Table 4.10.

Clinical Outcome	CVD	T2D	Dementia (inc. AD)	Stroke	Survival
Mean overall DNA methylation (%)					
No disease/ Alive	10.35	10.55	10.50	10.49	10.46
With disease/ Died	10.87	10.34	10.82	10.70	10.90
t-test p-value	0.057	0.323	0.303	0.309	0.150

Table 4.10: Comparison of PBL DNA methylation levels in association with age-related clinical outcomes in the Newcastle 85+ study participants (initial sample set)

The Student's t-test was used to analyse differences in overall DNA methylation levels between participants with and without cardiovascular disease, type II diabetes, dementia including Alzheimer's disease and stroke and a comparison of participants who had died or were still alive after the study follow up period of three years. There were no significant differences ($p<0.05$).

A number of previous studies had proposed a link between DNA methylation levels of specific genes and CVD, although these studies typically focused on global methylation (Sharma *et al.*, 2008; Kim *et al.*, 2010a), or in mouse models of atherosclerosis (Zaina *et al.*, 2005). Therefore, the potential link between DNA methylation and cardiovascular disease was investigated for gene specific methylation levels in the Newcastle 85+ population. For this analysis, a participant was defined as having CVD if they had any one or more of the following incidents or procedures: myocardial infarction, angina, coronary angioplasty/stent, coronary artery bypass graft.

The results suggested no significant difference in DNA methylation levels between participants who have had incidents or procedures indicative of CVD (Table A3.2), although there was a slight trend towards higher levels of DNA methylation in participants with CVD (Table 4.10). Despite a relatively large number of participants with CVD (n=91), since absolute difference in methylation levels were small,

confirmation of this trend utilising independent samples is required to determine whether this difference in DNA methylation levels is reproducible. In contrast, methylation levels at *HOXA5* were significantly lower in participants with cardiovascular events compared to those with none (CVD=58.43%, none=63.44%, p=0.013) (Table A3.2).

Unlike Type I diabetes which is often diagnosed in childhood, or young adolescence, the prevalence of T2D increases with age and a possible role for differential DNA methylation has been proposed (Groop and Lyssenko, 2008). In this analysis diabetes patients exhibited a trend towards reduced methylation levels of all genes analysed compared to those individuals without a diagnosis of diabetes (as shown in Table A3.3), however these did not approach statistical significance (Table 4.10). Overall these results suggest no clear correlation between type II diabetes and the measured methylation levels, although the comparatively small number of individuals with a diagnosis of type II diabetes would have limited the power of the study to identify small differences.

Dementia, including Alzheimer's disease, is rare in the younger population, and the incidence increases with age, particularly over the age of 75 years (Matthews and Brayne, 2005). From this analysis, there was no evidence of a significant difference between the overall mean DNA methylation levels of participants with and without dementia (Table 4.10), although there was a trend towards higher methylation of *HOXA5* in participants with dementia (Table A3.4). Interestingly, individuals with dementia appeared to have significantly higher methylation of *EPHA10* than those without the disease (Dementia=12.01%, none=10.48%, p=0.009) (Table A3.4), but the reliability of this observation at a single gene in a relatively small number of participants is unclear, and would need to be replicated in an independent set of samples.

The risk of stroke is significantly increased in an elderly population, with a lifetime risk of suffering stroke estimated to be around 21% after age 55 (Hollander *et al.*, 2003). Although a tendency towards higher DNA methylation levels at several loci was identified in participants with stroke (Table A3.5), the T-test analysis showed no significant difference between the overall mean DNA methylation levels at any gene

(Table 4.10), suggesting that altered DNA methylation in PBL DNA is not a feature of participants with an incidence of stroke.

Finally, since the risk of dying inevitably increases with advancing age, we examined whether there was any differences in DNA methylation measurements between participants who had died prior to phase III of the study (36 months later), or those who were still alive. A trend towards higher DNA methylation levels was observed at most loci in participants who did not survive to phase III (Table A3.6) however this was not a significant relationship at the overall mean (Table 4.10). Again this suggests that DNA methylation levels are not significantly different in participants who died within three years of sample donation compared to those individuals who were still alive.

Performing a Kaplan-Meier survival analysis could have been used as an alternative method for analysing the relationship between DNA methylation levels and survival. Participants could be grouped based on their overall methylation level (high or low) and combined with survival data, to uncover any significant differences between those participants with high DNA methylation compared to those with low levels of DNA methylation in terms of survival over 36 months. This may possibly be a more powerful analysis as it may identify significant differences in the length of time survived between individuals with differential DNA methylation levels.

4.3 Validation analysis

To further test our hypothesis that individuals with higher levels of methylation may be more susceptible to cancer, participants who developed cancer during the study period, after DNA samples were taken at phase I, were selected for analysis of their DNA methylation levels. In total 72 participants were diagnosed with cancer within the 3 years after sample collection (future cancer). Nineteen of these individuals also had a previous history of cancer so were included in the initial comparison of methylation levels in participants who had a previous cancer. Seventeen individuals were unknowing tested as part of the initial control group but were excluded from analysis. A total of 36 participants with a future cancer were tested.

A number of factors were identified as demonstrating potential associations with methylation levels in the initial analysis of 270 Newcastle 85+ study participants including gender, smoking behaviour, cardiovascular events, type II diabetes, dementia, stroke and survival. Such associations may have arisen by chance due to multiple testing, so an additional group of Newcastle 85+ study participants were tested as a validation sample set, in an attempt to confirm the findings using independent group. Often, the trends observed were not statistically significant so testing additional samples may increase the strength of the association.

The validation sample set was comprised of all 24 current smokers, 36 dementia patients, and 57 participants diagnosed with Type II diabetes. This represents a total of 103 unique participants due to an overlap of participants in more than one group, for example individuals with both Type II diabetes and dementia. In addition, the 36 participants with future cancer were included alongside a control group comprised of 71 participants randomly selected from all remaining participants. In total, the number of DNA samples analysed in the validation sample set was 210.

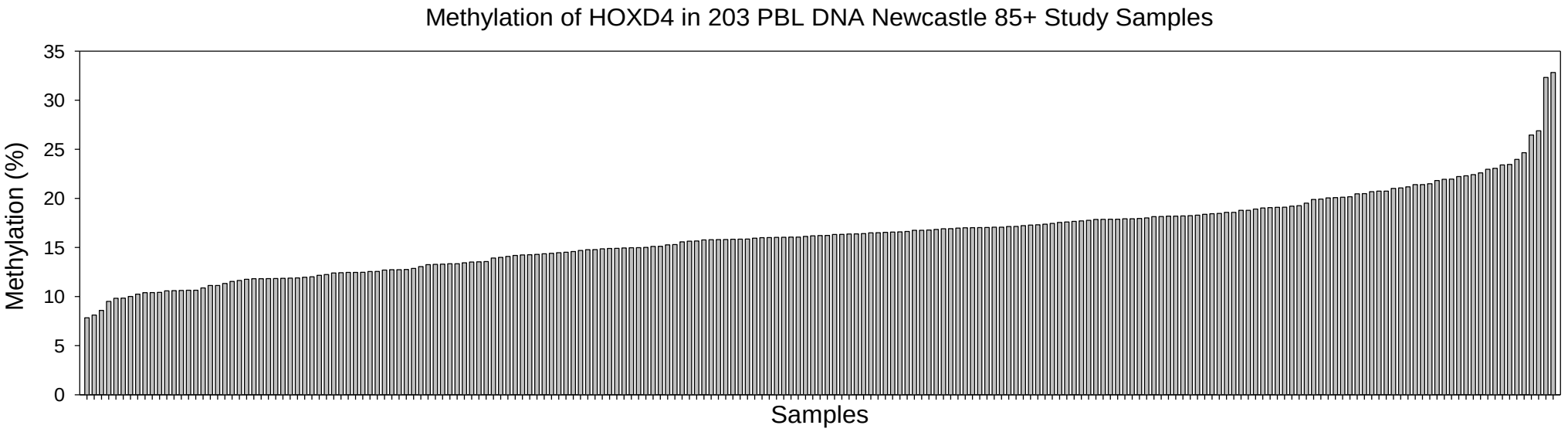
4.3.1 Pyrosequencing results – validation analysis

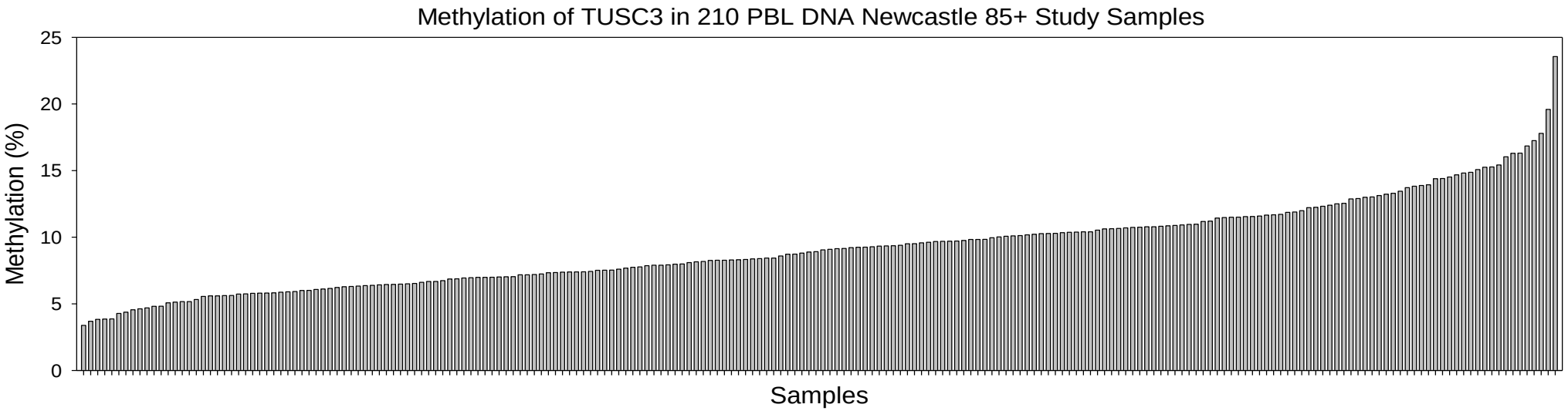
DNA methylation levels were once more measured at the six same genes; *HOXD4, TUSC3, TWIST2, HAND2*, *EPHA10* and *HOXA5*. For the 210 participants in the validation set, the results again demonstrated highly variable DNA methylation levels at all loci and the range of DNA methylation measurements, mean and standard deviation was very similar at all gene promoters to those observed in the previous 226 participants used for the initial analysis (Table 4.11 and Figure 4.6).

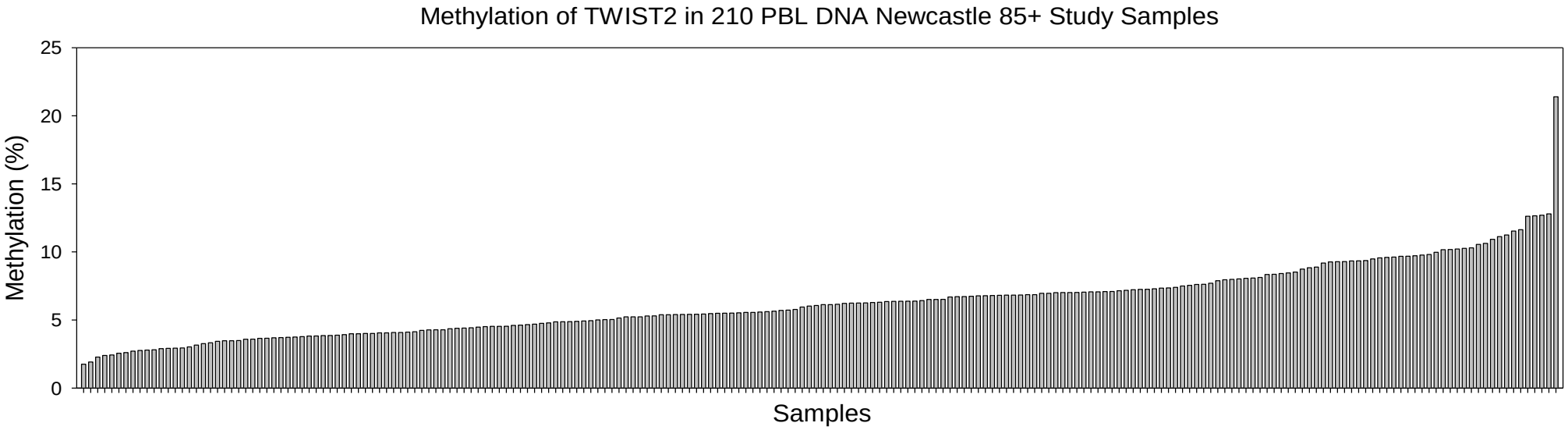
Gene	Methylation Range (%)	Mean Methylation (%)	Standard Deviation (%)
HOXA5	30-95	62.8	13.5
HOXD4	8-33	16.3	4.0
TUSC3	3-24	9.3	3.3
TWIST2	2-21	6.3	2.6
EPHA10	6-24	12.3	3.3
HAND2	4-17	9.0	2.1

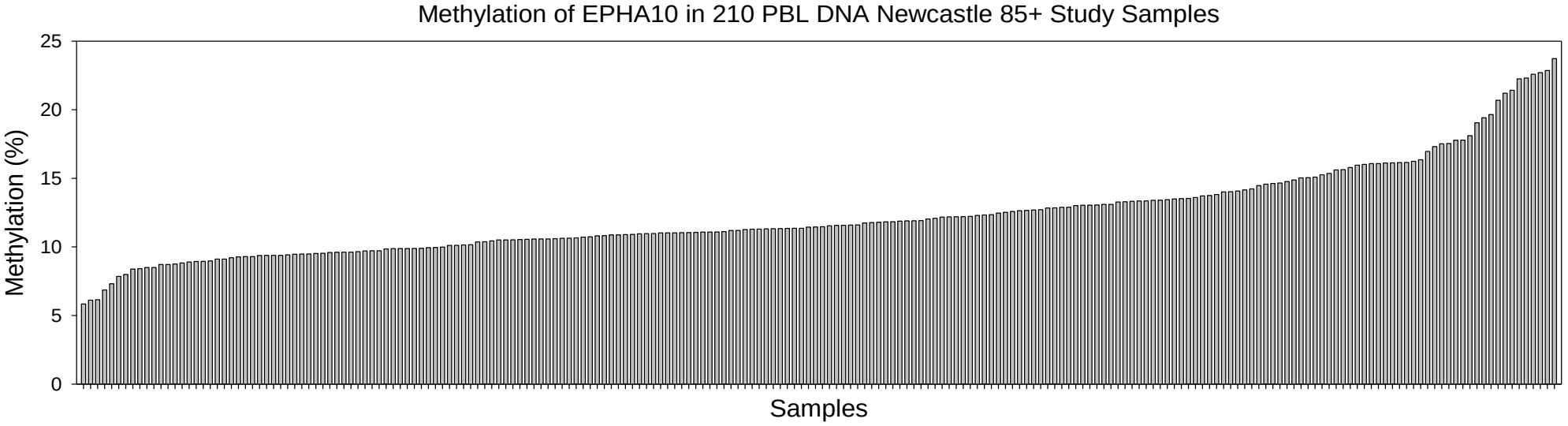
Table 4.11: Properties of PBL DNA methylation measurements in the validation sample set of 210 participants in the Newcastle 85+ study

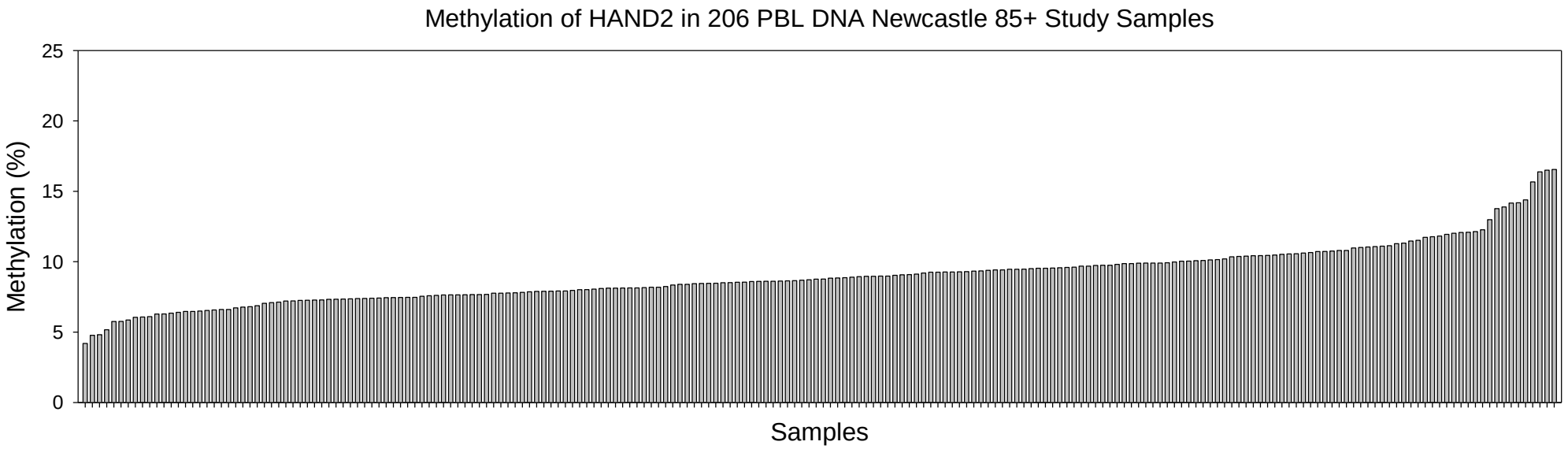
For each gene, the minimum and maximum methylation level in the DNA samples is shown for all participants, together with the mean methylation calculated from 3-6 consecutive CpG sites and the standard deviation as a measure of variation.


Methylation of HOXD4 in 203 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
30
35
Samples


Methylation of TUSC3 in 210 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
Samples


Methylation of TWIST2 in 210 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
Samples


Methylation of EPHA10 in 210 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
0
5
10
15
20
25
Samples


Methylation of HAND2 in 206 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
25
20
15
10
5
0
Samples

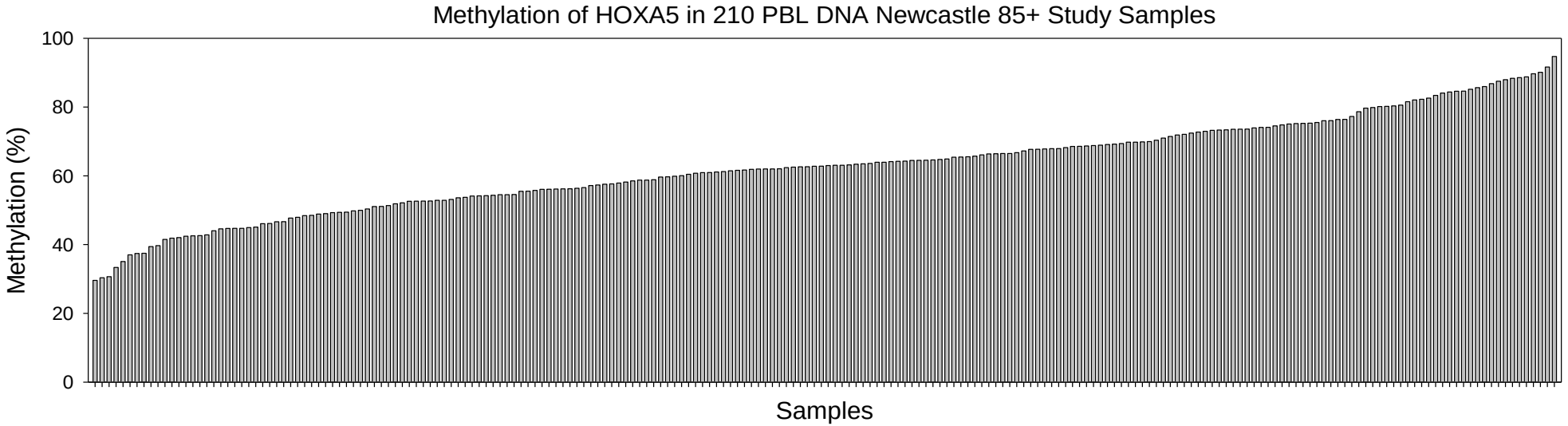

Methylation of HOXA5 in 210 PBL DNA Newcastle 85+ Study Samples
Methylation (%)
100
80
60
40
20
0
Samples

Figure 4.6: PBL DNA methylation levels of *HOXD4, TUSC3, TWIST2, EPHA10, HAND2* and *HOXA5* in the validation sample set of 210 participants from the Newcastle 85+ Study

Samples are shown in ascending order of DNA methylation measurements at each particular gene. All genes exhibit highly variable DNA methylation levels between individuals. Seven results from *HOXD4* and four results at the *HAND2* promoter were excluded due to a lack of reproducible methylation measurements.

4.3.2 Pairwise correlations of DNA methylation measurements, correlation with lymphocyte percentage and correction of measurements to account for lymphocyte percentage – validation analysis

As in all previous analysis, calculation of the Pearson correlation coefficient and corresponding p value for each pairwise comparison of DNA methylation levels revealed strong positive correlations between methylation levels of *HOXD4, TUSC3, TWIST2, HAND2* and *EPHA10* in the 210 participants, which were highly significant (Table A3.7). Similarly, there was no association between methylation levels of HOXA5 and those at any other gene.

Additionally, the calculated Pearson correlation coefficient again indicated a highly significant relationship between the proportion of lymphocytes and DNA methylation levels at *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2,* but not *HOXA5* in 210 participants (Figure A3.8 and Table A3.9).

Once more, DNA methylation measurements for each individual across all five loci were adjusted so that each gene contributes an equal proportion to the overall mean. The average methylation level across five genes in the validation set of 210 participants was 10.63%. Linear regression analysis indicated that every 1% increase in lymphocytes was responsible for a 0.1947% increase in methylation levels in this validation sample set (Figure A3.10) and a specific lymphocyte normalisation factor (Figure A3.11) was applied to the DNA measurements of *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* for each individual participant, to account for the contribution of lymphocytes to DNA methylation levels. All analysis performed after this point used the lymphocyte corrected DNA methylation values. Changes to the mean, standard deviation and range of methylation measurements for each gene after this correction can be found in Table

A3.12. As for previous participants, *HOXA5* results remained unchanged as DNA methylation measurements at this gene are not associated with lymphocytes percentage.

4.3.3 DNA methylation levels of Newcastle 85+ study participants diagnosed with cancer throughout the three year study duration, compared to those with no cancer

The participants who had a previous incidence of cancer had higher levels of DNA methylation compared to cancer-free individuals, which was statistically significant at several loci including *TUSC3* and *HOXD4*, as well as the overall sample mean. Next, DNA methylation levels measurements from the 72 participants who developed cancer during the 3 year follow up of the study (future cancer cases) were compared to the 71 control group participants in the validation sample set, in addition to the 141 cancer free participants from the first analysis, using the Student's t-test. DNA samples were taken prior to cancer diagnosis.

The results revealed that methylation levels at individual loci, *TUSC3*, *TWIST2*, *HOXD4*, *EPHA10*, *HAND2* and the overall mean DNA methylation were higher in those participants with a subsequent diagnosis of cancer compared to those who remained cancer free, reflecting the trend observed in the initial analysis comparing previous cancer incidence this no cancer (Table 4.12).

The differences were statistically significant when analysing the overall mean DNA methylation at these genes, as well as at *HAND2* and *EPHA10* individually. Similarly, *HOXA5* promoter methylation was also increased in participants with a future diagnosis of cancer (Table 4.12). These results are intriguing, as it suggests that DNA methylation increases may be evident before the onset of clinically detectable disease in the Newcastle 85+ study participants.

DNA methylation	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
No cancer (%)	141	**9.90**	10.61	**10.14**	10.04	10.37	**10.21**	62.42
Previous cancer (%)	113	**10.78**	10.77	**10.79**	10.48	10.72	**10.71**	60.40
t-test p-value		**0.021***	0.376	**0.041***	0.122	0.158	**0.043***	0.188
No cancer (%)	212	10.20	10.75	10.29	**10.23**	**10.32**	**10.36**	62.02
Future cancer (%)	72	10.82	10.99	10.75	**11.18**	**11.11**	**10.96**	64.19
t-test p-value		0.076	0.337	0.128	**0.016***	**0.014***	**0.034***	0.152

Table 4.12: PBL DNA methylation levels of participants with a previous or subsequent diagnosis of cancer, compared to cancer-free individuals in the Newcastle 85+ study.

Student's t-test was used to analyse difference in mean methylation levels between the two groups. In the first analysis, *TUSC3, HOXD4* and the overall mean were more highly methylated in participants with a previous cancer incidence. In the validation sample set, there was a statistical significance in *HAND2, EPHA10* and the overall mean methylation level (*$p < 0.05$).

4.3.3.1 Participants with subsequent diagnosis of haematological cancer had highly increased DNA methylation levels

Only 2 of the total 72 participants who were subsequently diagnosed with any cancer during the study had a haematological cancer. Participant NE00000472 was diagnosed with lymphoma and participant NE00001741 was diagnosed with leukaemia approximately one year and two and a half years after DNA samples were obtained, respectively. Participant NE00000472 had the highest overall DNA methylation (average of *TUSC3, TWIST2, HOXD4, EPHA10* and *HAND2)* and participant NE00001741 had the third highest overall DNA methylation of all 480 participants

analysed. Although this is based on a very small number of individuals, it suggests that high methylation levels of these specific gene loci in PBL DNA samples could predict individuals at risk from a haematological based cancer before the onset of clinically detectable disease. Methylation levels of all six loci are detailed in Table 4.13, alongside the population average measurements.

However, further analysis showed that these two participants had two of the three highest measurements of B-cell number measured by flow cytometry within the total Newcastle 85+ study population. Although at a population level in the Newcastle 85+ study, DNA methylation levels were not associated with B-cell number, previous analysis has suggested that B-cells harbour particularly high levels of methylation compared to other haematopoietic cell populations (see chapter 4.2.4). This means that the high methylation levels in these two individuals may simply be reflecting an abnormally increased level of B-cells.

Sample	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Mean (%) (n=480)	10.4	8.7	13.1	9.5	11.2	10.6	62.5
NE00000472 (rank)	22.6 (2)	25.4 (2)	41.1 (1)	27.5 (1)	23.6 (1)	28.0 (1)	-
NE00001741 (rank)	20.0 (8)	18.2 (12)	19.3 (40)	18.7 (5)	19.1 (9)	19.1 (3)	88.8 (6)

Table 4.13: Methylation levels measured in PBL DNA samples from two participants with subsequent diagnosis of haematological cancer in the Newcastle 85+ study

At each gene, the methylation measurements for participants NE00000472 and NE00001741 alongside the Newcastle 85+ study population average measurement. At each locus, the rank position of the participant when all 480 participants were sorted based on highest to lowest DNA methylation level is shown in brackets after the measurement. *HOXA5* methylation was not measured for individual NE00000472.

4.3.4 Association between DNA methylation levels with gender, smoking behaviour and levels of folate, vitamin B12 and homocysteine in the Newcastle 85+ study – validation analysis

The same associations made in the first analysis were re-analysed using the validation set of 210 participants, in an attempt to confirm any significant findings in an independent sample set.

4.3.4.1 DNA methylation and gender – validation analysis

Previous analysis indicated a significant difference in methylation levels at several loci such that males had higher levels of DNA methylation than females. The validation analysis confirmed this trend at most loci, however only the increased methylation levels of males at *TUSC3* and the *HOXD4* loci reached statistical significance (Table 4.14). When examining the overall mean, the difference between genders was close to statistical significance (p=0.09). These results combined with the first analysis add support to previous observations of higher levels of gene specific DNA methylation in males.

DNA methylation	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Initial sample set								
Males (%)	113	**10.97**	10.97	10.82	**10.81**	**11.28**	**10.96**	**64.96**
Females (%)	157	**9.92**	10.76	10.27	**10.02**	**10.08**	**10.21**	**59.79**
t-test p-value		**0.007***	0.35	0.072	**0.030***	**0.0001***	**0.008***	**0.008***
Validation sample set								
Males (%)	85	**11.06**	10.64	**11.18**	10.91	10.64	10.88	63.02
Females (%)	125	**10.34**	10.62	**10.30**	10.48	10.63	10.49	62.64
t-test p-value		**0.044***	0.480	**0.003***	0.110	0.491	0.094	0.420

Table 4.14: Comparison of methylation levels in PBL DNA samples between male and female participants in the initial and validation sample sets in the Newcastle 85+ study

Student's t-test was used to analyse difference in methylation levels between the sexes. In the validation analysis, methylation levels of *TUSC3* and *HOXD4* were significantly different between the two groups *($p < 0.05$). The overall mean methylation level failed to reach significance.

4.3.4.2 DNA methylation and smoking behaviour – validation analysis

In the validation analysis, DNA methylation levels of current smokers were lower than both former and never smokers at many loci, which is the opposite trend to that in the first analysis (Table 4.15). Although several significant difference between current, former and never smokers were implied, the majority of these were not evident in the first analysis and so are unlikely to represent real associations.

However, both *HAND2* and *HOXA5* demonstrated the same relationship in both the previous and validation analysis suggesting these loci could be worthy for further investigation of their altered methylation levels in smokers. The Student's t-test analysis showed significantly increased methylation of current smokers compared to former smokers at *HOXA5*, and reduced methylation at the *HAND2* locus in both current and former smokers compared never smokers (Table 4.16).

DNA methylation	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Initial sample set								
Current Smokers (%)	18	10.74	12.03	10.68	10.64	10.97	11.01	**66.67**
Former Smokers (%)	156	10.37	10.47	10.35	**10.01**	10.64	10.37	**60.26**
Never Smokers (%)	95	10.30	11.23	10.72	**10.87**	10.42	10.70	62.75
Validation sample set								
Current Smokers (%)	24	**9.64**	**9.44**	10.29	**10.12**	10.53	9.99	**67.13**
Former Smokers (%)	122	**10.82**	10.60	**10.95**	**10.47**	10.68	10.71	**62.21**
Never Smokers (%)	64	10.64	**11.14**	**10.19**	**11.12**	10.57	10.77	62.28

Table 4.15: Methylation levels in PBL DNA from current, former and never smokers in the initial and validation sample sets of Newcastle 85+ Study participants

Methylation levels of *TUSC3*, *TWIST2*, *HOXD4*, *HAND2*, *EPHA10*, the overall mean and *HOXA5*. There was a lack of consistent trend between participants in the initial and validation sample sets.

Student's t-test	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Initial sample set							
Current v. Never	0.32	0.28	0.48	0.41	0.19	0.35	0.11
Current v. Former	0.33	0.06	0.32	0.19	0.32	0.18	**0.04***
Former v Never	0.44	0.08	0.18	**0.02***	0.27	0.17	0.14
Validation sample set							
Current v. Never	0.084	**0.027***	0.423	**0.041***	0.471	0.062	0.078
Current v. Former	**0.039***	0.079	0.104	0.263	0.403	0.060	**0.046***
Former v Never	0.347	0.179	**0.020***	**0.025***	0.390	0.424	0.486

Table 4.16: Significance of Student's t-test for differences between PBL DNA methylation levels of current, former and never smokers in the initial and validation set of Newcastle 85+ study participants

Student's t-test was used to analyse difference in DNA methylation levels between current and never smokers, current and former smokers, and former versus never smokers. Significant associations *($p < 0.05$).

4.3.4.3 DNA methylation and levels of folate, vitamin B12 and homocysteine – validation analysis

As for the previous analysis, individual Z-scores were calculated for each of participants in the validation set of 210 participants to examine the relationship between levels of homocysteine, vitamin B12 and red cell folate. Initial analysis revealed no relationship between levels of folate, vitamin B12 and homocysteine and DNA methylation. The Pearson correlation coefficient was calculated for the validation sample set and there was a weak but statistically significant positive relationship between the two variables at the *HAND2* locus, suggesting a higher folate, vitamin B12 and homocysteine score may be weakly associated with increased methylation at this gene (Figure 4.7), however this is likely to be coincidental, since no relationship was observed in the first analysis, and none of the other genes show a similar association (Table 4.17).

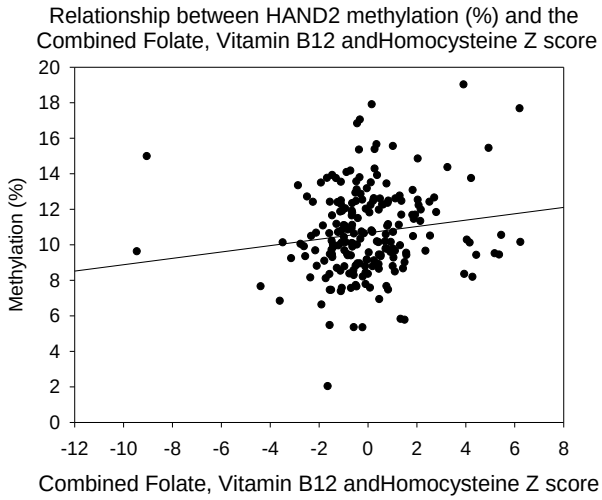


Figure 4.7: Relationship between *HAND2* PBL DNA methylation levels and combined homocysteine, vitamin B12 and red cell folate Z-score in the validation sample set of the Newcastle 85+ study participants

There is a weak but statistically significant positive correlation between the two variables.

	TUSC3	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Initial sample set							
Combined Z – Score	r = -0.02 p = 0.72	r = 0.03 p =0.58	r = 0.01 p = 0.98	r = 0.09 p = 0.14	r = 0.02 p = 0.78	r = 0.02 p = 0.73	r = 0.01 p = 0.90
Validation sample set							
Combined Z – Score	r = -0.10 p = 0.16	r = -0.06 p =0.38	r = -0.12 p = 0.10	**r = 0.15** **p = 0.03***	r = 0.08 p = 0.24	r = -0.02 p = 0.78	r = -0.05 p = 0.52

Table 4.17: Relationship between PBL DNA methylation levels and combined homocysteine, vitamin B12 and red cell folate Z-score in the initial and validation sample sets of the Newcastle 85+ study participants

The Pearson correlation coefficient (r) and corresponding p values are shown. There were no associations in the first analysis but a weak positive association between the combined homocysteine, vitamin B12 and red cell folate Z-score and methylation of *HAND2* in the validation sample set, which was statistically significant *($p<0.05$).

4.3.5 Validation of observed relationships between DNA methylation measurements and clinical outcomes measured in the Newcastle 85+ Study

Once again, overall DNA methylation levels in individuals with CVD, T2D, Dementia/Alzheimer's disease, stroke or participants who had died within three years was compared to disease free individuals (Table 4.18).

The validation analysis revealed no significant differences in DNA methylation measurements between overall mean DNA methylation levels of participants with CVD compared to those with no disease. This was unlike the initial analysis which suggested increased overall DNA methylation in participants with CVD compared to those without, which was close to statistical significance (p=0.057) (Table 4.18). At several individual loci, including *HOXA5*, the trends observed in the first analysis were reversed in validation analysis (Table A3.13), implying that differences in gene specific DNA methylation patterns cannot separate individuals with and without CVD.

The DNA methylation levels of 58 pre-selected participants with Type II diabetes (T2D) were compared to the control group of 71 in the validation sample set. In the first analysis, a trend for lower levels of DNA methylation was observed in participants diagnosed with Type II diabetes compared to those who did not have the disease, however this association was reversed in the validation sample set (Table 4.18) DNA methylation levels at all six gene loci were either no different, or higher in participants with Type II diabetes than those without the disease (Table A3.14), implying no clear relationships between Type II diabetes and levels of DNA methylation.

For the validation analysis, DNA methylation levels of the 36 pre-selected participants with dementia or Alzheimer's disease were compared to the control group of 71 participants. In the first analysis, there was no evidence for differential methylation in participants with and without dementia with the exception of significantly increased *EPHA10* promoter methylation in dementia. In this analysis, there was little difference in overall mean DNA methylation levels in participants with and without dementia (Table 4.18), however, the significantly higher methylation of *EPHA10* in participants with dementia was replicated (dementia = 11.13%, none = 10.23, p=0.043) (Table A3.15). These results suggest that although overall patterns of methylation are not

affected, *EPHA10* demonstrates consistently higher methylation in PBL DNA samples from participants with dementia.

Clinical Outcome	CVD	T2D	Dementia (inc. AD)	Stroke	Survival
Mean overall DNA methylation (%)					
Initial sample set					
No disease/ Alive	10.35	10.55	10.50	10.49	10.46
With disease/ Died	10.87	10.34	10.85	10.70	10.90
t-test p-value	0.057	0.323	0.301	0.309	0.150
Validation sample set					
No disease/ Alive	10.65	10.65	10.68	10.67	10.66
With disease/ Died	10.64	10.93	10.46	10.46	10.62
t-test p-value	0.479	0.240	0.305	0.318	0.456

Table 4.18: Comparison of methylation levels in PBL DNA samples in association with age-related clinical outcomes in the initial and validation set of Newcastle 85+ study participants

The Student's t-test was used to analyse differences in overall DNA methylation levels between participants with and without cardiovascular disease, type II diabetes, dementia including Alzheimer's disease, stroke and survival. There were no significant differences in overall DNA methylation in any analysis of either the initial or validation set of samples ($p<0.05$).

Despite initial results indicating a tendency towards higher methylation levels in participants with stroke, subsequent validation analysis indicated no significant differences of the overall mean of DNA methylation between participants (Table 4.18).

At most gene loci, DNA methylation levels were either lower or not different in participants with stroke (Table A3.16), again suggesting that DNA methylation patterns are not altered by stroke in this population.

Initial indications suggested that the methylation levels at the majority of gene loci were higher in participants who did not survive to phase III compared to survivors, although this was not a significant association. In the validation analysis however, there was no difference between the overall mean DNA methylation level of survivors or non-survivors (Table 4.18). Mean DNA methylation levels of non-survivors in the validation sample set were significantly higher than survivors at *HOXA5* (non-survivors = 66.66%, survivors = 61.88%, p=0.022) (Table A3.17), although there was only a small methylation difference at this gene in the initial analysis which was not statistically significant (non-survivors = 62.63%, survivors = 61.68%, p=0.37).

4.4 Construction of a linear regression analysis model

For all the analysis described above, all associations were tested individually. Performing the analysis in this manner was useful for generating an initial hypothesis of associations which may be related to DNA methylation levels. The reproducibility of these associations was tested in a separate set of participants from the 85+ study. One issue caused by examining a number of associations individually is that of multiple testing, which increases the likelihood of identifying a seemingly significant association purely by chance. In addition, individual testing fails to account for the fact that associations may be dependent on another factor in the analysis.

In order to perform a more robust analysis, a linear regression analysis model was constructed using non-lymphocyte corrected data from all 480 participants tested, to identify any variable which had a significant effect on DNA methylation levels at *HOXD4, TUSC3, TWIST2, HAND2, EPHA10,* the overall mean of these 5 genes and *HOXA5*. Variables included lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking behaviour, CVD, T2D, dementia, stroke and survival, which were built into a step-down model. Non-significant associations were removed (exclusion criteria $p>0.10$) and only the effects of significant associations on methylation levels were retained (retention criteria $p<0.05$). This

removes seemingly significant associations which are likely caused by additional factors.

Each individual locus was analysed in turn as the dependent variable, whilst all other variables were input as independent variables. Many associations which were significant in individual analysis, such as the increased methylation in participants with a previous incidence of cancer, were no longer significantly different when all variables were built into the step-down model. The percentage of lymphocytes had a clearly significant effect on methylation levels ($p<0.000$) at *TUSC3* (Table 4.19), *TWIST2* (Table 4.20), *HOXD4* (Table 4.21), *HAND2* (Table 4.22), *EPHA10* (Table 4.23) and the overall mean of these five loci (Table 4.24), which was previously recognised in the correlation analysis and methylation levels were corrected in an attempt to account for this. In addition, gender had a significant influence on methylation levels at *TUSC3* ($p=0.001$, Table 4.19), *HOXD4* ($p=0.02$, Table 4.21) the overall mean ($p=0.006$, Table 4.24), and a borderline association at *EPHA10* ($p=0.057$, Table 4.23). At the *HAND2* locus, other associations which remained significant in the model included increased methylation in participants with a future cancer diagnosis ($p=0.033$) and the presence of reduced methylation levels in current or former smokers ($p=0.005$). An association with stroke at this gene did not reach statistical significance, but was just outside of the exclusion criteria ($p=0.092$) (Table 4.22). Increased methylation of *EPHA10* in participants with dementia retained significance ($p=0.024$), although the increased levels of methylation associated with a future cancer diagnosis at this gene demonstrated borderline statistical significance ($p=0.088$) (Table 4.23). The model revealed no variables with a significant effect on levels of *HOXA5* methylation (Table 4.25).

These results indicate that affect of many associations which were significantly associated with differential DNA methylation levels when analysed alone, such as previous or future incidence of cancer, gender and smoking behaviour, may be mediated by the presence of other factors within the population, since they lose significance in a combined analysis.

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
9	(Constant)	3.901	.524		7.447	.000
	Lymphocytes_Percent	.250	.017	.563	14.440	.000
	Gender	-1.071	.307	-.136	-3.490	.001

a. Dependent Variable: TUSC3

Table 4.19: Significant associations with *TUSC3* methylation levels after construction of a step-down linear regression model

All 480 participants were included in this analysis. Independent variables built into the model include lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking behaviour, history of cardiovascular events, type II diabetes, dementia, stroke and survival. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
10	(Constant)	2.595	.629		4.127	.000
	Lymphocytes_Percent	.273	.021	.518	13.082	.000

a. Dependent Variable: TWIST2

Table 4.20: Significant associations with *TWIST2* methylation levels after construction of a step-down linear regression model

Independent variables built into the model include lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking behaviour, history of cardiovascular events, type II diabetes, dementia, stroke and survival. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
9	(Constant)	5.923	.445		13.321	.000
	Lymphocytes_Percent	.172	.015	.485	11.667	.000
	Gender	-.610	.261	-.097	-2.335	.020

a. Dependent Variable: HOXD4

Table 4.21: Significant associations with *HOXD4* methylation levels after construction of a step-down linear regression model

All 480 participants were included in this analysis. Independent variables built into the model include lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking behaviour, history of cardiovascular events, type II diabetes, dementia, stroke and survival. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
7	(Constant)	7.524	.518		14.529	.000
	Lymphocytes_Percent	.114	.016	.320	7.254	.000
	Smoking	-.375	.132	-.126	-2.849	.005
	Future_Cancer	.812	.378	.095	2.144	.033
	Stroke	.658	.390	.074	1.688	.092

a. Dependent Variable: HAND2

Table 4.22: Significant associations with *HAND2* methylation levels after construction of a step-down linear regression model

All 480 participants were included in this analysis. Independent variables built into the model include lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking behaviour, history of cardiovascular events, type II diabetes, dementia, stroke and survival. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
7	(Constant)	6.912	.434		15.920	.000
	Lymphocytes_Percent	.128	.014	.396	9.130	.000
	Dementia	.883	.390	.098	2.268	.024
	Gender	-.478	.251	-.084	-1.905	.057
	Future_Cancer	.584	.342	.074	1.709	.088

a. Dependent Variable: EPHA10

Table 4.23: Significant associations with *EPHA10* methylation levels after construction of a step-down linear regression model

All 480 participants were included in this analysis. Independent variables built into the model include lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking behaviour, history of cardiovascular events, type II diabetes, dementia, stroke and survival. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
9	(Constant)	5.369	.374		14.361	.000
	Lymphocytes_Percent	.190	.012	.589	15.398	.000
	Gender	-.602	.219	-.105	-2.748	.006

a. Dependent Variable: OVERALL_MEAN

Table 4.24: Significant associations with overall mean methylation levels after construction of a step-down linear regression model

All 480 participants were included in this analysis. Independent variables built into the model include lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking behaviour, history of cardiovascular events, type II diabetes, dementia, stroke and survival. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

Excluded Variables[k]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics Tolerance
11	Type2Diabetes	.014	.268	.789	.014	1.000
	Dementia	.031	.590	.556	.031	1.000
	Survival	.040	.762	.447	.040	1.000
	Cardiovascular	-.044	-.835	.404	-.044	1.000
	Stroke	.045	.846	.398	.045	1.000
	Future_Cancer	.067	1.278	.202	.067	1.000
	Lymphocytes_Percent	.038	.728	.467	.038	1.000
	Gender	-.060	-1.146	.252	-.060	1.000
	Smoking	.068	1.294	.196	.068	1.000
	Previous_Cancer	-.076	-1.453	.147	-.076	1.000

k. Dependent Variable: HOXA5

Table 4.25: No significant associations with *HOXA5* methylation levels after construction of a step-down linear regression model

All 480 participants were included in this analysis. Independent variables built into the model include lymphocyte percentage, previous history of cancer, subsequent diagnosis of cancer, gender, smoking behaviour, history of cardiovascular events, type II diabetes, dementia, stroke and survival. There were no significant associations ($p<0.05$), so all variables were excluded from the model.

4.5 Discussion

In total, 436 Newcastle 85+ study participants were tested in this part of the analysis, but results were combined with those from the initial pilot study creating a total of 480 individuals. In the initial analysis, 226 DNA were from participants with and without a previous history of cancer were compared and also utilised for generating hypotheses of associations of DNA methylation levels with other diseases. The remaining 210 participants formed a validation sample set and were comprised of individuals with a future diagnosis of cancer, individuals with specific age-related diseases and a disease free control group to enable confirmation of associations from the first analysis.

The Pyrosequencing results for this expanded set of Newcastle 85+ study participants revealed that DNA methylation levels at *HOXD4, TUSC3, TWIST2, HAND2, EPHA10* and *HOXA5* were highly variable within the 85 year old population, comparable to the observations from the pilot study. Furthermore, DNA methylation levels at the *HOXD4, TUSC3, TWIST2, HAND2* and *EPHA10* promoter regions demonstrated strong positive correlations with each other, with exceptionally high statistical significance. Again this co-methylation of genes suggests certain individuals are more susceptible to increased methylation at numerous gene promoters throughout the genome. An alternative explanation for the strong correlation between DNA methylation levels of multiple promoters may be reflecting a shifting proportion of a specific lymphocyte cell sub fraction in certain individuals at an elderly age. This seems unlikely however, given the magnitude of DNA methylation levels at some of these promoters which means that the proportion of such a specific leukocyte subtype would have to become significantly larger in order to contribute towards such an increase in DNA methylation levels.

Additional analysis of correlation revealed that the proportion of lymphocytes within a particular individual has a small but highly significant effect on the levels of DNA methylation at the *HOXD4, TUSC3, TWIST2, HAND2* and *EPHA10* promoters. The strength of these relationships was significantly greater than in the pilot study, possibly due to an increased sample size. In order to avoid the possibility of identifying methylation differences between individuals which were solely reflecting differences in haematopoietic cell populations, methylation levels at all five genes for all participants were corrected to account for the effect caused by differing lymphocyte percentages between individuals

The main hypothesis under investigation this part of the study was that pre-existing methylation may underlie susceptibility to disease, particularly cancer where aberrant methylation patterns are frequently observed. Therefore, it was expected that an individual with either a previous history of any cancer, or a subsequent diagnosis of cancer within the Newcastle 85+ study duration would have higher levels of DNA methylation than participants who were free of cancer, This may be particularly at the five loci, *HOXD4, TUSC3, TWIST2, HAND2* and *EPHA10* as we believe these genes might acts as 'sensors' for similar methylation changes which are occurring throughout the genome, since they show a very similar pattern of DNA methylation acquisition..

All forms of cancer were examined together, since categorising participants based on type of cancer could not yield groups with large enough sample sizes to make convincing comparisons. Cases of non-melanoma skin cancer were not included as either a previous or future occurrence of cancer as it typically has a specific environmental cause (sunlight exposure), is relatively common and has a very good prognosis, a high cure rate and rare metastasis. Types and frequency of cancer in the Newcastle 85+ study participants are detailed in Table 4.26. Some participants had more than one diagnosis of cancer at different times. Individuals with non-melanoma skin cancer were only included if they had an additional cancer diagnosis at a different site.

When performing the individual analysis of incidences of any previous cancer compared to a control group with no history of cancer (excluding non-melanoma skin cancer), a trend for increased DNA methylation of *HOXD4, TUSC3, TWIST2, HAND2* and *EPHA10* in the group of participants with a previous cancer was observed (statistically significant difference at *TUSC3*, *HOXD4* and the overall mean), which agrees with our hypothesis that individuals with increased levels of DNA methylation may be at increased risk of cancer. However, this increased methylation may not necessarily be pre-existing and may arise as a consequence of the cancer itself or treatment with chemotherapy drugs which can trigger hypermethylation and resistance (Nyce, 1997). These concerns can be avoided by examining a DNA sample taken from participants before their cancer diagnosis, and so not subjected to treatment. This was possible for 72 individuals in the Newcastle 85+ study. All gene loci (*HOXD4, TUSC3, TWIST2, HAND2, EPHA10* and *HOXA5)* again show the same trend. Mean methylation

levels were higher in the group of participants who subsequently developed cancer, which was a statistically significant difference at *HAND2, EPHA10* and the overall mean. Again, these results are in agreement with our initial hypothesis and suggest that increased methylation may exist prior to the development of cancer. It is possible that since the source of DNA is PBL cells, if a haematopoietic malignancy is present, neoplastic transformation may have already commenced and initiated altered methylation changes, although the frequency of haematological malignancy is very low in the Newcastle 85+ study population (Table 4.26). Even if such DNA methylation changes are triggered by cancer initiation but are evident before the onset of clinically detectable cancer, then differential DNA methylation levels at these specific genes may still prove a useful marker for possible early diagnosis of the disease in the elderly.

There are several caveats to the analysis presented here. Firstly, the DNA samples were taken from peripheral blood and the majority of cancer is occurring in other tissues (Table 4.26). However, if the inter-individual differences in DNA methylation have underlying genetic basis, then this may well affect DNA on methylation of genes similarly across multiple tissues. Thus the increased DNA methylation we observed in peripheral blood may also be associated with increased DNA methylation in other tissues. Additionally, if inter-individual DNA methylation differences are caused by an environmental exposure then blood will be a good surrogate for the likely exposure levels of other exposed tissues. Also, blood is a more widely available tissue for study as it can be retrieved in a non-invasive manner, unlike other tissue samples.

Secondly the absolute difference in DNA methylation levels between the groups was small. For this reason it is important to consider where such differences arise. The absolute methylation levels may not be particularly important, but instead may highlight differences in methylation patterns (e.g. at the allelic level of methylation density) in the participants who are at risk for developing cancer. For example, a recent specific case detailing the effect of differential DNA methylation on cancer risk described an MZ twin pair discordant for childhood lymphoblastic leukaemia with secondary cancer of the thyroid. Analysis revealed differential methylation of *BRCA1* in the affected twin compared to unaffected twin (12% and 3% respectively). Closer inspection of the allelic pattern of methylation found that this was comprised of 13% of alleles which were completely hypermethylated, whereas the unaffected twin had no dense methylation and only single CpG sites methylated. This was associated with reduced BRCA1 protein

levels in the affected twin (Galetzka *et al.*, 2012). This study highlights the impact that such relatively small increases can have on DNA methylation levels. The amount of *BRCA1* methylation in the affected twin is similar, if not lower than some levels observed in the Newcastle 85+ study participants, and has a clear impact on protein production, and possibly cancer risk.

Site of Cancer	Previous Incidence (n)				Future Diagnosis (n)		
	1st	2nd	3rd	4th	1st	2nd	3rd
Lip, oral cavity or pharynx	1						
Oesophagus					1		
Stomach	2	1			3		
Colon	11	2			10		
Recto-sigmoid junction	1						
Rectum	7				3		
Liver		1			2		1
Pancreas					1		
Bile duct/Gall bladder					1		
Lung	5	2	2	1	9	3	1
Non-melanoma skin cancer	10	30	17	8	1	1	
Melanoma skin cancer	11	1	1				
Breast	16	2			5		
Cervix	4						
Uterus	3						
Ovary	2				3		
Prostate	14	2			8	1	
Kidney	4				1	1	
Bladder	13	2			12	2	
Thyroid	2						
Lymphoma	5				1	1	
Leukaemia					1		
Multiple myeloma			1				
Brain		1			2		
Eye		1					
Secondary metastatic, ill-defined	1				5	8	1
Other	1				1		1

Table 4.26: Types of cancer in the Newcastle 85+ Study participants

Since there was a large amount of clinical information accompanying the Newcastle 85+ study participants, associations between DNA methylation levels and factors such as gender and smoking behaviour or the presence of age related diseases were investigated to generate new hypotheses concerning the role of DNA methylation in disease. An attempt to validate these initial associations was then attempted in a further set of Newcastle 85+ study participants.

There appeared to be a significant effect of gender on DNA methylation levels. All loci were more methylated in males than females (statistically significant at *TUSC3*, *HAND2*, *EPHA10* and *HOXA5*), and the same trend was confirmed in an additional sample set (statistically significant at *TUSC3* and *HOXD4*), suggesting this could be a real relationship. Higher levels of methylation at specific genes in males have also been reported by others (Sarter *et al.*, 2005; El-Maarri *et al.*, 2007). This association is particularly interesting given the differential life expectancy and discordance in disease rates between males and females. Could differential DNA methylation in the sexes underlie such differences?

There were conflicting effects on DNA methylation levels in current, former or never smokers in the two separate sample sets analysed. The first analysis indicated higher DNA methylation levels in current smokers at except *HAND2*, where current smokers appeared to have the lowest level of methylation. In addition, former smokers displayed the least mean methylation most loci. The validation analysis revealed that *HAND2* methylation was still lowest in current smokers compared to never smokers. The only other association which was replicated in both sample sets was the increased methylation of *HOXA5* in current compared to former smokers.

Previous studies have linked both hypermethylation and hypomethylation of specific genes to current smoking behaviour. For example, frequent methylation of several genes including *RARβ2, CDH13, p16* and *RASSF1A* was noted in the upper aerodigestive tract epithelia of healthy current heavy smokers compared to non-smokers (Zochbauer-Muller *et al.*, 2003). A recent comprehensive genome wide analysis of over 27,000 CpG sites in peripheral blood identified a highly significant single site in the *F2RL3* locus which was methylated at a lower level in smokers than non-smokers (Breitling *et al.*, 2011).

Overall, this analysis suggests there is little effect of smoking behaviour on DNA methylation levels, except perhaps at the *HAND2* or *HOXA5* locus. However, this analysis may complicated by reported observations that the quantity of cigarettes smoked by current smokers may alter DNA methylation levels. Also, for former smokers, time since quitting appears to have a significant impact on DNA methylation. A recent genome wide analysis described two loci, *F2RL3* and *GPR15*, which were hypomethylated in current smokers. Increased smoke exposure (number of pack years) in current smokers was significantly associated with progressively reducing methylation of CpG sites in these genes; whereas for former smokers, increasing time since quitting was linked to increasing levels of methylation (Wan *et al.*, 2012). This suggests that aberrant methylation patterns caused by smoking may be reversible. Neither of these factors is considered in our analysis but may be important for determining methylation levels within the groups of current and former smokers, and explain why we were unable to identify more significant associations.

Nutritional intake, particularly of methyl donors is known to impact upon DNA methylation (see section 1.6.3.1). Previous examination has linked folate deficiency to genomic hypomethylation of DNA (Rampersaud *et al.*, 2000) and increased levels of homocysteine (Jacob *et al.*, 1998), whilst folate supplementation may correct such changes by increasing DNA methylation levels (Jacob *et al.*, 1998), although this has not been replicated by all studies (Rampersaud *et al.*, 2000). Our analysis of methyl donor levels in the Newcastle 85+ study population revealed no relationship between the combined folate, vitamin B12 and homocysteine score and DNA methylation; although a weak but significant positive correlation was observed at the *HAND2* locus in the validation sample set. These observations suggest that differential nutritional intake of methyl donors is unlikely to be significantly contributing towards variable DNA methylation in the Newcastle 85+ study participants.

Analysis of participants with dementia revealed significantly increased methylation levels at the *EPHA10* promoter, which was confirmed in the validation sample set, suggesting that this loci could be worthy of further investigation in dementia. Previous examination of altered DNA methylation in dementia has mostly focused on brain tissue from Alzheimer's disease patients (Siegmund *et al.*, 2007), although a recent study has suggested that differential methylation of specific genes such as *SIRT1* and *APP* is evident in PBL DNA samples taken from Alzheimer's disease patients compared to

controls, which impacts upon expression levels (Hou, 2012). Since no other gene we examined demonstrated differential DNA methylation in participants with dementia, it is unlikely that a general deregulation of DNA methylation is associated with the disease. The difference in *EPHA10* methylation could potentially be driven by a nearby SNP which affects the methylation state of this specific gene in all tissues, and it may be the SNP rather than altered DNA methylation patterns, which is associated with dementia.

In the first analysis, trends of either increased or decreased DNA methylation were associated with clinical outcomes such as cardiovascular disease, type II diabetes, stroke and survival. However we were unable to confirm these same trends in the validation sample set. For this reason, it is likely that there is no real effect of these outcomes on DNA methylation levels at the specific loci examined. For some outcomes, although the same direction of methylation change was observed with disease in both the initial and validation sample sets, DNA methylation levels were not significantly different from those of individuals without the disease, possibly due to the relatively small number of participants diagnosed with many of these age-related diseases. Although trends were implied, the sample size is not large enough for a statistically significant association to be established.

One problem with the analysis presented here is that of multiple testing. The association between DNA methylation levels and the presence of clinical outcomes were analysed individually using a standard significance level of p=0.05, however since multiple outcomes were analysed in the same sample set, this increases the likelihood that significant associations may appear by chance at the level of p=0.05. Therefore, a Bonferroni correction may have been appropriate to improve the stringency of the statistical tests. As an alternative, a linear regression model was constructed which reduced the number of comparisons being made in multiple tests to a single model.

Once all variables were built into the linear regression model, several associations retained a significant effect on DNA methylation levels, including proportion of lymphocytes at all loci, higher methylation in males at *TUSC3, HOXD4* and the overall mean, increased methylation of *HAND2* in participants with a subsequent diagnosis of cancer, reduced methylation of *HAND2* in current or former smokers and increased methylation of *EPHA10* in individuals with dementia, which suggests that these

relationships are not dependent on other confounding factors. This significance of these relationships requires further investigation.

There was no longer a significant relationship between previous incidences or future occurrence of cancer and increased DNA methylation levels at most loci except *HAND2*. One possible explanation for this in case of future cancer is that there are more males (58%) than females (42%), and males have been shown to have increased levels of DNA methylation. When gender is factored into the model, the association of increased methylation in participants with a subsequent diagnosis of cancer is lost. However, this is likely to be a relatively small effect. A more likely explanation is that in the linear regression analysis model, comparisons are made between participants with either a previous history of cancer compared to all other individuals, or a future incidence of cancer compared to all other participants, therefore both future and previous cancers would be included in the respective control groups and lessen apparent associations. This is clearly different to the way the data was analysed in the first instance, with two strict control groups, containing no participants with either a previous cancer, or a diagnosis of cancer within 3 years of DNA sample donation.

Another possibility is that although DNA methylation increases are evident in blood, they are more apparent in the tissue in which the cancer originates. This is well illustrated by the only two subsequent diagnoses of haematological cancer (leukaemia and lymphoma) which occurred in individuals with some of the highest methylation levels of the total population (Table 4.13). Since the majority of both previous and future incidences of cancer in this study are occurring in other tissues, it is possible that a stronger association between increased DNA methylation and cancer would be observed in these tissues rather than blood. However, it is often impractical to obtain normal tissue samples, thus for a large study such as this, blood remains the best source of DNA for analysis.

Chapter 5 – Genome wide methylation analysis of lymphoma cell lines to identify genes susceptible to aberrant methylation in both cancer and ageing

5.1. Introduction

The number of genes with reports of abnormally increased levels of DNA methylation in the cancer epigenome is vast (Costello *et al.*, 2000). Since recently published data has indicated that genes which become hypermethylated in cancer are also susceptible to altered methylation during ageing, this implies that a large number of genes may be susceptible to such aberrant methylation in an elderly individual.

The results presented so far were based on a panel of genes which were pre-selected due to either their known hypermethylation in leukaemia, or their age-associated acquisition of methylation, and these genes also demonstrated highly variable DNA methylation in a population of 85 year olds. These genes also display a strong co-methylation, suggesting that the propensity to methylate promoter associated CpG islands varies between individuals, and that methylation levels of these genes could act as a "sensor" for methylation changes which are occurring at promoter associated CpG islands throughout the genome.

However, since these genes were pre-selected, it is possible that some features may be related to this non-random selection and not applicable more generally to a larger set of genes. Therefore, our aim was to identify hypermethylated genes from an unbiased screen of a haematological cancer and examine whether these loci were methylated in a similar manner to those loci already examined in normal haematopoietic cells.

In order to identify potentially novel methylation changes in haematological malignancies, a genome wide analysis was performed. This was achieved by isolating the methylated fraction of DNA using a methyl-binding domain (MBD) based method. DNA fragments containing methylated cytosines are recognised and bound by the MBD protein component of MeCP2, whilst unmethylated fragments remain unbound. DNA fragments were then subjected to high-throughput sequencing for identification of these hypermethylated regions.

For our analysis, four lymphoma cell lines: DB, HT, Karpas-442 and SUDHL8 were selected. DNA from all four cell lines was pooled to form a single sample rather than examined individually in order to detect aberrant methylation which is present in multiple cell lines. This increases the likelihood that the identified DNA methylated changes are truly cancer associated, rather than being unique to a single cell line, as such changes may be less relevant in terms of disease and possibly due to cell type specific variation. Furthermore, although malignant cell lines generally have more frequent methylation than primary samples, this appears to be largely due to an increased frequency of methylation of the same set of genes found to be methylated in primary samples, as opposed to the existence of large numbers of cell line specific methylated loci (Toyota *et al.*, 1999).

Finally, the methylation levels of genes identified by this method will be tested in a subset of the Newcastle 85+ study participants, to confirm that genes susceptible to both age-related and cancer associated methylation are overlapping. Further examination will reveal whether these loci also show a similar pattern of co-methylation to genes in the pilot study, suggesting a shared mechanism for the propensity to methylate CpG island sequences at numerous loci throughout the genome.

5.1.1 Enrichment of methylated regions by MBD binding

After DNA was prepared from the four individual lymphoma cell lines by standard phenol chloroform extraction, DNA was combined to form a single sample and fragments of DNA were generated firstly using the Bioruptor® Plus sonication system; however gel electrophoresis and the fragment profile generated by Bioanalyzer analysis suggested that some of the fragments may be too large, ranging from 150-650bp with a mean length of approximately 550bp (Figure 5.1A and B). The optimum fragment size for enrichment with MBD protein ranges from 200-500bp, whereas the ideal size for Illumina GAIIx sequencing is around 100-300bp, including the additional sequencing adapters. For this reason, the Covaris system was also used for DNA shearing which allowed the breakdown of DNA fragments which were too large without causing addition fragmentation to sections of DNA which were the correct size. The fragmented DNA was used to prepare a sequencing library involving the ligation of distinct sequences to the 5’ and 3’ ends of the fragments. After the library preparation, the fragment profile was again checked using the Bioanalyzer and the mean fragment length

was 241bp (Figure 5.2), an ideal length for performance of both MethylCap enrichment for DNA fragments containing methylated CpG sites and subsequent sequencing. The adapter ligated DNA was subjected to methylated DNA precipitation by MBD protein binding, using the MethylCap method. Using this technique, DNA can be precipitated on the basis of CpG methylation density by increasing the concentration of salt in the elution buffer.

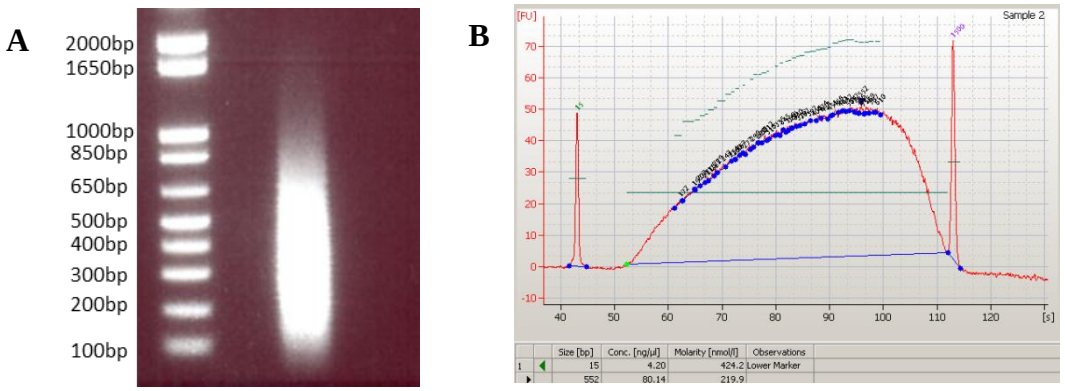


Figure 5.1: (A) gel electrophoresis image and (B) Bioanalyzer analysis of sonicated DNA to determine fragment size

(A) 500ng (5µl) of sonicated DNA was run on a 1% agarose gel with 1 Kb Plus DNA Ladder. The smear indicates fragmented DNA in the range of 150-650bp.

(B) 100ng (1µl) of sonicated DNA was examined by Bioanalyzer analysis. This allows more accurate determination of fragment sizes then electrophoresis, and quantitation of DNA by examination of the area under the curve, which indicated a fragment range from 150-620bp with a mean length of 552bp. Accuracy was determined by inclusion of internal controls at 15bp and 1500bp.

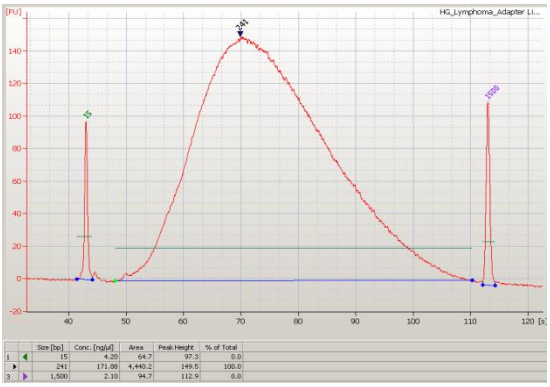


Figure 5.2: Bioanalyzer analysis of sonicated DNA to determine fragment size after ligation of sequencing adapters

1µl of sonicated DNA was analysed by the Bioanalyzer allowing accurate determination of fragment size and quantity of DNA by examination of the area under the curve, which indicated a mean fragment size of 241bp and a DNA concentration of 172ng/µl. Accuracy was determined by inclusion of internal controls at 15bp and 1500bp.

5.1.2 Validation of enrichment by quantitative PCR (qPCR) analysis of known methylated and unmethylated genes

In order to validate the MethylCap technique for isolating methylated regions of the genome, successful enrichment was determined by qPCR analysis of a known methylated gene, *HOXA5;* and a known unmethylated gene, *GAPDH*. Each sample was tested in triplicate, and an average Ct value was calculated from three values. In order to determine the amount of methylated gene template relative to unmethylated gene template in the input DNA as well as the low, medium and high salt concentration elution, the delta (Δ) Ct value was calculated by the following formula:

$$\Delta Ct = Ct\ (GAPDH) - Ct\ (HOXA5)$$

As the salt concentration of elution buffers was increased, the amount of *HOXA5* template relative to *GAPDH* template also increased (demonstrated by an increasing ΔCt value) (Figure 5.3); confirming that *HOXA5*, a methylated gene, is enriched after the MBD protein based precipitation of DNA. In the high salt elution, this typically resulted in an enrichment equivalent to 10 cycles, which represents approximately 1000-fold enrichment for a known methylated region, compared to the input fraction containing DNA which has not been subjected to MBD enrichment. The DNA obtained

from the high salt fraction was used for high throughput sequencing on the Illumina GAIIx platform as this fraction demonstrated the largest fold enrichment for a known methylated sequence.

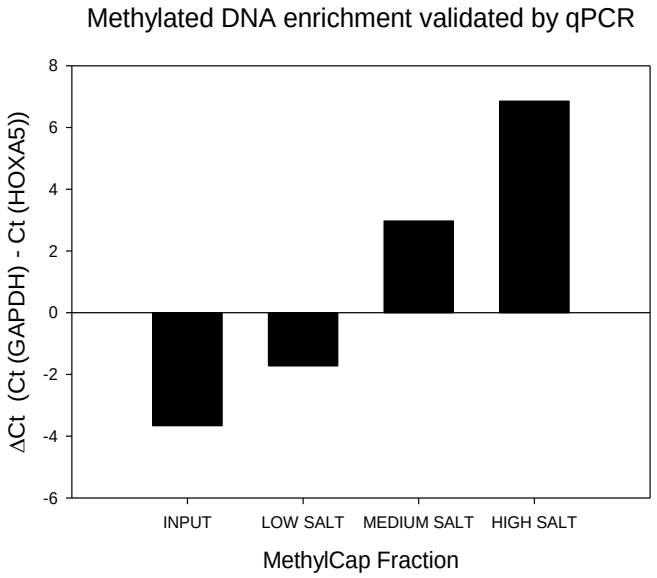


Figure 5.3: Validation of the MethylCap technique for methylated DNA enrichment by qPCR analysis

The Ct Value of *HOXA5* and *GAPDH* were compared in each of the different salt concentration precipitation fractions, plus input DNA. The Y axis represents the delta (Δ) Ct value of *HOXA5* Ct values compared to *GAPDH* Ct values as this represents the magnitude of HOXA5 enrichment (A higher delta (Δ) Ct value indicates the presence of more of the methylated target compared to unmethylated). In the input fraction, no enrichment of methylated DNA is expected and there is more *GAPDH* template than *HOXA5* (shown by a negative ΔCt value). *HOXA5* enrichment is enhanced by increasing salt concentrations (shown by progressively higher ΔCt values).

5.1.3 Sequencing and Bioinformatics analysis

Once enrichment for methylated CpG containing sequences was confirmed by PCR analysis, the DNA fragments were subjected to 75 cycles of paired end sequencing on the Illumina GAIIx platform, generating two 75bp sequencing reads (forward and reverse).

Bioinformatic analysis was executed by Dr Simon Cockell. Quality control analysis removed unreliable sequencing reads prior to mapping onto the human reference

genome for identification of enriched fragments peaks from the 4 lymphoma cell lines. A gene list was created based on the 10,000 sequencing peaks which either crossed the transcriptional start site, or were within 300bp of annotated genes. In order to restrict the number of potential target genes for validation, two additional gene lists were created containing the top 10% and bottom 10% of peaks which mapped near genes, based on the accompanying p-value. The gene list featuring the top 10% of peaks still contained over 800 genes, whereas the bottom 10% of peaks mapped near just over 350 genes since a greater proportion of these peaks matched pseudo genes which were not included in the analysis (for further explanation see section 2.15). Genes were selected from list of the bottom 10% to compare to those from the top 10% to determine whether there was any differences in methylation status of the genes associated with peaks with more significant p-values. It was expected that peaks with more strongly significantly p-values (top 10%) have greater confidence and may be better candidates for true hypermethylated cancer-associated genes than those with lower p-values (bottom 10%).

5.2 Validation of methylation status of specific genes in lymphoma cell lines by COBRA

Genome wide analysis provided a large list of possible target genes for further analysis, therefore a reduced list from which to the select target genes for validation of their methylation status in primary samples was required. Consequently, the methylation status of genes contained within the top 10% and bottom 10% of all sequencing peaks based on p-value were compared using an alternative technique for determining DNA methylation (COBRA).

From the gene list representing the top 10% of peaks, seven genes were selected at random for further validation and analysis. These genes were *APP, CACNA1B, CELF4, CYGB, ERN2, HAP1* and *SH2D4A.* Additionally, four genes were chosen from the gene list featuring the bottom 10% of peaks, which were *EBF4, HCK, PDCD5* and *TCL1B.*

Initially, the DNA methylation status of target genes in the individual cells lines was established by COBRA analysis of the promoter regions. This involved restriction enzyme digestion of each PCR amplified amplicon with enzymes specifically targeting sequences containing a CpG site. The sequence change (C→T conversion) induced by bisulfite modification at unmethylated sites means that the enzyme cutting site is no

longer recognised and the amplicon remains undigested. Methylated sequences are recognised and cut into smaller fragments which can be detected by agarose gel electrophoresis.

Of the genes within the top 10%, the COBRA results indicated that all four cell lines were highly methylated at the promoter regions of the *CANA1B, CELF4* and *CYGB* genes. Both the HT and SUDHL8 cell lines were unmethylated at the *APP* locus whilst DB and Karpas-442 were both hypermethylated. The Karpas-442 cell line demonstrated only partial methylation at the *ERN2* promoter, whilst the remaining cell lines were highly methylated at this gene. The HT cell line was unmethylated at the *HAP1* promoter, whilst the other three cell lines were highly, although not completely, methylated. Finally, whilst the DB, HT and Karpas-442 lines were completely hypermethylated at the *SH2D4A* promoter, SUDHL8 was only partially methylated (Figure 5.4). These results demonstrate that all of these seven randomly selected genes are highly methylated in multiple cell lines. 3/7 were hypermethylated in all four cell lines, 3/7 were highly methylated in three of the cell lines and just 1 gene was only methylated in two of the cell lines.

Conversely, of the genes selected from the gene list representing the bottom 10% of sequencing peaks, none of the cell lines were methylated at the *PDCD5* locus. *HCK* was unmethylated in the SUDHL8 cell line, and showed only partial methylation in both the DB and Karpas-442 line whilst HT was only partially methylated at BstUI digested sites at this locus. The Karpas-442 cell line showed hypermethylation of *EBF4* however only partial methylation of specific sites was detected in DB and SUDHL8 whilst HT was largely unmethylated. *TCL1B* was hypermethylated in all four cell lines; however it is also highly methylated in peripheral blood making it an uninformative candidate for hypermethylation in leukaemia (Figure 5.5). These results suggest that genes within the bottom 10% of identified peaks are methylated at a lower frequency in the cell lines compared to those in the top 10%, therefore may be more representative of cell specific methylation changes as opposed to cancer specific changes.

Consequently, the genes featured in the top 10% of sequencing peaks may represent better candidates for hypermethylated loci in primary leukaemia samples, than those in the bottom 10%, and these gene targets identified by our genome wide methylation analysis are representative of methylation changes occurring in the majority of the cell lines.

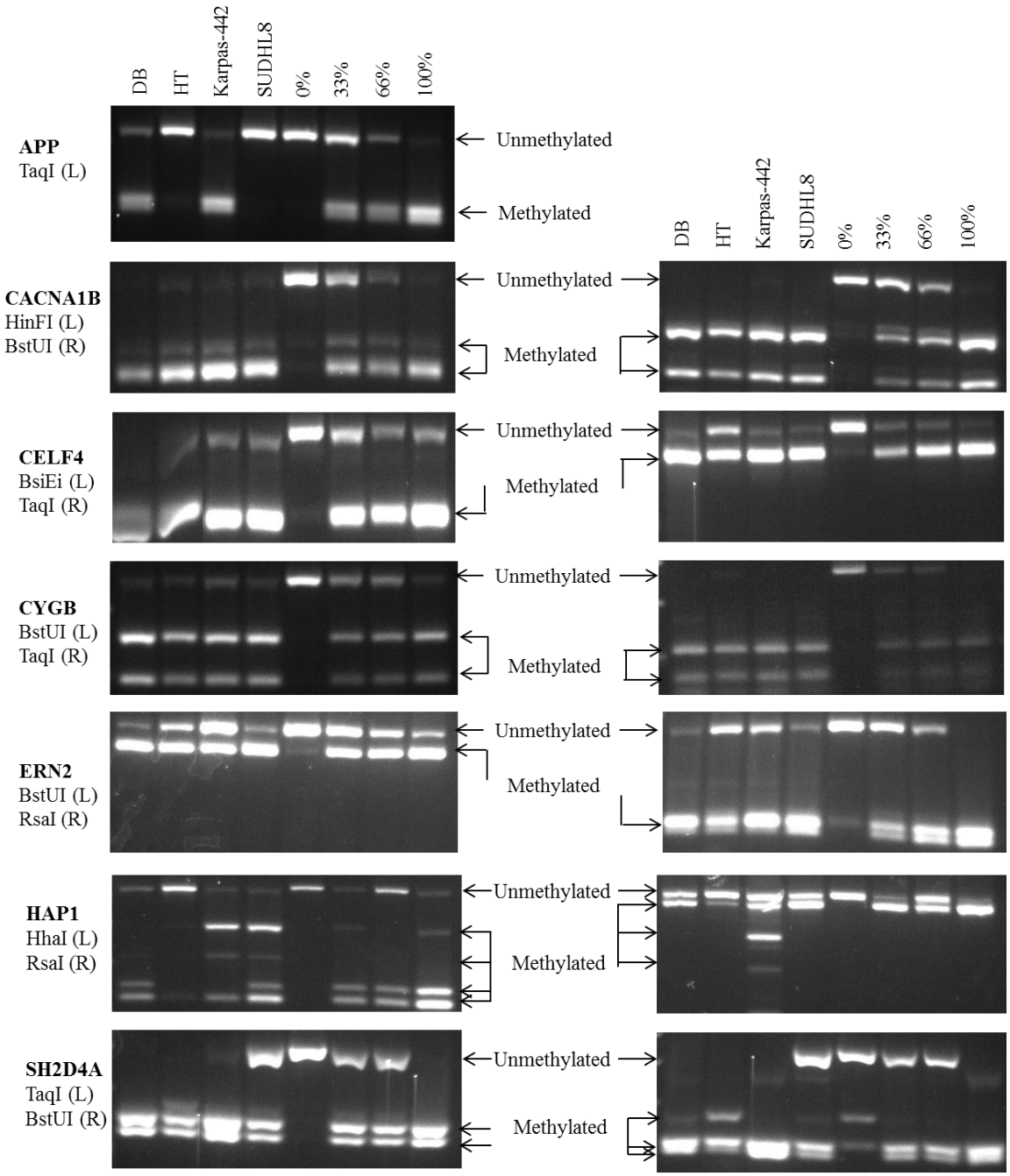


Figure 5.4: COBRA analyses of seven genes selected from the top 10% of sequenced peaks in the individual lymphoma cell lines: DB, HT, Karpas-442 and SUDHL8

Restriction enzyme digestion of the *APP, CACNA1B, CELF4, CYGB, ERN2, HAP1* and *SH2D4A* amplicons in the four individual cell lines, alongside methylation standards of 0%, 33%, 66% and 100%. Restriction enzymes used for DNA digestion are shown and detailed in chapter 2 (Table 2.6). Only the TaqI digest is shown at the *APP* locus. (L) = left, (R) = right

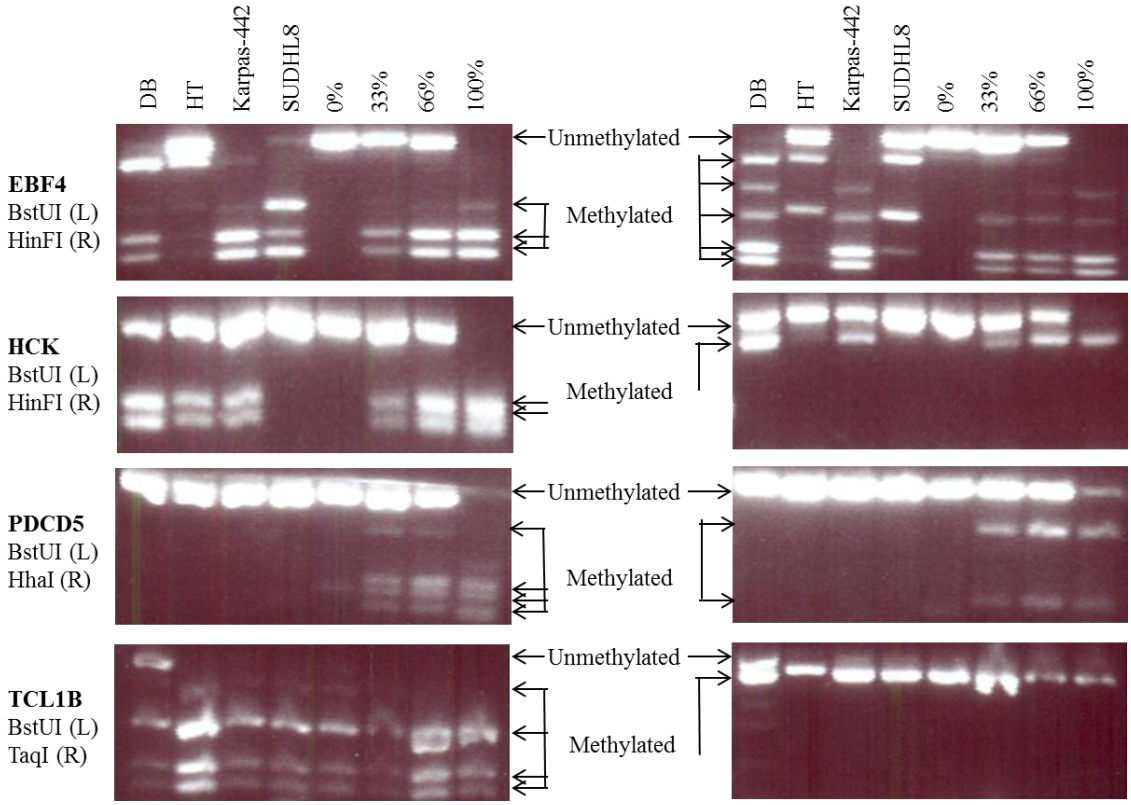


Figure 5.5: COBRA analyses of four genes selected from the bottom 10% of sequenced peaks in the individual lymphoma cell lines: DB, HT, Karpas-442 and SUDHL8

Restriction enzyme digestion of the *EBF4*, *HCK*, *PDCD5* and *TCL1B* amplicons in the four individual cell lines are shown, alongside known methylation standards of 0%, 33%, 66% and 100%. Restriction enzymes used for DNA digestion are shown and detailed in chapter 2 (Table 2.6) (L) = left, (R) = right.

5.3 Validation of methylation status of specific genes in primary lymphoma/ALL samples by COBRA

Once the methylation status of *APP*, *CACNA1B*, *CELF4*, *CYGB*, *ERN2*, *HAP1* and *SH2D4A* was confirmed by COBRA analysis in the cell lines, a panel of primary adult acute lymphoblastic leukaemia (ALL) samples were selected for analysis of their methylation status. It is important to confirm that methylation exists in primary samples and therefore may be cancer associated, rather than being an artefact of cell culture. We did not have sufficient primary lymphoma samples to test, however lymphoid leukaemias are closely related to lymphomas as both affect cells of the lymphoid

lineage (Swisher *et al.*, 1991) and previous analysis in the lab had suggested a strong overlap in the genes methylated in lymphoid leukaemia and lymphoma. The four genes selected from the bottom 10% of identified targets were not taken forward for further analysis due to the lack of frequency with which they were methylated in the individual cell lines. A total of eighteen primary ALL samples were amplified and digested with restriction enzymes for each gene, alongside 0%, 33%, 66% and 100% controls (Figure 5.6). A sample was defined as hypermethylated by the presence of greater than 50% methylation. A sample containing >10% but <50% was categorised as partially methylated, whilst a sample with <10% methylation was unmethylated.

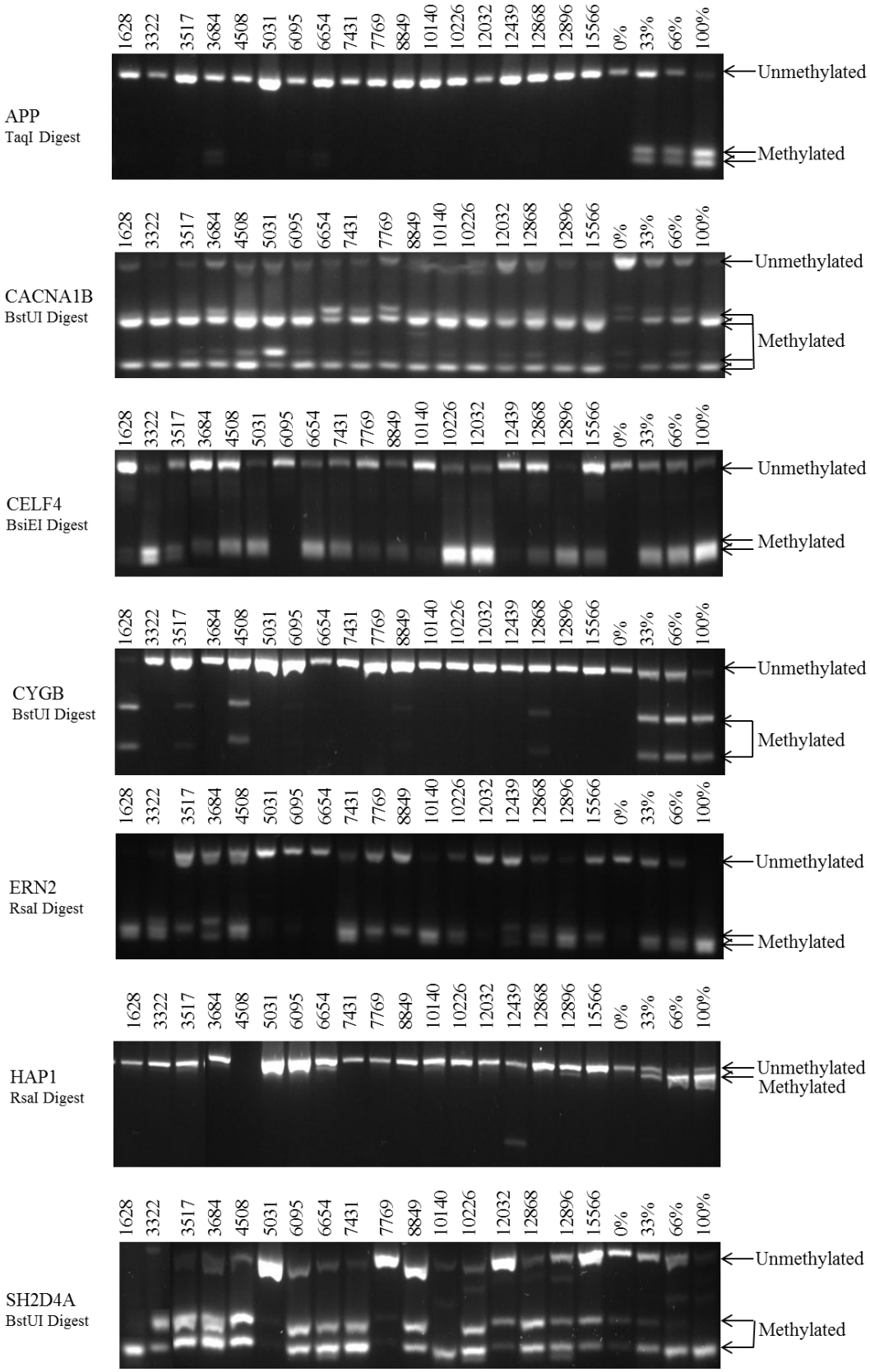


Figure 5.6: DNA methylation of *APP, CACNA1B, CELF4, CYGB, ERN2, HAP1* and *SH2D4A* in primary acute lymphoblastic leukaemia samples by COBRA

Restriction enzyme digestion of the *APP, CACNA1B, CELF4, CYGB, ERN2, HAP1* and *SH2D4A* amplicons in primary ALL samples alongside methylation standards of 0%, 33%, 66% and 100%. Only one restriction enzyme digestion is shown for each locus as an example.

The results demonstrated that methylation of the *CACNA1B* locus appeared to be a very frequent event in ALL, with hypermethylation of all samples evident. Likewise, the majority of ALL samples were highly methylated at *SH2D4A.* At both the *CELF4* and *ERN2* genes partial or hypermethylation was common. In contrast, methylation was much less frequent at the *CYGB* promoter and furthermore, there was no detectable methylation of the any of the *APP* or *HAP1* promoter in any of the adult ALL samples digested with both restriction enzymes (Table 5.1).

Gene	Restriction Enzymes	Hypermethylated (n)	Partial methylation (n)	Unmethylated (n)
APP	TaqI	0	0	18
	HpyCH4IV	0	0	18
CACNA1B	HinFI	16	1	0
	BstUI	17	0	0
CELF4	BsiEI	7	8	3
	TaqI	4	13	1
CYGB	TaqI	1	4	13
	BstUI	1	4	13
ERN2	RsaI	8	6	4
	BsiEI	4	7	7
HAP1	HhaI	0	0	18
	RsaI	0	0	18
SH2D4A	BstUI	11	4	2
	TaqI	10	5	2

Table 5.1: Results of COBRA assay digestion of *APP, CACNA1B, CELF4, CYGB, ERN2, HAP1* and *SH2D4A* in primary ALL samples

Each gene was digested with two restriction enzymes. The number of samples which were defined as being hypermethylated (>50%), partially methylated (>10% and <50%) and unmethylated (>10%) with each digest are shown in the Table.

5.4 Measuring the methylation status of *APP, CACNA1B, CELF4, CYGB, ERN2, HAP1* and *SH2D4A* in the Newcastle 85+ Study by Pyrosequencing

In order to contribute to the hypothesis that age associated and cancer associated methylation changes are overlapping, we examined whether these genes which are hypermethylated to differing degrees in lymphoid malignancies, also exhibit age associated variability in DNA methylation. DNA methylation levels were measured in a subset of the Newcastle 85+ study participants by Pyrosequencing.

5.4.1 Pyrosequencing assay design

Suitable forward, reverse and sequencing primers corresponding to the region surrounding transcriptional start site were designed using PSQ assay design software as previously described (see section 2.2).

5.4.2 Validation of Pyrosequencing assays

Previous COBRA analysis demonstrated that all seven genes were largely unmethylated in normal peripheral blood from young volunteers (0% methylated control – see Figure 5.6). Validation of the Pyrosequencing assays was performed by mixing unmethylated normal peripheral blood with varying quantities of *in vitro* methylated DNA, to create known methylation standards of 0%, 5%, 10%, 15%, 20%, 30%, 50%, 60%, 70% and 100%. This known methylation was plotted against methylation values obtained by Pyrosequencing measurements. For all genes, average methylation was calculated across 3-5 consecutive CpG sites and all samples were tested in duplicate. Successful assay validation was achieved for the *APP, CYGB, ERN2, HAP2* and *SH2D4A* genes, shown by the strong, positive correlation between observed and expected methylation measurements (Figure 5.7). Unfortunately, validation of *CACNA1B* and *CELF4* was not possible, despite numerous redesigns of sequencing primers and forward and reverse primer sequences. Therefore the DNA methylation status of these two genes in the Newcastle 85+ population remains undetermined.

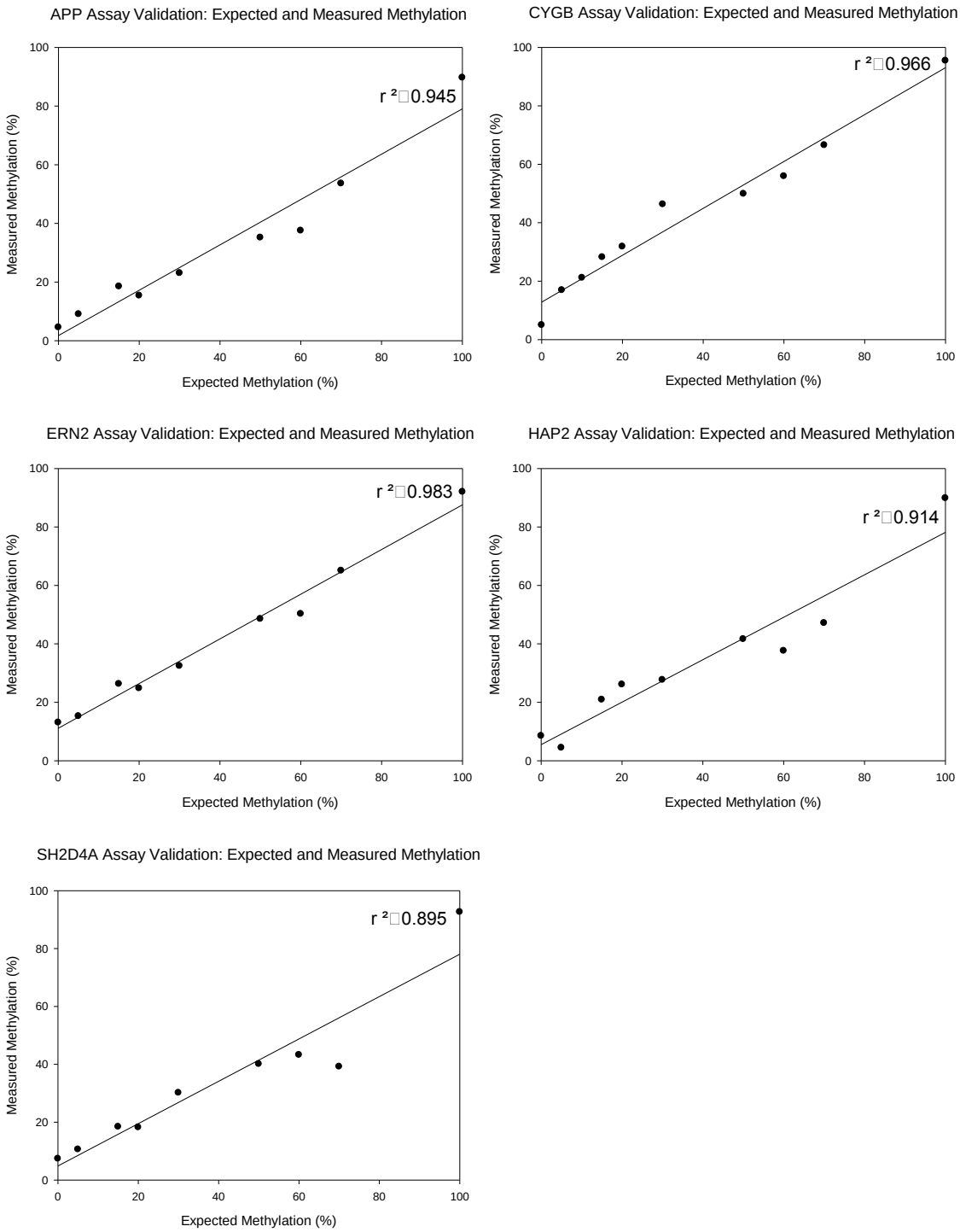


Figure 5.7: Validation of Pyrosequencing assays for *APP, CYGB, ERN2, HAP2* and *SH2D4A* genes identified from genome wide methylation analysis of lymphoma cell lines

Linear regression analysis was performed by plotting the expected methylation of each sample against the methylation levels at the *APP, CYGB, ERN2, HAP2* and *SH2D4A* loci measured by the PyroMark MD instrument, with all assays demonstrating a robust positive relationship between these measurements (respective r^2 values shown on each graph). All samples were tested in duplicate and an average of 3-5 CpG sites calculated.

5.4.3 Pyrosequencing results of *APP, CYGB, ERN2, HAP2* and *SH2D4A* methylation levels in PBL DNA samples from the Newcastle 85+ study population

A sample size of 50 participants was selected for the methylation analysis of *APP, CYGB, ERN2, HAP2* and *SH2D4A* in the Newcastle 85+ study, as this quantity of participants was successful in the pilot study analysis for both identifying the extent of methylation levels at individual loci, and to make an assessment of inter-individual variation in methylation at specific genes.

The Pyrosequencing results are presented in Table 5.2, and show the most highly methylated and most variable of these five genes was *ERN2* (11.8 ± 3.6% [mean ± SD] – Figure 5.8), which is similar to the mean methylation and considerable variability seen for the most variably methylated genes in the pilot study which were; *HOXD4* (14.8 ± 6.0%), *EPHA10* (12.0 ± 3.9%), *TUSC3* (8.9 ± 3.4%), *TWIST2* (8.4 ± 3.3%) and *HAND2* (7.3 ± 2.8%).

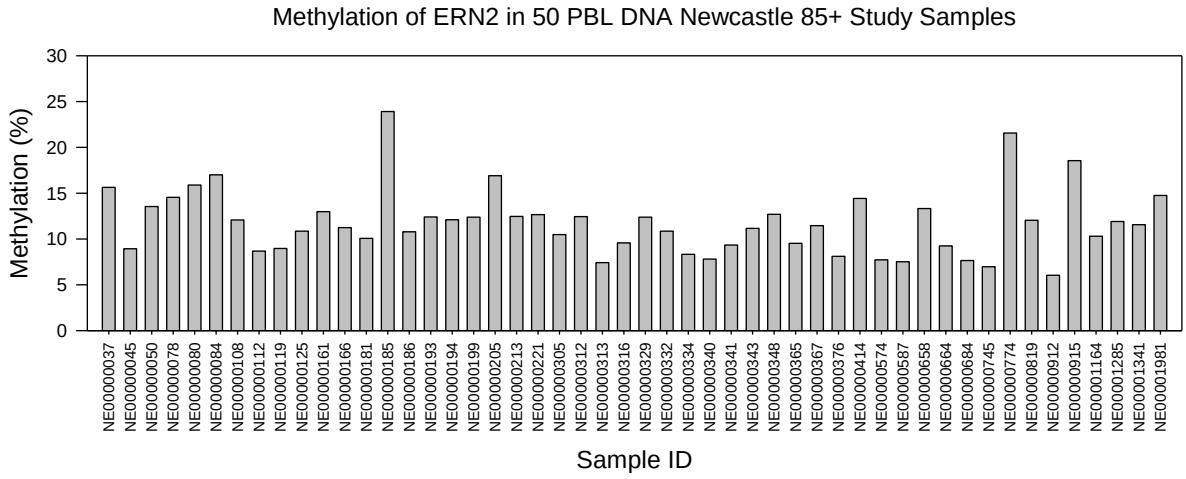
The other four genes: *HAP* (5.1 ± 3.0%), *APP* (4.0 ± 2.1%), *SH2D4A* (4.6 ± 1.7%), and *CYGB* (4.1 ± 1.4%), exhibited lower average DNA methylation measurements, although some variation was apparent (Figure 5.8). These methylation patterns at four genes were comparable to *HLXB9* (3.3 ± 2.1%), *APOE* (8.4 ± 1.9%), *p15* (4.9 ± 1.9%) and *ESR1* (5.5 ± 1.6%) from the pilot study which showed some evidence of moderate variability.

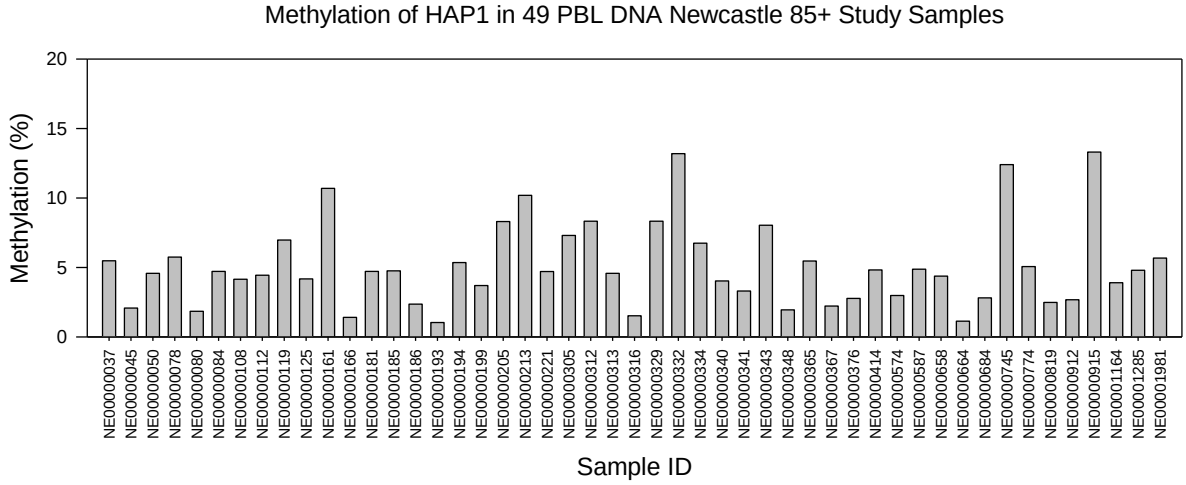
The Pyrosequencing results indicate that the *APP, CYGB, ERN2, HAP2* and *SH2D4A* promoter regions, identified from a genome wide analysis of hypermethylated regions in lymphoma cell lines and ALL primary samples, also show some evidence of becoming methylated in normal elderly PBL DNA samples, suggesting that they may also be susceptible to acquiring DNA methylation during ageing as well as in cancer. However, the levels of methylation are not as high as several of the genes tested during the pilot study.

Gene	Methylation range (%)	Mean Methylation (%)	Standard Deviation (%)
ERN2	6-24	11.8	3.6
HAP1	1-13	5.1	3.0
APP	1-14	4.0	2.1
SH2D4A	2-11	4.6	1.7
CYGB	1-8	4.1	1.4

Table 5.2: DNA methylation properties of *APP, CYGB, ERN2, HAP2* and *SH2D4A* analysed by Pyrosequencing in 50 PBL samples from the Newcastle 85+ study participants

For each gene, the minimum and maximum methylation levels in the samples is shown, alongside average methylation calculated from 3-5 consecutive CpG sites and the standard deviation as a measure of variation.





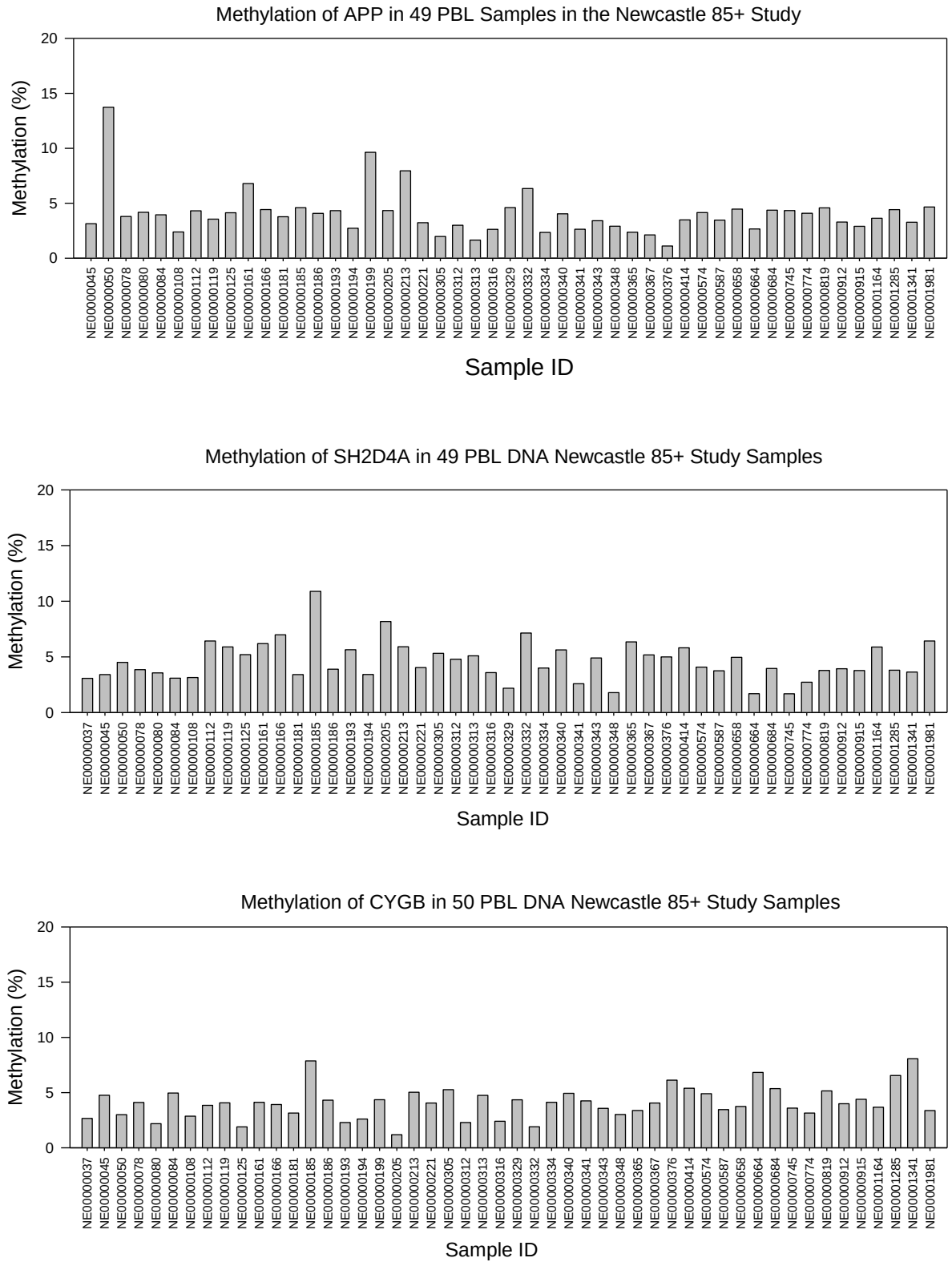


Figure 5.8: DNA Methylation levels of *ERN2, HAP2, APP, SH2D4A* and *CYGB* in 50 PBL DNA samples from the Newcastle 85+ study measured by Pyrosequencing

The five genes demonstrate variable methylation levels. One sample was excluded from *HAP1, APP* and *SH2D4A* results due to a lack of reproducible methylation measurements.

5.4.4 Correlation of *ERN2, HAP2, APP, SH2D4A* and *CYGB* methylation levels with methylation levels of genes measured in the pilot study of Newcastle 85+ study participants

Finally, the methylation levels of *APP, CYGB, ERN2, HAP2* and *SH2D4A* were assessed to determine whether these genes displayed co-methylation, in a similar manner to those genes displaying variable methylation in the pilot study. As well as the pairwise correlation between *APP, CYGB, ERN2, HAP2* and *SH2D4A* DNA methylation levels, a correlation was assessed with the eight genes which were strongly correlated with each other in the pilot study. The purpose of this was to further test the hypothesis that certain individuals are more susceptible to acquiring methylation at numerous loci throughout the genome, which may explain why methylation levels are extremely variable within the population. The Pearson correlation coefficient was calculated based on methylation levels for all fifty participants at each pair of loci (Table 5.3). This analysis revealed that, *ERN2* methylation levels were positively correlated with methylation levels of *TUSC3, HOXD4, HAND2, HLXB9, ESR1* and *p15* with high significance, and the association with *TWIST2* was borderline statistically significant. Likewise, methylation of *SH2D4A* demonstrated a significant and positive relationship with methylation of *HOXD4, HLXB9* and *ESR1*, and also a positive correlation with *TWIST2, TUSC3* and *p15* although these failed to reach the significance level required to compensate for multiple testing. Correlations concerning *CYGB* methylation and *TWIST2, HOXD4* and *HLXB9* methylation were suggested but did not reach statistical significance. Conversely, there was no indication of any relationships between *APP* and *HAP1* methylation levels with any other gene.

Both *ERN2* and *SH2D4* were frequently methylation in primary ALL samples and the *ERN2* promoter demonstrated the most extensive variability in the 85+ population out of the five gene targets identified by genome wide DNA methylation analysis. In contrast neither the *APP* or *HAP1* promoter was hypermethylated in any of the primary ALL samples; suggesting aberrant methylation of these genes is not common in leukaemia.

This suggests that two of the genes, *ERN2* and *SH2D4*, identified by the genome wide methylation analysis for cancer associated loci are also methylated in an aged population, and similarly fit the pattern of a "CpG island methylator phenotype"

observed in the pilot study whereby certain individuals are predisposed to methylate a large number of loci in normal haematopoietic cells.

	APP	*CYGB*	*ERN2*	*HAP1*	*SH2D4A*
APP					
CYGB	r = -0.11 p = 0.44				
ERN2	r = 0.21 p = 0.14	r = -0.01 p = 0.90			
HAP1	r = 0.18 p = 0.22	r = -0.15 p = 0.29	r = 0.17 p = 0.23		
SH2D4A	r = 0.19 p = 0.19	r = 0.03 p = 0.82	r = 0.26 p = 0.07	r = 0.16 p = 0.28	
TWIST2	r = -0.02 p = 0.86	r = 0.36 p = 0.01	r = 0.34 p = 0.01	r = 0.13 p = 0.36	r = 0.38 p = 0.006
TUSC3	r = 0.04 p = 0.74	r = 0.21 p = 0.14	r = 0.48 p = 0.0004	r = -0.06 p = 0.68	r = 0.41 p = 0.003
HOXD4	r = -0.003 p = 0.97	r = 0.31 p = 0.03	r = 0.47 p = 0.0006	r = 0.08 p = 0.56	r = 0.48 p = 0.0004
HAND2	r = 0.05 p = 0.71	r = 0.10 p = 0.48	r = 0.71 p = 5.5e-09	r = 0.18 p = 0.22	r = 0.38 p = 0.007
EPHA10	r = -0.04 p = 0.74	r = 0.24 p = 0.09	r = 0.21 p = 0.15	r = 0.08 p = 0.59	r = 0.28 p = 0.08
HLXB9	r = 0.015 p = 0.92	r = 0.44 p = 0.002	r = 0.51 p = 0.0002	r = 0.001 p = 0.98	r = 0.46 p = 0.001
ESR1	r = 0.04 p =0.84	r = 0.36 p =0.08	r = 0.68 p =0.0002	r = 0.20 p =0.34	r = 0.67 p =0.0006
p15	r = -0.01 p = 0.94	r = 0.11 p = 0.41	r = 0.50 p = 0.0004	r = 0.004 p = 0.97	r = 0.30 p = 0.04

Table 5.3: Pair-wise correlations between PBL DNA methylation levels of genes identified from genome wide analysis and genes analysed in the pilot study of Newcastle 85+ study participants

DNA methylation data for fifty participants (25 individuals for *ESR1*) was correlated at each pair of loci using Pearson test of correlation. The Pearson correlation coefficient (r) and corresponding p values are shown. Methylation levels at *ERN2* and *SH2D4A* are highly significantly correlated with methylation at several other loci. Black p value = NS, Orange p value = significant at <0.05 level, Red p value = significant at <0.0010 level after performing Bonferroni correction for multiple comparisons.

5.5 Discussion

The aim of this part of the study was to strengthen the previously noted overlap between genes which are susceptible to hypermethylation in both cancer and ageing. Initially, a genome wide DNA methylation analysis of lymphoma cell lines was undertaken.

The use of cell lines in this analysis has disadvantages, such as the propensity for cell lines to exhibit exceptionally abnormal methylation profiles, more so than their counterpart primary cancer cells (Smiraglia *et al.*, 2001). However, a consequence of this high number of aberrantly methylated genes is increased frequency of potentially cancer associated loci, which is beneficial for the identification of novel targets. Cell lines are also useful for providing an almost infinite source of DNA, as opposed to primary samples for which abundant DNA can be difficult to obtain.

A possible problem caused by the use of cancer cell lines is that of ploidy. Although normal human cells are diploid, immortalised cell lines and cancer cells often exhibit aneuploidy (Pellman, 2007). This can result in certain genes being present at a different copy number to the usual two per cell. When cell lines are mixed such as in the analysis presented here, the enrichment for hypermethylated genes and subsequent identification by genome wide sequencing may be artificially enhanced due to multiple copies of the same gene caused by an increased number of chromosomes in certain cell lines. The possible bias caused by aneuploidy could be avoided by controlling for copy number alterations. Aneuploidy could be identified by sequencing the input genomic DNA of the cell lines without any enrichment for DNA methylation to identify whether the signal for certain chromosomes was significantly different to the average signal across the genome.

The captured of methylated DNA fragments by MBD proteins; in combination with high throughout sequencing has many advantages over other techniques for methylation analysis. This method required no sodium bisulfite conversion of the DNA as the protein acts directly on 5'methylcytosine, which removes potential over-estimation of DNA methylation levels due to incomplete modification of DNA. De novo sequencing for the identification of methylated regions of the genome was determined to be the most informative analysis since there is no preference towards certain gene targets,

unlike an array for example, where a limited number of genes or individual CpG sites are represented.

The validation of hypermethylated targets identified from the genome wide methylation analysis of lymphoma cell lines was performed in primary samples of acute lymphoblastic leukaemia origin. Similarities exist between lymphomas and lymphoblastic leukaemia as both originate in lymphoid cells, and they share expression of some cell surface markers (Swisher *et al.*, 1991), therefore it is likely that many genes exist which are susceptible to aberrant methylation in both malignancies. In accordance with this, it has been demonstrated that the genes subjected to hypermethylation in cases of lymphoma and lymphoid leukaemia can be overlapping (Corn *et al.*, 1999; Takahashi *et al.*, 2004). However, more recent and comprehensive microarray analysis of over 700 genes in 367 haematopoietic neoplasms of different subtypes found important differences in the methylation profiles of all subtypes including lymphoblastic leukaemia and lymphomas (Martin-Subero *et al.*, 2009). Also, gene expression arrays have revealed differential transcriptional profiles of lymphoblastic leukaemia and lymphoma, suggesting that distinctive gene regulatory mechanisms, including DNA methylation levels, may be responsible for such differences between the two types of malignancy (Raetz *et al.*, 2006). This may explain why some of the genes identified as being hypermethylated in the lymphoma cell lines, were not methylated in primary leukaemia samples. None of the ALL samples showed evidence of any methylation at the *APP* or *HAP1* loci, and only one sample was hypermethylated at the *CYGB* locus. Either using primary samples for the initial genome wide methylation analysis, or validating the genes hypermethylated in lymphoma cell lines using primary lymphoma tissues may have increased the specificity of identified targets, but sensitivity for new targets would be reduced unless a large number of primary samples could be analysed for recurring DNA methylation changes.

The Pyrosequencing results of 50 participants from the Newcastle 85+ study showed variation at all loci, although overall DNA methylation levels were quite low at most loci. The methylation levels were highest at *ERN2*, which was one of the genes for which several ALL samples were hypermethylated. Even the three loci, *APP*, *HAP1* and *CYGB* which were unmethylated or methylated very infrequently in the ALL samples, showed some evidence of variable methylation in the 85 year old population.

Out of the seven genes, only four (*CACNA1B*, *CELF4*, *ERN2* and *SH2D4A*) were methylated with a high degree of frequency in primary leukaemia samples. Unfortunately, no assessment of *CACNA1B* and *CELF4* could be made in the Newcastle 85+ study participants, but pair-wise correlation analysis of *ERN2* and *SH2D4A* revealed strong positive and highly significant correlations between both *ERN2* and *SH2D4A*, and the methylation levels at several of the loci measured in the pilot, as well as a positive relationship between methylation levels of these two loci although this was of borderline statistical significance.

This is similar to the observation made in the pilot study whereby genes which are methylated in leukaemia/lymphoma, exhibit variable methylation in the elderly and a pattern of methylation acquisition which is similar at multiple loci across the genome. The fact that methylation levels of the loci which were confirmed to be frequently methylated in leukaemia in our analysis, were correlated with methylation levels at other loci in normal cells, suggests a mechanism affecting the acquisition of methylation which could potentially be affecting hundreds of loci simultaneously (an age-related "CpG island methylator phenotype"). Some individuals may be more susceptible to acquiring methylation, explaining their high methylation levels of numerous loci. Further validation and testing of other gene targets identified from the genome wide methylation analysis could confirm whether this is the case. My results also suggest the existence of a second type of increased methylation during ageing, of those genes which are infrequently methylated in leukaemia. The methylation levels at these genes do not correlate with each other, so DNA methylation at these promoters may be sporadic.

For the *SH2D4A* gene, there is some reported evidence of altered expression in malignancies, which could potentially be mediated by hypermethylation. Although there is no direct evidence in the literature regarding the presence of hypermethylation of *SH2D4A* in leukaemia; however the number of reports of activity of this gene is low. The SH2D4A protein has a high level of homology to two other proteins, TSAd and ALX, which function in the regulation of T cell receptor signalling. Both the generation of SH2D4A deficient mice, and the knockdown of *SH2D4A* by RNA interference in human T cells appeared to have no effect of normal functioning, suggesting a non-essential role in signalling (Lapinski *et al.*, 2008). As well as being expressed in T cells, *SH2D4A* is universally expression in numerous tissues and has been suggested to be involved in estrogen receptor alpha signalling pathways, acting as an inhibitor of

proliferation (Li *et al.*, 2009). In addition to this, a region on chromosome 8p containing the *SH2D4A* gene is deleted in incidences of hepatocellular carcinoma (HCC) with poor prognosis. Re-expression of *SH2D4A* in HCC cell lines inhibited cell growth and injection of SH2D4A into nude mice reduced tumour incidence (Roessler et al., 2012). This evidence suggests a tumour suppressive role for *SH2D4A*, which could be potentially silenced by hypermethylation.

For the *ERN2*, a role in cancer remains to be elucidated. However, overexpression of *Ern2* in mice is associated with activation of the unfolded protein response, a stress response of the endoplasmic reticulum, which could also trigger apoptosis (Kaufman, 1999). This observation proposes a possible hypothesis whereby silencing expression of *ERN2* by a mechanism such as hypermethylation could be beneficial for survival of a tumour cell.

Chapter 6 – Biological significance of DNA methylation levels in the 85+ study participants

One of the key advantages of performing this analysis within the context of the 85+ study is that numerous other analyses were also being performed on the same or overlapping sample sets, thus allowing additional opportunities for assessing the potential biological significance of identified variation in methylation, both at a molecular and a clinical level.

6.1 Identification of candidate genes which may control the differential development of age related CpG island methylation.

6.1.1 SNPs and DNA methylation

In an attempt to uncover a possible explanation for variable DNA methylations levels in this population, the role of genetic variation was assessed by integration of SNP genotyping data and methylation data to determine which, if any, SNPs or genomic regions correlate with DNA methylation levels in the Newcastle 85+ study participants. The primary purpose of this analysis was to identify candidate genes which may control global levels of CpG island methylation.

In order to examine the possibility that allelic variants caused by SNPs are associated with particular traits in the Newcastle 85+ study, a genome wide association study (GWAS) analysis was undertaken by Dr Kristin Ayers and Dr Heather Cordell.

Genotyping of 733,202 SNPs in 642 Newcastle 85+ study participants was performed using an Illumina OmniExpress array at the Estonian Genome Center, University of Tartu, Estonia. This was combined with DNA methylation data set, providing a total of 411 participants with both measurements available.

Since methylation measurements at *TWIST2, HOXD4, TUSC3, EPHA10* and *HAND2* are correlated with each other, and levels at these genes seem to act as a sensor for methylation changes at multiple loci, the overall mean methylation of these five genes was used to represent a single DNA methylation value for each participant, which was then correlated with the SNP data.

Since only a fraction of all potential SNPs were assayed by genotyping in this analysis, information about additional SNPs was inferred using the presence linkage disequilibrium (LD) within the local region, using the International Haplotype Map Project (HapMap) as a reference data set, as it contains a large number of genotyped SNPs. This enabled the likely alleles of SNPs which were not directly genotyped to be determined and these are referred to as imputed SNPs.

The large number of variables in GWAS can lead to the identification of false relationships due to multiple testing. Because of this, previous GWAS analyses have typically applied a very stringent level of significance for associations with genotype at around 5×10^{-8} (corrected from $p<0.05$) (Risch and Merikangas, 1996; Pearson and Manolio, 2008; Johnson *et al.*, 2010), although the exact threshold varies depending on the exact number of SNPs in an individual study. However, for this analysis a less strict level of significance was used ($p<1 \times 10^{-5}$). This allowed a preliminary identification of SNPs which were suggested to be associated with DNA methylation levels, whilst reducing the likelihood of identifying false associations. Further validation of such associations will be required

6.1.2 Relationship between SNPs and DNA methylation levels – results

The negative logarithm of the significance level for associations between both genotyped and imputed SNPs and levels of DNA methylation was calculated and are displayed using a Manhattan plot (Figure 6.1). Strongly related SNPs have the smallest p values and therefore highest negative logarithm of the p value.

The results showed that several genotyped SNPs on chromosomes 1, 4, 7, 11, 14 and 21 and imputed SNPs on chromosomes 1, 4, 6, 7, 11, 14 and 21, may be associated with DNA methylation levels with genome wide significance. Table 6.1 details the position and allele frequencies of identified SNPs.

Next, in order to further characterise the genomic location of methylation-associated SNPs, the LocusZoom tool (URL http://csg.sph.umich.edu/locuszoom/) (Pruim *et al.*, 2010) was utilised to identify genes within the region. The genotyped SNP with the smallest p-value on chromosomes 1, 4, 11, 14 and 21 was designated as the central SNP

from which linkage disequilibrium of other SNPs was determined (r^2 value). The SNPs on chromosomes 1, 4 and 14 were not located either within or nearby to genes (>50kb away from nearest gene) (Figure 6.2). Interestingly, SNPs located on chromosome 11 and 21 were both positioned within the 3' end of the genes, Down syndrome cell adhesion molecule like 1 (*DSCAML1*) (Figure 6.3) and Down syndrome cell adhesion molecule (*DSCAM*) (Figure 6.4) respectively. Further investigation of these two loci revealed that they are highly related. An analysis of protein similarity using NCBI BLAST (Basic Local Alignment Search Tool) revealed a homologous protein structure between DSCAM and DSCAML1 (59% similarity) with no other strongly similar proteins (maximum similarity was only 33% over relatively small portions of DSCAM). The high level of homology between these two proteins suggests that they will likely have similar functions.

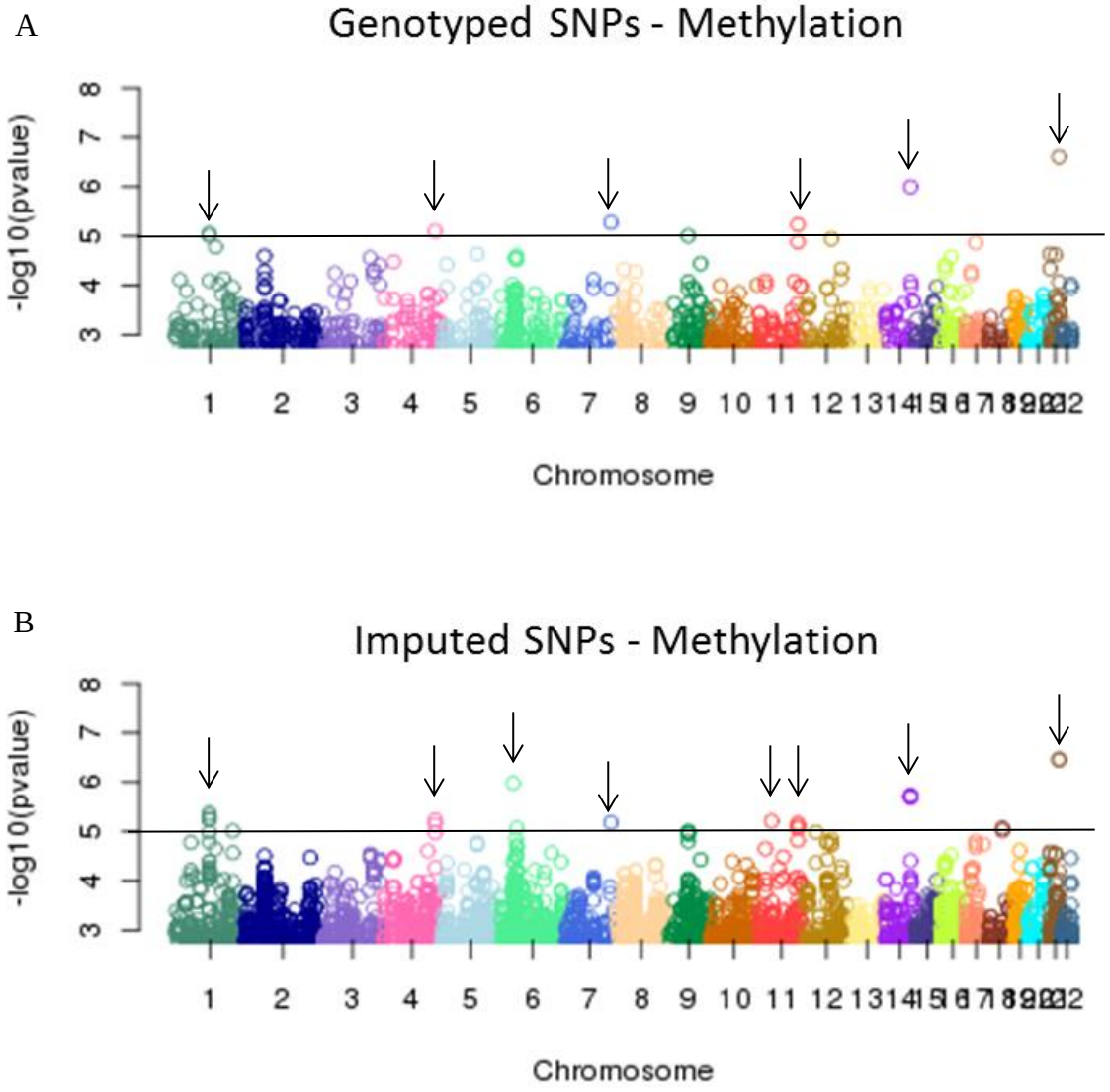


Figure 6.1: Manhattan plots displaying significance of associations between individual SNPs and DNA methylation levels in the Newcastle 85+ study

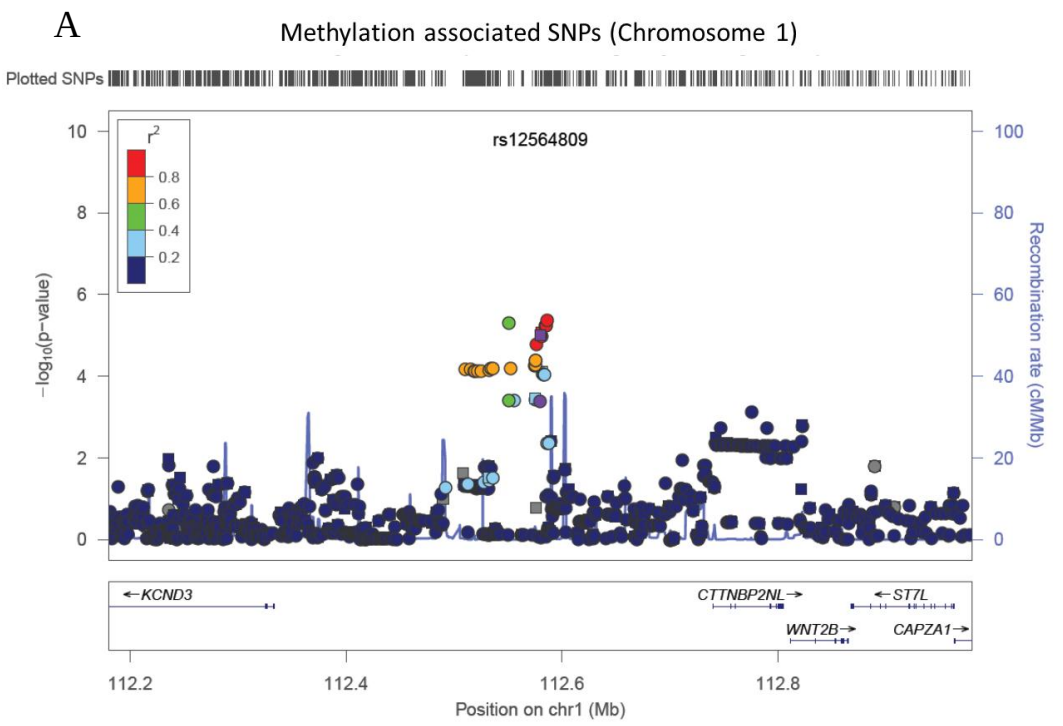
The plots show the association between methylation and both genotyped SNPs (A) and imputed SNPs (B). Genomic location is displayed along the X-axis and negative logarithm of the p value for the association with methylation at each SNP shown on the Y-axis. SNPs showing potentially significant association are indicated by an arrow (p-value>negative logarithm 1×10^{-5}).

Chr.	rsID	Position	A	B	Frequency (A)	p-value	Genotyped
1	rs1116438	112551210	C	T	0.316	4.84E-06	No
	rs12564809	**112580127**	**G**	**A**	**0.292**	**9.72E-06**	**Yes**
	rs12033004	112581191	G	A	0.293	8.64E-06	Yes
	rs2788420	112585128	A	G	0.293	5.69E-06	No
	rs2119192	112585504	A	C	0.293	5.68E-06	No
	rs7534975	112586441	C	T	0.297	4.19E-06	No
4	rs7669062	175328361	A	G	0.332	7.15E-06	No
	rs13102070	175329014	C	T	0.334	5.76E-06	No
	rs7657652	**175330825**	**C**	**A**	**0.347**	**7.80E-06**	**Yes**
6	rs708017	37104751	C	G	0.460	8.43E-06	No
7	rs10237793	146069234	A	G	0.022	5.22E-06	Yes
11	rs11216542	117155935	C	T	0.469	7.71E-06	No
	rs7940750	**117158560**	**A**	**G**	**0.402**	**5.76E-06**	**Yes**
	rs7124184	117160273	C	T	0.390	8.52E-06	No
14	rs8020379	93385069	A	G	0.394	1.99E-06	No
	rs8021173	93385453	C	T	0.394	1.93E-06	No
	rs11160147	93386595	A	G	0.406	1.87E-06	No
	rs1956661	**93388655**	**G**	**A**	**0.396**	**9.96E-07**	**Yes**
	rs753642	93390339	C	T	0.394	1.93E-06	No
	rs910784	94964112	A	G	0.458	1.02E-04	No
	rs1187751	94964331	A	G	0.458	1.04E-04	Yes
21	rs1001972	41069632	C	G	0.190	3.36E-07	No
	rs1571724	**41070706**	**G**	**A**	**0.191**	**2.45E-07**	**Yes**

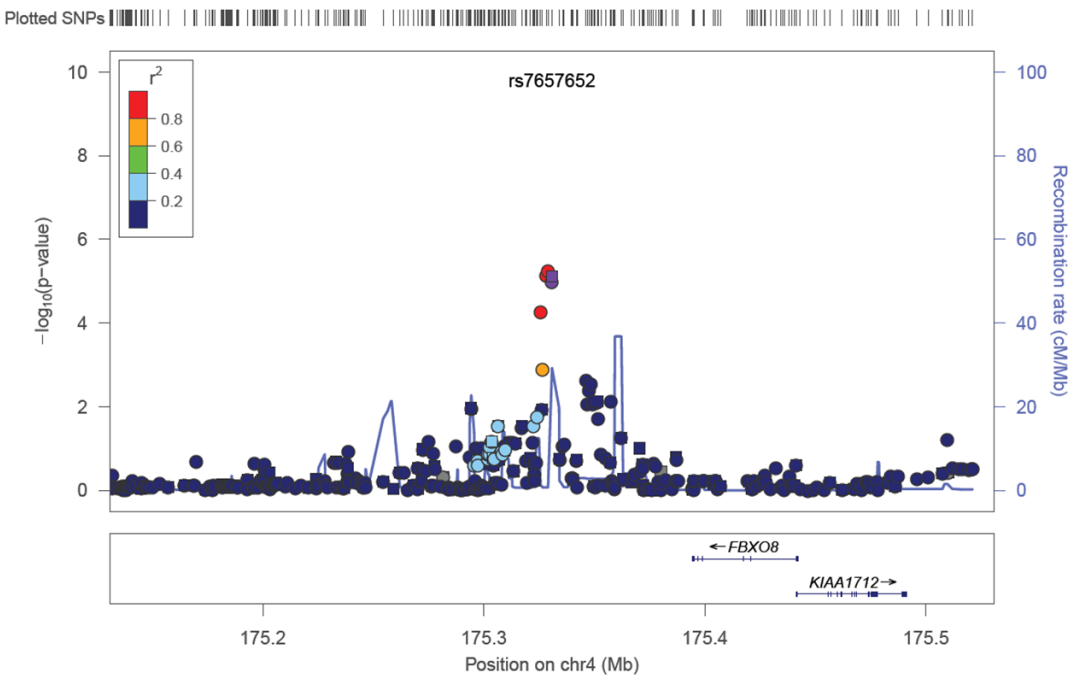
Table 6.1: GWAS identification of SNPs potentially associated with DNA methylation levels in the Newcastle 85+ study participants.

For each SNP, the chromosomal position, allelic variants, minor allele frequency and p-value for the association with DNA methylation level are shown. Information about non-genotyped (imputed) SNPs was inferred by the presence of LD within the local region. The local genomic regions of genotyped SNPs highlighted in bold font were investigated. Chr. = Chromosome.

A


Methylation associated SNPs (Chromosome 1)
Plotted SNPs
rs12564809
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
KCND3
CTTNBP2NL
ST7L
WNT2B
CAPZA1
112.2
112.4
112.6
112.8
Position on chr1 (Mb)

B


Methylation associated SNPs (Chromosome 4)
Plotted SNPs
rs7657652
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
FBXO8
KIAA1712
175.2
175.3
175.4
175.5
Position on chr4 (Mb)

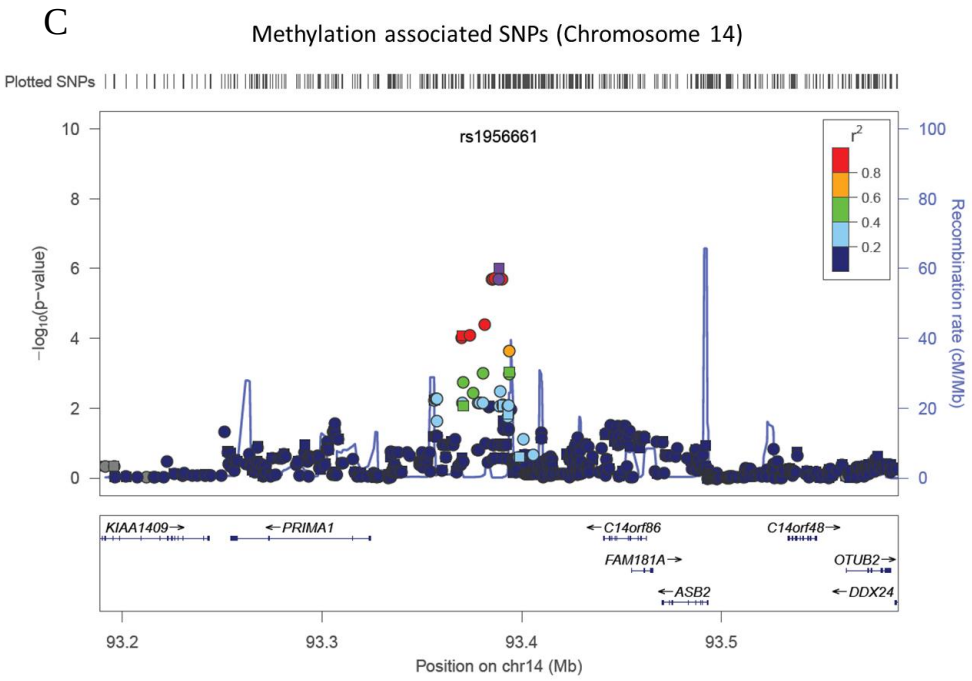


Figure 6.2: LocusZoom plots of DNA methylation associated SNPs on chromosomes 1, 4 and 14

The genotyped SNP with the smallest p value was allocated as the central SNP (purple square) from which linkage disequilibrium was calculated (r^2 value). Genotyped SNPs are represented by squares and imputed SNPs by circles. SNPs on chromosome 1 (A), chromosome 4 (B) and chromosome 14 (C) were not located nearby to genes (>50kb distance from nearest gene).

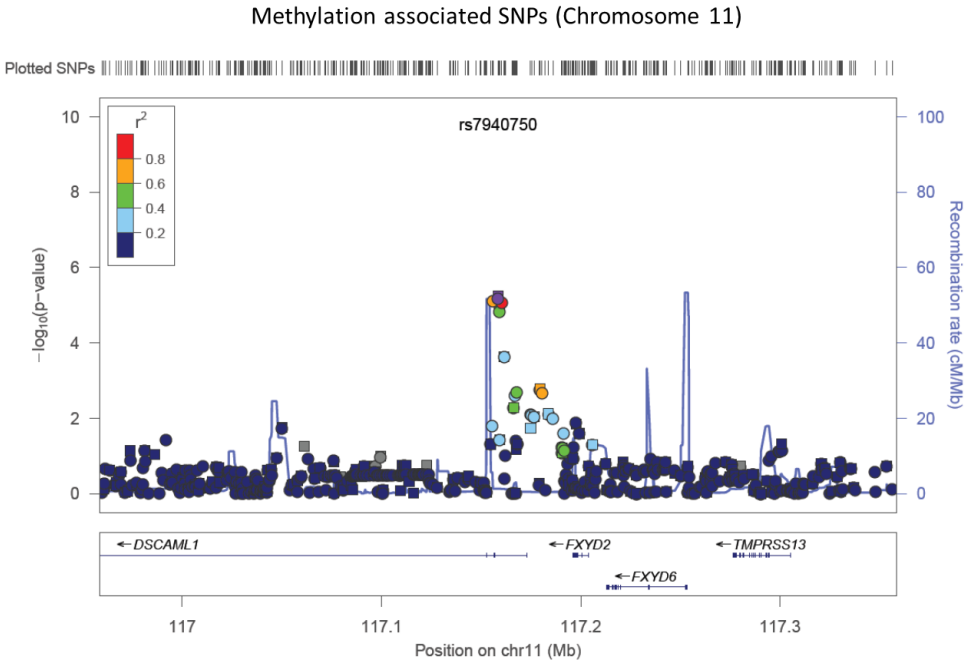


Figure 6.3: LocusZoom plot of methylation associated SNPs on chromosome 11

The genotyped SNP with the smallest p value was allocated as the central SNP (purple square) from which linkage disequilibrium was calculated (r^2 value). Genotyped SNPs are represented by squares and imputed SNPs by circles. SNPs were located within the *DSCAML1* gene.

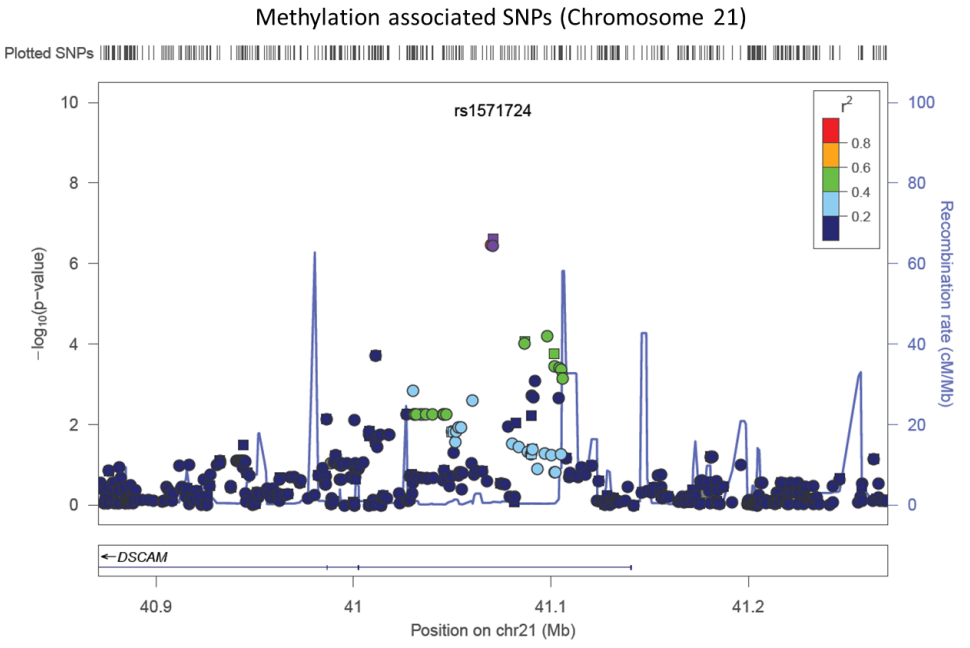


Figure 6.4: LocusZoom plot of methylation associated SNPs on chromosome 21

The genotyped SNP with the smallest p value was allocated as the central SNP (purple square) from which linkage disequilibrium was calculated (r^2 value). Genotyped SNPs are represented by squares and imputed SNPs by circles. SNPs were located within the *DSCAM* gene.

6.1.3 Discussion of the association between SNPs and DNA methylation levels

A recent analysis of human cell lines examined the differential methylation status of specific alleles caused by the presence of a heterozygous SNP by bisulfite sequencing. There were two categories of SNP identified by this study. Firstly were those which affected methylation only at the site which was overlapping the SNP, since the nucleotide change resulted in a loss of potential to methylate that particular site. This appeared to be the most frequent consequence of SNPs. Secondly were examples of SNPs which despite only affecting a single site, were associated with similar pattern of methylation at nearby CpG sites, suggesting these SNPs may be located in regions involved in the regulation of DNA methylation levels (Shoemaker *et al.,* 2010), possibly affecting CpG sites within 2Kb of the most strongly SNP-associated CpG site (Bell *et al.*, 2011).

Alternatively, SNPs may cause differential expression of a gene or genes involved in determining methylation patterns. This would have consequences for methylation levels at loci throughout the genome as opposed to only those genes with a nearby SNP. An examination of approximately 3 million SNPs and 27,000 CpG sites in lymphoblastoid cell lines from HapMap individuals found that genotype associated methylation was mostly due to nearby SNPs within 50Kb. A SNP (rs10876043) located in *DIP2B,* a gene previously implicated in DNA methylation as it contains binding sites for DNA methyltransferase 1 – associated protein 1, which in turn interacts with DNMT1, was seemingly related to overall DNA methylation levels; however this was a fairly weak association (Bell *et al.,* 2011).

SNPs in two genes, *DSCAM* and *DSCAM1* were identified from the GWAS as being potentially associated with DNA methylation levels in the Newcastle 85+ study population. Whilst the role of proteins involved in cell adhesion in determining DNA methylation patterns is not clear, the fact that the only two SNPs within genes which were significantly associated with methylation levels were located in two highly homologous genes of the same family suggests that this observation may well be functionally significant, although confirmatory studies in other sample sets will be required.

Additionally, there is evidence that the *DSCAM* enhancer is bound by FOXA1 (Carroll *et al.*, 2005). FOXA1 is a forkhead DNA binding protein and has been recently identified in our lab as a possible candidate for mediating the altered DNA methylation patterns observed in undifferentiated and differentiated cells (Strathdee and Thathia, unpublished data).

6.2 Association between mitochondrial haplogroups and DNA methylation levels in the Newcastle 85+ study participants

6.2.1 Mitochondria, reactive oxygen species and DNA methylation

Reactive oxygen species (ROS) are often a by-product of normal cellular processes, including mitochondrial oxidative phosphorylation (Cooke *et al.*, 2003). However, ROS can damage cellular macromolecules including DNA, triggering a DNA damage response which is typically repaired by base excision repair (Cooke *et al.*, 2003).

Several publications have suggested that DNA methylation patterns could be influenced by oxidative stress (reviewed in (Franco *et al.*, 2008)). A recent study described a possible mechanism for this association. Treatment of cell lines with hydrogen peroxide to induce oxidative damage led to the formation of large multi-protein silencing complexes containing DNMT1, DNMT3B, SIRT1 and members of the PRC4, which are recruited to chromatin at sites of oxidative DNA damage, particularly those in promoter associated CpG islands (O'Hagan *et al.*, 2011). A reduction in the active chromatin mark (H3K4me3), increased DNA methylation specifically at genes containing CpG island promoters and reduced expression was detected. Additionally, a link was observed between the genes which exhibit this oxidative damage response and those susceptible to hypermethylation and silencing in cancer (O'Hagan *et al.*, 2011). This study suggested that different levels of exposure to oxidative stress could mediate differential DNA methylation levels.

Mitochondrial haplogroups reflect SNP differences in mitochondrial DNA (mtDNA) sequences which represent different lineages from a single common ancestor and there is evidence linking mtDNA variants with ageing, longevity and disease (De Benedictis *et al.*, 1999; Takasaki, 2009; Nishigaki *et al.*, 2010), however the mechanisms underlying such relationships was not determined.

It is thought that mtDNA haplotype may influence mitochondrial function via effects on oxidative phosphorylation capability. For example, haplogroup J has been suggested to have lower oxygen consumption and ATP production than haplogroup H (Gomez-Duran *et al.*, 2012). Analysis suggests that individuals with haplogroup H have the greatest maximum oxygen consumption and highest levels of mitochondrial oxidative damage, a sign of increased ROS production (Martinez-Redondo *et al.*, 2010).

The possibility that the mtDNA genome can influence DNA methylation of the nuclear genome was emphasised by the disrupted DNA methylation patterns observed in Rho^0 cells, which lack a mitochondrial genome. Such changes could be partially reversed by the replacing mtDNA in these cells (Smiraglia *et al.*, 2008), which suggests that mtDNA is somehow involved in regulating epigenetic mechanisms.

As both the mtDNA genome and oxidative stress have been linked to the establishment of DNA methylation, and differing mtDNA haplogroups appear to exhibit functional differences, it is possible that the differing ROS levels generated by specific haplogroups can lead to differential DNA methylation levels between individuals.

6.2.2 Association between mitochondrial haplogroups and DNA methylation levels – results

In order to investigate the association between DNA methylation levels and mtDNA, DNA methylation measurements from the Newcastle 85+ study participants were compared dependant on their mtDNA haplogroup. MtDNA haplogroups were determined by Dr Angela Pyle and were available for 444 Newcastle 85+ study participants whose DNA methylation levels were known. Average overall DNA methylation measurements of participants sharing the same haplogroups are shown in Table 6.2.

A step-down linear regression analysis model was constructed using non-lymphocyte corrected data from all 480 participants tested, to identify any mtDNA haplogroup which had significantly different DNA methylation levels at *HOXD4, TUSC3, TWIST2, HAND2, EPHA10* and the overall mean of these 5 genes. Other variables included in the model were lymphocyte percentage and gender. Non-significant associations were

removed (exclusion criteria $p>0.10$) and only mtDNA haplogroups significantly associated with DNA methylation levels were retained (retention criteria $p<0.05$).

There were no significant associations between overall DNA methylation level and any of the mtDNA haplogroups remaining in the step-down model, only the proportion of lymphocytes and gender were related to DNA methylation levels, which was expected based on previous results (Table 6.3).

MtDNA Haplogroup	Number of individuals	*TWIST2*	*TUSC3*	*HOXD4*	*HAND2*	*EPHA10*	Overall mean (%)
H	196	10.75	10.23	10.43	10.57	10.48	10.50
I	8	10.40	9.27	11.01	9.38	9.5	9.92
J	47	9.78	10.80	10.59	10.45	9.98	10.32
K	29	10.76	10.75	10.57	10.31	11.2	10.73
T	51	9.37	9.49	10.08	9.72	10.57	9.84
U	70	11.02	11.20	10.77	10.56	10.88	10.90
V	17	11.55	11.13	11.91	12.02	11.33	11.58
W	14	10.44	11.71	10.87	11.03	10.30	10.87
X	12	8.98	9.96	9.62	9.25	9.56	9.48
No haplogroup	36						

Table 6.2: Mitochondrial DNA haplogroups and DNA methylation levels in the Newcastle 85+ study participants

DNA methylation levels of *TWIST2*, *TUSC3*, *HOXD4*, *EPHA10*, *HAND2* and the overall mean in Newcastle 85+ study participants with different mtDNA haplogroups.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	6.818	.387		17.620	.000
	Lymphocytes	.141	.013	.466	10.859	.000
	Gender	-.451	.248	-.078	-1.817	.070

a. Dependent Variable: OVERALL_MEAN

Table 6.3: Significant associations with overall mean DNA methylation levels from the step-down linear regression model analysis of mtDNA haplogroups.

None of the mtDNA haplogroups were associated with overall DNA methylation levels. Both gender and lymphocyte percentage was related to levels of DNA methylation which was expected based on previous analysis. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

In addition, the relationship between mtDNA haplogroup and DNA methylation levels of each of the genes was tested individually by the linear regression analysis model. Added to the expected link with lymphocytes and gender, individuals with haplogroup T had significantly lower levels of DNA methylation at *TUSC3* ($p=0.049$) (Table 6.4). Individuals with the mtDNA haplogroup V demonstrated significantly higher methylation at the *HAND2* locus ($p=0.031$) (Table 6.5). At all other loci there were no associations between DNA methylations levels and mtDNA haplogroup.

Based on such results however, there is little evidence to support the hypothesis that specific mtDNA haplogroups may exhibit differential methylation patterns in the Newcastle 85+ study population, at least for the set of loci analysed in this study.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	5.989	.544		11.012	.000
	Lymphocytes	.185	.018	.441	10.174	.000
	Gender	-.892	.347	-.112	-2.575	.010
	hap_T	-1.041	.528	-.084	-1.972	.049

a. Dependent Variable: TUSC3

Table 6.4: Significant associations with *TUSC3* methylation levels from the step-down linear regression model analysis of mtDNA haplogroups

The mtDNA haplogroup T was associated with lower levels of *TUSC3* DNA methylation. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
9	(Constant)	8.235	.458		17.965	.000
	Lymphocytes	.074	.016	.219	4.746	.000
	hap_V	1.647	.762	.100	2.161	.031

a. Dependent Variable: HAND2

Table 6.5: Significant associations with *HAND2* methylation levels from the step-down linear regression model analysis of mtDNA haplogroups

The mtDNA haplogroup V was associated with increased *HAND2* DNA methylation. The magnitude of the effect on DNA methylation levels is indicated by the coefficient B. Significant associations ($p<0.05$).

6.2.3 Discussion of the association between mitochondrial haplogroups and DNA methylation levels

The results of the linear regression analysis suggested that mitochondrial haplogroups are unlikely to be associated with the highly variable DNA methylation patterns observed in the Newcastle 85+ study population. A link was suggested between altered DNA methylation levels of two specific loci, *TUSC3* and *HAND2* with the T and V mtDNA haplotype respectively, however, these associations were not strongly significant, and the same significance could not be achieved at other loci. Participants with the V haplogroup actually had increased levels of methylation at all five loci compared to participants in all other haplogroups (Table 6.2), but the very small number of individuals with this haplogroup (n=17), make it hard to assess the potential significance of this observation. Thus while this study presents little evidence of a link to mitochondrial haplotypes, a link between methylation and some of the rarer haplotypes, particularly the V haplotype, cannot be excluded.

A recent study examining levels of global methylation in peripheral blood from elderly individuals with different mitochondrial haplogroups, found that people with the J haplogroup appeared to have a significantly higher global level of DNA methylation than those individuals with other haplogroups (Bellizzi *et al.*, 2012a). Analysis was also performed by *in vitro* experiments using cybrid cells, containing identical nuclear genomic DNA but mtDNA from different haplogroups (H, J, U, X and T). Methylation levels were variable between the haplogroups and once again, the J haplogroup was associated with the highest levels. The J cybrids also exhibited lower levels of ROS, ATP and increased expression of *MAT1A*, a gene involved in the synthesis of SAM, This suggests that increased SAM availability may be a possible mechanism for the higher DNA methylation status linked with carriers of this mtDNA variant (Bellizzi *et al.*, 2012a). We did not observe the same trend for high DNA methylation associated with the J haplogroup in the specific genes examined in Newcastle 85+ study participants.

6.3 Association between frailty and DNA methylation levels in the Newcastle 85+ study participants

6.3.1 Frailty and DNA methylation

Frailty encompasses a multitude of symptoms associated with declining function and increasing vulnerability to disability and death in the elderly (Weiss, 2011). There is no universally accepted method to define frailty, and many means of classifying frailty have been suggested. Two of the foremost methods applied to characterise frailty in elderly individuals are those proposed by Fried and Rockwood.

Fried applies a fairly simplistic estimation of frailty considering few variables. Frailty is defined as a clinical syndrome comprised of three or more of the following characteristics: weight loss, exhaustion, low physical activity, muscle weakness or slow walking speed (Fried *et al.*, 2001).

The Rockwood approach is more focused on an accumulation of health defects, measured by the presence of disease-related and clinical symptoms and abnormal clinical test results (Rockwood *et al.*, 2005). A standardised protocol was developed to determine a consistent measure of frailty, the Rockwood frailty index (RFI) (Searle *et al.*, 2008).

In an attempt to uncover the consequences of inter-individual variability of DNA methylation in the elderly, a recent study examined the relationship between global DNA methylation levels and the presence of frailty in 318 individuals aged 65-105 years (Bellizzi *et al.*, 2012b). Frailty was assessed using a cluster analysis applied by Montesanto et al (2010), using a combination of physical (activities of daily living, hand grip strength), cognitive (mini mental state exam (MMSE) score) and psychological (depression) markers to develop a frailty phenotype. Individuals were categorised as frail, pre-frail or non-frail (Montesanto *et al.*, 2010). Frail individuals had significantly decreased levels of global DNA methylation compared to those who were pre-frail or non-frail. In addition, progressively deteriorating frailty after a 7 year follow up was associated with further reductions in global DNA methylation (Bellizzi *et al.*, 2012b). This suggests a link between altered DNA methylation and functional decline in the

elderly; however the importance of global DNA methylation changes is largely uncertain.

6.3.2 Association between frailty and DNA methylation levels – results

Therefore, an analysis investigating the association between gene specific DNA methylation levels (which have a role in gene regulation) and frailty in the Newcastle 85+ study population was undertaken. Analysis of frailty in the Newcastle 85+ study population using both the Rockwood and Fried approaches was performed by Dr Joanna Collerton.

For the Newcastle 85+ study participants, Fried frailty characteristics were determined by BMI measurement (<18.5, underweight), questionnaires (negative responses to questions regarding energy levels), hand grip strength and a timed "up and go" test (rise from chair, walk short distance). Participants with stroke, Parkinson's disease or a MMSE score <18 were excluded, as these disease states may affect frailty variables. The Fried frailty status (FFS) for each individual was calculated as: 3 or more characteristics = frail, 1-2 = pre-frail, 0 = non-frail. A total of 321 Newcastle 85+ study participants had both DNA methylation data and FFS measurement.

For Rockwood frailty, a total of 40 deficit variables were measured individually in the Newcastle 85+ study population with scores ranging from 0 (no deficit) to 1 (full deficit) and partial deficiencies scored within this range. The RFI represents the score of combined deficits divided by total deficits measured (e.g. 20 deficits present, 20/40 = 0.5 RFI). Examples of deficit variables included activities of daily living (e.g. washing, feeding), presence of disease (e.g. CVD, cancer), geriatric symptoms (e.g. incontinence, impaired hearing), cognitive function and other symptoms (chronic pain, depression). A total of 471 participants had both RFI scores and DNA methylation measurements available.

6.3.2.1. Relationship between FFS and DNA methylation levels

An ordinal linear regression model was utilised to analyse the possible relationship between FFS and DNA methylation levels in the Newcastle 85+ participants. FFS was the dependent variable, overall DNA methylation was the independent variable and the

output was an odds ratio representing the risk of being in a frailer FFS category. The model was adjusted for the effects of gender and percentage of lymphocytes, since these variables are known to influence DNA methylation measurements (model 1). A second model was adjusted for gender, percentage of lymphocytes, smoking behaviour, education and disease count (model 2). DNA methylation values were either categorised into quartiles, or log transformed methylation values used as a continuous variable.

The results show evidence of an association between DNA methylation levels and frailty as determine by Fried using both the adjusted ordinal linear regression models (Table 6.6). When the lowest quartile was used as a reference category, the calculated odds ratios demonstrate that individuals in higher quartiles (increased methylation) had a greater risk of being in a frailer FFS category, although this was only statistically significant for the 2^{nd} and 4^{th} quartiles. Similarly, when the middle 2 quartiles were used as a reference, those participants with the lowest quartile of DNA methylation levels had a significantly reduced risk of being in a more frail category. With DNA methylation as a continuous variable, there was also a significant association between increasing level of DNA methylation and worsening frailty in model 1, although this was just outside the level of significance for model 2.

Overall, this analysis suggests that in the Newcastle 85+ study, elderly individuals with increased DNA methylation levels at the specific gene loci investigated are more likely to be in a frailer category as classified by FFS, than those with lower levels of DNA methylation.

6.3.2.2. Relationship between RFI and DNA methylation levels

Since RFI is measured as a continuous variable, the relationship between Rockwood frailty scores and DNA methylation levels were firstly examined by Spearman's rank correlation. This revealed no association (Spearman's rho -0.038; p-value 0.415).

Next, a linear regression model was constructed with RFI values, subjected to square root transformation in order to best fit the model, as the dependent variable. As for previous analysis, DNA methylation either as quartiles or a continuous variable was the independent variable. The outcome of the model, the unstandardized regression coefficient, represents the difference between the mean square root RFI for a particular

category compared to that of the reference category. The model was adjusted for the effects of gender and percentage of lymphocytes. The results revealed no significant association between DNA methylation levels and frailty as classified by Rockwood frailty scores (Table 6.7), although there was a trend towards increased frailty in those individuals with higher level of DNA methylation.

Overall DNA methylation (%)		Model 1 odds ratio (95% CI)	p-value	Model 2 odds ratio (95% CI)	p-value
Quartiles – lowest quartile as reference	≤ 8.50	Reference		Reference	
	8.51-10.14	**2.39 (1.27 to 4.50)**	**0.007***	**2.42 (1.24 to 4.74)**	**0.010***
	10.15-11.95	1.83 (0.94 to 3.56)	0.076	1.86 (0.91 to 3.79)	0.090
	≥ 11.96	**3.32 (1.55 to 7.12)**	**0.002***	**3.17 (1.38 to 7.26)**	**0.006***
Quartiles – middle quartiles as reference	≤ 8.50	**0.47 (0.27 to 0.83)**	**0.009***	**0.46 (0.25 to 0.85)**	**0.013***
	8.51-11.95	Reference		Reference	
	≥ 11.96	1.63 (0.89 to 2.97)	0.113	1.54 (0.80 to 2.95)	0.198
Continuous variable		**3.19 (1.06 to 9.62)**	**0.039***	3.21 (0.98 to 10.51)	0.054

Table 6.6: Ordinal linear regression model analysis of FFS and DNA methylation

Odds ratios represent the risk of being in a frailer FFS category. Model 1 was adjusted for gender and percentage of lymphocytes, model 2 was adjusted for gender, percentage of lymphocytes, smoking behaviour, education and disease count. Significant associations *($p<0.05$)

Overall DNA methylation (%)		Unstandardized regression coefficient (95% CI)	p-value
Quartiles – lowest quartile as reference	≤ 8.50	Reference	
	8.51-10.14	0.009 (-0.022 to 0.041)	0.564
	10.15-11.95	0.033 (0.000 to 0.067)	0.052
	≥ 11.96	0.033 (-0.004 to 0.070)	0.077
Quartiles – middle quartiles as reference	≤ 8.50	-0.020 (-0.048 to 0.009)	0.169
	8.51-11.95	Reference	
	≥ 11.96	0.010 (-0.019 to 0.040)	0.490
Continuous variable		0.003 (-0.002 to 0.008)	0.245

Table 6.7: Linear regression model analysis of RFI and DNA methylation

The unstandardized regression coefficient, represents the difference between the mean square root RFI for a particular category compared to that of the reference category. The model was adjusted for gender and percentage of lymphocytes. There were no significant associations ($p<0.05$)

6.3.3 Discussion of the association between frailty and DNA methylation levels

The results of this analysis suggested that DNA methylation levels were significantly increased in Newcastle 85+ participants who were categorized as being more frail by the definition proposed by Fried. However, there was no significant relationship using the Rockwood definition of frailty. This suggests that the approach used to define frailty is of great importance and investigations of frailty in relation to other biomarkers could either miss associations or lack reproducibility if different definitions are used.

Although there was a significant association between DNA methylation level and Fried frailty score, but not between methylation and the Rockwood frailty index, this may be due to the different sample sets used for both analyses. A total of 150 Newcastle 85+ study participants were excluded from the Fried frailty analysis due to certain pre-existing medical conditions such as stroke, Parkinson's disease or dementia. An additional analysis could be performed to examine the relationship between DNA methylation and Rockwood frailty index in the same participants who were included in the Fried analysis. If this analysis resulted in a significant association, it would indicate that DNA methylation is associated with frailty in people without stroke, Parkinson or dementia.

An additional problem is that a number of criteria often used to define frailty are self-reported, for example using questionnaires meaning responses can be subjective, making it more difficult to establish consistent associations with other biomarkers.

From this investigation, it is impossible to decipher causation i.e. whether individuals with higher DNA methylation are more susceptible to becoming frail, or whether DNA methylation changes are induced subsequent to this. However, it is certainly possible that increased DNA methylation may silence gene expression in a number of cells to a level which is detrimental to normal tissue functioning, affecting multiple systems and this manifests itself as frailty.

To my knowledge there is only one previous study detailing a link between frailty and global DNA methylation (Bellizzi *et al.*, 2012b). However, the method applied by this study, which depends on HpaII digestion, only assessed a small fraction of all CpG sites. In addition, the results seem to indicate extremely variable levels of global

methylation, which seem unlikely to be biologically accurate. Previous studies of global DNA methylation levels using HPLC have identified much lower amounts of variability in the normal population (Fuke *et al.*, 2004). Therefore, it would be interesting to attempt to confirm the Bellizzi *et al.* findings using an alternative assay for global methylation and the Fried and Rockwood definitions for frail individuals in the Newcastle 85+ study population. Ageing is typically associated with increased methylation at CpG islands within a background of reduced global methylation and so the direction of the associations identified; that is increased frailty in those with reduced global methylation (Bellizzi *et al.*, 2012b) and increased CpG island methylation (this study) is consistent with the idea that increased disruption of methylation patterns is associated with detrimental effects on healthy ageing.

Chapter 7 – Final Discussion

The role of DNA methylation in the regulation of gene expression is well established and the silencing of tumour suppressor genes by hypermethylation is now known to be a major mechanism behind the altered gene expression profiles evident in tumour cells (Esteller, 2007). It is also apparent that similar DNA methylation changes are accrued with increasing age but the consequences are not well understood, since such changes are not as dramatic as those observed in cancer.

Methylation changes are typically less widespread during ageing, as some genes appear to be uniquely methylated in cancer tissue (Toyota *et al.*, 1999). In addition, the extent of methylation is often lower, since genes are often completely methylated in cancer tissues. The underlying cause of such differences is likely to be differences in cellular composition between tumours and ageing tissue. Tumours are typically composed of a clonally expanded cell (monoclonal), therefore if the original cancer stem cell had a particular tumour suppressor gene inactivated by methylation, then all progeny cells will similarly by methylated at that site. Consequently, a quantitative measure of methylation at that site in tumour DNA will reveal 100% methylation.

Ageing tissues are derived from many different stem cells, and so a sample of cells from the tissue represents many genomes with differing patterns of methylation. Such patterns are likely to be the result of effects of "damage" and stochastic methylation changes. Detection of 20% methylation at a tumour suppressor gene in ageing tissues means that around 20% of cells are methylated at that site. The number of cells in which expression of the gene is silenced by hypermethylation may reach a level which is detrimental for normal tissue function without being completely methylated in all cells.

According to the Office for National Statistics (ONS), the greatest increase in population growth over the next 20 years will be represented by individuals aged over 75 years old, particularly the 85 and over category (Figure 7.1). Therefore, identification of factors underlying health at an elderly age is of great importance to prevent an overwhelming burden on health service providers and ensure a good quality of life as the elderly population grows.

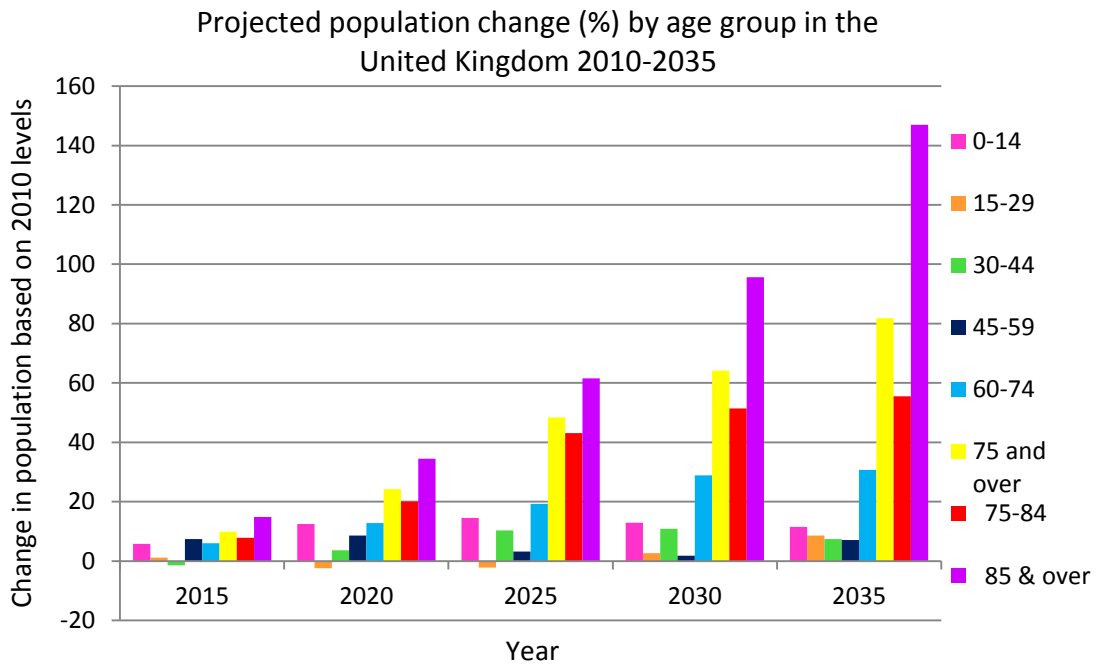


Figure 7.1: Projected population change by age group in the United Kingdom 2010-2035

For each age group and time point, change in population was calculated compared to 2010 population measurements. Greatest population growth is predicted in individuals aged over 75. Data obtained from the ONS National Population Projections, 2010-based projections (URL http://www.ons.gov.uk/ons/rel/npp/national-population-projections/2010-based-projections/index.html)

The Newcastle 85+ study is a unique resource as it is one of the largest longitudinal based studies of the very elderly available for analysis. In addition to the investigation of DNA methylation levels in this population, a number of other biomarkers of health status have been measured in an effort to identify the features underlying healthy ageing in this population of the oldest old.

In an attempt to more clearly understand age-related patterns of DNA methylation, we quantified methylation levels in normal PBLs from individuals aged 85 years, at the promoter region of genes either known to be hypermethylated in leukaemia, or exhibited age-related methylation increases in tissues rather than blood (See Chapter 3.1). These genes were not chosen due to a particular functional role but selected as likely candidates for demonstrating some degree of methylation in normal aged DNA, since an age-associated effect had already been demonstrated in other tissues, or they were susceptible to methylation increases in a malignancy of the same cell type. We

found that those genes which were known to be hypermethylated in leukaemia exhibited highly variable methylation levels in PBL DNA from individuals aged 85, suggesting that similar genes acquire methylation during ageing and cancer. However, we did not detect any appreciable level of DNA methylation of non-leukaemia associated genes in PBL DNA from the elderly, suggesting that these genes are only methylated in solid tissues during ageing.

7.1 Critique of the Pyrosequencing technique

The analysis of DNA methylation levels in the Newcastle 85+ study participants was performed by Pyrosequencing, which is one of the best techniques available for quantification of site specific DNA methylation, even allowing small methylation changes to be accurately measured. In addition, several consecutive CpG sites can be measured within the same region and there is no sequence bias for which sites can be analysed, unlike other techniques such as COBRA where only sites incorporated within a restriction enzyme cut site can be examined (Shen and Waterland, 2007). One disadvantage of using this method was that specific genes had to be targeted for analysis. It is not possible to look at large number of genes due to the time consuming nature of validating each assay before use. A genome wide approach such as methylation arrays would possibly have identified a greater number of candidate genes with altered methylation during ageing, but genome wide techniques are unlikely to be sensitive enough to detect the small inter-individual differences in DNA methylation which we were able to distinguish by Pyrosequencing in the Newcastle 85+ study (Shen and Waterland, 2007). Another disadvantage of Pyrosequencing is that information regarding allele specific DNA methylation cannot be gained, although this was overcome by performing sodium bisulfite sequencing of individual sequences to allow an assessment of allele specific methylation patterns in the Newcastle 85+ study participants.

7.2 Critique of Newcastle 85+ study samples

The Newcastle 85+ study is an excellent resource for studying people of a very elderly age. Presently, people who survive beyond age 85 are a minority but with increasing life expectancy, a greater proportion of people will reach 85 years of age and soon they may be the majority. However, as we want to know about changes during the ageing process,

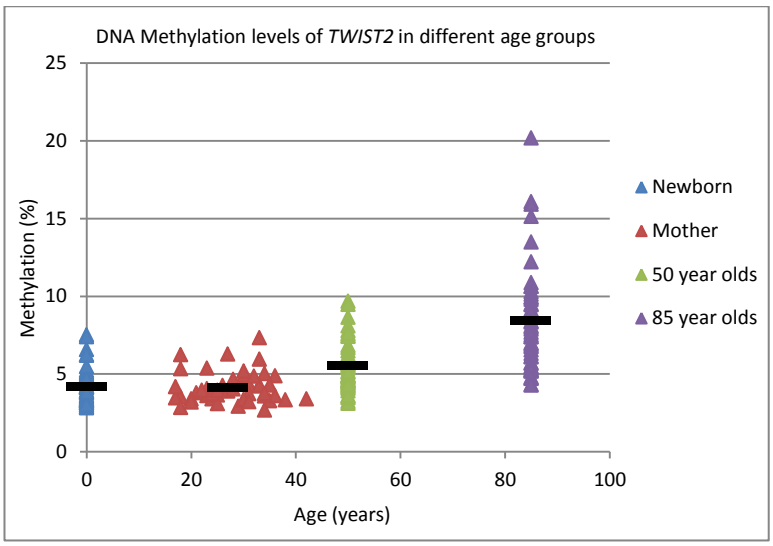
we want to ensure that such changes are evident throughout ageing, and not unique to this very elderly group of individuals. Therefore, we need to study people of different ages Thus, to test whether increased and variable DNA methylation levels were also evident in younger age groups, DNA methylation levels of *TWIST2, TUSC3, HOXD4, HAND2* and *EPHA10* were measured in samples of cord blood from newborns, PBLs from their mothers (aged 17-42) and a separate population of 50 year olds by Sanne van Otterdijk (Figure 7.2). The results revealed that at all loci; there is a clear increase in DNA methylation levels with advancing age although a significant amount of methylation can also be detected at birth. There is also a lot of variation in DNA methylation measurements within each age group; although it does become more pronounced with increasing age (Unpublished data from van Otterdijk, Mathers and Strathdee). These results suggest that increasing DNA methylation is evident all throughout life not just at very elderly ages.

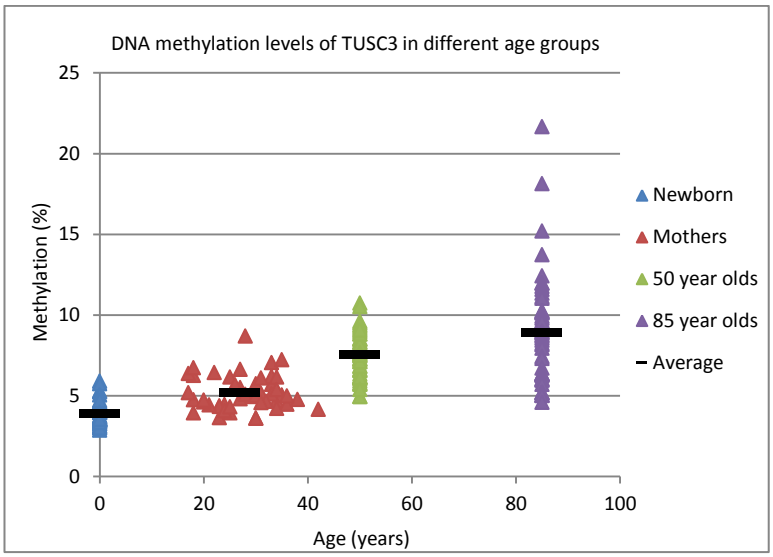
A recent study examined DNA methylation levels in peripheral blood at over 27,000 CpG sites in boys aged 3-17 and discovered over 2000 loci which exhibit age-related methylation changes in this young population. In addition, a comparison with adult populations (24-85 years) suggested a high degree of overlap between those loci which accrued methylation in both populations (Alisch *et al.*, 2012). Interestingly, there did not appear to be a direct linear accumulation of DNA methylation at these loci throughout life, suggesting that some regions of the genome may be particularly sensitive to changeable methylation during childhood or adolescence with smaller changes during adulthood. A similar finding was observed in an analysis of 300 CpG sites flanking the TSS of genes in mouse brain, muscle and kidney. DNA methylation changes occurred more rapidly in the period prior to adulthood, than during adulthood to old age (Takasugi, 2011).

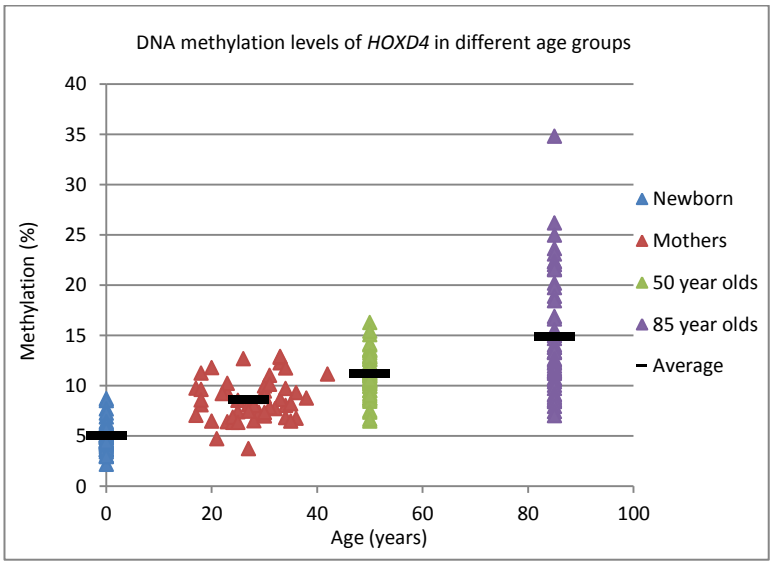
For the purposes of our analysis, the very old age of the Newcastle 85+ study participants was a benefit as they have had more time to accumulate DNA aberrant methylation meaning it is easier to detect. Secondly, the higher incidence of age-related disease in this population increased the ability to detect associations between disease and DNA methylation levels with lower participant numbers than would be required with younger cohorts. For example, as the incidence of cancer increases rapidly with advancing age (Cancer Research UK, URL http://info.cancerresearchuk.org/cancerstats/incidence/age/), it is likely that there would

be fewer incidences of cancer in younger aged population both prior to study recruitment and during the study follow up of three years.

For all Newcastle 85+ volunteers, PBL was used as a DNA source. This can complicate the interpretation of results due to the presence of different cell types within the sample. A recent genome wide analysis examined the methylation status over 450 thousand CpG sites in whole blood, as well as specific cell populations such as mononuclear cells, granulocytes, monocytes, neutrophils, eosinophils, natural killer (NK), B and T cells. This revealed extensive differential methylation of CpG sites between cell lineages; although the methylation states of cells types within the same lineage (i.e. lymphoid or myeloid cells) are most similar. Differential DNA methylation was particularly evident in genes with an immune-related cell type specific function (Reinius *et al.*, 2012). Our results should therefore be interpreted with caution, as the inter-individual differences in DNA methylation levels of specific genes which we observed in PBL DNA may be reflecting differences in NK, B and T cell numbers between individuals. We attempted to counteract such effects by accounting for the proportion of overall lymphocyte cells in the Newcastle 85+ study PBL DNA sample, but not specific types of lymphocyte. The isolation of specific cell types requires cell sorting typically by flow cytometry or using antibodies against cell surface markers, however samples suitable for such cell isolations were not available and analysing multiple samples from each individual would have significantly reduced the total number of individuals who could have been screened. Furthermore the study was based on the hypothesis that methylation levels in PBL samples would, at least in part, reflect methylation levels in other tissues (as this would be influenced by shared genetic and environmental factors). This would be especially likely to be the case for different cells in the same tissue (i.e. PBL). This is further supported by the relatively consistent effect of ageing on DNA methylation levels seen at different loci (see Figure 7.2). Additionally, examination of multiple cell types increases the likelihood of identifying universal changes in DNA methylation, rather than cell type specific alterations of DNA methylation.


DNA Methylation levels of *TWIST2* in different age groups
Methylation (%)
Age (years)
Newborn
Mother
50 year olds
85 year olds


DNA methylation levels of TUSC3 in different age groups
Methylation (%)
Age (years)
Newborn
Mothers
50 year olds
85 year olds
Average


DNA methylation levels of *HOXD4* in different age groups
Methylation (%)
Age (years)
Newborn
Mothers
50 year olds
85 year olds
Average

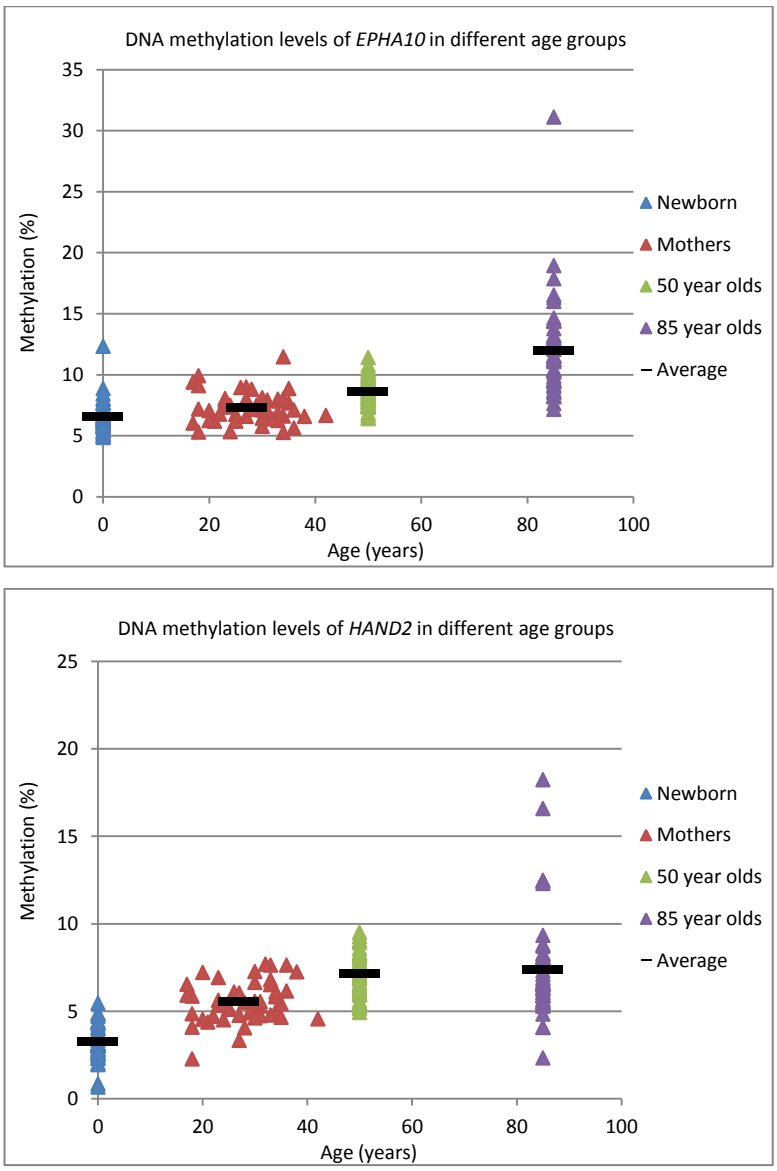


Figure 7.2: DNA methylation levels of *TWIST2, TUSC3, HOXD4, EPHA10* and *HAND2* in different age groups

DNA methylation in the promoter region of *TWIST2*, *TUSC3*, *HOXD4*, *EPHA10* and *HAND2* were measured by Pyrosequencing in 50 newborn cord blood samples and PBL DNA from their respective mothers aged 17-42 (average age 27) from the North Cumbrian Community Genetics Project and 50 PBL DNA samples from the Newcastle Thousand Families study (Unpublished data from van Otterdijk, Mathers and Strathdee), together with the 50 pilot study PBL DNA samples from the Newcastle 85+ study. DNA methylation levels, as well as inter-individual variability are clearly increased with advancing age.

7.3 Role of DNA methylation in ageing and development of cancer

Observations that the same types of genes are susceptible to increased and variable methylation during ageing and in cancer cells have been made by others since this study began (Rakyan *et al.*, 2010; Teschendorff *et al.*, 2010), and we were able to identify the same overlap at specific genes in the Newcastle 85+ study population (see Chapter 3.5.2). Although these specific genes were initially chosen as they were known to be hypermethylated in leukaemia, a genome wide analysis for identification of novel gene targets in haematological malignancies also showed a similar overlap when their methylation levels were investigated in the 85 year old individuals (see Chapter 5.4.3). This proposes a hypothesis that cancer-associated aberrant DNA methylation may directly originate from the age-related methylation which accumulates to variable degrees in normal individuals, and may partly explain why cancer risk is greater with increasing age, and why certain individuals are more susceptible than others. DNA methylation is usually much more dramatically increased in cancer than in ageing and this may because tumour tissue is typically monoclonal whereas ageing tissue is comprised of numerous cells. Cancer may originate by clonal expansion of an already abnormally methylated stem cell, and this may be more likely in an individual who has a greater proportion of abnormally methylated cells (higher level of methylation quantified by Pyrosequencing) than in a person with lower levels.

The co-methylation of genes in the Newcastle 85+ population is suggestive of a mechanism such that multiple genes are simultaneously affected by aberrant DNA methylation (see Chapter 3.5.1 and 5.5.4). Although not all of these targets may be expressed in haematopoietic cells, it is likely that for some genes, their expression would be detrimentally affected by the presence of promoter associated DNA methylation increases. Silenced expression may result in cells where DNA methylation appears to be particularly dense in individual alleles (see Chapter 3.5.3).

Interestingly, our results revealed a small but significant increase in DNA methylation measurement, in Newcastle 85+ participants with either a previous history, or future incidence of cancer, at the overall mean of five genes (*TWIST2, TUSC3, HOXD4, EPHA10* and *HAND2*) which are known by exhibit increased and variable methylation in ageing (as well as being hypermethylated in leukaemia) (see Chapter 4.2.5 and 4.3.3). Greater methylation was apparent in the 2 participants who went on to be diagnosed

with lymphoma or leukaemia (see chapter 4.3.3.1). However, since there were such few incidences of haematological cancer, care should be taken in concluding that higher methylation is a predictor of imminent risk of haematological cancer. It will be important to attempt to replicate this finding in other prospective analysis of leukaemia or lymphoma incidence, to provide further evidence of the possible predictive value of methylation of these genes a marker of individuals at high risk of haematological malignancy.

This prospective study has shown that differences in DNA methylation are evident before cancer can be diagnosed; therefore methylation changes may exist prior to cancer onset and subsequent treatment. Understanding how such DNA methylation differences arise in apparently normal samples could be vital to uncovering the role of pre-existing methylation in risk of cancer. If these small increases are due to an increased proportion of densely methylated alleles for example, the consequence is a greater number of cells with abnormal methylation patterns and altered gene transcription. However, it is not just densely methylated promoters which are associated with loss of expression. Recent results from the ENCODE project analysing over 1 million CpG sites in cells lines and tissue found the most variably methylated sites were located outside of promoter regions and also revealed DNA methylation located at some distance from the gene may be important in regulating transcription (Bernstein *et al.*, 2012). The early appearance of methylation abnormalities, prior to clinical disease development, would be consistent with a direct functional role for altered methylation in early stages of cancer development.

The absolute DNA methylation differences at the five loci measured were quite small between the groups of participants with previous cancer, future cancer or no cancer. The genes known to be hypermethylated leukaemia which were investigated in the 85 year old population, appeared to be co-ordinately methylated within an individual. This suggests the existence of a mechanism which controls the acquisition of DNA methylation across the genome within an individual and could therefore affect other loci similarly. The mechanism responsible for determining individual DNA methylation levels could be due to underlying genetic sequence, activity of the methylation machinery such as DNMTs or differential environmental exposures (See section 1.9). Therefore it is possible that the genes which were originally selected for analysis were not the most optimal. Other genes may have demonstrated similar, but more dramatic

DNA methylation differences between previous cancer, future cancer and individuals with no cancer, but were not investigated as part of this study.

Also, we chose to measure CpG sites within the promoter region of genes, as methylation covering the TSS is most well understood to cause transcriptional silencing (Eckhardt *et al.*, 2006; Bell *et al.*, 2011). However there are several reports that DNA methylation in alternative genomic regions, such as introns and exons, could control expression of an associated gene, particularly in a tissue specific manner (Strathdee *et al.*, 2004a; Wu *et al.*, 2010; Brenet *et al.*, 2011). Therefore, differential DNA methylation at different genomic locations could play an important role in gene expression.

7.4 Role of DNA methylation in age-related clinical outcomes

Our analysis revealed no evidence for any association between DNA methylation levels and CVD, T2D, dementia, stroke, smoking status or the 3 year survival of participants in the Newcastle 85+ study. This suggests that the mechanism underlying altered DNA methylation levels in this population may not be related to these clinical factors. However, we may have failed to identify the effect for several reasons, such as too few observations, failure to examine the right genes or the most informative regions in the genome.

Although altered DNA methylation patterns have been proposed to impact upon markers of healthy ageing (Liu *et al.*, 2011) and a role for aberrant methylation suggested in many age-related diseases (Barrachina and Ferrer, 2009; Ling and Groop, 2009; Baccarelli *et al.*, 2010; Kim *et al.*, 2010a), the significance of such changes in diseases other than cancer is not clearly understood. Thus it is not known whether aberrant DNA methylation may increase susceptibility to certain diseases, or whether DNA methylation changes may be triggered after the disease has initiated, for example inflammation is often present in CVD, T2D and stroke and has been implicated in causing altered DNA methylation (Stenvinkel *et al.*, 2007).

7.5 Future studies

We observed that genes which are methylated in leukaemia/lymphoma, exhibit variable methylation in the elderly. This was found in our pilot study analysis based on a panel of genes known to be methylated in leukaemia, and in two new targets identified by genome wide DNA methylation analysis of haematological malignancies. The level of DNA methylation across these loci was highly correlated within the 85 year old individuals.

The supposed presence of a mechanism affecting the acquisition of DNA methylation at genes susceptible to both age and cancer associated methylation would be further strengthened by examining other genes hypermethylated in haematological malignancies. This would require investigation of the other gene targets from the genome wide DNA methylation screen either by establishing successful Pyrosequencing assays for the other 2 genes known to be methylated in ALL samples by COBRA analysis (*CACNA1B* and *CELF4*); or alternatively by picking new targets from the genome wide methylation study, confirming their methylation status in primary leukaemia samples and establishing new Pyrosequencing assays to test DNA methylation levels in the Newcastle 85+ study participants.

An important finding from our pilot study analysis was the presence of densely methylated alleles in apparently normal samples. The majority of previous studies investigating the role of altered DNA methylation in malignancies, or during ageing have utilised techniques for measuring average methylation across all cells in a sample. However, methylation is known to act at the level of individual alleles and thus, the presence of different allelic patterns of methylation at individual alleles in both ageing and cancer should be investigated, as these patterns are likely to have different biological consequences. The cloning and bisulfite sequencing technique used in this study is time consuming and only allowed investigation of a relatively small number of alleles, which is unlikely to reveal the true complexity of allelic methylation patterns. A new procedure is currently being developed within our laboratory, combining multiplexed gene specific amplification of sodium bisulfite modified DNA with ultra-deep DNA sequencing allowing potentially thousands of alleles to be simultaneously characterised (Unpublished data from van Otterdijk and Strathdee). Subsequently, changes to the allelic patterns of methylation during ageing or in malignancies could be

examined to possibly identify particular types of methylation patterns associated with cancer, or age related disease risk.

It has been acknowledged that methylation covering the promoter region of a gene will result in transcriptional silencing, particularly evident at tumour suppressor genes in cancer. However, as the dense methylation we observed in normal elderly PBL DNA did not affect all alleles in the sample, it is not known to what extent expression of the gene would be affected. Certainly, not all the genes we examined in the pilot study would be expected to be expressed in normal haematopoietic cells, but since the methylation pattern we observed at these genes appears to be affecting numerous loci it is likely that some of these would normally be expressed. At least two of set of five genes used in most of the studies, *TWIST2* and *TUSC3*, are known to be expressed in PBL cells. Measuring gene expression levels in samples with differing proportions of densely methylated alleles could be performed to characterise the biological consequence of the variable DNA methylation levels in the Newcastle 85+ study population.

Our results have suggested co-methylation of genes in the Newcastle 85+ study participants is occurring at the levels of an individual and we hypothesise this is because multiple loci are hypermethylated in same cell, rather than different genes being hypermethylated in different cells. Experiments to confirm this hypothesis of concordant methylation would involve analysis of a single cell. Single cells can be isolated by flow cytometry, laser capture microdissection or by performing limiting dilutions, but subsequent analysis is technically difficult due to the low concentration of DNA in the starting material and high sensitivity to contamination with foreign DNA. This can possibly be overcome by clonal expansion of a single cell *in vitro* to a density which would generate enough DNA for accurate quantification of DNA methylation levels. However, this runs the risk of acquisition of inter-cell methylation differences as demonstrated by Kim and Shibata (2002) (Kim and Shibata, 2002). A recent publication described a technique allowing high-throughput methylation analysis of a single cell without the need for clonal expansion involving cleavage of DNA with methylation sensitive restriction enzymes and subsequent gene specific amplification utilising two forward primers either side of the enzyme cut site. Methylation status can be determined by the size of products formed depending on whether the site was methylated/uncleaved or unmethylated/cleaved (Kantlehner *et al.*, 2011). A criticism of this approach is that

analysis of CpG sites is limited to those which lie within a methylation sensitive enzyme recognition sequence, and it is not possible to examine multiple genes in the same cell. If single gene methylation analysis confirmed that numerous genes are hypermethylated in a single cell, it would strengthen the hypothesis that clonal expansion from these types of cells may be responsible for the abnormal methylation patterns seen in cancer, without requiring any additional methylation alterations.

This study focused entirely on methylation gains at gene promoter regions, although the presence of global DNA hypomethylation is frequently associated with both advancing cancer and advancing age (Ehrlich, 2002; Jintaridth and Mutirangura, 2010). Therefore it would be interesting to measure global DNA methylation levels in the Newcastle 85+ study participants, and look for any significant associations between global DNA methylation levels and the incidence of cancer or other clinical outcomes in the elderly age group. A study measuring LINE-1 and Alu repetitive elements in an elderly population as a surrogate for global methylation found that individuals with low LINE-1 methylation appeared to be at increased risk of cancer, with greater mortality (Zhu *et al.*, 2011).

Genome wide SNPs were characterised in the Newcastle 85+ study participants which allowed us to perform a hypothesis generating GWAS for identification of SNPs which may be associated with determination of DNA methylation levels. This analysis revealed several SNPs which were related to DNA methylation although the level of significance for the associations was not particularly stringent since there are a huge number of comparisons being made in genome wide analysis, therefore these associations require replication in an independent population to confirm the effect of these particular SNPs in driving DNA methylation differences within the population.
Two of the identified SNPs were each located within two genes, *DSCAM* and *DSCAML1*, which encode highly related protein products. In an attempt to characterise the effects of these SNPs on DNA methylation levels, the relationship between the presence of these SNPs at *DSCAM/DSCAML1* and the accumulation of abnormally high levels of DNA methylation will be examined in leukaemia patients.

In addition, Fried frailty status appeared to be significantly related to DNA methylation levels in the Newcastle 85+ study participants. However, an alternative definition of frailty (Rockwood) was unrelated to DNA methylation levels in the same individuals.

Further investigation could uncover which specific frailty characteristics are linked to differential DNA methylation. Replication of the relationship between DNA methylation in additional population, perhaps with differing definitions of frailty would strength this finding.

Chapter 8 – Appendices

Appendix 1 – Methodology

A1.1 Epigentek Methylamp™ One-Step DNA Modification Kit Protocol

Before starting, add 15 ml of 100% ethanol to G5 to make final cleaning buffer. Prepare 90% ethanol.

1. Add 1.1ml of G2 (DNA modification solution) to 1 vial of G1 (DNA modification powder). Vortex until solution is clear or saturated (about 2 min). Add 40µl of G3 (balance solution) to the solution, lightly vortex.

2. Add 110µl of the mixed G1/G2/G3 solution to 10µl of DNA sample (200ng). Vortex and place the vial in a thermal cycler with a program of 99°C for 6 min followed by 65°C for 90 min.

3. Place a spin column into a 2ml collection tube. Add 300µl of G4 (modified DNA capture buffer) to the column, and then transfer the sample (from step 2) to the column containing G4. Centrifuge at 12,000 rpm for 30 sec. Remove the column from the collection tube and discard the flow through. Replace column to the collection tube.

4. Add 200µl of G5 solution (modified DNA cleaning buffer) to the column, and centrifuge at 12,000 rpm for 30 sec.

5. Add 10µl of G3 to 1.1ml of 90% ethanol, mix. Add 50µl of the mixed G3/ethanol to the column. Sit for 8 min at room temperature, then centrifuge at 12,000 rpm for 20 sec.

6. Add 200µl of 90% ethanol to the column, centrifuge at 12,000 rpm for 30 sec., remove the column from the collection tube and discard the flow through. Replace column to the collection tube. Add 200µl of 90% ethanol to the column again and centrifuge at 12,000 rpm for 1 minute.

7. Place the column in a new 1.5 ml vial. Add 15µl of G6 (modified DNA elution buffer) directly to the column filter, centrifuge at 12,000 rpm for 30 sec to elute modified DNA.

A1.2 Recipe for TAE (Tris base, acetic acid and EDTA) DNA Electrophoresis Buffer (50x and 1x) and EDTA (Ethylenediaminetetraacetic acid) 0.5 M (pH 8)

TAE (50 X) (500ml)

121g Tris base

28.6ml acetic acid

50ml 0.5M EDTA (Ethylenediaminetetraacetic acid) pH 8

Add ddH_2O (double deionized water) to 500ml

TAE (1X) (10 Litres)

200 ml TAE (50x) buffer

Add 9800ml ddH_2O.

EDTA 0.5 M (pH8.0) (1 litre)

148 g EDTA

~30-40 g NaOH (Sodium hydroxide) to adjust to pH 8

Add ddH_2O to 1000ml

A1.3 Pyrosequencing Protocol

Materials

Ethanol (70%)

Streptavidin Sepharose beads (GE Healthcare)

Binding buffer (Qiagen)

Annealing buffer (Qiagen)

Denaturation buffer (BDH)

Wash buffer (Tris acetate buffer) (10x) (Qiagen)

18.2 mΩ ddH_2O

Sequencing primer (MWG Eurofins)

Pyro gold SQA reagents (Qiagen)

96-well PCR plate

96-well Pyrosequencing plate

Pipettes

Pipette tips

Heat block

Plate shaker

Buffer Solutions

Wash buffer (1x) (1 litre)

100ml (10x) wash buffer

900ml 18.2 mΩ ddH_2O

Ethanol (70%) (1 litre)

700ml 100% Ethanol

300ml 18.2 mΩ ddH_2O

Denaturing buffer (1 litre)

8g NaOH (Sodium hydroxide)

1000ml 18.2 mΩ ddH_2O

Set up Pyro work station as follows

	Probe block 18.2 mΩ ddH_2O
Denaturing buffer (NaOH)	18.2 mΩ ddH_2O
70% Ethanol	Wash buffer (1x)
PCR plate	Pyrosequencing plate

Remove reagents from fridge and allow solutions to reach room temperature.

Heat block set to 80ºC

Immobilisation of PCR products

Biotinylated PCR products are immobilised on streptavidin-coated beads.

1. Shake bottle of streptavidin coated sepharose beads to obtain a homogenous solution
2. Prepare binding buffer solution master mix (per sample)

   Beads 2µl

   Binding buffer 38µl

   18.2 mΩ ddH_2O 30µl
3. Add 70µl to PCR plate, shake master mix every few wells so beads don't settle
4. Add 10µl PCR product to the binding buffer in plate
5. Seal plate and shake for at least 5 minutes
6. Prepare Pyrosequencing plate annealing master mix (per sample)

   Annealing buffer 11.5µl

   Sequencing primer (10pmol/µl) 0.5µl
7. Add 12µl to the Pyrosequencing plate wells in corresponding locations to those on the PCR plate

Strand separation

Switch on the vacuum pump and prime probes by sucking through 18.2 mΩ ddH_2O for a few seconds.

1. Remove plate from shaker and immediately apply vacuum prep tool. Make sure all sample is taken up.
2. Place vacuum tool in ethanol trough. Wait for liquid to flow through, then count 5 seconds. Repeat for denaturation buffer for 5 seconds and wash buffer for 5 seconds.
3. Turn off vacuum. Place tool in Pyrosequencing plate to release single stranded PCR product into annealing solution.
4. Place plate on heat block for 2 minutes. Clean vacuum by sucking through 18.2 mΩ ddH_2O.
5. Remove plate from block and leave to cool for a few minutes.

Pyrosequencing

Fill reagent cartridge

1. Switch on the computer.
2. (If necessary create a new assay, see step 1-4 below)
3. Go to tools, volume information. It gives the volume for needed for all reagents.
4. Reconstitute enzyme and substrate mixtures with 18.2 mΩ ddH_2O.
5. Fill Enzyme and substrate tips with appropriate volume for assay.
6. Fill nucleotide tips with care –air bubbles can interfere with dispensation- flick to remove bubbles.
7. Tips are inserted into cartridge in alphabetical order

E S

A C G T

Creating new assay

1. Pyro Q CpG → new assay
2. Type in sequence and generate dispensation order (30-35 nucleotides max)
3. Tick analyse boxes
4. Add bisulfite control
5. Start a Pyrosequencing run
6. New run.
7. Instrument parameters MD or HS(A) code 004 NDT

8. Use date in plate ID
9. Load assays into appropriate wells.

Starting a new Pyrosequencing run

1. Select Pyro Q CpG from the start menu, then select New Run
2. Select Instrument parameters MD or HS(A) code 004 NDT
3. Use date in plate ID
4. In appropriate well, right click and select Load Assay. This will open up files of previously save assays. Double click the required assay and it will appear in the well
5. Copy the assay by dragging the bottom corner of the first well to all appropriate wells.
6. Open and close the pyrosequencer using the 'open process chamber' and 'close process chamber' buttons on screen.
7. To run the assay click the play button

Always test tip dispensing before starting the run – Insert a SEALED Pyrosequencing plate into machine and select test dispensation – should observe 6 dots (one for each tip)

After the assay

1. Analyse all
2. Report, analysis reports, all wells, save
3. Report, programme, all wells, landscape, save
4. Instrument, shut down instrument.

A1.4 Recipe for LB (Lysogeny broth) media and agar with Kanamycin (50µg/ml) (1 litre)

1% tryptone
0.5% yeast extract
1% NaCl (Sodium chloride)
pH 7

10g tryptone
5g yeast extract
10g NaCl
950ml of ddH2O
Adjust pH to 7 with NaOH (Sodium hydroxide) & increase volume to 1 litre
Autoclave and allow to cool
Add 1ml of stock kanamycin (50mg/ml)

If making agar, add 15g prior to autoclaving, cooling and addition of kanamycin
Pour plates, leave to harden, invert and store at 4ºC.

A1.5 QIAprep Spin Miniprep Kit Protocol

After overnight incubation, centrifuge cells at 5000r.p.m. at 4°C for 10 minutes to form pellet. Remove media and invert tubes to drain

Add provided RNase A to Buffer P1 then store Buffer P1 at 2-8°C.

Add the provided LyseBlue reagent to Buffer P1 and mix before use. Use one vial LyseBlue per bottle of Buffer P1 to achieve a 1:1000 dilution. Shake Buffer P1 vigorously before use.

All remaining protocol steps carried out at room temperature.

1. Resuspend pelleted bacterial cells in 250 µl Buffer P1 and transfer to a microcentrifuge tube.

No cell clumps should be visible after resuspension of the pellet; bacteria should be resuspended completely by vortexing or pipetting up and down.

2. Add 250µl Buffer P2 and mix thoroughly by inverting the tube 4–6 times.
Do not vortex, as this will result in shearing of genomic DNA. Do not allow this lysis reaction to proceed for more than 5 min.
Cell suspension will turn blue after addition of Buffer P2. Continue mixing to achieve a homogeneously coloured suspension.

3. Add 350µl Buffer N3 and mix immediately and thoroughly by inverting the tube 4–6 times.
The suspension should be mixed until all trace of blue has gone and the suspension is colourless and cloudy. A homogeneous colourless suspension indicates that the sodium dodecyl sulphate (SDS) has been effectively precipitated.

4. Centrifuge for 10 min at 13,000 r.p.m. in a microcentrifuge. A compact white pellet will form.

5. Remove the supernatant from step 4 and add to the QIAprep spin column.

6. Centrifuge at 13,000 r.p.m. for 1 minute. Discard the flow-through.

7. Wash the QIAprep spin column by adding 500µl Buffer PB and centrifuging at 13,000 r.p.m. for 1 minute. Discard the flow-through.

8. Wash QIAprep spin column by adding 500µl Buffer PE and centrifuging at 13,000 r.p.m. for 1 minute. Discard the flow-through. Centrifuge at 13,000 r.p.m. for another 1 minute to remove residual wash buffer.

9. Place the QIAprep column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 50µl Buffer EB (10 mM Tris·Cl, pH 8.5) to the center of each QIAprep spin column, stand for 5 minutes, and centrifuge at 13,000 r.p.m. for 1 minute.

A1.6 Recipe for Reagent B (Nucleon) for DNA Extraction, Tris-HCl 1M and NaCl 1M

400mM Tris/HCl

60mM EDTA(Ethylenediaminetetraacetic acid)

150mM NaCl (Sodium Chloride)

1% SDS (Sodium dodecyl sulfate)

pH 8.0

1M Tris-HCl (1 litre)

121g Tris base

HCl to pH 8

Add ddH_2O to 1000ml

1M NaCl (1 litre)

58g NaCl (Sodium Chloride)

Add ddH_2O to 1000ml

A1.7 Recipe for TE (Tris EDTA) Buffer (1x) (1 litre)

10 mM Tris-HCl, pH 8

1 mM EDTA

10ml 1 M Tris-HCl

2ml 0.5M EDTA

Add ddH_2O to 1000ml

A1.8 QIAquick PCR Purification Kit/MinElute PCR Purification Kit Protocol

Add 100% ethanol to Buffer PE before use.

All centrifugation steps carried out at 13,000 r.p.m. in a microcentrifuge at room temperature.

Add 1:250 volume pH indicator I to Buffer PB (120µl pH indicator I to 30ml Buffer PB). The yellow colour of Buffer PB with pH indicator I indicates a pH of 7.5.

1. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix.

2. If pH indicator I has been added to Buffer PB, check that the colour of the mixture is yellow.

3. Place a QIAquick/MinElute column in a provided 2 ml collection tube.

4. To bind DNA, add the sample to the QIAquick/MinElute column and centrifuge for 1 minute.

5. Discard flow-through. Place the QIAquick/MinElute column back into the same tube.

6. To wash, add 750µl Buffer PE to the QIAquick/MinElute column and centrifuge for 1 minute.

7. Discard flow-through and place the QIAquick/MinElute column back in the same tube. Centrifuge the column for an additional 1 minute for complete removal of buffers.

8. Place QIAquick/MinElute column in a clean 1.5 ml microcentrifuge tube.

9. To elute DNA, add an appropriate volume of Buffer EB (10 mM Tris·Cl, pH 8.5) to the centre of the QIAquick/MinElute membrane, stand for 1 minute and centrifuge the column for 1 min.

Appendix 2 – Results Chapter 3

A2.1 Reproducibility of duplicate Pyrosequencing runs of the same *TWIST2* PCR product

Sample	Average of 4 CpG sites (%)	Difference (%)
37	7.28	
37	7.19	0.09
45	10.58	
45	10.17	0.41
50	4.97	
50	4.45	0.53
78	7.75	
78	7.55	0.20
80	3.11	
80	3.42	0.31
84	9.03	
84	9.00	0.03
108	4.13	
108	4.19	0.06
112	5.18	
112	5.07	0.11
119	8.26	
119	8.54	0.28
125	3.39	
125	3.46	0.07
128	9.27	
128	9.37	0.11
148	4.52	
148	4.52	0.00
161	7.81	
161	7.53	0.29
166	12.29	
166	11.78	0.51
181	11.74	
181	11.76	0.02
185	19.03	
185	20.08	1.05
186	4.78	
186	5.70	0.92
193	7.46	
193	7.31	0.15
194	5.19	
194	5.06	0.13
205	11.77	
205	12.49	0.72
213	14.37	
213	14.90	0.52
221	6.17	
221	6.23	0.06
312	5.63	
312	5.48	0.15
Average difference between 23 samples		0.29

For 23 of the 85+ study participants, an aliquot of the same *TWIST2* PCR product was used for two separate Pyrosequencing runs. For each sample, methylation measurements at four consecutive CpG sites was used to calculate the average, and the methylation (%) difference between the two repeats is shown.

A2.2 Reproducibility between of Pyrosequencing measurements between two duplicate *TWIST2* PCR repeats

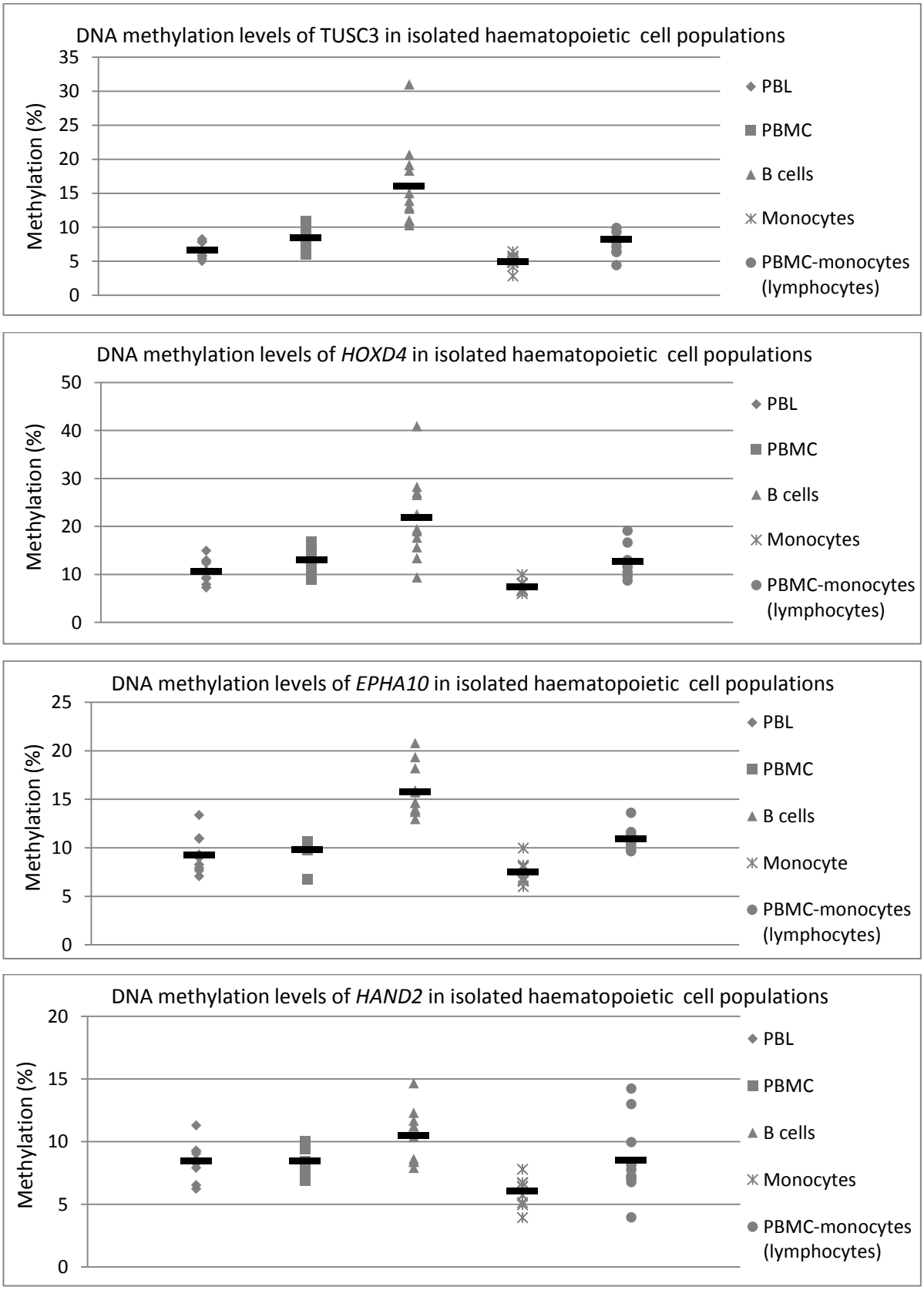
Sample	Average of 4 CpG sites (%)	Difference (%)
37	6.80	
37	7.24	0.44
45	9.90	
45	10.38	0.48
50	4.92	
50	4.71	0.21
78	7.65	
78	7.41	0.24
80	9.01	
80	10.86	1.85
84	7.81	
84	9.02	1.21
108	6.42	
108	4.16	2.26
112	6.26	
112	5.13	1.13
119	9.63	
119	8.40	1.23
125	5.19	
125	3.42	1.77
128	5.27	
128	9.32	4.05
148	5.00	
148	4.52	0.47
161	6.23	
161	7.67	1.45
166	12.04	
166	9.73	2.31
181	8.89	
181	11.75	2.87
185	20.85	
185	19.55	1.30
186	3.44	
186	5.24	1.80
193	4.62	
193	5.95	1.32
194	5.12	
194	5.30	0.18

Sample	Average of 4 CpG sites (%)	Difference (%)
199	6.27	
199	6.84	0.57
205	12.33	
205	12.13	0.19
213	17.56	
213	14.63	2.93
221	4.22	
221	6.20	1.98
305	5.87	
305	6.42	0.56
312	9.26	
312	5.55	3.71
Average difference between 25 samples		1.46

For 25 of the 85+ study participants, two distinct *TWIST2* PCR amplifications were performed and the two products were used for two separate Pyrosequencing runs. For each sample, methylation measurements at four consecutive CpG sites were used to calculate the average, and the methylation (%) difference between the two PCR repeats is shown.

Appendix 3 – Results Chapter 4

Figure A3.1: DNA methylation levels of *TUSC3, HOXD4, EPHA10* and *HAND2* in isolated haematopoietic cell populations



DNA was isolated from PBLs (lymphocytes, neutrophils, and monocytes), PBMCs (lymphocytes and monocytes), B cells and monocytes and methylation levels of the *TUSC3*, *HOXD4*, *EPHA10* and *HAND2* promoter measured by Pyrosequencing. PBMC-monocytes represent total lymphocytes (T cells, B cells and NK cells). DNA methylation levels are lowest in monocytes and highest in B cells.

Table A3.2: Comparison of PBL DNA methylation levels between participants with and without history of cardiovascular events in the Newcastle 85+ study (initial sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
None (%)	179	10.20	10.56	10.37	10.17	10.47	10.35	63.44
CVD (%)	91	10.68	11.40	10.76	10.69	10.80	10.87	58.43
t-test p-value		0.142	0.070	0.162	0.113	0.167	0.057	**0.013***

Student's t-test was used to analyse difference in methylation levels between participants with a previous history of myocardial infarction, angina, coronary angioplasty/stent, coronary artery bypass graft, or none of these. *HOXA5* methylation levels were significantly different between the two groups *($p < 0.05$).

Table A3.3: Comparison of PBL DNA methylation levels between participants with and without Type II diabetes in the Newcastle 85+ study (initial sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
None (%)	234	10.41	10.87	10.54	10.36	10.60	10.55	61.99
T2D (%)	36	10.08	10.71	10.20	10.28	10.45	10.34	60.45
t-test p-value		0.300	0.419	0.267	0.447	0.374	0.323	0.326

Student's t-test was used to analyse difference in methylation levels between participants with Type II diabetes compared to none. There were no significant differences ($p < 0.05$).

Table A3.4: Comparison of PBL DNA methylation levels between participants with and without Dementia in the Newcastle 85+ study (initial sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
None (%)	252	10.36	10.83	10.51	10.35	10.48	10.50	61.24
Dementia (%)	18	10.41	11.07	10.28	10.34	12.01	10.82	66.54
t-test p-value		0.475	0.411	0.390	0.495	**0.009***	0.303	0.060

Student's t-test was used to analyse difference in methylation levels between participants with and without Dementia including Alzheimer's disease. There was significantly increased methylation of *EPHA10* in participants with dementia *($p<0.05$).

Table A3.5: Comparison of PBL DNA methylation levels between participants with and without stroke in the Newcastle 85+ study (initial sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
None (%)	228	10.34	10.89	10.47	10.22	10.55	10.49	61.30
Stroke (%)	42	10.46	10.61	10.66	11.07	10.73	10.70	64.44
t-test p-value		0.423	0.350	0.358	0.063	0.346	0.309	0.135

Student's t-test was used to analyse difference in methylation levels between participants with and without stroke. There were no significant differences ($p<0.05$).

Table A3.6: Comparison of PBL DNA methylation levels between surviving and non- surviving (by Phase III) participants of the Newcastle 85+ study (initial sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Survivors (%)	229	10.24	10.89	10.41	10.22	10.51	10.46	61.68
Non-survivors (%)	41	11.02	10.58	10.96	11.11	10.97	10.90	62.63
t-test p-value		0.09	0.337	0.149	0.060	0.164	0.150	0.373

Student's t-test was used to analyse difference in methylation levels between participants who had survived to phase III (36 months later) or not. There were no significant differences ($p<0.05$).

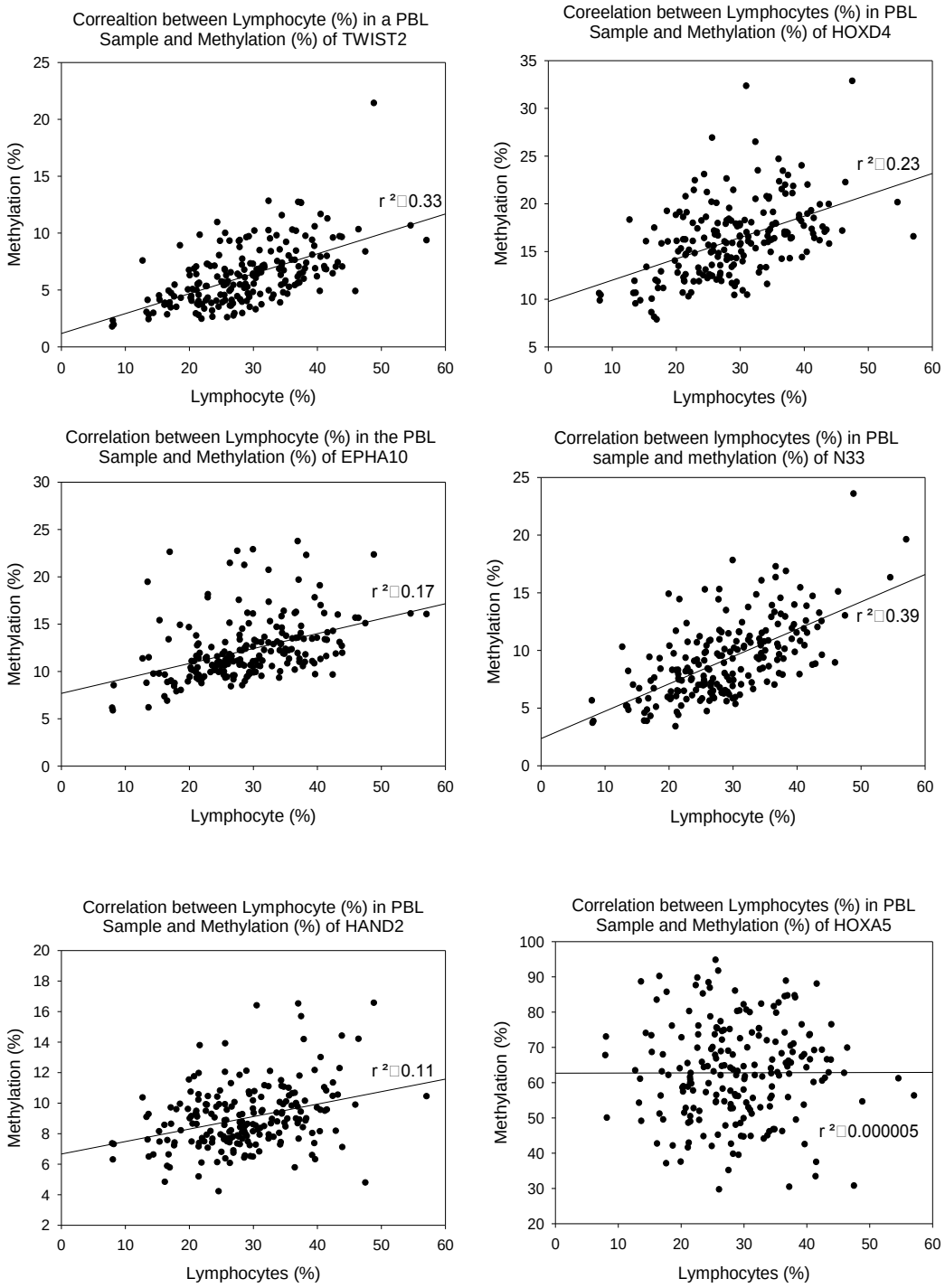
Table A3.7: Pair-wise correlations of *TWIST2, HOXD4, EPHA10, TUSC3, HAND2* and *HOXA5* DNA methylation levels in 210 validation PBL samples from the Newcastle 85+ study participants.

	TWIST2	*HOXD4*	*EPHA10*	*TUSC3*	*HAND2*
TWIST2					
HOXD4	r = 0.72 p < 1e-16				
EPHA10	r = 0.48 p = 1.3e-13	r = 0.38 p = 1.4e-08			
TUSC3	r = 0.84 p < 1e-16	r = 0.71 p < 1e-16	r = 0.54 p < 1e-16		
HAND2	r = 0.52 p = 1.1e-15	r = 0.34 p = 7.4e-07	r = 0.19 p = 0.005	r = 0.48 p = 1.9e-13	
HOXA5	r = -0.018 p = 0.79	r = 0.05 p = 0.51	r = 0.011 p = 0.87	r = -0.047 p = 0.50	r = 0.009 p = 0.89

Methylation data for 210 participants was correlated at each pair of loci using Pearson test of correlation. The Pearson correlation coefficient (r) and corresponding p values are shown. Methylation levels at *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* are strongly positively correlated. *HOXA5* demonstrates no relationship.

Black p value = NS, Red p value = significant at <0.003 level after performing Bonferroni correction for multiple comparisons.

Figure A3.8: Correlation between PBL DNA methylation levels and lymphocyte percentage in the peripheral blood of 210 Newcastle 85+ study participants (validation sample set).



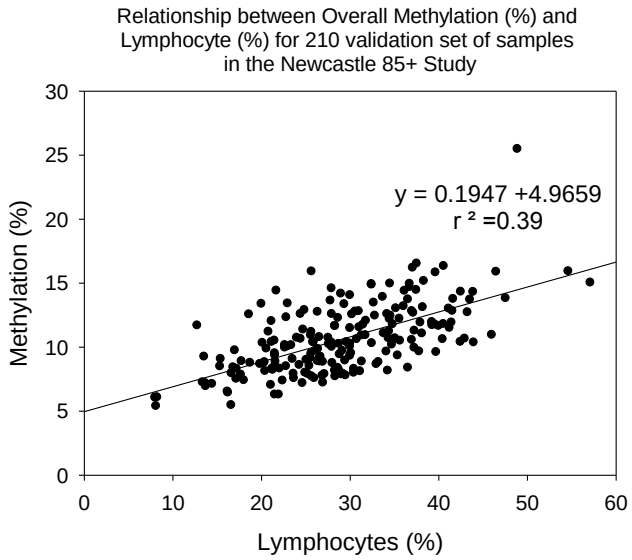
Methylation values of *TWIST2, HOXD4, EPHA10, TUSC3, HAND2* and *HOXA5* were plotted against lymphocyte percentage calculated from full blood counts (Chapter 2.1). The linear regression line and r^2 value indicate a positive relationship for all genes except *HOXA5*, for which there is no association.

Table A3.9: Correlation between the DNA methylation levels of *TWIST2, HOXD4, EPHA10, TUSC3* and *HAND2* and the percentage of lymphocytes in the peripheral blood of 210 participants in the Newcastle 85+ study (validation sample set)

	TWIST2	*HOXD4*	*TUSC3*	*EPHA10*	*HAND2*	*HOXA5*
Lymphocyte (%)	r = 0.58 p = 0	r = 0.48 p = 5.1e-13	r = 0.62 p = 0	r = 0.41 p = 6.6e-10	r = 0.33 p = 1.6e-06	r = 0.002 p = 0.97

Pearson correlation coefficient (r) and corresponding p value is given for each locus. Black p value = NS, Red p value = significant at <0.0083 level after performing Bonferroni correction for multiple comparisons.

Figure A3.10: Relationship between overall mean DNA methylation and percentage of lymphocytes in the peripheral blood of 210 participants from the Newcastle 85+ study (validation sample set)



Linear regression analysis was performed by plotting the overall mean methylation of each sample against the percentage of lymphocytes. A positive relationship is indicated by the r^2 value (0.39). The equation of the regression y = 0.1947x + 4.9659 indicates that every 1% increase in lymphocytes (x), is responsible for a 0.1947% increase in methylation levels (y).

Figure A3.11: The lymphocyte normalisation factor calculated and applied to DNA methylation measurements of 210 participants from the Newcastle 85+ study (validation sample set)

Sample lymphocytes (%) – Population Mean lymphocytes (29.0%) * Methylation increase determined by lymphocytes (0.1947%)

Table A3.12: DNA methylation measurements after correction for the proportion of lymphocytes in the peripheral blood of 210 participants from the Newcastle 85+ study (validation sample set)

Gene	Methylation Range (%)	Mean Methylation (%)	Standard Deviation (%)
HOXD4	5-21	10.63	2.2
TUSC3	4-27	10.63	3.8
TWIST2	3-36	10.63	4.4
EPHA10	5-20	10.63	2.8
HAND2	5-20	10.63	2.5

For each gene, the corrected minimum and maximum methylation levels in the samples is shown, alongside the normalised average methylation level calculated as described above, and the standard deviation as a measure of variation.

Table A3.13: Comparison of PBL DNA methylation levels between individuals with and without history of cardiovascular events in 210 participants from the Newcastle 85+ study (validation sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
None (%)	142	10.67	10.51	10.68	10.78	10.60	10.65	62.31
CVD (%)	68	10.55	10.89	10.58	10.40	10.70	10.64	63.79
t-test p-value		0.394	0.243	0.395	0.150	0.391	0.479	0.230

Student's t-test was used to analyse difference in methylation levels between participants with a previous history of myocardial infarction, angina, coronary angioplasty/stent, coronary artery bypass graft, or none of these. There were no significant associations ($p < 0.05$).

Table A3.14: Comparison of PBL DNA methylation levels between individuals with and without Type II diabetes in 210 participants from the Newcastle 85+ study (validation sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
None (%)	71	10.79	11.01	10.59	10.62	10.23	10.65	61.53
T2D (%)	58	10.93	11.05	10.93	10.72	10.77	10.93	63.92
t-test p-value		0.404	0.477	0.226	0.413	0.090	0.240	0.172

Student's t-test was used to analyse difference in methylation levels between participants with Type II diabetes compared to none. There were no significant differences ($p < 0.05$).

Table A3.15: Comparison of PBL DNA methylation levels between individuals with and without Dementia in 210 participants from the Newcastle 85+ study (validation sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
None (%)	70	10.83	11.09	10.60	10.63	10.23	10.68	61.49
Dementia (%)	36	10.10	10.11	10.27	10.68	11.13	10.46	61.51
t-test p-value		0.123	0.087	0.242	0.461	**0.043***	0.305	0.497

Student's t-test was used to analyse difference in methylation levels between participants with and without Dementia including Alzheimer's disease. Information regarding dementia status was missing for one participant. There was significantly increased methylation of *EPHA10* in participants with dementia *($p<0.05$)

Table A3.16: Comparison of PBL DNA methylation levels between individuals with and without stroke in 210 participants from the Newcastle 85+ study (validation sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
None (%)	186	10.72	10.67	10.70	10.58	10.65	10.67	62.83
Stroke (%)	24	9.91	10.34	10.29	11.24	10.52	10.46	62.52
t-test p-value		0.107	0.342	0.208	0.108	0.413	0.318	0.458

Student's t-test was used to analyse difference in methylation levels between participants with and without stroke. There were no significant differences ($p<0.05$)

Table A3.17: Comparison of PBL DNA methylation levels between surviving and non- surviving (by phase III) participants in the Newcastle 85+ study (validation sample set)

	n	*TUSC3*	*TWIST2*	*HOXD4*	*HAND2*	*EPHA10*	Overall Mean	*HOXA5*
Survivors (%)	170	10.63	10.69	10.71	10.59	10.62	10.66	61.88
Non-survivors (%)	40	10.63	10.37	10.39	10.96	10.69	10.62	66.66
t-test p-value		0.50	0.309	0.215	0.202	0.434	0.456	**0.022***

Student's t-test was used to analyse difference in methylation levels between participants who had survived to phase III (36 months later) or not. There was a significantly higher mean methylation level of *HOXA5* in non-survivors *($p<0.05$).

Chapter 9 – References

Agalioti, T., Chen, G. and Thanos, D. (2002) 'Deciphering the transcriptional histone acetylation code for a human gene', *Cell*, 111(3), pp. 381-92.

Agrawal, S., Unterberg, M., Koschmieder, S., zur Stadt, U., Brunnberg, U., Verbeek, W., Buchner, T., Berdel, W.E., Serve, H. and Muller-Tidow, C. (2007) 'DNA methylation of tumor suppressor genes in clinical remission predicts the relapse risk in acute myeloid leukemia', *Cancer Res*, 67(3), pp. 1370-7.

Ahlbom, A., Lichtenstein, P., Malmstrom, H., Feychting, M., Hemminki, K. and Pedersen, N.L. (1997) 'Cancer in twins: genetic and nongenetic familial risk factors', *J Natl Cancer Inst*, 89(4), pp. 287-93.

Ahmadian, A., Ehn, M. and Hober, S. (2006) 'Pyrosequencing: history, biochemistry and future', *Clin Chim Acta*, 363(1-2), pp. 83-94.

Ahuja, N., Li, Q., Mohan, A.L., Baylin, S.B. and Issa, J.P. (1998) 'Aging and DNA methylation in colorectal mucosa and cancer', *Cancer Res*, 58(23), pp. 5489-94.

Al-Ghnaniem, R., Peters, J., Foresti, R., Heaton, N. and Pufulete, M. (2007) 'Methylation of estrogen receptor alpha and mutL homolog 1 in normal colonic mucosa: association with folate and vitamin B-12 status in subjects with and without colorectal neoplasia', *Am J Clin Nutr*, 86(4), pp. 1064-72.

Alisch, R.S., Barwick, B.G., Chopra, P., Myrick, L.K., Satten, G.A., Conneely, K.N. and Warren, S.T. (2012) 'Age-associated DNA methylation in pediatric populations', *Genome Res*, 22(4), pp. 623-32.

Avraham, A., Sandbank, J., Yarom, N., Shalom, A., Karni, T., Pappo, I., Sella, A., Fich, A., Walfisch, S., Gheber, L. and Evron, E. (2010) 'A similar cell-specific pattern of HOXA methylation in normal and in cancer tissues', *Epigenetics*, 5(1), pp. 41-6.

Baccarelli, A., Wright, R., Bollati, V., Litonjua, A., Zanobetti, A., Tarantini, L., Sparrow, D., Vokonas, P. and Schwartz, J. (2010) 'Ischemic heart disease and stroke in relation to blood DNA methylation', *Epidemiology*, 21(6), pp. 819-28.

Baccarelli, A., Wright, R.O., Bollati, V., Tarantini, L., Litonjua, A.A., Suh, H.H., Zanobetti, A., Sparrow, D., Vokonas, P.S. and Schwartz, J. (2009) 'Rapid DNA methylation changes after exposure to traffic particles', *Am J Respir Crit Care Med*, 179(7), pp. 572-8.

Bai, A.H., Tong, J.H., To, K.F., Chan, M.W., Man, E.P., Lo, K.W., Lee, J.F., Sung, J.J. and Leung, W.K. (2004) 'Promoter hypermethylation of tumor-related genes in the progression of colorectal neoplasia', *Int J Cancer*, 112(5), pp. 846-53.

Barbot, W., Dupressoir, A., Lazar, V. and Heidmann, T. (2002) 'Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction', *Nucleic Acids Res*, 30(11), pp. 2365-73.

Barker, D.J. (2004) 'The developmental origins of adult disease', *J Am Coll Nutr*, 23(6 Suppl), pp. 588S-595S.

Barrachina, M. and Ferrer, I. (2009) 'DNA methylation of Alzheimer disease and tauopathy-related genes in postmortem brain', *J Neuropathol Exp Neurol*, 68(8), pp. 880-91.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F. and Niehrs, C. (2007) 'Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation', *Nature*, 445(7128), pp. 671-5.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007) 'High-resolution profiling of histone methylations in the human genome', *Cell*, 129(4), pp. 823-37.

Baylin, S.B. (2005) 'DNA methylation and gene silencing in cancer', *Nat Clin Pract Oncol*, 2 Suppl 1, pp. S4-11.

Belinsky, S.A., Palmisano, W.A., Gilliland, F.D., Crooks, L.A., Divine, K.K., Winters, S.A., Grimes, M.J., Harms, H.J., Tellez, C.S., Smith, T.M., Moots, P.P., Lechner, J.F., Stidley, C.A. and Crowell, R.E. (2002) 'Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers', *Cancer Res*, 62(8), pp. 2370-7.

Bell, A.C. and Felsenfeld, G. (2000) 'Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene', *Nature*, 405(6785), pp. 482-5.

Bell, J.T., Pai, A.A., Pickrell, J.K., Gaffney, D.J., Pique-Regi, R., Degner, J.F., Gilad, Y. and Pritchard, J.K. (2011) 'DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines', *Genome Biol*, 12(1), p. R10.

Bell, J.T., Tsai, P.C., Yang, T.P., Pidsley, R., Nisbet, J., Glass, D., Mangino, M., Zhai, G., Zhang, F., Valdes, A., Shin, S.Y., Dempster, E.L., Murray, R.M., Grundberg, E., Hedman, A.K., Nica, A., Small, K.S., Dermitzakis, E.T., McCarthy, M.I., Mill, J., Spector, T.D. and Deloukas, P. (2012) 'Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population', *PLoS Genet*, 8(4), p. e1002629.

Bellizzi, D., D'Aquila, P., Giordano, M., Montesanto, A. and Passarino, G. (2012a) 'Global DNA methylation levels are modulated by mitochondrial DNA variants', *Epigenomics*, 4(1), pp. 17-27.

Bellizzi, D., D'Aquila, P., Montesanto, A., Corsonello, A., Mari, V., Mazzei, B., Lattanzio, F. and Passarino, G. (2012b) 'Global DNA methylation in old subjects is correlated with frailty', *Age (Dordr)*.

Berdasco, M. and Esteller, M. (2012) 'Hot topics in epigenetic mechanisms of aging: 2011', *Aging Cell*, 11(2), pp. 181-6.

Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C. and Snyder, M. (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57-74.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber,

S.L. and Lander, E.S. (2006) 'A bivalent chromatin structure marks key developmental genes in embryonic stem cells', *Cell*, 125(2), pp. 315-26.

Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V. (1988) 'Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases', *J Mol Biol*, 203(4), pp. 971-83.

Bird, A. (2007) 'Perceptions of epigenetics', *Nature*, 447(7143), pp. 396-8.

Bjornsson, H.T., Sigurdsson, M.I., Fallin, M.D., Irizarry, R.A., Aspelund, T., Cui, H., Yu, W., Rongione, M.A., Ekstrom, T.J., Harris, T.B., Launer, L.J., Eiriksdottir, G., Leppert, M.F., Sapienza, C., Gudnason, V. and Feinberg, A.P. (2008) 'Intra-individual change over time in DNA methylation with familial clustering', *Jama*, 299(24), pp. 2877-83.

Bock, C., Paulsen, M., Tierling, S., Mikeska, T., Lengauer, T. and Walter, J. (2006) 'CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure', *PLoS Genet*, 2(3), p. e26.

Bocklandt, S., Lin, W., Sehl, M.E., Sanchez, F.J., Sinsheimer, J.S., Horvath, S. and Vilain, E. (2011) 'Epigenetic predictor of age', *PLoS One*, 6(6), p. e14821.

Boks, M.P., Derks, E.M., Weisenberger, D.J., Strengman, E., Janson, E., Sommer, I.E., Kahn, R.S. and Ophoff, R.A. (2009) 'The relationship of DNA methylation with age, gender and genotype in twins and healthy controls', *PLoS One*, 4(8), p. e6767.

Bollati, V., Schwartz, J., Wright, R., Litonjua, A., Tarantini, L., Suh, H., Sparrow, D., Vokonas, P. and Baccarelli, A. (2009) 'Decline in genomic DNA methylation through aging in a cohort of elderly subjects', *Mech Ageing Dev*, 130(4), pp. 234-9.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S. and Jacobsen, S.E. (2007) 'UHRF1 plays a role in maintaining DNA methylation in mammalian cells', *Science*, 317(5845), pp. 1760-4.

Bourc'his, D., Xu, G.L., Lin, C.S., Bollman, B. and Bestor, T.H. (2001) 'Dnmt3L and the establishment of maternal genomic imprints', *Science*, 294(5551), pp. 2536-9.

Boyes, J. and Bird, A. (1991) 'DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein', *Cell*, 64(6), pp. 1123-34.

Brait, M., Ford, J.G., Papaiahgari, S., Garza, M.A., Lee, J.I., Loyo, M., Maldonado, L., Begum, S., McCaffrey, L., Howerton, M., Sidransky, D., Emerson, M.R., Ahmed, S., Williams, C.D. and Hoque, M.O. (2009) 'Association between lifestyle factors and CpG island methylation in a cancer-free population', *Cancer Epidemiol Biomarkers Prev*, 18(11), pp. 2984-91.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. and Cedar, H. (1994) 'Sp1 elements protect a CpG island from de novo methylation', *Nature*, 371(6496), pp. 435-8.

Brandt, A., Bermejo, J.L., Sundquist, J. and Hemminki, K. (2008) 'Age of onset in familial cancer', *Ann Oncol*, 19(12), pp. 2084-8.

Breitling, L.P., Yang, R., Korn, B., Burwinkel, B. and Brenner, H. (2011) 'Tobacco-smoking-related differential DNA methylation: 27K discovery and replication', *Am J Hum Genet*, 88(4), pp. 450-7.

Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A.J., Socci, N.D. and Scandura, J.M. (2011) 'DNA Methylation of the First Exon Is Tightly Linked to Transcriptional Silencing', *PLoS One*, 6(1).

Breton, C.V., Byun, H.M., Wenten, M., Pan, F., Yang, A. and Gilliland, F.D. (2009) 'Prenatal tobacco smoke exposure affects global and gene-specific DNA methylation', *Am J Respir Crit Care Med*, 180(5), pp. 462-7.

Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earabino, C., Lipford, J., Lindblom, A. and et al. (1994) 'Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer', *Nature*, 368(6468), pp. 258-61.

Buckley, J.D., Buckley, C.M., Breslow, N.E., Draper, G.J., Roberson, P.K. and Mack, T.M. (1996) 'Concordance for childhood cancer in twins', *Med Pediatr Oncol*, 26(4), pp. 223-9.

Butler, M.G. (2009) 'Genomic imprinting disorders in humans: a mini-review', *J Assist Reprod Genet*, 26(9-10), pp. 477-86.

Cameron, E.E., Baylin, S.B. and Herman, J.G. (1999) 'p15(INK4B) CpG island methylation in primary acute leukemia is heterogeneous and suggests density as a critical factor for transcriptional silencing', *Blood*, 94(7), pp. 2445-51.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S. and Zhang, Y. (2002) 'Role of histone H3 lysine 27 methylation in Polycomb-group silencing', *Science*, 298(5595), pp. 1039-43.

Carew, J.S., Giles, F.J. and Nawrocki, S.T. (2008) 'Histone deacetylase inhibitors: mechanisms of cell death and promise in combination cancer therapy', *Cancer Lett*, 269(1), pp. 7-17.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., Fox, E.A., Silver, P.A. and Brown, M. (2005) 'Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1', *Cell*, 122(1), pp. 33-43.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P. and Workman, J.L. (2005) 'Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription', *Cell*, 123(4), pp. 581-92.

Casillas, M.A., Jr., Lopatina, N., Andrews, L.G. and Tollefsbol, T.O. (2003) 'Transcriptional control of the DNA methyltransferases is altered in aging and neoplastically-transformed human fibroblasts', *Mol Cell Biochem*, 252(1-2), pp. 33-43.

Cedar, H. and Bergman, Y. (2009) 'Linking DNA methylation and histone modification: patterns and paradigms', *Nat Rev Genet*, 10(5), pp. 295-304.

Chase, A. and Cross, N.C. (2011) 'Aberrations of EZH2 in cancer', *Clin Cancer Res*, 17(9), pp. 2613-8.

Chen, P.Y., Feng, S., Joo, J.W., Jacobsen, S.E. and Pellegrini, M. (2011) 'A comparative analysis of DNA methylation across human embryonic stem cell lines', *Genome Biol*, 12(7), p. R62.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y. and Li, E. (2007) 'Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells', *Nat Genet*, 39(3), pp. 391-6.

Chim, C.S., Liang, R., Tam, C.Y. and Kwong, Y.L. (2001a) 'Methylation of p15 and p16 genes in acute promyelocytic leukemia: potential diagnostic and prognostic significance', *J Clin Oncol*, 19(7), pp. 2033-40.

Chim, C.S., Tam, C.Y., Liang, R. and Kwong, Y.L. (2001b) 'Methylation of p15 and p16 genes in adult acute leukemia: lack of prognostic significance', *Cancer*, 91(12), pp. 2222-9.

Cho, Y.H., Yazici, H., Wu, H.C., Terry, M.B., Gonzalez, K., Qu, M., Dalay, N. and Santella, R.M. (2010) 'Aberrant promoter hypermethylation and genomic hypomethylation in tumor, adjacent normal tissues and blood from breast cancer patients', *Anticancer Res*, 30(7), pp. 2489-96.

Christensen, B.C., Houseman, E.A., Marsit, C.J., Zheng, S., Wrensch, M.R., Wiemels, J.L., Nelson, H.H., Karagas, M.R., Padbury, J.F., Bueno, R., Sugarbaker, D.J., Yeh, R.F., Wiencke, J.K. and Kelsey, K.T. (2009) 'Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context', *PLoS Genet*, 5(8), p. e1000602.

Christman, J.K. (2002) '5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy', *Oncogene*, 21(35), pp. 5483-95.

Clark, S.J., Statham, A., Stirzaker, C., Molloy, P.L. and Frommer, M. (2006) 'DNA methylation: bisulphite modification and analysis', *Nat Protoc*, 1(5), pp. 2353-64.

Collerton, J., Barrass, K., Bond, J., Eccles, M., Jagger, C., James, O., Martin-Ruiz, C., Robinson, L., von Zglinicki, T. and Kirkwood, T. (2007) 'The Newcastle 85+ study: biological, clinical and psychosocial factors associated with healthy ageing: study protocol', *BMC Geriatr*, 7, p. 14.

Collerton, J., Davies, K., Jagger, C., Kingston, A., Bond, J., Eccles, M.P., Robinson, L.A., Martin-Ruiz, C., von Zglinicki, T., James, O.F. and Kirkwood, T.B. (2009) 'Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study', *Bmj*, 339, p. b4904.

Cooke, M.S., Evans, M.D., Dizdaroglu, M. and Lunec, J. (2003) 'Oxidative DNA damage: mechanisms, mutation, and disease', *Faseb J*, 17(10), pp. 1195-214.

Corn, P.G., Kuerbitz, S.J., van Noesel, M.M., Esteller, M., Compitello, N., Baylin, S.B. and Herman, J.G. (1999) 'Transcriptional silencing of the p73 gene in acute lymphoblastic leukemia and Burkitt's lymphoma is associated with 5' CpG island methylation', *Cancer Res*, 59(14), pp. 3352-6.

Corn, P.G., Smith, B.D., Ruckdeschel, E.S., Douglas, D., Baylin, S.B. and Herman, J.G. (2000) 'E-cadherin expression is silenced by 5' CpG island methylation in acute leukemia', *Clin Cancer Res*, 6(11), pp. 4243-8.

Costa, F.F. (2008) 'Non-coding RNAs, epigenetics and complexity', *Gene*, 410(1), pp. 9-17.

Costello, J.F., Berger, M.S., Huang, H.S. and Cavenee, W.K. (1996) 'Silencing of p16/CDKN2 expression in human gliomas by methylation and chromatin condensation', *Cancer Res*, 56(10), pp. 2405-10.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C., Schuller, D.E., Yu, L., Bloomfield, C.D., Caligiuri, M.A., Yates, A., Nishikawa, R., Huang, H.J.S., Petrelli, N.J., Zhang, X.L., O'Dorisio, M.S., Held, W.A., Cavenee, W.K. and Plass, C. (2000) 'Aberrant CpG-island methylation has non-random and tumour-type-specific patterns', *Nat Genet*, 24(2), pp. 132-138.

Coulondre, C., Miller, J.H., Farabaugh, P.J. and Gilbert, W. (1978) 'Molecular basis of base substitution hotspots in Escherichia coli', *Nature*, 274(5673), pp. 775-80.

Crooks, G.M., Fuller, J., Petersen, D., Izadi, P., Malik, P., Pattengale, P.K., Kohn, D.B. and Gasson, J.C. (1999) 'Constitutive HOXA5 expression inhibits erythropoiesis and increases myelopoiesis from human hematopoietic progenitors', *Blood*, 94(2), pp. 519-28.

Cui, K., Zang, C., Roh, T.Y., Schones, D.E., Childs, R.W., Peng, W. and Zhao, K. (2009) 'Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation', *Cell Stem Cell*, 4(1), pp. 80-93.

Dammann, R., Schagdarsurengin, U., Seidel, C., Strunnikova, M., Rastetter, M., Baier, K. and Pfeifer, G.P. (2005) 'The tumor suppressor RASSF1A in human carcinogenesis: an update', *Histol Histopathol*, 20(2), pp. 645-63.

Daskalos, A., Nikolaidis, G., Xinarianos, G., Savvari, P., Cassidy, A., Zakopoulou, R., Kotsinas, A., Gorgoulis, V., Field, J.K. and Liloglou, T. (2009) 'Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer', *Int J Cancer*, 124(1), pp. 81-7.

De Benedictis, G., Rose, G., Carrieri, G., De Luca, M., Falcone, E., Passarino, G., Bonafe, M., Monti, D., Baggio, G., Bertolini, S., Mari, D., Mattace, R. and Franceschi, C. (1999) 'Mitochondrial DNA inherited variants are associated with successful aging and longevity in humans', *Faseb J*, 13(12), pp. 1532-6.

De Bustos, C., Ramos, E., Young, J.M., Tran, R.K., Menzel, U., Langford, C.F., Eichler, E.E., Hsu, L., Henikoff, S., Dumanski, J.P. and Trask, B.J. (2009) 'Tissue-specific variation in DNA methylation levels along human chromosome 1', *Epigenetics Chromatin*, 2(1), p. 7.

De Smet, C., Lurquin, C., Lethe, B., Martelange, V. and Boon, T. (1999) 'DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter', *Mol Cell Biol*, 19(11), pp. 7327-35.

DeChiara, T.M., Robertson, E.J. and Efstratiadis, A. (1991) 'Parental imprinting of the mouse insulin-like growth factor II gene', *Cell*, 64(4), pp. 849-59.

Den Elzen, W.P., Martin-Ruiz, C., von Zglinicki, T., Westendorp, R.G., Kirkwood, T.B. and Gussekloo, J. (2011) 'Telomere length and anaemia in old age: results from the Newcastle 85-plus Study and the Leiden 85-plus Study', *Age Ageing*, 40(4), pp. 494-500.

Duncan, B.K. and Miller, J.H. (1980) 'Mutagenic deamination of cytosine residues in DNA', *Nature*, 287(5782), pp. 560-1.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., Haefliger, C., Horton, R., Howe, K., Jackson, D.K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K. and Beck, S. (2006) 'DNA methylation profiling of human chromosomes 6, 20 and 22', *Nat Genet*, 38(12), pp. 1378-85.

Egger, G., Jeong, S., Escobar, S.G., Cortez, C.C., Li, T.W., Saito, Y., Yoo, C.B., Jones, P.A. and Liang, G. (2006) 'Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival', *Proc Natl Acad Sci U S A*, 103(38), pp. 14080-5.

Ehrlich, M. (2002) 'DNA methylation in cancer: too much, but also too little', *Oncogene*, 21(35), pp. 5400-13.

Ehrlich, M. (2003) 'The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease', *Clin Immunol*, 109(1), pp. 17-28.

El-Maarri, O., Becker, T., Junen, J., Manzoor, S.S., Diaz-Lacava, A., Schwaab, R., Wienker, T. and Oldenburg, J. (2007) 'Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males', *Hum Genet*, 122(5), pp. 505-14.

Esteller, M. (2007) 'Epigenetic gene silencing in cancer: the DNA hypermethylome', *Hum Mol Genet*, 16 Spec No 1, pp. R50-9.

Esteller, M., Silva, J.M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., Bussaglia, E., Prat, J., Harkes, I.C., Repasky, E.A., Gabrielson, E., Schutte, M., Baylin, S.B. and Herman, J.G. (2000a) 'Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors', *J Natl Cancer Inst*, 92(7), pp. 564-9.

Esteller, M., Sparks, A., Toyota, M., Sanchez-Cespedes, M., Capella, G., Peinado, M.A., Gonzalez, S., Tarafa, G., Sidransky, D., Meltzer, S.J., Baylin, S.B. and Herman, J.G. (2000b) 'Analysis of adenomatous polyposis coli promoter hypermethylation in human cancer', *Cancer Res*, 60(16), pp. 4366-71.

Feil, R. (2006) 'Environmental and nutritional effects on the epigenetic regulation of genes', *Mutat Res*, 600(1-2), pp. 46-57.

Ferguson, S., Gautrey, H.E. and Strathdee, G. (2011) 'The dual role of HLXB9 in leukemia', *Pediatr Blood Cancer*, 56(3), pp. 349-52.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S. and Reik, W. (2011) 'Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation', *Nature*, 473(7347), pp. 398-402.

Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., Tallman, M.S., Sun, Z., Wolniak, K., Peeters, J.K., Liu, W., Choe, S.E., Fantin, V.R., Paietta, E., Lowenberg, B., Licht, J.D., Godley, L.A., Delwel, R., Valk, P.J., Thompson, C.B., Levine, R.L. and Melnick, A. (2010) 'Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation', *Cancer Cell*, 18(6), pp. 553-67.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Urioste, M., Benitez, J., Boix-Chornet, M., Sanchez-Aguilera, A., Ling, C., Carlsson, E., Poulsen, P., Vaag, A., Stephan, Z., Spector, T.D., Wu, Y.Z., Plass, C. and Esteller, M. (2005a) 'Epigenetic differences arise during the lifetime of monozygotic twins', *Proc Natl Acad Sci U S A*, 102(30), pp. 10604-9.

Fraga, M.F., Ballestar, E., Villar-Garea, A., Boix-Chornet, M., Espada, J., Schotta, G., Bonaldi, T., Haydon, C., Ropero, S., Petrie, K., Iyer, N.G., Perez-Rosado, A., Calvo, E., Lopez, J.A., Cano, A., Calasanz, M.J., Colomer, D., Piris, M.A., Ahn, N., Imhof, A., Caldas, C., Jenuwein, T. and Esteller, M. (2005b) 'Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer', *Nat Genet*, 37(4), pp. 391-400.

Franco, R., Schoneveld, O., Georgakilas, A.G. and Panayiotidis, M.I. (2008) 'Oxidative stress, DNA methylation and carcinogenesis', *Cancer Lett*, 266(1), pp. 6-11.

Freedman, N.D., Leitzmann, M.F., Hollenbeck, A.R., Schatzkin, A. and Abnet, C.C. (2008) 'Cigarette smoking and subsequent risk of lung cancer in men and women: analysis of a prospective cohort study', *Lancet Oncol*, 9(7), pp. 649-56.

Fried, L.P., Tangen, C.M., Walston, J., Newman, A.B., Hirsch, C., Gottdiener, J., Seeman, T., Tracy, R., Kop, W.J., Burke, G. and McBurnie, M.A. (2001) 'Frailty in older adults: evidence for a phenotype', *J Gerontol A Biol Sci Med Sci*, 56(3), pp. M146-56.

Fuke, C., Shimabukuro, M., Petronis, A., Sugimoto, J., Oda, T., Miura, K., Miyazaki, T., Ogura, C., Okazaki, Y. and Jinno, Y. (2004) 'Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study', *Ann Hum Genet*, 68(Pt 3), pp. 196-204.

Fuller, J.F., McAdara, J., Yaron, Y., Sakaguchi, M., Fraser, J.K. and Gasson, J.C. (1999) 'Characterization of HOX gene expression during myelopoiesis: role of HOX A5 in lineage commitment and maturation', *Blood*, 93(10), pp. 3391-400.

Fulop, T., Larbi, A., Witkowski, J.M., McElhaney, J., Loeb, M., Mitnitski, A. and Pawelec, G. (2010) 'Aging, frailty and age-related diseases', *Biogerontology*, 11(5), pp. 547-63.

Gagnon-Kugler, T., Langlois, F., Stefanovsky, V., Lessard, F. and Moss, T. (2009) 'Loss of human ribosomal gene CpG methylation enhances cryptic RNA polymerase II transcription and disrupts ribosomal RNA processing', *Mol Cell*, 35(4), pp. 414-25.

Galetzka, D., Hansmann, T., El Hajj, N., Weis, E., Irmscher, B., Ludwig, M., Schneider-Ratzke, B., Kohlschmidt, N., Beyer, V., Bartsch, O., Zechner, U., Spix, C. and Haaf, T. (2012) 'Monozygotic twins discordant for constitutive BRCA1 promoter methylation, childhood cancer and secondary cancer', *Epigenetics*, 7(1), pp. 47-54.

Garcia-Manero, G., Daniel, J., Smith, T.L., Kornblau, S.M., Lee, M.S., Kantarjian, H.M. and Issa, J.P. (2002) 'DNA methylation of multiple promoter-associated CpG islands in adult acute lymphocytic leukemia', *Clin Cancer Res*, 8(7), pp. 2217-24.

Garcia-Ramirez, M., Rocchini, C. and Ausio, J. (1995) 'Modulation of chromatin folding by histone acetylation', *J Biol Chem*, 270(30), pp. 17923-8.

Gardiner-Garden, M. and Frommer, M. (1987) 'CpG islands in vertebrate genomes', *J Mol Biol*, 196(2), pp. 261-82.

Gervin, K., Hammero, M., Akselsen, H.E., Moe, R., Nygard, H., Brandt, I., Gjessing, H.K., Harris, J.R., Undlien, D.E. and Lyle, R. (2011) 'Extensive variation and low heritability of DNA methylation identified in a twin study', *Genome Res*, 21(11), pp. 1813-21.

Gervin, K., Vigeland, M.D., Mattingsdal, M., Hammero, M., Nygard, H., Olsen, A.O., Brandt, I., Harris, J.R., Undlien, D.E. and Lyle, R. (2012) 'DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes', *PLoS Genet*, 8(1), p. e1002454.

Ghosh, S., Yates, A.J., Fruhwald, M.C., Miecznikowski, J.C., Plass, C. and Smiraglia, D. (2010) 'Tissue specific DNA methylation of CpG islands in normal human adult somatic tissues distinguishes neural from non-neural tissues', *Epigenetics*, 5(6), pp. 527-38.

Giannoukakis, N., Deal, C., Paquette, J., Goodyer, C.G. and Polychronakos, C. (1993) 'Parental genomic imprinting of the human IGF2 gene', *Nat Genet*, 4(1), pp. 98-101.

Gibney, E.R. and Nolan, C.M. (2010) 'Epigenetics and gene expression', *Heredity*.

Glozak, M.A., Sengupta, N., Zhang, X. and Seto, E. (2005) 'Acetylation and deacetylation of non-histone proteins', *Gene*, 363, pp. 15-23.

Goelz, S.E., Vogelstein, B., Hamilton, S.R. and Feinberg, A.P. (1985) 'Hypomethylation of DNA from benign and malignant human colon neoplasms', *Science*, 228(4696), pp. 187-90.

Goll, M.G. and Bestor, T.H. (2005) 'Eukaryotic cytosine methyltransferases', *Annu Rev Biochem*, 74, pp. 481-514.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E. and Bestor, T.H. (2006) 'Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2', *Science*, 311(5759), pp. 395-8.

Gomez-Duran, A., Pacheu-Grau, D., Martinez-Romero, I., Lopez-Gallardo, E., Lopez-Perez, M.J., Montoya, J. and Ruiz-Pesini, E. (2012) 'Oxidative phosphorylation differences between mitochondrial DNA haplogroups modify the risk of Leber's hereditary optic neuropathy', *Biochimica Et Biophysica Acta-Molecular Basis of Disease*, 1822(8), pp. 1216-1222.

Goyal, R., Reinhardt, R. and Jeltsch, A. (2006) 'Accuracy of DNA methylation pattern preservation by the Dnmt1 methyltransferase', *Nucleic Acids Res*, 34(4), pp. 1182-8.

Graff, J.R., Herman, J.G., Myohanen, S., Baylin, S.B. and Vertino, P.M. (1997) 'Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methylation', *J Biol Chem*, 272(35), pp. 22322-9.

Grant, P.A. (2001) 'A tale of histone modifications', *Genome Biol*, 2(4), p. REVIEWS0003.

Greenblatt, M.S., Bennett, W.P., Hollstein, M. and Harris, C.C. (1994) 'Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis', *Cancer Res*, 54(18), pp. 4855-78.

Gronniger, E., Weber, B., Heil, O., Peters, N., Stab, F., Wenck, H., Korn, B., Winnefeld, M. and Lyko, F. (2010) 'Aging and chronic sun exposure cause distinct epigenetic changes in human skin', *PLoS Genet*, 6(5), p. e1000971.

Groop, L. and Lyssenko, V. (2008) 'Genes and type 2 diabetes mellitus', *Curr Diab Rep*, 8(3), pp. 192-7.

Hammons, G.J., Yan, Y., Lopatina, N.G., Jin, B., Wise, C., Blann, E.B., Poirier, L.A., Kadlubar, F.F. and Lyn-Cook, B.D. (1999) 'Increased expression of hepatic DNA methyltransferase in smokers', *Cell Biol Toxicol*, 15(6), pp. 389-94.

Hayashi, A., Horiuchi, A., Kikuchi, N., Hayashi, T., Fuseya, C., Suzuki, A., Konishi, I. and Shiozawa, T. (2010) 'Type-specific roles of histone deacetylase (HDAC) overexpression in ovarian carcinoma: HDAC1 enhances cell proliferation and HDAC3 stimulates cell migration with downregulation of E-cadherin', *Int J Cancer*, 127(6), pp. 1332-46.

Hazra, A., Fuchs, C.S., Kawasaki, T., Kirkner, G.J., Hunter, D.J. and Ogino, S. (2010) 'Germline polymorphisms in the one-carbon metabolism pathway and DNA methylation in colorectal cancer', *Cancer Causes Control*, 21(3), pp. 331-45.

Heijmans, B.T., Kremer, D., Tobi, E.W., Boomsma, D.I. and Slagboom, P.E. (2007) 'Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus', *Hum Mol Genet*, 16(5), pp. 547-54.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Slagboom, P.E. and Lumey, L.H. (2008) 'Persistent epigenetic differences associated with prenatal exposure to famine in humans', *Proc Natl Acad Sci U S A*, 105(44), pp. 17046-9.

Hellman, A. and Chess, A. (2007) 'Gene body-specific methylation on the active X chromosome', *Science*, 315(5815), pp. 1141-3.

Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.P., Markowitz, S., Willson, J.K., Hamilton, S.R., Kinzler, K.W., Kane, M.F., Kolodner, R.D., Vogelstein, B., Kunkel, T.A. and Baylin, S.B. (1998) 'Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma', *Proc Natl Acad Sci U S A*, 95(12), pp. 6870-5.

Hervouet, E., Vallette, F.M. and Cartron, P.F. (2009) 'Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation', *Epigenetics*, 4(7), pp. 487-99.

Heyn, H., Li, N., Ferreira, H.J., Moran, S., Pisano, D.G., Gomez, A., Diez, J., Sanchez-Mut, J.V., Setien, F., Carmona, F.J., Puca, A.A., Sayols, S., Pujana, M.A., Serra-Musach, J., Iglesias-Platas, I., Formiga, F., Fernandez, A.F., Fraga, M.F., Heath, S.C., Valencia, A., Gut, I.G., Wang, J. and Esteller, M. (2012) 'Distinct DNA methylomes of newborns and centenarians', *Proc Natl Acad Sci U S A*.

Hofhuis, W., de Jongste, J.C. and Merkus, P.J. (2003) 'Adverse health effects of prenatal and postnatal tobacco smoke exposure on children', *Arch Dis Child*, 88(12), pp. 1086-90.

Hollander, M., Koudstaal, P.J., Bots, M.L., Grobbee, D.E., Hofman, A. and Breteler, M.M. (2003) 'Incidence, risk, and case fatality of first ever stroke in the elderly population. The Rotterdam Study', *J Neurol Neurosurg Psychiatry*, 74(3), pp. 317-21.

Holler, M., Westin, G., Jiricny, J. and Schaffner, W. (1988) 'Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated', *Genes Dev*, 2(9), pp. 1127-35.

Holz-Schietinger, C., Matje, D.M. and Reich, N.O. (2012) 'Mutations in DNA methyltransferase (DNMT3A) observed in acute myeloid leukemia patients disrupt processive methylation', *Journal of Biological Chemistry*.

Holz-Schietinger, C. and Reich, N.O. (2010) 'The inherent processivity of the human de novo methyltransferase 3A (DNMT3A) is enhanced by DNMT3L', *J Biol Chem*, 285(38), pp. 29091-100.

Holzenberger, M., Dupont, J., Ducos, B., Leneuve, P., Geloen, A., Even, P.C., Cervera, P. and Le Bouc, Y. (2003) 'IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice', *Nature*, 421(6919), pp. 182-7.

Hou, Y.C., H.; He, Q.; Jiang, W.; Luo, T.; Duan, J.; Mu, N.; He, Y. and Wang, H. (2012) 'Changes in methylation patterns of multiple genes from peripheral blood leukocytes of Alzheimer's disease patients', *Acta Neuropsychiatrica*.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., Humphray, S., Cox, T., Langford, C. and Bird, A. (2008) 'A novel CpG island set identifies tissue-specific methylation at developmental gene loci', *PLoS Biol*, 6(1), p. e22.

Illingworth, R.S. and Bird, A.P. (2009) 'CpG islands--'a rough guide'', *FEBS Lett*, 583(11), pp. 1713-20.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R. and Bird, A.P. (2010) 'Orphan CpG islands identify numerous conserved promoters in the mammalian genome', *PLoS Genet*, 6(9).

Ingrosso, D., Cimmino, A., Perna, A.F., Masella, L., De Santo, N.G., De Bonis, M.L., Vacca, M., D'Esposito, M., D'Urso, M., Galletti, P. and Zappia, V. (2003) 'Folate

treatment and unbalanced methylation and changes of allelic expression induced by hyperhomocysteinaemia in patients with uraemia', *Lancet*, 361(9370), pp. 1693-9.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J.B., Sabunciyan, S. and Feinberg, A.P. (2009) 'The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores', *Nat Genet*, 41(2), pp. 178-86.

Irving, L., Mainou-Fowler, T., Parker, A., Ibbotson, R.E., Oscier, D.G. and Strathdee, G. (2011) 'Methylation markers identify high risk patients in IGHV mutated chronic lymphocytic leukemia', *Epigenetics*, 6(3), pp. 300-6.

Issa, J.P., Ottaviano, Y.L., Celano, P., Hamilton, S.R., Davidson, N.E. and Baylin, S.B. (1994) 'Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon', *Nat Genet*, 7(4), pp. 536-40.

Issa, J.P., Vertino, P.M., Boehm, C.D., Newsham, I.F. and Baylin, S.B. (1996a) 'Switch from monoallelic to biallelic human IGF2 promoter methylation during aging and carcinogenesis', *Proc Natl Acad Sci U S A*, 93(21), pp. 11757-62.

Issa, J.P., Zehnbauer, B.A., Civin, C.I., Collector, M.I., Sharkis, S.J., Davidson, N.E., Kaufmann, S.H. and Baylin, S.B. (1996b) 'The estrogen receptor CpG island is methylated in most hematopoietic neoplasms', *Cancer Res*, 56(5), pp. 973-77.

Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science*, 333(6047), pp. 1300-3.

Jacob, R.A., Gretz, D.M., Taylor, P.C., James, S.J., Pogribny, I.P., Miller, B.J., Henning, S.M. and Swendseid, M.E. (1998) 'Moderate folate depletion increases plasma homocysteine and decreases lymphocyte DNA methylation in postmenopausal women', *Journal of Nutrition*, 128(7), pp. 1204-1212.

Javierre, B.M., Fernandez, A.F., Richter, J., Al-Shahrour, F., Martin-Subero, J.I., Rodriguez-Ubreva, J., Berdasco, M., Fraga, M.F., O'Hanlon, T.P., Rider, L.G., Jacinto, F.V., Lopez-Longo, F.J., Dopazo, J., Forn, M., Peinado, M.A., Carreno, L., Sawalha, A.H., Harley, J.B., Siebert, R., Esteller, M., Miller, F.W. and Ballestar, E. (2010) 'Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus', *Genome Res*, 20(2), pp. 170-9.

Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., Rossi, D.J., Inlay, M.A., Serwold, T., Karsunky, H., Ho, L., Daley, G.Q., Weissman, I.L. and Feinberg, A.P. (2010) 'Comprehensive methylome map of lineage commitment from haematopoietic progenitors', *Nature*, 467(7313), pp. 338-42.

Jia, D., Jurkowska, R.Z., Zhang, X., Jeltsch, A. and Cheng, X. (2007) 'Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation', *Nature*, 449(7159), pp. 248-51.

Jintaridth, P. and Mutirangura, A. (2010) 'Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences', *Physiol Genomics*, 41(2), pp. 194-200.

Johnson, R.C., Nelson, G.W., Troyer, J.L., Lautenberger, J.A., Kessing, B.D., Winkler, C.A. and O'Brien, S.J. (2010) 'Accounting for multiple comparisons in a genome-wide association study (GWAS)', *BMC Genomics*, 11.

Jones, P.A. and Baylin, S.B. (2002) 'The fundamental role of epigenetic events in cancer', *Nat Rev Genet*, 3(6), pp. 415-28.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) 'Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription', *Nat Genet*, 19(2), pp. 187-91.

Jurkowska, R.Z., Jurkowski, T.P. and Jeltsch, A. (2011) 'Structure and function of mammalian DNA methyltransferases', *Chembiochem*, 12(2), pp. 206-22.

Kaminsky, Z.A., Tang, T., Wang, S.C., Ptak, C., Oh, G.H., Wong, A.H., Feldcamp, L.A., Virtanen, C., Halfvarson, J., Tysk, C., McRae, A.F., Visscher, P.M., Montgomery, G.W., Gottesman, II, Martin, N.G. and Petronis, A. (2009) 'DNA methylation profiles in monozygotic and dizygotic twins', *Nat Genet*, 41(2), pp. 240-5.

Kanfi, Y., Naiman, S., Amir, G., Peshti, V., Zinman, G., Nahum, L., Bar-Joseph, Z. and Cohen, H.Y. (2012) 'The sirtuin SIRT6 regulates lifespan in male mice', *Nature*, 483(7388), pp. 218-21.

Kantlehner, M., Kirchner, R., Hartmann, P., Ellwart, J.W., Alunni-Fabbroni, M. and Schumacher, A. (2011) 'A high-throughput DNA methylation analysis of a single cell', *Nucleic Acids Res*, 39(7), p. e44.

Kaufman, R.J. (1999) 'Stress signaling from the lumen of the endoplasmic reticulum: coordination of gene transcriptional and translational controls', *Genes Dev*, 13(10), pp. 1211-33.

Kawai, H., Li, H., Avraham, S., Jiang, S. and Avraham, H.K. (2003) 'Overexpression of histone deacetylase HDAC1 modulates breast cancer progression by negative regulation of estrogen receptor alpha', *Int J Cancer*, 107(3), pp. 353-8.

Kelkar, A. and Deobagkar, D. (2010) 'Methylation profile of genes on the human X chromosome', *Epigenetics*, 5(7), pp. 612-8.

Kim, K.M. and Shibata, D. (2002) 'Methylation reveals a niche: stem cell succession in human colon crypts', *Oncogene*, 21(35), pp. 5441-9.

Kim, M., Long, T.I., Arakawa, K., Wang, R.W., Yu, M.C. and Laird, P.W. (2010a) 'DNA Methylation as a Biomarker for Cardiovascular Disease Risk', *PLoS One*, 5(3).

Kim, S.Y., Hwang, S.H., Song, E.J., Shin, H.J., Jung, J.S. and Lee, E.Y. (2010b) 'Level of HOXA5 hypermethylation in acute myeloid leukemia is associated with short-term outcome', *Korean J Lab Med*, 30(5), pp. 469-73.

Klauke, K. and de Haan, G. (2011) 'Polycomb group proteins in hematopoietic stem cell aging and malignancies', *Int J Hematol*, 94(1), pp. 11-23.

Klose, R.J. and Bird, A.P. (2006) 'Genomic DNA methylation: the mark and its mediators', *Trends Biochem Sci*, 31(2), pp. 89-97.

Kondo, Y., Shen, L., Cheng, A.S., Ahmed, S., Boumber, Y., Charo, C., Yamochi, T., Urano, T., Furukawa, K., Kwabi-Addo, B., Gold, D.L., Sekido, Y., Huang, T.H. and Issa, J.P. (2008) 'Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation', *Nat Genet*, 40(6), pp. 741-50.

Kuang, S.Q., Bai, H., Fang, Z.H., Lopez, G., Yang, H., Tong, W., Wang, Z.Z. and Garcia-Manero, G. (2010) 'Aberrant DNA methylation and epigenetic inactivation of Eph receptor tyrosine kinases and ephrin ligands in acute lymphoblastic leukemia', *Blood*, 115(12), pp. 2412-9.

Kuroda, A., Rauch, T.A., Todorov, I., Ku, H.T., Al-Abdullah, I.H., Kandeel, F., Mullen, Y., Pfeifer, G.P. and Ferreri, K. (2009) 'Insulin gene expression is regulated by DNA methylation', *PLoS One*, 4(9), p. e6953.

Kwabi-Addo, B., Chung, W., Shen, L., Ittmann, M., Wheeler, T., Jelinek, J. and Issa, J.P. (2007) 'Age-related DNA methylation changes in normal human prostate tissues', *Clin Cancer Res*, 13(13), pp. 3796-802.

Lagger, G., O'Carroll, D., Rembold, M., Khier, H., Tischler, J., Weitzer, G., Schuettengruber, B., Hauser, C., Brunmeir, R., Jenuwein, T. and Seiser, C. (2002) 'Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression', *Embo J*, 21(11), pp. 2672-81.

Lakshmikuttyamma, A., Scott, S.A., DeCoteau, J.F. and Geyer, C.R. (2010) 'Reexpression of epigenetically silenced AML tumor suppressor genes by SUV39H1 inhibition', *Oncogene*, 29(4), pp. 576-88.

Lancaster, J.M., Carney, M.E. and Futreal, P.A. (1997) 'BRCA 1 and 2--A Genetic Link to Familial Breast and Ovarian Cancer', *Medscape Womens Health*, 2(2), p. 7.

Lapinski, P.E., Oliver, J.A., Kamen, L.A., Hughes, E.D., Saunders, T.L. and King, P.D. (2008) 'Genetic analysis of SH2D4A, a novel adapter protein related to T cell-specific adapter and adapter protein in lymphocytes of unknown function, reveals a redundant function in T cells', *J Immunol*, 181(3), pp. 2019-27.

Lee, K.K. and Workman, J.L. (2007) 'Histone acetyltransferase complexes: one size doesn't fit all', *Nat Rev Mol Cell Biol*, 8(4), pp. 284-95.

Lee, S., Hwang, K.S., Lee, H.J., Kim, J.S. and Kang, G.H. (2004) 'Aberrant CpG island hypermethylation of multiple genes in colorectal neoplasia', *Lab Invest*, 84(7), pp. 884-93.

Lennartsson, A. and Ekwall, K. (2009) 'Histone modification patterns and epigenetic codes', *Biochim Biophys Acta*, 1790(9), pp. 863-8.

Leonhardt, H., Page, A.W., Weier, H.U. and Bestor, T.H. (1992) 'A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei', *Cell*, 71(5), pp. 865-73.

Leu, Y.W., Yan, P.S., Fan, M., Jin, V.X., Liu, J.C., Curran, E.M., Welshons, W.V., Wei, S.H., Davuluri, R.V., Plass, C., Nephew, K.P. and Huang, T.H. (2004) 'Loss of estrogen receptor signaling triggers epigenetic silencing of downstream targets in breast cancer', *Cancer Res*, 64(22), pp. 8184-92.

Levine, A.J., Figueiredo, J.C., Lee, W., Poynter, J.N., Conti, D., Duggan, D.J., Campbell, P.T., Newcomb, P., Martinez, M.E., Hopper, J.L., Le Marchand, L., Baron, J.A., Limburg, P.J., Ulrich, C.M. and Haile, R.W. (2010) 'Genetic variability in the MTHFR gene and colorectal cancer risk using the colorectal cancer family registry', *Cancer Epidemiol Biomarkers Prev*, 19(1), pp. 89-100.

Li, E., Beard, C. and Jaenisch, R. (1993) 'Role for DNA methylation in genomic imprinting', *Nature*, 366(6453), pp. 362-5.

Li, E., Bestor, T.H. and Jaenisch, R. (1992) 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality', *Cell*, 69(6), pp. 915-26.

Li, T., Li, W., Lu, J., Liu, H., Li, Y. and Zhao, Y. (2009) 'SH2D4A regulates cell proliferation via the ERalpha/PLC-gamma/PKC pathway', *BMB Rep*, 42(8), pp. 516-22.

Lin, C.H., Hsieh, S.Y., Sheen, I.S., Lee, W.C., Chen, T.C., Shyu, W.C. and Liaw, Y.F. (2001) 'Genome-wide hypomethylation in hepatocellular carcinogenesis', *Cancer Res*, 61(10), pp. 4238-43.

Ling, C. and Groop, L. (2009) 'Epigenetics: a molecular link between environmental factors and type 2 diabetes', *Diabetes*, 58(12), pp. 2718-25.

Linhart, H.G., Lin, H., Yamada, Y., Moran, E., Steine, E.J., Gokhale, S., Lo, G., Cantu, E., Ehrich, M., He, T., Meissner, A. and Jaenisch, R. (2007) 'Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing', *Genes Dev*, 21(23), pp. 3110-22.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B. and Ecker, J.R. (2009) 'Human DNA methylomes at base resolution show widespread epigenomic differences', *Nature*, 462(7271), pp. 315-22.

Liu, J., Morgan, M., Hutchison, K. and Calhoun, V.D. (2010) 'A study of the influence of sex on genome wide methylation', *PLoS One*, 5(4), p. e10028.

Liu, L., van Groen, T., Kadish, I., Li, Y., Wang, D., James, S.R., Karpf, A.R. and Tollefsbol, T.O. (2011) 'Insufficient DNA methylation affects healthy aging and promotes age-related health problems', *Clin Epigenetics*, 2(2), pp. 349-60.

Longo, V.D. and Kennedy, B.K. (2006) 'Sirtuins in aging and age-related disease', *Cell*, 126(2), pp. 257-68.

Lorincz, M.C., Dickerson, D.R., Schmitt, M. and Groudine, M. (2004) 'Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells', *Nat Struct Mol Biol*, 11(11), pp. 1068-75.

Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W. and Cui, Q. (2008) 'An analysis of human microRNA and disease associations', *PLoS One*, 3(10), p. e3420.

Maegawa, S., Hinkal, G., Kim, H.S., Shen, L., Zhang, L., Zhang, J., Zhang, N., Liang, S., Donehower, L.A. and Issa, J.P. (2010) 'Widespread and tissue specific age-related DNA methylation changes in mice', *Genome Res*, 20(3), pp. 332-40.

Margueron, R. and Reinberg, D. (2011) 'The Polycomb complex PRC2 and its mark in life', *Nature*, 469(7330), pp. 343-9.

Martin-Ruiz, C., Jagger, C., Kingston, A., Collerton, J., Catt, M., Davies, K., Dunn, M., Hilkens, C., Keavney, B., Pearce, S.H., den Elzen, W.P., Talbot, D., Wiley, L., Bond, J., Mathers, J.C., Eccles, M.P., Robinson, L., James, O., Kirkwood, T.B. and von Zglinicki, T. (2011) 'Assessment of a large panel of candidate biomarkers of ageing in the Newcastle 85+ study', *Mech Ageing Dev*, 132(10), pp. 496-502.

Martin-Subero, J.I., Ammerpohl, O., Bibikova, M., Wickham-Garcia, E., Agirre, X., Alvarez, S., Bruggemann, M., Bug, S., Calasanz, M.J., Deckert, M., Dreyling, M., Du, M.Q., Durig, J., Dyer, M.J., Fan, J.B., Gesk, S., Hansmann, M.L., Harder, L., Hartmann, S., Klapper, W., Kuppers, R., Montesinos-Rongen, M., Nagel, I., Pott, C., Richter, J., Roman-Gomez, J., Seifert, M., Stein, H., Suela, J., Trumper, L., Vater, I., Prosper, F., Haferlach, C., Cruz Cigudosa, J. and Siebert, R. (2009) 'A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms', *PLoS One*, 4(9), p. e6986.

Martinez-Redondo, D., Marcuello, A., Casajus, J.A., Ara, I., Dahmani, Y., Montoya, J., Ruiz-Pesini, E., Lopez-Perez, M.J. and Diez-Sanchez, C. (2010) 'Human mitochondrial haplogroup H: The highest VO2max consumer - Is it a paradox?', *Mitochondrion*, 10(2), pp. 102-107.

Mastroeni, D., Grover, A., Delvaux, E., Whiteside, C., Coleman, P.D. and Rogers, J. (2011) 'Epigenetic mechanisms in Alzheimer's disease', *Neurobiol Aging*, 32(7), pp. 1161-80.

Matthews, F. and Brayne, C. (2005) 'The incidence of dementia in England and Wales: findings from the five identical sites of the MRC CFA Study', *PLoS Med*, 2(8), p. e193.

McGowan, P.O., Sasaki, A., D'Alessio, A.C., Dymov, S., Labonte, B., Szyf, M., Turecki, G. and Meaney, M.J. (2009) 'Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse', *Nat Neurosci*, 12(3), pp. 342-8.

McKay, J.A., Williams, E.A. and Mathers, J.C. (2004) 'Folate and DNA methylation during in utero development and aging', *Biochem Soc Trans*, 32(Pt 6), pp. 1006-7.

Melki, J.R., Vincent, P.C. and Clark, S.J. (1999) 'Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia', *Cancer Res*, 59(15), pp. 3730-40.

Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K.W., Vogelstein, B. and Nakamura, Y. (1992) 'Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer', *Cancer Res*, 52(3), pp. 643-5.

Milagro, F.I., Campion, J., Garcia-Diaz, D.F., Goyenechea, E., Paternain, L. and Martinez, J.A. (2009) 'High fat diet-induced obesity modifies the methylation pattern of leptin promoter in rats', *J Physiol Biochem*, 65(1), pp. 1-9.

Mill, J., Dempster, E., Caspi, A., Williams, B., Moffitt, T. and Craig, I. (2006) 'Evidence for monozygotic twin (MZ) discordance in methylation level at two CpG sites in the promoter region of the catechol-O-methyltransferase (COMT) gene', *Am J Med Genet B Neuropsychiatr Genet*, 141B(4), pp. 421-5.

Montesanto, A., Lagani, V., Martino, C., Dato, S., De Rango, F., Berardelli, M., Corsonello, A., Mazzei, B., Mari, V., Lattanzio, F., Conforti, D. and Passarino, G. (2010) 'A novel, population-specific approach to define frailty', *Age (Dordr)*, 32(3), pp. 385-95.

Morgan, D.K. and Whitelaw, E. (2008) 'The case for transgenerational epigenetic inheritance in humans', *Mamm Genome*, 19(6), pp. 394-7.

Morgan, H.D., Sutherland, H.G., Martin, D.I. and Whitelaw, E. (1999) 'Epigenetic inheritance at the agouti locus in the mouse', *Nat Genet*, 23(3), pp. 314-8.

Morse, B., Rotherg, P.G., South, V.J., Spandorfer, J.M. and Astrin, S.M. (1988) 'Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma', *Nature*, 333(6168), pp. 87-90.

Murrell, A., Ito, Y., Verde, G., Huddleston, J., Woodfine, K., Silengo, M.C., Spreafico, F., Perotti, D., De Crescenzo, A., Sparago, A., Cerrato, F. and Riccio, A. (2008) 'Distinct Methylation Changes at the IGF2-H19 Locus in Congenital Growth Disorders and Cancer', *PLoS One*, 3(3).

Nakagawa, H., Nuovo, G.J., Zervos, E.E., Martin, E.W., Jr., Salovaara, R., Aaltonen, L.A. and de la Chapelle, A. (2001) 'Age-related hypermethylation of the 5' region of MLH1 in normal colonic mucosa is associated with microsatellite-unstable colorectal cancer development', *Cancer Res*, 61(19), pp. 6991-5.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) 'Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex', *Nature*, 393(6683), pp. 386-9.

Nguyen, C.T., Weisenberger, D.J., Velicescu, M., Gonzales, F.A., Lin, J.C., Liang, G. and Jones, P.A. (2002) 'Histone H3-lysine 9 methylation is associated with aberrant gene silencing in cancer cells and is rapidly reversed by 5-aza-2'-deoxycytidine', *Cancer Res*, 62(22), pp. 6456-61.

Niehrs, C. and Schafer, A. (2012) 'Active DNA demethylation by Gadd45 and DNA repair', *Trends Cell Biol*, 22(4), pp. 220-7.

Nile, C.J., Read, R.C., Akil, M., Duff, G.W. and Wilson, A.G. (2008) 'Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis', *Arthritis Rheum*, 58(9), pp. 2686-93.

Nilsen, H., Haushalter, K.A., Robins, P., Barnes, D.E., Verdine, G.L. and Lindahl, T. (2001) 'Excision of deaminated cytosine from the vertebrate genome: role of the SMUG1 uracil-DNA glycosylase', *Embo J*, 20(15), pp. 4278-86.

Nilsen, T.W. (2007) 'Mechanisms of microRNA-mediated gene regulation in animal cells', *Trends Genet*, 23(5), pp. 243-9.

Nishigaki, Y., Fuku, N. and Tanaka, M. (2010) 'Mitochondrial haplogroups associated with lifestyle-related diseases and longevity in the Japanese population', *Geriatr Gerontol Int*, 10 Suppl 1, pp. S221-35.

Nyce, J.W. (1997) 'Drug-induced DNA hypermethylation: a potential mediator of acquired drug resistance during cancer chemotherapy', *Mutat Res*, 386(2), pp. 153-61.

O'Hagan, H.M., Wang, W., Sen, S., Destefano Shields, C., Lee, S.S., Zhang, Y.W., Clements, E.G., Cai, Y., Van Neste, L., Easwaran, H., Casero, R.A., Sears, C.L. and Baylin, S.B. (2011) 'Oxidative damage targets complexes containing DNA methyltransferases, SIRT1, and polycomb members to promoter CpG Islands', *Cancer Cell*, 20(5), pp. 606-19.

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., Berman, D.M., Jenuwein, T., Pruitt, K., Sharkis, S.J., Watkins, D.N., Herman, J.G. and Baylin, S.B. (2007) 'A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing', *Nat Genet*, 39(2), pp. 237-42.

Okano, M., Bell, D.W., Haber, D.A. and Li, E. (1999) 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development', *Cell*, 99(3), pp. 247-57.

Okano, M., Xie, S. and Li, E. (1998) 'Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases', *Nat Genet*, 19(3), pp. 219-20.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., Cheng, X. and Bestor, T.H. (2007) 'DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA', *Nature*, 448(7154), pp. 714-7.

Ordovas, J.M. and Smith, C.E. (2010) 'Epigenetics and cardiovascular disease', *Nat Rev Cardiol*, 7(9), pp. 510-9.

Ottaviano, Y.L., Issa, J.P., Parl, F.F., Smith, H.S., Baylin, S.B. and Davidson, N.E. (1994) 'Methylation of the estrogen receptor gene CpG island marks loss of estrogen receptor expression in human breast cancer cells', *Cancer Res*, 54(10), pp. 2552-5.

Pearson, T.A. and Manolio, T.A. (2008) 'How to interpret a genome-wide association study', *Jama-Journal of the American Medical Association*, 299(11), pp. 1335-1344.

Peleg, S., Sananbenesi, F., Zovoilis, A., Burkhardt, S., Bahari-Javan, S., Agis-Balboa, R.C., Cota, P., Wittnam, J.L., Gogol-Doering, A., Opitz, L., Salinas-Riester, G., Dettenhofer, M., Kang, H., Farinelli, L., Chen, W. and Fischer, A. (2010) 'Altered Histone Acetylation Is Associated with Age-Dependent Memory Impairment in Mice', *Science*, 328(5979), pp. 753-756.

Pellman, D. (2007) 'Cell biology: aneuploidy and cancer', *Nature*, 446(7131), pp. 38-9.

Peters, I., Vaske, B., Albrecht, K., Kuczyk, M.A., Jonas, U. and Serth, J. (2007) 'Adiposity and age are statistically related to enhanced RASSF1A tumor suppressor gene promoter methylation in normal autopsy kidney tissue', *Cancer Epidemiol Biomarkers Prev*, 16(12), pp. 2526-32.

Petronis, A., Gottesman, II, Kan, P., Kennedy, J.L., Basile, V.S., Paterson, A.D. and Popendikyte, V. (2003) 'Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance?', *Schizophr Bull*, 29(1), pp. 169-78.

Pineda, M., Mur, P., Iniesta, M.D., Borras, E., Campos, O., Vargas, G., Iglesias, S., Fernandez, A., Gruber, S.B., Lazaro, C., Brunet, J., Navarro, M., Blanco, I. and Capella,

G. (2012) 'MLH1 methylation screening is effective in identifying epimutation carriers', *Eur J Hum Genet*.

Plaschke, J., Schwanebeck, U., Pistorius, S., Saeger, H.D. and Schackert, H.K. (2003) 'Methylenetetrahydrofolate reductase polymorphisms and risk of sporadic and hereditary colorectal cancer with or without microsatellite instability', *Cancer Lett*, 191(2), pp. 179-85.

Poynter, J.N., Siegmund, K.D., Weisenberger, D.J., Long, T.I., Thibodeau, S.N., Lindor, N., Young, J., Jenkins, M.A., Hopper, J.L., Baron, J.A., Buchanan, D., Casey, G., Levine, A.J., Le Marchand, L., Gallinger, S., Bapat, B., Potter, J.D., Newcomb, P.A., Haile, R.W. and Laird, P.W. (2008) 'Molecular characterization of MSI-H colorectal cancer by MLHI promoter methylation, immunohistochemistry, and mismatch repair germline mutation screening', *Cancer Epidemiol Biomarkers Prev*, 17(11), pp. 3208-15.

Protacio, R.U., Li, G., Lowary, P.T. and Widom, J. (2000) 'Effects of histone tail domains on the rate of transcriptional elongation through a nucleosome', *Mol Cell Biol*, 20(23), pp. 8866-78.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) 'LocusZoom: regional visualization of genome-wide association scan results', *Bioinformatics*, 26(18), pp. 2336-7.

Qin, Y., Liu, J.Y., Li, B., Sun, Z.L. and Sun, Z.F. (2004) 'Association of low p16INK4a and p15INK4b mRNAs expression with their CpG islands methylation with human hepatocellular carcinogenesis', *World J Gastroenterol*, 10(9), pp. 1276-80.

Raetz, E.A., Perkins, S.L., Bhojwani, D., Smock, K., Philip, M., Carroll, W.L. and Min, D.J. (2006) 'Gene expression profiling reveals intrinsic differences between T-cell acute lymphoblastic leukemia and T-cell lymphoblastic lymphoma', *Pediatr Blood Cancer*, 47(2), pp. 130-40.

Rakyan, V.K., Beyan, H., Down, T.A., Hawa, M.I., Maslau, S., Aden, D., Daunay, A., Busato, F., Mein, C.A., Manfras, B., Dias, K.R., Bell, C.G., Tost, J., Boehm, B.O., Beck, S. and Leslie, R.D. (2011) 'Identification of type 1 diabetes-associated DNA methylation variable positions that precede disease diagnosis', *PLoS Genet*, 7(9), p. e1002300.

Rakyan, V.K., Chong, S., Champ, M.E., Cuthbert, P.C., Morgan, H.D., Luu, K.V. and Whitelaw, E. (2003) 'Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission', *Proc Natl Acad Sci U S A*, 100(5), pp. 2538-43.

Rakyan, V.K., Down, T.A., Maslau, S., Andrew, T., Yang, T.P., Beyan, H., Whittaker, P., McCann, O.T., Finer, S., Valdes, A.M., Leslie, R.D., Deloukas, P. and Spector, T.D. (2010) 'Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains', *Genome Res*, 20(4), pp. 434-9.

Rakyan, V.K., Down, T.A., Thorne, N.P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E.M., Backdahl, L., Johnson, N., Herberth, M., Howe, K.L., Jackson, D.K., Miretti, M.M., Fiegler, H., Marioni, J.C., Birney, E., Hubbard, T.J., Carter, N.P., Tavare, S. and Beck, S. (2008) 'An integrated resource for genome-wide identification and analysis of

human tissue-specific differentially methylated regions (tDMRs)', *Genome Res*, 18(9), pp. 1518-29.

Rampersaud, G.C., Kauwell, G.P.A., Hutson, A.D., Cerda, J.J. and Bailey, L.B. (2000) 'Genomic DNA methylation decreases in response to moderate folate depletion in elderly women', *American Journal of Clinical Nutrition*, 72(4), pp. 998-1003.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P. and Jaenisch, R. (2000) 'Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a', *Proc Natl Acad Sci U S A*, 97(10), pp. 5237-42.

Randhawa, G.S., Cui, H., Barletta, J.A., Strichman-Almashanu, L.Z., Talpaz, M., Kantarjian, H., Deisseroth, A.B., Champlin, R.C. and Feinberg, A.P. (1998) 'Loss of imprinting in disease progression in chronic myelogenous leukemia', *Blood*, 91(9), pp. 3144-7.

Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D. and Pfeifer, G.P. (2009) 'A human B cell methylome at 100-base pair resolution', *Proc Natl Acad Sci U S A*, 106(3), pp. 671-8.

Raval, A., Lucas, D.M., Matkovic, J.J., Bennett, K.L., Liyanarachchi, S., Young, D.C., Rassenti, L., Kipps, T.J., Grever, M.R., Byrd, J.C. and Plass, C. (2005) 'TWIST2 demonstrates differential methylation in immunoglobulin variable heavy chain mutated and unmutated chronic lymphocytic leukemia', *J Clin Oncol*, 23(17), pp. 3877-85.

Reinius, L.E., Acevedo, N., Joerink, M., Pershagen, G., Dahlen, S.E., Greco, D., Soderhall, C., Scheynius, A. and Kere, J. (2012) 'Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility', *PLoS One*, 7(7), p. e41361.

Rhee, I., Jair, K.W., Yen, R.W., Lengauer, C., Herman, J.G., Kinzler, K.W., Vogelstein, B., Baylin, S.B. and Schuebel, K.E. (2000) 'CpG methylation is maintained in human cancer cells lacking DNMT1', *Nature*, 404(6781), pp. 1003-7.

Richardson, B. (2003) 'Impact of aging on DNA methylation', *Ageing Res Rev*, 2(3), pp. 245-61.

Richon, V.M. and O'Brien, J.P. (2002) 'Histone deacetylase inhibitors: a new class of potential therapeutic agents for cancer treatment', *Clin Cancer Res*, 8(3), pp. 662-4.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E. and Chang, H.Y. (2007) 'Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs', *Cell*, 129(7), pp. 1311-23.

Risch, N. and Merikangas, K. (1996) 'The future of genetic studies of complex human diseases', *Science*, 273(5281), pp. 1516-7.

Rockwood, K., Song, X., MacKnight, C., Bergman, H., Hogan, D.B., McDowell, I. and Mitnitski, A. (2005) 'A global clinical measure of fitness and frailty in elderly people', *CMAJ*, 173(5), pp. 489-95.

Rodriguez, J., Frigola, J., Vendrell, E., Risques, R.A., Fraga, M.F., Morales, C., Moreno, V., Esteller, M., Capella, G., Ribas, M. and Peinado, M.A. (2006)

'Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers', *Cancer Res*, 66(17), pp. 8462-9468.

Roessler, S., Long, E.L., Budhu, A., Chen, Y., Zhao, X., Ji, J., Walker, R., Jia, H.L., Ye, Q.H., Qin, L.X., Tang, Z.Y., He, P., Hunter, K.W., Thorgeirsson, S.S., Meltzer, P.S. and Wang, X.W. (2012) 'Integrative genomic identification of genes on 8p associated with hepatocellular carcinoma progression and patient survival', *Gastroenterology*, 142(4), pp. 957-966 e12.

Roman-Gomez, J., Jimenez-Velasco, A., Agirre, X., Cervantes, F., Sanchez, J., Garate, L., Barrios, M., Castillejo, J.A., Navarro, G., Colomer, D., Prosper, F., Heiniger, A. and Torres, A. (2005) 'Promoter hypomethylation of the LINE-1 retrotransposable elements activates sense/antisense transcription and marks the progression of chronic myeloid leukemia', *Oncogene*, 24(48), pp. 7213-23.

Roseboom, T., de Rooij, S. and Painter, R. (2006) 'The Dutch famine and its long-term consequences for adult health', *Early Hum Dev*, 82(8), pp. 485-91.

Sarg, B., Koutzamani, E., Helliger, W., Rundquist, I. and Lindner, H.H. (2002) 'Postsynthetic trimethylation of histone H4 at lysine 20 in mammalian tissues is associated with aging', *J Biol Chem*, 277(42), pp. 39195-201.

Sarter, B., Long, T.I., Tsong, W.H., Koh, W.P., Yu, M.C. and Laird, P.W. (2005) 'Sex differential in methylation patterns of selected genes in Singapore Chinese', *Hum Genet*, 117(4), pp. 402-3.

Sasaki, H. and Matsui, Y. (2008) 'Epigenetic events in mammalian germ-cell development: reprogramming and beyond', *Nat Rev Genet*, 9(2), pp. 129-40.

Saxonov, S., Berg, P. and Brutlag, D.L. (2006) 'A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters', *Proc Natl Acad Sci U S A*, 103(5), pp. 1412-7.

Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C. and Leonhardt, H. (2007) 'Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation', *Nucleic Acids Res*, 35(13), pp. 4301-12.

Schneider, E., Pliushch, G., El Hajj, N., Galetzka, D., Puhl, A., Schorsch, M., Frauenknecht, K., Riepert, T., Tresch, A., Muller, A.M., Coerdt, W., Zechner, U. and Haaf, T. (2010) 'Spatial, temporal and interindividual epigenetic variation of functionally important DNA methylation patterns', *Nucleic Acids Res*.

Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C. and Kouzarides, T. (2004) 'Histone H3 lysine 4 methylation patterns in higher eukaryotic genes', *Nat Cell Biol*, 6(1), pp. 73-7.

Scholz, C., Nimmrich, I., Burger, M., Becker, E., Dorken, B., Ludwig, W.D. and Maier, S. (2005) 'Distinction of acute lymphoblastic leukemia from acute myeloid leukemia through microarray-based DNA methylation analysis', *Ann Hematol*, 84(4), pp. 236-44.

Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J., Bell, S.P. and Groudine, M. (2004) 'The histone modification pattern of active genes revealed through

genome-wide chromatin analysis of a higher eukaryote', *Genes Dev*, 18(11), pp. 1263-71.

Schuebel, K.E., Chen, W., Cope, L., Glockner, S.C., Suzuki, H., Yi, J.M., Chan, T.A., Van Neste, L., Van Criekinge, W., van den Bosch, S., van Engeland, M., Ting, A.H., Jair, K., Yu, W., Toyota, M., Imai, K., Ahuja, N., Herman, J.G. and Baylin, S.B. (2007) 'Comparing the DNA hypermethylome with gene mutations in human colorectal cancer', *PLoS Genet*, 3(9), pp. 1709-23.

Scott, J.M. (1999) 'Folate and vitamin B12', *Proc Nutr Soc*, 58(2), pp. 441-8.

Searle, S.D., Mitnitski, A., Gahbauer, E.A., Gill, T.M. and Rockwood, K. (2008) 'A standard procedure for creating a frailty index', *BMC Geriatr*, 8, p. 24.

Sharabi, A.B., Aldrich, M., Sosic, D., Olson, E.N., Friedman, A.D., Lee, S.H. and Chen, S.Y. (2008) 'Twist-2 controls myeloid lineage development and function', *PLoS Biol*, 6(12), p. e316.

Sharma, P., Kumar, J., Garg, G., Kumar, A., Patowary, A., Karthikeyan, G., Ramakrishnan, L., Brahmachari, V. and Sengupta, S. (2008) 'Detection of altered global DNA methylation in coronary artery disease patients', *DNA Cell Biol*, 27(7), pp. 357-65.

Sharma, S., Kelly, T.K. and Jones, P.A. (2010) 'Epigenetics in cancer', *Carcinogenesis*, 31(1), pp. 27-36.

Shen, J.C., Rideout, W.M., 3rd and Jones, P.A. (1994) 'The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA', *Nucleic Acids Res*, 22(6), pp. 972-6.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A. and Issa, J.P. (2007) 'Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters', *PLoS Genet*, 3(10), pp. 2023-36.

Shen, L.L. and Waterland, R.A. (2007) 'Methods of DNA methylation analysis', *Curr Opin Clin Nutr Metab Care*, 10(5), pp. 576-581.

Shimamoto, T., Ohyashiki, J.H. and Ohyashiki, K. (2005) 'Methylation of p15(INK4b) and E-cadherin genes is independently correlated with poor prognosis in acute myeloid leukemia', *Leuk Res*, 29(6), pp. 653-9.

Shoemaker, R., Deng, J., Wang, W. and Zhang, K. (2010) 'Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome', *Genome Res*, 20(7), pp. 883-889.

Siegmund, K.D., Connor, C.M., Campan, M., Long, T.I., Weisenberger, D.J., Biniszkiewicz, D., Jaenisch, R., Laird, P.W. and Akbarian, S. (2007) 'DNA Methylation in the Human Cerebral Cortex Is Dynamically Regulated throughout the Life Span and Involves Differentiated Neurons', *PLoS One*, 2(9).

Simpkins, S.B., Bocker, T., Swisher, E.M., Mutch, D.G., Gersell, D.J., Kovatich, A.J., Palazzo, J.P., Fishel, R. and Goodfellow, P.J. (1999) 'MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers', *Hum Mol Genet*, 8(4), pp. 661-6.

Sinclair, K.D., Allegrucci, C., Singh, R., Gardner, D.S., Sebastian, S., Bispham, J., Thurston, A., Huntley, J.F., Rees, W.D., Maloney, C.A., Lea, R.G., Craigon, J., McEvoy, T.G. and Young, L.E. (2007) 'DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status', *Proc Natl Acad Sci U S A*, 104(49), pp. 19351-6.

Skibola, C.F., Smith, M.T., Kane, E., Roman, E., Rollinson, S., Cartwright, R.A. and Morgan, G. (1999) 'Polymorphisms in the methylenetetrahydrofolate reductase gene are associated with susceptibility to acute leukemia in adults', *Proc Natl Acad Sci U S A*, 96(22), pp. 12810-5.

Smiraglia, D.J., Kulawiec, M., Bistulfi, G.L., Gupta, S.G. and Singh, K.K. (2008) 'A novel role for mitochondria in regulating epigenetic modification in the nucleus', *Cancer Biol Ther*, 7(8), pp. 1182-90.

Smiraglia, D.J., Rush, L.J., Fruhwald, M.C., Dai, Z., Held, W.A., Costello, J.F., Lang, J.C., Eng, C., Li, B., Wright, F.A., Caligiuri, M.A. and Plass, C. (2001) 'Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies', *Hum Mol Genet*, 10(13), pp. 1413-9.

Song, J., Noh, J.H., Lee, J.H., Eun, J.W., Ahn, Y.M., Kim, S.Y., Lee, S.H., Park, W.S., Yoo, N.J., Lee, J.Y. and Nam, S.W. (2005) 'Increased expression of historic deacetylase 2 is found in human gastric cancer', *APMIS*, 113(4), pp. 264-268.

Stein, R., Gruenbaum, Y., Pollack, Y., Razin, A. and Cedar, H. (1982) 'Clonal inheritance of the pattern of DNA methylation in mouse cells', *Proc Natl Acad Sci U S A*, 79(1), pp. 61-5.

Stenvinkel, P., Karimi, M., Johansson, S., Axelsson, J., Suliman, M., Lindholm, B., Heimburger, O., Barany, P., Alvestrand, A., Nordfors, L., Qureshi, A.R., Ekstrom, T.J. and Schalling, M. (2007) 'Impact of inflammation on epigenetic DNA methylation - a novel risk factor for cardiovascular disease?', *J Intern Med*, 261(5), pp. 488-99.

Strathdee, G. and Brown, R. (2002) 'Aberrant DNA methylation in cancer: potential clinical interventions', *Expert Rev Mol Med*, 4(4), pp. 1-17.

Strathdee, G., Davies, B.R., Vass, J.K., Siddiqui, N. and Brown, R. (2004a) 'Cell type-specific methylation of an intronic CpG island controls expression of the MCJ gene', *Carcinogenesis*, 25(5), pp. 693-701.

Strathdee, G., Holyoake, T.L., Sim, A., Parker, A., Oscier, D.G., Melo, J.V., Meyer, S., Eden, T., Dickinson, A.M., Mountford, J.C., Jorgensen, H.G., Soutar, R. and Brown, R. (2007a) 'Inactivation of HOXA genes by hypermethylation in myeloid and lymphoid malignancy is frequent and associated with poor prognosis', *Clin Cancer Res*, 13(17), pp. 5048-55.

Strathdee, G., Sim, A. and Brown, R. (2004b) 'Control of gene expression by CpG island methylation in normal cells', *Biochem Soc Trans*, 32(Pt 6), pp. 913-5.

Strathdee, G., Sim, A., Parker, A., Oscier, D. and Brown, R. (2006) 'Promoter hypermethylation silences expression of the HoxA4 gene and correlates with IgVh mutational status in CLL', *Leukemia*, 20(7), pp. 1326-9.

Strathdee, G., Sim, A., Soutar, R., Holyoake, T.L. and Brown, R. (2007b) 'HOXA5 is targeted by cell-type-specific CpG island methylation in normal cells and during the development of acute myeloid leukaemia', *Carcinogenesis*, 28(2), pp. 299-309.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. and Tajima, S. (2004) 'DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction', *J Biol Chem*, 279(26), pp. 27816-23.

Surface, L.E., Thornton, S.R. and Boyer, L.A. (2010) 'Polycomb group proteins set the stage for early lineage commitment', *Cell Stem Cell*, 7(3), pp. 288-98.

Swisher, E.M., Shawler, D.L., Collins, H.A., Bustria, A., Hart, S., Bloomfield, C., Miller, R.A. and Royston, I. (1991) 'Expression of shared idiotypes in chronic lymphocytic leukemia and small lymphocytic lymphoma', *Blood*, 77(9), pp. 1977-82.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. and Rao, A. (2009) 'Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1', *Science*, 324(5929), pp. 930-5.

Takahashi, T., Shivapurkar, N., Reddy, J., Shigematsu, H., Miyajima, K., Suzuki, M., Toyooka, S., Zochbauer-Muller, S., Drach, J., Parikh, G., Zheng, Y., Feng, Z., Kroft, S.H., Timmons, C., McKenna, R.W. and Gazdar, A.F. (2004) 'DNA methylation profiles of lymphoid and hematopoietic malignancies', *Clin Cancer Res*, 10(9), pp. 2928-35.

Takai, D. and Jones, P.A. (2002) 'Comprehensive analysis of CpG islands in human chromosomes 21 and 22', *Proc Natl Acad Sci U S A*, 99(6), pp. 3740-5.

Takasaki, S. (2009) 'Mitochondrial haplogroups associated with Japanese centenarians, Alzheimer's patients, Parkinson's patients, type 2 diabetic patients and healthy non-obese young males', *Journal of Genetics and Genomics*, 36(7), pp. 425-34.

Takasugi, M. (2011) 'Progressive age-dependent DNA methylation changes start before adulthood in mouse tissues', *Mech Ageing Dev*, 132(1-2), pp. 65-71.

Talens, R.P., Boomsma, D.I., Tobi, E.W., Kremer, D., Jukema, J.W., Willemsen, G., Putter, H., Slagboom, P.E. and Heijmans, B.T. (2010) 'Variation, patterns, and temporal stability of DNA methylation: considerations for epigenetic epidemiology', *Faseb J*, 24(9), pp. 3135-44.

Talens, R.P., Christensen, K., Putter, H., Willemsen, G., Christiansen, L., Kremer, D., Suchiman, H.E., Slagboom, P.E., Boomsma, D.I. and Heijmans, B.T. (2012) 'Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs', *Aging Cell*.

Teschendorff, A.E., Jones, A., Fiegl, H., Sargent, A., Zhuang, J.J., Kitchener, H.C. and Widschwendter, M. (2012) 'Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation', *Genome Med*, 4(3), p. 24.

Teschendorff, A.E., Menon, U., Gentry-Maharaj, A., Ramus, S.J., Weisenberger, D.J., Shen, H., Campan, M., Noushmehr, H., Bell, C.G., Maxwell, A.P., Savage, D.A., Mueller-Holzner, E., Marth, C., Kocjan, G., Gayther, S.A., Jones, A., Beck, S., Wagner,

W., Laird, P.W., Jacobs, I.J. and Widschwendter, M. (2010) 'Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer', *Genome Res*, 20(4), pp. 440-6.

Thathia, S.H., Ferguson, S., Gautrey, H.E., van Otterdijk, S.D., Hili, M., Rand, V., Moorman, A.V., Meyer, S., Brown, R. and Strathdee, G. (2012) 'Epigenetic inactivation of TWIST2 in acute lymphoblastic leukemia modulates proliferation, cell survival and chemosensitivity', *Haematologica*, 97(3), pp. 371-8.

Tong, W.G., Wierda, W.G., Lin, E., Kuang, S.Q., Bekele, B.N., Estrov, Z., Wei, Y., Yang, H., Keating, M.J. and Garcia-Manero, G. (2010) 'Genome-wide DNA methylation profiling of chronic lymphocytic leukemia allows identification of epigenetically repressed molecular pathways with clinical impact', *Epigenetics*, 5(6), pp. 499-508.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B. and Issa, J.P. (1999) 'CpG island methylator phenotype in colorectal cancer', *Proc Natl Acad Sci U S A*, 96(15), pp. 8681-6.

Tra, J., Kondo, T., Lu, Q., Kuick, R., Hanash, S. and Richardson, B. (2002) 'Infrequent occurrence of age-dependent changes in CpG island methylation as detected by restriction landmark genome scanning', *Mech Ageing Dev*, 123(11), pp. 1487-503.

Tyler, J.K. and Kadonaga, J.T. (1999) 'The "dark side" of chromatin remodeling: repressive effects on transcription', *Cell*, 99(5), pp. 443-6.

Ushijima, T., Watanabe, N., Okochi, E., Kaneda, A., Sugimura, T. and Miyamoto, K. (2003) 'Fidelity of the methylation pattern and its variation in the genome', *Genome Res*, 13(5), pp. 868-74.

van Scherpenzeel Thim, V., Remacle, S., Picard, J., Cornu, G., Gofflot, F., Rezsohazy, R. and Verellen-Dumoulin, C. (2005) 'Mutation analysis of the HOX paralogous 4-13 genes in children with acute lymphoid malignancies: identification of a novel germline mutation of HOXD4 leading to a partial loss-of-function', *Hum Mutat*, 25(4), pp. 384-95.

van Vlodrop, I.J., Niessen, H.E., Derks, S., Baldewijns, M.M., van Criekinge, W., Herman, J.G. and van Engeland, M. (2011) 'Analysis of promoter CpG island hypermethylation in cancer: location, location, location!', *Clin Cancer Res*, 17(13), pp. 4225-31.

Vanyushin, B.F., Nemirovsky, L.E., Klimenko, V.V., Vasiliev, V.K. and Belozersky, A.N. (1973) 'The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents', *Gerontologia*, 19(3), pp. 138-52.

Varley, K.E., Mutch, D.G., Edmonston, T.B., Goodfellow, P.J. and Mitra, R.D. (2009) 'Intra-tumor heterogeneity of MLH1 promoter methylation revealed by deep single molecule bisulfite sequencing', *Nucleic Acids Res*, 37(14), pp. 4603-12.

Vilkaitis, G., Suetake, I., Klimasauskas, S. and Tajima, S. (2005) 'Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase', *J Biol Chem*, 280(1), pp. 64-72.

Wagner, J.M., Hackanson, B., Lubbert, M. and Jung, M. (2010) 'Histone deacetylase (HDAC) inhibitors in recent clinical trials for cancer therapy', *Clin Epigenetics*, 1(3-4), pp. 117-136.

Wan, E.S., Qiu, W.L., Baccarelli, A., Carey, V.J., Bacherman, H., Rennard, S.I., Agusti, A., Anderson, W., Lomas, D.A. and Demeo, D.L. (2012) 'Cigarette smoking behaviors and time since quitting are associated with differential DNA methylation across the human genome', *Hum Mol Genet*, 21(13), pp. 3073-3082.

Wang, S.C., Oelze, B. and Schumacher, A. (2008) 'Age-specific epigenetic drift in late-onset Alzheimer's disease', *PLoS One*, 3(7), p. e2698.

Wang, Y. and Leung, F.C. (2004) 'An evaluation of new criteria for CpG islands in the human genome as gene markers', *Bioinformatics*, 20(7), pp. 1170-7.

Weaver, I.C., Diorio, J., Seckl, J.R., Szyf, M. and Meaney, M.J. (2004) 'Early environmental regulation of hippocampal glucocorticoid receptor gene expression: characterization of intracellular mediators and potential genomic target sites', *Ann N Y Acad Sci*, 1024, pp. 182-212.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M. and Schubeler, D. (2007) 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nat Genet*, 39(4), pp. 457-66.

Weisenberger, D.J., Campan, M., Long, T.I., Kim, M., Woods, C., Fiala, E., Ehrlich, M. and Laird, P.W. (2005) 'Analysis of repetitive element DNA methylation by MethyLight', *Nucleic Acids Res*, 33(21), pp. 6823-36.

Weiss, C.O. (2011) 'Frailty and chronic diseases in older adults', *Clin Geriatr Med*, 27(1), pp. 39-52.

Wessa, P. (2012) *Free Statistics Software, Office for Research Development and Education, version 1.1.23-r7*.

Whitelaw, N.C. and Whitelaw, E. (2006) 'How lifetimes shape epigenotype within and across generations', *Hum Mol Genet*, 15 Spec No 2, pp. R131-7.

Wiemels, J.L., Smith, R.N., Taylor, G.M., Eden, O.B., Alexander, F.E. and Greaves, M.F. (2001) 'Methylenetetrahydrofolate reductase (MTHFR) polymorphisms and risk of molecularly defined subtypes of childhood acute leukemia', *Proc Natl Acad Sci U S A*, 98(7), pp. 4004-9.

Wigler, M., Levy, D. and Perucho, M. (1981) 'The somatic replication of DNA methylation', *Cell*, 24(1), pp. 33-40.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J. and Helin, K. (2011) 'TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity', *Nature*, 473(7347), pp. 343-8.

Wilson, V.L., Smith, R.A., Ma, S. and Cutler, R.G. (1987) 'Genomic 5-methyldeoxycytidine decreases with age', *J Biol Chem*, 262(21), pp. 9948-51.

Wolff, G.L., Kodell, R.L., Moore, S.R. and Cooney, C.A. (1998) 'Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice', *Faseb J*, 12(11), pp. 949-57.

Wong, I.H., Ng, M.H., Huang, D.P. and Lee, J.C. (2000) 'Aberrant p15 promoter methylation in adult and childhood acute leukemias of nearly all morphologic subtypes: potential prognostic implications', *Blood*, 95(6), pp. 1942-9.

Wu, H., Coskun, V., Tao, J.F., Xie, W., Ge, W.H., Yoshikawa, K., Li, E., Zhang, Y. and Sun, Y.E. (2010) 'Dnmt3a-Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes', *Science*, 329(5990), pp. 444-448.

Wu, H.K., Weksberg, R., Minden, M.D. and Squire, J.A. (1997) 'Loss of imprinting of human insulin-like growth factor II gene, IGF2, in acute myeloid leukemia', *Biochem Biophys Res Commun*, 231(2), pp. 466-72.

Wutz, A. and Gribnau, J. (2007) 'X inactivation Xplained', *Curr Opin Genet Dev*, 17(5), pp. 387-93.

Xiao, Y., Word, B., Starlard-Davenport, A., Haefele, A., Lyn-Cook, B.D. and Hammons, G. (2008) 'Age and gender affect DNMT3a and DNMT3b expression in human liver', *Cell Biol Toxicol*, 24(3), pp. 265-72.

Xiong, Z. and Laird, P.W. (1997) 'COBRA: a sensitive and quantitative DNA methylation assay', *Nucleic Acids Res*, 25(12), pp. 2532-4.

Yang, X., Ferguson, A.T., Nass, S.J., Phillips, D.L., Butash, K.A., Wang, S.M., Herman, J.G. and Davidson, N.E. (2000) 'Transcriptional activation of estrogen receptor alpha in human breast cancer cells by histone deacetylase inhibition', *Cancer Res*, 60(24), pp. 6890-4.

You, J.S. and Jones, P.A. (2012) 'Cancer genetics and epigenetics: two sides of the same coin?', *Cancer Cell*, 22(1), pp. 9-20.

Zaina, S., Lindholm, M.W. and Lund, G. (2005) 'Nutrition and aberrant DNA methylation patterns in atherosclerosis: More than just hyperhomocysteinemia?', *Journal of Nutrition*, 135(1), pp. 5-8.

Zemlyakova, V.V., Zhevlova, A.I., Zborovskaya, I.B., Strelnikov, V.V., Laktionov, K.K., Zaletaev, D.V. and Nemtsova, M.V. (2003) 'Methylation profile of several tumor suppressor genes in non-small-cell lung cancer', *Molecular Biology*, 37(6), pp. 836-840.

Zhang, F.F., Cardarelli, R., Carroll, J., Fulda, K.G., Kaur, M., Gonzalez, K., Vishwanatha, J.K., Santella, R.M. and Morabia, A. (2011) 'Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood', *Epigenetics*, 6(5), pp. 623-9.

Zhu, Z.Z., Hou, L., Bollati, V., Tarantini, L., Marinelli, B., Cantone, L., Yang, A.S., Vokonas, P., Lissowska, J., Fustinoni, S., Pesatori, A.C., Bonzini, M., Apostoli, P., Costa, G., Bertazzi, P.A., Chow, W.H., Schwartz, J. and Baccarelli, A. (2012) 'Predictors of global methylation levels in blood DNA of healthy subjects: a combined analysis', *Int J Epidemiol*, 41(1), pp. 126-39.

Zhu, Z.Z., Sparrow, D., Hou, L., Tarantini, L., Bollati, V., Litonjua, A.A., Zanobetti, A., Vokonas, P., Wright, R.O., Baccarelli, A. and Schwartz, J. (2011) 'Repetitive element hypomethylation in blood leukocyte DNA and cancer incidence, prevalence, and mortality in elderly individuals: the Normative Aging Study', *Cancer Causes Control*, 22(3), pp. 437-47.

Ziller, M.J., Muller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., Gnirke, A. and Meissner, A. (2011) 'Genomic distribution and inter-sample variation of non-CpG methylation across human cell types', *PLoS Genet*, 7(12), p. e1002389.

Zochbauer-Muller, S., Lam, S., Toyooka, S., Virmani, A.K., Toyooka, K.O., Seidl, S., Minna, J.D. and Gazdar, A.F. (2003) 'Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers', *Int J Cancer*, 107(4), pp. 612-6.

Zou, B., Chim, C.S., Zeng, H., Leung, S.Y., Yang, Y., Tu, S.P., Lin, M.C., Wang, J., He, H., Jiang, S.H., Sun, Y.W., Yu, L.F., Yuen, S.T., Kung, H.F. and Wong, B.C. (2006) 'Correlation between the single-site CpG methylation and expression silencing of the XAF1 gene in human gastric and colon cancers', *Gastroenterology*, 131(6), pp. 1835-43.

Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., Grausenburger, R., Schweifer, N., Chiocca, S., Decker, T. and Seiser, C. (2006) 'Negative and positive regulation of gene expression by mouse histone deacetylase 1', *Mol Cell Biol*, 26(21), pp. 7913-28.