Methods to investigate 5-hydroxymethylcytosine in monozygotic twins discordant for Psoriasis

Masters Thesis by Aase Marie Cohen Kiel

Department of Molecular Biosciences
Faculty of Mathematics and Natural Sciences

UNIVERSITETET I OSLO

March 2013

2013

Methods to investigate 5-hydroxymethylcytosine in monozygotic twins discordant for Psoriasis

Supervisors: Robert Lyle, Gregor Gilfillan

http://www.duo.uio.no

Trykk: Reprosentralen, Universitetet i Oslo

Acknowledgements

The research for this thesis was done at the Department of Medical Genetics at Oslo University Hospital, Ullevål, in collaboration with the Norwegian Sequencing Center.

First and foremost I would like to thank my supervisors at Ullevål, Dr. Robert Lyle and Dr. Gregor Gilfillan for giving me the opportunity to work on this project. Without your guidance and assistance, especially during the trying times, this project would never have come to fruition. I would also like to thank my internal supervisor at UiO Prof. Paul Grini for support throughout the last year.

I would also like to acknowledge John-Arne Dahl and Arne Klungland at Rikshospitalet for the never ending supply of mouse brain DNA as well as my own mouse brains. Thank you to everyone at the Department of Medical Genetics for taking an interest in my project, and encouragement when nothing seemed to be working, especially my officemates Maria and Yngvild for the frequent breaks and entertaining discussions. Andreas Moan, thank you so much for getting me in touch with MedGen during my bachelor studies, if it wasn't for you I wouldn't be where I am today.

Lastly, I would like to thank all of my friends and family who have supported me through this process by trying to understand my project, giving immeasurable amounts of support, and keeping me well fed.

Abstract

While investigating methylcytosine content in Purkinjie neurons in 2009, researchers uncovered a previously forgotten base, 5-hydroxymethylcytosine (5hmC) in high concentrations. Since then, 5hmC has been found in various tissue types and has been implicated in biological processes such as development, gene regulation, DNA demethylation and tumorigenesis, leading researchers to conclude that 5hmC could be an epigenetic mark.

The lab where this research was conducted has previously studied methylcytosine in a cohort of monozygotic twins discordant for psoriasis, and in this project we aimed to do the same, only this time focusing on 5hmC. Unfortunately, due to technical hurdles, time did not allow study of 5hmC levels in the discordant twins. We were however able to show the presence of 5hmC in both mouse brain and a subset of monozygotic control twins using an ELISA assay. Furthermore, we demonstrated a novel genome-wide method to examine 5hmC at a subset of CpG dinucleotides, through the combined use of Zymo Research Corporation´s Quest 5hmC Detection Kit, appropriate size selection and high-throughput-sequencing.

Using this novel method, we were able to identify a small number of differentially hydroxymethylated peaks in our twin samples using MACS (Model-based analysis for ChIP-seq). Further sequencing to achieve a greater read depth is required to fully validate the method, but the method could conceivably now be applied towards the original goal of finding sites of differential hydroxymethylation in monozygotic twins discordant for Psoriasis.

Abbreviations

5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
BER	Base Excision Repair
BS	Bisulfite Sequencing
CGI	CpG Island
CMS	Cytosine 5-methylenesulfonate
DMR	Differentially Methylated Region
DNMT	DNA Methyltransferase
DZ	Dizygotic
ELISA	Enzyme-linked Immunosorbent Assay
ES cells	Embryonic stem cells
EtOH	Ethanol
gDNA	Genomic DNA
GLIB	Glycosylation Periodate Oxidation, Biotinylation
hMeDIP	Hydroxymethylated DNA Immunoprecipitation
HPLC	High Pressure Liquid Chromatography
JBP	J-base Binding Protein
$KRuO_4$	Potassium Perruthenate
MACS	Model-based analysis for ChIP-seq
MBD	Methyl-binding domain
MeDIP	Methylated DNA Immunoprecipitation
miRNA	micro RNA
MS	Mass Spectrometry
MZ	Monozygotic
ncRNA	Non-coding RNA
NTC	Non-Template Control
Ox-BS-Seq	Oxidative Bisulfite Sequencing
RRBS	Reduced Representation Bisulfite Sequencing
qPCR	Quantitative (Real-Time) Polymerase Chain Reaction
SNP	Single Nucleotide Polymorphism
TAB-Seq	TET Assisted Bisulfite sequencing
TDG	Thymine DNA glycosylase
TET	Ten-eleven Translocase
TSS	Transcription Start Site

Table of Contents

Acknowledgements V

Abstract VII

Abbreviations IX

1 Introduction 2

1.1 5-hydroxymethylcytosine 2

1.1.1 Biological Pathways 3

1.1.2 5-hydroxymethylcytosine and Development 5

1.1.3 5-hydroxymethylcytosine in Disease 7

1.2 Epigenetics 8

1.2.1 Epigenetics & Development 10

1.2.2 Epigenetics & Disease 11

1.2.3 Epigenetics & Psoriasis 12

1.3 Twin Studies 13

1.4 Investigating 5-hydroxymethylcytosine 15

1.4.1 Isolation and Enrichment of 5hmC 15

1.4.2 Direct detection of 5-hydroxymethylcytosine by DNA sequencing 16

1.5 Aims of the Thesis 19

2 Methods and Materials 20

2.1 ELISA – Quantification of 5hmC 20

2.2 Hydroxymethylated Immunoprecipitation 21

2.2.1 hMeDIP using Zymo Research Quest 5hmC DNA Enrichment Kit 22

2.2.2 Hydroxymethylated Immunoprecipitation using Active Motif hMeDIP Kit 22

2.2.3 Purification of DNA after hMeDIP 24

2.3 Oxidative Reduced Representation Bisulfite Sequencing 24

2.4 Sequence Specific Detection of 5-hydroxymethylcytosine 25

2.4.1 Preparation of control DNA from mouse brain 25

2.4.2 Sequence and Locus Specific Detection of 5hmC using Zymo Research Quest 5hmC Detection Kit 25

2.4.3 Preparation for Sequencing 26

2.5 qPCR Validation of Experiments 28

2.6 Sequencing 29

2.7 Analysis of Sequencing Results 29

3 Results 30

3.1 ELISA 30

3.2 Hydroxymethylated DNA Immunoprecipitation 31

3.2.1 Zymo Research Quest 5hmC Enrichment 32

3.2.2 Active Motif hMeDIP 33

3.3 Sequence Specific Detection of Hydroxymethylcytosine 36

3.3.1 DNA Isolation 36

3.3.2 Detection of 5-hydroxymethylcytosine using Zymo Quest-5hmC Detection Kit 37

3.3.3 Library Preparation 40

3.4 Sequencing 43

3.4.1 Quality of the Sequenced Reads 43

3.5 Sequencing Analysis 44

3.5.1 Mouse Brain Control ... 46
3.5.2 Control Twin Samples ... 47

4 Discussion ... 50
4.1 ELISA ... 50
4.2 hMeDIP ... 50
4.3 Sequence Specific Detection of 5hmC ... 52
4.4 Sequencing ... 53
4.5 Sequencing Analysis ... 53
4.6 Conclusion ... 54
4.7 Further Research ... 54

5 Appendix ... 56
5.1 Protocols ... 56
5.1.1 DNA Quantification ... 56
5.1.2 DNA Isolation, Cleanup and Concentration ... 57
5.1.3 PCR & qPCR Reactions ... 58
Standard qPCR Reaction ... 58
Zymo Quest 5hmC Enrichment Kit ... 59
5.1.4 Illumina Library Preparation ... 60
5.2 Primers ... 61
5.3 Sequencing Data ... 62

6 References ... 64

1 Introduction

1.1 5-hydroxymethylcytosine

5-hydroxymethylcytosine (5hmC) was first described in 1952 as being the sole cytosine base of bacteriophage T2 (Hershey and Chase 1952). Later, it was found in T4 bacteriophage, and shown to protect progeny phage DNA from degradation by host nuclease activity, which can degrade unmodified cytosines (Wiberg 1967). 5hmC was then found in mammalian brain and liver tissue (Penn, Suwalski et al. 1972), but until recently, its function hasn't been further researched in mammals due to technological limitations. While comparing the abundance of 5mC in Purkinje and granule cell nuclei, 5hmC was identified in mouse cerebellar DNA, as well as in many other tissue types, but at much lower concentrations. At the same time it was concluded that 5hmC could possibly be an epigenetic mark. (Kriaucionis and Heintz 2009).

5hmC has been found in its highest concentrations in the brain followed by breast and liver (Branco, Ficz et al. 2012). It is also present throughout all the other tissues in the body; and it was found that tissue type, not correlation to 5mC or TET expression (the enzymes responsible for 5hmC synthesis), predicts 5hmC concentration (Nestor, Ottaviano et al. 2012). However, locus specific enrichment is commonly found at the *H19/IGF2* and *HOXA* loci.

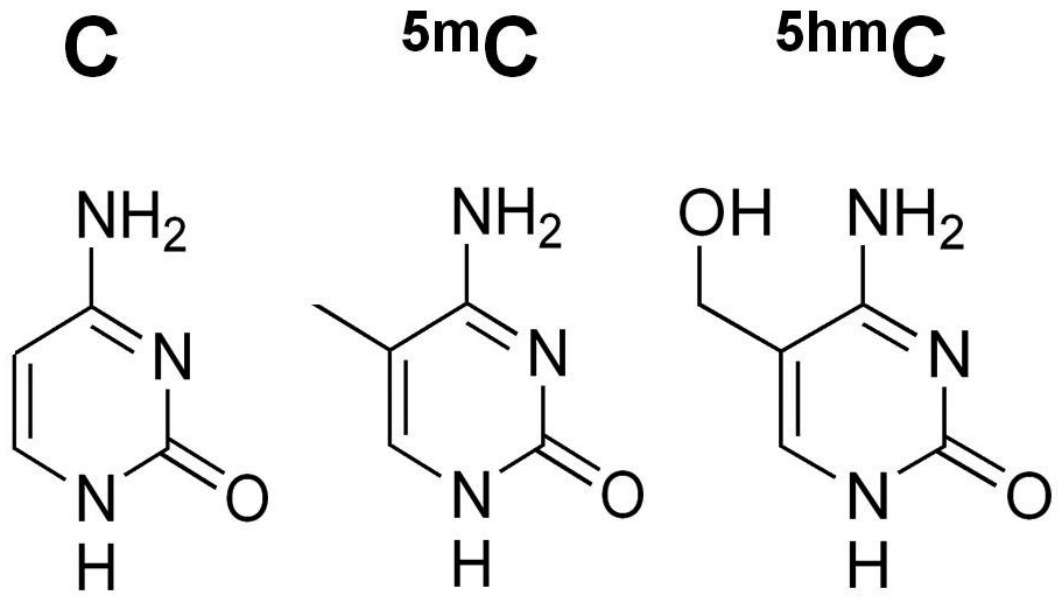


Figure 1: Structures of C, 5mC and 5hmC (Szulwach, Li et al. 2011)

1.1.1 Biological Pathways

5-hydroxymyethylcytosine is a further modified version of 5-methylcytosine; it contains a hydroxyl group connected to the methyl group at position 5 on the pyrimidine ring (Figure 1). Oxidation of the methyl group on 5mC converts it to 5hmC, and has been found to be facilitated by Ten-eleven translocation 1-3 (TET 1-3) proteins (Tahiliani, Koh et al. 2009; Ito, D'Alessio et al. 2010). The TET 1-3 family of proteins are 2-oxoglutarate (2OG) and Fe (II)-dependent dioxygenase enzymes, all of which have the capacity to catalyze the conversion of 5mC to 5hmC. They are also able to further oxidize 5hmC to 5-formylcytosine (5-fC), and 5-carboxycytosine (5-caC) (He, Li et al. 2011) (Figure 2).

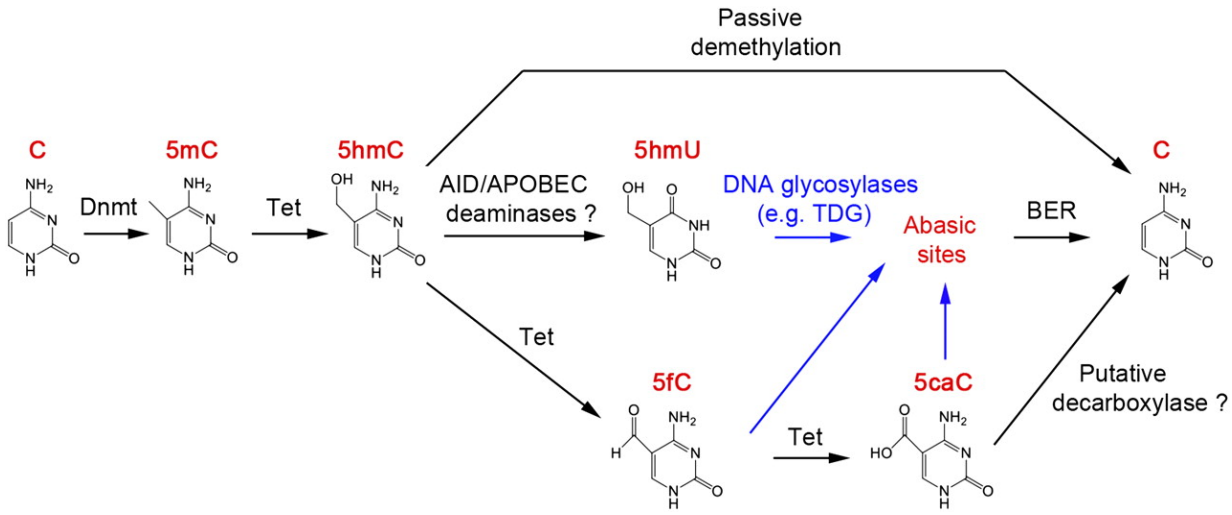


Figure 2: Proposed models of passive and active DNA demethylation (Wu and Zhang 2011)

All TET proteins contain a cysteine-rich region, a double-stranded β-helix fold, and a CXXC domain (Figure 3) – a DNA binding domain that has been shown to be a CpG-binding motif, and is likely involved in the recruitment of TET proteins to DNA (Szwagierczak, Bultmann et al. 2010). TET1 has been shown to mainly bind to gene rich regions of the genome, showing a preference for transcription start sites (TSSs). It has also been found to bind to CpG islands and bivalent promoters (ones simultaneously marked with both activating and repressing modifications) in mouse embryonic stem (ES) cells, consistent with the significant enrichment of 5hmC, but not 5mC at bivalent promoters (Matarese, Carrillo-de Santa Pau et al. 2011). TET1 binding also shows a near-perfect overlap with DNAse-1 hypersensitive sites found in ES cells – possibly providing evidence for activity of TET1 in transcriptional

regulation (Levasseur, Wang et al. 2008). Though TET1 is the primary enzyme for oxidation of 5mC to 5-fC, the other TET proteins, TET2 and TET3 are able to compensate for loss of TET1. The knockout of TET1 in ES cells leads to a slight reduction of 5hmC levels, but due to the compensatory action of TET2, pluripotency is not affected (Wossidlo, Nakamura et al. 2011). Similarly, TET3 has been shown to convert 5mC to 5hmC in the paternal pronucleus of mice (He, Li et al. 2011). Whereas TET proteins have distinct non-redundant roles in meiosis (Tan and Manley 2009), it is still unclear as to whether or not they have a redundant role in regards to 5hmC synthesis and further oxidation to 5-fC.

Figure 3: Conserved domains of TET proteins (Williams, Christensen et al. 2012)

Evidence suggests that both passive and active DNA demethylation may be facilitated by the oxidation of 5mC to 5hmC by TET proteins, as this can alter both global and locus-specific levels of 5mC (Wu and Zhang 2011). Proteins containing methyl-binding domains (MBD), such as methyl-CpG-binding protein 2 (MeCP2), and Dnmt1 (DNA methyltransferase) are unable to recognize 5hmC (Valinluck, Tsai et al. 2004; Valinluck and Sowers 2007; Jin, Kadam et al. 2010). Thus, when encountering a 5hmC instead of 5mC, Dnmt1 will be unable to maintain existing methylation patterns, leading to passive DNA demethylation. Though it is still unclear how 5hmC leads to active demethylation, multiple models have been proposed. The model with the most support proposes that oxidation of 5mC to 5hmC and further oxidation to 5fC and 5caC is followed by base excision repair (BER). This model is supported by data showing that Thymine DNA glycosylase (TDG) can efficiently remove 5fC and 5caC at CpG sites, and subsequent base repair introduces an unmodified cytosine (Maiti and Drohat 2011). Additional support for this model comes from the finding that loss of TDG in mice leads to elevated levels of methylation, and subsequent death (Cortellino, Xu et al. 2011). Moreover, the DNA methyltransferases Dnmt3a and Dnmt3b have been shown to have DNA dehydroxymethylase activity (Chen, Wang et al. 2012). A recent study of primordial germ cells showed that repression of Dnmt3a/Dnmt3b led to the prevention of re-methylation, thereby allowing the seemingly functionally redundant TET1/2 to further oxidate 5mC to 5hmC (Hackett, Sengupta et al. 2013).

To gain more insight on how TET1 and 5hmC affect transcription, multiple laboratories have mapped the genome-wide occupancy of TET1 in mouse ES cells (Williams, Christensen et al. 2011; Wu and Zhang 2011; Xu, Wu et al. 2011), and a comparison of all the data sets showed that 90% of the TET1 target genes overlap (Wu and Zhang 2011). In these datasets, it was shown that TET1 binds preferentially to CpG islands, and gene rich regions at TSSs and promoter regions, all of which have a high concentration of 5hmC. Gene promoters associated with bivalent domains that bear H3K4me3 and H3K27me3 marks are associated with both TET1 and 5hmC (Pastor, Pape et al. 2011). 5hmC was shown to be enriched at gene bodies of highly transcribed genes and at the promoters of PRC2-repressed genes (Ficz, Branco et al. 2011), supporting the possibility of it having a role in both transcriptional activation and repression. A recent study has proposed that when 5hmC is located in promoters it has an inhibitory role in regards to transcription, but when in the gene body, it has a neutral effect (Robertson, Robertson et al. 2011). There is a great deal of data showing that 5hmC is involved in regulation of transcription, but the mechanisms have not been clearly elucidated.

1.1.2 5-hydroxymethylcytosine and Development

While the status of DNA methylation during development is now well known, the role of 5hmC during development is still unclear. 5hmC has been shown to accumulate in the paternal pronucleus of mice, coinciding with the loss of 5mC (Iqbal, Jin et al. 2011) (Figure 4). The same data has also been found in both rabbit and bovine zygotes (Wossidlo, Nakamura et al. 2011); suggesting that 5hmC plays a role in DNA demethylation during zygote development.

TET3 has been shown to be expressed at high levels in oocytes and zygotes, but not in later stages of development (Figure 4). Knockdown or siRNA-mediated down-regulation of TET3 in female zygotes has been shown to block the oxidation of 5mC to 5hmC and leads to reduced developmental fitness and fetal survival (Gu, Guo et al. 2011), providing further evidence that oxidation of 5mC is an important event during development.

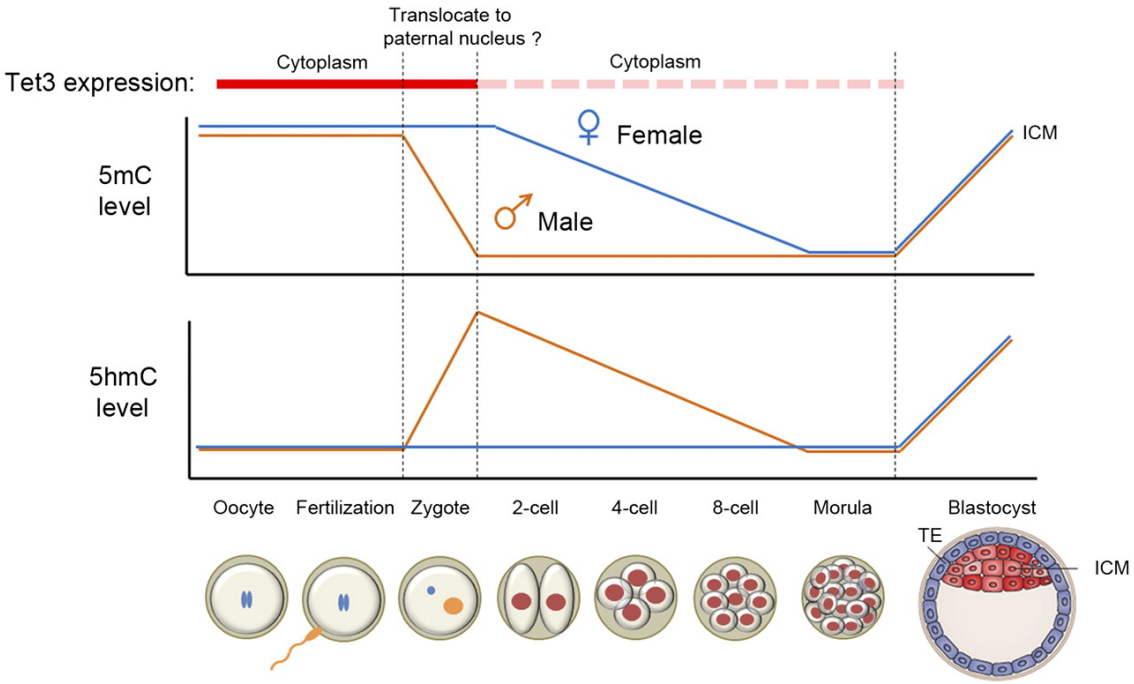


Figure 4: TET3 expression and amount of 5mC/5hmC present during development (Wu and Zhang 2011)

In the blastocyst stage of development in mouse zygotes, both TET1 and 5hmC are abundant (Ito, D'Alessio et al. 2010), but knockdown of TET1 still leads to differentiation of mouse ES cells, and TET1 null mutant ES cells are still able to proliferate under normal cell culture conditions (Dawlaty, Ganz et al. 2011). Because TET1 binds to numerous gene promoters of ES cell transcription factors, it is likely that TET1 is involved in the regulation of pluripotency (Wu and Zhang 2011). An RNA-seq analysis has confirmed that a TET1 and TET2 double knockdown in mouse ES cells led to down-regulation of several pluripotency factors (Ficz, Branco et al. 2011). However, other studies have shown that knockdown of TET1/TET2 do not affect pluripotency (Koh, Yabuuchi et al. 2011; Williams, Christensen et al. 2011). These discrepancies lead to an unclear understanding of the role of TET1 in development and pluripotency, though there may be other mechanisms that compensate for the loss of TET1/TET2 (Hackett, Sengupta et al. 2012).

As 5hmC is found in its highest concentrations in brain tissue, it is apparent that 5hmC must play a role in neurodevelopment (Nestor, Ottaviano et al. 2012). Enrichment of 5hmC at gene bodies has been shown to be positively correlated with gene expression in the frontal lobe tissue of the human brain (Jin, Wu et al. 2011). Purkinjie neurons, which are considered to be a primary organizer in the human brain, contain a greater amount of 5hmC than other

neurons, providing evidence for its function in neurodevelopment, and its disregulation may contribute to the pathogenesis of neurodevelopmental disorders (Wang, Pan et al. 2012).

1.1.3 5-hydroxymethylcytosine in Disease

Hypermethylation of promoters is an epigenetic hallmark of many types of cancer. It leads to the silencing of tumor suppressor genes, thereby enabling cancer proliferation (Jones and Laird 1999). DNA methyltransferases (DNMTs), which catalyze DNA methylation, have been correlated to poor prognosis in human cancers when their activity is increased.

Whereas deregulation of DNA methylation leading to abnormal gene expression is common in cancer, hydroxymethylation is always decreased (Kinney and Pradhan 2013), though its biological significance is still unknown. Decrease in hydroxymethylation is most likely due to loss of activity of TET enzymes. It has been reported that a loss-of-function mutation in the *TET2* gene leads to myeloid malignancies; suggesting that compromised DNA demethylation can play a role in tumorigenesis (Kudo, Tateishi et al. 2012). Solid tumors have also been shown to have lower 5hmC levels than normal tissues (Haffner, Chaux et al. 2011).

Using mammalian cells, it was shown that TET1 expression was decreased in 50% of colorectal cancers, and 73% of the tumors investigated had a reduced level of 5hmC in comparison to neighboring tissues. Reduced TET1 expression was also shown to be tightly associated with a decrease of TET3 mRNA. 88% of tumors that showed no *TET1* gene regulation showed increased expression of DNMT genes. It is likely that the suppression of TET1, as well as upregulation of DNMT1 affects the amount of 5hmC in cancerous tumors (Kudo, Tateishi et al. 2012).

Along with its disregulation in tumorigenesis, *TET2* mutations have been identified in B- and T-cell lymphoma (Quivoron, Couronne et al. 2011), and are primarily associated with a state of DNA hypermethylation, and low 5hmC levels (Figueroa, Abdel-Wahab et al. 2010). As seen in acute myeloid leukemia (AML), the function of TET2 can also be inhibited by *IDH1/2* mutations, which produce 2-hydroxyglutarate (2-HG) and compete with TET2's substrate 2-OG (Cimmino, Abdel-Wahab et al. 2011). Both *TET2* mutations and *IDH1/2*

mutations prevent oxidation of 5mC and lead to an increase in global 5mC levels (Figure 5) (Dang, White et al. 2009; Figueroa, Abdel-Wahab et al. 2010; Gross, Cairns et al. 2010).

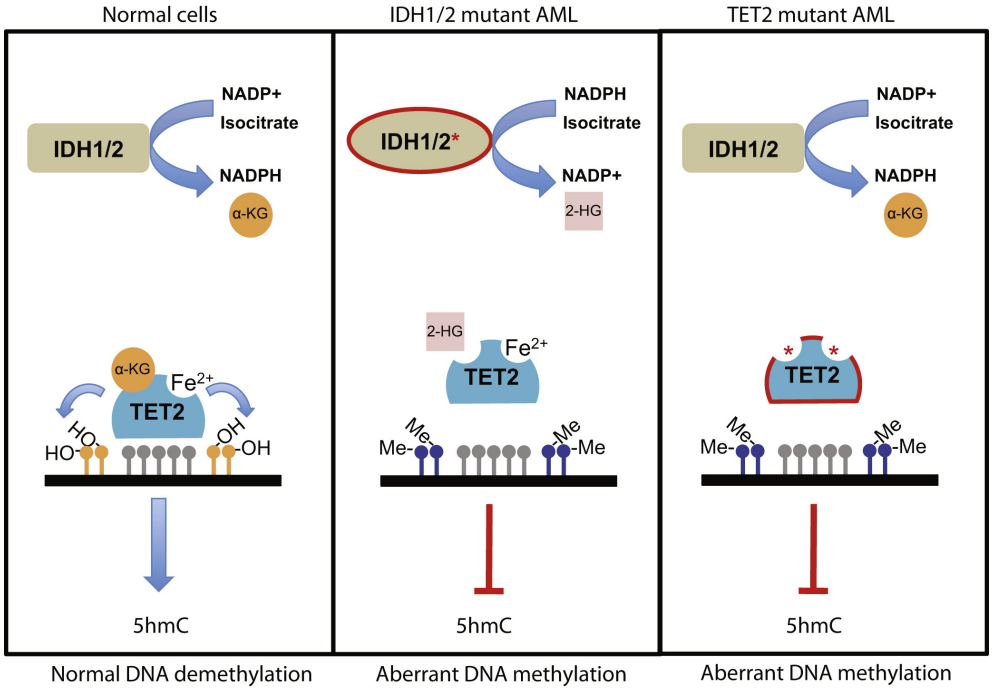


Figure 5: Both TET2 and IDH1/2 mutations lead to loss of 5hmC in AML, but are not seen to occur simultaneously (Cimmino, Abdel-Wahab et al. 2011).

Based on these studies it can be concluded that a global decrease in 5hmC can be accepted as a characteristic of malignancies, and that low levels of 5hmC encourage the emergence of both cancerous cells and solid tumors (Delhommeau, Dupont et al. 2009). Reduction of 5hmC may also be enhanced by active growth of cancer cells, possibly due to the activity of the maintenance of 5mC catalyzed by DNMT1(Valinluck and Sowers 2007; Inoue and Zhang 2011). As 5hmC is implicated as an intermediate in DNA demethylation, the loss of 5hmC in cancers may in fact lead to the increase in methylation that is seen in many types of cancer.

1.2 Epigenetics

Epigenetics has been defined as "heritable changes in gene expression that occur without a change in DNA sequence" (Wolffe and Matzke 1999), and is most prominently seen via DNA methylation and chromatin modifications. On a molecular level, DNA methyltransferases, methyl-CpG-binding proteins, chromatin remodeling factors, histone-modifying enzymes, chromosomal proteins and transcriptional factors all act together to

create the epigenetic landscape (Nakao 2001). Epigenetic states of daughter cells are inherited from their parent cells during DNA replication and cell division, thus showing that the epigenetic state is retained throughout the cell cycle (Rakyan, Preis et al. 2001; Groth, Rocha et al. 2007).

DNA methylation is the best-characterized chemical modification of DNA; in mammals, nearly all methylation is found on CpG dinucleotides, and areas with high CpG density, termed CpG islands. Methylation of CpG islands is associated with transcriptional repression (Goll and Bestor 2005). Somatic cells are able to maintain and pass on their epigenetic profiles, while germline cells are reprogrammed during development. The enzymes responsible for DNA methylation are called DNA methyltransferases (DNMTs), and are classified into two groups: maintenance methyltransferases, which add methyl groups to hemimethylated DNA during replication, and de novo methylases, which add methyl groups after replication (Bestor 2000). DNA methylation also plays a role in formation of heterochromatin, X chromosome inactivation, and mammalian imprinting (Yang and Kuroda 2007).

Chromatin structure can be influenced by covalent histone modifications, and is able to influence gene expression patterns (Jenuwein and Allis 2001). The 'histone code' (Figure 6) refers to the numerous different modifications found on histone tails such as acetylation, phosphorylation, methylation, ubiquitilation and more. These modifications in single or combinations will lead to changes in gene expression such as transcriptional activation or repression (Strahl and Allis 2000). Interestingly, 5hmC was found to be enriched at enhancers containing H3K4me1 marks, but not at active enhancers containing H3K27Ac marks, suggesting that 5hmC may also lead to an open chromatin state (Sun, Jolyon et al. 2013).

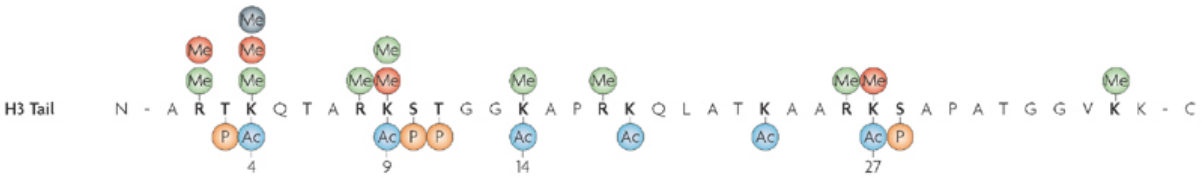


Figure 6: Possible modifications on H3 tail (Sims and Reinberg 2008) The different modifications are indicated as following: P – phosphorylation, orange; Ac – acetylation, blue; Me – methylation, green indicates an activating mark, red indicates a repressive mark, black indicates a mark not associated with transcription.

Noncoding RNAs (ncRNAs) are also involved in epigenetic phenomena. For example, ncRNA *Xist* regulates X-chromosome inactivation in females (Mikkelsen, Wakefield et al. 2007), and several genes are subject to post-transcriptional gene silencing by miRNAs (Wiklund, Kjems et al. 2010). Interestingly, while DNA methylation and histone modifications occur before gene transcription, ncRNAs can act both prior to and after transcription.

1.2.1 Epigenetics & Development

Mammalian development is characterized by many epigenetic events, the first being active demethylation of the paternal genome, taking place in the zygote (Mayer, Niveleau et al. 2000). Demethylation continues throughout the first few early embryonic replication cycles, though this may be a passive event due to Dnmt1 moving from the nucleus to the cytoplasm (Carlson, Page et al. 1992). The active demethylation does not remove all methyl groups, as DMRs (differentially methylated regions) at imprinting centers and other specific regions are able to conserve their methylation status (Bourc'his, Xu et al. 2001; Smallwood, Tomizawa et al. 2011).

Around the time of embryo implantation, Dnmt3a and Dnmt3b are present in high concentrations, concurrent with a dramatic increase in genome-wide methylation. Some sequences can become up to 80% methylated, though CpG islands remain unmethylated (Okano, Bell et al. 1999; Laurent, Wong et al. 2010). CpG islands are most likely protected from methylation by Sp1 motifs, which also have the ability to protect sequences not located in CpG islands, as shown after being introduced as a transgene in one-cell embryos (Brandeis, Frank et al. 1994; Siegfried, Eden et al. 1999).

After implantation, additional changes occur in methylation, most notably the silencing of the pluripotency genes Oct-3/4 and Nanog, allowing embryonic differentiation to proceed (Scholer 1991). The pluripotency genes are turned off first through a repression-factor mechanism, followed by recruitment of the histone methyltransferase G9a to their promoters, facilitation histone deacetylation, with subsequent methylation of H3K9, which binds HP1, forming heterochromatin (Feldman, Gerson et al. 2006; Cedar and Bergman 2012). The G9a complex is also able to recruit Dnmt3a/b, leading to the methylation of pluripotency gene

promoters (Epsztejn-Litman, Feldman et al. 2008). Thus, the pluripotency genes become permanently silenced.

In females, inactivation of the second X chromosome is a major epigenetic event, and inactivation occurs in a multi-stage process, beginning with the expression of *Xist*, a non-coding gene exclusively expressed from the inactive X chromosome (Keohane, O'Neill L et al. 1996). Eventually, *Xist* RNA coats the entire inactive X chromosome, and recruits silencing factors to keep it repressed (Jeon, Sarma et al. 2012). Histone methylases are responsible for generating heterochromatin and recruiting Dnmts that methylate CpG island promoters, which occurs several days after the initial X-inactivation (Cedar and Bergman 2009).

Tissue-specific genes that have been silenced through methylation at the beginning of development must be unmethylated for development to proceed. The genes must be recognized by cell-type-specific factors that must recruit the appropriate molecules to demethylated their promoters (Yisraeli, Adelstein et al. 1986). Tissue-specific de-novo methylation of previously protected sequences in CpG islands also occurs, though these regions are not associated with promoters, but with coding regions, where methylation is associated with gene activation (Straussman, Nejman et al. 2009).

1.2.2 Epigenetics & Disease

Diseases such as cancer, developmental disorders, and autoimmune disorders can be characterized by epigenetic abnormalities. While most of the focus lies on DNA methylation; other factors such as chromatin remodeling, histone modifications and non-coding RNAs are also implicated in many diseases.

A representative epigenetic characteristic of cancer is a global loss of 20-60% of DNA methylation in concurrence with hypermethylation at specific promoters at CpG islands (Goelz, Vogelstein et al. 1985). Hypomethylation occurs most frequently at repetitive sequences, leading to translocations, chromosomal instability and disruption of gene expression; which can lead to abnormal expression of oncogenes (Eden, Gaudet et al. 2003). Hypermethylation of promoters leads to disruption of genes associated with DNA repair, vitamin response, Ras signaling, cell cycle control, the p53 network, and apoptosis (Esteller

2007). miRNAs, which are downregulated in cancer, are hypermethylated at their promoters, which is also linked to metastasis development (Lujambio, Calin et al. 2008).

Epigenetic factors are significant players in key steps of nervous system development, specifically when neural cells lose their multipotency (Yoo, Staahl et al. 2009). Point mutations in the MBD protein MeCP2 lead to Rett Syndrome, a neurodevelopmental disorder leading to regression of motor and language skills after apparently normal development for the first 6-18 months after birth (Zachariah and Rastegar 2012).

Autoimmune disease results from the loss of immune tolerance to self-specific antigens (Portela and Esteller 2010). Immunodeficiency-Centromeric Instability (ICF) syndrome, one of the most well known autoimmune diseases is caused by mutations in *DNMT3B*, leading to extreme immunodeficiency from defective Dnmt3b enzyme activity (Robertson 2005). Most autoimmune diseases are not due to loss of functioning methylases; instead, they stem from hypomethylation of immune-related genes. Such is the case for systemic lupus erythematosus, where hypomethylation is found in the *PRF1, CD70, CD154* and *IFGNR2* genes, and also in the rRNA gene promoter 18S and 28S. miRNAs are also shown to mediate progression of systemic lupus erythematosus; miR-21 and miR-148a target DNMT1 and contribute to DNA hypomethylation (Pan, Zhu et al. 2010).

1.2.3 Epigenetics & Psoriasis

Psoriasis is a chronic T-cell-mediated disease of the skin occurring equally in men and women; approximately 2% of the population is affected. It is characterized by erythematous plaques, often covered by a silvery scaling, and enhanced keratinocyte proliferation (Wrone-Smith and Nickoloff 1996; Nickoloff and Nestle 2004). There is evidence suggesting that psoriasis has a genetically heritable component, and there are several genes that have been linked to the disease. There are also several environmental triggers associated with the pathogenesis of psoriasis, leading to the hypothesis that epigenetic factors also play a role in disease proliferation. These triggers include trauma, infections, drugs, smoking and alcohol consumption. (Zhang, Su et al. 2012)

Aberrant DNA methylation has been seen to disrupt gene expression, and this phenomenon has been seen in psoriasis. Both the p15 and p21 genes have been found to be

hypomethylated in individuals with psoriasis (Zhang, Zhang et al. 2009). A previous study using monozygotic twins discordant for psoriasis showed that changes in gene expression for a large number of immune response related genes (*LDHC, IL13, TNFSF11*, and more) were correlated to changes in methylation status of the gene (Gervin, Vigeland et al. 2012). Another example of atypical DNA methylation in psoriasis involves the SHP-1 promoter 2 region; SHP-1 regulates growth and proliferation processes in skin, and has been found to be extensively demethylated at the promoter 2 region in keratinocytes from psoriatic lesions (Ruchusatsawat, Wongpiyabovorn et al. 2006).

In addition to irregularities in DNA methylation, histone modifications and miRNA aberrations are also implicated in psoriasis. Histone deacetylase 1 (HDAC-1) has been shown to be overexpressed in psoriatic tissues, and HDAC SIRT1 (silent mating type information regulation 2 homologue 1) may prevent normal proliferation of keratinocytes (Tovar-Castillo, Cancino-Diaz et al. 2007; Blander, Bhimavarapu et al. 2009). A miRNA that can activate inflammatory cytokines – miR203 has been found to be overexpressed in patients with psoriasis (Sonkoly, Wei et al. 2007).

These results support the notion that epigenetic disregulation of genes related to the immune system can lead to autoimmune disease proliferation. There is currently no research on 5hmC in regards to any type of autoimmune disease, most likely due to the relative newness of 5hmC research.

1.3 Twin Studies

In comparison to dizygotic (DZ) twins who only share half of their DNA, monozygotic (MZ) twins, which account for 1 in 250 live births, share the same DNA sequence, (Hall 1996). Division of the zygote is not hereditary and is seen as abnormal (Ballestar 2010). Interestingly, both MZ and DZ twins exhibit phenotypic differences. In DZ twins this can be attributed to genetic differences, while phenotypic differences in MZ twins can be attributed to environment (Petronis 2006). This makes MZ twins the ideal model for researching environmental epigenetics, as the abundance of differences such as single-nucleotide polymorphisms (SNPs) found in DZ twins interfere with analysis.

MZ twins are identical in their DNA profiles, and their epigenetic profiles can provide insight into phenotypic differences. In a previous twin study of a group of 40 twin pairs ranging in age from 3-74 years, it was found that 65% of the MZ twins had nearly identical global 5mC levels and acetylation of histones H3 and H4 (Fraga, Ballestar et al. 2005). However, the other 35% had significant differences in their epigenetic profiles; interestingly enough, this group was comprised of the older twin pairs, suggesting that older twin pairs have greater epigenetic differences than younger ones. It was also observed that the older twins pairs had up to 2.5 times as many 5mC differences in CpG islands, and the result of a gene expression microarray analysis also showed that there were four times as many differentially expressed genes in the older twin pairs than the younger (Fraga, Ballestar et al. 2005). Later studies have confirmed that changes in methylated status are due to aging; age-related locus specific variation in methylation has been shown in multiple tissues in twins (Rakyan, Down et al. 2010). Thus, the importance of age must be taken into consideration when studying epigenetics both in twins and individuals.

Twin cohorts are fundamental in human studies, as using MZ twins makes comparison of phenotypic concordance easy to estimate heritability and identify disease specific genes, as well as being a powerful indicator of the influence of environmental factors on phenotype (Ballestar 2010). Disease concordance can be expressed as either pairwise or probandwise concordance, where pairwise concordance is the proportion of the affected twin pairs concordant for disease (e.g. 30 of 100 pairs are affected, resulting in 30% concordance); probandwise concordance estimates the risk of a twin developing a disease if the co-twin has already been diagnosed (Bogdanos, Smyk et al. 2012). A high rate of concordance presumes a predominant genetic influence, whereas low concordance rate is assumed to be due to environmental factors. The concordance rates of many autoimmune diseases have been studied in MZ twins; Type I diabetes has been found to have a concordance rate of between 13-47.4% (Kaprio, Tuomilehto et al. 1992; Matsuda and Kuzuya 1994), 60-75% for Coeliac disease (Bardella, Fredella et al. 2000; Greco, Romino et al. 2002), and between 35 and 65% in Psoriasis (Brandrup, Hauge et al. 1978; Duffy, Spelman et al. 1993).

Though there is a high concordance for many types of heritable disease in MZ twins, rates of discordance can also be very high, generally over 50%, implying that epigenetics is an important contributing factor to the discordant phenotype (Kendler and Prescott 1999;

MacGregor, Snieder et al. 2000). Discordance rates increase inversely with disease prevalence, where discordance is calculated as a function of prevalence. The discordant MZ twin model is a robust tool that is able to detect phenotypic risk factors while controlling for unknown confounding variables (Bell and Spector 2011). For example, schizophrenia and rheumatoid arthritis have been found to have a discordance rate of about 80%, while only being present in 1% of the population (MacGregor, Snieder et al. 2000; Sullivan, Kendler et al. 2003). However, osteoarthritis has a discordance rate of about 40%, and a prevalence of 20% (Spector, Cicuttini et al. 1996).

Epigenetic studies in MZ discordant twins can be used to identify differences in the epigenetic profiles of co-twins, to identify disease specific genes. In a previous discordant MZ twin study looking at schizophrenia, it was found that there were greater differences in methylation of the dopamine D2 receptor gene (*DRD2*) between the discordant twins than in unrelated cases (Petronis, Gottesman et al. 2003). Subsequent studies have found phenotype-related methylation changes in discordant MZ twins in bipolar disorder (Kuratomi, Iwamoto et al. 2008), caudal duplication anomaly (Oates, van Vliet et al. 2006), and the catechol-O-methyltransferase gene (Mill, Dempster et al. 2006).

1.4 Investigating 5-hydroxymethylcytosine

There are an increasing number of methods using to investigate 5hmC as it has only recently become of interest to researchers again. Methods such as Mass Spectrometry, High-Performance Liquid Chromatography, and ELISA assays (Section 2.1) can be used to quantify and locate 5hmC on a global level. Sequence-specific detection of 5hmC has proven to be more difficult, but one is able to locate 5hmC at a single base resolution using methods such as hMeDIP-seq, CMS or GLIB followed by DNA sequencing, selective chemical labeling, Ox-BS-Seq and TAB-seq. (Kriaucionis and Heintz 2009).

1.4.1 Isolation and Enrichment of 5hmC

Several methods rely on the specific recognition of 5hmC by antibodies, which offer the advantage of requiring little specialized lab equipment. However, there are two big disadvantages stemming from the usage of antibodies themselves, first being the fact that the antibodies may cross react with either unmodified or methylated cytosine bases, and secondly, antibodies are specific for regions highly enriched with 5hmC, thereby being less

likely to capture regions with few 5hmC bases. hMeDIP (Hydroxymethylated DNA Immunoprecipitation) is comparable to the older and more widely practiced MeDIP (Sorensen and Collas 2009). There exist already several commercial suppliers of kits to perform hMeDIP, though the results generally have only moderate resolution and give only relative quantification information (Booth, Branco et al. 2012).

Chemically modifying 5hmC bases before pull-down is another method of isolating fragments containing 5hmC. GLIB involves transferring glucose molecules to 5hmC via T4 phage β-glucosyltransferase, with subsequent oxidation using sodium periodate and further chemical modification yielding two biotin molecules at each 5hmC site, allowing pull-down using streptavidin beads. Another type of chemical modification of 5hmC yields 5-methylenesulfonate (CMS) from sodium bisulfite conversion, followed by immunoprecipitation of CMS-containing DNA with an antibody specific to CMS (Pastor, Pape et al. 2011).

DNA isolated and enriched using the above methods can be subject to a number of downstream applications such as qPCR (Section 2.5), Microarray Analysis, and Next Generation Sequencing.

1.4.2 Direct detection of 5-hydroxymethylcytosine by DNA sequencing

With the help of next-generation sequencing technologies, one can now sequence 5hmC at a single base resolution using a variety of methods. The biggest challenge is being able to differentiate between all of the modified cytosine bases. SMRT (Single Molecule Real-Time) sequencing is able to identify 5hmC at a single nucleotide resolution, though it has a high rate of sequencing errors (Eid, Fehr et al. 2009). Two alternative methods have been created for the more common Illumina sequencing technology; Ox-BS-seq and TAB-seq.

Oxidative Bisulfite Sequencing (Ox-BS-seq) builds upon a previously published method for investigating methylation – Reduced Representation Bisulfite Sequencing (RRBS) (Meissner, Gnirke et al. 2005). In short, RRBS uses the methylation insensitive restriction enzyme *Msp*I to cleave DNA at CpG sequences, followed by treatment with sodium bisulfite, which converts unmodified cytosine bases to uracil, while leaving 5mC intact. This allows one to

look at methylation in areas rich in CpG islands. Recently it was found that traditional bisulfite (BS) sequencing is unable to distinguish between 5mC and 5hmC, and that the product of BS conversion of 5hmC is CMS (Figure 7), which has also been shown to stall DNA polymerases during PCR, and indicates that previous methylation analyses may underrepresent areas of dense hydroxymethylation (Huang, Pastor et al. 2010). To combat this Ox-Bs-Seq uses an additional oxidation step at the beginning of the protocol, where 5hmC is oxidize to 5fC using Potassium Perruthenate ($KRuO_4$), which reacts specifically with 5hmC and has no interaction with Cytosine or 5mC (Booth, Branco et al. 2012). The same protocol is then followed allowing one to look at 5hmC in a CpG context.

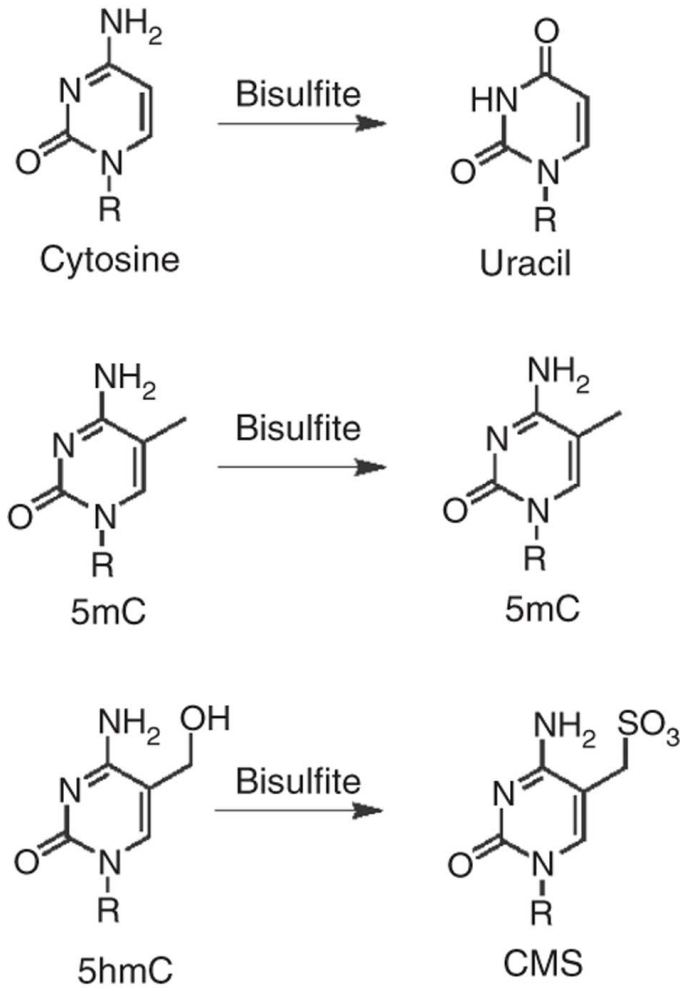


Figure 7: Effects of Sodium Bisulfite treatment on Cytosine, 5mC, and 5hmC (Huang, Pastor et al. 2012)

Experiments using Ox-BS-seq have shown that 5mC and 5hmC have an inverse correlation with CpG density across CpG islands. Intra- and intergenic CpG islands have a higher amount of both 5mC and 5hmC than CpG islands overlapping transcription start sites (TSS).

LINE1 elements have also been shown to have ~5% 5hmC content, which could indicate the demethylation of specific repeat classes (Booth, Branco et al. 2012).
Tet-Assisted Bisulfite Sequencing (TAB-seq) uses β-glucosyltransferase to tag 5hmC, protecting it during subsequent bisulfite conversion (Figure 8). After 5hmC is tagged and protected with a glucose molecule, TET proteins oxidize 5mC to 5caC. During bisulfite conversion, 5caC behaves as an umodified cytosine base would, while the 5hmC bases are protected from conversion, leading 5hmC to be read as cytosine, and 5caC to be read as thymine (Yu, Hon et al. 2012).

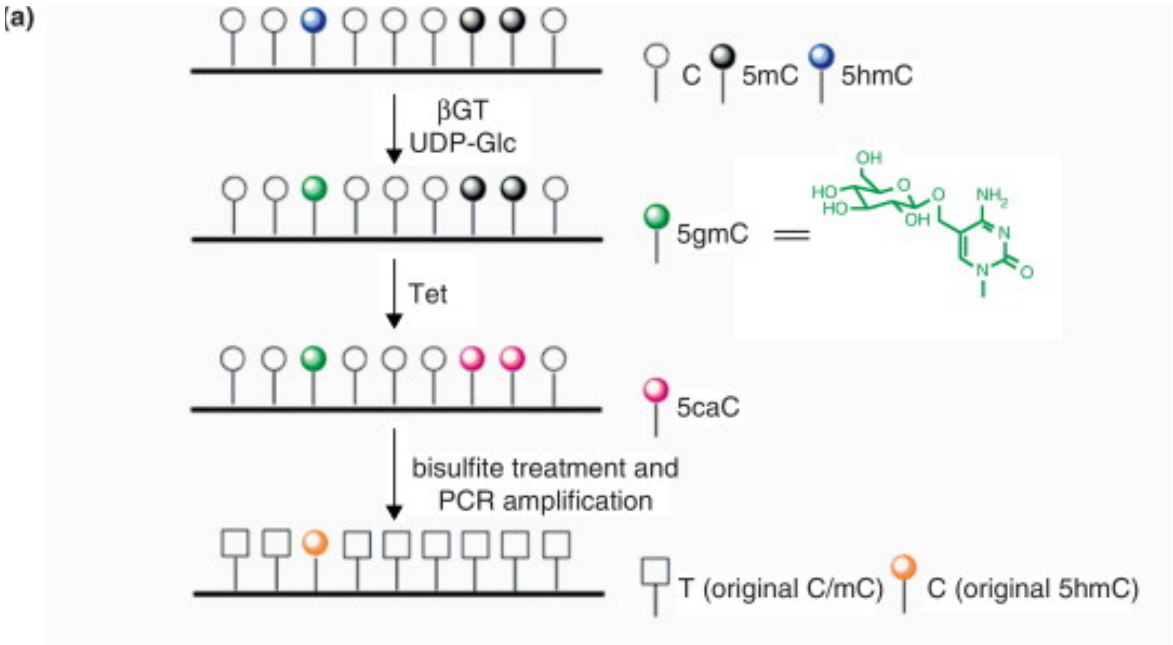


Figure 8: TAB-seq flow chart (Fu and He 2012)

Though Ox-BS-seq and TAB-seq are suitable methods for both single nucleotide resolution and quantifying relative abundance of 5hmC, there is one setback one must take into account for both methods. Both methods require one to run a BS-seq in parallel, as to be able to extract positional information on 5hmC, which considerably increases the costs of such experiments.

High-resolution sequencing of 5hmC can also be achieved using DNA-modification-dependent restriction enzymes coupled with sequencing. In a recently published method, 5hmC was glycosylated and then subject to incubation with the restriction enzyme *AbaSI*, which specifically cleaves at a range of distances from the glycosylated 5hmC and doesn't cleave 5mC or C (Sun, Jolyon et al. 2013). By mapping the cleaved ends of the fragments,

the exact location of 5hmC can be found (Wang, Guan et al. 2011). A similar method has also been published using *Msp*I restriction enzyme (Section 2.3 – Sequence Specific Detection of 5-Hydroxymethylcytosine).

1.5 Aims of the Thesis

By studying the epigenetic context of 5hmC in MZ twins discordant for psoriasis, we aimed to acquire insight into the location of 5hmC distributed throughout the genome, and if the location of 5hmC had any implications in regards to psoriasis.

The aims of this study were to:

1. Quantify the amount of 5hmC found in CD4+ cells isolated from blood.
2. Investigate the location of 5hmC on the human genome in a cohort of twins discordant for psoriasis.
3. Perform a comparison of the hydroxymethylation profiles between the co-twins, and see if any differences can be associated with psoriasis.

2 Methods and Materials

2.1 ELISA – Quantification of 5hmC

ELISA (Enzyme-linked immunosorbent assay) is a method of detecting an antigen in a solution by using an antibody against it. In short, antigens from a sample are bound to a surface, which is subject to incubation with an antibody recognizing the antigen. The antibody is linked to an enzyme, which when exposed to its substrate changes color. Therefore, a change in color shows the presence of the antigen, and using the depth of the color change and calculating absorbance, the amount of antigen present can be quantified. In this experiment, 5-hydroxymethylcytosine was detected and quantified using MethylFlash™ Hydroxymethylated DNA Quantification Kit (Epigentek). The protocol from the manufacturer was followed as described below (Figure 9):

A dilution series was made using a supplied control DNA containing 5hmC; this creates a standard curve, which can be used to calculate the amount of 5hmC in the experimental sample. 200ng of purified CD4+ DNA and the standard curve DNA were bound to the assay wells by incubating the reaction in a binding buffer at 37°C for 90 minutes, followed by removing the binding solution and washing the wells. An antibody specific to 5hmC was added and incubated at room temperature for 1 hour, followed by washing and addition of the detection antibody. After an additional wash, an enhancer solution was added, followed by a developer solution. If 5hmC is detected, the developer solution will cause a color change from yellow to blue. Once color change is detected, a stop solution is added and the absorbance is read at 450nm on a microplate reader (Bio-Rad).

Once the absorbance is read, the absolute amount of hydroxymethylated DNA in each experimental sample can be calculated in reference to the standard curve.

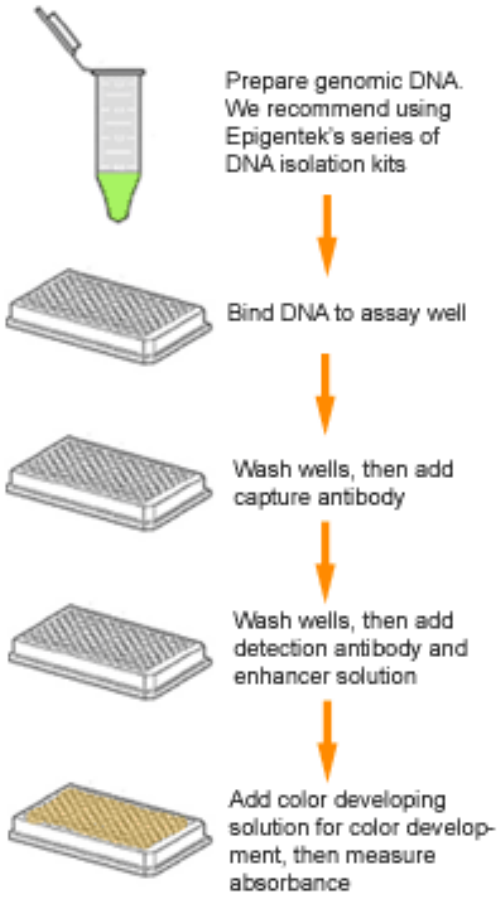


Figure 9 – Workflow of Epigentek MethylFlash Hydroxymethylated DNA Quantification Kit (Epigentek)

2.2 Hydroxymethylated Immunoprecipitation

Immunoprecipitation is a technique that makes use of a specific antibody to bind a protein, which can then be separated from a solution. Hydroxymethylated Immunoprecipitation (hMeDIP) uses an antibody against 5hmC to recover fragments of DNA that contain 5hmC. The recovered DNA enriched in 5hmC can then be checked for enrichment using qPCR, and fold enrichment can be calculated. Immunoprecipitated DNA could also be sequenced to generate a genome-wide measure of 5hmC distribution. In this study, both the Active Motif hMeDIP Kit and Zymo Research Quest 5hmC DNA Enrichment Kits were used. The kits use different types of antibodies, but follow the same principal. Both kits follow the same DNA fragmentation procedure prior to starting the kit, as well as the same cleanup methods post elution.

2.2.1 hMeDIP using Zymo Research Quest 5hmC DNA Enrichment Kit

Zymo Research's Quest 5hmC DNA Enrichment Kit utilizes J-base binding protein (JBP1) for enrichment of DNA containing 5hmC. JBP1 is a protein that binds to the highly modified base β-glucosyl-5-hydroxymethyluracil, often referred to as the J-base and found in African trypanosomes. JBP1 binds to the β-glucosyl moiety of the J-base, as well as glucosyl moieties on other bases, such as glycosylated 5hmC.

In the experiment, fragmented DNA is incubated with 5hmC Glucosyltransferase, which glycosylates 5hmC. Magnetic beads tagged with JBP1 are added to the solution containing glucose-tagged 5hmC and bind the DNA fragments containing 5hmC. After washing the beads, the DNA fragments enriched in 5hmC are eluted (Figure 10). The protocol was followed according to the manufacturers instructions, using 4µg of input DNA.

Figure 10 – Workflow of Quest 5hmC DNA Enrichment Kit (From Zymo Research Corp.)

To evaluate enrichment, a spike of 5hmC control DNA can be added to a sample as a positive control. The 5hmC control DNA is a mix of two plasmids chemically modified to contain either 5mC or 5hmC. The plasmids also differ in that the 5hmC-modified plasmid contains a *Kpn*I restriction site. From these spiked plasmid templates, a 400bp PCR fragment containing the *Kpn*I restriction site can be amplified. Successful enrichment is shown by cleavage of the 400bp amplicon to 200bp fragments, which can be seen on a gel. A negative control reaction should be included where the protocol is followed, but 5hmC Glucosyltransferase is not added.

2.2.2 Hydroxymethylated Immunoprecipitation using Active Motif hMeDIP Kit

Instead of a glycosylation reaction prior to capture, as in the Zymo Research kit, Active Motif's hMeDIP kit uses a highly specific purified 5-hydroxymethylcytidine antibody to capture DNA containing 5hmC. Briefly, fragmented DNA is incubated overnight in a solution containing a 5-hydroxymethylcytidine antibody. The next day, magnetic beads,

which bind to the 5-hydroxymethylcytidine antibody, are added to the solution and incubated for 2 hours. The DNA is then eluted from the beads, yielding DNA fragments enriched for 5hmC (Figure 11). The protocol from the manufacturer was followed, using 1µg input DNA for all reactions. In this experiment, hMeDIP reactions were done using several different control DNA spikes (described below) as well as using a Rabbit IgG antibody for negative controls, allowing one to calculate enrichment after qPCR.

For validation of the reaction, a spike of control DNA – an oligo containing either 5hmC, 5mC or unmodified cytosines can be added to the reaction, and can be analyzed using qPCR using the provided APC locus primer mix. Because the APC locus is not hydroxymethylated in human DNA, enrichment shown in the qPCR should only be derived from the control DNA. Spikes of unmethylated and methylated DNA were also be added to reactions to ensure that the antibody does not immunoprecipitate those bases. A negative control reaction was also run using Rabbit IgG instead of the 5-hydroxymethylcytidine antibody, to control the specificity of the provided 5hmC antibody.

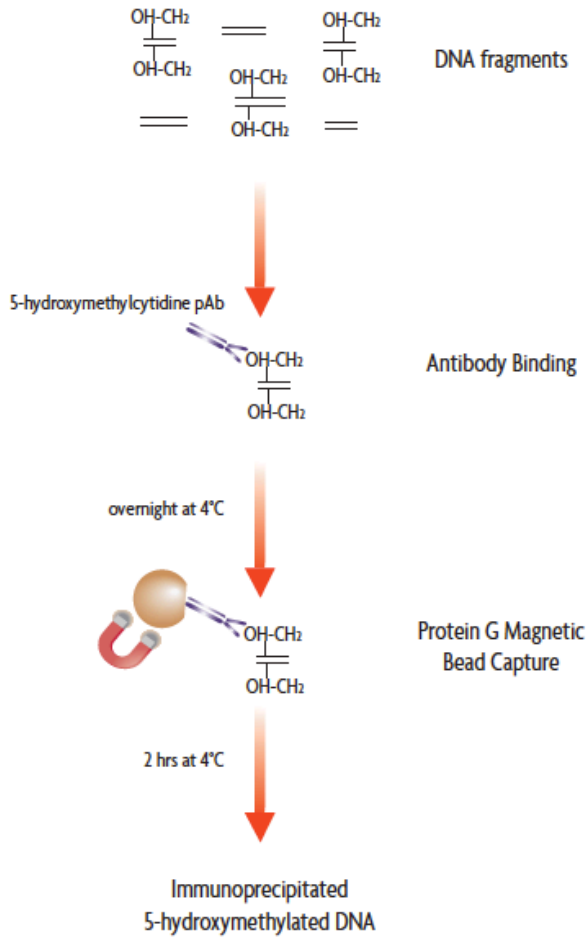


Figure 11 – Workflow of Active Motif hMeDIP Kit (Active Motif, Inc.)

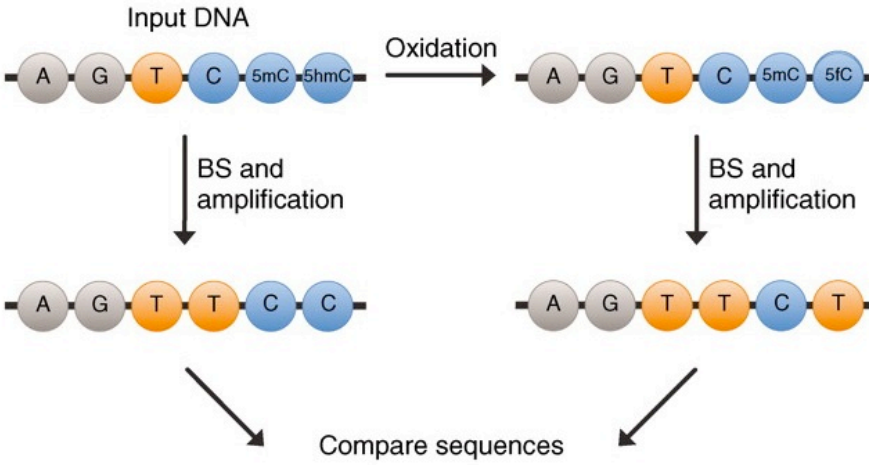
2.2.3 Purification of DNA after hMeDIP

Following the hMeDIP protocols, the volume of DNA eluted was found to be too high, leaving the DNA at a concentration that is too low to be measured using the Qubit reagents (Appendix 5.1.1) and unsuitable for downstream applications. Eluted hMeDIP DNA was therefore concentrated to a lower volume by using either a QIAQuick spin column (Qiagen) or Ampure XP beads (Beckman Coulter) and eluted in 10 µl dH_2O (Appendix 5.1.2).

2.3 Oxidative Reduced Representation Bisulfite Sequencing

Reduced Representation Bisulfite Sequencing is a method allowing for deep, selective sequencing of parts of the genome highly enriched for CpG islands, by using the methylation-insensitive restriction enzyme *Msp*I to generate fragments enriched with CpG dinucleotides at the ends (Meissner, Gnirke et al. 2005). The DNA is then treated with sodium bisulfite solution, which converts unmodified cytosine bases to thymine, so that when sequenced, an unmodified cytosine base is read as thymine, and 5mC is read as cytosine (Figure 12). This method was published before 5hmC was considered to be an epigenetic mark, and does not take into account that both 5mC and 5hmC are found as cytosine in sequencing data. (Matarese, Carrillo-de Santa Pau et al. 2011)

Oxidative Reduced Representation Bisulfite Sequencing (ox-RRBS) starts with an oxidation step, converting 5hmC to 5fC, while leaving 5mC and C in their original states. With subsequent bisulfite conversion, both unmodified cytosine and 5fC are converted to thymine, while only 5mC is read as cytosine. By comparing ox-RRBS sequencing reads to RRBS reads, one can define where 5hmC is located in the genome.



Base	Sequence	BS Sequence	oxBS Sequence
C	C	T	T
5mC	C	C	C
5hmC	C	C	T

Figure 12 – Diagram and table outlining RRBS and ox-RRBS methods (Booth, Branco et al. 2011)

2.4 Sequence Specific Detection of 5-hydroxymethylcytosine

2.4.1 Preparation of control DNA from mouse brain

Frozen whole mouse brain was obtained from John Arne Dahl at Rikshospitalet and genomic DNA isolated using a Qiagen DNEasy Blood & Tissue kit, which incubates tissue in a lysis buffer until homogenous, and then removes proteins and concentrates DNA by passing it through a spin column (Appendix 5.1.2) When protein contaminants were detected by UV spectroscopy (NanoDrop), the DNA was further cleaned up by Phenol: Chloroform treatment followed by EtOH precipitation (Appendix 5.1.2).

2.4.2 Sequence and Locus Specific Detection of 5hmC using Zymo Research Quest 5hmC Detection Kit

Similar to the hMeDIP protocol from Zymo Research (Section 2.2.2), this protocol starts with using 5hmC Glucosyltransferase, an enzyme that tags 5-hydroxymethylcytosine with a glucose moiety. A parallel reaction is run without glucosyltransferase treatment. Both of the

samples are then subject to treatment with *Msp*I, a Glucosyl-5hmC Sensitive Restriction Endonuclease (GSRE), which cleaves at C<u>C</u>GG. *Msp*I is a methylations insensitive enzyme; it will not cut sites that have a glycosylated 5hmC base, but will cut 5mC and unmodified cytosine bases (Figure 13). After *Msp*I digestion, DNA was cleaned up and reconcentrated, rendering it usable for downstream applications. One can expect that the sample treated with the glucosyltransferase will have longer fragments than the untreated sample, as there will be fewer places for *Msp*I to cut.

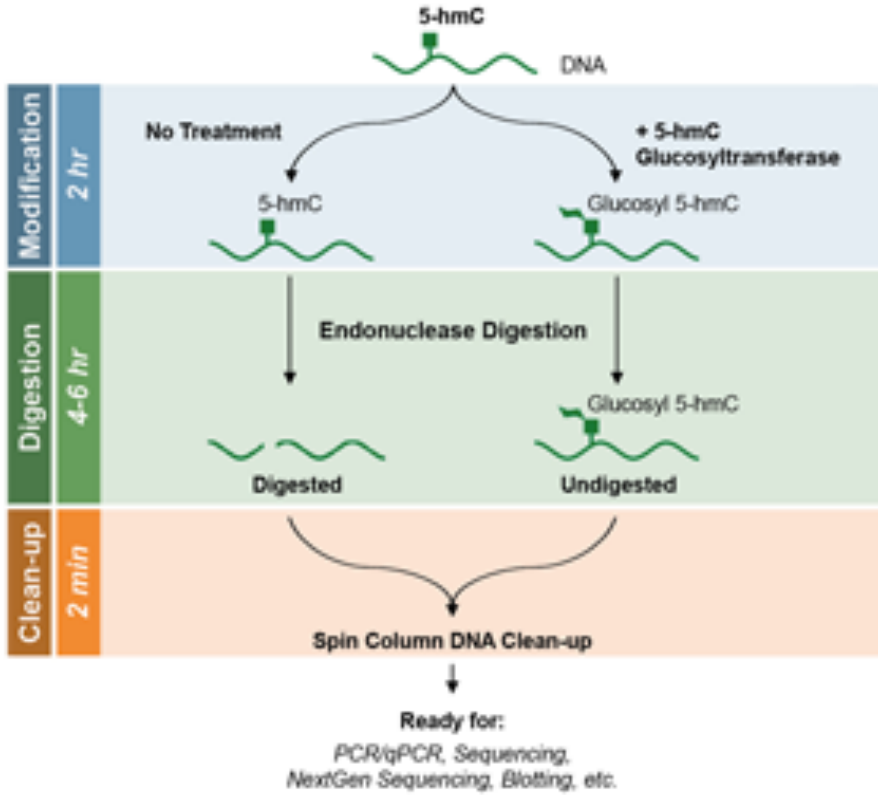


Figure 13: Workflow of Quest-5hmC Detection Kit (Zymo Research Corp.)

A positive control oligo with 5hmC in *Msp*I cleavage sites is included in the kit and was used to check if the reaction worked correctly.

2.4.3 Preparation for Sequencing

Sequencing libraries were made using TruSeq DNA Sample Preparation (Illumina) reagents, in a protocol modified for low input. The purpose of this protocol was to prepare DNA fragments for sequencing by adding indexed adapters to each end of the DNA fragments. Reaction conditions can be found in Appendix 5.1.4.

First, the overhangs on the DNA fragments were converted to blunt ends using an End Repair Mix, which both removes the 3' overhangs, and fills in the 5' overhangs. To remove traces of enzymes and nucleotides, the DNA was cleaned using a DNA Clean & Concentrator – 5 (Zymo Research) spin column, and eluted in 20µl Resuspension Buffer (RSB). Next, a single 'A' nucleotide was added to the 3' ends of the DNA fragments – preventing them from ligating to each other during adapter ligation and increasing ligation efficiency to the adapter molecules by introducing a single base overhang. Specific adapters with indexes were then ligated to the fragments, allowing them to hybridize onto a flow cell. A stop ligase mix was added to the solution to end the ligation process. Once again, the DNA fragments, now 122 bp longer due to the addition of adapters, were cleaned up using a DNA Clean & Concentrator – 5 (Zymo Research) spin column and eluted in 12µl RSB.

To remove adapter dimers and size-select DNA fragments for sequencing, the indexed DNA fragments were pooled and then run on a 2% TAE agarose gel for ca. 45 min at 100V. Two adjacent bands were then excised; the first band from 175-300 bp, the second from 300-766 bp. Samples were split into two size fractions to allow more even sequence coverage on the Illumina HiSeq 2000: The clustering stage of sequencing relies on PCR, which favors amplification of smaller products, therefore combining fragments from the entire size range in a single clustering reaction would result in the vast majority of sequences being derived from the smallest fragments. DNA was extracted from the gel and cleaned using spin columns and reagents from Qiagen's MinElute Gel Extraction Kit. The DNA was eluted from the spin column in 22.5µl of the supplied Elution Buffer. Next, the DNA was amplified in a PCR reaction using TruSeq reagents. The lowest possible number of PCR cycles was used to minimize PCR-induced bias and duplicate reads, in this case 13 cycles. Finally, the PCR reaction was cleaned up in a 1:1 Ampure XP bead cleanup, and DNA recovered in 15µl RSB.

Quality control of the libraries was performed by checking the DNA concentration on the Qubit (Invitrogen), and checking that adapter dimers had been removed during the gel extraction step by running the samples on a BioAnalyzer High Sensitivity Chip (Agilent).

2.5 qPCR Validation of Experiments

Quantitative (Real-Time) PCR is a method that simultaneously amplifies and quantifies DNA molecules. SYBR Green (Applied Biosystems), a fluorescent dye that binds to double stranded DNA was used in the reaction for quantification. It binds to dsDNA at the beginning of the reaction, and is released during the denaturing step. After the polymerization step when the PCR amplicon is generated, the dye binds to the amplicon, resulting in a net gain of fluorescence in the reaction, which is detected by the qPCR system (Figure 14).

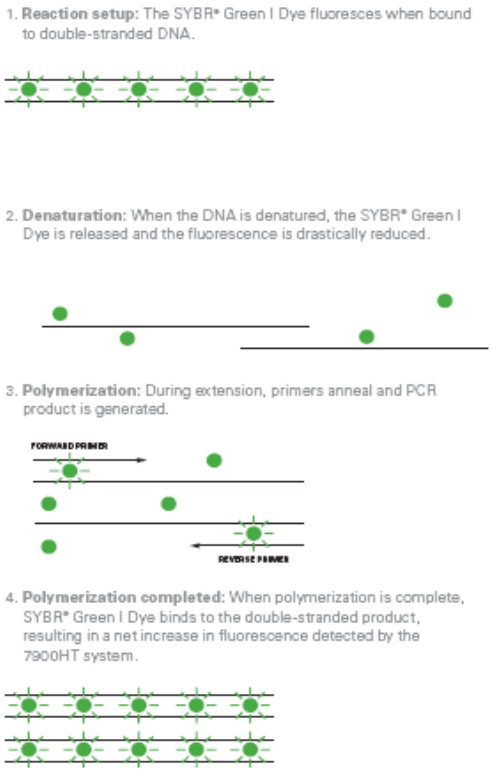


Figure 14: SYBR Green Dye Assay Chemistry (Applied Biosystems)

qPCR was used to check for enrichment of 5hmC containing fragments at specific loci. There are currently very few loci known to be enriched in 5hmC, making primer design difficult. Primers used for qPCR are listed in Appendix 5.2.

All qPCR reactions were run in triplicate on an ABI 7900 (Applied Biosystems) Real-Time PCR system. The absolute quantification method was used, allowing one to quantitate unknown samples based on a known quantity – a standard curve. The standard curve was in most cases made from DNA samples with the following concentrations: 20, 2, 0.2 and 0.02 ng/µl. Following each qPCR reaction, a melting curve analysis was performed.

2.6 Sequencing

Sequencing libraries were prepared such that compatible indexes were applied to samples of different insert sizes (either 150-300 bp or 300-766 bp), allowing these to be run together in a single lane. In total, two lanes of 100 bp paired end sequencing were run – one for each size range. The samples were clustered using TruSeq PE Cluster Kit v3-cBot-HS (Illumina), and sequenced using TruSeq SBS Kit v3 (200 cycles; Illumina) on a HiSeq 2000 instrument (Illumina). Sequencing was performed by the Norwegian Sequencing Centre (www.sequencing.uio.no).

2.7 Analysis of Sequencing Results

Image analysis and base calling were performed using Illumina's RTA software version 1.12. Reads were filtered to remove those with low base call quality using Illumina's default chastity criteria. In addition, due to the lower quality of bases at the 3' end of the sequence reads (Figure 29), reads were trimmed to 50bp using fastx-trimmer from the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Reads from the two size fractions 30-180 bp and 180-646 bp (lengths after removal of adapter sequences) were combined for further analysis.

Genome sequences for mouse (mm10) and human (hg19) were downloaded from Ensembl (http://www.ensembl.org/info/data/ftp/index.html). Sequence reads were aligned to the appropriate genome index using bwa (http://bio-bwa.sourceforge.net). A summary of the number of reads and mapping results are given in Table 1 (Section 3.5).

To identify CpG sites with 5hmC, it is necessary to identify locations with read map differences between the samples with +/- glycosylation reaction. To do this, we used MACS (Model-based analysis for ChIP-seq; http://liulab.dfci.harvard.edu/MACS/), with + glycosylation sample as treatment and -glycosylation sample as control. MACS identifies peaks that are significantly different between the two samples, and thus *Msp*I sites which contain 5hmC. A summary of the number of significant peaks is shown in Figure 31 (Section 3.5).

3 Results

3.1 ELISA

ELISA experiments were run using both control DNA from mouse brain, as well as twin DNA isolated from CD4+ T-cells. Mouse brain DNA was chosen as a positive control as it is known to have a higher than average 5hmC content.

The first few ELISA experiments were run with a standard curve made from the recommended values from the kit of 10.0, 5.0, 2.0, 1.0 and 0.5 ng/µl. This resulted in a non-linear curve, mostly from the sample with 10ng of hydroxymethylated DNA. Later experiments used a standard curve with concentrations of 1.0, 0.5, 0.2, 0.1 and 0.05ng/µl, resulting in a linear curve, which was used to extrapolate the amount of 5hmC in each DNA sample in ng, and from there calculate the percentage of 5hmC in the DNA samples.

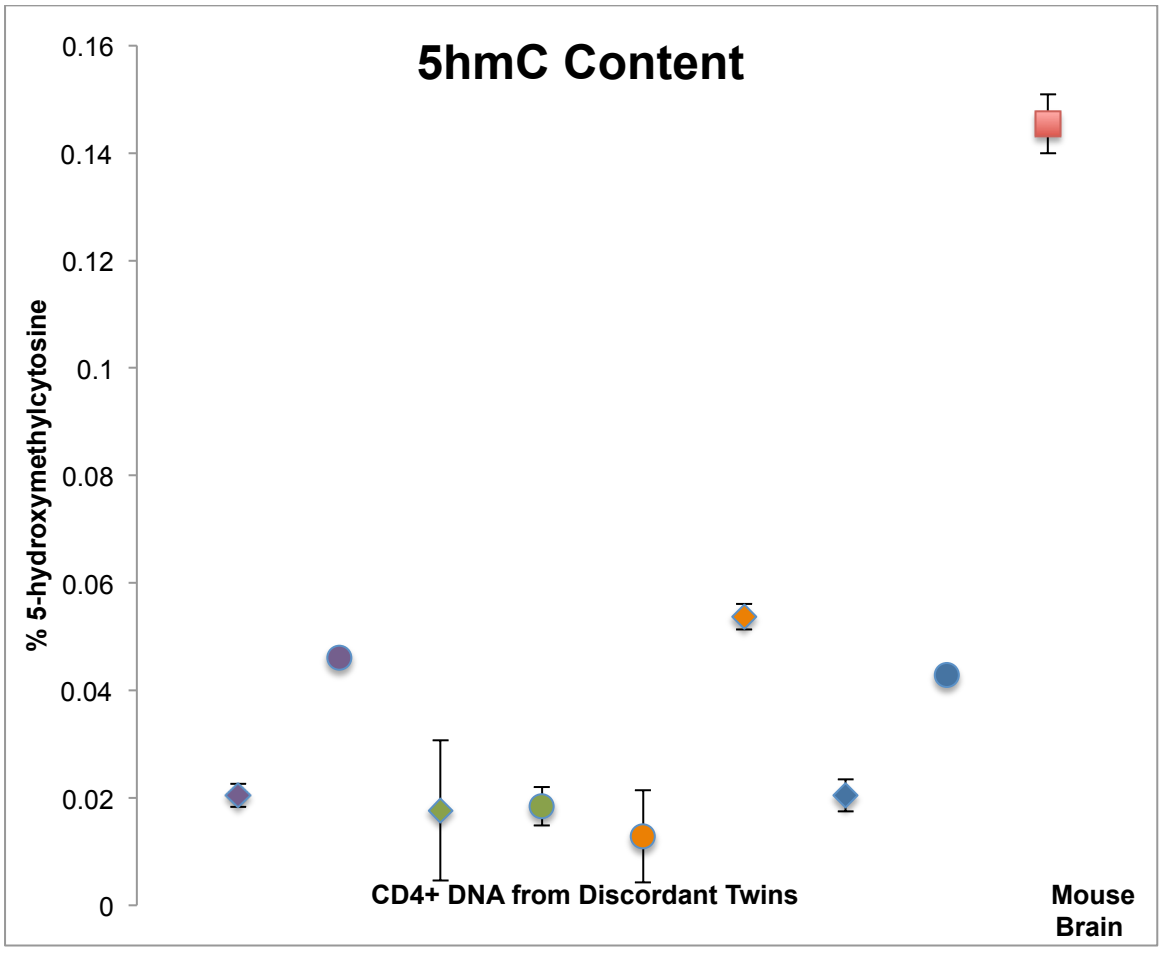


Figure 15: 200ng of CD4+ DNA from 4 twin pairs, and 200ng of mouse brain DNA was run in triplicate. DNA samples are shown along the X-axis, while % Hydroxymethylcytosine is on the Y-axis. The percent of 5-hydroxymethylcytosine of each sample and standard error were calculated from triplicate values. Twin pairs are shown in the same color. The healthy twin is denoted by a circular mark, the affected twin by a triangle.

The mouse brain DNA, which served as a positive control was determined to have approximately 0.15% 5hmC in the sample, while the CD4+ DNA, which varied within twin samples, was shown to have 0.02-0.06% 5hmC content. (Figure 15)

Significant differences (t-test $p<0.05$) were detected within three of the discordant twin pairs tested, though the second twin pair was not found to be significantly different ($p=0.603$). Therefore, we could detect global differences in 5hmC content within the twin pairs. However, there was no consistent relationship between disease state and hydroxymethylation level. Given the low number of twin pairs, it is only possible to draw minimal conclusions from the data.

Despite the low concentrations of 5hmC in the CD4+ DNA, we felt justified in proceeding with further experiments to examine 5hmC using more comprehensive assays.

3.2 Hydroxymethylated DNA Immunoprecipitation

All hMeDIP experiments were done using what we believe to be DNA from mouse brain (C57BL/6) provided by Arne Klungland at Rikshospitalet, and human genomic DNA (gDNA) from an anonymous donor (P-152) used as control DNA at the Department of Medical Genetics. The mouse brain served as a positive control in addition to the control oligos found in the respective hMeDIP kits. However, upon completion of the experiments detailed below, we were informed that we had mistakenly been given rat brain DNA. Nonetheless, further experiments were performed with genuine mouse brain DNA (see below) and produced identical results.

The lack of previously identified loci enriched in 5hmC made designing amplicons for qPCR validation a difficult task. At the time of running the experiment, only one publication was found with information in regards to specific loci in human tissues enriched in 5hmC. The regions identified to be the most enriched throughout various human tissues included *IGF2*, *H19*, and the *HOXA* cluster (Nestor, Ottaviano et al. 2012). Despite these regions showing considerably lower levels of enrichment in blood, we were forced to proceed in designing qPCR amplicons in these regions due to the lack of other sources of information. For human samples, one locus at the *IGF2* region and three loci from *H19* were chosen as likely positive controls, where qPCR amplicons were designed. *Tex19.1* was used to make a mouse qPCR

primer set for the mouse control samples, as it has also been previously identified as containing 5hmC (Nestor, Ruzov et al. 2010). *GAPDH* was used as a negative control for both species. After performing a melting curve analysis to check for primer efficiency, we found that the *IGF2* primer set amplified in the non-template control (NTC) wells, while the other primer pairs amplified only in wells with DNA. The *IGF2* primer set was discarded for further experiments, and only the *H19* primer sets were used. The first *Tex19.1* primer set designed failed to amplify, presumably due to the template DNA originating from the wrong species. Another set was therefore designed and fortuitously amplified correctly.

3.2.1 Zymo Research Quest 5hmC Enrichment

hMeDIP experiments were repeated several times using the Zymo Research Quest-5hmC Enrichment kit. The first run using 1µg mouse (in fact, rat) DNA did not lead to any pull down, confirmed by Qubit measurements and lack of amplification by qPCR. This led us to repeat the experiment with 4µg input of mouse (rat) DNA in parallel with 4µg, 3µg, 2µg and 1µg human gDNA. Once again, there was no measurable enrichment from checking DNA concentration on the Qubit or qPCR.

To make sure the kit was performing correctly, we ran a control hMeDIP using 1µg human DNA along with 200ng provided control DNA from the kit. A reaction was also run in parallel using the same input with no spike-in. After enrichment, the captured DNA was amplified and subject to *Kpn*I treatment. Because the control oligo has *Kpn*I recognition sites, a gel can be run with the samples and there will be a visible difference in size when comparing the treated and untreated samples. A 1% TBE gel was made and the samples were run at 100V for 35 minutes.

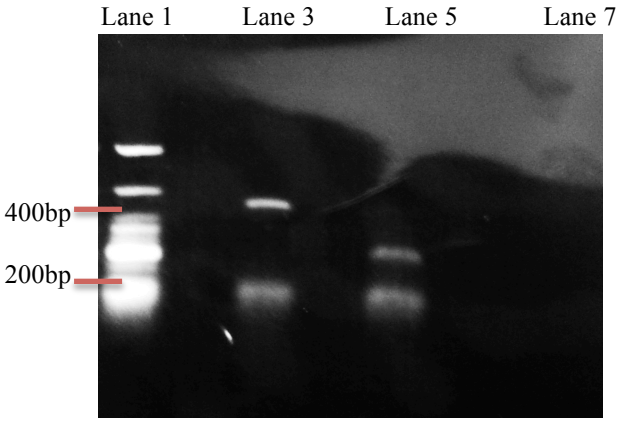


Figure 16: Control Zymo hMeDIP. Lane 1- 100bp Ladder, Lane 3 - Enriched control oligo, no *Kpn*I treatment, Lane 5 - Enriched control oligo, *Kpn*I treated, Lane 7 - Human gDNA sample, no spike or KpnI treatment

In Figure 16 we can see visible bands in both of the lanes from the hMeDIP with the control oligo spike, - a band at 400bp in Lane 3 for the untreated sample, and a band at 200bp in Lane 5, but no visible band in the lane without the control spike. If the hMeDIP had pulled down significant amounts of 5hmC-enriched fragments from the gDNA sample, there would be a visible band or smear at approx. 500bp (the size of the sheared input DNA) in lane 7. There is an obvious enrichment of the control fragment from the hMeDIP reaction, as shown by the visible bands in both Lane 3 and 5, and by the difference in sizes between the sample treated with *Kpn*I and the untreated sample. Because there was no cleanup reaction after running the PCR and subsequent restriction enzyme digest, we see leftover primer dimers in both samples at the bottom of the gel.

These results led us to conclude that while the kit was able to pull down the spiked-in control plasmid sequences, we were unable to pull down sufficient gDNA fragments enriched in 5hmC. This led us to try another type of kit before looking into other methods.

After receiving genuine mouse brain from which to extract fresh DNA (Section 3.3), we repeated the experiment in hope of a positive result, but once again we found no measurable enrichment of DNA using the Qubit instrument. Using the pulled-down DNA as template, we also ran a qPCR assay to detect potential 5hmC enrichment at positive control loci using the primer sequences supplied to us from Zymo for the Sequence Specific Detection of 5hmC (Section 3.3), but were unable to see any enrichment.

3.2.2 Active Motif hMeDIP

Because the Zymo hMeDIP kit didn't lead to any positive results, we decided to try a different kit, which uses an antibody directly specific to 5hmC. Once again, mouse brain (in fact, rat) DNA served as a positive control.

We first ran a reaction using 1µg (the maximum input) rat DNA, which yielded no enrichment, so a reaction was set up with 1µg (the maximum input) rat DNA, 1µg human gDNA, and 1µg human gDNA spiked with 50pg of the control oligo from the kit. Again, there was no obvious enrichment from Qubit measurements, though we were able to see enrichment from the spiked control, confirming that the kit was able to enrich it's own control oligo.

In order to calculate enrichment levels at various loci, we set up a new set of IP (Immunoprecipitation) reactions, this time including a negative control using rabbit IgG. Human gDNA samples were spiked with three different control oligos, the first being an oligo where the cytosine bases were all hydroxymethylated, the second with only methylated cytosine, and the third with unmodified cytosines (unDNA). A negative control sample was also included in the experiment, comprised of human gDNA and the 5hmC control DNA, but was subject to Rabbit IgG instead of the 5-hydroxymethylcytidine antibody during immunoprecipitation. The control DNA samples can all be amplified during qPCR using a primer set that amplifies the human *APC* locus, which contains only unmodified cytosines. The *APC* locus has never been shown to be methylated in human gDNA, so the only expected enrichment should come from the spiked sample. A rat hMeDIP was also run in parallel to the human hMeDIP – 1µg of rat DNA alongside an identical sample, which was subject to rabbit IgG instead of the 5-hydroxymethylcytidine antibody. No control DNA was added to the rat reactions, as the primers that amplify it are made for human gDNA. Once again, we saw no significant signs of enrichment from the unspiked samples from both Qubit measurements and the qPCR.

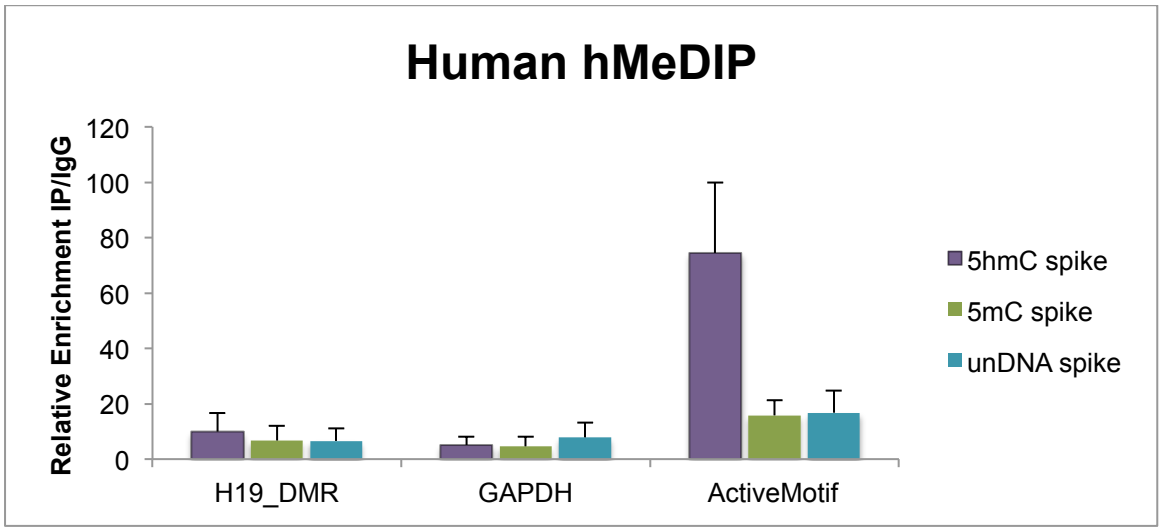


Figure 17: qPCR assay measuring enrichment of spiked human gDNA samples from the Active Motif hMeDIP kit – relative enrichment on the Y-axis represents the amount of IP DNA recovered from each spiked sample normalized against the negative control rabbit IgG reaction. Primers used in the control experiment are listed along the X-axis, where Active Motif refers to the APC primer set supplied with the kit. Error bars represent standard error. UnDNA spike refers to oligo without any modified cytosine bases.

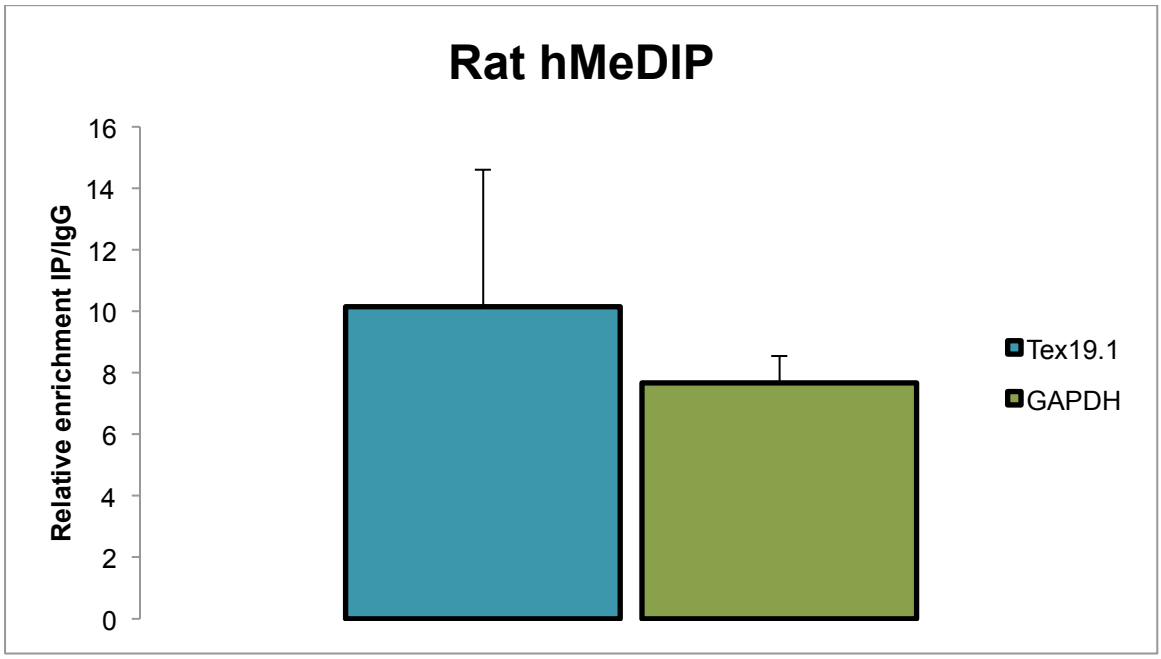


Figure 18: Enrichment of rat brain DNA samples from Active Motif hMeDIP kit – relative enrichment on the Y-axis represents the amount of IP DNA recovered from each sample normalized against the negative control Rabbit IgG reaction. qPCR products are shown along the X-axis. Error bars represent standard error.

Relative enrichment of each spiked reaction was calculated from qPCR data. As expected, we see significant enrichment from the gDNA sample spiked with 5hmC control DNA (t-test $p<0.05$), but we also see similar enrichment from the samples containing methylated and unmodified spikes (Figure 17). Both the *H19_DMR* and *GAPDH* loci also show insignificant levels of enrichment from all of the spiked samples, though these levels are much lower than from the Active Motif primer set, which amplified the *APC* locus. There should be an increase in enrichment in spiked samples as compared to genomic DNA alone, but the only amplification we saw was from the spiked samples, none in the gDNA samples.

The rat brain DNA hMeDIP showed similar results to the human gDNA hMeDIP, though there is no spike-in control to compare them to. The relative enrichment levels at both the *Tex19.1* and *GAPDH* loci are comparable to the enrichment levels of *H19_DMR* and *GAPDH* in the human hMeDIP with control DNA. Enrichment at *Tex19.1* and *GAPDH* was calculated to be insignificant using a t-test (Figure 18).

Before moving on to a different method, we tried concentrating our DNA samples to 12µl and 10µl using both Qiagen spin columns and Ampure XP beads to hopefully see some enrichment. Unfortunately we were still unable to detect any DNA on the Qubit. In a final

effort to understand why the hMeDIP reactions weren't working correctly, we ran a gel to check for DNA degradation. As seen in the gel image below (Figure 19), no obvious degradation is visible. Although disappointing, we were nonetheless forced to concede that the commercial hMeDIP kits were not sensitive and or efficient enough to successfully pull down 5hmC from DNA other than artificial oligos. We therefore decided to focus our efforts on a different method.

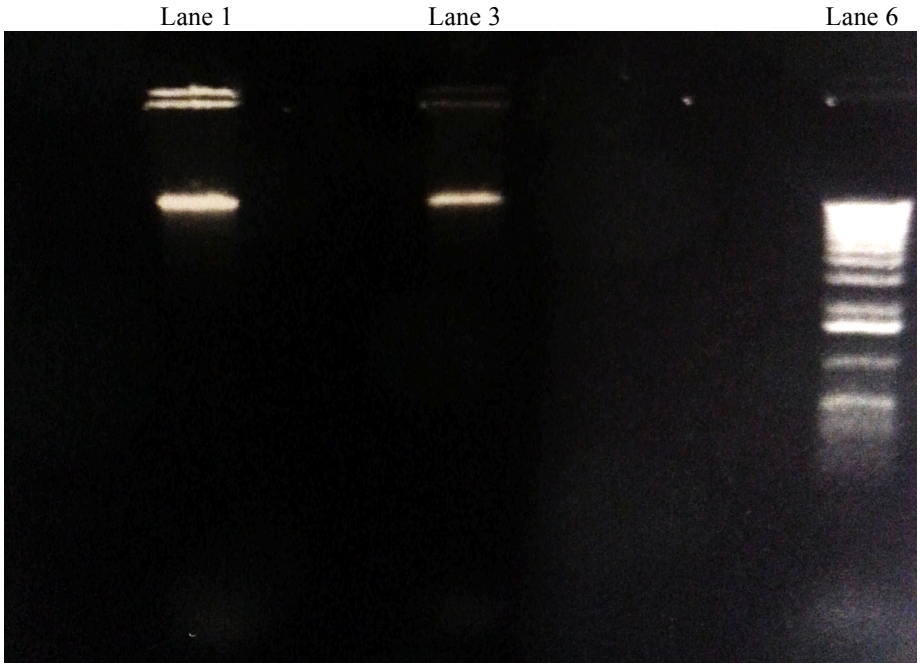


Figure 19: 8% TBE Agarose Gel, 100V, 30 min. Lanes 1&3: Mouse brain DNA supplied by Rikshospitalet. Lane 6: 1 KB ladder

3.3 Sequence Specific Detection of Hydroxymethylcytosine

3.3.1 DNA Isolation

Having exhausted our supply of mouse brain (in fact, rat) DNA for use as a positive control, we were required to isolate more. To this end, we obtained whole mouse brain from John Arne Dahl at Rikshospitalet, from which we isolated DNA. This material was used as a positive control for subsequent experiments.

After isolation of DNA from mouse brain using the DNEasy Blood & Tissue Kit (Qiagen), the purity of the DNA was checked on the NanoDrop (Thermo Scientific). A 260/280 ratio of ~1.8 and a 260/230 ratio of ~2.2 is expected for DNA samples. Contamination from proteins,

which absorb at 270nm, can impact the 260/280 ratio, while phenol absorbs at 230 and 270nm, so it can impact both the 260/280 ratio and 260/230 ratio. In Figure 20, the 260/280 ratio is much lower than expected, indicating the presence of a protein contaminant. Removing it required a phenol:chloroform extraction followed by EtOH precipitation.

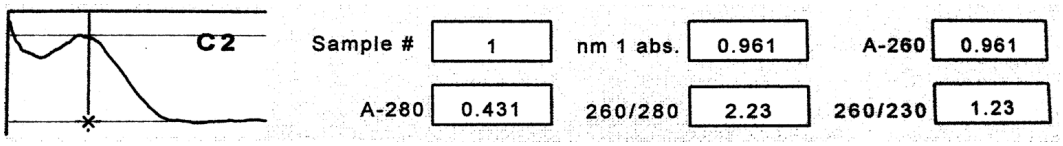


Figure 20: NanoDrop result before Phenol: Chloroform cleanup; notable protein contamination.

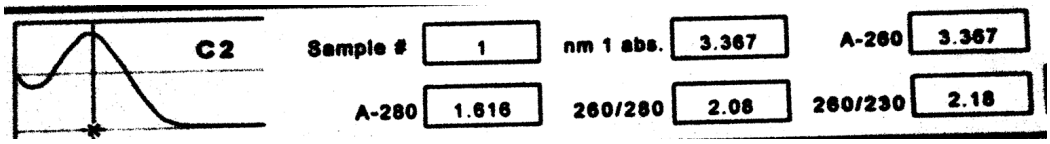


Figure 21: NanoDrop result after Phenol: Chloroform cleanup, no obvious sources of contamination.

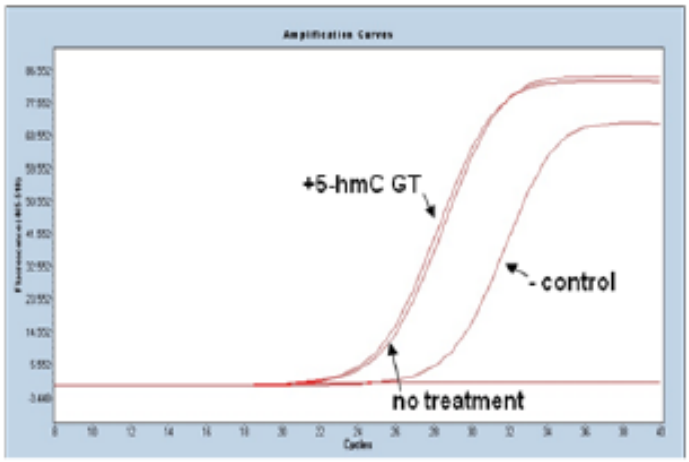
Post phenol:chloroform extraction, one can see in Figure 21 that the 260/280 ratio has normalized, and is closer to 1.8. We determined that the isolated DNA was cleaned up and potential contaminants had been removed, and proceeded on to the next step, detecting 5hmC in a locus-specific context.

3.3.2 Detection of 5-hydroxymethylcytosine using Zymo Quest-5hmC Detection Kit

The first reaction we ran was comprised of only the control oligo supplied from the kit, as to not waste valuable mouse brain DNA. 50ng of control oligo was incubated with 5hmC Glucosyltransferase, and a reaction was ran in parallel without the enzyme. Next, both samples were subject to a restriction enzyme digest by *Msp*I, and subsequently cleaned up using a spin column. 5hmC glucosyltransferase tags 5hmC with a glucose moiety, and blocks *Msp*I from digestion when the inner cytosine at the recognition site (C<u>C</u>GG) is glycosylated.

When checking for enrichment of 5hmC at a particular locus using qPCR, differences in amplification efficiencies will be seen between the glycosylated and unglycosylated samples and indicate the presence of 5hmC. As seen in Figure 22, the negative control sample (-control) indicates a level of 0% hydroxymethylation at the locus, and a no treatment sample

(no 5hmC glucosyltransferase or *Msp*I treatment) represents complete hydroxymethylation at the interrogation site. The experimental sample should lie between these two values.



MspI Digestion qPCR

Sample	Cp Value
no treatment	25.1
- control	28.5
+5-hmC Glucosyltransferase	24.9

Figure 22: Detection of 5hmC by qPCR (Zymo Research Corp.)

We ran a qPCR using the provided qPCR primers on the control oligo, a 90bp oligo with one *Msp*I recognition site with the inner cytosine at the recognition site being hydroxymethylated. An untreated sample was used calculate 100% hydroxymethylation and the negative control used to calculate 0% hydroxymethylation. The experimental sample was then calculated to have 136% hydroxymethylation at the interrogation site (Figure 23).

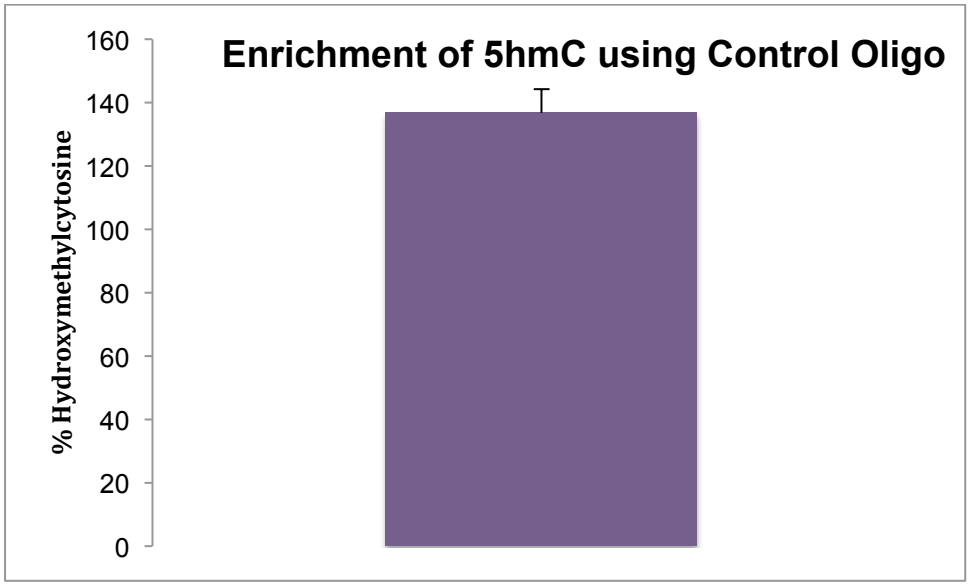


Figure 23: Locus specific detection of 5hmC using Zymo control oligo and control primer set.
% Hydroxymethylcytosine detected on the Y-axis. Error bar represents standard deviation.

The mouse brain sample was then subject to incubation with first 5hmC Glucosyltransferase, and thereafter digestion by *Msp*I (+glycosylation reaction). A reaction run in parallel without

the glucosyltransferase was also run (-glycosylation reaction). As expected, we were able to detect DNA at concentrations of 16.6 ng/µl (+ glycosylation) and 17.9 ng/µl (-glucosoylation). The variation is most likely due to the usage of spin columns to remove leftover enzyme and buffers from the reaction.

We received the sequences of four qPCR primers (two sets of positive controls and two sets of negative controls) used by Zymo Research to detect enrichment at specified loci in mouse brain. A qPCR was then run in the same fashion as with the control oligo in hopes of seeing enrichment in our brain tissue samples, which would allow us to move on to library preparation of our samples for sequencing.

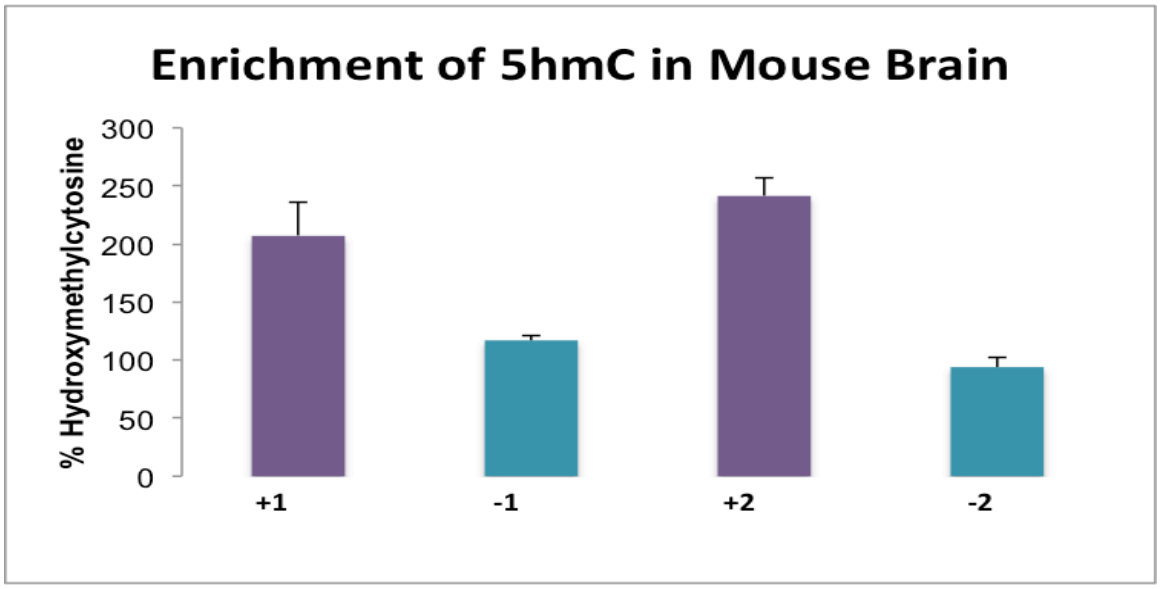


Figure 24: Enrichment of 5hmC in Mouse brain using qPCR primers provided from Zymo Research. Primer sets used are represented on the X-axis (Mouse Brain +1, -1, +2, -2), and % Hydroxymethylcytosine on Y-axis. Error bars represent standard error.

We can see from Figure 24 that we were indeed able to detect significant levels of enrichment when comparing the percent 5hmC at the positive and negative control loci (t-test, $p > 0.05$). Just as in our qPCR reaction with the control oligo, we detected hydroxymethylation values above 100% for each amplicon. However, because we saw a significant difference in enrichment between the positive and negative loci, we concluded that the reaction had worked correctly, and we continued with the same protocol this time

using CD4+ DNA from three sets of control co-twins (six DNA samples altogether), yielding 12 human DNA samples and 2 mouse DNA samples ready for library preparation.

3.3.3 Library Preparation

As we were able to show the enrichment of 5hmC for the mouse control samples using the Quest-5hmC Detection Kit (Zymo), we proceeded with library preparation for sequencing. 100ng of each sample was subject to library preparation using TruSeq DNA reagents as outlined in section 2.4.4. Each sample was ligated to a specific adapter. Directly before the size-separation step, the +glycosylation and –glycosylation samples were pooled to avoid the introduction of false-positive differences caused by slight differences in size selection. Two bands from a 2% TAE Agarose gel were excised; one from 150-300bp, and the second from 300-766bp (Figures 25 & 26). The most intense signals, which appear as "bands" in the gel-like image, indicate repetitive elements cut by the MspI enzyme.

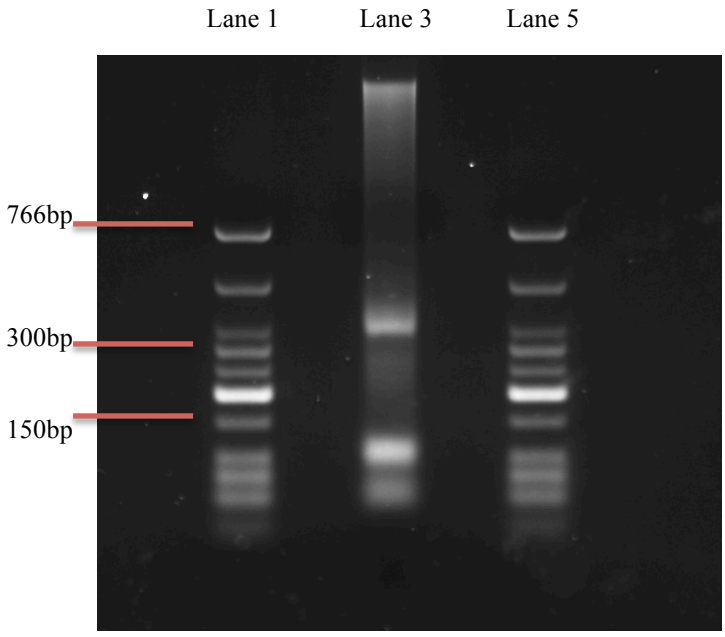


Figure 25: Image of 2% TBE gel before size selection, Lanes 1 and 3 contain Low Molecular Weight Ladder (NEB), and Lane 3 contains the sample.

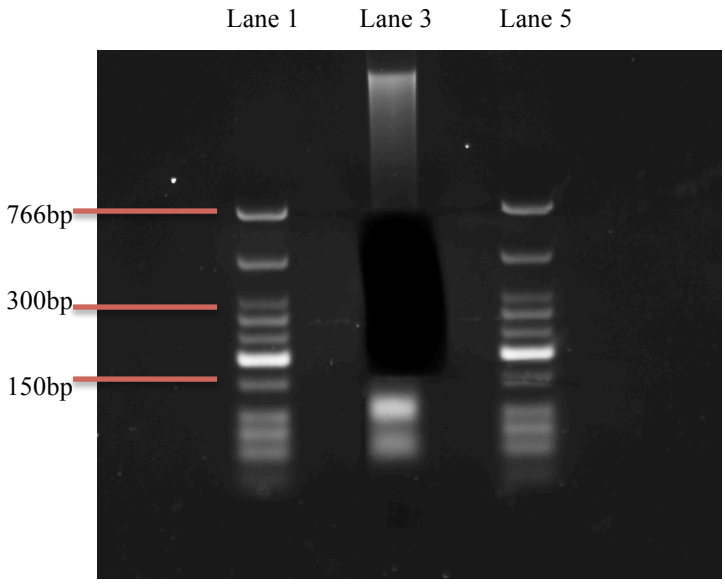


Figure 26: Image of 2% TBE gel after excision of 150-300bp and 300-766bp fragments. Contents of the lanes are the same as Figure 25

Proper size selection assures that correctly sized fragments are available for sequencing, as well as removing primer dimers, which will interfere with sequencing, and amplify in the subsequent PCR reaction. While optimizing the library preparation, we were also able to reduce the number of PCR cycles from 13 to 12, with the aim of minimizing PCR-induced bias during sequencing.

After PCR and the following Ampure XP bead cleanup, finished libraries were run on the Bioanalyzer 2100 using High Sensitivity chips and reagents to check for primer dimer contamination, and that the libraries were the correct sizes. In Figures 27 and 28, we can see that the libraries are of the correct sizes, and there is no visible peak at 120bp, which would indicate primer dimer contamination.

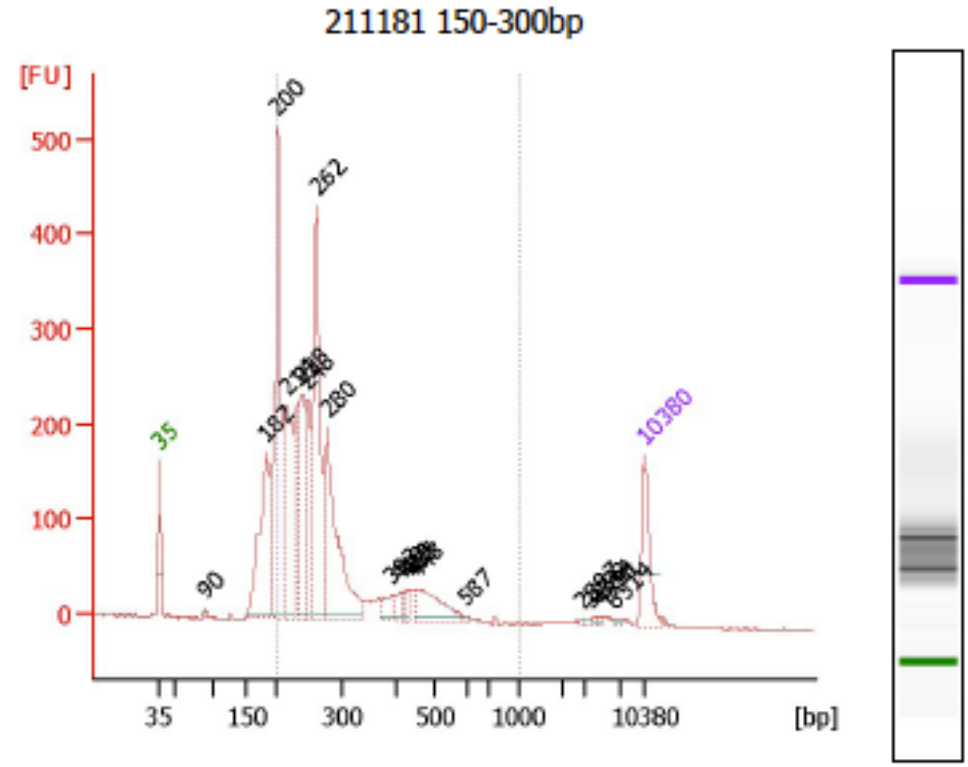


Figure 27: BioAnalyzer image of pooled + and – glycosylation reactions for twin 211181, with a spread from roughly 150-300bp, with peaks at 200 and 262bp. BP is represented on the X-axis, and Fluorescence Units on the Y-axis.

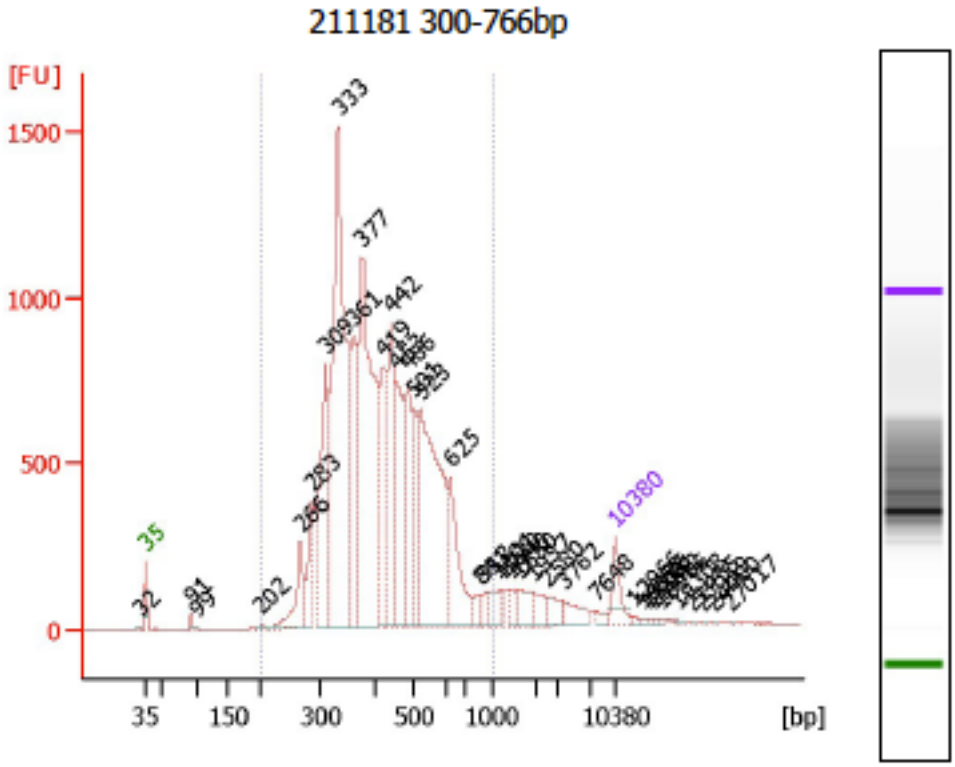


Figure 28: BioAnalyzer image of pooled + and – glycosylation reactions for twin 211181, with a spread from roughly 300-766bp with a large peak at 333bp and a lesser peak at 377bp. X- and Y-axes are the same as above.

3.4 Sequencing

3.4.1 Quality of the Sequenced Reads

A FASTQ file, where the quality for each base of a read is scored, was provided for all of the sequenced samples. The quality scores can be used to exclude low quality reads from analysis. As the samples were sequenced in a 100bp paired-end reaction, quality scores are provided for each member of the pair. Although all the sequenced samples passed quality control, we ended up with fewer reads than expected, and qualities that declined more rapidly than is normally seen. A quality score of 20 (indicative of a 1 in 100 chance of error) is considered acceptable, while a quality score above 28 is considered very good. A quality score below 20 is considered insufficient.

We can see from the Per Base Sequence Quality graph (Figure 29) that the quality of the reads decreases quite rapidly after position 70 on the graph, and while a decrease of quality at the end of a read is normally seen when using Illumina technology, this is a much more rapid decrease than usually seen.

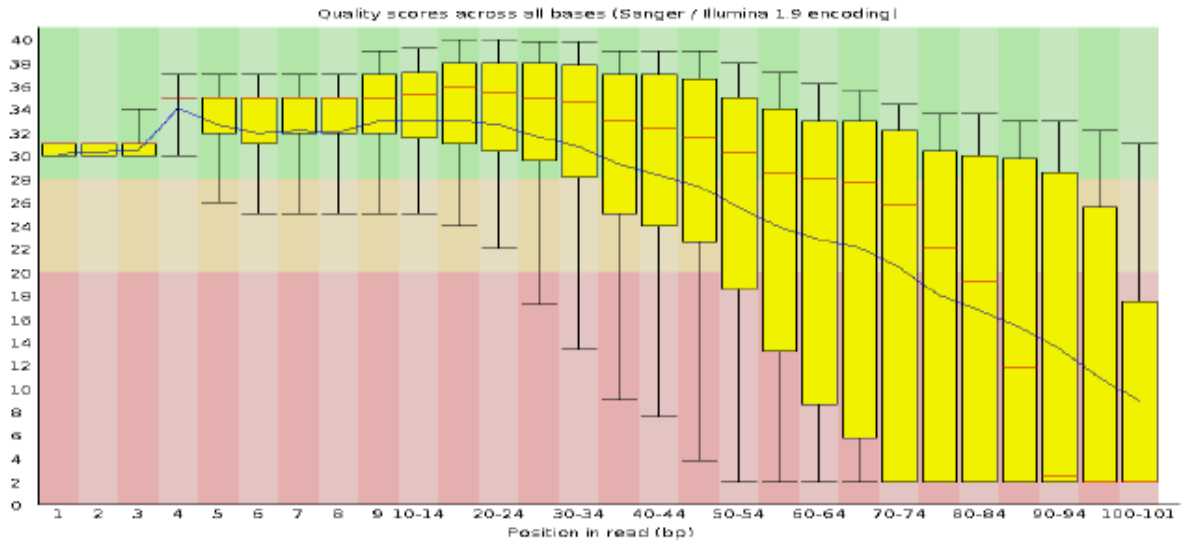


Figure 29: Per Base Sequence Quality of Twin Sample 14511, insert size 150-300bp fragments, read 1. The range of quality scores across all bases at each position in the FASTQ file is shown. Quality scores are plotted on the Y-axis, and read position on the X-axis. Yellow boxes represent the IQR (25-75%) range of the reads, and the whiskers represent 10% and 90%. The red and blue lines display the median and mean of the quality values.

The distribution of quality scores for all reads should also be examined to see the spread of the quality of reads from the sample, and potentially allow us to exclude samples with too low quality from analysis. Ideally we would see a narrow peak at 38, indicating that all of the reads have a high quality. Instead, we see in Figure 30 that the spread of our quality scores is quite broad, with a peak at quality score 29. Large portions of our reads have a quality score below 20, which is the bottom range of a read having acceptable quality. During analysis, we excluded low quality reads and trimmed our reads to 50bp to remove the low quality portions of the reads (Section 2.7).

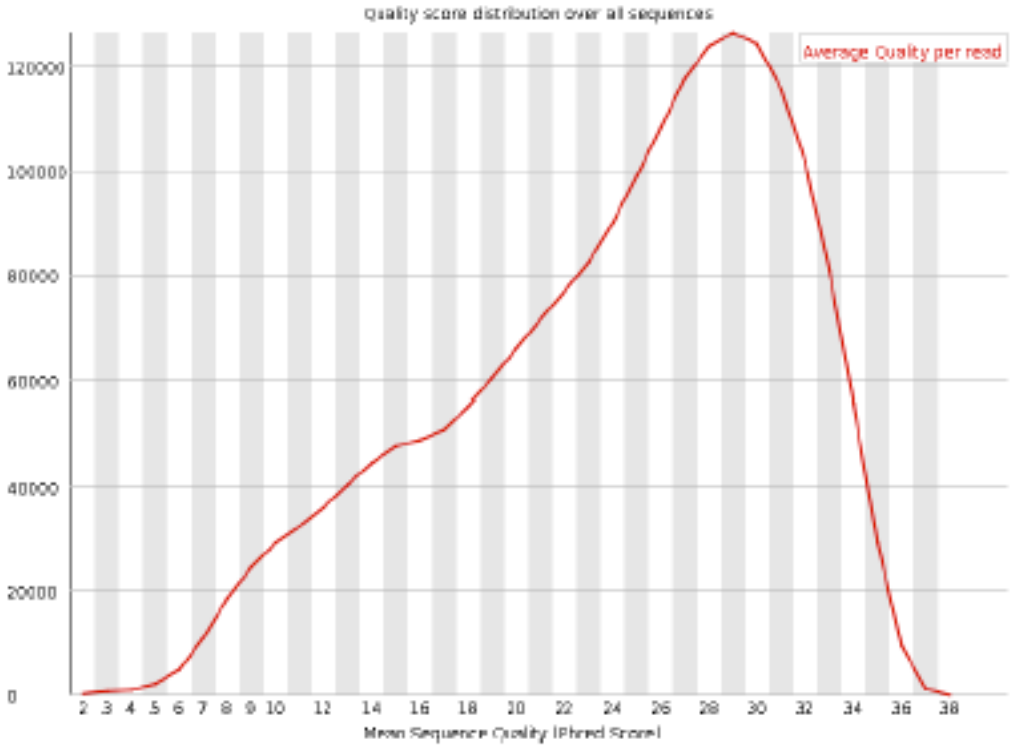


Figure 30: Per Sequence Quality Score over all sequences. Average quality score is plotted on the X-axis and number of reads on the Y-axis.

3.5 Sequencing Analysis

After the reads were trimmed and aligned to the respective reference genomes, we were able to see the type of coverage we received from sequencing. An overview of the mapped reads for each sequenced sample is seen in Figure 31, and we can see that a significant portion of our reads are mapping uniquely to the reference genomes. Interestingly, we saw a trend of

our treatment samples having on average a 20% decrease in uniquely mapped reads as compared to our control samples. Full details of the reads can be found in Appendix 5.3, Table 4.

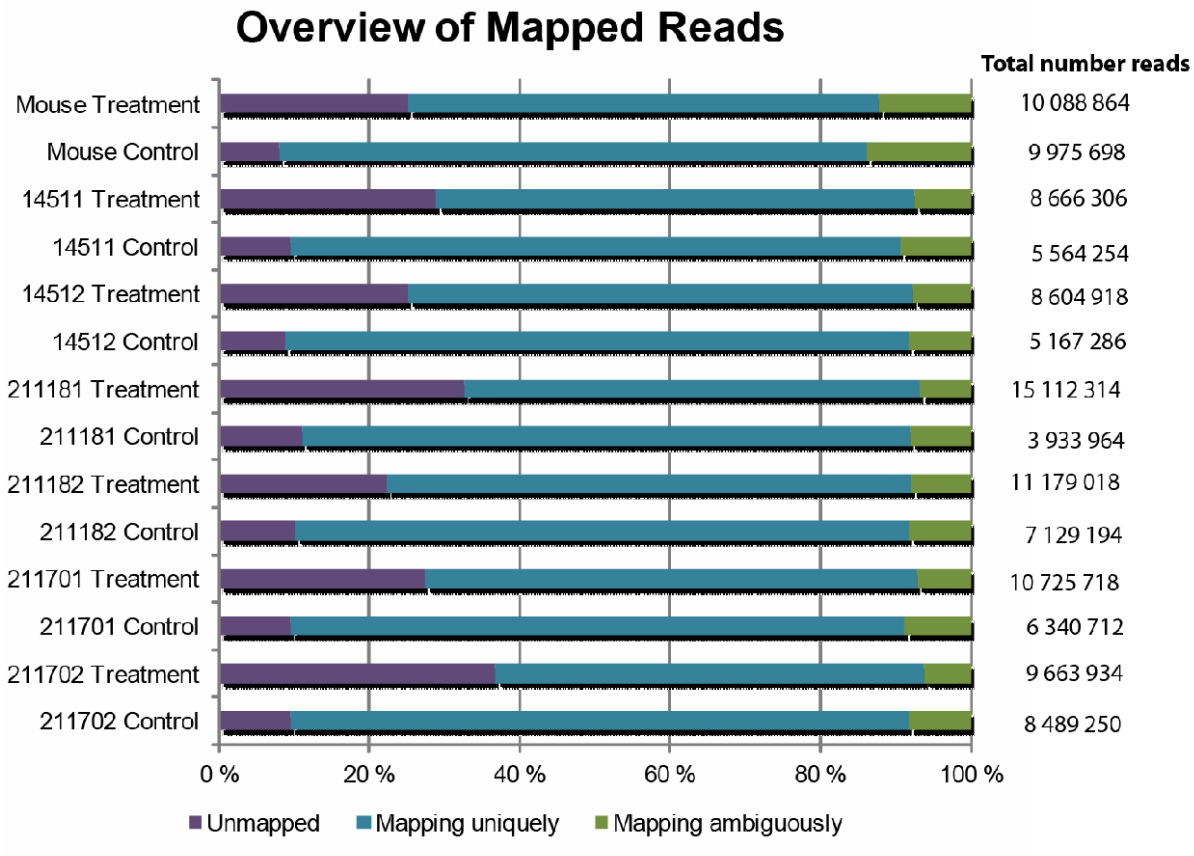


Figure 31: Overview of mapped reads from sequencing. Samples are listed on the X-axis, and percentage of reads that were unmapped, mapped uniquely and mapped ambiguously are shown on the Y-axis. Treatment refers to samples treated with both 5hmC glucosyltransferase and *Msp*I, and Control refers to samples only treated with *Msp*I.

After alignment, MACS was used to identify significant peaks. While MACS called a large number of significant peaks for each sample, there were only a few that were differentially called between the control and treatment samples (Table 1).

Sample	Number of peaks	Different +/- (FDR 5%)
Mouse	3338	0
14511	9010	5
14512	12579	10
21181	4823	0
21182	7814	5
211701	8972	0
211702	6302	5

Table 1: Summary of significant peaks called using MACS for each sequenced sample. Twin samples are identified through ID-numbers. Sample name in column one, number of peaks called by MACS in column two and number of peaks with significant differences between the + glycosylation and - glycosylation control in column 3.

3.5.1 Mouse Brain Control

Because the experimental sample treated with 5hmC glucosyltransferase blocked *Msp*I from cutting at the recognition site when 5hmC is present, we expected to see a difference between the experimental and control samples. While the treated sample will show coverage at an area where 5hmC is present, the control sample will lack coverage, due to *Msp*I cleavage. Alternatively, the 5hmC glucosyltransferase treated samples are too large, and in that case we would only identify the control samples. Another possible outcome from sequencing is opposite of the aforementioned scenario, where two *Msp*I recognition sites are close together the control samples are too small (minimum size for sequencing is 30 bp) for sequencing, and we would only see coverage from the treated sample.

Though we were unable to identify any significant peaks in the mouse control samples, in Figure 32 we can see nonetheless see two such examples identified by visual inspection showing differences in coverage at *Msp*I recognition sites. In *Cdk2, Msp*I was unable to cut at the recognition site in the glycosylated sample, resulting in two distinct peaks. The control sample has two corresponding regions with a lack of coverage. It is likely that increased sequencing depth would result in a lack of visible difference at this site. We can also see that the identified regions are located in a CpG island, thus showing that we have positively identified the presence of 5hmC in our sample. We expected to find that most of the regions

identified would have no visible differences between the control and experimental samples, an example of such is seen in Figure 33 at *Atp5e*.

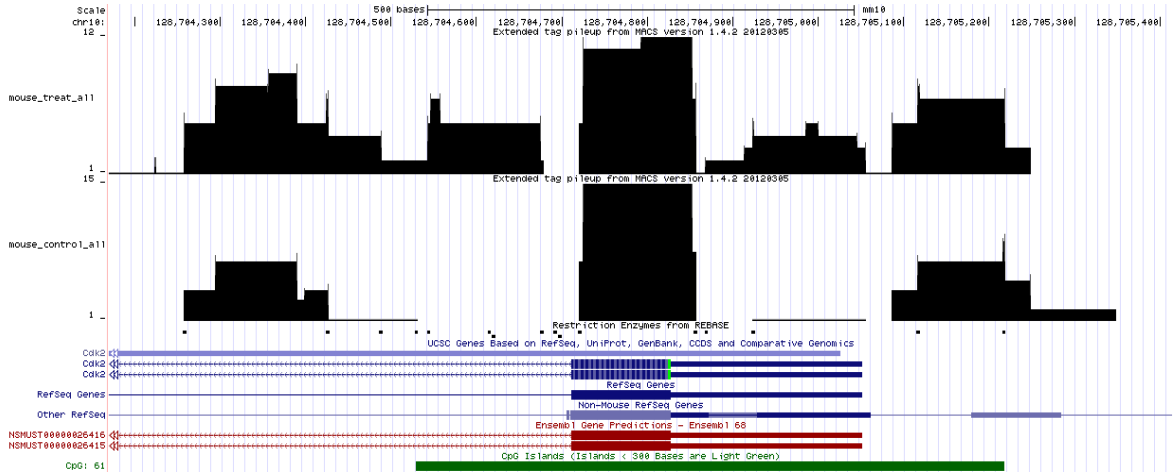


Figure 32: Screenshot from UCSC Genome Browser (genome.ucsc.edu) at *Cdk2* in the mouse genome. From top to bottom, the tracks included in the image are the experimental sample treated with 5hmC Glucosyltransferase, the control sample, *Msp*I recognition sites from REBASE, gene tracks from UCSC, RefSeq and Ensembl, and at the bottom, CpG islands.

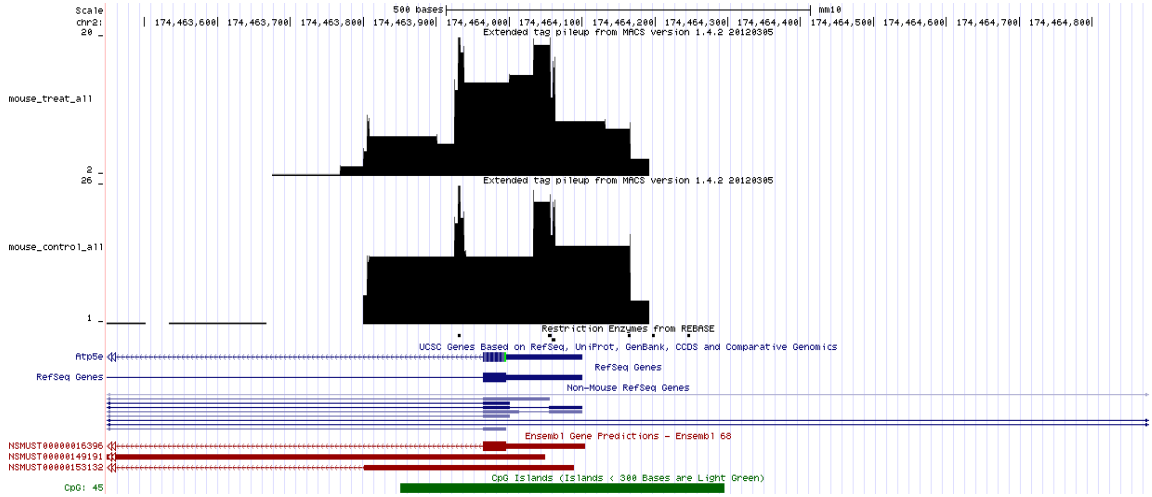


Figure 33: Screenshot from UCSC Genome Browser at *Atp5e* in the mouse genome. Tracks are the same as in Figure 31.

3.5.2 Control Twin Samples

In addition to mouse brain DNA, the Quest-sequencing approach was also applied to genomic DNA isolated from CD4+ cells from 3 pairs of human monozygotic twins. These twin pairs were not discordant for any disease phenotype, but represent control individuals. As in the mouse samples, we expected to detect sites of 5hmC based on the differences between glycosylated and non-glycosylated samples. While we were unable to identify any

significant peaks in the mouse samples, our human samples had a number of significant peaks. Below in Figure 34 we see such a region in *REXO1L1*, which codes for RNA Exonuclease 1-homolog (S. cerivisiae)-like 1.In the middle of the figure, in the promoter region of the gene, there is a region with significant enrichment in the treatment sample, with no corresponding enrichment in the control sample, showing the presence of 5hmC. While there is a visible difference in enrichment between the peaks on the right-hand side of the figure, it was not deemed significant.

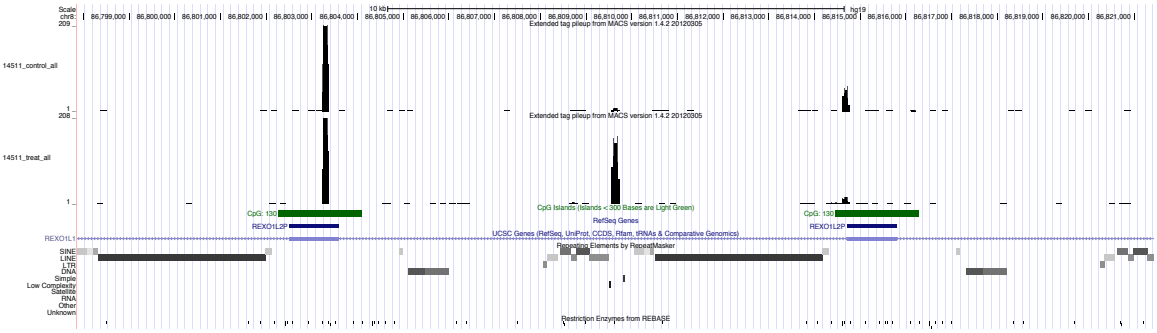


Figure 34: Screenshot of *REXO1L1* coding region from UCSC Genome Browser. The tracks in ascending order are; control sample, treatment sample, CpG Islands, RefSeq Genes, Repetitive Elements from Repeat Masker and *Msp*I sites from REBASE.

We expected to see significant differences in enrichment in CpG islands, due to the nature of *Msp*I cleaving at CpG sequences, which tend to overlap with promoter regions in genes. This was not the case, as most of the significant peaks mapped to non-coding regions of the genome, but we were able to see regions with good coverage. One such region is shown in Figure 35 – the *IFG2* coding region, which was used as one of our positive control loci for our hMeDIP qPCR experiments. While we were unable to detect any significant differences between the control and experimental samples in this region, we do see that we have good coverage from the sequencing. We did not see any enrichment in the region of our *IGF2* qPCR amplicon (shown in red on image) used as a positive control for our hMeDIP experiments. This could be due to lack of sequencing depth, but a different region with *Msp*I sites that show enrichment from our sequencing results can be used to make a better positive control for future experiments.

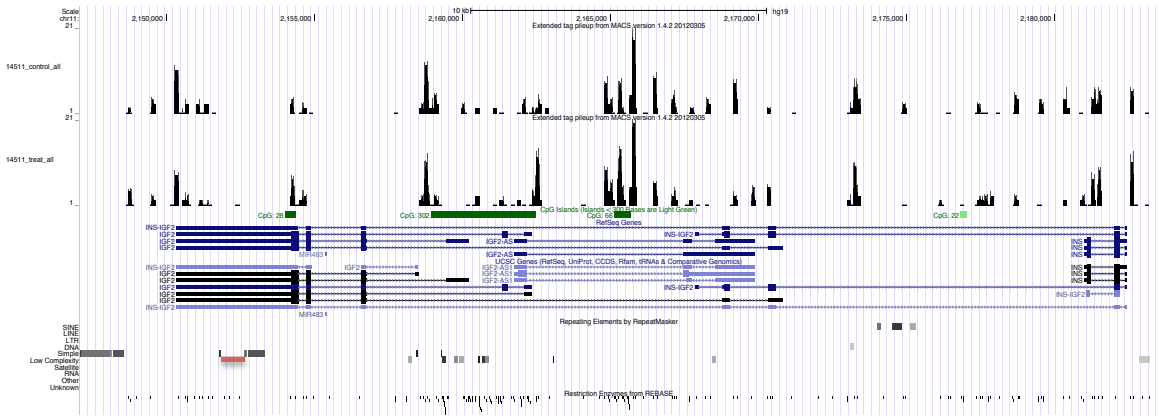


Figure 35: Screenshot of *IGF2* coding region from UCSC Genome Browser. Tracks are the same as in Figure 34. The red bar on the bottom left of the image corresponds to our qPCR amplicon used as a positive control in our hMeDIP experiments.

We were only able to detect significant peaks in four of our twin samples, and only one of the twin pairs. All four of the twin samples had significant peaks at *LINC00601*, three had a significant peak at *PWRN1*, and two shared a significant peak at *DGKB*. The only significant peak found in common between a twin pair was at *LINC00601*.

4 Discussion

4.1 ELISA

The detection of 5hmC by ELISA was considered an essential prerequisite to this project, as proceeding to the other methods would have been futile if no 5hmC was detectable in human DNA isolated from CD4+ cells. Encouragingly, we were able to detect 5hmC in both the mouse brain and human twin CD4+ samples. It was interesting to note that the differences in hydroxymethylation in the diseased and sick twins were not significant. This suggests that gross differences in 5hmC levels do not underlie the disease phenotype. However, it remains possible that locus-specific differences could be important, the detection of which requires a more sensitive and specific assay.

The values we found for 5hmC concentration in mouse brain tissue and CD4+ DNA differed from published results. While we found mouse brain to have a 0.15% 5hmC content, and CD4+ DNA to have a 0.02-0.06% 5hmC content, previous studies have identified 0.6% overall hydroxymethylation in brain tissue (Kriaucionis and Heintz 2009), and about 1000 times lower in blood (Nestor, Ottaviano et al. 2012). Our experimentally determined value for 5hmC content in brain tissue was about four times lower than the previously determined amount, while our CD4+ DNA had over 100 times the amount of the published values. This discrepancy is most likely due to the assay type used. Other reported values of 5hmC in a tissue or blood specific context have mainly been quantified using Mass Spectrometry, which is much more sensitive than an ELISA assay. Another possible source of error while performing the ELISA experiment could have been overdevelopment during the detection step of the protocol, where the directions ambiguously stated to wait until the solution had turned ‘medium blue’ to add the stop solution.

4.2 hMeDIP

Although there have been positive results in identifying 5hmC enriched loci using hMeDIP published (Nestor, Ottaviano et al. 2012), we were unable to obtain positive results from our hMeDIP experiments. This was most surprising for our mouse brain samples, as we expected to see significant enrichment due to the higher concentrations of 5hmC than our other samples. While there has been seen a global reduction of 5hmC in cell cultures (Nestor, Ottaviano et al. 2012), our control DNA was isolated directly from brain tissue, and shouldn’t

be subject to this phenomenon. In the future, if further experiments were to be done using the isolated CD4+ DNA twin material from cell cultures, then this would need to be taken into account.

A likely possibility for the lack of enrichment in the hMeDIP could be due to problems with the antibody used. Both the Zymo Research and Active Motif hMeDIP kits were both able to enrich their own positive control oligos, which have 19 hydroxymethylated cytosines in a 90 bp oligo and many hydroxymethylated cytosines in an unspecified plasmid control, respectively. It is reasonable to assume that the density and high percentage of 5hmC in these artificial control constructs led to positive results from the immunoprecipitations. 5hmC isn't found in these quantities or densities in our experimental DNA samples, which most likely had an effect on the lack of results we saw. In the case of the 5-hydroxymethylcytidine antibody from the Active Motif kit, which also pulled down significant amounts of the non-hydroxymethylated control oligos, the antibody may be rather non-specific, though this leads to even more confusion as to why we were unable to detect any enrichment.

Our qPCR experiments were also unable to detect enrichment at the loci we chose to make qPCR amplicons. We may have had more success with different loci, but we were unable to find other enriched loci at the time of the experiment. Other possible regions in human DNA that we could have used for human controls include the *HOXA* cluster and *TEX14* (Nestor, Ottaviano et al. 2012). As 5hmC is further studied, more regions should be identified as being 5hmC enriched, and should represent suitable positive controls for such qPCR reactions

It is of notable interest that both the *IGF2* and *H19* genes were both previously identified as being enriched for 5hmC, as they are reciprocally imprinted and expressed almost exclusively from the paternal and maternal chromosomes, respectively, during both zygote and postnatal development (Kaffer, Grinberg et al. 2001). As 5hmC also plays an important role in development, coinciding with global demethylation, further research could provide insight into the role of 5hmC and imprinting. Interestingly our qPCR experiments were unable to detect any enrichment at loci. This is a cause for concern, as these loci were supposed to be used as positive controls, though we may have been able to see significant differences between the treatment and control samples if we had greater sequencing depth.

After finishing the hMeDIP experiments we were informed (by a technician at Rikshospitalet) that we had actually received rat brain DNA instead of mouse brain DNA. While the *GAPDH* qPCR primer set created for mouse mapped to the rat genome, the *Tex19.1* primer set did not. Interestingly, we saw amplification when making a standard curves for the qPCR experiments using the mouse *Tex19.1* primer, due to conserved sequences. Rat brain DNA has been shown to have similar amounts of 5hmC in comparison to human and mouse brain tissue (Wang, Guan et al. 2011), so one could expect that the DNA isolated from rat brain should also have worked with our hMeDIP kits. The hMeDIP procedure was also repeated using the DNA isolated from mouse brain in our lab, but we failed to detect enrichment by qPCR.

4.3 Sequence Specific Detection of 5hmC

As we were unable to show any enrichment of 5hmC using our qPCR primer sets, we contacted Zymo Research and received from them the sequences of four primer sets that they had previously tested and showed detection of 5hmC in mouse brain tissue. The first primer set, comprised of a positive and negative control both mapped to *PAX9*, which codes for a transcription factor active in early embryo development in mice. The second set, also contained positive and negative control primers, but mapped to noncoding regions of the genome. Most likely it had been previously determined that the *Msp*I sites in the positive control primers had a 5hmC base at the *Msp*I recognition site, and the negative control primers had no hydroxymethylation at the recognition site, making them suitable for detecting the presence of 5hmC at these sites.

While we were able to show that our glycosylation reactions had worked, and had blocked *Msp*I from cutting at hydroxymethylated cytosine residues at the recognition site, we were puzzled as to why we had values that showed over 100% hydroxymethylation at specified loci, both using the control oligo as well as our experimental samples. The protocol from the kit states that from a qPCR reaction, a level of 100% hydroxymethylation is established by the no treatment control DNA, and 0% hydroxymethylation established by the negative control, and the glycosylated sample should therefore lie in between those values. We see in Figure 22 that the glycosylated sample has a slightly lower Cp value than the glycosylated sample, meaning that the producers of the kit also should have seen values of over 100% hydroxymethylation in the glycosylated sample. We are unsure as to why this is the case, but

a possibility is the treatment of the sample DNA during the protocol, specifically being passed over spin columns during cleanup, results in a greater loss of the untreated sample.

4.4 Sequencing

We assume that the quality of the reads was low due to the fact that all samples were treated with *Msp*I, leading to all our reads starting with the same sequence. This sequence bias in the initial cycles can lead to a loss of fidelity in cluster identification, thereby causing a reduction in the number of reads (Krueger, Andrews et al. 2011). Interestingly, this phenomenon hasn't been previously seen in sequencing of RRBS samples (also cut with *Msp*I) in our lab. To mitigate this problem, a PhiX (viral genome) control can be added to the libraries during clustering, shifting the base composition of the library and adding diversity to the base composition of the library for future experiments.

4.5 Sequencing Analysis

While we were able to identify a distinctive number of peaks in each sample, we saw at most 10 significant peaks in a specific sample. The most reasonable explanation for the lack of significant peaks is low sequencing depth. If the experiment were to be repeated with the addition of a PhiX control during cluster generation, we can expect to have higher quality sequences, and therefore a higher sequencing depth. It is also of interest that our treatment samples had a decrease in uniquely mapped reads in comparison to the control samples, as the treatment samples had on average, more reads than the control samples. The control samples were cut more often than the treatment with *Msp*I leading to shorter fragments, so a possibility is that the shorter fragments were more successful in mapping to the reference genome.

When investigating the locations of the significant peaks in our human samples, we found that all of them had a significant peak at *LINC00601*, a long interspersed non-coding RNA. Three of the four samples with significant peaks also called a significant peak at *PWRN1*, a non-protein coding RNA in the Prader-Willi region. Both of these peaks are found in repetitive regions, and interestingly enough LINE1 elements have previously been found to contain 5% 5hmC content (Booth, Branco et al. 2012), and these results could further strengthen the hypothesis that 5hmC may be involve in demethylation of specific repeat

classes. Locations of other significant peaks called by MACS can be found in Table 5 in Appendix 5.3.

Due to the lack of depth from sequencing, we were unable to extract more positional information from the peaks, though if the samples were to be re-sequenced with PhiX, we can assume that many more significant peaks would be found in CpG islands near coding regions.

4.6 Conclusion

We were able to show the presence of 5hmC from DNA isolated from both mouse brain tissue and CD4+ cells, though there was a discrepancy in that the mouse brain sample had a lower amount than previously published, while the CD4+ DNA samples had a higher amount than previously seen. Again, this is most likely due to sensitivity of the assay used.

While we had great difficulty in finding a method for investigating 5hmC at a single base resolution, we can conclude that we were able to do so using Zymo's Quest 5hmC Detection Kit in combination with appropriate size selection of the digested DNA, and high throughput sequencing. In the future this combination of reactions could be optimized and become a new method for detecting 5hmC, and different restriction enzymes can be used for different coverage. We were able to identify a number of regions that had significant differences between the control and treatment samples. While many of these regions were found in repetitive sequences and noncoding regions, they were identified in multiple samples. Further sequencing is required to fully validate the method.

4.7 Further Research

Unfortunately due to time constraints, we were unable to investigate 5hmC in the discordant twin pairs. As we were able to identify 5hmC in both the mouse and twin samples, as well as see similarities in the twin samples, the next logical step would be to repeat the experiment, with the inclusion of the PhiX control during clustering, and using the discordant twins, hopefully identify differences in them that have implications in disease causing genes in psoriasis. Additionally, positional data of 5hmC could be compared to data from previous methylation analyses to see if any cytosines that were previously identified as being methylated are in fact hydroxymethylated.

We experienced unexpected difficulties in the project, most specifically getting the hMeDIP reaction to yield enriched 5hmC fragments. This was particularly surprising, given that the reagents used were commercially available kits sold specifically for this purpose. As 5hmC is a relatively new DNA modification to study, in the future there will hopefully be a number of more robust methods conceived. As it was most likely the low concentration of 5hmC in our sample DNA which led our hMeDIP reactions to fail, it would be advantageous for future research if a method was developed to take into account the minute amount of 5hmC found in DNA, especially blood and cell extracts.

While the mechanism for how 5mC is converted to 5hmC has been explained, the question of what regulatory role, if any, 5hmC performs in the genome still exists. It has been shown that an increase of 5hmC during development coincides with global demethylation (Wu and Zhang 2011), though it is still not understood exactly what role 5hmC has during development. Similarly, its function in gene expression has yet to be elucidated. The TET enzymes responsible for the conversion of 5mC to 5hmC have been shown to have different roles in other biological processes (Tan and Manley 2009), but it is still unclear as to whether or not they are functionally redundant in regards to 5hmC.

Lastly, the reason for the high concentration of 5hmC in Purkinjie neurons is still unknown. High levels of 5hmC in frontal lobe tissue has been shown to be correlated to gene expression, and further research on the topic could provide insight into neurodevelopment.

5 Appendix

5.1 Protocols

5.1.1 DNA Quantification

Qubit

Qubit is a fluorometric method of quantifying DNA concentration.

1. Prepare a working solution of 199 µl Qubit dsDNA HS Buffer and 1 µl Qubit dsDNA HS Reagent per sample.
2. Add 190 µl of the working solution to the tubes used for standards, 199 µl for each sample tube.
3. Add 10 µl of each Qubit standard or 1 µl sample to the appropriate tubes.
4. Vortex 2-3 seconds to mix.
5. Incubate samples at room temperature for 2 minutes.
6. Calibrate Qubit Fluorometer using standards.
7. Insert sample tubes one by one into Qubit Fluorometer and read off DNA concentration.

Bioanalyzer

The Bioanalyzer (Agilent) is a microfluidics platform that allows quantification, sizing, and quality control of nucleic acid samples. In this project, both DNA High Sensitivity and DNA 1000 chips were used. High sensitivity chips were used for quality control of libraries, and DNA 1000 chips were used for sizing of genomic DNA.

Fragmentation of DNA

Prior to the hMeDIP protocol, DNA must be fragmented into ca. 500 bp in length; otherwise extremely long fragments of DNA will the pulled down, making the subsequent sequencing costly and of low resolution. DNA was fragmented using a Covaris S220 sonicator. First, DNA is brought up to 130 µl using 1X TE Buffer, and then sonicated for 80 seconds at a 10% duty factor, 140-peak power, 200 cycles per burst. Post-sonication, the sample was reconcentrated using a QIAQuick PCR Purification (Qiagen) column and eluted in 30 µl (Appendix 5.1.2). To ensure that the DNA has been correctly fragmented, samples were run on a BioAnalyzer (Agilent) DNA 1000 Chip.

5.1.2 DNA Isolation, Cleanup and Concentration

Isolation of DNA from Mouse Brain

DNA from mouse brain was prepared using DNeasy Blood & Tissue kit from Qiagen. Previously dissected brain from a mouse were brought to room temperature, and cut up into smaller 25 mg pieces. The pieces of brain were then incubated in a lysis solution for 8 hours at 56°C until the solution was homogenic. The solution was then passed over a provided spin column, and yielded 30µg DNA. The DNA concentration and purity of the sample was checked on the Qubit Fluorometer (Invitrogen) and NanoDrop (Thermo Scientific), respectively (Appendix 5.1.1).

Phenol: Chloroform Cleanup of DNA

When NanoDrop measurements indicated an excess of protein left over from the column-based cleanup using the DNeasy Blood & Tissue Kit (Qiagen), the DNA was subjected to additional cleanup by Phenol: Chloroform Extraction (Chomczynski and Sacchi 2006). Briefly, an equal volume of Phenol: Chloroform is added to the DNA solution, vortexed, and spun in a centrifuge for 2 minutes at top speed. The top aqueous layer containing DNA is transferred over into a fresh tube, while the protein-containing interphase and organic layers are discarded. The procedure is then repeated using pure Chloroform, and again, the aqueous phase containing DNA is kept, while the other phases are discarded.

To reconcentrate DNA, a simple Ethanol precipitation protocol is followed. Along with 0.1 volumes of NaOAc, 3 volumes of chilled (-20°C) 100% EtOH are added to the aqueous phase from the Phenol: Chloroform cleanup. The solution is spun in a centrifuge at 6°C for 30 minutes, and then a DNA pellet should be visible. The supernatant is removed, and the pellet is washed with chilled (-20°C) 70% EtOH. The solution is once again spun in a centrifuge at top speed for 15 minutes, 4°C. Once finished, the supernatant is removed, and the pellet is air-dried on the lab bench. Once all traces of alcohol are gone, the pellet is resuspended in 50µl TE Buffer. The quality and concentration of the DNA can then again be verified on the NanoDrop and Qubit (Appendix 5.1.1).

Ampure XP Cleanup

AmpureXP beads were used in either a 1:1 or 1.8:1 bead to DNA solution ratio, depending on the required outcome. A 1.8:1 ratio will remove all products less than 100bp from the solution, while a 1:1 ratio removes fragments around 200bp and lower.

1. Add Ampure XP beads to DNA solution in desired ratio.
2. Incubate the solution for 15 minutes at room temperature.
3. Place tube on magnetic stand and incubate for 5 minutes.
4. Remove supernatant without disturbing the bead pellet.
5. Add 200 µl 80% EtOH, incubate 30 seconds, and remove supernatant.
6. Repeat wash step.
7. Air dry pellet to remove traces of ethanol for 7-8 minutes
8. Remove from magnetic rack and resuspend beads in desired volume.
9. Incubate solution 2 minutes, room temperature.
10. Place tube back on magnetic rack and incubate for 5 minutes or until solution is clear.
11. Transfer supernatant containing DNA into fresh tube.

5.1.3 PCR & qPCR Reactions

Standard qPCR Reaction

Master Mix:

iTaq Universal SYBR Green Supermix:	5 µl
10X Primer Mix (10 pmol/µl) (F+R):	1 µl
ddH_20:	2 µl
DNA Standard/Sample:	2 µl
Total:	**10 µl**

Amplification Conditions

50°C – 2 min

95°C – 10 min

95°C – 15s
60°C – 10s } 40 cycles

95°C – 15s
60°C – 15s } Melting Curve Analysis
95°C – 15s

Zymo Quest 5hmC Enrichment Kit

Reaction setup:

5hmC enriched DNA:	1µl
Control Primers:	1µl
QuestTaq Premix:	10µl
dH_20:	8µl
Total:	**20µl**

Amplification Conditions

95°C – 3 min

95°C – 30s
59°C – 30s } 28 Cycles
72°C – 30s

72°C – 1 min

4°C – hold

***Kpn*I Enzyme Cut**

NEB 1 Buffer:	1µl
Amplified DNA from PCR:	5µl
BSA:	1µl
*Kpn*I Enzyme:	1µl
dH_2O:	2µl
Total:	**10µl**

5.1.4 Illumina Library Preparation

End Repair

Set up reaction on ice; incubate at 30C for 30 min:

100ng 5hmC DNA:	x µl
Resuspension Buffer (RSB):	x µl
End Repair Mix (ERP):	20 µl
Total:	**50 µl**

Adenylate 3' Ends

Set up reaction on ice; incubate at 37C for 30 min, followed by 70C for 5 minutes. Cool on bench to room temperature, and then place on ice.

DNA from End Repair:	17.5 µl
A-tailing Mix (ATL):	12.5 µl
Total:	**30 µl**

Ligate Adapters

Set up reaction on ice; incubate at 30C for 10 minutes.

Adenylated DNA:	30 µl
RSB:	2.5 µl
Ligase Mix (LIG):	2.5 µl
1:4 Diluted Adapters:	2.5 µl
Total:	**37.5 µl**

PCR Amplification

Set up reaction on ice and amplify according to specified conditions.

DNA from Adapter Ligation:	20 µl
TruSeq PCR primer cocktail:	5 µl
TruSeq PCR master mix:	25 µl
Total:	**50 µl**

Amplification Conditions

98C – 30s

98C – 10s

60C – 30s } 13 cycles

72C – 30s

72C – 5 min

5.2 Primers

Table 1: qPCR primers - Human

Oligo Name	Sequence 5'→3'	Amplicon Size
H19_DMR_A1 FWD	GATCTCGGCCCTAGTGTGAA	188bp
H19_DMR_A1 REV	GTGATGTGTGAGCCTGCACT	
H19_genic_4 FWD	GCCAGCTACACCTCCGTTG	137bp
H19_genic_4 REV	AGCTAGGGCTGGAAAGAAGG	
IFG2_genic_1 FWD	CATGAAATTTGGGGGTTCC	112bp
IGF2_genic_1 REV	GGGAGTTCTGGGGTAGGAAG	
H19_promoter_2 FWD	CCTGGAATTCTCCAAAGACG	115bp
H19_promoter_2 REV	AGTGGTCTGGGAGGGAGAAG	
GADPH FWD	CGGCTACTAGCGGTTTTACG	189bp
GADPH REV	AAGAAGATGCGGCTGACTGT	

Table 2: qPCR primers – Mouse

Oligo Name	Sequence 5'→3'	Amplicon Size
Tex19.1 FWD	AAAATGGGCCACCCACATCTC	184bp
Tex19.1 REV	CCACTGGCCCTTGGACCAGAC	
GADPH FWD	CCTGCGACTTCAACAGCAACTCCCA	135bp
GADPH REV	TGAGGTCCACCACCCTGTTGCTGTA	

Table 3: qPCR Primers from Zymo for Mouse Brain

Oligo Name	Sequence 5'→3'	Amplicon Size
Mouse Brain +1 FWD	GCAGTTTCGTCTCAGCATCC	223 bp
Mouse Brain +1 REV	GCAGAAGCGGTCACAGAATG	
Mouse Brain +2 FWD	ATGGGTTCTCTGGCTCGATT	191 bp
Mouse Brain +2 REV	CTGGCTGATTTTTGGAAGGA	
Mouse Brain -1 FWD	CTTGTCCAAGTGGCGTTTTC	218 bp
Mouse Brain -1 FWD	CACTGCACTGCCTCTGTCAA	
Mouse Brain -2 FWD	GGCATTGGGATGTGAACAGT	207 bp

Mouse Brain -2 REV	CCCACCTCACCTAAAATCAGTG	

5.3 Sequencing Data

Table 4: Overview of Mapped Reads from Sequencing. Treatment reactions are denoted with a suffix of 1, while control reactions have a suffix of 2.

Reads	Unmapped	Mapping uniquely	Mapping ambiguously	Total
Mouse_1	2521996 (25%)	6338995 (62.8%)	1227873 (12.2%)	10088864
Mouse_2	789215 (7.9%)	7802009 (78.2%)	1384474 (13.9%)	9975698
14511_1	2501516 (28.9%)	5508094 (63.6%)	656696 (7.6%)	8666306
14511_2	532938 (9.6%)	4506496 (81%)	524820 (9.4%)	5564254
14512_1	2159998 (25.1%)	5775229 (67.1%)	669691 (7.8%)	8604918
14512_2	450999 (8.7%)	4293034 (83.1%)	423253 (8.2%)	5167286
211181_1	4917059 (32.5%)	9185952 (60.8%)	1009303 (6.7%)	15112314
211181_2	436104 (11.1%)	3181015 (80.9%)	316845 (8.1%)	3933964
211182_1	2491774 (22.3%)	7809323 (69.9%)	877921 (7.9%)	11179018
211182_2	715289 (10%)	5817984 (81.6%)	595921 (8.4%)	7129194
211701_1	2930641 (27.3%)	7025473 (65.5%)	769604 (7.2%)	10725718
211701_2	606819 (9.6%)	5169886 (81.5%)	564007 (8.9%)	6340712
211702_1	3541944 (36.7%)	5525601 (57.2%)	596389 (6.2%)	9663934
211702_2	813594 (9.6%)	6967312 (82.1%)	708344 (8.3%)	8489250

Table 5: Locations of peaks identified to be significant by MACS

Sample	chr	start	end	length	fold_enrichment	FDR(%)
14511	chr1	228767450	228767608	159	7.84	0
14511	chr10	128106460	128106604	145	2.49	0
14511	chr15	24768644	24768768	125	6.58	0
14511	chr7	14728958	14729096	139	9.48	0
14511	chr8	86809573	86809730	158	33.14	0
14512	chr10	128106462	128106602	141	5.96	0
14512	chr12	84068408	84068731	324	808.7	0
14512	chr16	32169678	32169834	157	592.89	0
14512	chr19	36788825	36788998	174	6.63	0
14512	chr19	36791011	36791230	220	29.22	0
14512	chr21	9720851	9721001	151	114.73	0
14512	chr3	185671850	185672097	248	16.9	0
14512	chr5	102476275	102476578	304	126.48	0
14512	chr6	131863139	131863296	158	9.6	0
14512	chr7	1077150	1077295	146	330.43	0
21182	chr10	128106452	128106612	161	3.44	0
21182	chr11	68243257	68243801	545	32.82	0
21182	chr11	112005739	112005874	136	2.84	0
21182	chr15	24768640	24768788	149	2.95	0

21182	chrY	4277181	4277506	326	10.7	0
211702	chr10	128106442	128106622	181	2.17	0
211702	chr15	24768635	24768798	164	2.34	0
211702	chr16	71197186	71197322	137	220.91	0
211702	chr3	15790547	15790915	369	648.94	0
211702	chr7	14728953	14729096	144	2.25	0

6 References

Ballestar, E. (2010). "Epigenetics lessons from twins: prospects for autoimmune disease." Clinical reviews in allergy & immunology **39**(1): 30-41.

Bardella, M. T., C. Fredella, et al. (2000). "Gluten sensitivity in monozygous twins: a long-term follow-up of five pairs." The American journal of gastroenterology **95**(6): 1503-1505.

Bell, J. T. and T. D. Spector (2011). "A twin approach to unraveling epigenetics." Trends in genetics : TIG **27**(3): 116-125.

Bestor, T. H. (2000). "The DNA methyltransferases of mammals." Human molecular genetics **9**(16): 2395-2402.

Blander, G., A. Bhimavarapu, et al. (2009). "SIRT1 promotes differentiation of normal human keratinocytes." J Invest Dermatol **129**(1): 41-49.

Bogdanos, D. P., D. S. Smyk, et al. (2012). "Twin studies in autoimmune disease: genetics, gender and environment." Journal of autoimmunity **38**(2-3): J156-169.

Booth, M. J., M. R. Branco, et al. (2012). "Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution." Science.

Bourc'his, D., G. L. Xu, et al. (2001). "Dnmt3L and the establishment of maternal genomic imprints." Science **294**(5551): 2536-2539.

Branco, M. R., G. Ficz, et al. (2012). "Uncovering the role of 5-hydroxymethylcytosine in the epigenome." Nature reviews. Genetics **13**(1): 7-13.

Brandeis, M., D. Frank, et al. (1994). "Sp1 elements protect a CpG island from de novo methylation." Nature **371**(6496): 435-438.

Brandrup, F., M. Hauge, et al. (1978). "Psoriasis in an unselected series of twins." Archives of dermatology **114**(6): 874-878.

Carlson, L. L., A. W. Page, et al. (1992). "Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting." Genes & development **6**(12B): 2536-2541.

Cedar, H. and Y. Bergman (2009). "Linking DNA methylation and histone modification: patterns and paradigms." Nature reviews. Genetics **10**(5): 295-304.

Cedar, H. and Y. Bergman (2012). "Programming of DNA methylation patterns." Annual review of biochemistry **81**: 97-117.

Chen, C. C., K. Y. Wang, et al. (2012). "The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases." The Journal of biological chemistry **287**(40): 33116-33121.

Chomczynski, P. and N. Sacchi (2006). "The single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty-something years on." Nature protocols **1**(2): 581-585.

Cimmino, L., O. Abdel-Wahab, et al. (2011). "TET family proteins and their role in stem cell differentiation and transformation." Cell stem cell **9**(3): 193-204.

Cortellino, S., J. Xu, et al. (2011). "Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair." Cell **146**(1): 67-79.

Dang, L., D. W. White, et al. (2009). "Cancer-associated IDH1 mutations produce 2-hydroxyglutarate." Nature **462**(7274): 739-744.

Dawlaty, M. M., K. Ganz, et al. (2011). "Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development." Cell stem cell **9**(2): 166-175.

Delhommeau, F., S. Dupont, et al. (2009). "Mutation in TET2 in myeloid cancers." N Engl J Med **360**(22): 2289-2301.

Duffy, D. L., L. S. Spelman, et al. (1993). "Psoriasis in Australian twins." Journal of the American Academy of Dermatology **29**(3): 428-434.

Eden, A., F. Gaudet, et al. (2003). "Chromosomal instability and tumors promoted by DNA hypomethylation." Science **300**(5618): 455.

Epsztejn-Litman, S., N. Feldman, et al. (2008). "De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes." Nature structural & molecular biology **15**(11): 1176-1183.

Esteller, M. (2007). "Epigenetic gene silencing in cancer: the DNA hypermethylome." Human molecular genetics **16 Spec No 1**: R50-59.

Feldman, N., A. Gerson, et al. (2006). "G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis." Nature cell biology **8**(2): 188-194.

Ficz, G., M. R. Branco, et al. (2011). "Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation." Nature **473**(7347): 398-402.

Figueroa, M. E., O. Abdel-Wahab, et al. (2010). "Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation." Cancer cell **18**(6): 553-567.

Fraga, M. F., E. Ballestar, et al. (2005). "Epigenetic differences arise during the lifetime of monozygotic twins." Proceedings of the National Academy of Sciences of the United States of America **102**(30): 10604-10609.

Fu, Y. and C. He (2012). "Nucleic acid modifications with epigenetic significance." Current opinion in chemical biology **16**(5-6): 516-524.

Gervin, K., M. D. Vigeland, et al. (2012). "DNA methylation and gene expression changes in monozygotic twins discordant for psoriasis: identification of epigenetically dysregulated genes." PLoS genetics **8**(1): e1002454.

Goelz, S. E., B. Vogelstein, et al. (1985). "Hypomethylation of DNA from benign and malignant human colon neoplasms." Science **228**(4696): 187-190.

Goll, M. G. and T. H. Bestor (2005). "Eukaryotic cytosine methyltransferases." Annual review of biochemistry **74**: 481-514.

Greco, L., R. Romino, et al. (2002). "The first large population based twin study of coeliac disease." Gut **50**(5): 624-628.

Gross, S., R. A. Cairns, et al. (2010). "Cancer-associated metabolite 2-hydroxyglutarate accumulates in acute myelogenous leukemia with isocitrate dehydrogenase 1 and 2 mutations." The Journal of experimental medicine **207**(2): 339-344.

Groth, A., W. Rocha, et al. (2007). "Chromatin challenges during DNA replication and repair." Cell **128**(4): 721-733.

Gu, T. P., F. Guo, et al. (2011). "The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes." Nature **477**(7366): 606-610.

Hackett, J. A., R. Sengupta, et al. (2012). "Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine." Science.

Hackett, J. A., R. Sengupta, et al. (2013). "Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine." Science **339**(6118): 448-452.

Haffner, M. C., A. Chaux, et al. (2011). "Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers." Oncotarget **2**(8): 627-637.

Hall, J. G. (1996). "Twins and twinning." American journal of medical genetics **61**(3): 202-204.

He, Y. F., B. Z. Li, et al. (2011). "Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA." Science **333**(6047): 1303-1307.

Hershey, A. D. and M. Chase (1952). "Independent functions of viral protein and nucleic acid in growth of bacteriophage." J Gen Physiol **36**(1): 39-56.

Huang, Y., W. A. Pastor, et al. (2010). "The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing." PloS one **5**(1): e8888.

Huang, Y., W. A. Pastor, et al. (2012). "The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine." Nature protocols **7**(10): 1897-1908.

Inoue, A. and Y. Zhang (2011). "Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos." Science **334**(6053): 194.

Iqbal, K., S. G. Jin, et al. (2011). "Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine." Proceedings of the National Academy of Sciences of the United States of America **108**(9): 3642-3647.

Ito, S., A. C. D'Alessio, et al. (2010). "Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification." Nature **466**(7310): 1129-1133.

Jenuwein, T. and C. D. Allis (2001). "Translating the histone code." Science **293**(5532): 1074-1080.

Jeon, Y., K. Sarma, et al. (2012). "New and Xisting regulatory mechanisms of X chromosome inactivation." Current opinion in genetics & development **22**(2): 62-71.

Jin, S. G., S. Kadam, et al. (2010). "Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine." Nucleic acids research **38**(11): e125.

Jin, S. G., X. Wu, et al. (2011). "Genomic mapping of 5-hydroxymethylcytosine in the human brain." Nucleic acids research **39**(12): 5015-5024.

Jones, P. A. and P. W. Laird (1999). "Cancer epigenetics comes of age." Nat Genet **21**(2): 163-167.

Kaffer, C. R., A. Grinberg, et al. (2001). "Regulatory mechanisms at the mouse Igf2/H19 locus." Molecular and cellular biology **21**(23): 8189-8196.

Kaprio, J., J. Tuomilehto, et al. (1992). "Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a population-based cohort of twins in Finland." Diabetologia **35**(11): 1060-1067.

Kendler, K. S. and C. A. Prescott (1999). "A population-based twin study of lifetime major depression in men and women." Archives of general psychiatry **56**(1): 39-44.

Keohane, A. M., P. O'Neill L, et al. (1996). "X-Inactivation and histone H4 acetylation in embryonic stem cells." Developmental biology **180**(2): 618-630.

Kinney, S. R. and S. Pradhan (2013). "Ten eleven translocation enzymes and 5-hydroxymethylation in mammalian development and cancer." Advances in experimental medicine and biology **754**: 57-79.

Koh, K. P., A. Yabuuchi, et al. (2011). "Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells." Cell stem cell **8**(2): 200-213.

Kriaucionis, S. and N. Heintz (2009). "The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain." Science **324**(5929): 929-930.

Krueger, F., S. R. Andrews, et al. (2011). "Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by deferred cluster calling." PloS one **6**(1): e16607.

Kudo, Y., K. Tateishi, et al. (2012). "Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation." Cancer Sci **103**(4): 670-676.

Kuratomi, G., K. Iwamoto, et al. (2008). "Aberrant DNA methylation associated with bipolar disorder identified from discordant monozygotic twins." Molecular psychiatry **13**(4): 429-441.

Laurent, L., E. Wong, et al. (2010). "Dynamic changes in the human methylome during differentiation." Genome research **20**(3): 320-331.

Levasseur, D. N., J. Wang, et al. (2008). "Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells." Genes & development **22**(5): 575-580.

Lujambio, A., G. A. Calin, et al. (2008). "A microRNA DNA methylation signature for human cancer metastasis." Proceedings of the National Academy of Sciences of the United States of America **105**(36): 13556-13561.

MacGregor, A. J., H. Snieder, et al. (2000). "Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins." Arthritis and rheumatism **43**(1): 30-37.

Maiti, A. and A. C. Drohat (2011). "Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites." The Journal of biological chemistry **286**(41): 35334-35338.

Matarese, F., E. Carrillo-de Santa Pau, et al. (2011). "5-Hydroxymethylcytosine: a new kid on the epigenetic block?" Mol Syst Biol **7**: 562.

Matsuda, A. and T. Kuzuya (1994). "Diabetic twins in Japan." Diabetes research and clinical practice **24 Suppl**: S63-67.

Mayer, W., A. Niveleau, et al. (2000). "Demethylation of the zygotic paternal genome." Nature **403**(6769): 501-502.

Meissner, A., A. Gnirke, et al. (2005). "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis." Nucleic acids research **33**(18): 5868-5877.

Mikkelsen, T. S., M. J. Wakefield, et al. (2007). "Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences." Nature **447**(7141): 167-177.

Mill, J., E. Dempster, et al. (2006). "Evidence for monozygotic twin (MZ) discordance in methylation level at two CpG sites in the promoter region of the catechol-O-methyltransferase (COMT) gene." American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics **141B**(4): 421-425.

Nakao, M. (2001). "Epigenetics: interaction of DNA methylation and chromatin." Gene **278**(1-2): 25-31.

Nestor, C., A. Ruzov, et al. (2010). "Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA." Biotechniques **48**(4): 317-319.

Nestor, C. E., R. Ottaviano, et al. (2012). "Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes." Genome research **22**(3): 467-477.

Nickoloff, B. J. and F. O. Nestle (2004). "Recent insights into the immunopathogenesis of psoriasis provide new therapeutic opportunities." J Clin Invest **113**(12): 1664-1675.

Oates, N. A., J. van Vliet, et al. (2006). "Increased DNA methylation at the AXIN1 gene in a monozygotic twin from a pair discordant for a caudal duplication anomaly." American journal of human genetics **79**(1): 155-162.

Okano, M., D. W. Bell, et al. (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.

Pan, W., S. Zhu, et al. (2010). "MicroRNA-21 and microRNA-148a contribute to DNA hypomethylation in lupus CD4+ T cells by directly and indirectly targeting DNA methyltransferase 1." Journal of immunology **184**(12): 6773-6781.

Pastor, W. A., U. J. Pape, et al. (2011). "Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells." Nature **473**(7347): 394-397.

Penn, N. W., R. Suwalski, et al. (1972). "The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid." Biochem J **126**(4): 781-790.

Petronis, A. (2006). "Epigenetics and twins: three variations on the theme." Trends in genetics : TIG **22**(7): 347-350.

Petronis, A., Gottesman, II, et al. (2003). "Monozygotic twins exhibit numerous epigenetic differences: clues to twin discordance?" Schizophrenia bulletin **29**(1): 169-178.

Portela, A. and M. Esteller (2010). "Epigenetic modifications and human disease." Nature biotechnology **28**(10): 1057-1068.

Quivoron, C., L. Couronne, et al. (2011). "TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis." Cancer cell **20**(1): 25-38.

Rakyan, V. K., T. A. Down, et al. (2010). "Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains." Genome research **20**(4): 434-439.

Rakyan, V. K., J. Preis, et al. (2001). "The marks, mechanisms and memory of epigenetic states in mammals." The Biochemical journal **356**(Pt 1): 1-10.

Robertson, J., A. B. Robertson, et al. (2011). "The presence of 5-hydroxymethylcytosine at the gene promoter and not in the gene body negatively regulates gene expression." Biochemical and biophysical research communications **411**(1): 40-43.

Robertson, K. D. (2005). "DNA methylation and human disease." Nature reviews. Genetics **6**(8): 597-610.

Ruchusatsawat, K., J. Wongpiyabovorn, et al. (2006). "SHP-1 promoter 2 methylation in normal epithelial tissues and demethylation in psoriasis." J Mol Med (Berl) **84**(2): 175-182.

Scholer, H. R. (1991). "Octamania: the POU factors in murine development." Trends in genetics : TIG **7**(10): 323-329.

Siegfried, Z., S. Eden, et al. (1999). "DNA methylation represses transcription in vivo." Nature genetics **22**(2): 203-206.

Sims, R. J., 3rd and D. Reinberg (2008). "Is there a code embedded in proteins that is based on post-translational modifications?" Nature reviews. Molecular cell biology **9**(10): 815-820.

Smallwood, S. A., S. Tomizawa, et al. (2011). "Dynamic CpG island methylation landscape in oocytes and preimplantation embryos." Nature genetics **43**(8): 811-814.

Sonkoly, E., T. Wei, et al. (2007). "MicroRNAs: novel regulators involved in the pathogenesis of psoriasis?" PLoS One **2**(7): e610.

Sorensen, A. L. and P. Collas (2009). "Immunoprecipitation of methylated DNA." Methods in molecular biology **567**: 249-262.

Spector, T. D., F. Cicuttini, et al. (1996). "Genetic influences on osteoarthritis in women: a twin study." BMJ **312**(7036): 940-943.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." Nature **403**(6765): 41-45.

Straussman, R., D. Nejman, et al. (2009). "Developmental programming of CpG island methylation profiles in the human genome." Nature structural & molecular biology **16**(5): 564-571.

Sullivan, P. F., K. S. Kendler, et al. (2003). "Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies." Archives of general psychiatry **60**(12): 1187-1192.

Sun, Z., T. Jolyon, et al. (2013). "High-Resolution Enzymatic Mapping of Genomic 5-Hydroxymethylcytosine in Mouse Embryonic Stem Cells." Cell reports.

Szulwach, K. E., X. Li, et al. (2011). "5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging." Nature neuroscience **14**(12): 1607-1616.

Szwagierczak, A., S. Bultmann, et al. (2010). "Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA." Nucleic acids research **38**(19): e181.

Tahiliani, M., K. P. Koh, et al. (2009). "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1." Science **324**(5929): 930-935.

Tan, A. Y. and J. L. Manley (2009). "The TET family of proteins: functions and roles in disease." Journal of molecular cell biology **1**(2): 82-92.

Tovar-Castillo, L. E., J. C. Cancino-Diaz, et al. (2007). "Under-expression of VHL and over-expression of HDAC-1, HIF-1alpha, LL-37, and IAP-2 in affected skin biopsies of patients with psoriasis." Int J Dermatol **46**(3): 239-246.

Valinluck, V. and L. C. Sowers (2007). "Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1." Cancer research **67**(3): 946-950.

Valinluck, V. and L. C. Sowers (2007). "Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1." Cancer Res **67**(3): 946-950.

Valinluck, V., H. H. Tsai, et al. (2004). "Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2)." Nucleic acids research **32**(14): 4100-4108.

Wang, H., S. Guan, et al. (2011). "Comparative characterization of the PvuRts1I family of restriction enzymes and their application in mapping genomic 5-hydroxymethylcytosine." Nucleic acids research **39**(21): 9294-9305.

Wang, T., Q. Pan, et al. (2012). "Genome-wide DNA hydroxymethylation changes are associated with neurodevelopmental genes in the developing human cerebellum." Human molecular genetics **21**(26): 5500-5510.

Wiberg, J. S. (1967). "Amber mutants of bacteriophage T4 defective in deoxycytidine diphosphatase and deoxycytidine triphosphatase. On the role of 5-hydroxymethylcytosine in bacteriophage deoxyribonucleic acid." The Journal of biological chemistry **242**(24): 5824-5829.

Wiklund, E. D., J. Kjems, et al. (2010). "Epigenetic architecture and miRNA: reciprocal regulators." Epigenomics **2**(6): 823-840.

Williams, K., J. Christensen, et al. (2012). "DNA methylation: TET proteins-guardians of CpG islands?" EMBO reports **13**(1): 28-35.

Williams, K., J. Christensen, et al. (2011). "TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity." Nature **473**(7347): 343-348.

Wolffe, A. P. and M. A. Matzke (1999). "Epigenetics: regulation through repression." Science **286**(5439): 481-486.

Wossidlo, M., T. Nakamura, et al. (2011). "5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming." Nat Commun **2**: 241.

Wrone-Smith, T. and B. J. Nickoloff (1996). "Dermal injection of immunocytes induces psoriasis." J Clin Invest **98**(8): 1878-1887.

Wu, H. and Y. Zhang (2011). "Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation." Genes & development **25**(23): 2436-2452.

Wu, H. and Y. Zhang (2011). "Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells." Cell cycle **10**(15): 2428-2436.

Xu, Y., F. Wu, et al. (2011). "Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells." Molecular cell **42**(4): 451-464.

Yang, P. K. and M. I. Kuroda (2007). "Noncoding RNAs and intranuclear positioning in monoallelic gene expression." Cell **128**(4): 777-786.

Yisraeli, J., R. S. Adelstein, et al. (1986). "Muscle-specific activation of a methylated chimeric actin gene." Cell **46**(3): 409-416.

Yoo, A. S., B. T. Staahl, et al. (2009). "MicroRNA-mediated switching of chromatin-remodelling complexes in neural development." Nature **460**(7255): 642-646.

Yu, M., G. C. Hon, et al. (2012). "Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome." Cell **149**(6): 1368-1380.

Zachariah, R. M. and M. Rastegar (2012). "Linking epigenetics to human disease and Rett syndrome: the emerging novel and challenging concepts in MeCP2 research." Neural plasticity **2012**: 415825.

Zhang, K., R. Zhang, et al. (2009). "Promoter methylation status of p15 and p21 genes in HPP-CFCs of bone marrow of patients with psoriasis." Eur J Dermatol **19**(2): 141-146.

Zhang, P., Y. Su, et al. (2012). "Epigenetics and psoriasis." J Eur Acad Dermatol Venereol **26**(4): 399-403.