EPIGENETIC EFFECTS ON GENE REGULATION DURING CELL DIFFERENTIATION, DEVELOPMENT AND DISEASE

LIM YEN CHING (A0021761R)

(MSc, National University of Singapore)

A THESIS SUBMITTED FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

SCHOOL OF MEDICINE

DEPARTMENT OF BIOCHEMISTRY

NATIONAL UNIVERSITY OF SINGAPORE

2015

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

_____

Lim Yen Ching

2015

ACKNOWLEDGEMENTS

The completion of this thesis will form one of the most satisfying and rewarding milestones in my life. The academically challenging task will never be accomplished without a group of inspiring, supportive and encouraging people, whom I owe my greatest gratitude towards.

I would like to thank my supervisor, ***Dr. Xu Feng***, whose expertise, vast scientific knowledge and generosity in offering wise opinions, has enabled me to progress well in the pursuit of my PhD. I cannot express my gratitude enough for all the help that he has given me.

I would also like to thank express my greatest gratitude to my co-supervisor, ***Dr. Ding Chunming***. His passion for scientific research has always been a huge source of inspiration and motivation to push myself harder and to work more independently. Despite his busy schedule, he is never impatient to spend time and to discuss about various novel bioinformatics ideas to ensure that I can develop both creatively and intellectually as a bioinformatician and researcher.

I must also express my deepest appreciation to my co-supervisor, ***Dr. Sun Lei***, who is always approachable for academic discussion and has never shown any

signs of impatience regardless how elementary my questions sounded. I have benefitted tremendously under his guidance.

Next, I would like to thank members in my Thesis Advisory Committee, *Professor Fu Xin-Yuan* and *Professor Han Weiping*. Despite their hectic schedules, both Professor Fu and Professor Han always ensure they make time to attend my TAC meetings. All the useful comments given during the TAC provided great biological insights and is a major contributor to shaping my thesis. For that, I cannot express my appreciation enough.

A very special thanks goes out to *Dr. Jin Shengnan*. An encouraging, smart yet humble scientist whom I can always turn to for help and advices.

In addition, I would like to thank my past and present colleagues, especially, *Dr. Jaydon Lee Yew Kok*, *Dr. Zheng Zejun, Dr. Xu Dan* and *Miss Chia Sook Yoong* for all the guidance and support.

Last but not least, to my family, especially my parents. Thank you for always being there, good and bad times.

To all others whom I have missed out, my sincere thanks.

To all I have mentioned, my heartfelt thanks once again. Without all of you, I would not have gone this far.

TABLE OF CONTENTS

SUMMARY

Epigenetic mechanisms such as DNA methylation play profound and complex roles in affecting gene expression and contribute to phenotypic plasticity. DNA methylation has been shown to be implicated in a diverse array of functions such as X-chromosome inactivation, development, cell differentiation, cell type specificity and alternative splicing. High throughput sequencing technologies in combination with bioinformatics data mining methods allow for systematic genome-wide analysis between different -omics data types to elucidate complex interactive relationships.

To quantify genome-wide DNA methylation while ensuring cost-effectiveness, we made improvements to existing reduced representation bisulfite sequencing protocol and developed an automated software to calculate CpG methylation from high throughput sequencing data. We further analysed results obtained from this pipeline to understand the impact of DNA methylation in a diversity of biological systems, namely (i) mouse white and brown adipocytes differentiation, (ii) human placenta tissues at early and late gestational age and (iii) human placenta tissues from pregnancies carrying Down's syndrome (DS) and normal foetuses.

Methylome profiling of both white and brown adipogenesis processes in mouse revealed a general hypermethylation during cell differentiation, with strongest predominance of hypermethylation in promoter regions. When comparing methylome between white and brown adipocytes at each point of differentiation, there was a consistent hypermethylation in white adipocytes. Contrastingly to cell differentiation, predominance of hypermethylation was strongest in non-promoter regions. In addition, I identified a number of *Hox* family genes with consistent promoter

methylation difference between white and brown adipocytes where DNA methylation anti-correlates with gene expression throughout adipogenesis.

Next, through genome-wide correlation study between DNA methylation and gene expression in early and late trimester human placenta samples, I demonstrated that associations between DNA methylation and gene expression was highly complex and genomic context dependent.

Lastly, we examined differences in placenta villi methylome between DS and normal samples and observed a genome-wide hypermethylation associated with DS phenotype. Genes with promoter hypermethylation were associated with functions related to DS phenotypes. DNA hypermethylation may be partially attributable to down regulation of *TET* family genes and *REST/NRSF*.

Taken together, I have identified the existence of unique DNA methylation footprints in multiple biological processes such as cell differentiation, cell specificity and development. In addition, aberrant DNA methylation was also observed in conditions of genetic disorder. Further research can be focused on biological manipulations with either epigenetic status or gene expression to test hypothesis derived from the above omics studies.

LIST OF TABLES

LIST OF FIGURES

6

LIST OF SYMBOLS

5CaC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-Hydroxymethylcytosine
5hmU	5-hydroxylmethy-uracil
5mC	5-methylcytosine
AID	Activation-induced cytosine deaminase
APOBEC1	Apolioprotein B mRNA editing enzyme, catalytic polypeptide 1
BA	Brown adipocytes
BAT	Brown adipocyte tissue
BER	Base excision repair
C2T	Cytosine to thymine
CGI	CpG Islands
CGS	CpG Shores
Chr21	Chromosome 21
DMC	Differentially methylated CpG
DMP	Differentially methylated promoter
DNMT	DNA methyltransferase
DS	Down syndrome
ESC	Embryonic stem cell
G2A	Guanine to adenine
GWAS	Genome wide association studies
HTS	High throughput sequencing
IAP	Intracisternal A particle
ICR	Imprinting control region
LINE	Long Interspersed Nuclear Elements
MeDIP	Methylated DNA immunoprecipitation
MRE	Methylation sensitive restriction digestion
PCA	Principle Component Analysis
PGCs	Primordial germ cells
RE	Restriction Enzyme
RRBS	Reduced Representation Bisulfite Sequencing
SAM	S-adenosylmethionine
SINE	Short Interspersed Nuclear Elements
SMRT	Single-molecule-real-time
SNPs	Single nucleotide polymorphisms
T21	Trisomy 21
TDG	Thymine DNA glycosylase
TET	Ten-eleven translocation methylcytosine dioxygenase
WA	White adipocytes
WAT	White adipocyte tissue
WGBS	Whole genome bisulfite sequencing

PUBLICATIONS

The main body of the thesis is based on the following publications and manuscripts (in preparation):

1. Lee, Y. K., S. Jin, S. Duan, *Y. C. Lim*, D. P. Ng, X. M. Lin, G. Yeo and C. Ding (2014). Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples. Biol Proced Online **16**(1): 1.

2. Jin, S., Y. K. Lee, *Y. C. Lim*, Z. Zheng, X. M. Lin, D. P. Ng, J. D. Holbrook, H. Y. Law, K. Y. Kwek, G. S. Yeo and C. Ding (2013). Global DNA hypermethylation in down syndrome placenta. PLoS Genet **9**(6): e1003515.

3. A Complex Association between DNA Methylation and Gene Expression in Human Placenta at First and Third Trimesters (manuscript in preparation)

4. Epigenome-wide DNA methylation landscapes reveals systematic differences during adipogenesis and define cell type specificity (manuscript in preparation)

The following publications and manuscripts provided important compuational and biological education for relavant background knowledge in the construction and building of the thesis.

1. Pan, H., L. Chen, S. Dogra, A. L. Teh, J. H. Tan, Y. I. Lim, *Y. C. Lim*, S. Jin, Y. K. Lee, P. Y. Ng, M. L. Ong, S. Barton, Y. S. Chong, M. J. Meaney, P. D. Gluckman, W. Stunkel, C. Ding and J. D. Holbrook (2012). Measuring the methylome in clinical samples: improved processing of the Infinium Human Methylation450 BeadChip Array. Epigenetics **7**(10): 1173-1187.

2. Das, R., Y. K. Lee, R. Strogantsev, S. Jin, *Y. C. Lim*, P. Y. Ng, X. M. Lin, K. Chng, G. Yeo, A. C. Ferguson-Smith and C. Ding (2013). DNMT1 and AIM1 Imprinting in human placenta revealed through a genome-wide screen for allele-specific DNA methylation. BMC Genomics **14**: 685.

3. Zhang, R. R., Q. Y. Cui, K. Murai, *Y. C. Lim*, Z. D. Smith, S. Jin, P. Ye, L. Rosa, Y. K. Lee, H. P. Wu, W. Liu, Z. M. Xu, L. Yang, Y. Q. Ding, F. Tang, A. Meissner, C. Ding, Y. Shi and G. L. Xu (2013). Tet1 regulates adult hippocampal neurogenesis and cognition. Cell Stem Cell **13**(2): 237-245.

4. Jin, S., *Y. C. Lim*, D. P. Ng, H. Y. Law, K. Y. Kwek, G. S. Yeo and C. Ding (2014). Accurate fetal chromosome dosage determination by shotgun sequencing of maternal plasma DNA without PCR amplification during library preparation. Clin Chem **60**(4): 690-692.

5. Alvarez-Dominguez, J. R., Z. Bai, D. Xu, B. Yuan, K. A. Lo, M. J. Yoon, *Y. C. Lim*, M. Knoll, N. Slavov, S. Chen, P. Chen, H. F. Lodish and L. Sun. De Novo Reconstruction of Adipose Tissue Transcriptomes Reveals Long Non-Coding Rna Regulators of Brown Adipocyte Development. Cell Metab 21, no. 5 (2015): 764-76.

CHAPTER 1 - INTRODUCTION

1.1.   Epigenetics – The missing link between genetics and phenotype

Genetics is defined as the scientific study of understanding how heritable traits were passed down from parents to their offspring. Such studies often include identifying genetic mutations causing and polymorphisms associated with human diseases in either families or cohorts (e.g. genome-wide association studies (GWAS)).

Through GWAS, thousands of genetic variants associated with over 80 diseases and traits have been identified (http://www.gwascentral.org) [1, 2]. Despite so, most of the SNPs identified by application of rigorous statistics on large sample sizes have so far only achieved small effect sizes (OR <1.5), suggesting that individual SNPs have limited heritable accountability for phenotypic variation [1-5].

Moreover, despite almost all cells in an organism contain the same genetic sequence, each cell type expresses a unique set of gene signature [6]. This non-Mendelian inheritance resulting from the presence or absence of epigenetic marks, mediates a systematic control to qualitative and quantitative gene expression changes that are essential for cell lineage definition, cell differentiation and normal development [7, 8].

"Epigenetics" was first introduced in 1942 by Conrad Waddington, who first defined it as the study of casual interaction between genes and their productions which allowed for phenotypic expression [9, 10]. Recently, the term has evolved to become more specifically defined as the study of mechanisms which can regulate gene

expression not caused by a change in DNA sequence [6, 9-11]. To date, main members of the epigenetics family include DNA methylation, histone modifications and non-coding RNA (ncRNA).

Although most analyses have typically focused on a single epigenetic mechanism, increasing evidence has suggested that different mechanisms can act synergistically to activate or repress gene expression by affecting the accessibility of transcription factor binding or chromatin remodelling. Changes in environmental cues also contribute to the dynamics of the epigenetics, leading to a keen interest among researchers to associate epigenetic marks with phenotypic plasticity in mammals [6, 10, 12, 13]. Summing up, combined properties of genetics and epigenetics have conferred the genome both stability and flexibility which are crucial to the mediation of diversified gene expression in various cell and tissues types [6, 10, 12, 13].

Advancements in technologies have brought us into the age of high throughput sequencing (HTS) and any sample can theoretically be sequenced at a reasonable cost. Hence, it is now possible to interrogate whole genome signals on DNA, transcriptome and protein-DNA interactions. Ironic to this convenience, we are faced with a bottleneck of converting such a massive influx of digital data into usable biological signals. To optimize between noise removal while retaining true biological signals, we need better bioinformatics and biostatistical techniques and new biological hypotheses to perform proper data mining.

Among the known epigenetic mechanisms, there is much interest to focus on DNA methylation due to its potential roles in regulating fundamental biological processes such as organism development, X-chromosome inactivation and genetic imprinting

and association with various diseases such as cancers [14, 15]. Furthermore, as DNA methylation is a chemical modification on the DNA, the easy accessibility and stability of the DNA makes DNA methylation a robust epigenetic marker for prognostic and diagnostic purposes [14, 16].

In my thesis, I perform bioinformatics analyses on HTS data generated from biological samples to study the impact of DNA methylation in different systems, namely, (i) cell differentiation, (ii) different cell types, (iii) development and (iv) genetic disorder.

1.2.   Introduction to DNA methylation

1.2.1. Writers and erasers of DNA methylation

Among all known epigenetic mechanisms, DNA methylation makes modification directly on the DNA [17]. The covalent addition of a methyl group from the donor, S-adenosylmethionine (SAM), to the 5'-position of a cytosine to form 5-methylcytosine (5mC) [6, 18] (Figure 1), is catalysed by DNA methyltransferase (*DNMT*). Newly added methyl group projects into the major groove of the DNA, resulting in structure changes that inhibit or facilitate DNA recognition by proteins [19].

The key methyltransferses in the *DNMT* family include *DNMT1*, *DNMT3A* and *DNMT3B* and all of them have a similar architecture of a N-regulatory domain attached to a C-catalytic domain [6].
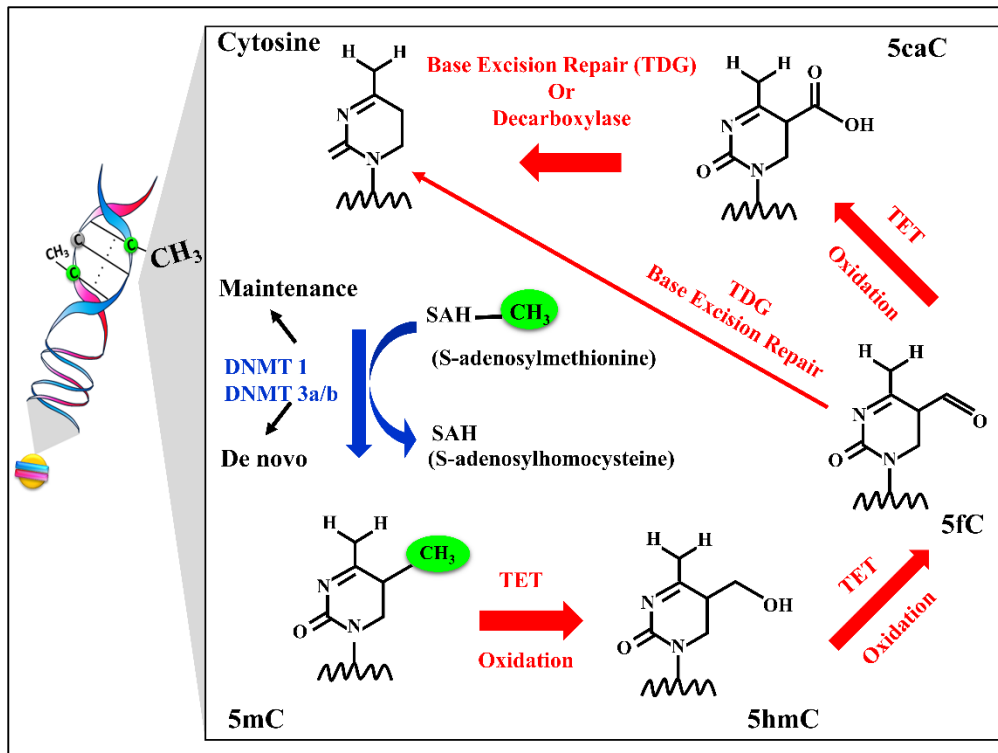
Figure 1: Biochemical reactions driving DNA methylation and demethylation.

DNMT1 is a maintenance methyltransferase which is ubiquitously expressed in proliferating cells [20]. It is recruited by *Np95/Uhrf1* [21, 22] during DNA replication and localises itself to the replication fork. This enzyme shows a 7-12 folds preference for hemimethylated DNA [23, 24] and copies the methylation marks from the parental template strand to the newly synthesized daughter strand [25, 26]. This not only ensures the symmetry of methylation marks on both Watson and Crick strands [17] but also makes sure the faithful preservation of original methylation patterns in a cell lineage during every cell division [6].

Evidence has supported the appropriate and time-specific expression of *DNMT1* is essential for normal mammalian development. Although *DNMT1* knockout mouse ES cells displayed lower 5mc levels and enhanced microsatellite instability, they remained viable [18, 27, 28] whereas mouse embryos deficient for *DNMT1* died in utero [29]. Conditional *DNMT1* mutant mice were also viable but exhibit hypomethylation in the cortical and hippocampal cells in the dorsal forebrain from

E13.5 [30]. In addition, these mice showed learning and memory defects later in life, implying that DNA methylation is crucial to the proper regulation of neuronal maturation in the central nervous system [30]. On the other hand, overexpressing *DNMT1* resulted in genomic hypermethylation, in turn caused a loss of imprinting marks [31] and finally led to a range of pregnancy complications or even embryo death [32, 33].

Contrary to *DNMT1*, *DNMT3A* demonstrates a threefold higher activity on unmethylated than hemimethylated DNA [23, 24]. *DNMT3A* and *DNMT3B* are known as *de novo* methyltransferase which catalyse the transfer of the methyl group onto naked DNA [6]. Both of these enzymes are believed to be important in early embryonic development by functioning as DNA methylation writers to re-establish methylation marks after embryonic implantation [9]. Mice heterozygous for *DNMT3A* or *DNMT3B* are normal and fertile, while *DNMT3B-/-* mutant mice are runted and died at about 4 weeks after birth [29].

Although DNA methylation marks are stable and heritable, mounting evidence generated in recent years has shown that these marks are not as static as once perceived [8]. In fact, the gain or loss of DNA methylation has been seen to occur either on a genome-wide scale according to developmental demands or on a gene-specific scale according to specific somatic cell signals [8].

Unlike DNA methylation, mechanism explaining the direct cleavage of the strong C-C bond binding methyl group from cytosine to account for DNA demethylation remains elusive and has yet to be discovered in mammals [6]. To date, evidence of DNA demethylation has been shown to occur either actively or passively. Passive demethylation occurs via an inhibition of *DNMT* resulting in loss of methylation marks over successive cell divisions [9, 10]. In contrast, several mechanisms have

been proposed for active demethylation, which generally involves a series of chemical modifications from 5mC to various intermediate products. These products eventually enter the base excision repair (BER) pathway which makes use of thymine DNA glycosylase (TDG) to cleave off the "inappropriate" residue and replaces it with an "appropriate" naked cytosine [6, 34-36]. However, much debate still goes on with regards to the fitting of the correct puzzle pieces to explain the intermediate steps. Two possible pathways have been proposed:

i.   Deamination of 5mC by *AID/APOBEC1* leading to a conversion from cytosine to thymine. This potentially results in a T-G mismatch and a subsequent call for base excision repair (BER).

ii.   Oxidation of 5mC to 5hmC by *TET*. From 5-hydroxylmethylcytosine (5hmC), the molecule either undergoes (a) iterative oxidation to 5-formylcytosine (5fC) and to 5-carboxylcytosine (5CaC) and finally to 5C (Figure 1), or (b) deamination to form 5-hydroxylmethy-uracil (5hmU) which then goes into BER pathway.

1.2.2. Landscape of DNA methylation

Approximately 3% of the cytosine are methylated in the human genome [37]. In mammals, cytosine methylation mostly occurs in the context of the palindromic CpG dinucleotide located in various genomic regions such as promoters, gene bodies, intergenic and repetitive regions [38]. Main regulatory regions such as promoters, enhancers and first exon tend to be hypomethylated relative to gene bodies, intergenic or repetitive regions [39-42].

Approximately 70-80% of the CpGs are methylated in mammals [13]. These methylated CpGs are predominantly located in repetitive regions such as transposons [43]. The remaining 20% are often found in clusters of high CpG density known as CpG islands (CGI) [44], defined by the following [45]:

i.   Length $\geq$ 200bp

ii.   GC content $\geq$ 50%

iii. $0.6 < \frac{Observed}{Expected} CpG = \frac{\#CpG}{\#C \times \#G} \times N$ over a moving average of 100bp

window at 1bp step size

Where N= total number of nucleotides

About 72% of gene promoters lie within regions CpG rich region [46], which show high levels of conservation between human and mouse [6]. Additionally, genes with CpG rich promoters include most of the housekeeping genes [45].

Gene expression is largely regulated by transcription factors [47], through their binding to genomic functional elements such as transcription factor binding motifs on the genome. Transcriptionally active genes are often marked by nucleosome-depleted regions characterized by H3K4me3 at their flanking nucleosome regions [48]. For a long time, promoter DNA methylation has been regarded as an epigenetic mark for gene silencing. Although such an association has been widely accepted, the cause-and-effect relationship between DNA methylation and gene silencing remains a controversial topic of discussion [47].

Early finding (1987) on mouse embryo reported that methylation of *Hprt* gene occurred after X chromosome inactivation, suggesting that DNA methylation could have acted as a reinforcement lock to ensure the repressed state of a gene [49].

Additionally, several cancer studies have also shown that genes silenced by the polycomb complex have a higher tendency to be hypermethylated, further supporting the hypothesis that gene silencing precedes DNA methylation [50-53].

From a second perspective (DNA methylation precedes gene silencing), presence of the methyl groups on the promoter regions physically impedes transcription factors from binding [54] and represses gene expression. Alternatively, various hypotheses linking DNA methylation and other epigenetic modifications have also been suggested. For example, methylated CpGs bind to methyl CpG binding protein 2 (*MeCP2*) which subsequently recruits *Sin3A* followed by histone deacetylase (*HDAC*) [55]. This enzyme then removes the acetyl group on the histone causing a tighter remodelling of the chromatin, leading to reduced permissiveness for gene expression.

Regardless of the sequential ordering of events, an anti-correlation between promoter DNA methylation and gene expression is generally accepted. However, such interferences cannot be extrapolated to other genomic regions. Before the advent of HTS, genome-wide correlations were not possible and inference of methylation-expression correlation was mainly made from the context of promoters. In recent years, HTS enables a more comprehensive correlation between DNA methylation and gene expression. Analyses on gene bodies DNA methylation using different human cell lines [56, 57] and X chromosome [58] have shown that DNA methylation tend to be positively correlated with gene expression. These results [59] suggest a far more complex role of DNA methylation in regulating gene expression, which is highly context dependent.

In terms of correlation between DNA methylation and gene expression, gene body typically starts from the end of the first exon due to this exon exhibiting the same gene repressive effect as the promoter [60]. Gene bodies typically have a considerably lower CpG content and are extensively methylated [47]. Thus, it is highly uncommon to locate CGI which are heavily methylated within gene bodies. [47]. Interestingly, even if such CGIs are identified, they do not block transcriptional elongation, implying that the gene silencing effect of DNA methylation is on transcription initiation and not on elongation [47].

Intragenic methylation is also speculated to be involved in regulating alternative splicing events. Various bioinformatics analyses performed on a genome-wide scale have reported hypermethylated CpGs in the exonic regions relative to the intronic counterparts [57, 61] . Such intriguing results has thus led to a series of questions. (i) Can DNA methylation be used as markers for defining exons [62]? (ii) Is DNA methylation a regulatory guide for directing alternative splicing? Although many of the mechanistic questions remain to be answered, Shukla et al. have proposed that DNA methylation could possibly affect CTCF binding to its target exons, leading to pausing of RNA pol II and a subsequent preference for the assembly of the co-transcriptional spliceosome at the upstream splice sites [63].

The impact of DNA methylation is not limited to regulatory elements in the genome. DNA methylation plays a crucial role in regulating transcription of intergenic genomic sequences as well. The mammalian genome is made up of approximately 45% repetitive and transposon elements [64], which when expressed, affects genomic stability and causes transcriptional dysregulation and DNA mutation [65-67]. Suppression of most of these potentially harmful genomic elements such as intracisternal A particle (IAP), short interspersed nuclear elements (SINE) and long

interspersed nuclear elements (LINE) elements are achieved by DNA methylation [68, 69].

1.3. Global reprogramming of DNA methylation define crucial mammalian development events

Whilst DNA methylation had been perceived as a static mark for decades, mounting evidence in recent years has refuted this belief. Studies have revealed that these stable and heritable marks were removed either genome-wide during early development or in gene-specific context in somatic cells [8] for defining cell lineage or during differentiation.

Extensive global reprogramming of DNA methylation takes place during mammalian development to direct cells into cell fate transition and developmental potency [70] (Figure 2), with the exceptions to specific genomic regions such as imprinting control regions (ICR) and IAP [8]. Such systematic programming of DNA methylome requires proper resetting and establishment of the epigenetic marks which are achieved by ten-eleven translocation methylcytosine dioxygenase (*TET*) and *DNMT* respectively (Figure 2).

Shortly after fertilization, the sperm-derived pronucleus undergoes active demethylation (proposed to be catalysed by *TET3*) leading to rapid drop in global methylation levels [71, 72]. The sharp decline in 5mC is accompanied by an immediate increase of its corresponding first oxidation product, 5hmC. Unexpectedly, 5hmC is not removed immediately by BER, but have persisted into early embryogenesis. This suggests that demethylation in paternal pronucleus is brought by combined actions of oxidation and passive demethylation [73]. Unlike

paternal-derived pronucleus, maternal-derived pronucleus shows gentler decline in methylation as it solely undergoes passive demethylation through a few cell cycles [72]. This first wave of global demethylation is thought to allow cells to return to the totipotency state [72]. After embryo implantation, *de novo* methylation sets in to establish a new pattern eventually leading to more than 70% of the CpGs (mainly non-CGI CpGs) being methylated [72, 74].

Following blastocyst implantation, embryoblast-derived epiblast gives rise to three germ layers (ectoderm, endoderm and mesoderm), which contain the precursors for all embryonic lineage [75]. In particular, the primordial germ cells (PGCs) undergoes a round of epigenetic reprogramming around E11.5 to E12.5 [76, 77] during which most of the DNA methylation marks will be erased by a combination of both active and passive demethylation [78-81]. Although exact mechanisms for global DNA methylation erasure remains elusive, recent work have reported the expressions of *TET1* and *TET2* in PGCs which peaked between E10.5 and E11.5 [79], suggesting that actions of these two enzymes could contribute to the phenomenon. Previous studies have shown that although mice deficient for *TET1* or *TET2* are viable and fertile, *TET1* deficient mice display reduced body mass [82] while *TET2* deficient mice show affected hematopietic stem cells [83, 84]. To test for redundancy functions of *TET1* and *TET2*, Dawlaty et al. generated DKO mice which was found to have higher methylation in 94 imprinted loci [85], implying deficiency in *TET1* and *TET2* affects the establishment of imprinting marks in gametes [85].
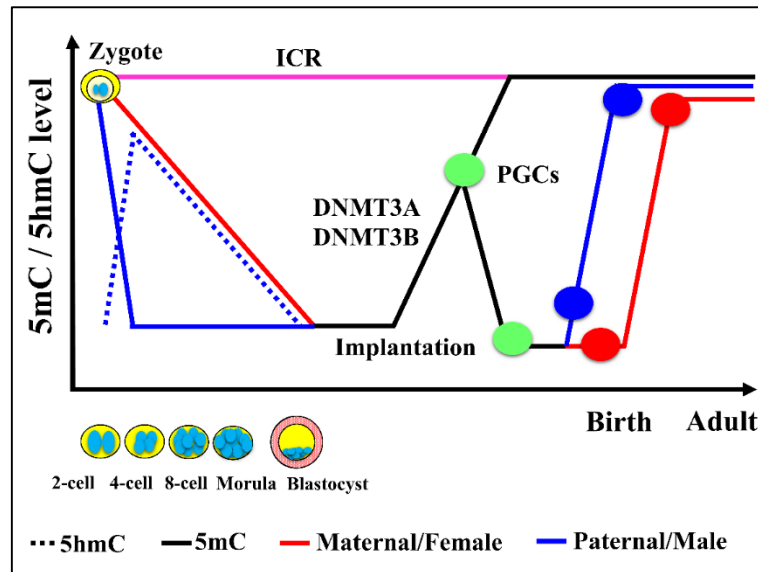
Figure 2: Tremendous methylation reprogramming takes place throughout mammalian development. This is based on the mouse model.

1.4. DNA methylation – Potentials for improved diagnostics and therapy

Disease may occur as a result of gene transcriptional dysregulation. Progression in technologies allow probing of gene expression for any sample at a relative low cost. Therefore, much of the bioinformatics work have been directed towards the selection of potential key genes that mark unique differences between disease and normal samples. Identification of these genes not only offers hope for new therapeutic targets, but also serves as biomarkers for early detection of diseases.

However, the search for biomarkers from gene expression profiles have faced much difficulty owing to the instability of RNA and high individual sample variation [11]. DNA methylation is a chemically stable mark and thus appears to be better candidate [11]. Recent studies conducted on cancer samples have shown that using DNA

23

methylation status of key genes provide a better discriminating sensitivity and specificity than either gene or protein expression [86-88].

Secondly, perturbations of DNA methylation have been associated with a vast array of disease such as cancers and imprinting disorders such as Beckwith-Wiedemann syndrome, Prader-Willi syndrome, Angelman syndrome and Transient neonatal diabetes mellitus [43]. Genomic imprinting is an example of non-Mendelian genetic inheritance for which genes are expressed in a parent-of-origin-specific manner. These genes are controlled by regulatory elements known as imprinting control regions (ICRs), which are characterised by differential DNA methylation between paternal and maternal alleles. When imprinted epigenetics marks are disrupted by either an inappropriate gain or loss of a methyl group, the non-imprinted allele expression maybe affected thereby causing imprinting disorders [43].

Lastly, DNA methylation of specific regions of interest such as promoters can also be profiled easily using loci-targeted platforms such as EpiTYPER assays, making it a convenient tool for molecular diagnostics [11].

1.5.   Studying DNA methylation

There are approximately 29 million CpGs on the human genome which carries a huge wealth of biological information. Methods for analysing DNA methylation have progressed from a local to a global level. A schematic diagram showing the main steps in typical DNA methylation studies is given in Figure 3.
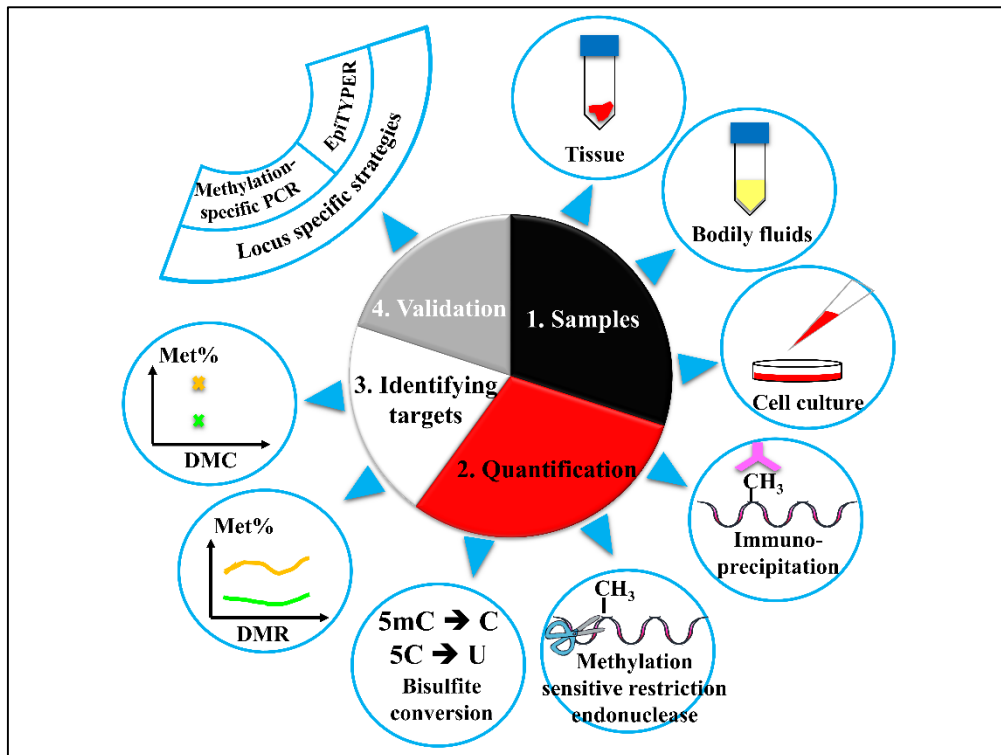
Figure 3: Schematic diagram of DNA methylation analysis. *DMR- Differentially methylated regions, DMC-differentially methylated CpGs.

To date, various methods such as immunoprecipitation, methylation sensitive restriction endonuclease and bisulfite conversion (Figure 3) have been developed to study DNA methylation, each with their unique pros and cons.

Methylated DNA immunoprecipitation (MeDIP) makes use of 5mC-specific monoclonal antibody to immunocapture DNA fragments which are enriched for 5mC. Subsequently, these purified 5mC-enriched genomic fragments can be input on a microarray (MeDIP-chip) or sequencing platform (MeDIP-seq) followed by bioinformatics analyses to identify 5mC-enriched genomic regions, relative to a control sample [11]. While this method allows for a comprehensive coverage of genomic regions, it shows preference for sites with high methylation levels [2].

Methylation sensitive restriction digestion (MRE) utilises both a methylation insensitive restriction enzyme (MspI) and a methylation sensitive restriction enzyme (HpaII) to identify methylated and unmethylated CpGs within the recognition site 5'-

CCGG-3'. For this, analysed CpGs are limited by the selection of restriction enzyme [11].

Lastly, bisulfite conversion makes use of sodium bisulfite to treat DNA which converts unmethylated cytosine to thymine while methylated cytosine (inclusive of 5hmC) shows resistance to the conversion and remain as cytosine. By comparing the ratios or relative binding of cytosine and thymine by either sequencing or array based method, DNA methylation for the cytosine can be estimated. As much as it is the preferred method, it comes with potential limitation too. Any incomplete conversion of unmethylated cytosine will be read as C in sequencing and deemed as 5mC. This results in an overestimation of true methylation levels. Bisulfite treated samples can be analysed on either on array or deep sequencing platform (Bi-seq). Besides being susceptible to batch effects, array-based platforms are inferior to sequencing-based approaches due to the pre-designed probes that limit detection of SNPs and coverage of certain CpG sites.

Before the development of the single molecule real time sequencing technique (SMRT), methylation based experiments were performed on a population of cells which are obtained from bodily fluids, tissues or cell cultures [89]. Although bodily fluids such as blood and tissues provide a more in vivo perspective than cell cultures, one should be cautious of the conclusions made from these studies. This is because such samples are made up of heterogeneous cell types for which the cell composition is dependent on the age of the organism. It is therefore highly recommended to first perform cell-sorting procedures followed by independent profiling for each separated cell type [89, 90].

Another potential problem with using cell populations for DNA methylation studies lies with connecting estimated values with actual biological scenarios. For example, a

50% methylation for a CpG could imply either allelic-specific methylation or 50% of all cells being methylated [89]. Thus, methods for detecting and quantification of DNA methylation must be carefully selected. Out of the three main methods (bisulfite conversion, methylation sensitive restriction endonuclease and immunoprecipitation) (Figure 3) commonly used to quantify DNA methylation levels [91], bisulfite conversion remains the golden standard [92] as it is the only method that can address the question described above [89].

Methods for identifying differential methylation have been broadly classified under two categories: (i) individual CpG sites or (ii) regional level, abbreviated as DMC and DMR respectively (Figure 3). Many statistical software have been developed to identify both DMCs and DMR and usage is highly dependent on sample size and experimental platforms.

In DMCs identification, fisher exact test or chi-squared test is the most straightforward method as it simply makes use of the number of methylated and unmethylated counts to generate a 2 by 2 contingency table for statistical testing. Despite its simplicity, this method fails to address biological variability [93].
For this, the beta-binomial model has been proposed recently for DMC analysis. This model not only models the methylation distribution of biological replicates, but also considers coverage uncertainty for each CpG site.

Although traditional studies have focused on DMCs, recent analysis have suggested DMR which appears to provide better predictability [94]. Furthermore, it is intuitive that a persistence of differential methylation over a region might be more robust and subjected to lower false discovery than reliance on a single CpG. In light of this, a number of DMR algorithms such as Bumphunter [95],Methylkit [93], MOABS [96], Methylsig [97] and Radmeth [98] have been developedThese algorithms typically

identify DMRs in two steps, beginning with the identification of DMCs followed by merging these sites into a genomic region by applying statistical techniques. For example, MOABS, Methylsig and Radmeth use beta-binomial model to identify DMCs followed by the application of Hidden Markov Model, tiling window and weight Z test for p-values respectively to merge DMCs into DMRs. For software such as Methylsig which uses a sliding window to detect DMR, caution should be taken to determine the appropriate size of the window. This is because the selection of too small a window might result fragmented DMRs which should have been merged together [14].

CHAPTER 2 – STUDY DESIGN

I hypothesize that novel biological insights and testable hypotheses can be derived from computational analyses of various epigenomics, genomics and transcriptomics data in cell differentiation, development and disease models.

This project aims to apply computational and statistical methods to interrogate how DNA methylation correlates with gene expression. Comprehensive genome-wide analysis will be performed in increasing biological complexity, progressively from cell differentiation to organ development and finally to disease models (Figure 4).
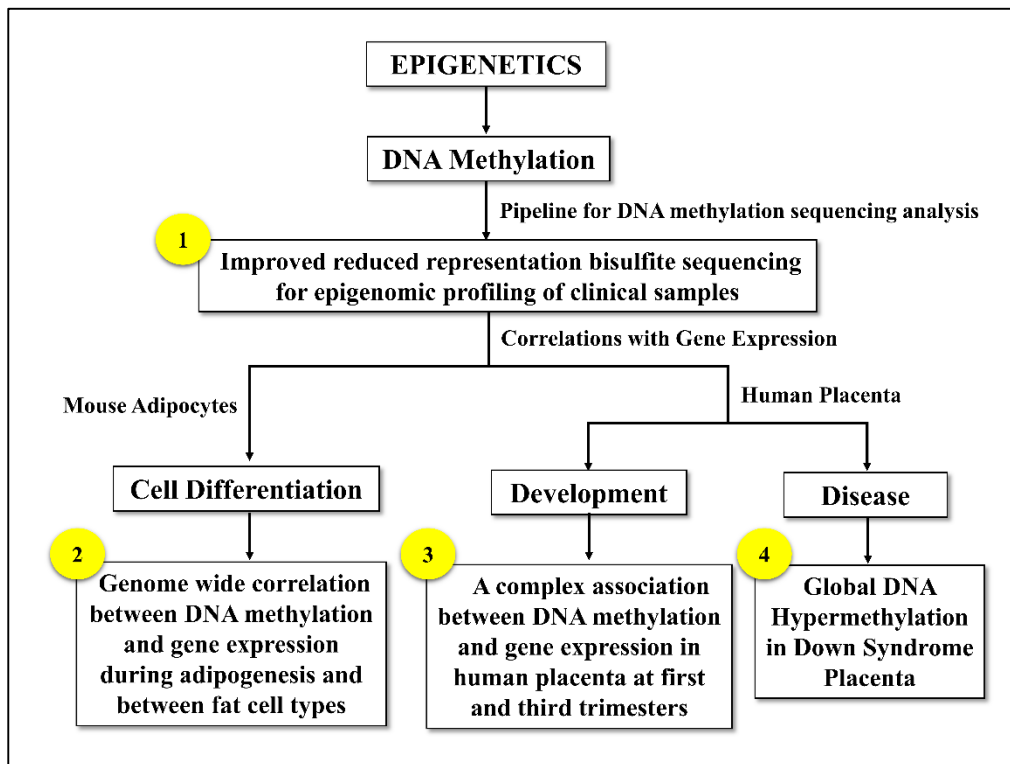
Figure 4: Project overview

2.1.   Specific Aims

2.1.1.   Study 1 – Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples (Chapter 3)

An overwhelming amount of biological signal is embedded within the mammalian methylome and analysed cytosine tend to be limited by platforms and computational pipeline. Different platforms such as (i) microarrays, (ii) immunoprecipitation, (iii) Meth-seq, have been developed [92] to increase the pool of analysable CpGs, while keeping cost at a manageable level.

To date, bisulfite sequencing (Bi-Seq) remains the golden standard to study DNA methylation [92]. Reduced representation bisulfite sequencing (RRBS) has emerged as the preferred method compared to whole genome bisulfite sequencing (WGBS) due to cost and genomic coverage considerations. There is increasing evidence to support that CpGs located in non-CpG rich regions exhibit correlation with gene expression. Thus, this calls for a need to improvise current protocol to select for such regions. In this project, we aim to

i.   Make use of a combination of two restriction enzymes to increase coverage on non-CpG rich regions

ii.   Improve existing protocol to increase number of unique and usable aligned reads by removal of repetitive regions

iii.   Develop a highly automated computational pipeline that allows user to quantify DNA methylation percentage for each sequenced CpG simply by inputting raw sequencing files and running through simple commands.

2.1.2. Study 2 – Epigenome-wide DNA methylation landscapes reveals systematic differences during adipogenesis and define cell type specificity (Chapter 4)

Obesity is defined as a phenotypic manifestation of abnormal or excessive fat accumulation. It results from both an increase in the adipocyte cell size and the development of new mature cells from undifferentiated precursors [99, 100]. Obesity has been known to be well associated with metabolic risk factors such as diabetes, dyslipidaemia and hypertension [101]; conditions which indirectly increase mortality [99].

The fat body distribution is variably distributed and its deposition affects metabolic risk towards the diseases [101]. Fat storage occurs in the adipocytes, which are the main cellular component of the adipose tissues. There are two main types of fat tissues, namely the (i) white adipose tissues (WAT) and (ii) brown adipose tissues (BAT). The former is the predominant type of fat tissue in human and is used as a storage depot for excess energy, whereas the latter is essential for classical non-shivering thermogenesis [102].

BAT is densely packed with mitochondria which expresses high levels of uncoupling protein 1 (*UCP1*). When activated, *UCP1* facilitates a proton leak across the inner membrane of mitochondria to mitochondria matrix without ATP synthesis, resulting in heat generation [103-105]. Such exclusive characteristic of BAT makes it an attractive anti-obesity therapeutic target [104, 106]. Interestingly, recent studies have shown that depots of WA (white adipocytes) showed morphological resemblance to BA (brown adipocytes) when subjected to stimuli such as β-adrenergic receptor agonist or proliferator-activated receptor-γ (*Ppar-γ*) [105, 106]. With these, it is of

31

intense interest on how to modify metabolism of WA to resemble BA-like properties to be used as an anti-obesity therapeutic target.

Before that, it is important to first understand the underlying biochemical and molecular differences between WAT and BAT, from a DNA methylation perspective. In this study, I aim to determine the importance of DNA methylation to adipocytes by addressing the following:

i.   Examine genome-wide DNA methylation changes during white and brown adipogenesis

ii.   Identify DNA-methylation regulated genes that are important in defining fat cell specificity

2.1.3.   Study 3 – A complex association between DNA methylation and gene expression in human placenta at first and third trimesters (Chapter 5)

The human placenta is a maternal-foetal organ essential for normal foetal development. During pregnancy, the placenta undergoes many structural and functional changes in response to foetal needs and environmental exposures. Previous studies have demonstrated widespread epigenetic and gene expression changes from early to late pregnancy. However, on the global level, how DNA methylation changes impact on gene expression in human placenta is not yet well understood. I seek to:

i.   Identify trends of DNA methylome changes in human placenta across gestational age, in different genomic regions including promoters and gene bodies

ii.   Understand the dynamic correlations between DNA methylation and gene expression changes at different gestational age.

iii.   Identify imprinted genes which are significantly changed from first to trimester in the placenta.

2.1.4.   Study 4 - Global DNA Hypermethylation in Down Syndrome Placenta (Chapter 6)

Down syndrome (DS) is a genetic disease caused by an extra partial or full copy of chromosome 21 (chr21). It occurs in approximately one out of 700 live births and is associated with over 80 clinically defined phenotypes. Extrinsically, this includes growth delays, characteristic facial features, and mild to moderate intellectual disability [107]. Intrinsically, organs such as central nervous system, heart, gastrointestinal tract and immune system are affected to varying penetrance [108]. Previous studies on human and mouse DS have reported on the gene dosage effect brought by an extra copy of chr21 [109-111]. Surprisingly, dysregulation of gene expression was also observed for genes located on other chromosomes [112-114].

Epigenetics is an important mechanism known to regulate many vital biological process and abnormalities have often lead to disease phenotypes. From this perspective, I investigated global DNA methylation differences between the placenta of mothers carrying normal and DS foetus to provide insights on the following:

i.   Understand potential perturbations of DNA methylation associated with DS

ii.   Investigate if these identified perturbations are functionally relevant to DS

CHAPTER 3 – IMPROVED REDUCED REPRESENTATION BISULFITE SEQUENCING FOR EPIGENOMIC PROFILING OF CLINICAL SAMPLES

3.1.   Background and Hypothesis

DNA methylation refers to the reversible biochemical process by which a methyl group is transferred from SAM to the 5' position of cytosine, producing 5mC [115, 116]. The forward and reverse reactions are catalysed by *DNMT* and *TET* enzymes respectively [6, 117].

The dynamic nature of DNA methylation allows for a switch in methylation marks at specific time-points that allows it to regulate cell differentiation, cell type specificity, X chromosome inactivation, parental imprinting and development [11, 27, 29, 118-122]. Additionally, DNA methylation can also be subjected to environmental regulation which may change the methylation status, causing a potential deviation from the normal phenotype [123, 124].

Precise and accurate quantification of DNA methylation is a crucial prerequisite to decipher and elucidate various mechanisms and pathways that could be perturbed by abnormalities in DNA methylation. The human genome contains about 29 million CpGs which are heavily methylated in genomic locations such as repetitive elements (SINE, LINES) and hypomethylated in CpG island associated promoters [11, 39, 125, 126]. Due to limitations in technologies and cost-prohibition, most of the studies have centred around CpG enriched promoters with little focus made on non-CGI rich regions.

Of the methods and platforms used to investigate DNA methylation, bisulfite sequencing (Bis-Seq) remains the golden standard since it allows quantification at allelic, contiguous and single base pair resolution levels [92, 127]. Despite the theoretical suitability of the technique, its practical application on genome-wide studies is faced with computational and financial challenges. In clinical studies where statistical power is crucial in a well-designed and well-constructed experimental design, it may require tens or hundreds of samples to be sequenced. Consequently, performing of whole genome bisulfite sequencing (WGBS) on all samples might pose cost impracticability for a single study. To curb this issue while keeping the main advantages of Bis-Seq, Meissner et al. [128-130] developed a technique, Reduced Representation Bisulfite Sequencing (RRBS), that makes use of a single restriction enzyme (RE) digestion (MspI) to select for CpG enriched genomic regions, thereby reducing the number of sequenced reads. A major drawback of this method is that it does not cover non-CpG rich regions such as gene bodies. Recent studies have shown CpG poor regions which are distal to core promoters can perform important regulatory functions [131]. Furthermore, gene bodies which are often less CpG enriched than promoters have been shown to have correlate positively with gene expression, possibly by modulating alternative splicing events or transcription initiation sites [57, 58, 132].

In this project, improvements have been made to the existing protocol by making use of a combination of two REs, MspI and Taq$^{\alpha}$I, to cover both CpG rich and reasonably, the non-CpG rich regions. Furthermore, a computational pipeline has been developed to analyse bisulfite converted samples for quantification of DNA methylation at each sequenced CpG site.

3.2   Material and methods

3.2.1. Samples

Many clinical samples such as human placenta [133], umbilical cord and leukocytes have been processed using the described improved RRBS pipeline. Data from one of the human buffy coat samples will be presented. This study is supported by a Bench to Bedside grant (09/1/50/19/622) from BMRC-NMRC.

3.2.2. Experimental protocol

A schematic view for the experimental protocol is shown in Figure 5.

Briefly, 1–5 μg of high molecular weight (>10 kb) genomic DNA was used for each library preparation. Each DNA sample was sequentially digested by MspI and Taq$^{\alpha}$I. The digested product was then purified, end-repaired, 3′-end-adenylated and adapter-ligated. DNA fragments were size selected using gel electrophoresis with 3% agarose gel. Previous in silico analysis performed showed enrichment of repetitive sequences lying between 198 to 206 bp (with adaptor) and were thus first removed using GeneCatcher Gel Excision kit. Subsequently, fragments of 150-197 bp and 207-230 bp were selected and purified by MinElute Gel Extraction Kit.

Each DNA library was analysed by two lanes of paired-end sequencing (2×36 bp) read on an Illumina Genome Analyzer II$_x$. The read quality was then discriminated by a Phred score, given by the formula:

$$Q(X) = -10 \log_{10}(P(\sim X))$$

where $P(\sim X)$ is the estimated probability a wrong base calling [134, 135]. A cut-off of 30 (Q30) was used, implying that probability of wrongly called base is 0.001.
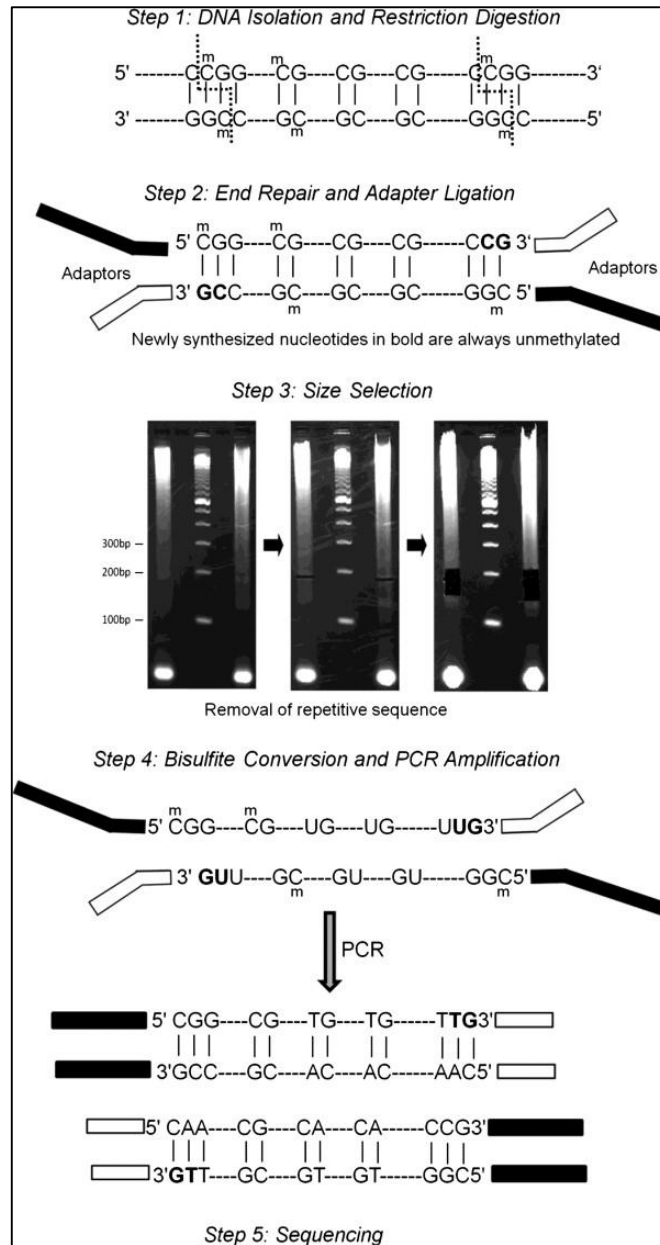
Figure 5: Key laboratory steps in RRBS. Figure originally from [136].

3.2.3.   Computational processing

A schematic view for the computational pipeline is shown in Figure 6.

The reference genome for the organism was converted independently twice into two

new references, (i) all cytosine converted to thymine (C2T converted genome) and

(ii) all guanine to adenosines (G2A converted genome). Usage of two references

allows the discrimination of sequenced CpGs into original bisulfite unconverted Watson and Crick strands.

High quality paired-end reads in Fastq format were first converted in silico into either C2T or G2A based on C/G base count ratio. These converted reads were then aligned to both converted genomes using the Bowtie program [137]. The bisulfite conversion rate was calculated by:

$$\text{Bisulfite Conversion Rate} = \frac{\text{non-CpG}\,C \to T}{\text{non-CpG}\,C \to C + \text{non-CpG}\,C \to T} \times 100\%$$

Where non-CpG C→T indicates the number of successful conversion of C to T in non-CpG sites, and non-CpG C→C indicates the number of failed conversion of C to T in non-CpG sites. This method uses only non-CpG cytosine which are mostly methylated, to estimate conversion rate. By such calculation, any unconverted cytosine is entirely attributed to incomplete conversion. Therefore, this is a conservative measurement that underestimates the conversion rate, especially in embryonic stem cells where non-CpG cytosine is observed [39, 138].

Due to the possible complications of polymorphic CpG sites, CpGs sites that had combined 'CG' and 'TG' for less than 80% of the reads were considered polymorphic and filtered from the analysis.

Finally, for each CpG site, DNA methylation was quantified by the ratio of C to (C+T) in the Watson strand and G to (G+A) in the Crick strand.
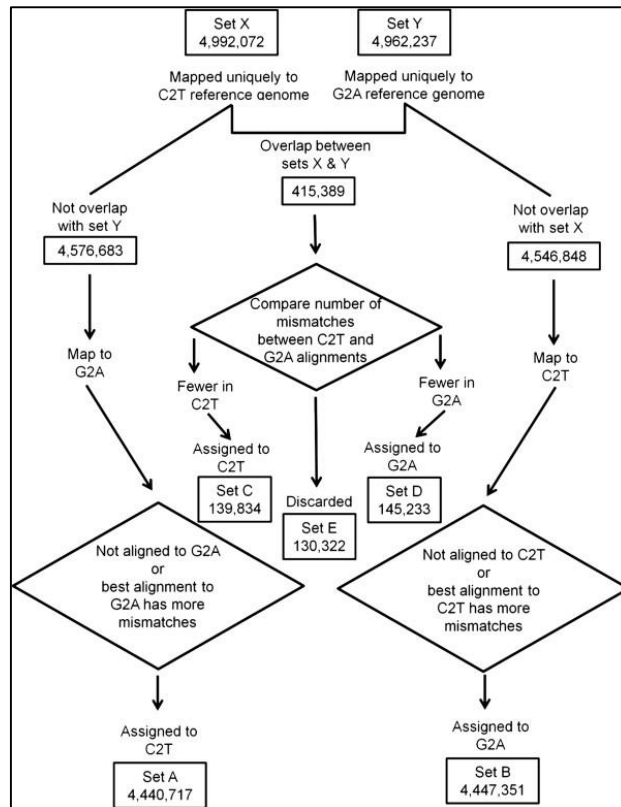
Figure 6: Computational steps in data processing. Figure originally from [136].

3.3   Results

A total of 289 motifs which were recognized by combinations of REs were analysed

and in silico analysis was performed using Hg19 as testing reference genome. The

assessment of the best combination was determined by the following criteria:

i.   Commercial availability of the REs and sensitivity of CpG methylation

ii.   Size distribution of the genomic fragments generate by REs

iii.   Obvious advantage brought by using double enzyme digestion from

        traditional single restriction enzyme digestion

iv.   Coverage of different genomic locations, specifically promoters (defined

        as -1kb to +500 bp relative to a transcription start site), gene bodies,

        transcription termination site (TTRs, defined by -500bp to +500 bp

        relative to transcription termination site), CGIs and CpG island shores

        (CGSs)

Finally, the combination of MspI and Taq$^{\alpha}$I were selected for the protocol.

3.3.1. *In silico* comparison of single vs double restriction enzyme digestion

The standard RRBS protocol uses a single RE MspI, which recognises 5' – CCGG – 3', as a cutting site to select for CpG enriched regions. In this improved protocol, a second RE, Taq$^{\alpha}$I, which targets 5' – TCGA – 3' was added to widen genomic coverage.

To assess the improvements brought by the addition of Taq$^{\alpha}$I, simulation of double RE digestion was done on the genome under the assumptions of 100% efficiency and no off-target products. With double RE, the number of targeted 80-160bp fragments increased from 263,890 to 450,689, representing an approximate increase in 25% of CpG sites (Table 1) that was consistent across all chromosomes (Hg19) (Figure 7).

	MspI Recognition site: CCGG	MspI, Taq$^{\alpha}$I Recognition site: CCGG, TCGA
Total fragments	2,297,220	3,810,058
Fragments of 80-160bp	263,890	450,689
#CpGs covered in fragments of 80-160bp	1,508,818	1,950,458

Table 1: In silico comparison of the fragments generation and CpGs covered using a single and double RE (considering only CpGs on the Watson strand)
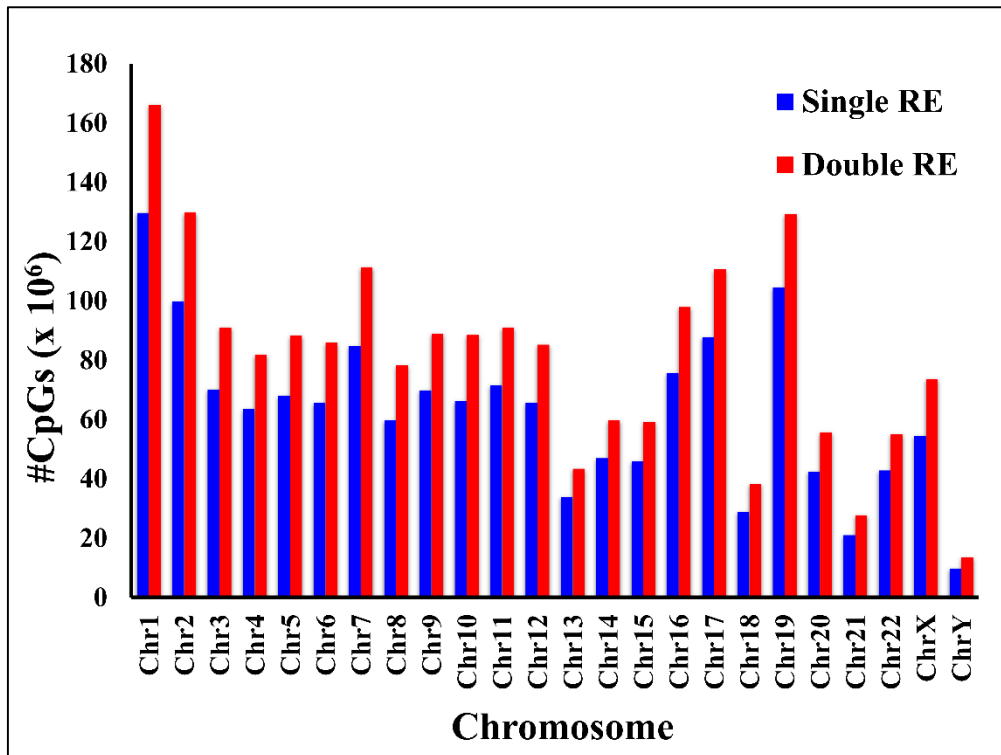
Figure 7: In silico comparison of covered CpGs in each chromosome (Hg19) using a single and double RE (considering only CpGs on the Watson strand)

In addition, there was a moderate increase in the coverage of CpGs in CGI (7.4%), CGIs (6.3%) and promoter (12.7%) regions and a marked escalation in CpGs that laid in non-CGI regions (41.8%) (Table 2).

	MspI	MspI, Taq$^\alpha$I	Percentage increase (%)
CpGs in CGI regions	1,098,462	1,180,058	7.4
CpGs in non-CGI regions	1,919,174	2,720,858	41.8
CGIs*	20,227	21,511	6.3
Promoters* (-1kb and 500bp from TSS)	24,520	27,633	12.7

Table 2: In silico comparison of the genomic coverage of CpGs comparing the use of 1 and 2 restriction enzymes. *At least 3 CpGs need to be present in the region. Table originally from [136].

3.3.2. Pipeline assessment using a buffy coat sample

The results for a buffy coat sample will be presented below.

3.3.2.1. Quality control measures

Paired-end sequencing was performed on Illumina Genome Analysed IIx platform. Samples were loaded onto eight-lane flow cell surface, on which each lane contains 120 tiles. At the end of each sequencing run, several quality controls measures were imposed to ensure data quality.

i.   The proportion of good quality to total sequenced reads will be is quantified for each tile and this typically lies between 70-90%.

ii.   The number of uniquely aligned reads to C2T or G2A (mutually exclusive) references will be assessed and the percentage generally ranges from 55-65%.

iii.   Bisulfite conversion rate which is an indicator of bisulfite efficiency will be evaluated. With the exception of embryonic stem cells which have higher levels of non-CpG cytosine methylation [39, 138], most samples typically have a conversion rate of 99%.

3.3.2.2. Assessment of library quality

The sequence specific cutting property of restriction enzyme digestion confers some unique features to the prepared RRBS library.

i.   Following the protocol principles of sequence specific cleavage followed by end repair, bisulfite conversion and PCR, the first three nucleotides for Read 1 should be CGG/TGG (MspI), or CGA/TGA (Taq$^\alpha$I) and CAA for read 2 (*Figure 8*). As predicted, 97.8% of Read 1 had CGG/TGG/CGA/TGA for first three nucleotides while 91.9% of Read 2 had CAA for the first three sequenced base.
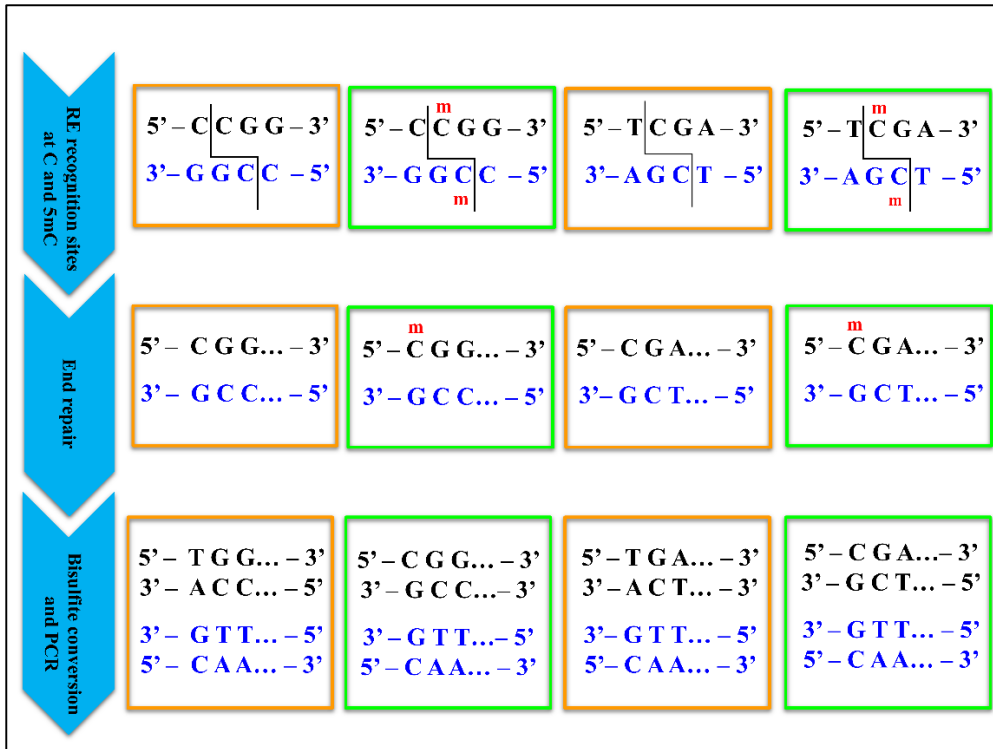
Figure 8: Four possible paired-read ends from combinations of two REs, bisulfite conversion and PCR.

ii. As illustrated in Figure 8, it is expected that Read 1 should either align to the positive strand of C2T reference or negative strand of the G2A reference. Correspondingly, Read 2 must be aligned in the exact opposite direction to Read 1, i.e. negative strand of C2T reference and positive strand of the G2A reference (Table 3).

Sample	C2TRef				G2ARef			
	Read 1		Read 2		Read 1		Read 2	
S1	2	-	3,412,630	-	3,406,274	-	3	-
	3,412,630	+	2	+	3	+	3,406,274	+
S2	1	-	3,143,620	-	3,139,936	-	1	-
	3,143,620	+	1	+	1	+	3,139,936	+
S3	3,283,607	+	3,283,607	-	3,274,004	-	2	-
					2	+	3,274,004	+
S4	1	-	3,157,175	-	3,144,735	-	1	-
	3,157,175	+	1	+	1	+	3,144,735	+
S6	3	-	3,945,575	-	3,915,666	-	5	-
	3,945,575	+	3	+	5	+	3,915,666	+
S7	3	-	3,955,564	-	3,931,783	-	3,931,783	+
	3,955,564	+	3	+				
S8	1	-	3,810,620	-	3,079,045	-	4	-
	3,810,620	+	1	+	4	+	3,079,045	+

Table 3: Number of reads aligned to the positive or negative strand of the two converted reference genomes. C2TRef: C2T reference genome; G2ARef: G2A reference genome. Table originally from [136].

iii.   When the relative ratio of cytosine to guanine is less than 1 in read 1, the corresponding read 2 would give a ratio of more than 1.

iv.   The protocol aims to select for genomic fragments of size 80-160bp for which repetitive sequences have been intentionally excluded. (128-136bp). Library size distribution analysis performed on seven samples (excluding phi control) showed two clear peaks separated by a dip at about 120 to 140 bp (Figure 9).
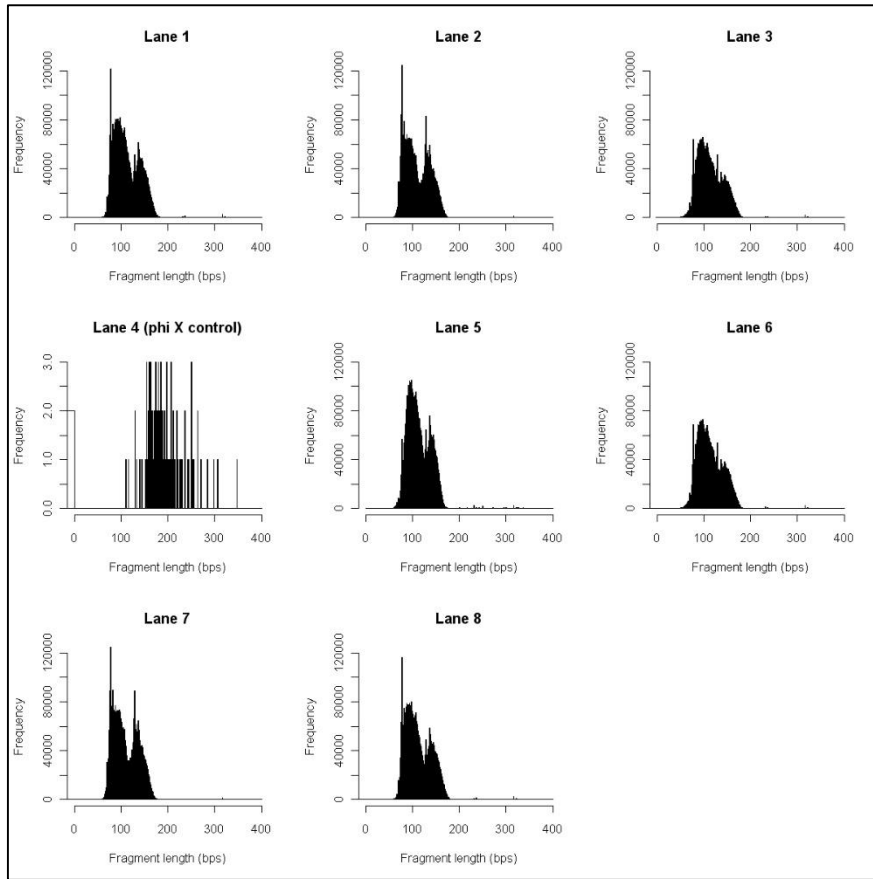
Figure 9: Distribution of Library Insert Length. Figure originally from [136].

3.3.2.3. Genomic coverage of CpGs

The sequencing depth of a CpG refers to the frequency of the site being sequenced. In RRBS, the choice of minimum sequencing depth is a fundamental yet crucial step in DNA methylation analysis. Its difficulty lies in optimizing the quality of DNA methylation estimation, yet not sacrificing the quantity of final analysed CpG list.

A minimum sequencing depth of 10 as the cutoff, which leaves approximately 1.8 million CpGs for subsequent analyses. This represented 3% of all CpGs in the genome, 76.7% of CGIs, 54.9% of CGSs and 52.2% of promoters (Table 4).

	Total in human genome	Covered by RRBS	Percentage (%)
CpGs	56,434,896	1,837,502	3.3
CGIs	27,718	21,252	76.7
CGSs	49,300	27,074	54.9
Promoters (-1kb and 500bp from TSS)	44,399	23,168	52.2

Table 4: Comparison of CpG coverages of specific genomic regions using RRBS to whole genome. At least three CpGs need to be present in each region. Table originally from [136].

In the context of CGIs, a total of 54% of these CpGs lie in either CGIs or CGSs (defined by 2kb upstream and downstream from CGI, Figure 10A). Alternatively, from the perspective of genomic context, most of the CpGs laid in intragenic regions (39%), followed by intergenic (37%), CGI promoters (18%), non-CGI promoters (4%) and TTRs (2%) (Figure 10B).

Under the consideration that single CpG effect might be random and sporadic, analysis were based on genomic regions with at least three covered CpGs. As shown in Figure 10C, 30% of the covered regions were located at CGIs and CGSs and remaining 70% lies in regions 2kb upstream and downstream away from CGIs. Separately, the covered regions were found most in intergenic regions (62%), followed by intragenic (28%), CGI promoters (6%), non-CGI promoters (3%) and finally TTRs (2%) (Figure 10D).
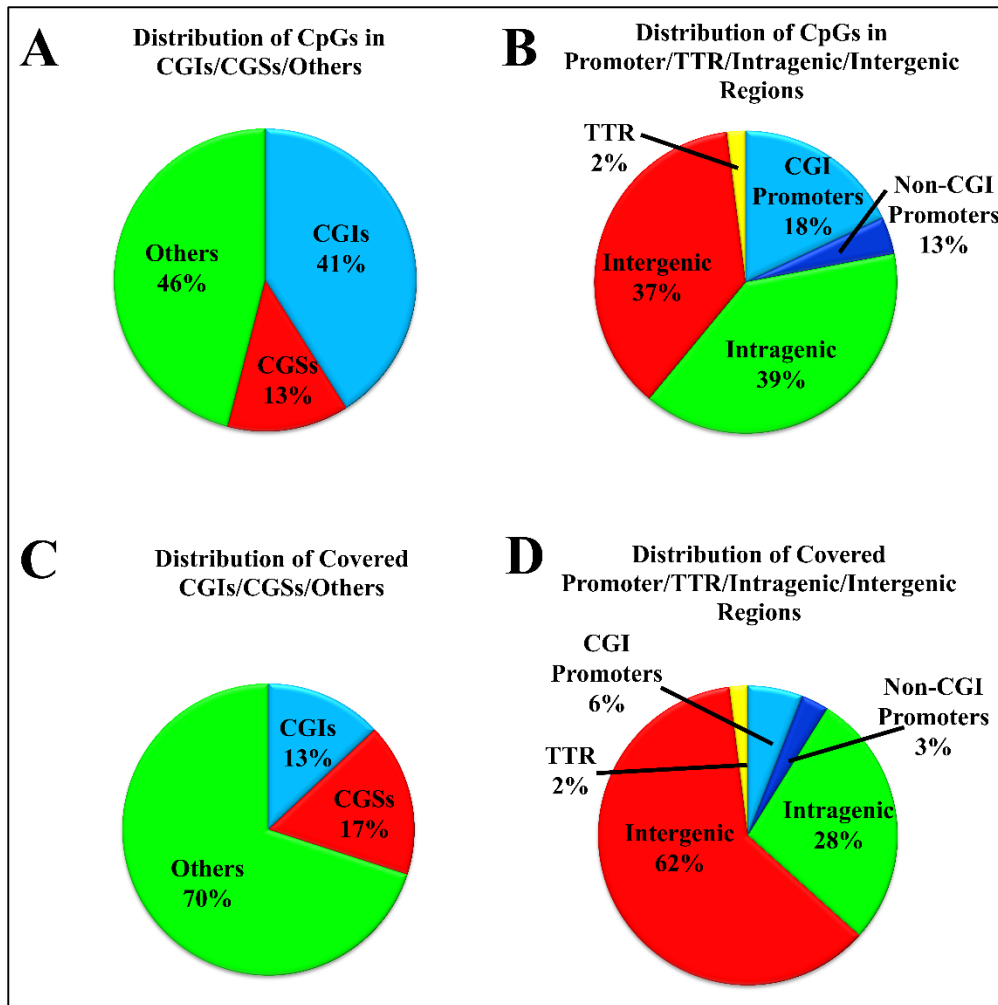
Figure 10: Genomic coverage of covered CpGs. *(A)* Distribution of CpGs in CGIs/CGSs/Others, using a sequencing depth ≥ 10 as the cutoff; *(B)* Distribution of CpGs in Promoter/TTR/Intragenic/Intergenic regions; *(C)* Distribution of genomic regions in CGIs/CGSs/others. *(D)* Distribution of genomic regions in promoter/TTR/Intragenic/Intergenic regions. A genomic region was considered covered if at least three CpGs within the region were sequenced at a depth ≥ 10. Figure adapted from [136].

3.4. Discussion

Through the addition of a second restriction enzyme to the existing RRBS protocol, it is now possible to cover both CpG rich and non-rich regions at a reasonable cost. This allows for genome-wide methylome analyses across large number of samples to be performed to decipher the role of DNA methylation in regulating fundamental biological processes such as cell differentiation, development and diseases.

3.5. Summary

Unlike most studies which typically focused only on either the experimental or computational aspects, this newly developed pipeline addresses improvements on both sides. Firstly, the usage of double RE digestion has greatly increased the number of analysable CpGs that are located outside of CGIs. Secondly, repetitive sequences have been deliberately removed to increase number of unique and usable reads for quantification. Thirdly, the computational pipeline is highly automated which could be operated by a few simple command lines on the Linux system.

CHAPTER 4 – EPIGENOME-WIDE DNA METHYLATION LANDSCAPE REVEALS SYSTEMATIC DIFFERENCES DURING ADIPOGENESIS AND DEFINE CELL TYPE SPECIFICITY

4.1. Background and Hypothesis

Improved quality of life and frequent consumption of energy-dense food are primary causes leading to an escalation of obesity rates (Figure 11) [139]. This alarming increase for the epidemic has brought much worldwide concerns due to its associated risks for diseases such as type 2 diabetes, cardiovascular diseases, hyperglycaemia, dyslipidaemia, hypertension and cancers [140-142]. Persistent rise in the morbidity eventually leads to increased financial burden on healthcare expenses and shorter life spans [143-145].
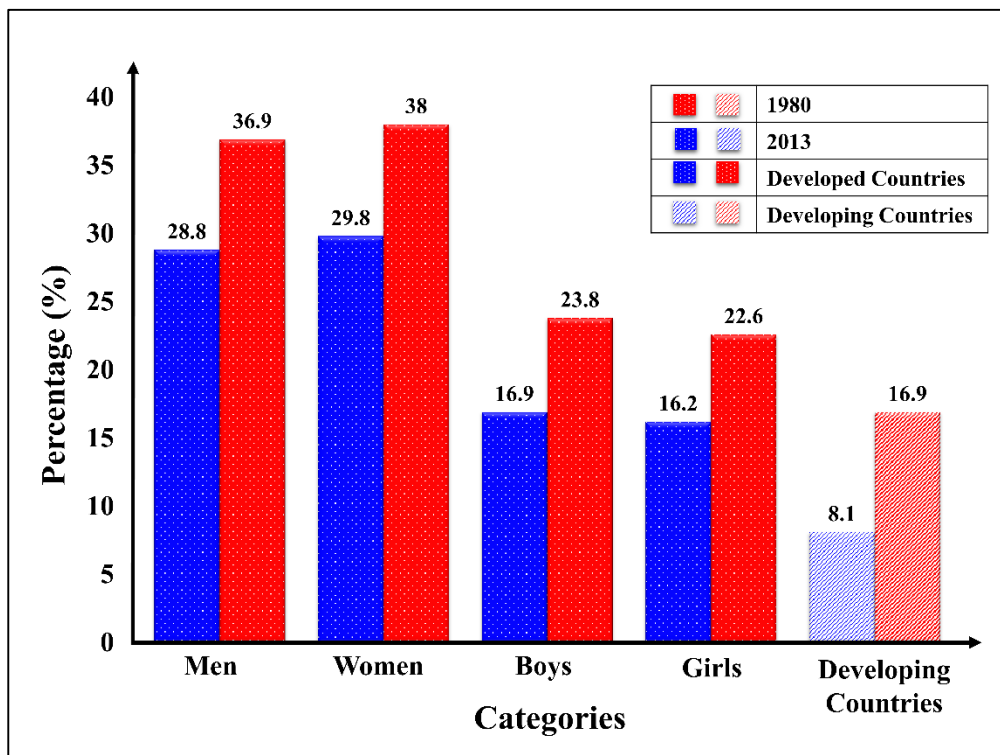
Figure 11: Prevalence of obesity from 1980 to 2013. Statistics taken from [139].

While white adipose tissues (WAT) function mainly as a primary organ of energy storage in the form of triglycerides to provide fuels during low glucose levels [142, 146], brown adipose tissues (BAT) is the dominant site for non-shivering thermogenesis. Brown adipocytes in BAT is densely packed with highly oxidative and naturally uncoupled mitochondria [146] which contains highly expressed *UCP1*. When activated via a cascade of triggering events starting from release of norepinephrine, *UCP1* facilitates a proton leak across the inner membrane of mitochondria to mitochondria matrix without ATP synthesis, resulting in heat generation [103-105] . Such exclusive characteristic of BAT makes it an attractive anti-obesity therapeutic target [104, 106].

The re-discovery of BAT present in adult human using positron emission tomography (PET) scans have challenged the long held belief that BAT did not play an important role in adult energy metabolism [147-149]. Adding on to the excitement in this field, a third potential type of fat cells, known as beige or brite adipocytes, are found in WAT. While these cells look indistinguishable from white adipocytes in the basal state, they take morphological resemblance to BA and express comparable levels of *UCP1* with BA when activated by β-adrenergic receptor agonist or proliferator-activated receptor-γ (*Ppar-γ*) (Figure 12) [105, 106].

Thus, elucidating molecular mechanisms that guide adipogenesis regulation and stimulation of browning in WA will build a complementary knowledge base in the identification of potential therapies or biomarkers in controlling obesity.
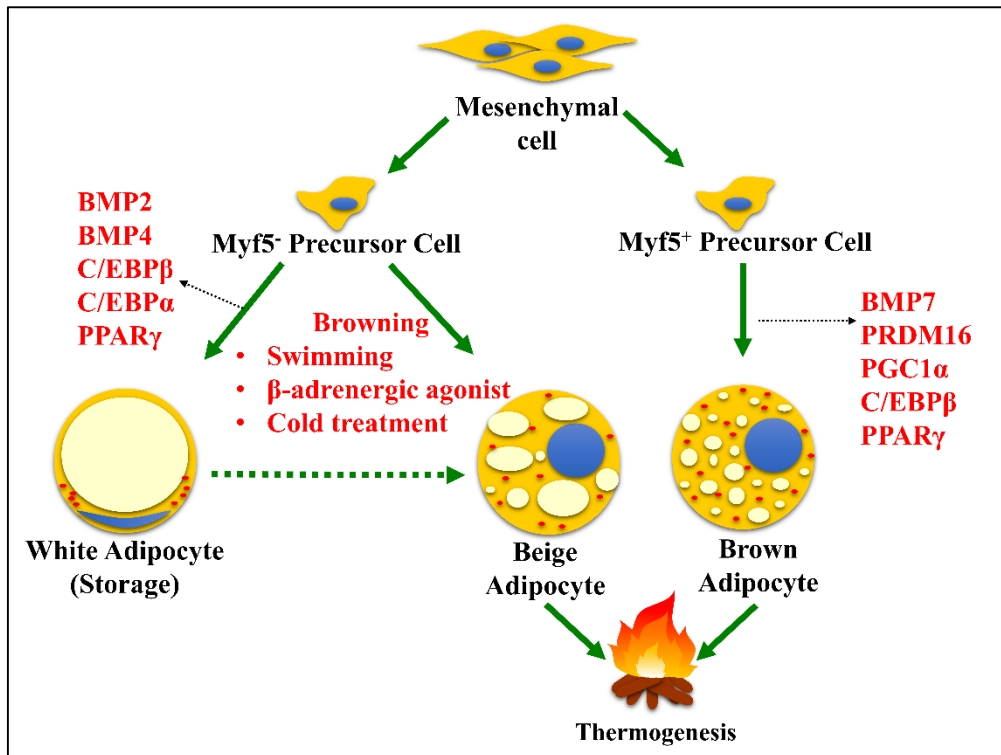
Figure 12: Origins of white, beige and brown adipocytes.

DNA methylation is one of the most extensively studied epigenetic mechanisms involved in regulating cell differentiation, cell type specificity, X chromosome inactivation, parental imprinting and development [11, 27, 29, 118-122]. Previous knockdown studies on DNA methyltransferases (*DNMT1*, *DNMT3A*, *DNMT3B*) have shown that while DNA methylation is dispensable for maintaining the undifferentiated state in embryonic stem cell (ESC), it is necessary in lineage commitment [150, 151]. Selective examination of methylation status of four adipogenic markers (*Leptin*, *Pparg2*, *Fabp4* and *LPL*) and three non-adipogenic makers (*MYOG*, *CD31* and *GAPDH*) in human mesenchymal stem cells, adipose tissue-derived stems cells and human ESCs showed that adipogenic promoters tend to be hypomethylated compared to non-adipogenic promoters, suggestive of an epigenetic control on expression of adipogenic specific genes [152]. Recent study by Gentile et al. [153] reported an accelerated adipogenesis in 3T3-L1 by silencing

DNMT1. This was accompanied by an apparent early induction of adipocyte-specific genes (*Glut4*, *Fabp4* and *PPARγ*), evident from gene expression data.

To understand if DNA methylation is important in regulating adipocytes differentiation and tracing fat cells lineage, possibility through transcription factors, I analysed the DNA methylome profiles of nine samples, covering both white, beige and brown adipocytes across specific differentiation states. For the first time, a DNA methylome landscape has been provided to understand (i) profile changes in white and brown adipogenesis and (ii) identifying methylome signatures for defining cell types. In summary, I observed a global hypermethylation as cells reach differentiation termination. I have also made comparisons between different cell types and observed that methylation profile was distinctively higher in white adipocytes (WA) than either brown-induced white adipocytes (BWA) or brown adipocytes (BA).

4.2. Material and methods

4.2.1. Cell isolation and cell culture

Brown and white preadipocytes were isolated from interscapular brown adipose tissues and inguinal white fat depot respectively. Eight BL6 mice were sacrificed at 3 weeks and the harvested fat tissues were minced, collagenase digested and fractionated. Pre-adipocytes which were enriched at the bottom stromal vascular fractions were collected, resuspended and cultured to confluence. Subsequently, white and brown pre-adipocytes were induced to differentiate into mature BA and WA by exposure to differentiation medium. In additional, an independent set of white pre-adipocytes was brown induced at day 0 by treatment with a Norepinephrine,

followed by differentiation induction. All cells types at days 0, 4 and 6 were used for both RRBS and RNA-seq library preparation.

4.2.2. DNA methylation

Differentiating cells at days 0, 4 and 6 for both white and brown adipogenesis and induced brown adipocytes at days 4 and 6 were included in this study, making up a total of nine samples. Sample preparation and computational processing followed the steps described in Chapter 3, Section 3.2.2 and Section 3.2.3 respectively. The mouse July 2007 (NCBI37/mm9) genome assembly was used throughout the study.

4.2.3. Differential DNA methylation analysis

Differential methylation analysis was performed at both single CpG and regional level. Only autosome CpGs (sequencing depth $\geq$10) common to all nine samples were included in all subsequent analyses.

Between a pair of samples, CpGs having methylation difference of at least 10% were selected for 2-sided fisher exact test. P values were then adjusted by the Benjamini Hochberg method. A CpG was considered significant if (i) difference between sample pair was at least 10% and (ii) FDR corrected p value <0.05. A promoter was considered significantly differentially methylated if it contains at least two significant CpGs, all of which must be regulated in the same direction (either all hypermethylated or hypomethylated).

All statistical analyses were performed using R package.

4.2.4. RNA-seq

Total RNA from adipocytes were extracted according to Qiagen miRNeasy kit. RNA-seq libraries were prepared according to NEBNext Ultra Directional RNA Library Prep Kit for Illumina and ran on Hiseq2000 sequencer platform. The 100bp paired-end reads were first quality checked with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), subsequently aligned to mm9, using Tophat (version tophat-2.0.11). The aligned reads were then input into Cufflinks (Version 2.1.1) to quantify gene expression into units known as fragments per kilobase of exon per million fragments mapped (FPKM). Genes with low expression (FPKM< 1 in all samples) were removed, leaving 14,491 autosomal genes.

4.2.5. q-PCR

Total RNA from cell samples was isolated as mentioned above. RNA was reverse transcribed into cDNA with random primers (SuperScript II Reverse Transcriptase, Invitrogen), followed by PCR amplification using gene specific primers. Sybr Green based qPCR was performed in an Applied Biosystems 7900HT Fast Real-time PCR System, using RPL23 as an internal control for normalization. Data were analysed by the relative quantification ($\Delta\Delta C_t$) method.

4.3. Results

4.3.1. Samples used

A total of nine samples were included in this study, covering three main adipogenic differentiation stages (Day 0, 4 and 6) for each type of adipocytes (BA, WA and BWA) (Figure 13).
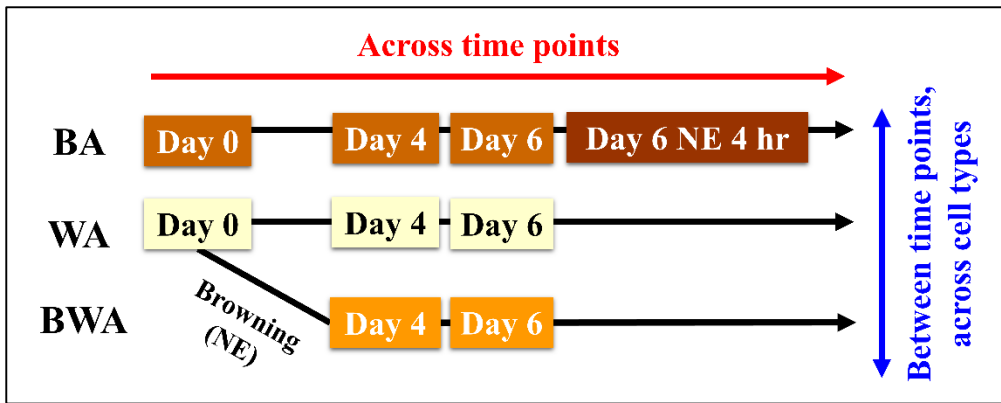
Figure 13: A total of 9 samples were included in this study. BA: Brown adipocytes, WA: white adipocytes, BWA: brown induced white adipocytes

The adipogenic model was first validated by quantitative PCR of general adipogenic and brown fat specific marker genes. All three general adipogenesis marker genes (*AdipoQ*, *Fabp4*, and Pparγ, Figure 14A-D) showed increased expression relative to day 0 for brown adipocytes (BA), white adipocytes (WA) and brown induced white adipocytes (BWA).

In addition, another three brown fat markers (*Pgc1a*, *Ucp1* and *Cidea*, Figure 14E-F) were selected and as expected, gene profiles for BA were distinctively different from WA and BWA. These data provided evidence that the cells had differentiated well into their designated respective cell types.
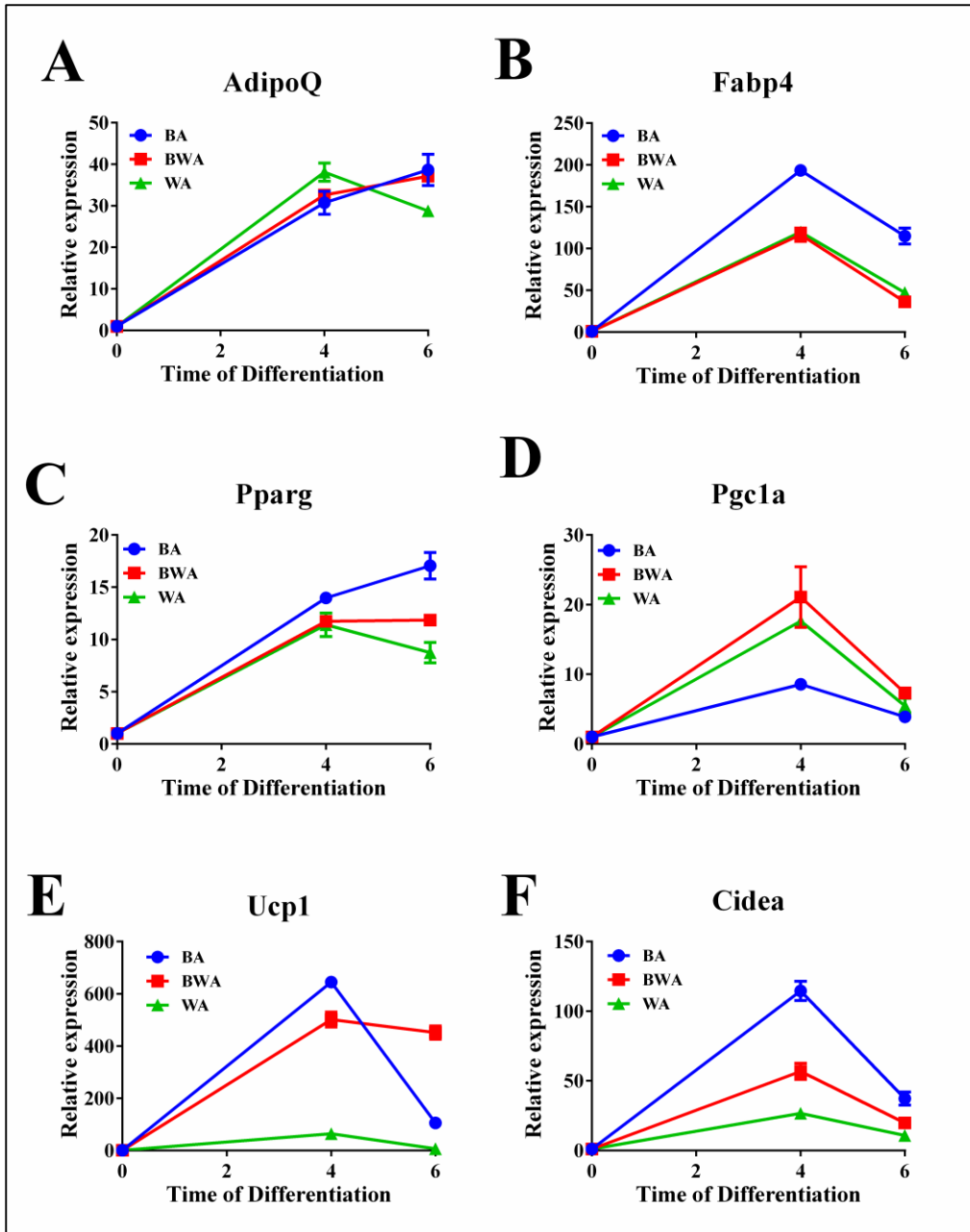
Figure 14: Quantitative PCR results for general fat markers (A, B, C) and brown fat specific markers (D, E, F).

4.3.2. Descriptive Statistics for RRBS quality

An improved version of RRBS [133, 136] was applied to interrogate genome-wide

DNA methylation profiles of (i) BA, (ii) WA and (iii) BWA at different stages of cell

differentiation and between cell types (Figure 13).

There was an average of 52.5 million pass filter reads per sample for nine samples, with an average alignment rate (C2T and G2A) of 60.0%. A high bisulfite conversion rate of above 99% was achieved for all the samples (Figure 15A). Using a minimum sequencing depth of 10 as the cut off, we obtained on average 1.3 million autosomal CpGs per sample (Figure 15B).
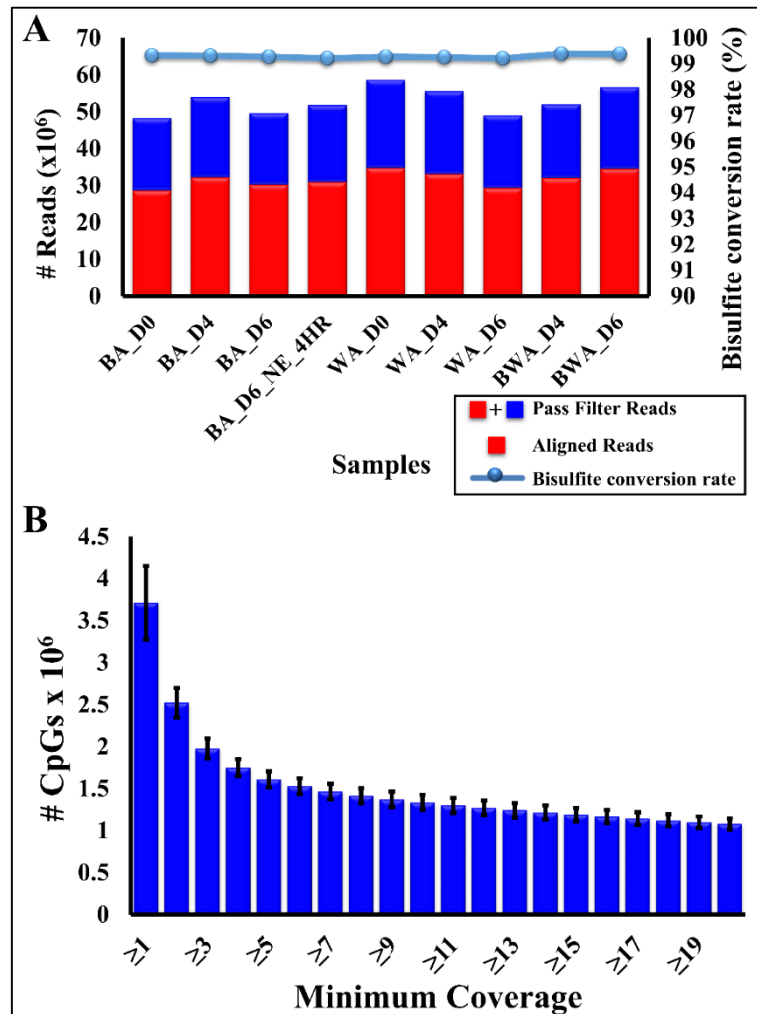
Figure 15: Descriptive statistics of the 9 samples. (A) For each sample, red bars represent the number of uniquely aligned reads, while blue bars are the non-uniquely or unaligned reads. The blue line gives the bisulfite rate for each sample. (B) The histogram gives the average number of CpGs sites for 9 samples, with varying minimum sequencing depths. Error bars represents standard deviation for 9 samples.

4.3.3. Genomic coverage of CpGs

To facilitate time series and cross cell type comparisons in an unbiased manner, only CpGs which with at least10× coverage for all nine samples were retained. This reduced the pool of analysable autosomal CpGs to 838,481. The CpGs were

approximately equally distributed in promoters (defined as 1kb upstream and 500bp from TSS, 287,243 CpGs, 33%), gene bodies (defined as from end of promoter to TTS, 315,733 CpGs, 40%) and intergenic regions (27%) (Figure 16A). About half of the CpGs were located within 6kb of annotated CGIs, comprising of CGIs (37%), CG shores (2kb up/downstream from CGI, 10%) and CG shelves (2kb up/downstream from CG shores, 3%) (Figure 16B). On a regional level, these CpGs covered 47.5% of promoters, 56.1% of gene bodies and 63.8% of CGIs on the mouse genome, for which each region required at least two covered CpGs. (Figure 16A-B).

Repetitive elements make up approximately 45% of the mouse genome (http://www.repeatmasker.org/species/mm.html) and DNA methylation changes in these regions have been associated with gene transcriptional regulation, differentiation and cancer [154]. The RRBS experimental protocol intentionally removes genomic fragments enriched for repetitive regions and thus only 14.7% of the analysed CpGs (122,818 CpGs) were mapped to repetitive elements. With this caveat, a total of 78% of all CpGs mapped to repetitive elements fell in class I of transposon elements (LINE, SINE and LTR), which was slightly lower than the 92.3% on a whole genome scale (http://www.repeatmasker.org/species/mm.html).
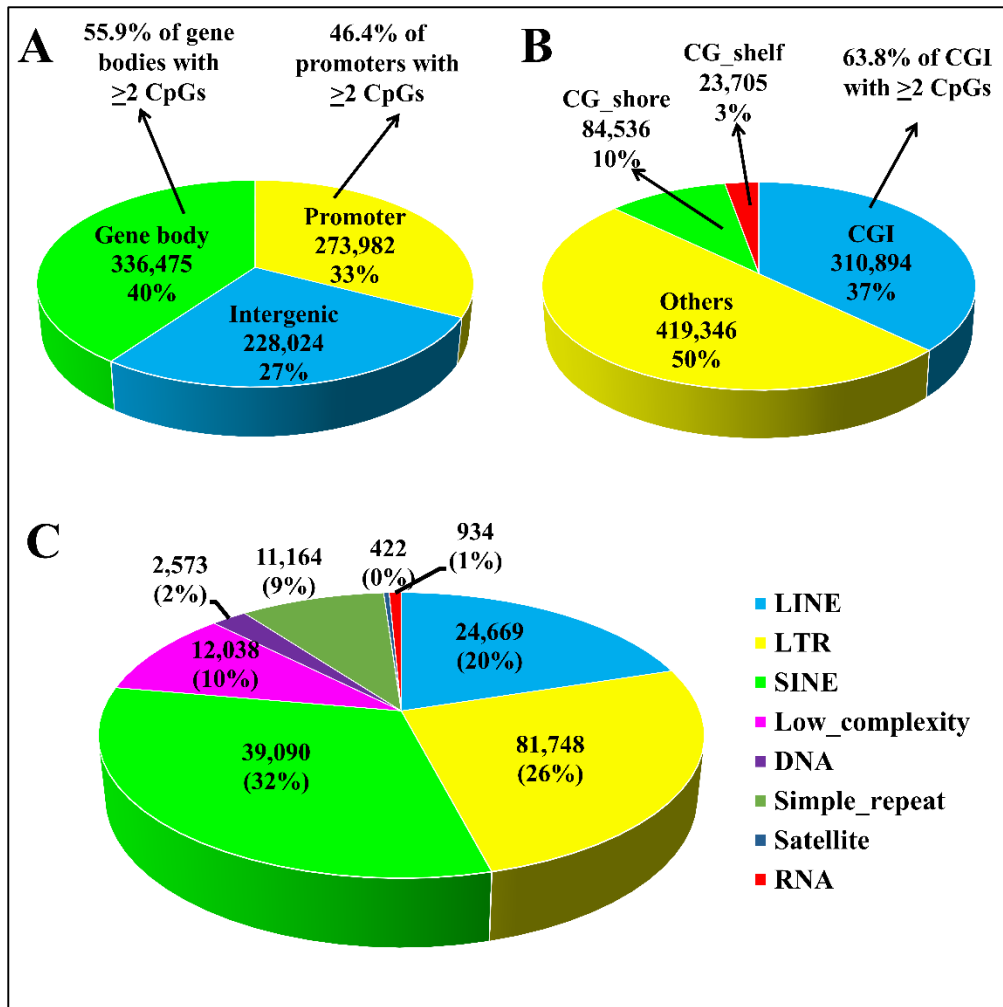
Figure 16: Coverage of CpGs for RRBS. A total of 838,481 CpGs were covered in all 9 samples, at a sequencing depth >10. (A) About one third of the CpGs are located in promoters, gene bodies and intergenic region each. Half of annotated promoters and gene bodies were covered. (B) A total of 50% of the CpGs lie in CpG rich and medium rich regions. (C) Combined percentage of LINE, SINE and LTR made up 78% of all 122,818 CpGs mapped to repetitive elements.

4.3.4. DNA methylation profile distinguishes samples by cell types while gene expression profile separates samples by stages of differentiation

To test if DNA methylome encrypt similar biological signals as transcriptome, I applied PCA and hierarchical clustering to the dataset. To do so, I used average methylation levels of multi-CpG genomic fragments created by merging nearby CpGs

as input for DNA methylome analysis while the time-matched RNA-seq data was used for gene expression analysis.

Although PCA and hierarchical clustering worked by different statistical and mathematical principles, they pointed towards consistent and interesting conclusions. By using methylome, samples were shown to be grouped by cell types than by stages of differentiation (Figure 17A-B). In addition, closer proximity of BWA to WA than BA suggested that BWA and WA shared greater epigenetic similarity. On the other hand, same analyses performed using gene expression profile clearly shows that samples were, instead, separated by stages of differentiation (Figure 17C-D).
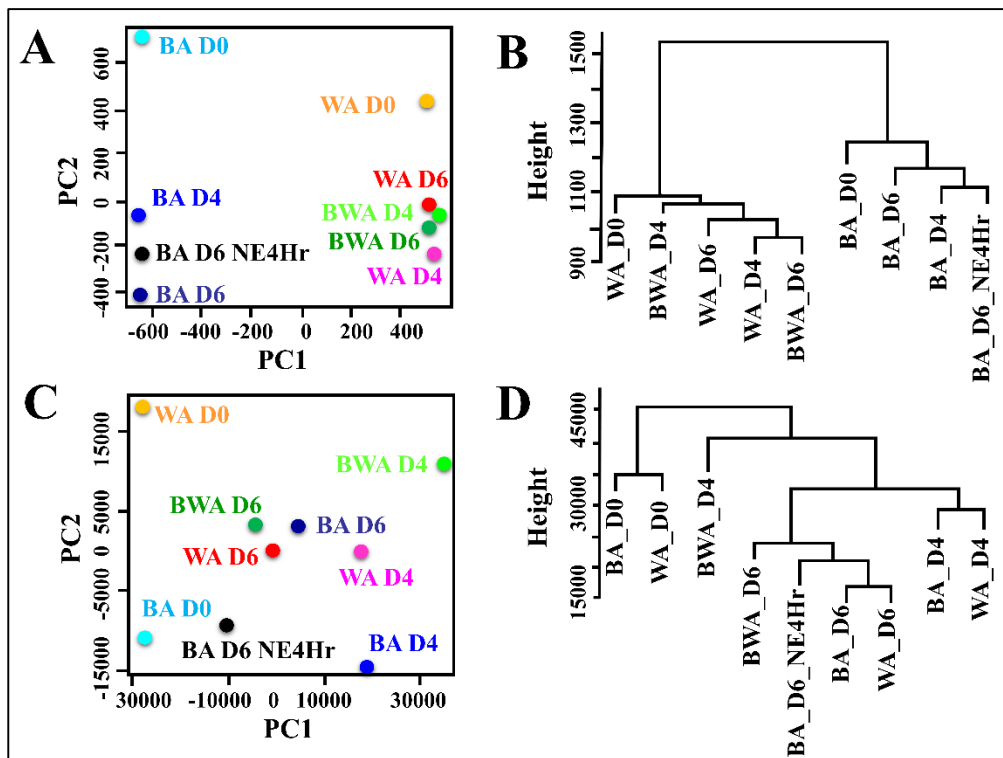
Figure 17: PCA and hierarchical clustering analyses of DNA methylation (A, B) and gene expression data (C, D) of all nine analysed samples.

4.3.5. Distinct hypermethylation was observed in WA than BA or BWA

Results from Figure 17 intrigued me to probe further test if there are any systematic differences in the DNA methylome between cell types so as to gain a deeper insight

to understanding the role of DNA methylation in defining adipocyte specificity. To do so, I made pairwise comparisons between various cell types for all stages of differentiation. I only considered CpGs having at least 10% difference to remove for noisy data. Consistent among all possible combinatorial pairs of comparisons (BA vs. WA, BA vs. BWA, BWA vs. WA) within each time point, greatest dissimilarity was observed between BA and WA. Additionally, across all analysed time points, there was a uniform trend of higher frequency of CpGs being hypermethylated in WA than BA (Figure 18A). Persistent skewed difference observed throughout adipogenesis suggested the regulatory role of DNA methylation in defining and maintaining cell lineage. Trend differences between BA and BWA were similar to Figure 18A, though to a relatively smaller extent (Figure 18B). In contrast, comparisons of hypermethylated and hypomethylated CpGs between WA and BWA showed more subtle difference in numbers (Figure 18C). These further supported conclusions made from PCA and hierarchical clustering diagrams from Figure 17A-B.
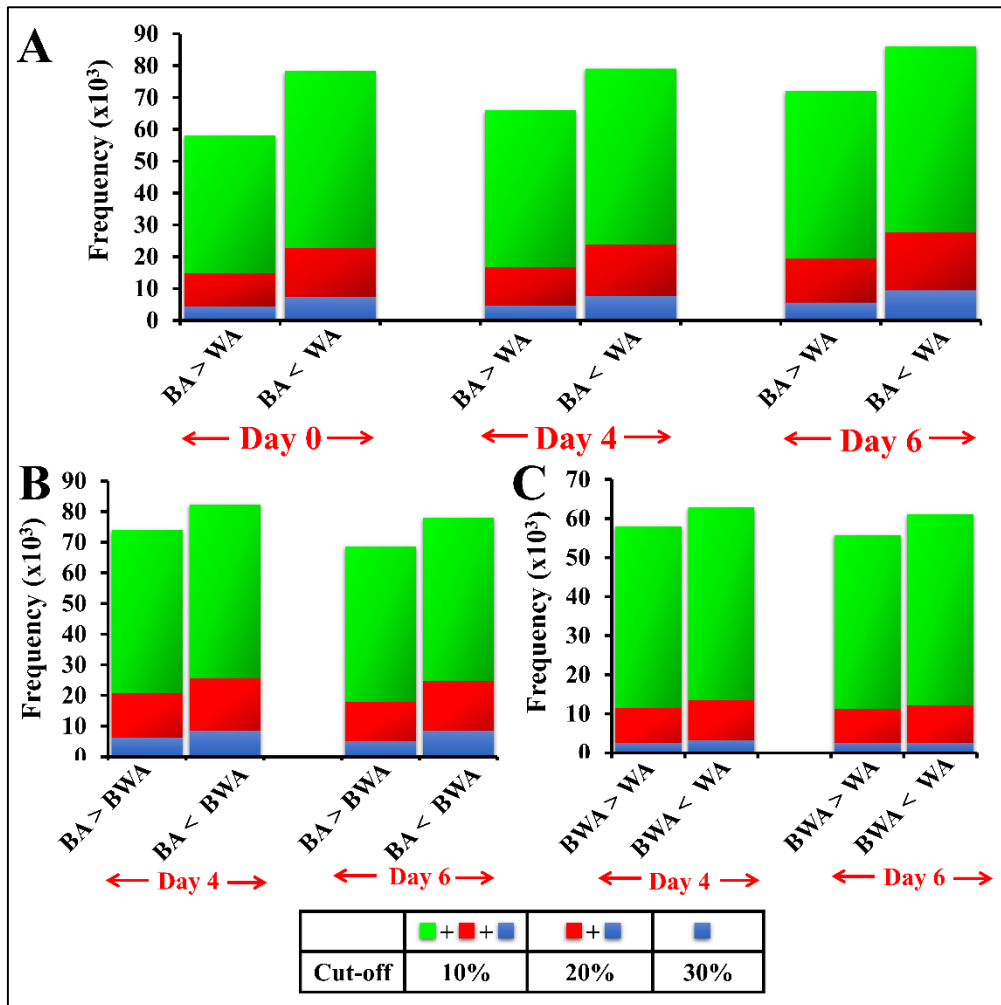
Figure 18: Frequencies of hypermethylated and hypomethylated CpGs between cell types at different stages of differentiation, using varying cutoffs at 10, 20 and 30%. (A) Higher frequency of hypermethylated CpGs in WA than BA at 10%, 20% and 30 cut-off at days 0, 4 and 6. (B) Higher frequency of hypermethylated CpGs in BWA than BA at 10%, 20% and 30 cut-off at days 4 and 6. (C) Higher frequency of hypermethylated CpGs in WA than BWA at 10%, 20% and 30 cut-off at days 4 and 6.

4.3.6. Dominance of hypermethylated DMCs in WA than BA in non-promoter regions

Starting with CpGs having at least 10% methylation difference, Fisher exact test followed by FDR were applied to identify significantly differentially methylated CpGs (DMCs) between all pairs of cell types (Figure 19). Across all differentiation stages, there were more significant DMCs in WA than BA and in WA than BWA. On the contrary, comparisons between BWA and WA on both days 4 and 6 generated

much fewer number of DMCs than the rest of comparisons. Consolidating the

observations, I inferred that the

    i)   DNA methylation was the highest in WA and lowest in BA.

    ii)   DNA methylome between WA and BWA were highly comparable

    iii)   Greatest cell type methylation difference occurred at day 4

    iv)   Least cell type methylation difference occurred at day 6, i.e. fully

              differentiated adipocytes
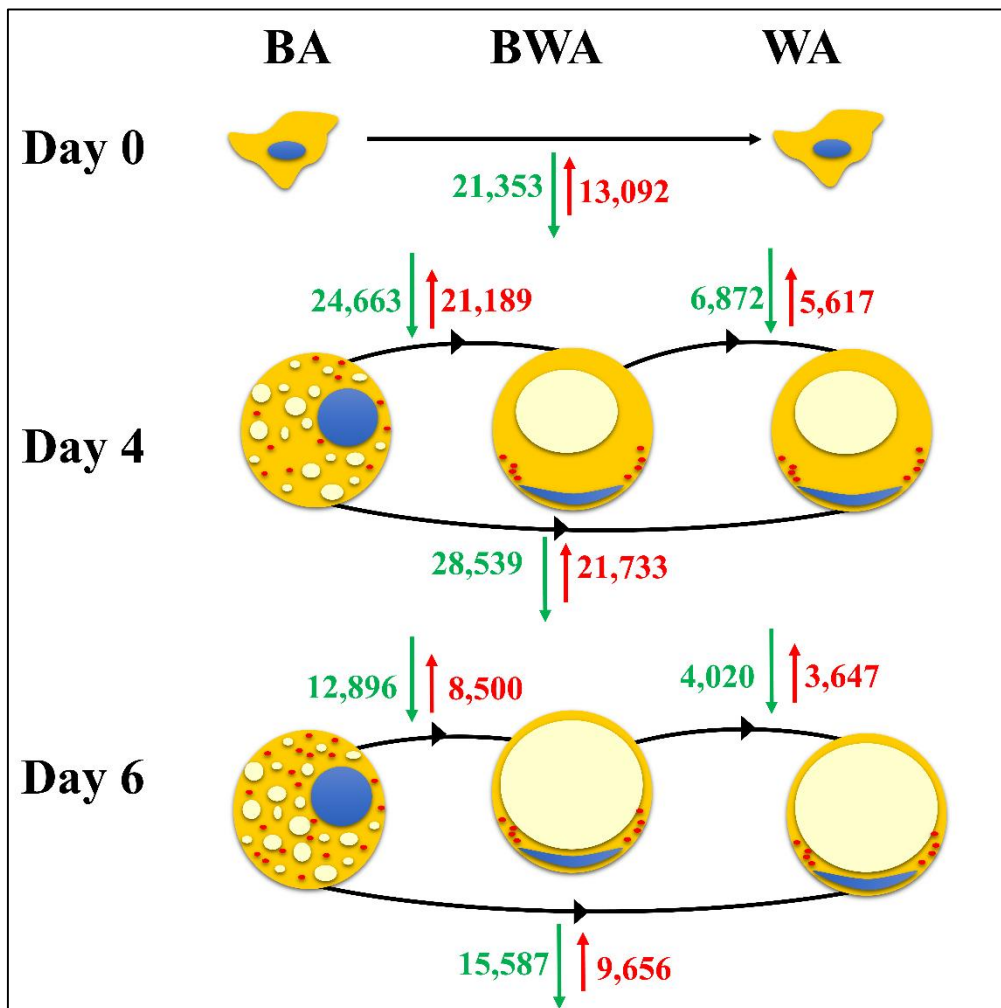
Figure 19: Frequency of significant CpGs by comparing different cell types at various stages of differentiation. Green arrows: hypomethylation. Red arrows: hypermethylation.

Next, I asked if these DMCs were enriched at any genomic locations. Upon mapping

these statistically significant DMCs to their respective genomic locations, most

changes occurred at non-promoter regions (Figure 20). To test if there was a

dominance of hypermethylation in all genomic locations, I examined the ratios of

hypermethylated and hypomethylated DMCs within each genomic category. Amongst

all categories, exons showed greatest deviation from equal proportions, followed by

introns, intergenic and finally promoters (Figure 20). This was consistent in all other

comparisons.
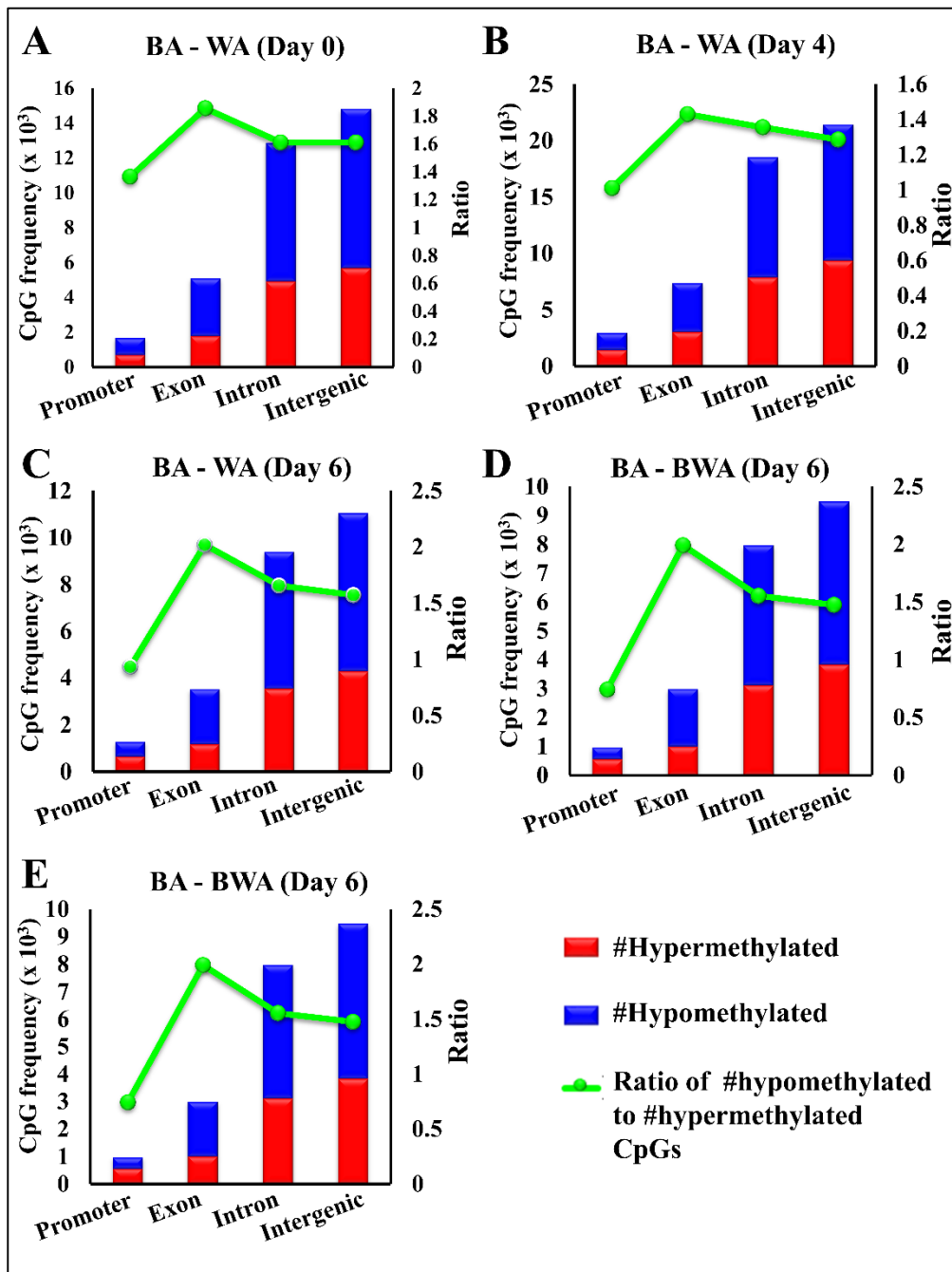
Figure 20: Proportions of significant differentially methylated CpGs in each genomic category. DMCs were mostly located in intron and intergenic regions. Ratio between hypermethylated and hypomethylated DMCs were highest in exonic regions.

4.3.7. Gene ontology analysis for DMPs

Extending from single CpGs, I went on to identify significantly differentially methylated promoters (DMPs). Similar to DMCs analyses, dissimilarity between cell types were most apparent between BA and WA (Table 5, Table 6 and Table 7), especially at day 4 (Table 5).

RNA classes	Day 0		Day 4		Day6	
	BA>WA	BA<WA	BA>WA	BA<WA	BA>WA	BA<WA
mRNA	70	86	144	138	58	54
miRNA	3	8	10	10	2	3
ncRNA	8	18	22	32	6	10
Total	81	113	176	180	66	67

Table 5: Frequencies of DMPs between BA and WA at days 0, 4 and 6 their respective RNA classes.

RNA classes	Day 4		Day6	
	BA>BWA	BA<BWA	BA>BWA	BA<BWA
mRNA	122	123	50	42
miRNA	8	6	0	6
ncRNA	20	25	2	5
Total	150	154	52	53

Table 6: Frequencies of DMPs between BA and BWA at days 0, 4 and 6 their respective RNA classes.

RNA classes	Day 4		Day6	
	BWA>WA	BWA<WA	BWA>WA	BWA<WA
mRNA	22	31	4	19
miRNA	3	2	1	0
ncRNA	6	4	0	2
Total	31	37	5	21

Table 7: Frequencies of DMPs between BWA and WA at days 0, 4 and 6 their respective RNA classes.

I next examined if DMPs identified between BA and WA showed any biological importance by using QIAGEN's Ingenuity IPA to implement gene ontology analysis. Upon separate analyses performed on hyper- and hypomethylated DMPs, two of five top regulatory networks in hypomethylated DMPs (BA<WA) showed relevance to BA functions. These networks were "Cancer, skeletal and Muscular disorders, Tissue morphology" and "Energy Production, Lipid Metabolism, Small molecule biochemistry" (Figure 21). Genes in the latter network (Figure 21*B*) showed direct

and indirect connections with *PPARα*, a brown fat specific marker that was proposed to be involved in the coordination of transcriptional activation for lipid oxidation and induction of *UCP1* [155].
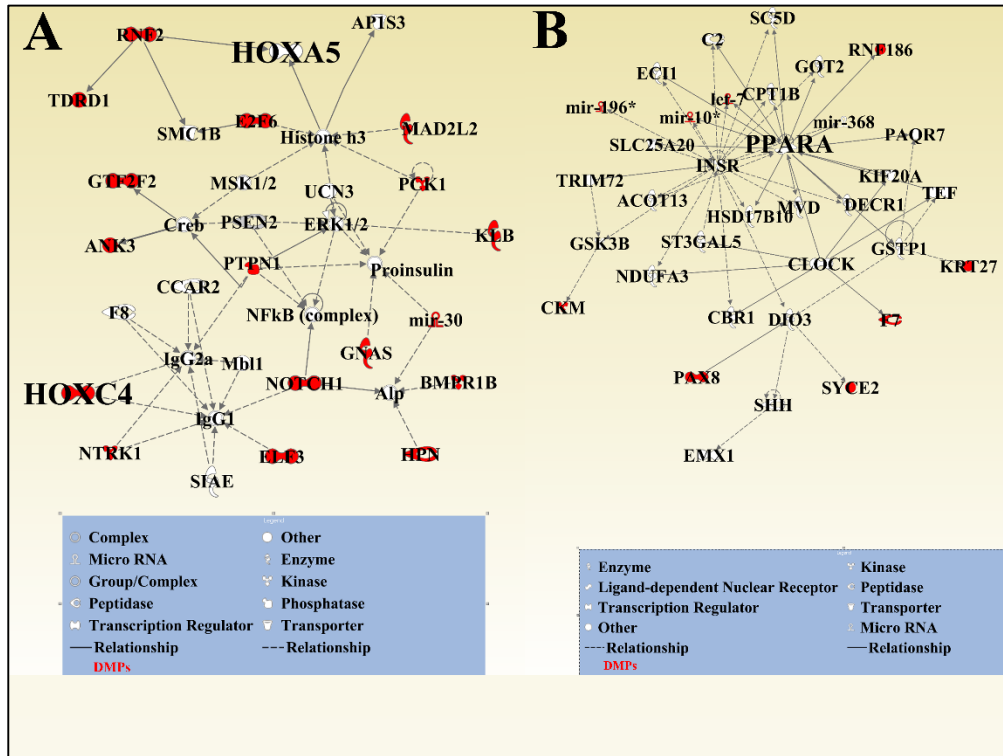
Figure 21: Two out of five top regulatory networks showed relevance to brown adipocytes functions. (A) Cancer, skeletal and Muscular disorders, Tissue morphology (B) Energy Production, Lipid Metabolism, Small molecule biochemistry

4.3.8. Strongest correlation between promoter DNA methylation and gene expression was seen at day 4

To test if the DMPs have any functional significance, sample matched RNA-seq data was added to the analysis. Up-regulated genes in the mature BA compared to mature WA showed enrichment in brown fat related functions, validating the quality of the RNA-seq dataset (Figure 22A-B).
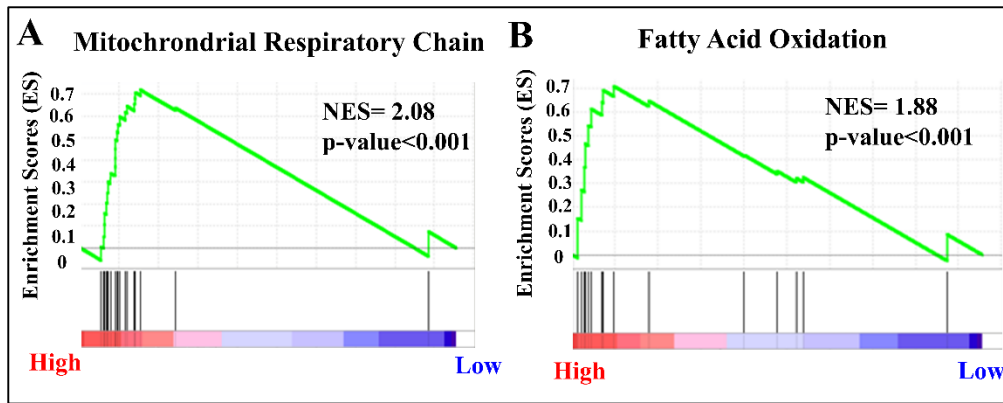
Figure 22: Gene set enrichment analysis showed that genes up-regulated in BA compared to WA at day 6 were enriched in (A) Mitochondrial respiratory chain and (B) Fatty acid oxidation.

Although most DMPs (Table 5) were not associated with gene expression changes (Figure 23), there was an increase in proportion of genes exhibiting the inverse relationship between these two variables along cell differentiation (49.1% at day 0, 60.3% at day 4, 61.5% at day 6).
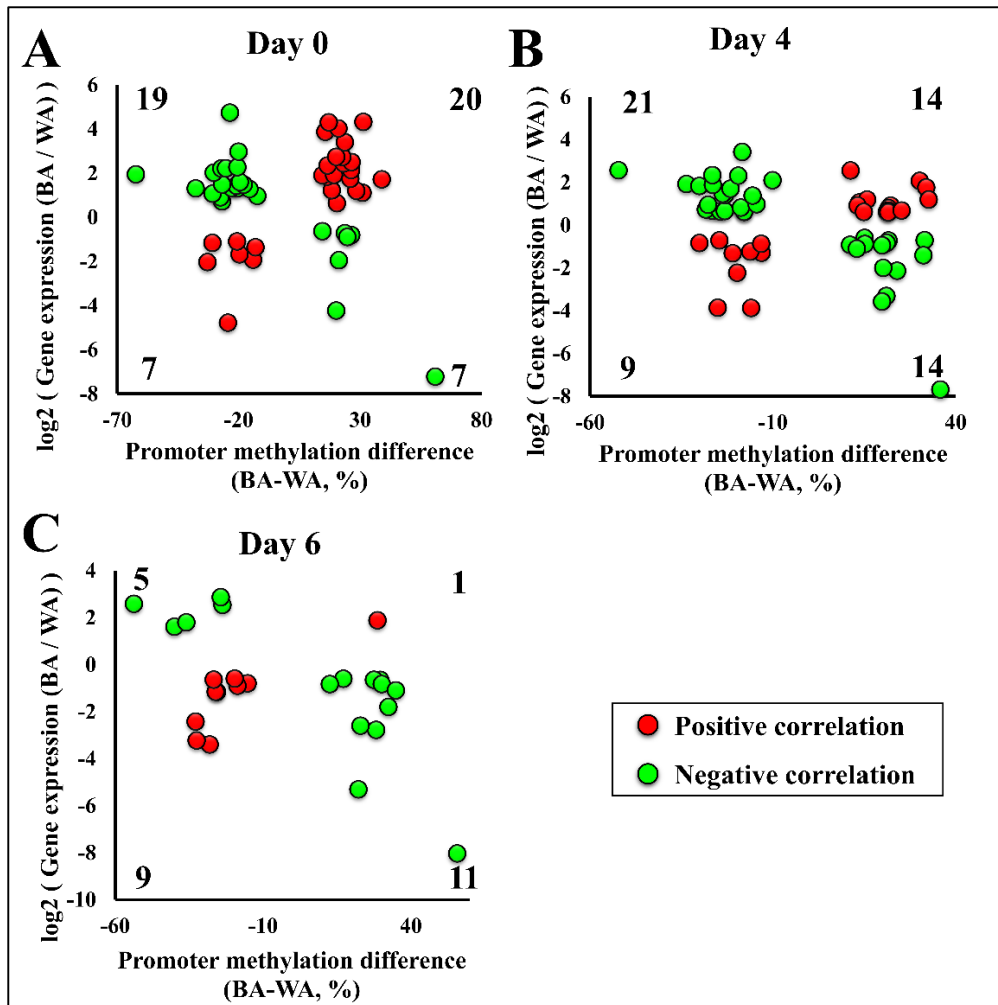
Figure 23: Scatterplot of gene expression changes (fold change $\geq$ 1.5) against promoter methylation differences for DMPs at (A) day 0, (B) day 4 and (C) day 6. Most of the genes showed anti-correlation between promoter methylation change and gene expression change.

4.3.9. Multiple *Hox* genes were identified in DMPs between cell types for which methylation difference was maintained throughout cell differentiation

To test if DNA methylation is important for tracing cell lineage, I intersected the lists of DMPs identified between WA and BA at three three analysed time points. (Figure 24). A total of 31 promoters (23 with BA<WA, 8 with BA > WA) were found to be significantly differentially methylated consistently at all three time points of differentiation (Figure 24A-B). The common set of 23 promoters having

68

hypomethylation in WA represented one-third of the 67 genes identified at day 0. This substantial overlap suggested that genes which showed potential ability in discriminating cell types were epigenetically controlled at the precursor stage and had their marks maintained throughout adipogenesis.
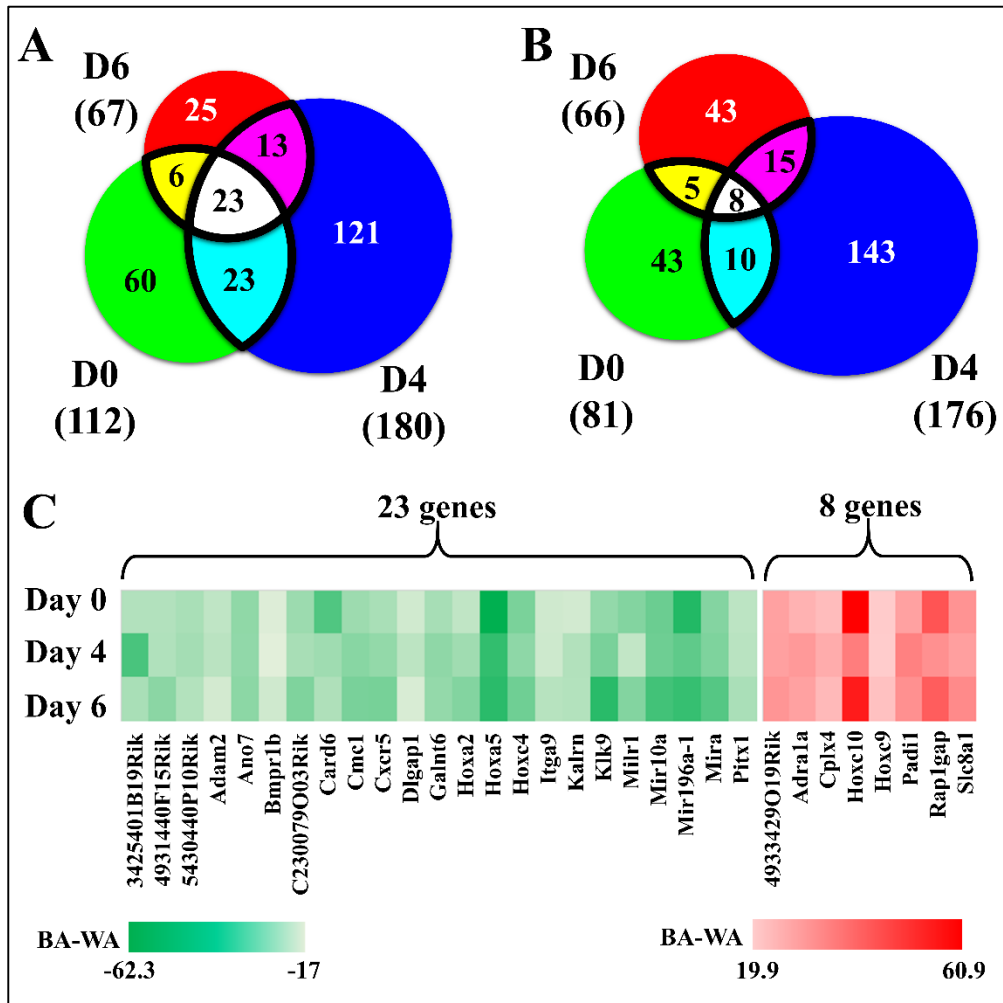
Figure 24: Overlaps of DMPs between BA and WA at days 0, 4 and 6. Promoters were split into (A) BA > WA and (B) BA < WA. (C) Heatmap representation of methylation differences for gene promoters (between BA and WA) which were either consistently hypo or hypermethylated from day 0 to 6 of cell differentiation.

Next, I extracted the identities of these 31 genes and represented the methylation difference (between BA and WA) in *Figure 24C*. Notably, five members from the homeotic (*Hox*) gene family (*Hoxa2*, *Hoxa5*, *Hoxc4*, *Hoxc9* and *Hoxc10*) appeared in the list, of which *Hoxc9* is a white adipocyte marker [156]. Intrigued by the

observation, the gene expression changes for these genes were extracted from RNA-seq results and further validated with q-PCR (*Figure 25*). Except for *Hoxc4*, four genes (*Hoxa2*, *Hoxa5*, *Hoxc9* and *Hoxc10*) showed consistent anti-correlation between DNA methylation and gene expression changes across the three time points of differentiation.
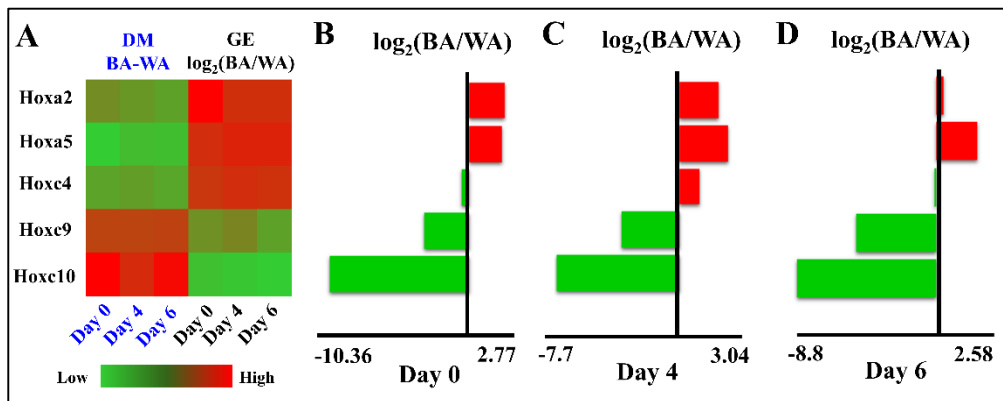
Figure 25: (A) Heatmap of promoter DNA methylation difference and gene expression changes for five *Hox* genes. (B-D) q-PCR results for same set of five Hox genes.

Products of the *Hox* genes are transcription factors which bind to DNA enhancers via homeodomain to either activate or suppress gene expression. Among many essential roles of the *Hox* genes include proper embryo development and control of cell death and cell proliferation [157, 158]. Recent study by Benton et al. [159] compared methylation changes before and after gastric bypass and weight loss human subcutaneous adipose and have identified *Hox* genes from multiple *Hox* clusters having significant methylation change.

4.3.10. General hypermethylation during adipogenesis

PCA analyses performed on white and brown adipogenesis separately revealed a distinctive sequential separation of samples by adipogenic progression just by PC1 (Figure 26A-B).

As I had analysed populations of cultured cells, the methylation status of each CpG was reflected on a continuous scale, ranging from 0 to 100%. Thus, I categorised these CpGs into lowly methylated (LM, <30%), partially methylated (PM, 30-70%) and highly methylated (HM, >70%) groups. In both white and brown adipogenesis, there was uniform decrease in proportions of CpGs in LM group, accompanied by a consistent increase in the proportions in both PM and HM groups (Figure 26C-D). These implied a shift of methylation status from LM to PM and HM as cells differentiate.
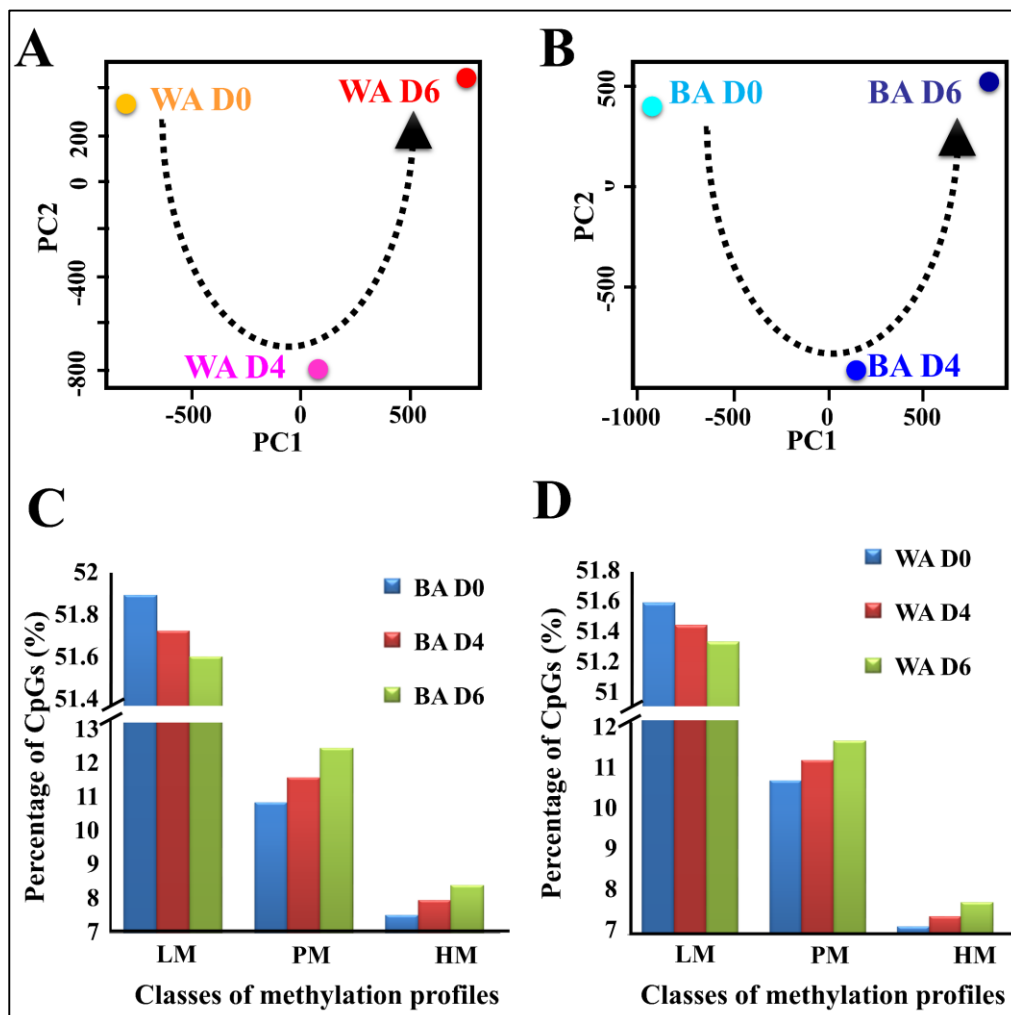
Figure 26: PCA analyses of (A) BA and (B) WA at day 0, 4 and 6. (C, D) CpGs were grouped by methylation levels into LM (<30%), PM (30-70%) and HM (>70%). Clear decreasing trends for LM and increasing trends for PM and HM were observed in both (C) brown and (D) white adipogenesis.

Next, I obtained the methylation difference for each CpG between time points and tabulated the frequencies using varying cutoffs of 10%, 20% and 30%. Concordant trends were observed for both adipogenesis processes for which the frequencies of CpGs showing hypermethylation at a later time point were consistently found in more abundance than hypomethylation (Figure 27).
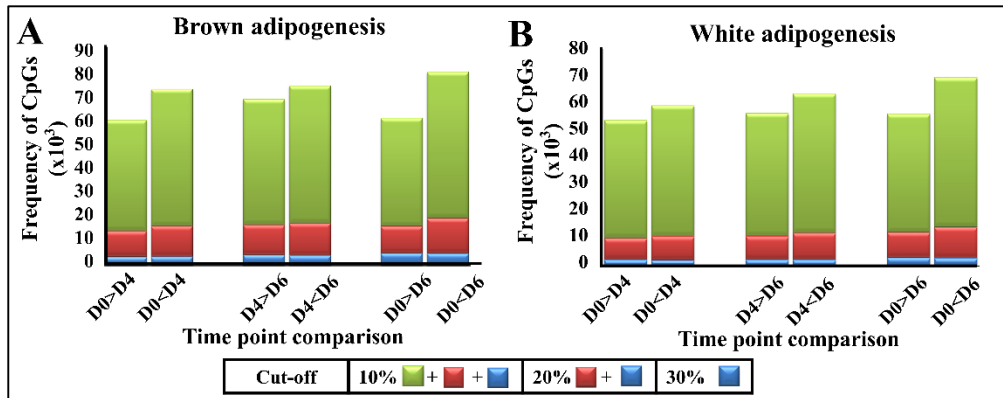
Figure 27: Frequencies of hypermethylated and hypomethylated CpGs between comparative time points, using varying cutoffs at 10, 20 and 30%.

4.3.11. Hypermethylated DMCs were predominant in promoter regions

To test if there is any systematic trend in methylation change during adipogenesis, I performed statistical test to identify DMCs. To do so, three pairwise comparisons were made for each cell type, (i) Day 0 vs Day 4, (ii) Day 4 vs Day 6, and (iii) Day 0 vs Day 6. Overall, I obtained more CpGs having an elevated DNA methylation across adipogenesis (Figure 28).
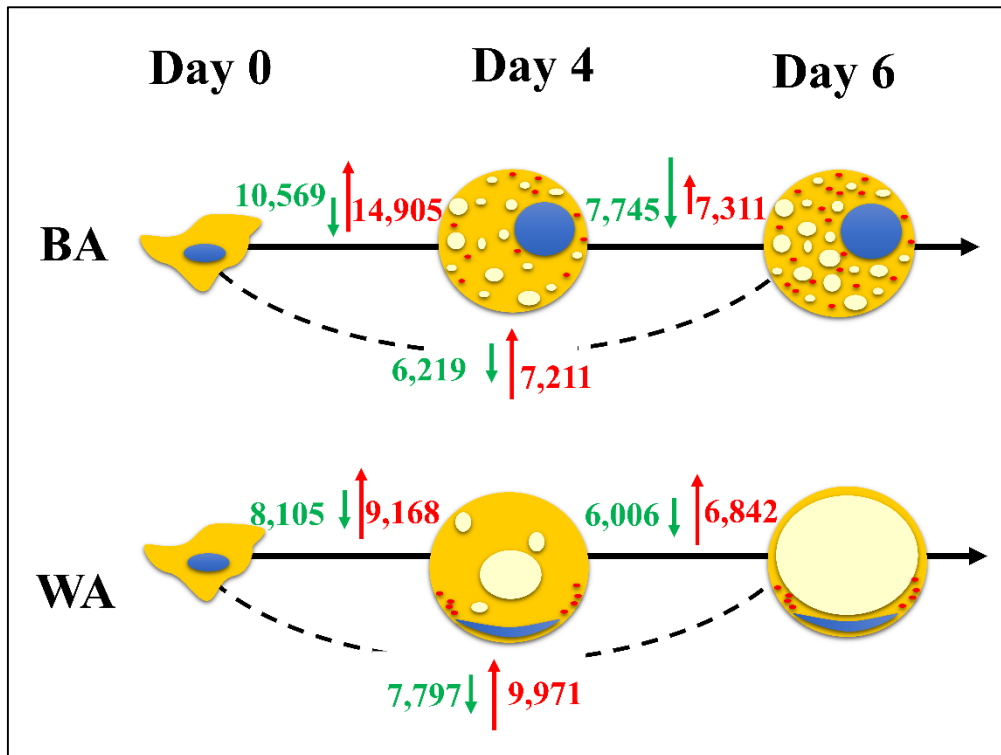
Figure 28: Frequency of significant CpGs by comparing stages of cell differentiation. Green arrows: hypomethylation. Red arrows: hypermethylation. Comparisons were made with respect to the earlier time point.

Next, I tested if these DMCs were enriched in any specific genomic locations. Consistent in all six comparisons, about 7% of the DMCs were located in promoters, 16-17% in exons, 36-38% in introns and 39-40% in intergenic regions. Notably, although a third of covered CpG are within promoter regions, only 7% of identified DMCs from promoter regions. Furthermore, dominance of hypermethylation was most pronounced in promoters, followed by exons, introns and finally intergenic regions (Figure 29). This is contrastingly different to cell type differences for which a dominance of hypermethylation was observed in non-promoter regions.

These observations suggested that during differentiation, promoter methylation is tightly regulated, with a strong preference towards DNA hypermethylation.
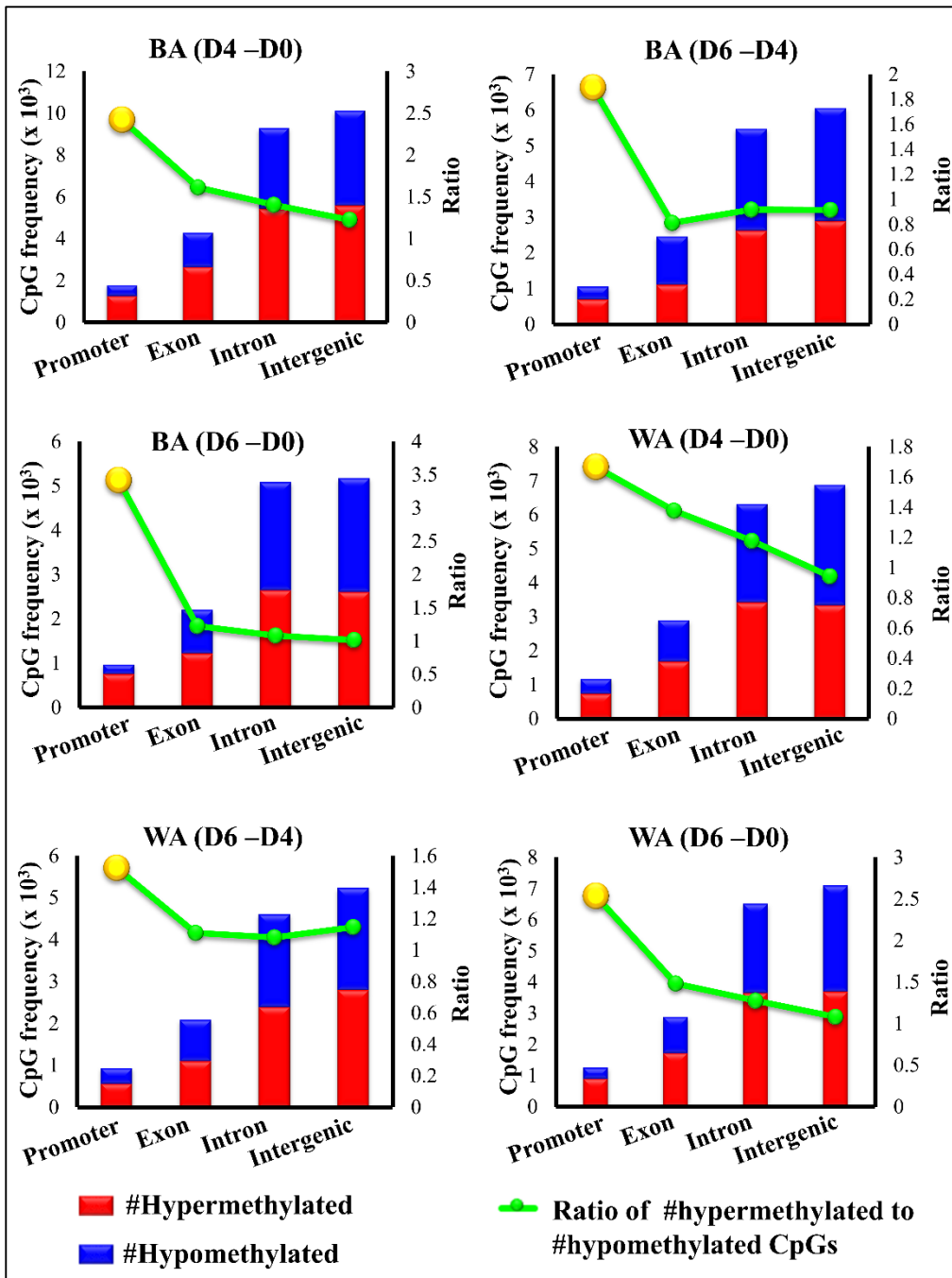
Figure 29: Proportions of significant differentially methylated CpGs in each genomic category. Ratios of hyper- to hypomethylated CpGs was highest in promoters for all comparisons.

To gain further insights on the dynamics of methylation change for these DMCs, I identified CpGs displaying significant methylation changes (i) from day 0 to 4, (ii) from day 4 to 6 or both. Most of the identified DMCs showed significant changes from Day 0 to 4, followed by a maintenance of the methylation marks thereafter (blue portion of the venn diagram in Figure 30).

Although the frequency of DMCs showing dynamic changes throughout adipogenesis was higher in brown adipocytes (3,691 CpGs) than white adipocytes (2,531 CpGs),(indicated by yellow in venn diagram, Figure 30), the relative proportions of these CpGs in various genomic regions were similar in both fat cells (~5% promoters, ~17% exons, ~36% introns, ~41% intergenic). Interestingly, of these DMCs, only a mere 212 (5.25%) and 63 (2.5%) CpGs for brown and white adipogenesis respectively had either continuous hyper or hypomethylation from Day 0 to Day 4, and then to Day 6. Vast majority of these CpGs showed a transient hyper or hypomethylation at Day 4, i.e. a hypermethylation from Day 0 to Day 4 is followed by a hypomethylation from Day 4 to Day 6 (or vice versa).
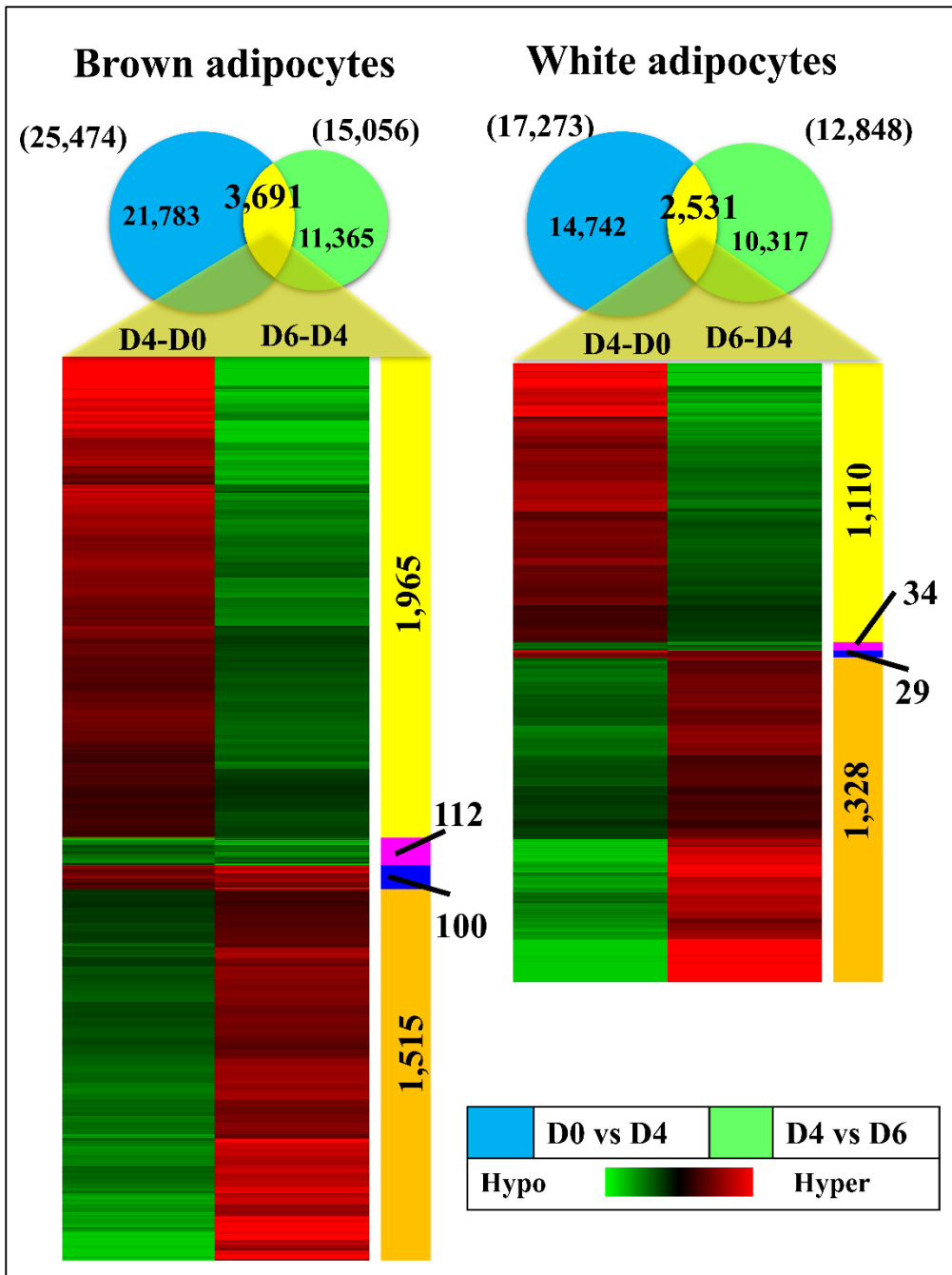
Figure 30: DMCs having significant methylation changes throughout adipogenesis. More than 90% of the DMCs did not show maintenance of methylation differences from day 0 to 6.

4.3.12. DMPs in adipogenesis were enriched for functions related to cell proliferation and cell differentiation

Similar to DMCs analysis, there was a dominance of hypermethylated events at a later time point (Table 8 and Table 9) in both brown and white adipogenesis.

Interestingly, while there was about twice the number of DMPs found in brown adipogenesis between days 0 to 4, comparison made between days 0 and 6 revealed more DMPs during white adipogenesis. This suggested that methylation changes in white adipogenesis were gradual and accumulative whereas brown adipogenesis which showed greatest differences from day 0 to 4.

Promoter RNA classes	Brown adipocytes		White adipocytes	
	D0 > D4	D0 < D4	D0 > D4	D0 < D4
mRNA	32	144	13	67
miRNA	3	9	2	5
ncRNA	5	17	5	10
Total	40	170	20	82

Table 8: Frequencies of DMPs between days 0 to 4 for brown and white adipogenesis and their respective RNA classes.

Promoter RNA classes	Brown adipocytes		White adipocytes	
	D4 > D6	D4 < D6	D4 > D6	D4 < D6
mRNA	12	47	26	42
miRNA	3	4	2	3
ncRNA	4	6	1	2
Total	19	57	29	47

Table 9: Frequencies of DMPs between days 4 to 6 for brown and white adipogenesis and their respective RNA classes.

Promoter RNA classes	Brown adipocytes		White adipocytes	
	D0 > D6	D0 < D6	D0 > D6	D0 < D6
mRNA	10	67	27	98
miRNA	2	3	2	2
ncRNA	3	7	4	17
Total	15	77	33	117

Table 10: Frequencies of DMPs between days 0 to 6 for brown and white adipogenesis and their respective RNA classes.

Independent functional enrichment analyses on DMPs of brown and white adipogenesis (day 0 vs day 4) showed significant enrichments in cell proliferation and cell differentiation (Figure 31A-B). There were significant overlaps of DMPs between white and brown adipogenesis from day 0 to day 4 (19 genes, hypergeometric test p<3.586×10^{-18}, Figure 31C,Table 11,), day 4 to 6 (5 genes, hypergeometric test p<7.441×10^{-6}, Figure 31B, Table 12) and from day 0 to day 6

(22 genes, hypergeometric test p<5.231 x 10^{-27}, Figure 31E, Table 13). These three

lists included a number of genes related to cell proliferation and cell differentiation

(Table 11, Table 12, Table 13). Noteworthy, *Hoxa1* and its antisense counterpart,

Hotairm1 have been identified to show significant hypermethylation at day 6 relative

to day 0 during both brown and white cell differentiation.
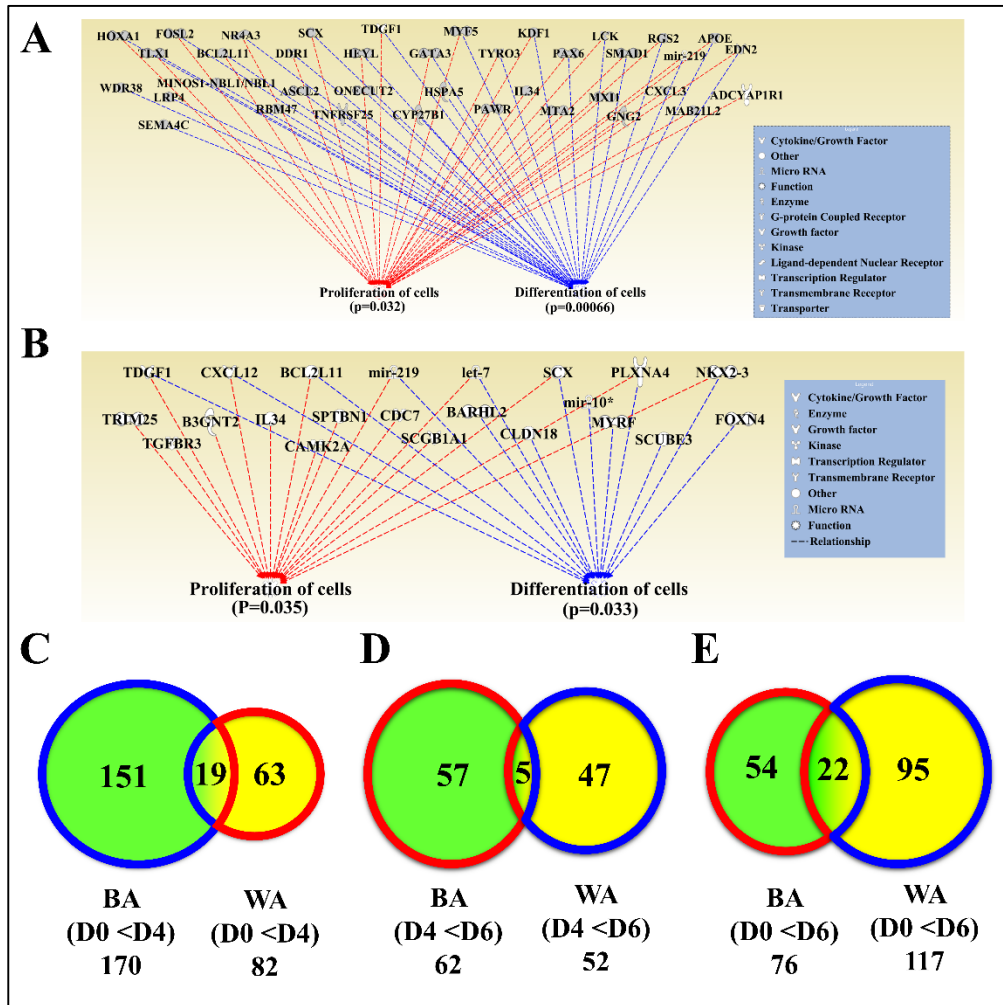
Figure 31: Hypermethylated DMPs of (A) BA and (B) WA from day 0 to 4 showed enriched functions in cell differentiation and cell proliferation. Overlaps of hypermethylated DMPs between BA and WA for comparison of (C) day 0 vs day 4, (D) day 4 vs day 6 and (E) day 0 vs day 6.

Gene	RNA category	Brown adipogenesis				White adipogenesis			
		#DMC	D0 (DMC)	D4 (DMC)	D4-D0 (DMC)	#DMC	D0 (DMC)	D4 (DMC)	D4-D0 (DMC)
Akap5	mRNA	8	14.07	34.66	20.59	7	6.61	25.65	19.04
Bcl2l11*	mRNA	3	46.11	69.27	23.16	2	4.61	17.98	13.37
Cldn3	mRNA	2	0.00	17.55	17.55	2	1.57	17.79	16.21
Cxcl3	mRNA	2	17.06	51.66	34.61	2	17.06	39.10	22.04
Cyba	mRNA	3	3.45	21.69	17.73	2	2.22	17.92	15.71
Derl3	mRNA	3	16.11	53.47	37.36	5	18.45	48.54	30.09
Des*	mRNA	2	18.51	48.23	16.42	3	5.17	22.12	16.95
Dio3os	ncRNA	3	22.18	45.52	23.34	2	18.13	43.38	25.25
Foxr1	mRNA	2	70.61	92.35	21.74	3	78.69	95.50	16.81
Gm12992	ncRNA	4	15.23	57.77	42.54	2	23.64	58.33	34.70
Gm16793	ncRNA	4	56.75	82.03	25.28	2	21.51	45.91	24.40
Il34*	mRNA	2	23.38	60.17	36.78	3	36.36	59.09	28.76
Inafm1	mRNA	3	11.72	32.97	21.24	2	24.36	52.20	27.84
Mir219a-2*	miRNA	3	60.45	82.37	21.91	2	64.87	84.04	19.18
Podxl2	mRNA	6	12.73	34.44	21.72	2	2.63	13.67	11.04
Scx*	mRNA	7	11.20	34.11	22.91	3	14.49	31.09	16.61
Slc5a5	mRNA	22	9.53	32.65	23.13	16	8.52	32.30	23.78
Syce2	mRNA	2	20.37	46.81	26.44	2	0.00	23.68	23.68
Tdgf1*	mRNA	7	52.18	79.79	27.61	2	33.75	59.40	25.65

Table 11: Methylation information for 19 promoters which were hypermethylated at day 4 wither respect to day 0 for both BA and WA. *Genes involved in cell proliferation or cell differentiation.

Gene	RNA category	Brown adipogenesis				White adipogenesis			
		#DMC	D4 (DMC)	D6 (DMC)	D6-D4 (DMC)	#DMC	D4 (DMC)	D6 (DMC)	D6-D4 (DMC)
Des*	mRNA	3	10.38	39.29	28.90	4	5.67	26.09	20.42
Madcam1*	mRNA	2	51.21	83.72	32.52	3	48.13	70.15	22.02
Npas4	mRNA	2	14.47	61.11	46.64	2	23.34	53.21	29.98
Rn45s	ncRNA	2	14.67	31.53	16.86	2	44.40	68.08	23,68
Tmem171	mRNA	2	0.86	20	19.14	2	15.96	45.81	29.85

Table 12: Methylation information for 5 promoters which were hypermethylated at day 6 with respect to day 4 for both BA and WA. *Genes involved in cell differentiation.

Gene	RNA category	Brown adipogenesis				White adipogenesis			
		#DMC	D0 (DMC)	D6 (DMC)	D6-D0 (DMC)	#DMC	D0 (DMC)	D6 (DMC)	D6-D0 (DMC)
1700124L16Rik	nrRNA	2	1.05	17.65	16.59	2	4.65	21.54	16.89
Akap5	mRNA	6	13.45	47.73	34.28	4	9.81	34.08	24.27
B3gnt2	mRNA	3	8.01	37.04	29.02	2	8.06	36.36	28.30
Ccdc8	mRNA	8	1.11	15.39	14.29	8	0.42	16.07	15.65

Chad	mRNA	2	34.22	79.19	44.97	4	30.66	60.66	30.00
Cpt1b	mRNA	2	17.67	43.03	25.36	2	5.13	39.06	33.93
Derl3	mRNA	6	11.50	49.14	37.64	4	11.26	42.46	31.20
Des*	mRNA	6	7.43	34.63	27.20	3	0.51	16.51	16.00
Fgr	mRNA	4	49.60	80.21	30.61	3	40.62	67.12	26.50
Gm128	mRNA	2	69.70	97.73	28.03	2	32.92	62.47	29.56
Gm12992	ncRNA	5	12.18	53.33	41.15	4	9.43	38.20	28.77
Gm16157	ncRNA	4	25.51	57.44	31.93	4	17.24	47.43	30.19
Hapln3	mRNA	2	2.91	23.33	20.43	2	3.97	21.43	17.46
Hotairm1	ncRNA	2	10.50	27.65	17.15	5	24.31	47.46	23.16
Hoxa1*	mRNA	2	10.50	27.65	17.15	5	24.31	47.46	23.16
Nbl1*	mRNA	6	12.20	46.88	34.67	3	15.68	39.06	23.38
Podxl2	mRNA	3	6.21	20.63	14.43	3	3.07	16.62	13.55
Rexo4	mRNA	2	11.74	32.78	21.04	2	26.40	58.89	32.49
Scx*	mRNA	4	6.94	27.21	20.26	2	11.54	36.84	25.30
Slc5a5	mRNA	14	10.35	39.74	29.39	12	8.36	37.77	29.40
Tex40	mRNA	2	13.50	32.02	18.51	2	19.64	50.79	31.15
Ttc16	mRNA	2	2.78	20.15	17.38	2	3.00	18.25	15.26

Table 13: Methylation information for 22 promoters which were hypermethylated at day 6 wither respect to day 0 for both BA and WA. *Genes involved in cell differentiation.

4.3.13. Correlating promoter DNA methylation with gene expression during adipogenesis

Next, I asked how promoter DNA methylation changes affects gene expression during adipocytes differentiation by integrating sample matched RNA-seq data to RRBS data. Comparing the transcriptome profiles of mature brown and white adipocytes with respect to their precursors (day 0 vs day 6), there was enrichment of "Mitochondrial Respiratory Chain" and "Lipid Metabolic Process" in up-regulated genes (Figure 32).
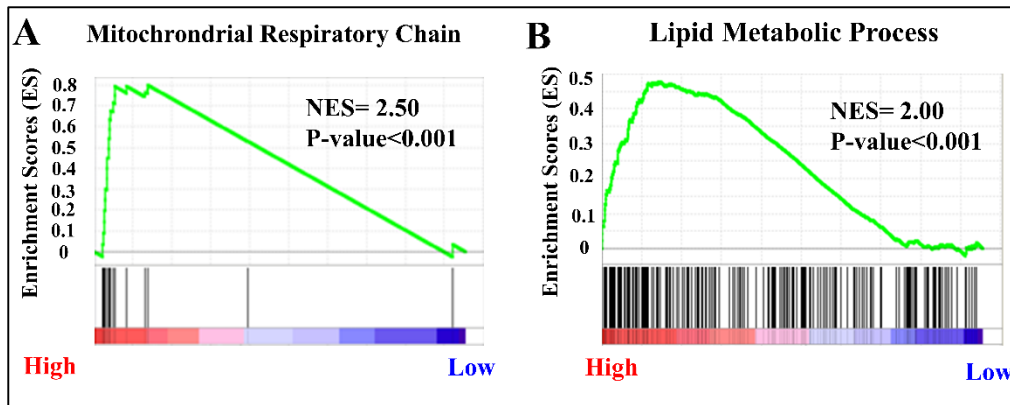
Figure 32: Gene set enrichment analysis on (A) up-regulated genes during brown adipogenesis and (B) white adipogenesis using RNA-seq data.

Among all correlations (Figure 33), expression changes during brown adipogenesis appears to be more correlated with DNA methylation changes than white adipogenesis, especially during earlier time points (day 0 versus day 4) (Figure 33A). In this category, as many as 48 out of 70 genes (68.5%) were anti-correlated between DNA methylation and gene expression changes, representing a statistically significant difference in frequencies between the positively and negatively correlations (2 sided binomial test, p=0.0095). Within this list, 16 genes (Table 14) with functional relevance to cell proliferation or cell differentiation were identified. Out of these, eleven of them showed anti-correlation between promoter methylation changes with gene expression changes.
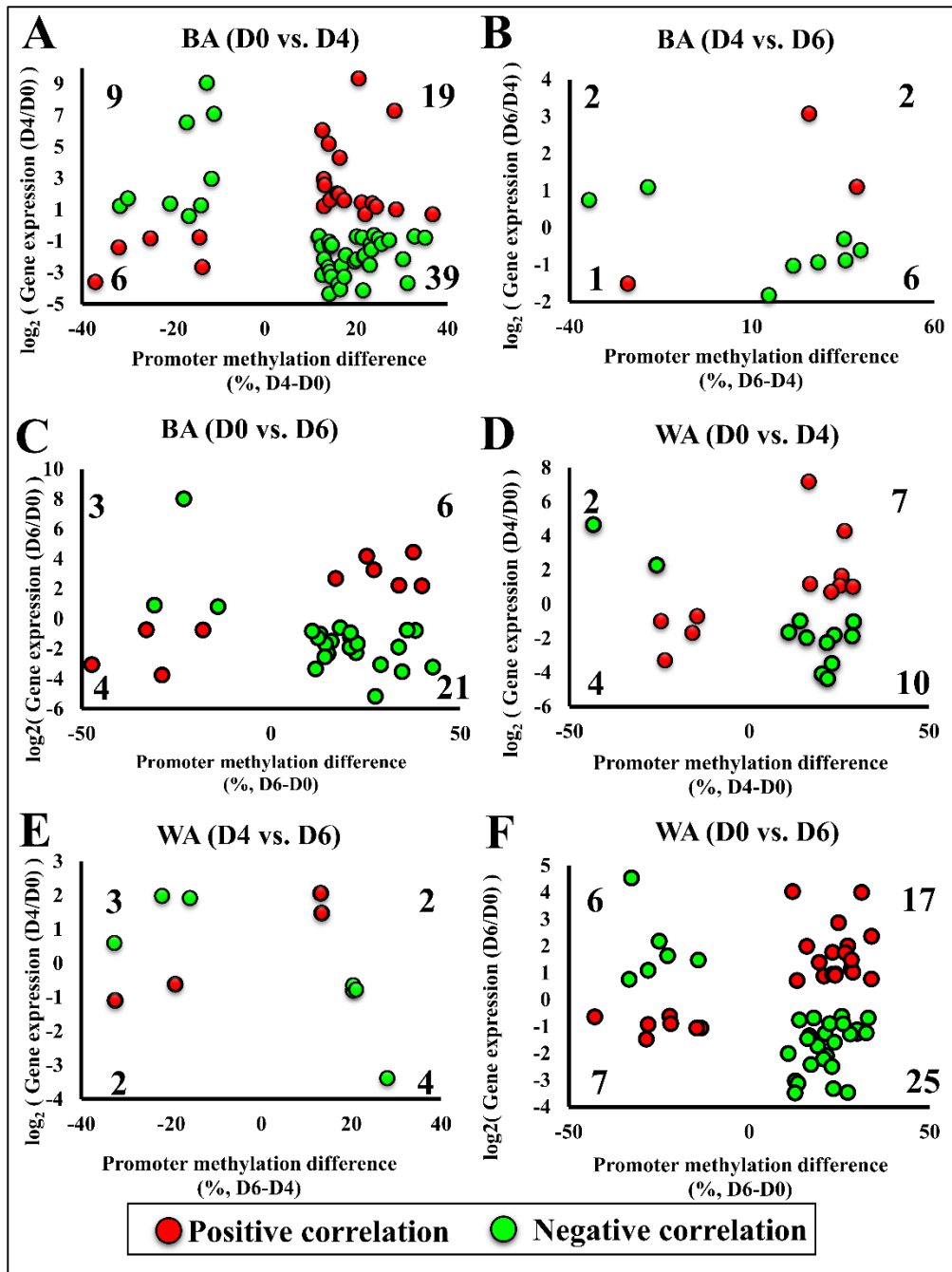
Figure 33: Scatterplot of gene expression changes against promoter methylation differences for brown adipogenesis (A) day0 vs day4 in BA, (B) day4 vs day6 in BA, (C) day0 vs day6 in BA, (D) day0 vs day6 in WA, (E) day4 vs day6 in WA, (F) day0 vs day6 in WA.

Genes	#DMCs	BA_D4-D0	RNA category	log_2(D4/D0)
Arhgef19	2	-37.07	mRNA	-3.61
Atp2c1*	3	-31.97	mRNA	-1.41
D130017N08Rik	3	-21.23	ncRNA	-0.75
Hoxc8*	2	-14.29	mRNA	-0.78
Nufip1	2	-25.00	mRNA	-0.84
Slc39a14	2	-13.64	mRNA	-2.66
8430408G22Rik	2	-11.08	mRNA	7.09

Genes	#DMCs	BA_D4-D0	RNA category	$\log_2$(D4/D0)
Aldh3b2	2	-17.06	mRNA	6.54
Gm2a	2	-31.68	mRNA	1.22
Homez	2	-29.97	mRNA	1.71
Letmd1	2	-11.63	mRNA	2.95
Mcts2	4	-20.70	mRNA	1.37
N6amt2	2	-16.58	mRNA	0.59
Psd	2	-13.95	mRNA	1.27
Tmem45b	3	-12.67	mRNA	9.06
6820431F20Rik	2	14.69	ncRNA	-1.25
Adcyap1r1	2	21.58	mRNA	-4.15
Aldh1a3	2	16.08	mRNA	-3.77
Apoe*	2	21.33	mRNA	-0.79
Bcl2l11*	3	23.16	mRNA	-1.21
C920025E04Rik	2	24.81	mRNA	-2.25
Camk1d	2	13.80	mRNA	-3.27
Ccdc8	4	13.00	mRNA	-2.15
Crocc	2	20.15	mRNA	-0.72
Cuedc1	2	27.23	mRNA	-0.96
Ddr1*	2	23.01	mRNA	-2.52
Dio3os	3	23.34	ncRNA	-1.56
Elmo2	2	35.14	mRNA	-0.79
Fads6	2	12.64	mRNA	-3.16
Fchsd1	2	14.32	mRNA	-1.33
Flrt2	4	14.16	mRNA	-4.37
Fosl2*	2	14.08	mRNA	-1.04
Galnt1	2	11.78	mRNA	-0.76
Ganab	2	23.97	mRNA	-0.63
Gng11	2	24.92	mRNA	-0.85
Htra3	4	14.47	mRNA	-1.29
Isyna1	2	13.92	mRNA	-2.67
Lck*	3	26.53	mRNA	-3.75
Lrp4*	3	16.55	mRNA	-4.07
Lvrn	2	20.14	mRNA	-2.17
Map6	2	14.63	mRNA	-3.29
Mmp23	2	17.39	mRNA	-3.29
Mtss1l	2	17.76	mRNA	-1.91
Nbl1*	3	31.29	mRNA	-3.69
Notum	2	11.86	mRNA	-0.69
Nrip2	2	25.57	mRNA	-1.18
Palm3	2	21.86	mRNA	-1.92
Pawr*	2	16.79	mRNA	-2.58
Pcnxl3	2	12.49	mRNA	-1.33
Podxl2	6	21.72	mRNA	-2.00

Genes	#DMCs	BA_D4-D0	RNA category	log₂(D4/D0)
Sema4c*	2	19.65	mRNA	-2.31
Smad1*	3	30.33	mRNA	-2.16
Ttll3	2	32.82	mRNA	-0.71
Tyro3*	3	14.27	mRNA	-2.96
Acvr1c	2	12.62	mRNA	6.04
Armt1	2	12.96	mRNA	1.22
Derl3	3	37.36	mRNA	1.49
Des	2	16.42	mRNA	4.29
Fcgrt	2	22.00	mRNA	0.70
Hotairm1	2	15.88	ncRNA	2.03
Hspa9	2	17.42	mRNA	1.58
Il34	2	36.78	mRNA	0.70
Inafm1	3	21.24	mRNA	1.45
Lrrc14b	2	24.42	mRNA	1.18
Mxi1*	2	13.15	mRNA	2.55
Nat8l	2	28.43	mRNA	7.29
Rbm47*	3	28.79	mRNA	1.01
Rgs2*	2	14.02	mRNA	5.18
Rmnd1	2	12.96	mRNA	2.93
Samm50	2	16.25	mRNA	1.96
Slc25a34	2	20.60	mRNA	9.34
Tmem229b	2	14.29	mRNA	1.60
Uba52	2	23.59	mRNA	1.39

Table 14: List of DMPs between day 0 and 4 of brown adipocytes with a gene expression fold change of at least 1.5. *Genes which are relevant to cell differentiation/ cell proliferation.

4.3.14. 5-Azacytidine treated BA and WA induced gene expressions in Hoxc9 and Hoxc10 for both BA and WA

Intrigued by the correlation between promoter methylation changes and gene expression changes (Figure 25), I investigated if the gene expression of these gene were indeed an effect of DNA methylation alterations. To test the hypothesis, both brown and white pre-adipocytes were treated with varying concentrations (5μm and 10μm) of DNA methylation inhibitor, 5-Azacytidine. As shown in Figure 34, both *Hoxc9* and *Hoxc10* were up-regulated in both BA and WA at day 5 when DNA

methylation is inhibited by drug treatment. These results implied that both *Hoxc9* and

Hoxc10 gene expression may be regulated by promoter methylation manipulation.
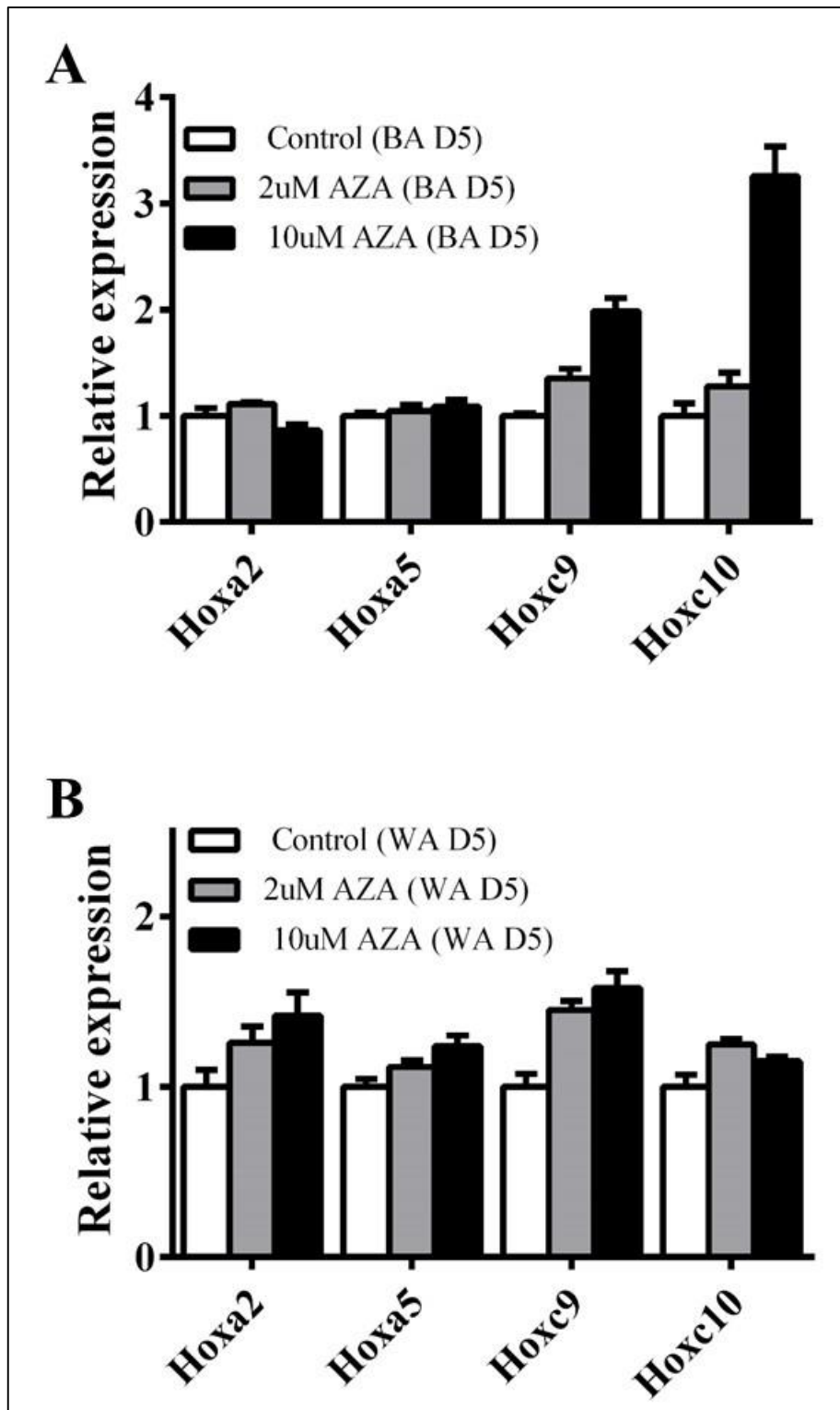
Figure 34: q-PCR results of five *Hox* genes being subjected to 2μm and 10μm of 5-Azacitidine treatment for (A) BA and (B) WA at day 5.

4.4.   Discussion

Comprehensive understanding of molecular mechanisms which are involved in (i) adipogenesis regulation and (ii) guiding differentiation of mesenchymal stem cells into distinct fat cells are necessary to identify optimal therapeutic targets for obesity. DNA methylation is one of the most well-studied epigenetic mechanisms and have been shown to be important in cell differentiation [152, 160] and defining cell lineages. Due to its versatile capacity, DNA methylation presents itself as an ideal target mechanism for this study. Here, I present the first comprehensive genome-wide DNA methylation landscape for both white and brown adipogenesis, with the aim of understanding how DNA methylation could affect both fat cells differentiation and cell fate commitment.

During adipogenesis, precursors loses its pluripotency and becomes specialised in function and morphology. Overall, DNA methylation increases during both white and brown adipogenesis. Significantly differentially methylation promoters from both adipogenesis were enriched for functional categories related to cell differentiation and cell proliferation. Interestingly, when comparing the ratios of hyper- and hypomethylated DMCs based on cell specificity (white vs. brown) and time series (at different time points during differentiation), we observed:

i.   Steepest ratio in promoter regions for time series,

ii.   Steepest ratio in non-promoter regions when comparing cell types.

Such opposite trends were intriguing and suggested that genomic context dependent DNA methylation changes may perform different biological functions.

I have identified a set of 31 promoters which showed consistent hyper- or hypomethylation between WA and BA for all time points of analyses (Figure 24).

86

Out of these 31 genes, five Hox genes (*Hoxa2*, *Hoxa5*, *Hoxc4*, *Hoxc9*, *Hoxc10*) were included which four of them (*Hoxa2*, *Hoxa5*, *Hoxc9*, *Hoxc10*) all showed negative correlation with gene expression changes (RNA-seq and q-PCR). *Hox* gene members are a super family of transcription factors [158] which are highly conservative in sequence and functions. Not only are they well known for their roles in normal in embryonic development [158], *Hox* genes have also gained much recognition in the adipogenic field such as displaying cell specific expression and involvement in metabolic diseases such as diabetes [158, 161]. My data suggested that *Hox* genes are important in defining cell lineage, which in turn may be regulated by DNA methylation.

A few issues were yet addressed in this study and should be considered for future work. Firstly, the RRBS approach to analysing DNA methylation covers only about 3% of CpGs in the genome and 50% of the promoters. Though being data-to-cost effective, biased selection of CpG rich regions might not fully represent true genomic features. Secondly, simple bisulfite sequencing does not allow discrimination between 5mC and 5hmC. Although the distinct functions of 5hmC have yet to be defined, this modified base was found in stem cells and brain, especially on the gene bodies within the genome [162-164]. Enrichment of 5hmC in the gene bodies have been suggested to be involved in facilitating transcription in olfactory sensory neurons [165]. Thus, to be able to distinguish among C, 5mC and 5hmC, single-molecule real-time (SMRT) sequencing could be enlisted [166].

Thirdly, so far analyses have been conducted on mouse adipogenic cell culture models, whose results might not be extrapolated to *in vivo* model or to human. Thus, future work could be targeted at perturbing DNA methylation to test for phenotypic variations in mouse models and using human cell models.

4.5.   Summary

In this study, we provided the first DNA methylome profiles for brown and white adipogenic differentiations. We aimed to understand the impacts of DNA methylation on adipogenesis and cell type specification. To do that, comprehensive comparisons were made

    i.   Across white adipogenesis

    ii.   Across brown adipogenesis

    iii.   Inter cell types (BA, BWA and WA) at each analysed time points.

We also did RNA-seq profiling and correlated it with the DNA methylation data. We made the following key findings:

    i.   DNA methylation was similar between WA and BWA

    ii.   Greatest methylation difference existed between WA and BA

    iii.   Five *Hox* genes promoters (*Hoxa2*, *Hoxa5*, *Hoxc4*, *Hoxc9*, and *Hoxc10*) were found to be uniformly hypo or hypermethylated in the same direction between WA and BA for days 0, 4 and 6. With the exception of Hoxc4, the rest showed anti-correlation with gene expression changes.

    iv.   General hypermethylation was observed with progression of cell differentiation in both WA and BA. DMPs identified showed enriched functions for cell proliferation and cell differentiation.

    v.   Predominance of hypermethylated DMCs were located in promoters for adipogenesis analyses while a predominance of hypomethylated DMCs (with respect to BA) were found in non-promoter regions.

CHAPTER 5 – A COMPLEX ASSOCIATION BETWEEN DNA METHYLATION AND GENE EXPRESSION IN HUMAN PLACENTA AT FIRST AND THIRD TRIMESTERS

5.1.   Background and Hypothesis

The human placenta is a temporary maternal-foetal organ essential for normal foetal development. It serves a number of functions such as exchange of oxygen, nutrients and waste products between the mother and fetus. The dysfunction of the placenta usually leads to dire consequences such as recurrent pregnancy loss, preterm birth, pre-eclampsia and intrauterine growth restrictions (IUGR). Conditions of preterm and IUGR often leads to low birth weight, which is associated with diseases such as hypertension and type 2 diabetes that might show up later in life [167-173]. During pregnancy, the human placenta undergoes tremendous growth in size, morphology and structure to cope with the development of the fetus [174-176].

Not surprisingly, extensive molecular changes occur during placenta development. A number of studies have investigated gene expression profiles at different structural locations of the placenta [177], and at different gestational ages of the placenta, with gene expression changes often correlating with functional changes at different gestational ages. However, the molecular mechanisms underlying such drastic gene expression changes remain to be elucidated.

Epigenetics is considered as a fundamental mechanism regulating gene expression during development. The placenta has long been a favourite organ for the study of epigenetics, particularly in genomic imprinting [178-183]. Epigenetics is also widely

considered as a mechanism for environmental factors to impact on development. For this reason, studying the epigenetics of the human placenta is particularly interesting as the placenta serves as the portal for the foetus to experience the external environment. Recently, a number of groups have investigated the DNA methylation changes of the placenta at different gestational ages [184] and due to foetal abnormalities [133].

In this study, I systematically analysed the transcriptomes and the DNA methylomes of human placenta samples derived from different gestational ages. Furthermore, we studied the dynamic correlations between gene expression and DNA methylation at different gestational ages.

5.2. Material and methods

5.2.1. Samples

Women with euploidy pregnancies who attended KK Women and Children's Hospital were recruited in this study. Chorionic villus samples from subjects at the first or early second trimesters of pregnancy were collected by chronic villus sampling (CVS). Placenta villi samples (foetal side) were collected from third trimester of pregnancy after delivery.

5.2.2. DNA methylation

Six first trimester and five third trimester samples were included in the study.

Sample preparation and computational processing followed the steps described in Chapter 3, Section 3.2.2 and Section 3.2.3 respectively.

5.2.3.   Differential DNA methylation analysis

Differential methylation analysis was performed at regional levels to identify differentially methylated promoters (defined as 1kb upstream and 500 bp downstream from transcription start site) and gene bodies. Autosome CpGs (sequencing depth $\geq$10) covered in at least 3 early and 3 late gestational samples were used for all subsequent analyses.

A 2-sided Mann Whitney U test was first performed at single CpG level and p values were adjusted within regions using the Benjamini Hochberg method. A promoter was considered significantly differentially methylated if 1) methylation difference between average first trimester and third trimester samples was at least 10% and 2) contained at least 2 CpGs with FDR corrected $p < 0.05$. For gene bodies, in additional to the above two criteria, we require that all significantly differential methylated fragments mapped to the gene body be regulated in the same direction (either all hypermethylated or hypomethylated).

All statistical analyses were performed using R package.

5.2.4.   RNA-seq

Five first and second trimester samples from women carrying normal foetuses were included in this study.

Gene expression was quantified using RNA-seq, details found in [133]. Five RNA samples from normal pregnancies and four samples from pregnancies carrying DS were included in the study. Briefly, 2–5 µg of total RNA was used for each library preparation. Each RNA sample was treated with DNase I, subsequently subjected to messenger RNA purification and fragmentation, complementary DNA synthesis, end-repair, 3′-end-adenylation and adapter-ligation. Adapter-ligated cDNA fragments were size-selected using a 3% agarose gel (200±25 bp). The DNA samples were then amplified by PCR and purified using 3% agarose gels and further quantified.

The single end reads from RNA-seq was then analysed using Illumina RNA-Seq pipeline, CASAVA software version 1.7. Alignment of the reads was done using the default parameters and performed step-wise on three references, (i) contaminant reference made up of mitochondrial DNA (chrM), (ii) genome assembly of the species of interest, and (iii) splice junction set created from the exon information of annotated genes. All the reference sequences were downloaded from UCSC website (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/).

5.2.5.   Differential RNA-seq analysis

RPKM values generated by running the Illumina RNA-seq pipeline represented gene expression. Average RPKM values for each gene in each sample group (early and late gestational age groups) were calculated. When the average RPKM for a gene is less than 0.5, the value was set as 0.5. A gene was considered to be differentially expressed between first and third samples when: 1) two-sided Mann Whitney U test p-value < 0.05; and 2) the ratio of gene expression between compared groups ≥ 2 or $\leq$ 0.5. R package was used for all statistical analyses.

5.3.   Results

5.3.1.   Descriptive statistics for RRBS quality

Using an improved version of reduced representation bisulfite sequencing (RRBS) [133, 136], we quantified DNA methylation of six first trimester and five third trimester placenta villi samples. There was an average of 37.5 million pass filter reads for the 11 samples, with an average alignment rate (C2T and G2A) of 58.5%. A high bisulfite conversion rate of above 99% was achieved for all the samples (Figure 35A). Using a minimum sequencing depth of 10 as the cut off, we obtained on average 1.8 million CpGs per sample (Figure 35B).
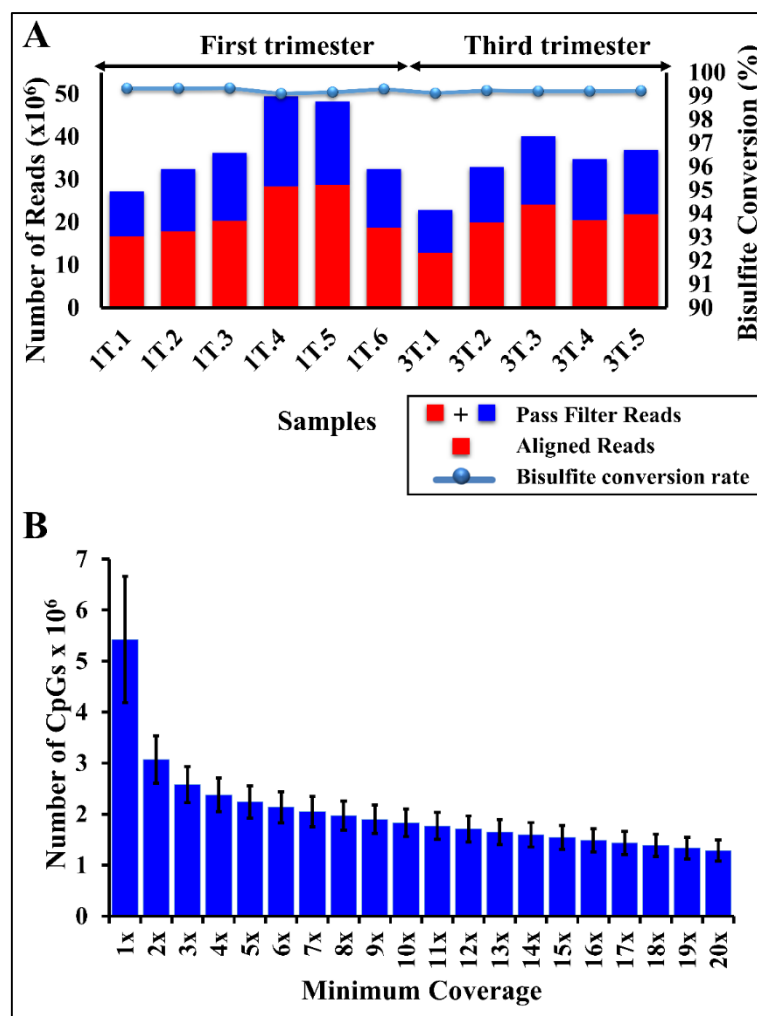
Figure 35: Descriptive statistics of the 11 samples. (A) For each sample, red bars represent the number of uniquely aligned reads, while blue bars are the non-uniquely or unaligned reads. The blue line gives the bisulfite rate for each sample. (B) The histogram gives the average

number of CpGs sites for 11 samples, with varying minimum sequencing depths. Error bars represents standard deviation for 11 samples.

5.3.2. Genomic coverage of the analysed CpGs

To facilitate cross gestation comparison, I further removed CpG sites that were on the sex chromosomes or were present in less than three samples in either the first or the third trimester group, resulting in 1.7 million CpG sites for further analysis. These CpGs represented about 3% of Hg19 autosomal CpGs, 78% of CGIs, 70.8% of core promoters (defined as -1kb upstream and +500bp downstream from a transcription start site) and 64.2% of gene bodies (defined as +1kb downstream from a transcription start site to transcription termination site) (*Figure 36*).
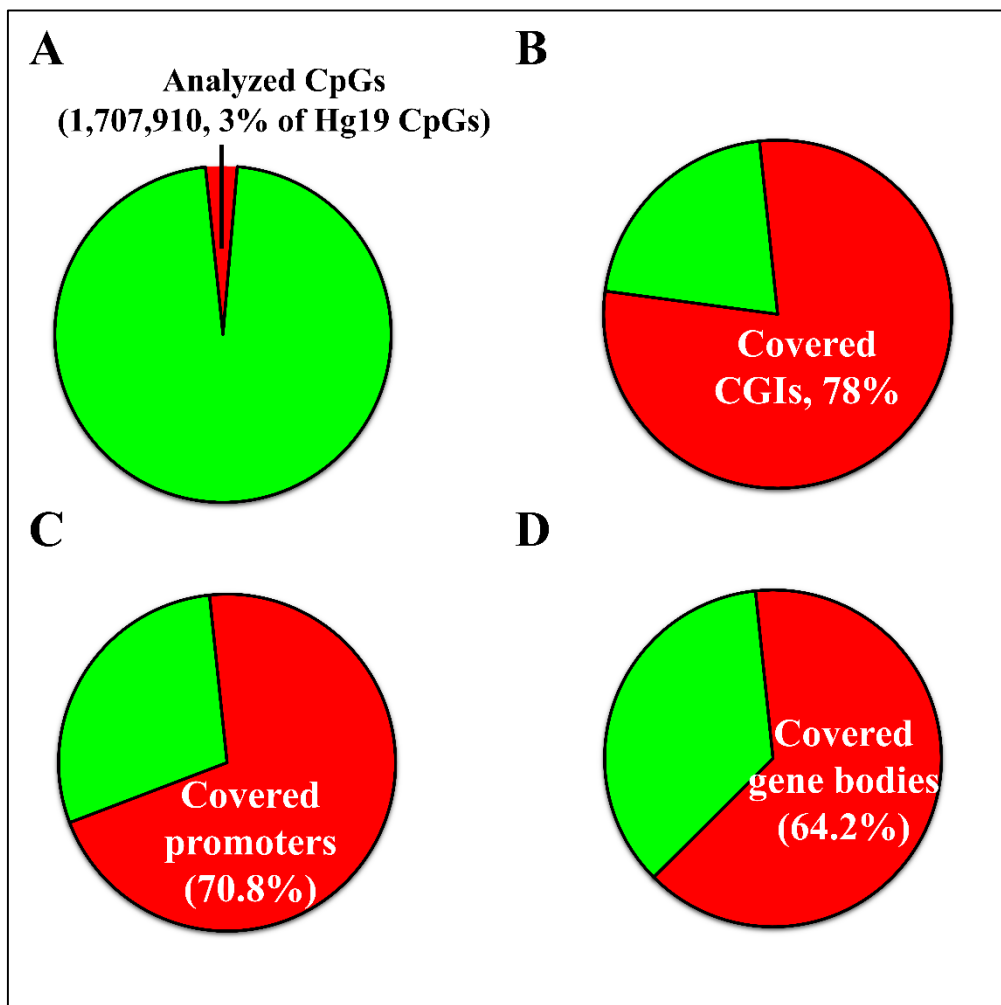
Figure 36: Coverage of CpGs for RRBS. (A) A total of 1,707,910 CpGs were covered in at least 3 first and 3 third trimester samples, at a sequencing depth ≥10. This covers 3% of all

94

CpGs on Hg19 (autosomes). (B) 78% of CGIs, (C) 70.8% of promoters and (D) 64.2% of gene bodies. All regions required at least 3 covered CpGs.

5.3.3. Global methylation profiles were different depending on CGI status and between placenta at different gestational ages

The distribution of individual CpG methylation levels were drastically different for CpG island (CGI) and non-CpG island (non-CGI) regions, as were shown in many earlier studies in placenta and other cell types [57, 129, 131, 133, 185, 186] (Figure 37A-B). Similar distribution was observed for DNA fragments from merging neighbouring CpGs (Figure 37C-D).

For both individual CpGs and regional levels, methylation profiles associated with CGIs have a peak at 0-5%, making the distributions heavily skewed to the right. In non-CGI regions, there was an apparent enrichment of highly methylated CpGs (95-100% methylation) in the third trimester samples (Figure 37B, Figure 37D), indicative of global differences in DNA methylation at different gestation ages.
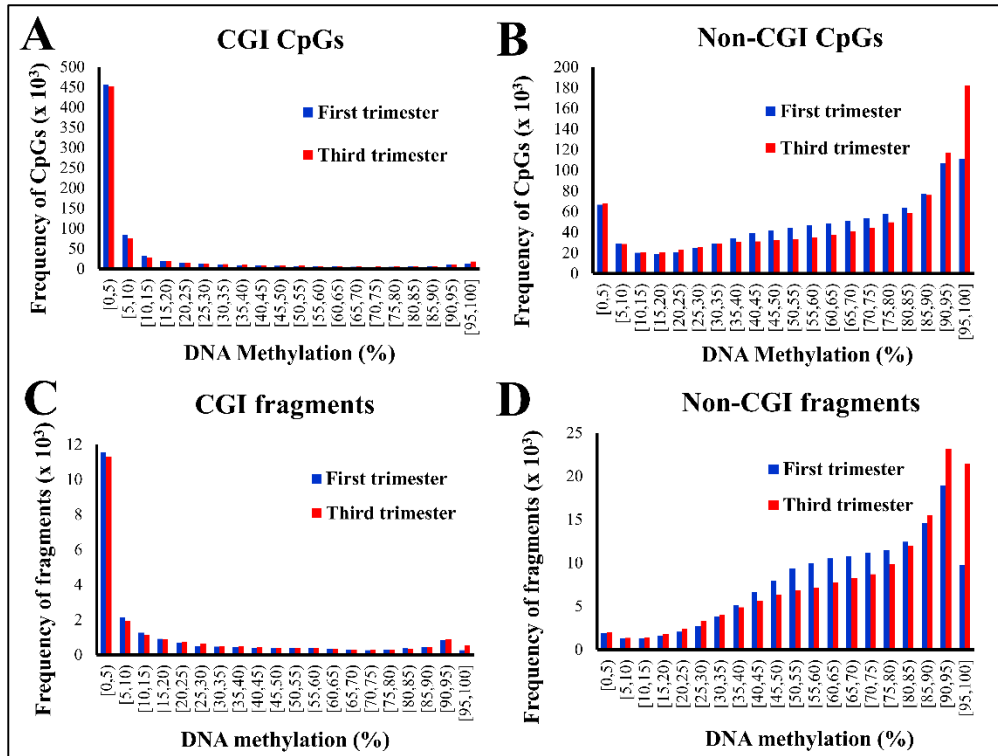
Figure 37: DNA methylation profiles of single and regional autosomes CpGs in CGIs and non-CGIs. Regions are created by merging nearby CpGs of less than 500bp together. (A) Distribution of the average DNA methylation by gestational age for single CpGs (723,727 CpGs sites) which lie in CGIs. (B) Distribution of the average DNA methylation by gestational age for single CpGs (984,183 CpGs sites) which do not lie in CGIs. (C) Distribution of the average DNA methylation by gestational age for regions (22,652 regions sites) which lie in CGIs. (D) Distribution of the average DNA methylation by gestational age for regions (153,563 regions) which do not lie in CGIs.

Principal component analysis showed distinct separation of samples based on gestation age (Figure 38A). Additionally, there was a significant increase in mean CpG methylation in third trimester samples (p = 0.028, 2-sided Mann-Whitney U test) (Figure 38B). Furthermore, at both individual CpG and genomic fragment level, hypermethylation was consistently and significantly more abundant than hypomethylation (Figure 38C, Figure 38D).
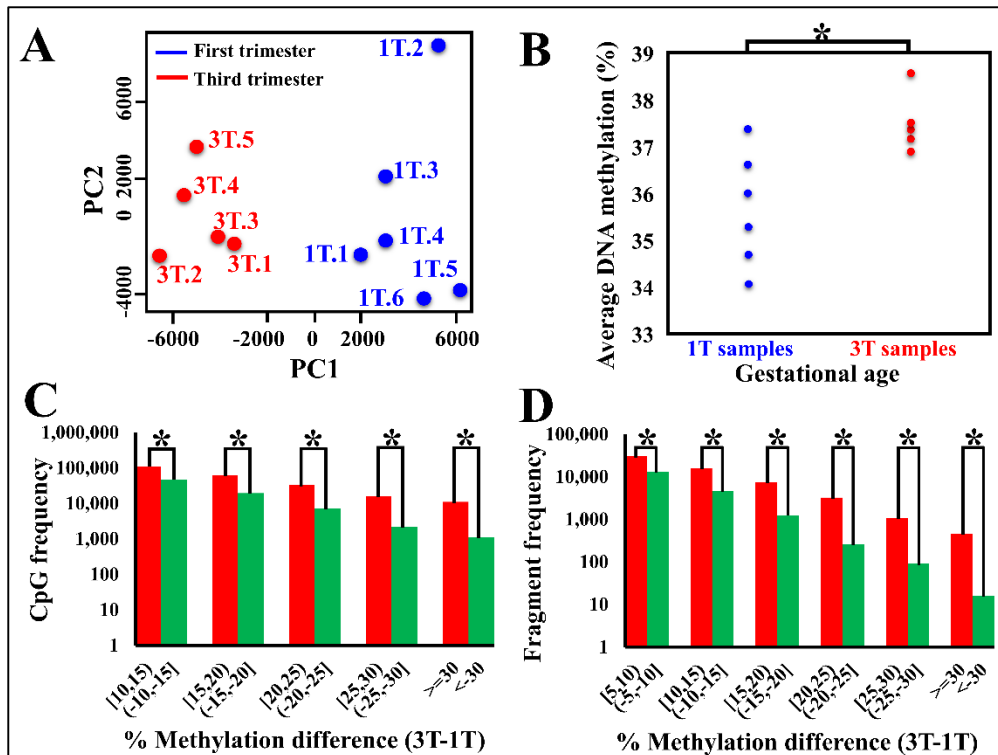
Figure 38: DNA methylation distinguishes samples by gestational age. (A) PCA plot on 1.7 million autosome CpGs shows separation between first and third trimester samples. (B) Dot plot of the average DNA methylation for each sample using 730,594 common CpGs across 11 samples. The third trimester samples show higher DNA methylation than the first trimester samples. Mann-Whitney U test on the average DNA methylation values show significant differences ($p = 0.028$). (C) Distribution of the difference in average DNA methylation between first and third trimester samples, at single CpG level. Strong evidence of hypermethylation was observed, supported by higher peaks in the red bars for all bins of DNA methylation difference. (D) Distribution of the difference in average DNA methylation between first and third trimester samples, at regional level. A similar hypermethylation observation in Figure 2C was observed at regional level.

5.3.4.   Identification of significantly differentially methylated promoters and gene bodies

DNA methylation changes in promoters and gene bodies were further analysed as methylation of these regions have been demonstrated to be associated with gene expression. A total of 199 promoters (corresponding to 189 genes) were found to be significantly differentially methylated between the first and third trimester samples, with 193 (corresponding to 183 genes) (96.8%) showing higher methylation in the third trimester group. I also identified 2,297 gene bodies to be significantly

differentially methylated, with 2,136 (93.0%) being hypermethylated in the third trimester samples. In both promoters and gene bodies, the hypermethylated counts greatly outnumbered the hypomethylated counts.

5.3.5.   RNA-seq analysis

Next, RNA-Seq analysis was carried out on five first/second trimester and four third trimester placenta villi samples. Genes with low expression levels (RPKM < 0.5) in both sample groups were filtered out, leaving 13,756 genes for differential gene expression analysis. A total of 2,447 genes were significantly differentially expressed between first and third trimester samples, of which, 1,889 (77.2%) were down-regulated and 588 (22.8%) were up-regulated in the third trimester samples (Figure 39). Gene ontology analysis with multiple testing correction (p < 0.05) using a commercial database (MetaCore from GeneGo Inc.) was performed independently on the down-regulated and up-regulated gene lists. The down-regulated genes were enriched mainly in the cell cycle pathways, with the top three pathways being "Cell cycle_The metaphase checkpoint", "Cell cycle_Role of APC in cell cycle regulation" and "Apoptosis and survival_DNA-damage-induced apoptosis" The up-regulated genes, on the other hand, were mainly related to immune response, with the top three pathways being "Immune response_Alternative complement pathway", "Immune response_Classical complement pathway" and "Immune response_Lectin induced complement pathway".
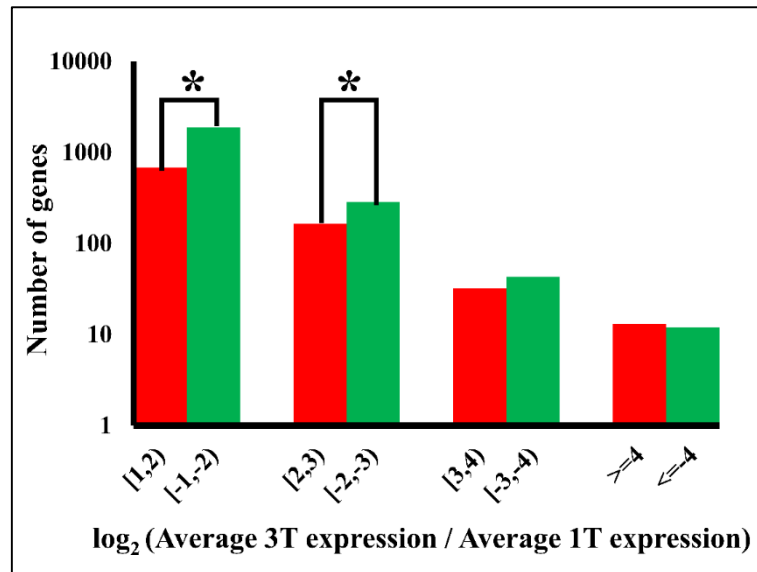
Figure 39: Expression changes during placental development across gestational age. Distribution of gene expression changes from first to third trimester for the genes that showed at least 2 fold changes. There is a general gene repression from first to third trimester, indicated by higher green bars.

Eleven imprinted genes were also differentially expressed, of which ten (*SLC22A18,*

PEG10, MEST, NAP1L5, MIMT1, PSIMCT-1, PEG3, LIN28B, DGCR6, PLAGL1)

showed higher expression for the first trimester samples and one showed higher

expression in the third trimester (*ANO1*). This represented 20.8% of the imprinted

genes that were expressed in placenta (average RPKM $\geq$ 0.5 in either early or late

gestational group)

5.3.6. Correlation between DNA methylation and gene expression

To explore the correlation between DNA methylation and gene expression, I first

equally separated the genes into 50 bins with increasing expression levels. The

average DNA methylation level of the promoters in each bin was then calculated.

Interestingly, for both CGI promoters (promoters overlapping with CGIs) and non-

CGI promoters, a non-linear correlation between DNA methylation and gene

expression was observed (*Figure 40*A-B). There was a clear anti-correlation for genes

with lower expression levels (expression bins 1 to 20). However, for genes at higher

expression levels (bins 21 and above), DNA methylation levels were largely similar

regardless of gene expression level.
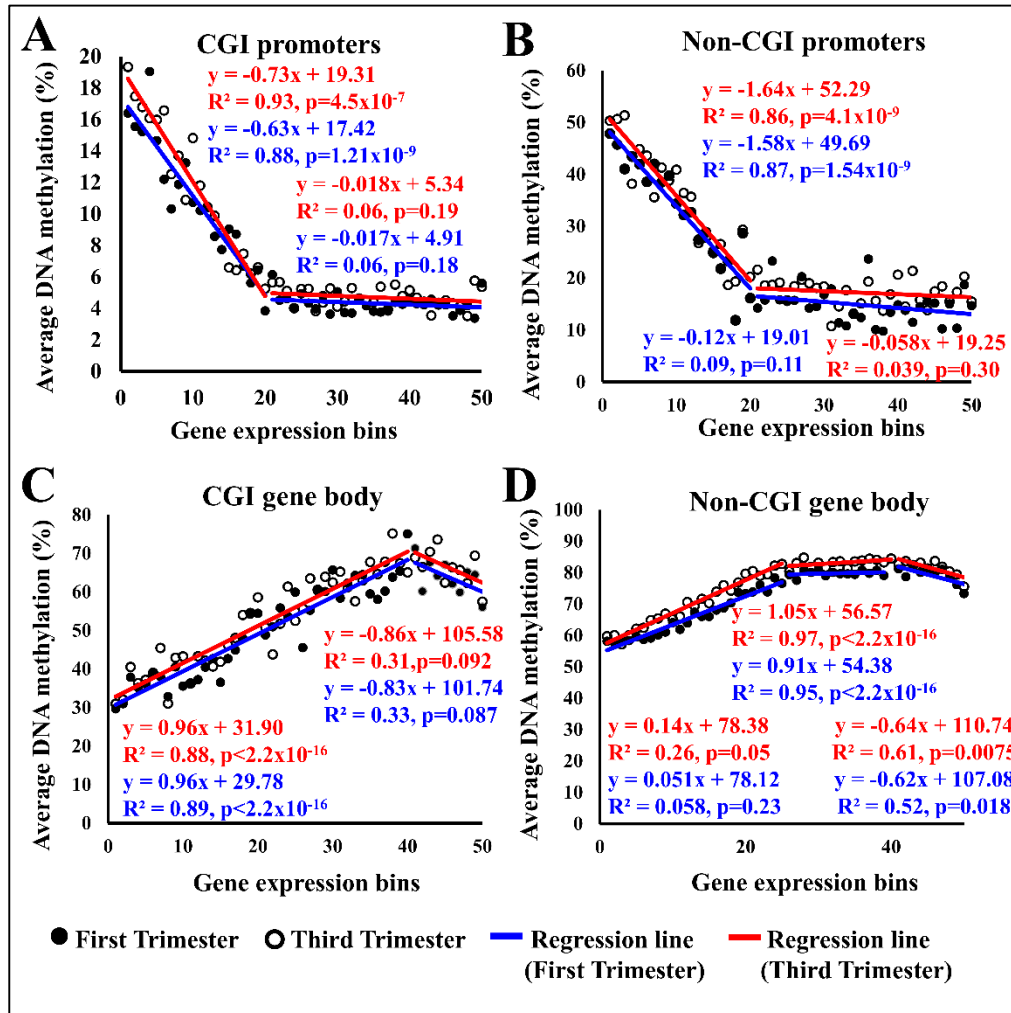
Figure 40: Correlation between DNA methylation and gene expression. Genes were grouped into 50 bins, in order of increasing gene expression. DNA methylation of the promoter or gene body fragments within each gene expression groups were then averaged to obtain the relationship. (A) Scatterplot of the DNA methylation of promoters in CGI against the gene expression showed anti-correlation for the lower expressed genes. (B) Scatterplot of the DNA methylation of promoters in non-CGI against the gene expression showed anti-correlation for the lower expressed genes. The non-CGI promoters showed higher DNA methylation than the CGI promoters. (C) Scatterplot of the DNA methylation of gene body fragments in CGI against the gene expression shows positive correlation. (D) Scatterplot of the DNA methylation of gene bodies in non-CGI against the gene expression shows positive correlation. The non-CGI gene bodies showed higher DNA methylation than the CGI gene bodies.

The correlation between DNA methylation at gene bodies and gene expression was

also non-linear, with CGI gene bodies and non-CGI gene bodies behaving somewhat

differently (Figure 40C-D). For CGI gene bodies, there was a positive correlation

between DNA methylation and gene expression, for genes in bin 1 to bin 40,

followed by a seemingly negative correlation for genes in bin 41 to 50. For non-CGI

gene bodies, the positive correlation was only observed for genes in bin 1 to bin 25.

DNA methylation levels for the genes falling within bin 26 to about 40 were largely

similar, beyond which a seemingly negatively correlation was observed (Figure 40).

We did not observe a difference in exonic and intronic regions (Figure 41).
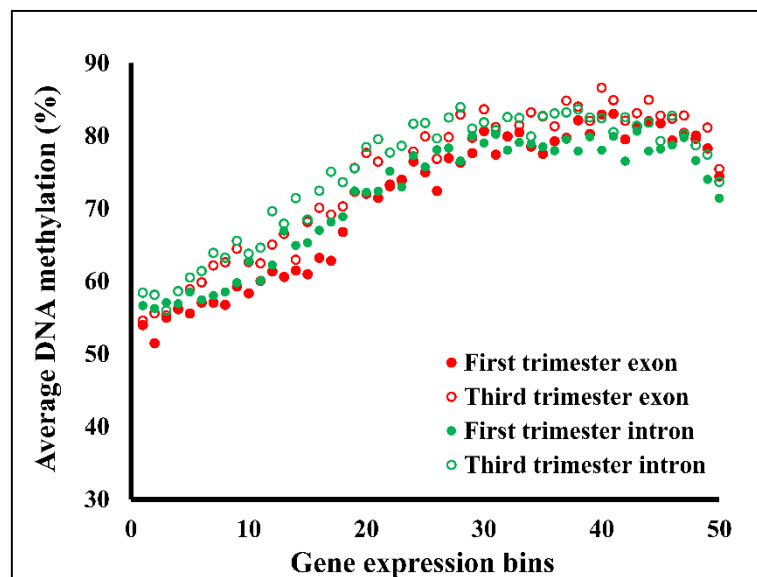
Figure 41: Correlation between DNA methylation and gene expression. Genes were grouped
into 50 bins, in order of increasing gene expression. DNA methylation of the exons/introns
fragments within each gene expression group was then averaged to obtain the relationship.
Scatterplot of the DNA methylation of gene body exons and introns against the gene
expression showed positive correlation. The DNA methylation in exons and introns did not
exhibit clear differences.

However, the DNA methylation levels of CGI exons were consistently higher than

CGI introns, regardless of gene expression levels (Figure 42A-Figure 42B).
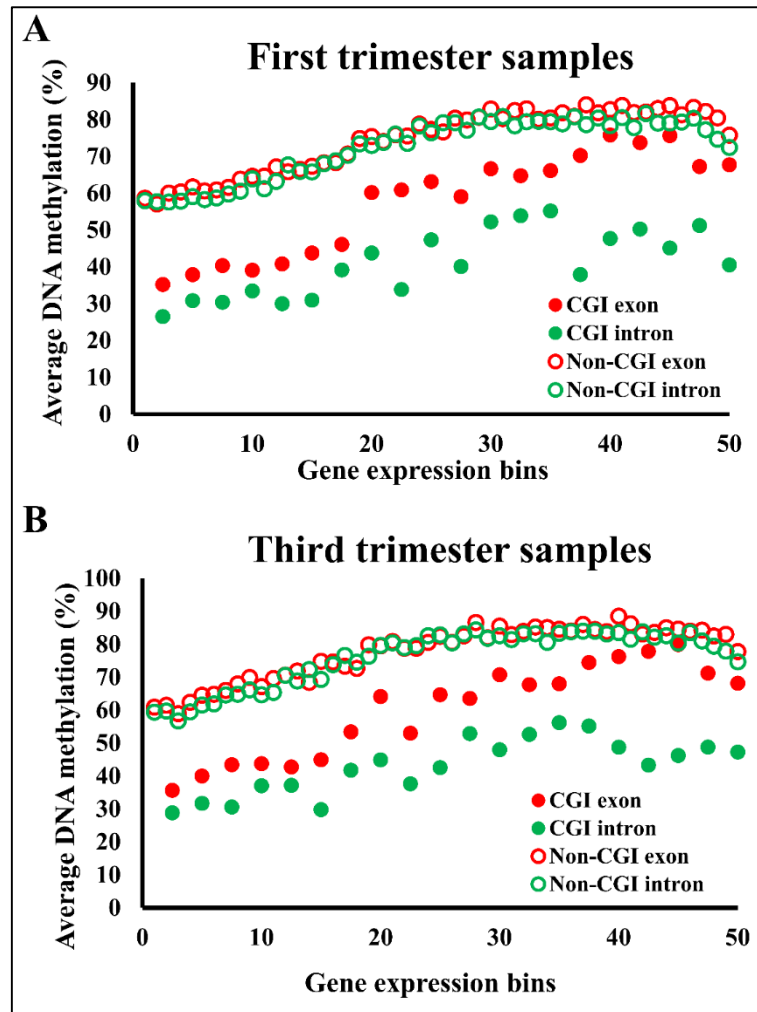
Figure 42: Correlation between DNA methylation and gene expression. Genes were grouped into 50 bins, in order of increasing gene expression. DNA methylation of the exons/introns fragments within each gene expression group was then averaged to obtain the relationship. The genes were divided into 4 groups where non-CGI introns and exons showed similar pattern while differences were observed between exons and introns in CGI gene bodies. First and third trimester samples showed similar patterns and trends.

Lastly, I asked how changes in DNA methylation from early gestation to late gestation in human placenta affects gene expression. A total of 25 genes (Table 15) showed both differential gene expression and differential DNA methylation in promoters when comparing the two gestation groups (Figure 43A). Of those, 19 genes (78%) showed anti-correlation between changes in gene expression and changes in promoter DNA methylation. There was a statistically significant difference between the positive and negative correlations (2-sided binomial test, $p =$ 0.015). We validated the results in five genes (*GJB5, LOC401109, BRDT, BIN2* and

ANGPTL2) using the dual luciferase assay by cloning the respective promoters into the reporter vectors. Gene expression repression was observed in all five genes when the vectors were treated with the methyltransferase M.SssI (Figure 43B).

Gene	Promoter DNA methylation change (3T-1T)	Gene Expression (log$_2$(3T/1T))
ANGPTL2	22.32	-2.20
BIN2	18.44	-3.19
BRDT	12.26	-3.01
CCRL2	23.06	-1.69
CYP2W1	15.00	-3.72
FAM111A	14.58	-1.10
FBXO17	16.93	-1.49
FGL2	26.25	-1.49
GJB5	-24.29	2.51
GREB1	17.90	-1.06
HSPB2	10.37	-1.28
ISLR	18.62	-1.23
LOC100289019	14.03	1.06
LOC401109	-13.71	1.16
MAL	10.43	2.36
PLEKHA6	20.11	1.58
PTPRE	10.36	-1.03
RAB42	18.59	-1.51
RHOBTB2	14.66	-1.38
SEMA6D	24.79	-1.56
SNORD110	13.68	-1.85
SNRPF	14.89	-1.43
ST5	19.15	1.36
STRA6	16.80	1.71
SYNPO	23.10	1.27

Table 15: List of genes having both significant promoter DNA methylation difference and significant gene expression between first and third trimester placenta samples.
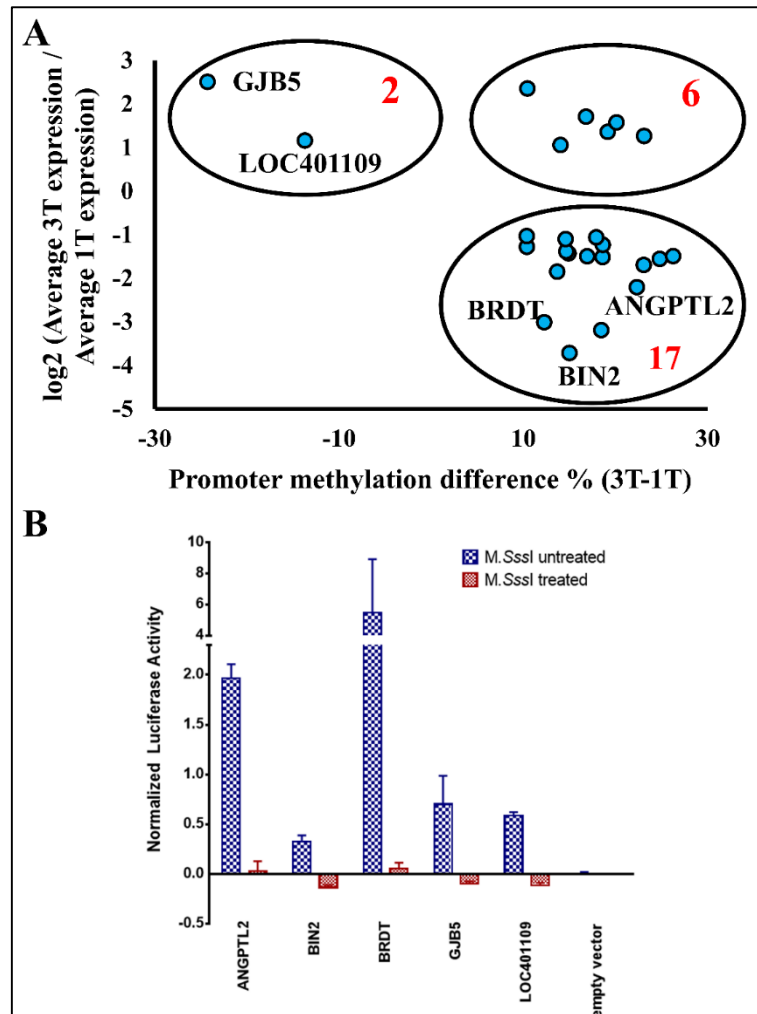
Figure 43: Correlation between significant differential DNA methylation and differential gene expression. (A) Scatterplot of differentially methylated promoters with differential gene expression between first and third trimester samples. 19 out of 25 promoters showed anti-correlation. (B) Experimental validation of the genes labelled in Figure 5A, showing DNA methylation is associated with gene repression. Dual luciferase assays were performed, using empty vector as negative control.

A total of 370 genes showed both differential gene expression and differential DNA methylation in gene bodies when comparing the two gestation groups (Figure 44). Of those, 233 (63%) showed negative correlation between changes in gene expression and changes in gene body DNA methylation. Similar to the promoters, there was a statistically significant difference between the positive and negative correlations (2-sided binomial test, $p = 6.85 \times 10^{-7}$).
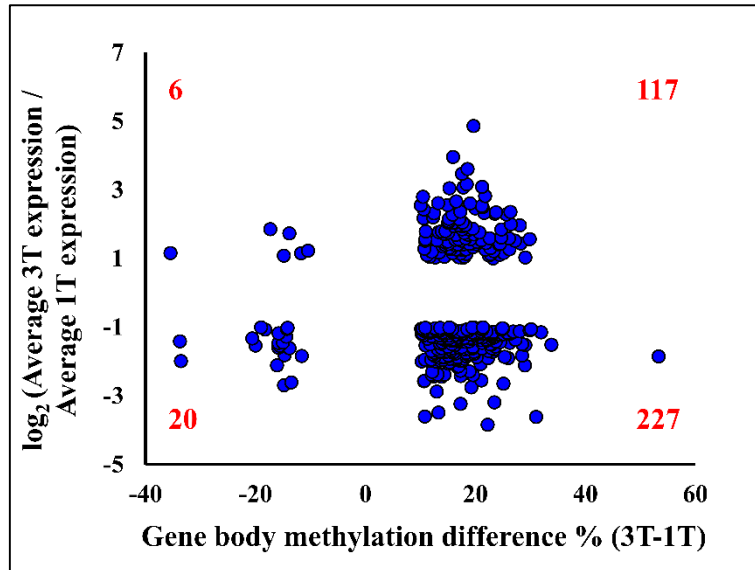
Figure 44: Scatterplot of differentially methylated gene bodies with differential gene expression between first and third trimester samples. Majority of the genes showed negative correlation.

Given that the correlation between gene body methylation and gene expression was positive for genes with relatively lower expression, but negative for genes with higher expression (Figure 40C-D), we hypothesized that the negatively correlated genes (233 genes) were of higher expression levels than the positively correlated genes. RNA-seq data confirmed our hypothesis (Figure 45).
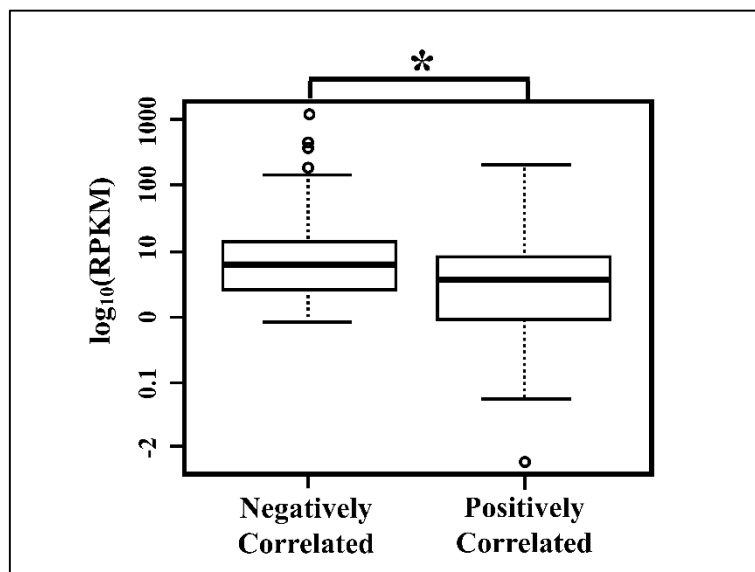
Figure 45: The differential expressed and methylated gene bodies from Figure 44 were grouped by positive and negative correlation. A boxplot comparing the initial gene expression at first trimester was given. The negatively correlated group showed elevated gene expression compared to the positively correlated group (2 sided p-value=8.38×10^{-6}).

5.4.   Discussion

In this study, next generation sequencing techniques was applied to study the gene expression (by RNA-Seq) and DNA methylation profiles (by RRBS) of placenta tissues derived from early and late gestations.

The RNA-seq results shows that a total of 2,477 genes, including eleven imprinted genes, were found to be differentially expressed between the early and late gestational age placenta samples (Table 16). Of the list of eleven imprinted genes, eight were paternally expressed, two were maternally expressed and the remaining one being random. All of the eight paternally expressed genes were down-regulated by the third/term trimester, which is coherent with the fully developed status of the foetus.

Gene	Average first trimester (RPKM)	Average third trimester (RPKM)	$\log_2$ (third trimester/first trimester)	p-value	Maternal/Paternal imprinted
SLC22A18	23.25	3.90	-2.58	0.016	Maternal
PEG10	612.10	188.34	-1.70	0.016	Paternal
MEST	291.60	93.04	-1.65	0.016	Paternal
NAPIL5	4.26	1.48	-1.53	0.016	Paternal
MIMT1	1.34	0.50	-1.42	0.032	Paternal
PSIMCT-1	3.246	1.35	-1.26	0.032	Paternal
PEG3	194.33	81.51	-1.25	0.016	Paternal
LIN28B	23.84	10.08	-1.24	0.016	Paternal
DGCR6	1.38	0.64	-1.11	0.032	Unknown
PLAGL1	119.87	56.75	-1.08	0.016	Paternal

Table 16: Information of the differentially expressed imprinted genes. The list was obtained from *http://www.geneimprint.com*. Expression was given relative to the third trimester samples.

Imprinted genes are essential to the normal growth and development of the mammalian foetus. Paternally and maternally expressed genes have been known to promote and repress foetal growth respectively [187]. Alterations in imprinted genes have been implicated in pregnancy complications such as intrauterine growth restriction (IUGR) [188, 189], preeclampsia (PE) [190, 191] and lethality [32, 33].

Prospectively, even if the foetus survives to birth, these effects may be exhibited chronically and are linked to increased risks for hypertension [192],cardiovascular disease [193-195], abnormalities in neuro [196] and renal development [197, 198].

Distinct profiles of DNA methylation were observed for the placenta samples at different gestational ages. A similar trend had been reported in a study by Novakovic *et al.* [184], whose group made use of the Illumina Infinium HumanMethylation 27 Beadchip to demonstrate the hypermethylation effects in promoters with gestational age increment.

Interestingly, DNA methylation changes in both promoters and gene bodies were predominantly hypermethylation in later pregnancy, coincident with largely gene expression repression. On the genomic level, this suggests that DNA hypermethylation may be used to reduce transcriptional activity in later pregnancy.

The association between DNA methylation and gene expression was found to be complex and dependent on at least two factors: genomic context (promoters or gene bodies) and gene expression level. Consistent with published results [184], we found a negative correlation between gene expression level and promoter methylation level in both early and late pregnancies, as well as a positive correlation between gene expression level and gene body methylation level. However, we also found that such negative and positive correlations were no longer present in genes with higher expression levels (Figure 40). In contrast, for genes with the highest expression levels, there was a negative correlation between gene expression and DNA methylation (Figure 40).

There are a few limitations in our study. First, the placenta is a complex organs with different cell types at different structural locations with different expression profiles [177]. Thus, the molecular profiling at these different locations may reveal different functional changes during pregnancies. Secondly, placenta samples from additional time points during pregnancy may provide more dynamic and detailed changes in gene expression and DNA methylation. Thirdly, other epigenetic mechanisms such as histone modification, transcriptional factor binding and nucleosome positioning may provide further insight into gene regulation.

5.5.   Summary

In this study, we profiled DNA methylation and transcriptome of human placental villi samples from early and late gestational ages. Comprehensive comparisons of DNA methylome were made by stratifying the genome into functional units and subsequently correlating it with gene expression profiles. The following key finding s were made:

i. Distinctive global hypermethylation in promoters and gene bodies with increasing gestational age

ii. Global gene repression with increasing gestational age

iii. Correlations between DNA methylation and gene expression were genomic context dependent and non-linear

CHAPTER 6– GLOBAL DNA HYPERMETHYLATION IN DOWN SYNDROME PLACENTA

6.1.   Background and Hypothesis

Down syndrome (DS) is a genetic disorder that is caused by an extra copy of chromosome 21 (chr21) as a result of random error that occurs during cell division. The incidence of DS occurs one in every 700 live births, across all ethnic groups and social status [108, 199].

Although the cause of DS is clearly defined, there is currently no cure for the condition. DS is mainly caused by three types of chr21 related abnormalities, (i) complete T21, (ii) translocation T21 and (iii) mosaic T21 [200-203], resulting in variable penetrance and expressivity of the aneuploidy effect. There have been over 80 clinically defined phenotypes associated with DS, ranging from affected intellectual ability, characteristic facial features, physical growth delays to affecting organs such as central nervous system, heart, gastrointestinal tract and immune system [108].

By the gene dosage imbalance hypothesis [108, 204, 205], it is unsurprising that most genes located on chr21 showed an increased expression in DS than normal samples. Even so, not all the chr21 genes with altered expression result in a phenotypic effect, possibly attributed to dosage compensation via other regulatory networks or negative feedback loops [108, 204, 205]. What is more intriguing is genes located on other chromosomes are also dys-regulated [112-114], leading to the question of how an extra chromosome is capable of causing a global gene expression change.

The search for therapeutics for DS is mainly been directed towards either silencing the entire extra chromosome [206] or bringing the elevated gene expression to normal levels. Epigenetic regulation such as DNA methylation is often viewed as an epigenetic switch that can silence gene expression. DNA methylation related enzyme such as *DNMT1*, *DNMT3A* and *TET*1 are abundantly expressed in the nervous system [207, 208] and epigenetic alterations have been frequently observed in intellectual disability syndromes [209].

It was reported that children with DS have perturbed homocysteine metabolism, which could result in lower levels of the methyl donor (SAM) and its corresponding reduced form (SAH) [210]. Under such condition, it was thus beyond expectation to see hypermethylation in DS, which indicates that SAM is unlikely is unlikely to be a key player in the DS phenotype [211].

Here, we would like to understand that at an epigenome level, if the perturbations are associated to DS and if such perturbations are functionally relevant to DS. To do so, we quantified the CpG methylation of 17 placenta villi samples, comprising of 11 DS and six normal samples. In addition, we also quantified the transcriptome of four DS and five normal placenta villi. We observed a genome-wide hypermethylation associated with DS phenotype. Furthermore, genes with promoter hypermethylation were associated with functions related to DS phenotypes. Our results thus support that perturbation in the DNA methylome maybe one of the regulative mechanisms explaining for the DS phenotypes.

6.2. Material and methods

6.2.1. Samples

Women with euploidy and DS pregnancies who attended KK Women's and Children's Hospital, Singapore, were recruited. Chorionic villus samples from subjects carrying a normal or DS foetus at the first or second trimesters of pregnancy were collected by chorionic villus sampling (CVS). Placenta villi samples (foetal side) from DS foetuses were collected from termination of pregnancy (TOP). This study by supported by a Bench to Bedside grant (09/1/50/19/622) from BMRC-NMRC.

6.2.2. DNA methylation

Six DNA samples from normal pregnancies and 11 samples from carrying DS foetuses were chosen for DNA methylation analysis by RRBS [133]. Sequencing data were deposited into the GEO database with accession numbers GSE42144. Sample preparation and computational processing followed the steps described in Chapter 3, Section 3.2.2 and Section 3.2.3 respectively.

6.2.3. Differential DNA methylation analysis

Differential DNA methylation between normal and DS samples were analysed at both single CpG and genomic level. Autosome CpGs (sequencing depth $\geq$10) covered in at least 3 normal and 6 DS samples were used for all subsequent analyses. A CpG was considered differentially methylated if the methylation difference between average DS and average normal samples was at least 10%; and 2) $p<0.05$ (Wilcoxon rank-

sum test, two-sided). For genomic regions, the same set of criteria follows, with an extra requirement that the region should include at least 6 CpGs.

The probability density functions (PDF) for comparison between DS and normal samples were calculated and plotted using R package.

6.2.4. RNA-seq

Gene expression was quantified using RNA-seq, details found in [133]. Five RNA samples from normal pregnancies and four samples from pregnancies carrying DS were included in the study. The protocol followed the steps from Chapter 5, Section 5.2.4.

6.2.5. Differential RNA-seq analysis

The expression level for each gene was represented by the reads per kilobase per million mapped reads (RPKM) value, using the formula below:

$$RPKM = \frac{\text{Number of aligned reads for a gene of interest}}{\text{Number of total aligned reads} \times \text{Transcript length for the gene (kb)}} \times 10^6$$

Average RPKM values for each gene in each sample group (normal and DS) were calculated. When the average RPKM for a gene is less than 0.5, the value was set as 0.5. A gene was considered to be differentially expressed between normal and DS samples when: 1) Binomial test with a Benjamini-Hochberg corrected p value of less than 0.01; and 2) the ratio of (Average DS/Average normal) ≥ 1.25 or ≤ 0.8. We used the R package to calculate the probability density functions (PDF) distributions for various gene groups with regard to the expression changes represented by $\log_2$(Average DS/Average normal).

6.2.6. DNA methylation validation by EpiTYPER assays

An independent set of gestational age matched samples were used for validating results from RRBS and computational analyses. 14 normal (gestational age: 17.41±3.77 weeks) and 17 DS (gestational age: 17.41±3.77 weeks) placental villi samples were included. All reagents and equipment were from Sequenom (San Diego, California, USA). Briefly, 1 µg of genomic DNA was bisulfite converted with EZ DNA Methylation Gold Kit (Zymo Research, USA) and subsequently amplified. The PCR products were then treated with shrimp alkaline phosphatase (SAP), followed by T-cleavage transcription/RNase A cocktail from EpiTYPER Reagent Kit (Sequenom). The products were then conditioned by Clean Resin and fragments analysed using MassARRAY system and data was analysed with EpiTYPER 1.2 software (Sequenom). For each sample, the DNA methylation level was obtained by taking the average of all analysed CpGs within the target amplicon.

6.2.7. Quantitative real-time PCR validation

An independent set of gestational age matched samples were used for validating results from RRBS and computational analyses. 8 normal (gestational age: 19.18±3.56 weeks) and 10 DS (gestational age: 178.37±2.70 weeks) placental villi samples were included. All reagents and equipment were from Life Technologies. Briefly, 0.5 to 1 µg total RNA was treated with DNase I and subsequently subjected to first strand DNA synthesis by SuperScript III First-Strand Synthesis SuperMix Kit. Quantitative real-time PCR was performed on Applied Biosystems 7900HT Fast Real-time PCR system with 384-well block module. Each assay were duplicated and the average Ct value was obtained using SDS version 2.3 software. GAPDH was used for normalisation using the formula:

$$\text{Expression level of a target gene} = 2^{-(Ct(targetgene) - Ct(GAPDH))}$$

6.3. Results

6.3.1. Descriptive statistics for RRBS quality

We quantified DNA methylation using RRBS on 11 DS and 6 normal samples. On average, there were approximately 37.5 million reads per sample with 21.4 million aligned reads (C2T or G2A references), corresponding to 57% alignment rate. The bisulfite conversion were above 99% for all the samples (Figure 46A).

We used a minimum sequencing depth of at least ten, leaving an average of about 1.7 million CpGs (Figure 46B) in each placenta villi samples.
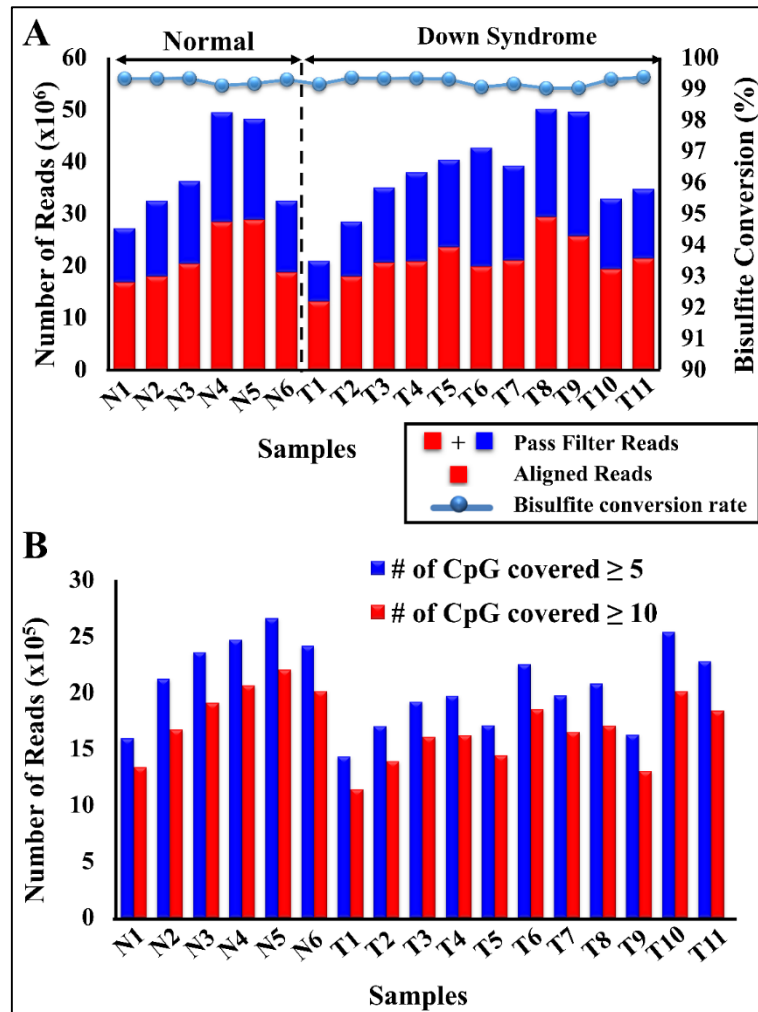
Figure 46: Descriptive statistics of the 17 samples. (A) For each sample, red bars represent the number of uniquely aligned reads, while blue bars are the non-uniquely or unaligned reads. The blue line gives the bisulfite rate for each sample. (B) Number of CpGs at 5x and 10x coverage

6.3.2. Genomic coverage of the analysed CpGs

The assayed CpG sites represented about 3% of all the CpG sites in Hg19, on both Watson and Crick strands (Figure 47A). On a genomic level (includes at least 3 CpGs), the assayed CpG sites covers 75.1% of annotated CGIs (Figure 47B), 50.8% of CGSs (Figure 47C) and 51.9% of promoter regions (Figure 47D).
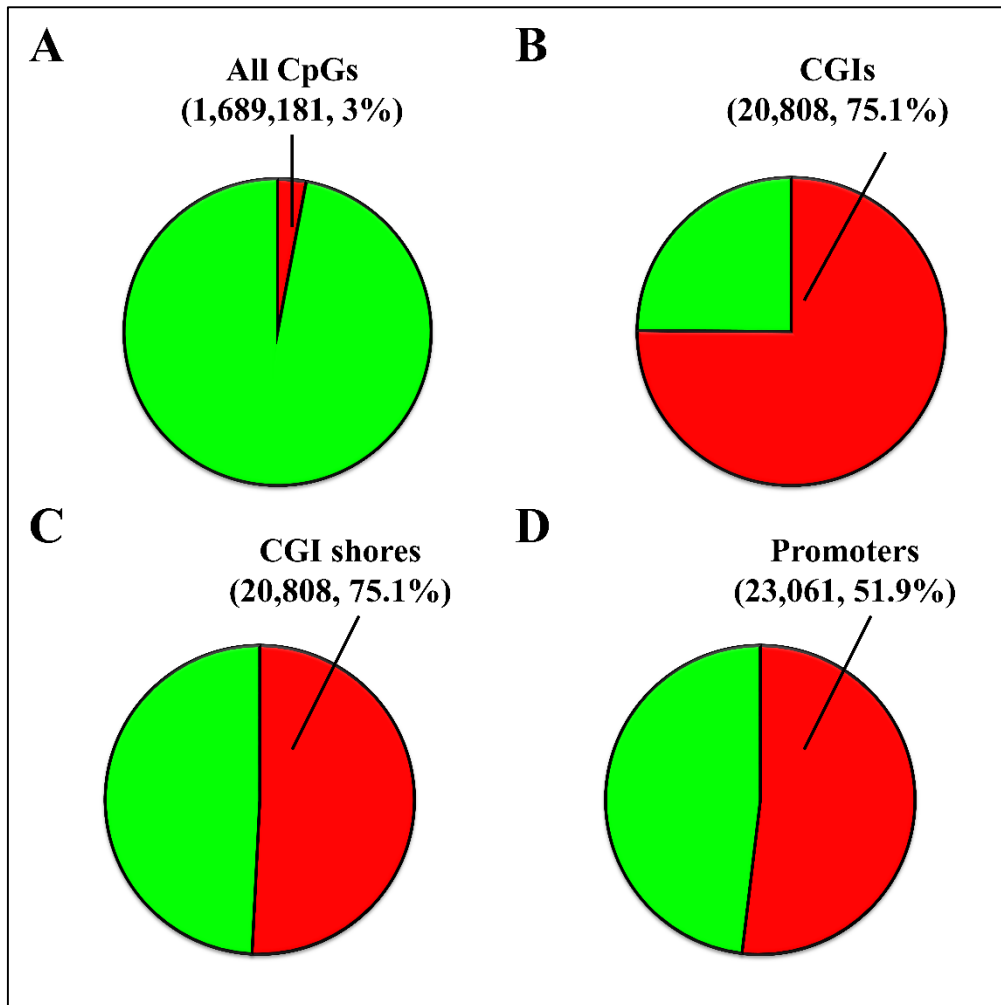
Figure 47: A CpG site was considered covered if the sequencing depth was ≥10. A genomic region (CGI, CGI shore or promoter) was considered covered if at least 3 CpGs within the region was sequenced at a depth ≥10. Figure adapted from [133].

Next, we examined the locations of the CpGs with respect to CpG content. A total of 56% of the CpGs were located in either CpG rich CGIs (731,924 CpGs corresponding to 43%) or CpG medium rich CGSs (218,659 corresponding to 13%) and the remaining 44% (738,598 CpGs) are found in CpG poor regions (Figure 48A). Independently, we also assessed the genomic locations of the CpGs. They were mostly located in intragenic regions (38%), followed by intergenic (36%), then promoters (defined as -1kb to +500bp from transcription start site, 23%) and lastly, transcription termination regions (TTR, defined as -500bp to +500bp from TTS, 2%) (Figure 48B).
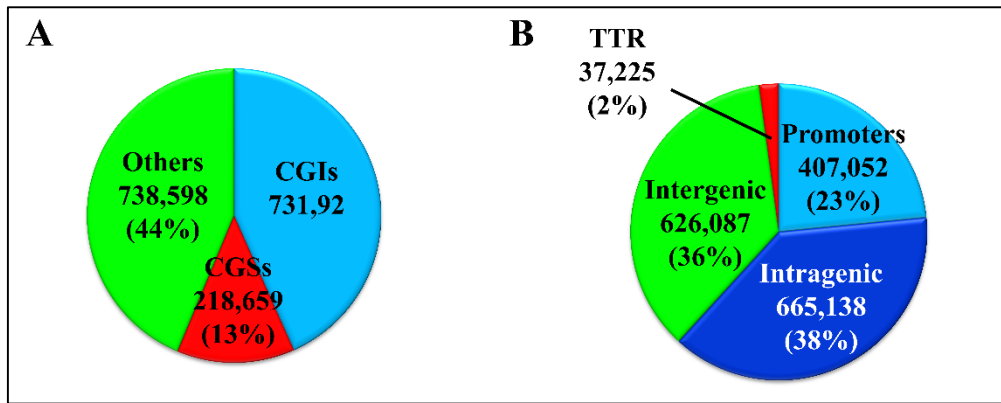
Figure 48: Distributions of covered CpGs in different functional regions. (A) Context of CpG richness (B) Genomic locations. Figure adapted from [133].

6.3.3. Methylation distribution profiles are unique to genomic location regardless of normal or DS status

The methylation of all the CpGs showed a distinct bimodal profile with prominent peaks at the two extreme ends (Figure 49A), consistent in other studies involving different cell types [39, 57, 129, 131, 185, 186]. ~30% of the CpGs have methylation 0-5% while another ~10% of them have methylation 95-100%. This made up a total of 40% of all analysed CpGs in a mere combined 10% interval. However, this is an underestimation as the fully methylated CpGs located in repetitive regions have been intentionally removed as a part of the RRBS feature.

Upon stratifying the genomic regions by the functional locations (promoters, TTR, intragenic and intergenic), we observed dramatic differences in the methylation profiles, with promoters being most distinctive from the rest (Figure 49B- E). Unlike TTR, intragenic and intergenic regions which have similar profiles, promoters have a positively skewed distribution where there is a strong enrichment (~70%) of the CpGs in 0-5% methylation. In contrast, the non-promoter regions have higher proportions of partially methylated CpGs (30-70%), as have been observed by other groups [131].

117
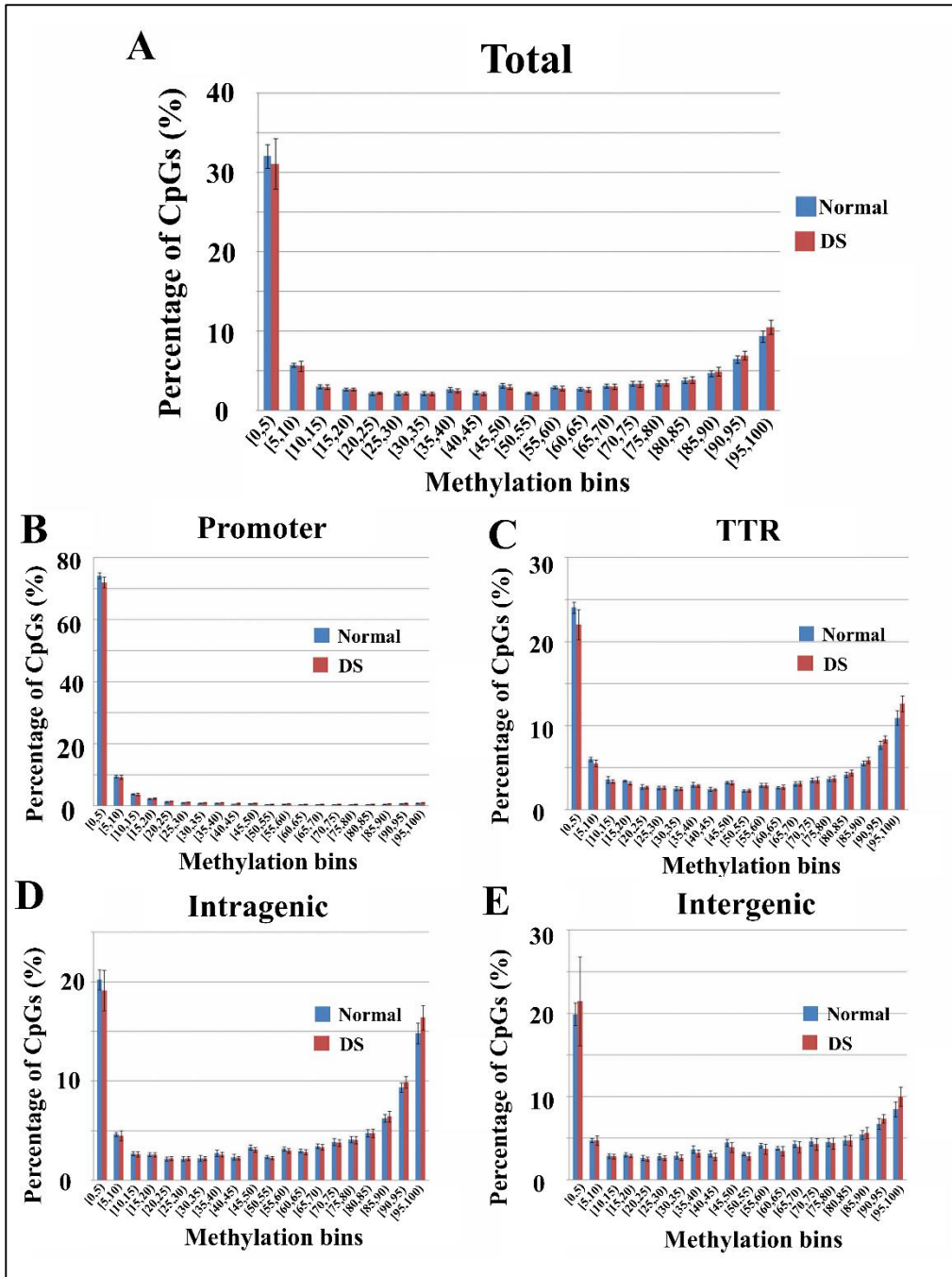
Figure 49: Distributions for individual CpG methylation at respective genomic regions. (A) All analysed CpGs; (B); Promoter regions (C) Transcription termination regions (TTR); (D) Intragenic regions; (E) Intergenic regions. The average methylation for normal and DS samples at each CpG site was used for the plots. Figure originally from [133].

6.3.4.   Inter-Individual Variability of Normal Foetuses

We next assessed the inter-individual variability by using CpGs from five normal

male foetuses, from the motivation that CpGs with high variability might be potential

biomarker candidates [212]. Thus, we only selected CpGs with partial methylation, as defined by those having average methylation of 30-70% as this range would have higher variability than the extreme values. Varying sequencing depth of 10, 20 and 50 were also used to demonstrate the impact of sequencing depths to the accurate quantification of DNA methylation.

Most CpGs (~50% to 75%, depending on minimum sequencing depth) have standard deviation among the five samples within 10% and almost all the CpGs have standard deviations within 27.5% (Figure 50). Unsurprisingly, increasing the cut off for sequencing depth gives rise to lower variability, suggesting that besides biological variability, sequencing depth is also an important contributor towards observed variation among similar samples.
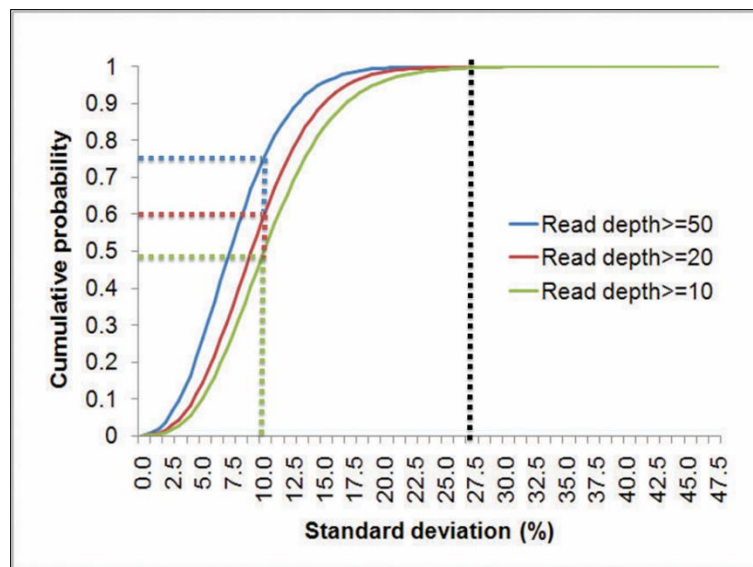
Figure 50: Inter-individual variability for CpGs. Only CpGs with methylation 30-70% for five normal male foetuses were included in the analysis. Analysis was performed at varying sequencing depth cut offs of 10, 20 and 50. Figure originally from [133].

6.3.5. Methylation status distinguishes samples by disease status

PCA is a statistical technique that is commonly used to linearly transform a set of possibly correlated variables (CpG methylation) into a set of linearly uncorrelated variables known as principle components. By doing so, it also reduces the dimension of the data. Typically, the first principle component accounts for most of the variability in the data.

PCA based on DNA methylation percentage of the CpGs revealed that the samples were separated by disease status and not on gender (Figure 51). The combined first and second principle component accounts for 20.6% of the variability.
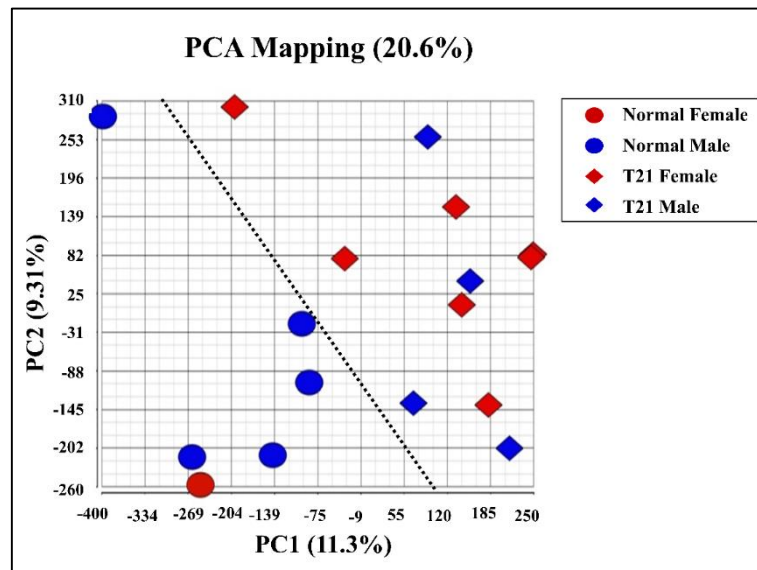
Figure 51: PCA plot using analysed CpGs. Samples were clearly separated by disease status. Figure originally from [133].

6.3.6. Global hypermethylation in DS Samples

Earlier studies performed on DS chorionic villus samples and leukocytes have shown a dominance of hypermethylated to the hypomethylated promoters [211, 213]. In our study, we made extended from promoter analysis and observed a global hypermethylation which was consistent in all defined genomic regions (CGIs, promoters, gene bodies, TTRs, intergenic) (Figure 52A-F). This observation was

demonstrated by an elevated probability density curve for hypermethylation (DS >

Normal, red curve) than the curve for hypomethylation (DS < Normal, blue curve). In

addition, such a phenomenon was not only limited to genomic location but across all

autosomes (Figure 52G). As shown in Figure 52H, the average DNA methylation of

CGIs, given by the average methylation of all CpGs (at least 6) falling within the

CGI, was significantly higher in the DS than normal samples (2 sided Wilcoxon rank-

sum test, $p<0.002$).
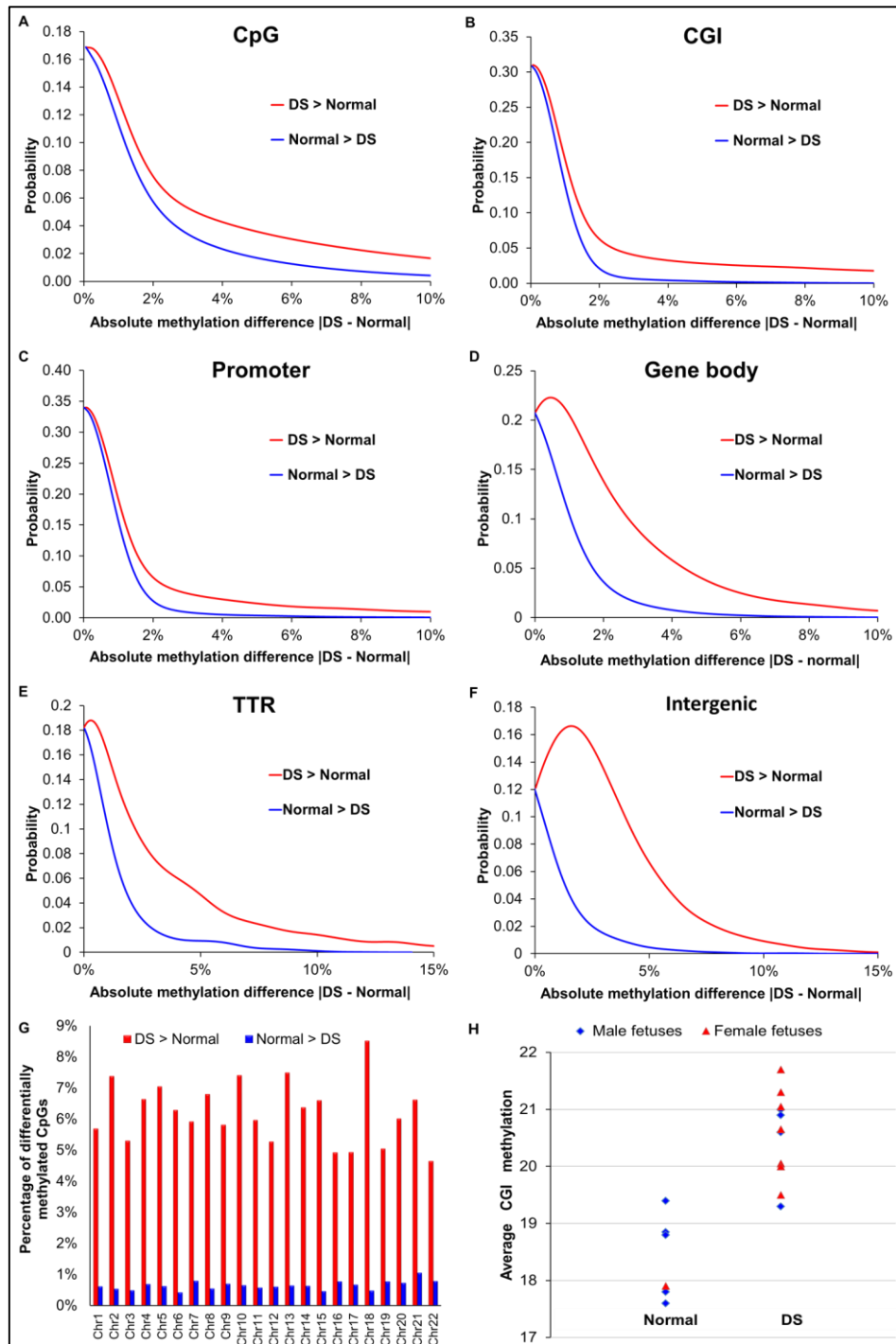
Figure 52: (A) Probability density function (PDF) distribution for methylation difference between DS and normal samples for individual CpGs. Similarly, the methylation difference values for (B) 18,939 CGIs (each CGI with at least 6 covered CpGs), (C) 19,479 promoters (each promoter with at least 6 covered CpGs), (D) 30,648 gene bodies (each gene body with at least 6 covered CpGs), (E) 3,215 TTRs (each TTR with at least 6 covered CpGs) and (F) 8,611 intergenic regions (each intergenic region with at least 6 covered CpGs) were used for calculating their respective PDF distributions. In (A–F), hypermethylation in DS (DS>Normal) occurs much more frequently than hypomethylation in DS (Normal>DS). (G) Percentages of hyper- and hypomethylated CpGs in each autosome. (H) Average CGI methylation was higher in DS than in normal samples (p<0.002, Wilcoxon rank-sum test, two-sided). Only CGIs with at least 6 covered CpGs were included. Figure originally from [133].

122

Next, we performed significant differential analysis on the CpGs, details provided in Section 6.2.2. Of all the genomic regions analysed, TTRs and intergenic regions have the highest proportion of significant DMC, followed by intragenic, non-CGI promoters and finally CGI promoters (Figure 53). Despite so, the ratio of hypermethylated to hypomethylated CpGs was the highest in promoters among all other categories (ratio of 56.2).

	Promoter (All)	Promoter (CGI)	Promoter (Non-CGI)	TTR	Intragenic	Intergenic	Total
# DS > Normal	18,937	14,500	4,437	3,415	38,436	33,944	94,732
# DS < Normal	721	258	463	260	5,298	3,862	10,141
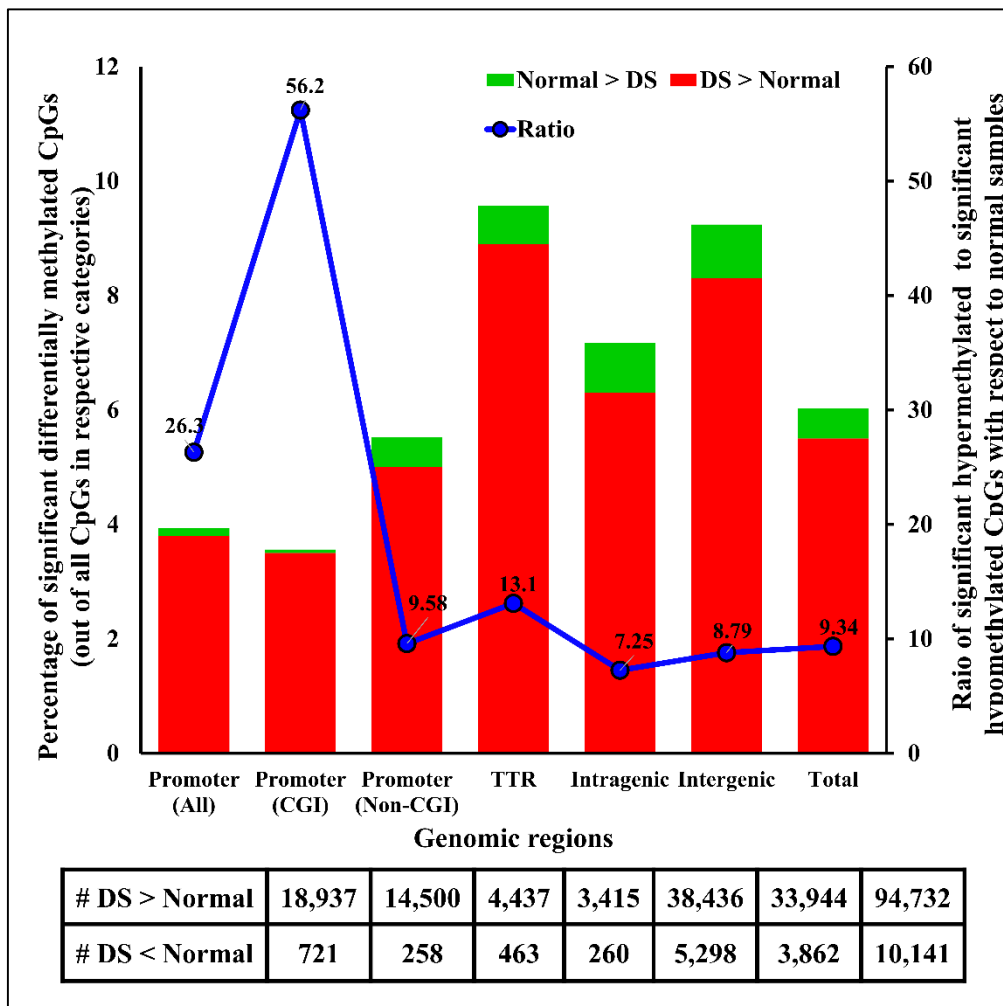
Figure 53: Proportion and frequencies of differentially methylated CpGs in different genomic regions.

6.3.7. DNA methylation perturbations in DS may occur early in development

We compared our list of significantly differentially promoters against a study by Kerkel et al. [213], in which the authors identified nine genes with differential methylation between peripheral blood leukocytes (PBLs) from DS and karyotypically normal controls. Out of this list of nine, three of them (*TCF7*, *FAM62C* and *CPT1B*) were also found in our results. The statistical significant overlap between studies involving different developmental origins suggests that perturbations in DNA methylation of DS phenotypes could have occurred during early development.

6.3.8. Hypermethylated genes largely correlated with gene repression

RNA-seq analysis was performed on five normal and four DS placenta villi samples. Gene expression was measured using RPKM. We removed the genes which have an average RPKM<0.5 in both normal and DS groups and reduced the analysed gene list to 14,090. Out of these, 1,413 were significantly higher expressed in DS which includes 73 chr21 genes. Meanwhile, out of 3,449 genes which were significantly up-regulated in normal samples, only three were chr21 genes.

We next compared the expression change profiles between chr21 and all background genes. There was an apparent right shift of the chr21 genes relative to all genes, corresponding to an average of 53% up regulation in gene expression in DS samples (Figure 54). This is consistent with gene dosage imbalance hypothesis [108, 204, 205] and other reports [109-111].

A total of 589 genes located on the autosomes were significantly hypermethylated on DS samples. Of these, 207 of them were non chr21 genes which passed the minimum gene expression threshold (RPKM $\geq$ 0.5 in either DS or normal group). There was significant down regulation of gene expression (two-sided Wilcoxon rank-sum test, p <0.05), indicating an anti-correlation with promoter DNA methylation changes. This was consistent with the left shift of the PDF for hypermethylated genes relative to "All genes" curve (Figure 54).
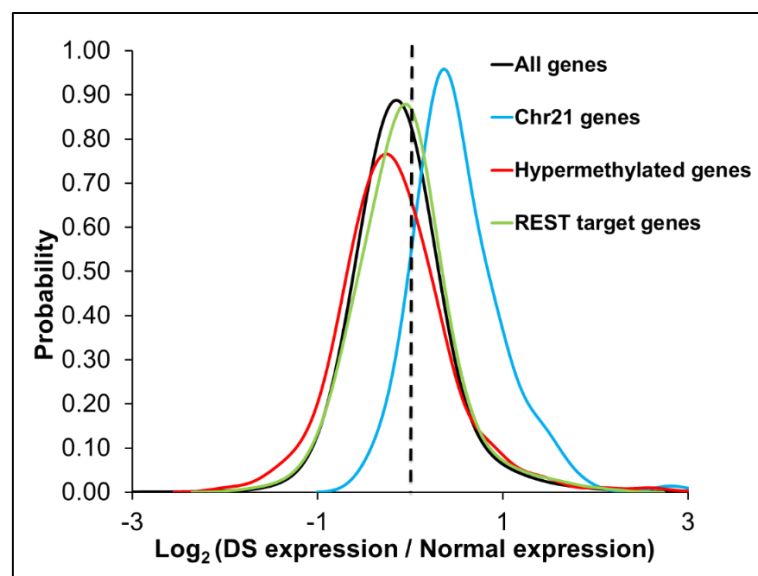
Figure 54: Proportion and frequencies of differentially methylated CpGs in different genomic regions. Figure originally from [133].

Four genes (*CS1*, *TFAP2E*, *CDH13*, *and NDN*) with hypermethylated promoters and corresponding gene repression were selected for validation with EpiTYPER assays and quantitative real-time PCR on an additional set of gestational age matched samples (Figure 55).
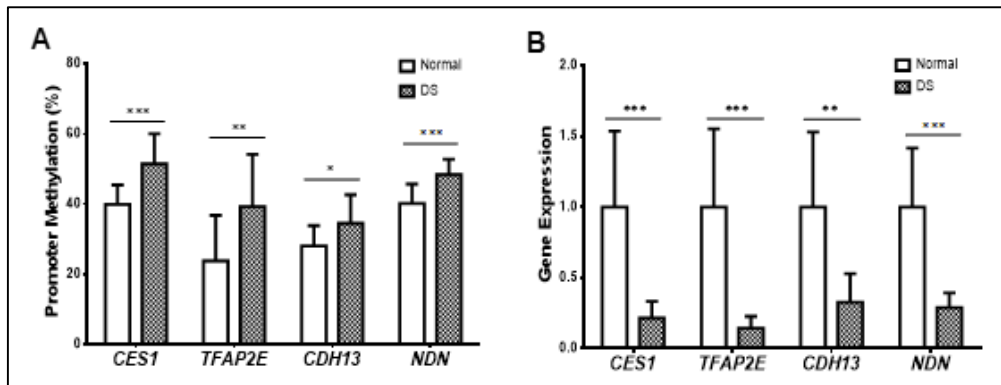
Figure 55: Genes with hypermethylated promoters are associated with expression down-regulation in DS. (*A*) Promoter hypermethylation in DS samples. (*B*) Down-regulation of gene expression in DS samples. *: p<0.05, **: p<0.01, ***: p<0.001, t-test, two-sided. Error bars represent standard deviations. EpiTYPER: normal n = 14, DS n = 17. Gene expression: normal n = 8, DS n = 10. Figure originally from [133].

6.3.9. Gene ontology analysis of significantly differentially methylated promoters

Gene ontology analysis was next carried out on the list of significant differentially methylated promoters (589 hypermethylated and 9 hypomethylated) using a commercial database (MetaCore from GeneGo Inc.). The top three enriched pathway maps were:

    i.   Immune response_Lectin induced complement pathway

    ii.   Neurophysiological process Dopamine D2 receptor signalling in CNS

    iii.   Cytoskeleton remodelling Neurofilaments

Additionally, the software also identified top three significantly enriched process networks being

    i.   Inflammation complement system

    ii.   Signal transduction Neuropeptide signalling pathways

    iii.   Developmental Neurogenesis Axonal Guidance

The results from these two analyses were highly correlated which implied an involvement of perturbations in the immune system and neural development, related to DS phenotypes.

6.3.10. Possible mechanisms to explain for genome-wide hypermethylation

Previous studies have reported of lowered levels of SAM in the plasma of DS individuals. However, we noticed a global hypermethylation which was opposite to what would be expected given a reduced availability of the methyl donor for methylation. This implied other mechanisms have acted to bring about methylome changes in the DS.

The *TET* family is involved in DNA demethylation [35, 208, 214, 215] and is made up of three main members (*TET1* on chr10, *TET2* on chr4 and *TET3* on chr2) which none are located on chr21. Our RNA-seq results showed that all three genes were significantly down regulated in DS (Table 17). We further validated the results by using quantitative real-time PCR on a new set of gestational matched samples (Figure 56), which confirmed that *TET1* and *TET2* were indeed significantly differentially expressed between normal and DS samples.

Gene	Chromosome	Average (Normal$_{RPKM}$)	Average (DS$_{RPKM}$)	(DS$_{RPKM}$/ Normal$_{RPKM}$)	Corrected p-value
TET1	Chr10	0.80	0.57	0.71	5.89E-03
TET2	Chr4	4.01	2.54	0.63	5.18E-19
TET3	Chr2	5.44	3.40	0.62	1.14E-29
REST	Chr4	5.73	3.41	0.60	1.30E-25

Table 17: RNA-seq expression values for *TET1*, *TET2*, *TET3* and *REST*
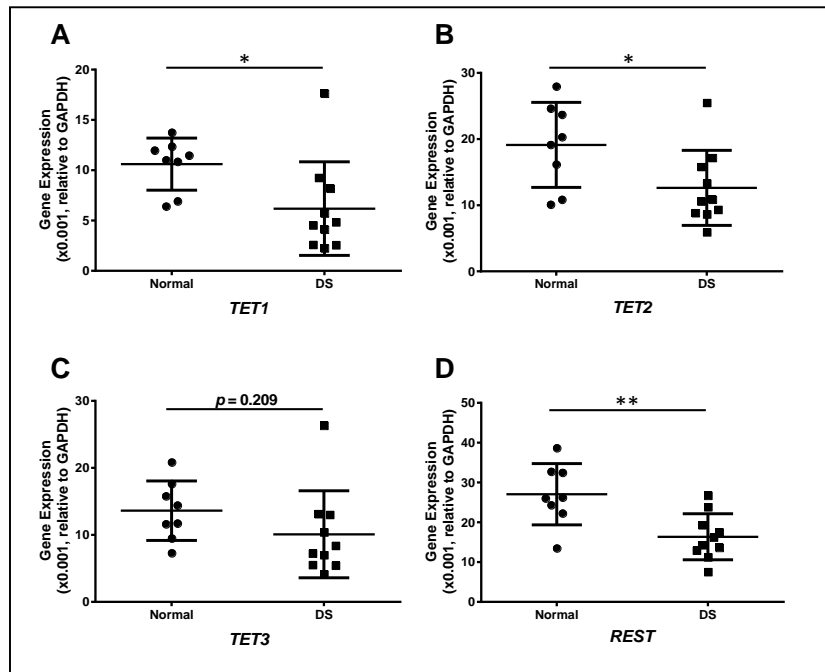
Figure 56: Quantitative real-time PCR validation of expression changes for (*A*) *TET1*, (*B*) *TET2*, (*C*) *TET3* and (*D*) *REST*. *: p<0.05, **: p<0.01. T-test, two-sided. Error bars represent standard deviations. Sample size: normal n = 8, DS n = 10. Figure originally from [133].

Repressor element 1 silencing transcription factor (*REST*) is a transcriptional repressor in that regulates gene expression in neuronal and non-neuronal cells. We observed significant down regulation of REST in both our RNA-seq data (Figure 57) and quantitative real-time PCR (c). In conjunction with recent studies showing (i) DNA hypermethylation observed in *REST* -/- cells and (ii) *REST* binding to the target regions is able to maintain the hypomethylated states in low methylated regions (10-50%) [131], we investigated the expression profiles of the *REST* target genes. The *REST* target genes were slightly up-regulated compared to the background gene list as shown in Figure 57. With a down-regulation of *REST* expression in DS, this could have caused a reduction in the binding of *REST* to the target genes, leading to DNA hypermethylation relative to the non-*REST* target genes.
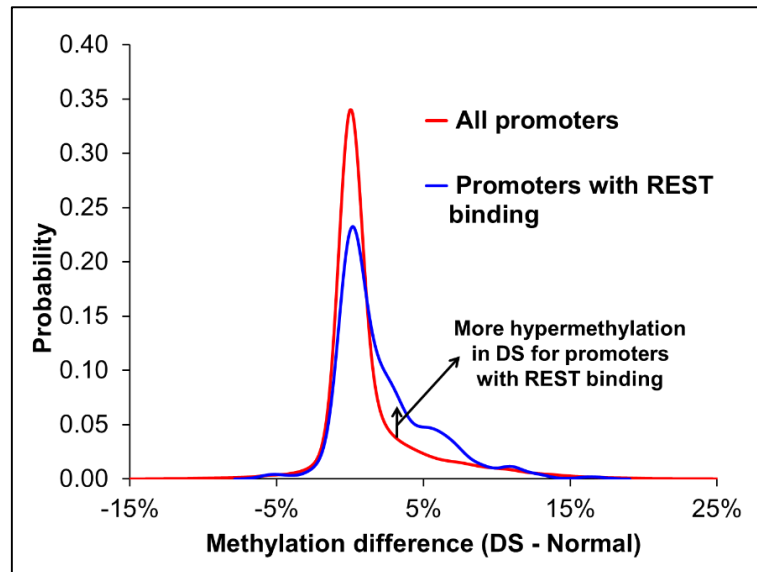
Figure 57: PDF distributions of methylation difference for all promoters and promoters targeted by *REST*. Figure originally from [133].

6.4.   Discussion

In this study, we analysed the DNA methylation profiles of human placenta villi tissues derived from mothers carrying normal and DS foetuses to investigate if perturbations in DNA methylation is a possible mechanism connecting Trisomy 21 and varying DS phenotypes.

Comparison made between placenta villi (our study) and PBL (Kerkel et al.) identified three common genes (out of nine genes from Kerkel et al.) being differentially methylated in a similar fashion. This suggests that epigenetic perturbation occurs early in DS embryo development and persists to adulthood. These epigenetic perturbations may occur either early or later during development, leading to varying predisposition to certain disease.

Data from other studies, together with our RNA-seq results, have provided possible explanations to the observed global methylation via manipulations to important genes which can affect DNA methylation.

i.   RNA-seq data from our study has shown that genes from the *TET* family were down regulated in DS. It may be possible that a decrease in *TET* gene expression may cause a hypermethylation to the target regions by a reduction in DNA demethylation.

ii.   *DYRK1A* is a chr21 gene located in DS critical region. This gene was up regulated in DS and may have induced global epigenetic changes by down regulating *REST*, which is a transcription factor important to maintain DNA hypomethylation in low methylated regions [131], thereby leading to hypermethylation of the REST target genes.

The study design of using clinical samples such as placenta villi to investigate DNA has to be carefully planned to avoid scrutiny from a statistical perspective. Being aware of possible confounding factors such as gestational age of the placenta, gender, potentially different cell mixtures from different samples, we have validated some genes for both DNA methylation and gene expression data on a new gestational age match dataset using EpiTYPER and quantitative PCR.

6.5.   Summary

We compared the DNA methylomes and expression profiles between the placental villi samples from subjects carrying a normal or DS foetus. The key findings have been summarised below:

i.   An overall hypermethylation was observed in the DS samples which was consistent across all genomic regions.

ii.   Genes located on chromosome 21 were on average up-regulated by 50% in DS samples.

iii.   Genes with hypermethylated promoters were associated with down-regulation

iv.   Comparative studies between placental villi and adult PBL showed a significant overlap, suggesting global epigenetic changes occurs early in development, leading to DS phenotypes

v.   Global hypermethylation could be partly attributed to reduced gene expression in *TET* and *REST* in the DS samples.

CHAPTER 7 – CONCLUSION

The regulations of the eukaryotic genome and gene expression are complex processes which involve multiple layers of mechanisms. Increasing evidence have argued that gene expression is no longer a simple reflection of DNA sequences; instead, modifications to DNA and histones known as epigenetics may play a fundamental role. In recent years, HTS technologies in combination with bioinformatics data mining methods allow for systematic genome-wide analysis among different -omics data types to elucidate complex interactive relationships.

Taken together, we have shown that DNA methylation is a simple yet powerful epigenetic mark for which its impact cannot be neglected in regulating fundamental biological process such as cell differentiation (Study 2), defining cell linage (Study 2) and development (Study 3) (Figure 58). Our data from these genome-wide studies have demonstrated that DNA methylation changes are tightly controlled both temporally and spatially, resulting in the uniqueness of methylated marks at each specific developmental time point and cell types.

DNA methylation have been implicated in both cell differentiation and cell lineage commitment. In our study (Study 2, Chapter 4), I have examined both questions in mouse adipogenic models. We found a predominance of hypermethylation during differentiation while a predominance of hypermethylation in WA compared to BA occurred in non-promoter regions. This strongly suggests that functional relevance of DNA methylation is highly context dependent.

I also demonstrated the hypothesis of context dependency DNA methylation effect on gene expression by investigating changes in DNA methylome across gestational ages in human placenta tissues (Study 3, Chapter 5). I performed an in-depth analysis by stratifying the genomic regions. On a broad picture, a positive correlation between gene expression and DNA methylation was found in gene bodies as opposed to negative correlation for promoter regions. Upon stratifying gene bodies by the CGI and exonic status, CGI exon and introns exhibited differences in correlation.

Conclusions made from the above studies have laid a foundation that DNA methylation is indeed important to fundamental process in organism survival. Thus, it became unsurprising that there was an aberrant genome-wide hypermethylation in the DS phenotype (Study 4, Chapter 6). In fact, genes with promoter hypermethylation were association with overall gene repression and were enriched in pathways related to neuronal activities. Down-regulation of genes involved in epigenetic regulation, particularly *TET* and *REST/NRSF*, may contribute to hypermethylation in DS phenotypes.
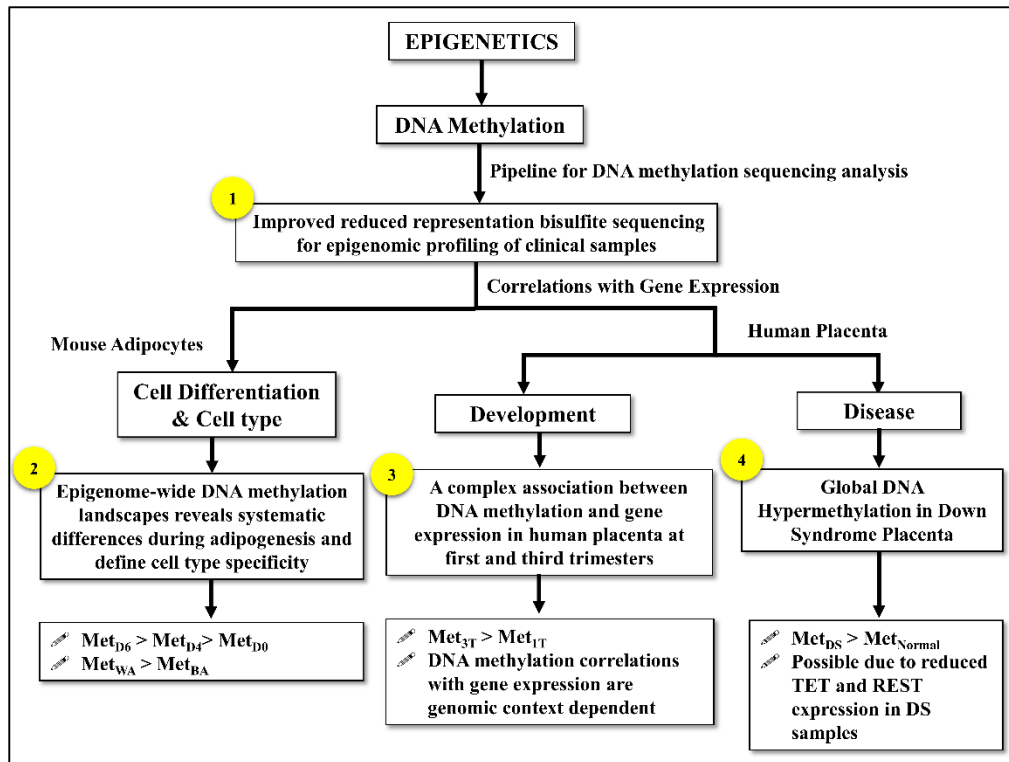
Figure 58: Schematic diagram binding four studies together

From the above conclusions, I have shown that DNA methylation is indeed important in various diverse biological processes by using bisulfite conversion techniques. However, this technique fails to distinguish between 5mC and 5hmC. Recent discovery of 5hmC as an oxidative product of 5mC via TET (Figure 1) [4-6] being highly abundant in brain than ESC has suggested that it is a stable epigenetic mark which might be important in carrying out cell specific mechanisms in the brain [164, 216-218]. Furthermore, recent studies have also found that the genomic locations of 5hmC are distinctly different between brain and ES cells which intrigues the field to understand more on the functions of 5hmC [164].

To distinguish between 5mC and 5hmC, Booth et al. developed oxBS-seq which makes use of a highly selective chemical oxidation to convert 5hmC to 5fC [219]. Subsequently, the converted 5fC will be deformylated and deaminated to uracil after bisulfite treatment. To identify 5mC alone, an independent performance of BS-seq

has to be performed and combining results from oxBS-seq, both 5mC and 5hmC can be discriminated [219].

Following detection of 5hmC, Booth et al developed another method, redBS-Seq to decode 5fC at single-base resolution [220]. Opposed to oxBS-seq, this technique is based on a selective chemical reduction of 5fC to 5hmC and followed by bisulfite treatment. By using a combination of oxBS-seq, BS-seq and redBS-seq, it is now possible to generate genome-wide landscape of 5mC, 5hmC and 5fC at single base resolution to provide biological insights of their respective functions [220].

Overall, I have established that epigenetic wide association studies (EWAS) via HTS coupled with bioinformatics analysis does bring us useful insights, such as biomarker discovery. In time to come, we will experience another paradigm shift from next generation sequencing to third generation sequencing such as the SMRT sequencing. These improved technologies, together with bioinformatics, will be crucial in helping us to seek answers to the many unknown questions in the epigenetics world.

BIBLIOGRAPHY

1.   Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9362-7.

2.   Ku, C.S., et al., *Studying the epigenome using next generation sequencing.* J Med Genet, 2011. **48**(11): p. 721-30.

3.   Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.

4.   Eichler, E.E., et al., *Missing heritability and strategies for finding the underlying causes of complex disease.* Nat Rev Genet, 2010. **11**(6): p. 446-50.

5.   Clarke, A.J. and D.N. Cooper, *GWAS: heritability missing in action?* Eur J Hum Genet, 2010. **18**(8): p. 859-61.

6.   Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.

7.   Gibney, E.R. and C.M. Nolan, *Epigenetics and gene expression.* Heredity (Edinb), 2010. **105**(1): p. 4-13.

8.   Wu, S.C. and Y. Zhang, *Active DNA demethylation: many roads lead to Rome.* Nat Rev Mol Cell Biol, 2010. **11**(9): p. 607-20.

9.   Zhang, G. and S. Pradhan, *Mammalian epigenetic mechanisms.* IUBMB Life, 2014. **66**(4): p. 240-56.

10.   Duncan, E.J., P.D. Gluckman, and P.K. Dearden, *Epigenetics, plasticity, and evolution: How do we link epigenetic change to phenotype?* J Exp Zool B Mol Dev Evol, 2014. **322**(4): p. 208-20.

11.   Fazzari, M.J. and J.M. Greally, *Introduction to epigenomics and epigenome-wide analysis.* Methods Mol Biol, 2010. **620**: p. 243-65.

12.   Laird, P.W., *Principles and challenges of genomewide DNA methylation analysis.* Nat Rev Genet, 2010. **11**(3): p. 191-203.

13.   Kohli, R.M. and Y. Zhang, *TET enzymes, TDG and the dynamics of DNA demethylation.* Nature, 2013. **502**(7472): p. 472-9.

14.   Sun, Z., et al., *Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis.* Epigenomics, 2015. **7**(5): p. 813-28.

15.   Gopalakrishnan, S., B.O. Van Emburgh, and K.D. Robertson, *DNA methylation in development and human disease.* Mutat Res, 2008. **647**(1-2): p. 30-8.

16.   Mikeska, T. and J.M. Craig, *DNA methylation biomarkers: cancer and beyond.* Genes (Basel), 2014. **5**(3): p. 821-64.

17.   Bock, C. and T. Lengauer, *Computational epigenetics.* Bioinformatics, 2008. **24**(1): p. 1-10.

18.   Newell-Price, J., A.J. Clark, and P. King, *DNA methylation and silencing of gene expression.* Trends Endocrinol Metab, 2000. **11**(4): p. 142-8.

19.   Prokhortchouk, E. and P.A. Defossez, *The cell biology of DNA methylation in mammals.* Biochim Biophys Acta, 2008. **1783**(11): p. 2167-73.

20.   Chen, T., et al., *Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b.* Mol Cell Biol, 2003. **23**(16): p. 5594-605.

21.   Sharif, J., et al., *The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA.* Nature, 2007. **450**(7171): p. 908-12.

22.   Bostick, M., et al., *UHRF1 plays a role in maintaining DNA methylation in mammalian cells.* Science, 2007. **317**(5845): p. 1760-4.

23.   Pradhan, S., et al., *Recombinant human DNA (cytosine-5) methyltransferase I. Expression, purification, and comparison of de novo and maintenance methylation.* Journal of Biological Chemistry, 1999. **274**(46): p. 33002-33010.

24.   Hermann, A., R. Goyal, and A. Jeltsch, *The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites.* Journal of Biological Chemistry, 2004. **279**(46): p. 48350-48359.

25.   Bestor, T.H., *Cloning of a mammalian DNA methyltransferase.* Gene, 1988. **74**(1): p. 9-12.

26.   Kim, J.K., M. Samaranayake, and S. Pradhan, *Epigenetic mechanisms in mammals.* Cell Mol Life Sci, 2009. **66**(4): p. 596-612.

27.   Li, E., T.H. Bestor, and R. Jaenisch, *Targeted mutation of the DNA methyltransferase gene results in embryonic lethality.* Cell, 1992. **69**(6): p. 915-26.

28.   Kim, M., et al., *Dnmt1 deficiency leads to enhanced microsatellite instability in mouse embryonic stem cells.* Nucleic Acids Res, 2004. **32**(19): p. 5742-9.

29.   Okano, M., et al., *DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.* Cell, 1999. **99**(3): p. 247-57.

30.   Hutnick, L.K., et al., *DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation.* Hum Mol Genet, 2009. **18**(15): p. 2875-88.

31.   Biniszkiewicz, D., et al., *Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality.* Mol Cell Biol, 2002. **22**(7): p. 2124-35.

32. Blake, W.J., et al., *Noise in eukaryotic gene expression.* Nature, 2003. **422**(6932): p. 633-7.

33. Fraser, H.B., et al., *Noise minimization in eukaryotic gene expression.* PLoS Biol, 2004. **2**(6): p. e137.

34. Citterio, E., et al., *Np95 is a histone-binding protein endowed with ubiquitin ligase activity.* Mol Cell Biol, 2004. **24**(6): p. 2526-35.

35. He, Y.F., et al., *Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.* Science, 2011. **333**(6047): p. 1303-7.

36. Bhutani, N., D.M. Burns, and H.M. Blau, *DNA demethylation dynamics.* Cell, 2011. **146**(6): p. 866-72.

37. Nafee, T.M., et al., *Epigenetic control of fetal gene expression.* BJOG, 2008. **115**(2): p. 158-68.

38. Bestor, T.H. and G.L. Verdine, *DNA methyltransferases.* Curr Opin Cell Biol, 1994. **6**(3): p. 380-9.

39. Lister, R., et al., *Human DNA methylomes at base resolution show widespread epigenomic differences.* Nature, 2009. **462**(7271): p. 315-22.

40. Schmidl, C., et al., *Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity.* Genome Res, 2009. **19**(7): p. 1165-74.

41. Rollins, R.A., et al., *Large-scale structure of genomic methylation patterns.* Genome Res, 2006. **16**(2): p. 157-63.

42. Bird, A.P., *DNA methylation and the frequency of CpG in animal DNA.* Nucleic Acids Res, 1980. **8**(7): p. 1499-504.

43. Robertson, K.D., *DNA methylation and human disease.* Nat Rev Genet, 2005. **6**(8): p. 597-610.

44. Bird, A., et al., *A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA.* Cell, 1985. **40**(1): p. 91-9.

45. Gardiner-Garden, M. and M. Frommer, *CpG islands in vertebrate genomes.* J Mol Biol, 1987. **196**(2): p. 261-82.

46. Saxonov, S., P. Berg, and D.L. Brutlag, *A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters.* Proc Natl Acad Sci U S A, 2006. **103**(5): p. 1412-7.

47. Jones, P.A., *Functions of DNA methylation: islands, start sites, gene bodies and beyond.* Nat Rev Genet, 2012. **13**(7): p. 484-92.

48. Kelly, T.K., et al., *H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes.* Mol Cell, 2010. **39**(6): p. 901-11.

49. Lock, L.F., N. Takagi, and G.R. Martin, *Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation.* Cell, 1987. **48**(1): p. 39-46.

50.   Gal-Yam, E.N., et al., *Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line.* Proc Natl Acad Sci U S A, 2008. **105**(35): p. 12979-84.

51.   Ohm, J.E., et al., *A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing.* Nat Genet, 2007. **39**(2): p. 237-42.

52.   Schlesinger, Y., et al., *Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer.* Nat Genet, 2007. **39**(2): p. 232-6.

53.   Widschwendter, M., et al., *Epigenetic stem cell signature in cancer.* Nat Genet, 2007. **39**(2): p. 157-8.

54.   Wade, P.A., *Methyl CpG-binding proteins and transcriptional repression.* Bioessays, 2001. **23**(12): p. 1131-7.

55.   Jones, P.L., et al., *Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription.* Nat Genet, 1998. **19**(2): p. 187-91.

56.   Jjingo, D., et al., *On the presence and role of human gene-body DNA methylation.* Oncotarget, 2012. **3**(4): p. 462-74.

57.   Laurent, L., et al., *Dynamic changes in the human methylome during differentiation.* Genome Res, 2010. **20**(3): p. 320-31.

58.   Hellman, A. and A. Chess, *Gene body-specific methylation on the active X chromosome.* Science, 2007. **315**(5815): p. 1141-3.

59.   Jones, P.A., *The DNA methylation paradox.* Trends Genet, 1999. **15**(1): p. 34-7.

60.   Brenet, F., et al., *DNA methylation of the first exon is tightly linked to transcriptional silencing.* PLoS One, 2011. **6**(1): p. e14524.

61.   Hodges, E., et al., *High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing.* Genome Res, 2009. **19**(9): p. 1593-605.

62.   Luco, R.F., et al., *Epigenetics in alternative pre-mRNA splicing.* Cell, 2011. **144**(1): p. 16-26.

63.   Shukla, S., et al., *CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing.* Nature, 2011. **479**(7371): p. 74-9.

64.   Schulz, W.A., C. Steinhoff, and A.R. Florl, *Methylation of endogenous human retroelements in health and disease.* Curr Top Microbiol Immunol, 2006. **310**: p. 211-50.

65.   Fazzari, M.J. and J.M. Greally, *Epigenomics: beyond CpG islands.* Nat Rev Genet, 2004. **5**(6): p. 446-55.

66.   Kuster, J.E., et al., *IAP insertion in the murine LamB3 gene results in junctional epidermolysis bullosa.* Mamm Genome, 1997. **8**(9): p. 673-81.

67.   Michaud, E.J., et al., *Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage.* Genes Dev, 1994. **8**(12): p. 1463-72.

68.   Gaudet, F., et al., *Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing.* Mol Cell Biol, 2004. **24**(4): p. 1640-8.

69.   Jin, B., Y. Li, and K.D. Robertson, *DNA methylation: superior or subordinate in the epigenetic hierarchy?* Genes Cancer, 2011. **2**(6): p. 607-17.

70.   Cedar, H. and Y. Bergman, *Linking DNA methylation and histone modification: patterns and paradigms.* Nat Rev Genet, 2009. **10**(5): p. 295-304.

71.   Seisenberger, S., et al., *Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1609): p. 20110330.

72.   Sanz, L.A., S.K. Kota, and R. Feil, *Genome-wide DNA demethylation in mammals.* Genome Biol, 2010. **11**(3): p. 110.

73.   Iqbal, K., et al., *Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3642-7.

74.   Jaenisch, R. and D. Jahner, *Methylation, expression and chromosomal position of genes in mammals.* Biochim Biophys Acta, 1984. **782**(1): p. 1-9.

75.   Auclair, G. and M. Weber, *Mechanisms of DNA methylation and demethylation in mammals.* Biochimie, 2012. **94**(11): p. 2202-11.

76.   Morgan, H.D., et al., *Epigenetic reprogramming in mammals.* Hum Mol Genet, 2005. **14 Spec No 1**: p. R47-58.

77.   Hajkova, P., et al., *Epigenetic reprogramming in mouse primordial germ cells.* Mech Dev, 2002. **117**(1-2): p. 15-23.

78.   Yamaguchi, S., et al., *Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming.* Cell Res, 2013. **23**(3): p. 329-39.

79.   Hackett, J.A., et al., *Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine.* Science, 2013. **339**(6118): p. 448-52.

80.   Kagiwada, S., et al., *Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice.* EMBO J, 2013. **32**(3): p. 340-53.

81.   Seisenberger, S., et al., *The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells.* Mol Cell, 2012. **48**(6): p. 849-62.

82.   Dawlaty, M.M., et al., *Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development.* Cell Stem Cell, 2011. **9**(2): p. 166-75.

83.   Ko, M., et al., *Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice.* Proc Natl Acad Sci U S A, 2011. **108**(35): p. 14566-71.

84.   Moran-Crusio, K., et al., *Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation.* Cancer Cell, 2011. **20**(1): p. 11-24.

85.   Dawlaty, M.M., et al., *Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development.* Dev Cell, 2013. **24**(3): p. 310-23.

86.   Paziewska, A., et al., *DNA methylation status is more reliable than gene expression at detecting cancer in prostate biopsy.* Br J Cancer, 2014. **111**(4): p. 781-9.

87.   Claus, R., et al., *Quantitative DNA methylation analysis identifies a single CpG dinucleotide important for ZAP-70 expression and predictive of prognosis in chronic lymphocytic leukemia.* J Clin Oncol, 2012. **30**(20): p. 2483-91.

88.   Issa, J.P., *DNA methylation as a clinical marker in oncology.* J Clin Oncol, 2012. **30**(20): p. 2566-8.

89.   Robinson, M.D., et al., *Statistical methods for detecting differentially methylated loci and regions.* Front Genet, 2014. **5**: p. 324.

90.   Houseman, E.A., et al., *DNA methylation arrays as surrogate measures of cell mixture distribution.* BMC Bioinformatics, 2012. **13**: p. 86.

91.   Gupta, R., A. Nagarajan, and N. Wajapeyee, *Advances in genome-wide DNA methylation analysis.* Biotechniques, 2010. **49**(4): p. iii-xi.

92.   Fouse, S.D., R.O. Nagarajan, and J.F. Costello, *Genome-scale DNA methylation analysis.* Epigenomics, 2010. **2**(1): p. 105-17.

93.   Akalin, A., et al., *methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles.* Genome Biol, 2012. **13**(10): p. R87.

94.   Weaver, I.C., et al., *Epigenetic programming by maternal behavior.* Nat Neurosci, 2004. **7**(8): p. 847-54.

95.   Jaffe, A.E., et al., *Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies.* Int J Epidemiol, 2012. **41**(1): p. 200-9.

96.   Sun, D., et al., *MOABS: model based analysis of bisulfite sequencing data.* Genome Biol, 2014. **15**(2): p. R38.

97.   Park, Y., et al., *MethylSig: a whole genome DNA methylation analysis pipeline.* Bioinformatics, 2014. **30**(17): p. 2414-22.

98. Dolzhenko, E. and A.D. Smith, *Using beta-binomial regression for high-precision differential methylation analysis in multifactor whole-genome bisulfite sequencing experiments.* BMC Bioinformatics, 2014. **15**: p. 215.

99. Gonzalez-Castejon, M. and A. Rodriguez-Casado, *Dietary phytochemicals and their potential effects on obesity: a review.* Pharmacol Res, 2011. **64**(5): p. 438-55.

100. Okamura, M., et al., *Role of histone methylation and demethylation in adipogenesis and obesity.* Organogenesis, 2010. **6**(1): p. 24-32.

101. Pinnick, K.E. and F. Karpe, *DNA methylation of genes in adipose tissue.* Proc Nutr Soc, 2011. **70**(1): p. 57-63.

102. Cannon, B. and J. Nedergaard, *Brown adipose tissue: function and physiological significance.* Physiol Rev, 2004. **84**(1): p. 277-359.

103. Park, A., W.K. Kim, and K.H. Bae, *Distinction of white, beige and brown adipocytes derived from mesenchymal stem cells.* World J Stem Cells, 2014. **6**(1): p. 33-42.

104. Alvarez-Dominguez, J.R., et al., *De Novo Reconstruction of Adipose Tissue Transcriptomes Reveals Long Non-coding RNA Regulators of Brown Adipocyte Development.* Cell Metab, 2015. **21**(5): p. 764-76.

105. Lee, Y.H., E.P. Mottillo, and J.G. Granneman, *Adipose tissue plasticity from WAT to BAT and in between.* Biochim Biophys Acta, 2014. **1842**(3): p. 358-69.

106. Harms, M. and P. Seale, *Brown and beige fat: development, function and therapeutic potential.* Nat Med, 2013. **19**(10): p. 1252-63.

107. Weijerman, M.E. and J.P. de Winter, *Clinical practice. The care of children with Down syndrome.* Eur J Pediatr, 2010. **169**(12): p. 1445-52.

108. Deitz SL, B.J., Solzak JP, Roper RJ, *Down Syndrome: A Complex and Interactive Genetic Disorder, Genetics and Etiology of Down Syndrome*, P.S. Dey, Editor. 2011, InTech.

109. Kahlem, P., et al., *Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of down syndrome.* Genome Res, 2004. **14**(7): p. 1258-67.

110. Lyle, R., et al., *Gene expression from the aneuploid chromosome in a trisomy mouse model of down syndrome.* Genome Res, 2004. **14**(7): p. 1268-74.

111. Mao, R., et al., *Global up-regulation of chromosome 21 gene expression in the developing Down syndrome brain.* Genomics, 2003. **81**(5): p. 457-67.

112. FitzPatrick, D.R., et al., *Transcriptome analysis of human autosomal trisomy.* Hum Mol Genet, 2002. **11**(26): p. 3249-56.

113. Costa, V., et al., *Massive-scale RNA-Seq analysis of non ribosomal transcriptome in human trisomy 21.* PLoS One, 2011. **6**(4): p. e18493.

114.   Rozovski, U., et al., *Genome-wide expression analysis of cultured trophoblast with trisomy 21 karyotype.* Hum Reprod, 2007. **22**(9): p. 2538-45.

115.   Berger, S.L., et al., *An operational definition of epigenetics.* Genes Dev, 2009. **23**(7): p. 781-3.

116.   Lande-Diner, L., et al., *Role of DNA methylation in stable gene repression.* J Biol Chem, 2007. **282**(16): p. 12194-200.

117.   Dahl, C., K. Gronbaek, and P. Guldberg, *Advances in DNA methylation: 5-hydroxymethylcytosine revisited.* Clin Chim Acta, 2011. **412**(11-12): p. 831-6.

118.   Bartolomei, M.S. and S.M. Tilghman, *Genomic imprinting in mammals.* Annu Rev Genet, 1997. **31**: p. 493-525.

119.   Li, E., C. Beard, and R. Jaenisch, *Role for DNA methylation in genomic imprinting.* Nature, 1993. **366**(6453): p. 362-5.

120.   Compere, S.J. and R.D. Palmiter, *DNA methylation controls the inducibility of the mouse metallothionein-I gene lymphoid cells.* Cell, 1981. **25**(1): p. 233-40.

121.   Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development.* Science, 1975. **187**(4173): p. 226-32.

122.   Meissner, A., *Epigenetic modifications in pluripotent and differentiated cells.* Nat Biotechnol, 2010. **28**(10): p. 1079-88.

123.   Weinhold, B., *Epigenetics: the science of change.* Environ Health Perspect, 2006. **114**(3): p. A160-7.

124.   Huang, C., et al., *Early life exposure to the 1959-1961 Chinese famine has long-term health consequences.* J Nutr, 2010. **140**(10): p. 1874-8.

125.   Herman, J.G. and S.B. Baylin, *Gene silencing in cancer in association with promoter hypermethylation.* N Engl J Med, 2003. **349**(21): p. 2042-54.

126.   Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nat Rev Genet, 2002. **3**(6): p. 415-28.

127.   Frommer, M., et al., *A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.* Proc Natl Acad Sci U S A, 1992. **89**(5): p. 1827-31.

128.   Meissner, A., et al., *Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.* Nucleic Acids Res, 2005. **33**(18): p. 5868-77.

129.   Meissner, A., et al., *Genome-scale DNA methylation maps of pluripotent and differentiated cells.* Nature, 2008. **454**(7205): p. 766-70.

130.   Smith, Z.D., et al., *High-throughput bisulfite sequencing in mammalian genomes.* Methods, 2009. **48**(3): p. 226-32.

131. Stadler, M.B., et al., *DNA-binding factors shape the mouse methylome at distal regulatory regions.* Nature, 2011. **480**(7378): p. 490-5.

132. Maunakea, A.K., et al., *Conserved role of intragenic DNA methylation in regulating alternative promoters.* Nature, 2010. **466**(7303): p. 253-7.

133. Jin, S., et al., *Global DNA hypermethylation in down syndrome placenta.* PLoS Genet, 2013. **9**(6): p. e1003515.

134. University of California, R. *IIGB HT Sequencing.* 27th April 2014]; Available from: http://illumina.bioinfo.ucr.edu/ht/documentation/file-formats.

135. Illumina. *Questions & Answers.* 27th April 2014]; Available from: https://support.illumina.com/sequencing/sequencing_instruments/genome_an alyzer_iix/questions.ilmn.

136. Lee, Y.K., et al., *Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples.* Biol Proced Online, 2014. **16**(1): p. 1.

137. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.

138. Ramsahoye, B.H., et al., *Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5237-42.

139. Ng, M., et al., *Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013.* Lancet, 2014. **384**(9945): p. 766-81.

140. Lloyd-Jones, D., et al., *Heart disease and stroke statistics--2009 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee.* Circulation, 2009. **119**(3): p. 480-6.

141. Bornfeldt, K.E. and I. Tabas, *Insulin resistance, hyperglycemia, and atherosclerosis.* Cell Metab, 2011. **14**(5): p. 575-85.

142. Sanchez-Gurmaches, J. and D.A. Guertin, *Adipocyte lineages: tracing back the origins of fat.* Biochim Biophys Acta, 2014. **1842**(3): p. 340-51.

143. Cai, L., et al., *The predicted effects of chronic obesity in middle age on medicare costs and mortality.* Med Care, 2010. **48**(6): p. 510-7.

144. Hammond, R.A. and R. Levine, *The economic impact of obesity in the United States.* Diabetes Metab Syndr Obes, 2010. **3**: p. 285-95.

145. Olshansky, S.J., et al., *A potential decline in life expectancy in the United States in the 21st century.* N Engl J Med, 2005. **352**(11): p. 1138-45.

146. Giralt, M. and F. Villarroya, *White, brown, beige/brite: different adipose cells for different functions?* Endocrinology, 2013. **154**(9): p. 2992-3000.

147.   Cypess, A.M., et al., *Identification and importance of brown adipose tissue in adult humans.* N Engl J Med, 2009. **360**(15): p. 1509-17.

148.   van Marken Lichtenbelt, W.D., et al., *Cold-activated brown adipose tissue in healthy men.* N Engl J Med, 2009. **360**(15): p. 1500-8.

149.   Virtanen, K.A., et al., *Functional brown adipose tissue in healthy adults.* N Engl J Med, 2009. **360**(15): p. 1518-25.

150.   Tsumura, A., et al., *Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b.* Genes Cells, 2006. **11**(7): p. 805-14.

151.   Schmidt, C.S., et al., *Global DNA hypomethylation prevents consolidation of differentiation programs and allows reversion to the embryonic stem cell state.* PLoS One, 2012. **7**(12): p. e52629.

152.   Noer, A., et al., *Stable CpG hypomethylation of adipogenic promoters in freshly isolated, cultured, and differentiated mesenchymal stem cells from adipose tissue.* Mol Biol Cell, 2006. **17**(8): p. 3543-56.

153.   Londono Gentile, T., et al., *DNMT1 is regulated by ATP-citrate lyase and maintains methylation patterns during adipocyte differentiation.* Mol Cell Biol, 2013. **33**(19): p. 3864-78.

154.   Su, J., et al., *Genome-wide dynamic changes of DNA methylation of repetitive elements in human embryonic stem cells and fetal fibroblasts.* Genomics, 2012. **99**(1): p. 10-7.

155.   Hondares, E., et al., *Peroxisome proliferator-activated receptor alpha (PPARalpha) induces PPARgamma coactivator 1alpha (PGC-1alpha) gene expression and contributes to thermogenic activation of brown fat: involvement of PRDM16.* J Biol Chem, 2011. **286**(50): p. 43112-22.

156.   Scheele, C., T.J. Larsen, and S. Nielsen, *Novel nuances of human brown fat.* Adipocyte, 2014. **3**(1): p. 54-7.

157.   Foronda, D., et al., *Function and specificity of Hox genes.* Int J Dev Biol, 2009. **53**(8-10): p. 1404-19.

158.   Procino, A. and C. Cillo, *The HOX genes network in metabolic diseases.* Cell Biol Int, 2013. **37**(11): p. 1145-8.

159.   Benton, M.C., et al., *An analysis of DNA methylation in human adipose tissue reveals differential modification of obesity genes before and after gastric bypass and weight loss.* Genome Biol, 2015. **16**: p. 8.

160.   Kulis, M., et al., *Whole-genome fingerprint of the DNA methylome during human B cell differentiation.* Nat Genet, 2015.

161.   He, D., et al., *Differentiation of PDX1 gene-modified human umbilical cord mesenchymal stem cells into insulin-producing cells in vitro.* Int J Mol Med, 2011. **28**(6): p. 1019-24.

162.	Jin, S.G., et al., *Genomic mapping of 5-hydroxymethylcytosine in the human brain.* Nucleic Acids Res, 2011. **39**(12): p. 5015-24.

163.	Song, C.X., et al., *Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine.* Nat Biotechnol, 2011. **29**(1): p. 68-72.

164.	Szulwach, K.E., et al., *5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging.* Nat Neurosci, 2011. **14**(12): p. 1607-16.

165.	Colquitt, B.M., et al., *Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity.* Proc Natl Acad Sci U S A, 2013. **110**(36): p. 14682-7.

166.	Flusberg, B.A., et al., *Direct detection of DNA methylation during single-molecule, real-time sequencing.* Nat Methods, 2010. **7**(6): p. 461-5.

167.	Pijnenborg, R., et al., *Attachment and differentiation in vitro of trophoblast from normal and preeclamptic human placentas.* Am J Obstet Gynecol, 1996. **175**(1): p. 30-6.

168.	Rogers, B.B., S.L. Bloom, and K.J. Leveno, *Atherosis revisited: current concepts on the pathophysiology of implantation site disorders.* Obstet Gynecol Surv, 1999. **54**(3): p. 189-95.

169.	Barker, D.J., et al., *Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth.* Diabetologia, 1993. **36**(1): p. 62-7.

170.	Boyko, E.J., *Proportion of type 2 diabetes cases resulting from impaired fetal growth.* Diabetes Care, 2000. **23**(9): p. 1260-4.

171.	Eriksson, J.G., et al., *Pathways of infant and childhood growth that lead to type 2 diabetes.* Diabetes Care, 2003. **26**(11): p. 3006-10.

172.	Curhan, G.C., et al., *Birth weight and adult hypertension and obesity in women.* Circulation, 1996. **94**(6): p. 1310-5.

173.	Ross, M.G. and M.H. Beall, *Adult sequelae of intrauterine growth restriction.* Semin Perinatol, 2008. **32**(3): p. 213-8.

174.	Jansson, T. and T.L. Powell, *Role of the placenta in fetal programming: underlying mechanisms and potential interventional approaches.* Clin Sci (Lond), 2007. **113**(1): p. 1-13.

175.	Maccani, M.A. and C.J. Marsit, *Epigenetics in the placenta.* Am J Reprod Immunol, 2009. **62**(2): p. 78-89.

176.	John, R. and M. Hemberger, *A placenta for life.* Reprod Biomed Online, 2012. **25**(1): p. 5-11.

177.	Sood, R., et al., *Gene expression patterns in human placenta.* Proc Natl Acad Sci U S A, 2006. **103**(14): p. 5478-83.

178.	Das, R., et al., *DNMT1 and AIM1 Imprinting in human placenta revealed through a genome-wide screen for allele-specific DNA methylation.* BMC Genomics, 2013. **14**: p. 685.

179.	Magalhaes, H.R., et al., *Placental hydroxymethylation vs methylation at the imprinting control region 2 on chromosome 11p15.5.* Braz J Med Biol Res, 2013. **46**(11): p. 916-919.

180.	Buckberry, S., et al., *Quantitative allele-specific expression and DNA methylation analysis of H19, IGF2 and IGF2R in the human placenta across gestation reveals H19 imprinting plasticity.* PLoS One, 2012. **7**(12): p. e51210.

181.	Barbaux, S., et al., *A genome-wide approach reveals novel imprinted genes expressed in the human placenta.* Epigenetics, 2012. **7**(9): p. 1079-90.

182.	Lambertini, L., et al., *Imprinted gene expression in fetal growth and development.* Placenta, 2012. **33**(6): p. 480-6.

183.	Lambertini, L., et al., *Differential methylation of imprinted genes in growth-restricted placentas.* Reprod Sci, 2011. **18**(11): p. 1111-7.

184.	Novakovic, B., et al., *Evidence for widespread changes in promoter methylation profile in human placenta in response to increasing gestational age and environmental/stochastic factors.* BMC Genomics, 2011. **12**: p. 529.

185.	Zhang, Y., et al., *DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution.* PLoS Genet, 2009. **5**(3): p. e1000438.

186.	Eckhardt, F., et al., *DNA methylation profiling of human chromosomes 6, 20 and 22.* Nat Genet, 2006. **38**(12): p. 1378-85.

187.	Moore, G.E., et al., *The role and interaction of imprinted genes in human fetal growth.* Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1663): p. 20140074.

188.	Diplas, A.I., et al., *Differential expression of imprinted genes in normal and IUGR human placentas.* Epigenetics, 2009. **4**(4): p. 235-40.

189.	McMinn, J., et al., *Unbalanced placental expression of imprinted genes in human intrauterine growth restriction.* Placenta, 2006. **27**(6-7): p. 540-9.

190.	Yu, L., et al., *The H19 gene imprinting in normal pregnancy and pre-eclampsia.* Placenta, 2009. **30**(5): p. 443-7.

191.	Tycko, B., *Imprinted genes in placental growth and obstetric disorders.* Cytogenet Genome Res, 2006. **113**(1-4): p. 271-8.

192.	Spence, D., et al., *Intra-uterine growth restriction and increased risk of hypertension in adult life: a follow-up study of 50-year-olds.* Public Health, 2012. **126**(7): p. 561-5.

193. Hecher, K., et al., *Assessment of fetal compromise by Doppler ultrasound investigation of the fetal circulation. Arterial, intracardiac, and venous blood flow velocity studies.* Circulation, 1995. **91**(1): p. 129-38.

194. Crispi, F., et al., *Cardiac dysfunction and cell damage across clinical stages of severity in growth-restricted fetuses.* Am J Obstet Gynecol, 2008. **199**(3): p. 254 e1-8.

195. Comas, M., et al., *Usefulness of myocardial tissue Doppler vs conventional echocardiography in the evaluation of cardiac dysfunction in early-onset intrauterine growth restriction.* Am J Obstet Gynecol, 2010. **203**(1): p. 45 e1-7.

196. Fattal-Valevski, A., et al., *Growth patterns in children with intrauterine growth retardation and their correlation to neurocognitive development.* J Child Neurol, 2009. **24**(7): p. 846-51.

197. Rodriguez, M.M., et al., *Histomorphometric analysis of postnatal glomerulogenesis in extremely preterm infants.* Pediatr Dev Pathol, 2004. **7**(1): p. 17-25.

198. Teeninga, N., et al., *Influence of low birth weight on minimal change nephrotic syndrome in children, including a meta-analysis.* Nephrol Dial Transplant, 2008. **23**(5): p. 1615-20.

199. Ranweiler, R., *Assessment and care of the newborn with Down syndrome.* Adv Neonatal Care, 2009. **9**(1): p. 17-24; Quiz 25-6.

200. Hook, E.B., *Unbalanced Robertsonian translocations associated with Down's syndrome or Patau's syndrome: chromosome subtype, proportion inherited, mutation rates, and sex ratio.* Hum Genet, 1981. **59**(3): p. 235-9.

201. Nadal, M., et al., *Clinical and cytogenetic characterisation of a patient with Down syndrome resulting from a 21q22.1-->qter duplication.* J Med Genet, 2001. **38**(1): p. 73-6.

202. Megarbane, A., et al., *The 50th anniversary of the discovery of trisomy 21: the past, present, and future of research and treatment of Down syndrome.* Genet Med, 2009. **11**(9): p. 611-6.

203. Parker, S.E., et al., *Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006.* Birth Defects Res A Clin Mol Teratol, 2010. **88**(12): p. 1008-16.

204. Antonarakis, S.E., et al., *Chromosome 21 and down syndrome: from genomics to pathophysiology.* Nat Rev Genet, 2004. **5**(10): p. 725-38.

205. Ait Yahya-Graison, E., et al., *Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes.* Am J Hum Genet, 2007. **81**(3): p. 475-91.

206. Jiang, J., et al., *Translating dosage compensation to trisomy 21.* Nature, 2013. **500**(7462): p. 296-300.

207. Feng, J., et al., *Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons.* Nat Neurosci, 2010. **13**(4): p. 423-30.

208. Guo, J.U., et al., *Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain.* Cell, 2011. **145**(3): p. 423-34.

209. Sanchez-Mut, J.V., D. Huertas, and M. Esteller, *Aberrant epigenetic landscape in intellectual disability.* Prog Brain Res, 2012. **197**: p. 53-71.

210. Pogribna, M., et al., *Homocysteine metabolism in children with Down syndrome: in vitro modulation.* Am J Hum Genet, 2001. **69**(1): p. 88-95.

211. Eckmann-Scholz, C., et al., *DNA-methylation profiling of fetal tissues reveals marked epigenetic differences between chorionic and amniotic samples.* PLoS One, 2012. **7**(6): p. e39014.

212. Jaffe, A.E., et al., *Significance analysis and statistical dissection of variably methylated regions.* Biostatistics, 2012. **13**(1): p. 166-78.

213. Kerkel, K., et al., *Altered DNA methylation in leukocytes with trisomy 21.* PLoS Genet, 2010. **6**(11): p. e1001212.

214. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* Science, 2009. **324**(5929): p. 930-5.

215. Ito, S., et al., *Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine.* Science, 2011. **333**(6047): p. 1300-3.

216. Santiago, M., et al., *TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they?* Genomics, 2014. **104**(5): p. 334-40.

217. Kriaucionis, S. and N. Heintz, *The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain.* Science, 2009. **324**(5929): p. 929-30.

218. Munzel, M., et al., *Quantification of the sixth DNA base hydroxymethylcytosine in the brain.* Angew Chem Int Ed Engl, 2010. **49**(31): p. 5375-7.

219. Booth, M.J., et al., *Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine.* Nat Protoc, 2013. **8**(10): p. 1841-51.

220. Booth, M.J., et al., *Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution.* Nat Chem, 2014. **6**(5): p. 435-40.