The role of CpG island methylation and MBD2 in immune cell gene regulation

Aimée M. Deaton

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2010

Contents

Declaration

I declare that this thesis was composed by me and that the research presented is my own work unless otherwise stated.

Aimée M. Deaton

2010

Acknowledgements

Firstly I have to thank Adrian Bird for all the support he's given me throughout my PhD. This, in combination with his sense of humour and many a pint bought for me from the Union, have helped to get me through tricky times in the lab. Thanks also to Christine Struthers for her help with administrative matters, not least of all getting paid. From the lab, I am particularly grateful to Rob Illingworth for help with all things Solexa-related and to Heather Owen for many a discussion on MBD2. Both Rob and Heather also deserve thanks for their constructive comments on this thesis. Lara Hing and Dina de Sousa provided assistance with some of the bisulfite sequencing presented here. Dina de Sousa, Jim Selfridge and Alan McClure helped, in various ways, with the mouse work. Thanks also to Jacky Guy for the use of unpublished brain expression data and Uli Gruenewald-Schneider for MBD protein production. The WTCCB bioinformaticians, Alastair Kerr and Shaun Webb, deserve special praise for their assistance in setting up resources to analyse our high throughput sequencing data. I am grateful to all Bird lab members, past and present, for a lot of fun, a plentiful supply of cake and at least some scientific discussion along the way!

My foray into immunology required the assistance of a number of experts on the subject. Martin Waterfall and Andrew Sanderson provided help with FACS sorting. Andrew MacDonald along with Peter Cook and Georgia Perona-Wright from his lab gave essential advice on immune cell purification and John Grainger was integral to the Treg experiments. This project would not have been possible without our collaboration with the Wellcome Trust Sanger Institute who carried out all the high throughput sequencing. Particular thanks must go to Katherine Auger for co-ordinating the sequencing project and Rob Andrews for mapping the data. The Illumina BeadChip experiments were carried out by Louise Evenden and Angie Fawkes at the Wellcome Trust Clinical Research Facility, Edinburgh.

There are many other people who have helped get me through my PhD. Thanks to my parents, Marian and John, for their continuous support throughout my education and for their many visits to Edinburgh over the last five years. I am especially grateful to all of my friends in Edinburgh for many good times surfing, climbing, walking and snowboarding which have helped me maintain a clear head during my PhD work. Special thanks go to Jen and Tom for putting a roof over my head and being there to listen at the end of a hard day. Finally, I thank Kevin Hatcher for a lot of support and patience during the last two years of my PhD, as well as for attempting to understand my napkin-based explanations of epigenetics! I can't wait to go on our next adventure together.

Last, but not least, I thank the Wellcome Trust for generous PhD funding which has helped support the surfing, climbing and snowboarding endeavours mentioned above.

Abstract

The phenomenon of cell type-specific DNA methylation has received much attention in recent years and a number of DNA methylation differences have been described between cells of the immune system. Of particular interest when studying DNA methylation are CpG islands (CGIs) which are distinct from the rest of the genome due to their elevated CpG content, generally unmethylated state and promoter association. In the instances when they become methylated this is associated with gene repression although it is unclear the extent to which differential methylation corresponds to differential gene expression. I have used an immune system model to assess the role of CGI methylation and the role of the methylation reader MBD2 in regulation of gene expression.

A relatively small number of DNA methylation differences were seen between immune cell types with the most developmentally related cells showing the fewest methylation differences. Interestingly, the vast majority of CGI-associated cell-specific methylation occurred at intragenic CGIs located, not at transcription start sites, but in the gene body. Increased intragenic CGI methylation tended to associate with gene repression, although the precise reason for this remains unclear. Most differentially methylated CGIs were depleted for the active chromatin mark H3K4me3 regardless of their methylation state but some of these were associated with the silencing mark H3K27me3 when unmethylated. These findings suggest that intragenic CGIs are a distinct class of genomic element particularly susceptible to cell type-specific methylation. I also looked at the effect of removing the methyl-CpG binding domain protein MBD2 from immune system cells. Immune cells from *Mbd2-/-* mice showed a number of previously uncharacterised phenotypes as well as a number of differences in gene expression compared to wild-type animals. Most of these genes increased their expression in the absence of MBD2 consistent with MBD2's role as a transcriptional repressor and *Mbd2-/-* Th1 cells showed increases in histone H3 acetylation compared to wild-type Th1 cells. This work provides an insight into the role played by cell-specific CGI methylation and MBD2 in regulating gene expression.

Figures

Tables

Data access

Sequencing and microarray data produced for this thesis are publically available at the following location http://bifx-core.bio.ed.ac.uk:8080/library. Sample information is given below and additional details are available in the file "Sample information.txt".

File name	Sample name	Technique	Cell type	Genotype
DC_MAP	DC_MAP	MAP-seq	dendritic cell	wild-type
Bcell_MAP	Bcell_MAP	MAP-seq	B cell	wild-type
CD4_1_MAP	CD4_1_MAP	MAP-seq	CD4 cell	wild-type
CD4_2_MAP	CD4_2_MAP	MAP-seq	CD4 cell	wild-type
Th1_MAP	Th1_MAP	MAP-seq	Th1 cell	wild-type
Th2_MAP	Th2_MAP	MAP-seq	Th2 cell	wild-type
Brain_MAP	Brain_MAP	MAP-seq	brain (cerebellum)	wild-type
CD4_Mbd2_MAP	CD4_Mbd2_MAP	MAP-seq	CD4 cell	Mbd2-/-
DC_H3K4me3	DC_H3K4me3	ChIP-seq	dendritic cell	wild-type
Th1_H3K4me3	Th1_H3K4me3	ChIP-seq	Th1 cell	wild-type
Brain_H3K4me3	Brain_H3K4me3	ChIP-seq	brain (whole)	wild-type
CD4_CXXC	CD4_CXXC	CAP-seq	CD4 cell	wild-type
Th1_H3Ac	Th1_H3Ac	ChIP-seq	Th1 cell	wild-type
Th1_Mbd2_H3Ac	Th1_Mbd2_H3Ac	ChIP-seq	Th1 cell	Mbd2-/-
Illumina_BeadChip_Sample_Probe_Profile	wt Th1 1-3	Illumina BeadChip array (MouseWG-6)	Th1 cell	wild-type
Illumina_BeadChip_Sample_Probe_Profile	null Th1 1-3	Illumina BeadChip array (MouseWG-6)	Th1 cell	Mbd2-/-
Illumina_BeadChip_Sample_Probe_Profile	wt Th2 1-3	Illumina BeadChip array (MouseWG-6)	Th2 cell	wild-type
Illumina_BeadChip_Sample_Probe_Profile	null Th2 1-3	Illumina BeadChip array (MouseWG-6)	Th2 cell	Mbd2-/-
Illumina_BeadChip_Sample_Probe_Profile	wt naïve 1-3	Illumina BeadChip array (MouseWG-6)	naïve CD4 cell	wild-type
Illumina_BeadChip_Sample_Probe_Profile	null naïve 1-3	Illumina BeadChip array (MouseWG-6)	naïve CD4 cell	Mbd2-/-
Illumina_BeadChip_Sample_Probe_Profile	wt DC 1-3	Illumina BeadChip array (MouseWG-6)	dendritic cell	wild-type
Illumina_BeadChip_Sample_Probe_Profile	null DC 1-3	Illumina BeadChip array (MouseWG-6)	dendritic cell	Mbd2-/-
Illumina_BeadChip_Sample_Probe_Profile	wt B 1-3	Illumina BeadChip array (MouseWG-6)	B cell	wild-type
Illumina_BeadChip_Sample_Probe_Profile	null B 1-3	Illumina BeadChip array (MouseWG-6)	B cell	Mbd2-/-
Illumina_BeadChip_Sample_Probe_Profile	A1, B1, C1	Illumina BeadChip array (MouseWG-6)	brain	wild-type
NimbleGen_oesophagus_all_norm_calls	wt 1	NimbleGen mouse expression array (mm8_60mer_expr)	oesophagus	wild-type
NimbleGen_oesophagus_all_norm_calls	null 1	NimbleGen mouse expression array (mm8_60mer_expr)	oesophagus	Mbd2-/-
NimbleGen_oesophagus_all_norm_calls	wt 2	NimbleGen mouse expression array (mm8_60mer_expr)	oesophagus	wild-type
NimbleGen_oesophagus_all_norm_calls	null 2	NimbleGen mouse expression array (mm8_60mer_expr)	oesophagus	Mbd2-/-
NimbleGen_oesophagus_all_norm_calls	wt 3	NimbleGen mouse expression array (mm8_60mer_expr)	oesophagus	wild-type
NimbleGen_oesophagus_all_norm_calls	null 3	NimbleGen mouse expression array (mm8_60mer_expr)	oesophagus	Mbd2-/-

Abbreviations

5azaC	5-azacytidine
5' RACE	5’ rapid amplification of cDNA ends
A	adenine
ac	acetyl-
ATP	adenosine triphosphate
bp	base pairs
C	cytosine
CAGE	cap analysis of gene expression
CAP	CxxC affinity purification
CAP-seq	CAP followed by high-throughput sequencing
cDNA	complementary DNA
CFP1	CxxC finger protein 1
CGI	CpG island
ChIP	chromatin immunoprecipitation
ChIP-seq	ChIP followed by high-throughput sequencing
chr	chromosome
CpG	CG dinucleotide
DC	dendritic cell
DNA	deoxyribonucleic acid
Dnmt	DNA methyltransferase
ES	embryonic stem
FACS	fluorescence activated cell sorting
FRAP	fluorescence recovery after photobleaching
G	guanine
GO	gene ontology
H2A/2B/3/4	histone H2A/2B/3/4
HDAC	histone deacetylase
hmC	5-hydroxymethylcytosine
IFN-γ	interferon-gamma
IL-4	interleukin-4
IP	immunoprecipitation
K	lysine
kb	kilobase pairs
MACS	magnetic cell sorting
MAP	MBD affinity purification
MAP-seq	MAP followed by high-throughput sequencing
MBD	methyl-CpG binding domain
MBD2	methyl-CpG binding domain protein 2

mC	5-methylcytosine
me	methyl-
MeCP2	methyl-CpG-binding protein 2
mRNA	messenger RNA
MTF	mouse tail fibroblast
ncRNA	non-coding RNA
NuRD	nucleosome remodelling and deacetylase complex
PcG	polycomb group
PCR	polymerase chain reaction
Pol II	RNA Polymerase II
PRC1	Polycomb Repressive Complex 1
PRC2	Polycomb Repressive Complex 2
qPCR	quantitative PCR
qRT-PCR	quantitative reverse transcription-PCR
RNA	ribonucleic acid
rRNA	ribosomal RNA
SAM	S-adenosyl methionine
T	thymine
Th1 cell	T helper 1 cell
Th2 cell	T helper 2 cell
Treg	regulatory T cell
tRNA	transfer RNA
TrxG	trithorax group
TSS	transcription start site

Chapter 1 Introduction

The protein-coding capacity of mammalian genomes does not, at first glance, fit with organismal complexity. The human genome has approximately 22,000 protein-coding genes whilst the genome of the simple invertebrate *Caenorhabditis elegans* contains approximately 20,000 (source http://www.ensembl.org). However, less than 5% of the 3 billion base pair human genome codes for protein (Lander et al., 2001). The remainder includes a large amount of repetitive sequence as well as regulatory sequences, such as transcription factor binding sites and enhancers, and non-coding RNA genes, which contribute to complex patterns of gene expression in the organism. As well as the genetic information carried by the DNA sequence itself, additional complexity resides in chemical modification of DNA in the form of cytosine methylation and in the way DNA is packaged into chromatin. Such information stored outside the primary DNA sequence is commonly referred to as "epigenetic" although more rigorous definitions require epigenetic phenomena to be heritable through cell division (Russo et al., 1996). Heritability, whilst holding true for DNA methylation and the polycomb-trithorax system, does not apply to all chromatin states. Epigenetic phenomena contribute to developmental and tissue-specific gene expression which allows specification of the hundreds of distinct cell types that make up the human body. Cells of the immune system, which are the focus of this work, have the property of responding rapidly to infection yet can also develop into memory cells to fight future infections. Immune cells are frequently differentiating and specialising so it is not surprising that epigenetic mechanisms have been found to be important in their regulation. This thesis investigates the role CpG island methylation and a protein which "reads" this mark, MBD2, in immune system cells.

1.1 DNA methylation

The cytosine base of DNA can be chemically modified by addition of a methyl-group to the carbon-5 position of its pyrimidine ring (Figure 1.1.1A). This bestows an additional layer of information on the DNA sequence as DNA methylation patterns can be maintained through cell division. Vertebrate genomes are extensively methylated (Ehrlich, 1982), perhaps reflective of their genomic complexity, and DNA methylation is essential for life (Li et al., 1992; Okano et al., 1999). Methylation does not affect the coding potential of DNA but when it occurs at promoters it is associated with transcriptional repression (Laurent et al., 2010; Lister et al., 2009; Stein et al., 1982; Vardimon et al., 1982). The best-characterised roles of DNA methylation in mammals are in X chromosome inactivation, where CpG island promoters gain methylation, and in imprinting, where methylation directs monoallelic gene expression.

1.1.1 Vertebrate genomes are globally methylated

Vertebrates have cytosine methylation throughout their genomes while invertebrates have less DNA methylation. *Drosophila melanogaster* and *Caenorhabditis elegans* have very low levels of cytosine methylation (Hodgkin, 1994; Lyko et al., 2000) whilst some invertebrates such as *Ciona intestinalis* have a mosaic methylation pattern consisting of large domains of methylated or unmethylated DNA (Suzuki et al., 2007; Tweedie et al., 1997). In organisms such as *C. intestinalis* methylation seems to be enriched in constitutively expressed "housekeeping genes" which is in contrast to vertebrate genes where methylation is present regardless of activity. The origin of vertebrates coincides with this transition from partial to global cytosine methylation with even the most primitive vertebrates (lamprey and hagfish) showing global DNA methylation (Tweedie et al., 1997). An attractive proposal is that global DNA methylation arose to facilitate appropriate gene regulation and dampen transcriptional noise in the more complex vertebrate genome (Bird, 1995).

1.1.2 Methylation occurs in the context of CpG dinucleotides

The vast majority of cytosine methylation occurs in the context of CpG dinucleotides. As this site is symmetrical, methylation occurs on both strands of the DNA and can be copied during DNA replication making it a heritable mark (Figure 1.1.1B). Whilst being heritable, DNA methylation patterns can change, for example during development and differentiation (Meissner et al., 2008; Mohn et al., 2008; Morgan et al., 2005) suggesting that CpG methylation can be regulated to influence expression of different genes in different cellular contexts. Mammalian cytosine methylation in contexts other than CpG has been documented, predominantly in embryonic stem (ES) cells. This does not, however, have the heritable properties of CpG methylation (Laurent et al., 2010; Lister et al., 2009; Ramsahoye et al., 2000).

In vertebrates approximately 70% of CpG sites are methylated (Ehrlich et al., 1982). This global pattern of CpG methylation has led to a partitioning of the genome into two distinct fractions with respect to CpG content and methylation status. Methyl-CpG is mutagenic as methyl-cytosine (mC) spontaneously deaminates to thymine (T) (Figure 1.1.1A) resulting in a T.G mismatch. Although there is cellular machinery to repair this (Hendrich et al., 1999b; Millar et al., 2002; Wiebauer and Jiricny, 1990), repair is inefficient, and over evolutionary time CpGs have been lost in the genome. Human bulk genomic DNA contains approximately one-fifth the number of CpGs than would be expected from its G+C content (Bird, 1980). A small fraction of the genome (~1%) is maintained in an unmethylated state and therefore shows elevated CpG and G+C content compared to the bulk (Bird et al., 1985). These functionally important regions are termed CpG islands (CGIs), often associate with gene promoters and are discussed in section 1.5.

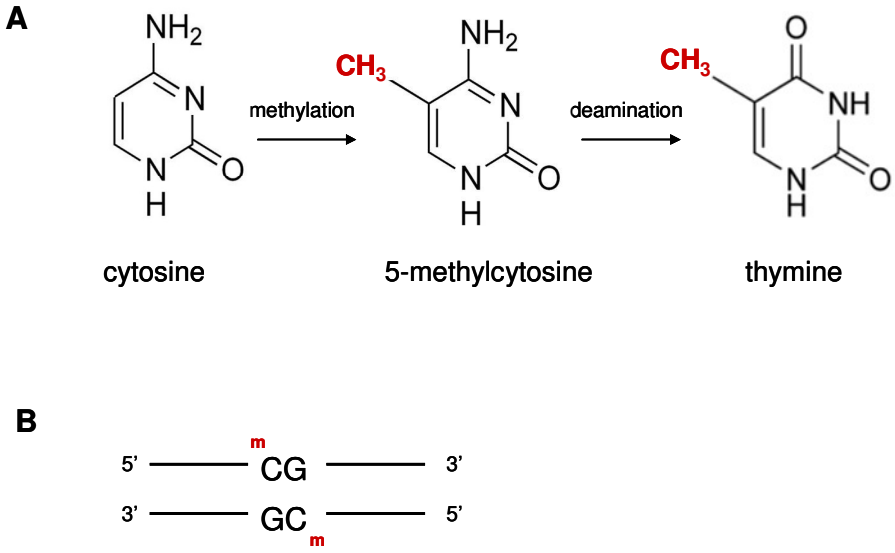


Figure 1.1.1 Cytosine methylation

A) Cytosine can be methylated at the carbon-5 position to form 5-methylcytosine which can deaminate to thymine.

B) The CpG dinucleotide is symmetrically methylated giving methylation on both DNA strands (m = methyl-group).

1.1.3 Cytosine is methylated by the DNA methyltransferase enzymes

The DNA methyltransferase enzymes (Dnmts) are responsible for establishing and maintaining the DNA methylation pattern. They act by transferring methyl-groups from S-adenosyl methionine (SAM) to cytosine. The maintenance methyltransferase Dnmt1 re-establishes the CpG methylation state after DNA replication and the *de novo* methyltransferases Dnmt3a and 3b act on previously unmethylated CpGs [Dnmts are reviewed in (Hermann et al., 2004)]. The function of another Dnmt family member, Dnmt2, appears to be in methylating tRNA rather than DNA (Goll et al., 2006).

Dnmt1 restores the symmetrical pattern of CpG methylation following DNA replication. When DNA is replicated during cell division the newly synthesised DNA is initially unmethylated while the template strand remains methylated, a state known as hemi-methylation. Dnmt1 associates with the replication machinery to methylate

the daughter DNA strand and is therefore necessary to propagate the DNA methylation state through cell division. Knock out of the *Dnmt1* gene in mouse is embryonically lethal demonstrating the essential role of DNA methylation in development (Li et al., 1992).

Dnmt3a and Dnmt3b catalyse the *de novo* methylation of CpGs and, like Dnmt 1, are essential for mouse development. Dnmt3a and 3b have partially overlapping roles in *de novo* methylation as a Dnmt3a/3b double knockout mouse has a more severe phenotype than either single deletion (Okano et al., 1999). Dnmt3a and 3b are highly expressed in early embryonic development where, after erasure of the parental methylation pattern, they act to re-establish methylation (Morgan et al., 2005). Promiscuous activity of Dnmt3a and 3b due to this high expression level is also thought to be responsible for the non-CpG methylation found in ES cells (Laurent et al., 2010; Lister et al., 2009; Ramsahoye et al., 2000). Dnmt3a and 3b do have specialised functions and differ in their mechanism of action (Gowher and Jeltsch, 2002). Dnmt3b alone methylates the pericentromeric repeats while there have been suggestions that Dnmt3a is involved in targeting methylation to discrete genomic loci (Hermann et al., 2004). Dnmt3a also seems to have a more essential role in imprinting (Kaneda et al., 2004).

Dnmt3L is a member of the Dnmt3 family that lacks a catalytic domain. It has been identified as an interacting partner for Dnmt3a and Dnmt3b and is essential for imprinting (Hata et al., 2002). Dnmt3L enhances the catalytic activity of the *de novo* methyltransferases (Gowher et al., 2005; Suetake et al., 2004) and is also required for mouse spermatogenesis (Webster et al., 2005).

1.1.4 DNA demethylation and 5-hydoxymethylcytosine

DNA methylation can be removed either by passive or active demethylation. Passive demethylation results from the failure of Dnmt1 to re-establish the methylation pattern after replication. This leads to loss of the DNA methylation pattern through

successive rounds of cell division. Active demethylation involves removal of methylation by an enzyme without the need for DNA replication. There is evidence for active demethylation of the paternal genome in the pre-implantation embryo (Mayer et al., 2000; Oswald et al., 2000). Demethylation of a region in the *interleukin-2* (*Il2*) promoter after T cell activation but prior to DNA replication has also been shown (Bruniquel and Schwartz, 2003) and a similar phenomenon has been observed at the Th2 locus control region, although here demethylation kinetics are slower (Kim et al., 2007). The identity of the DNA demethylation activity has, however, remained elusive. The methyl-binding domain protein, MBD2, was reported to act as a demethylase (Bhattacharya et al., 1999) but this has been widely disputed as paternal demethylation occurs normally in embryos deficient for MBD2 and a number of studies have failed to find evidence for demethylase activity *in vitro* (Hendrich et al., 2001; Ng et al., 1999; Santos et al., 2002). The most likely mechanism of demethylation is thought to be modification of methylated cytosine (for example by deamination to T or conversion to a novel base) followed by removal by a base excision repair mechanism and replacement with unmethylated cytosine as has been shown in zebrafish and plants (Gong et al., 2002; Penterman et al., 2007; Rai et al., 2008). A cytidine deaminase (AID) has been found to be required for DNA demethylation during reprogramming of human fibroblasts into induced pluripotent stem (iPS) cells (Bhutani et al., 2009).

Recently a potential intermediate in DNA demethylation, 5-hydroxymethylcytosine (hmC), was identified (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). mC is converted to hmC by the CxxC protein TET1 and this modified base is detectable in mouse ES cells and purkinje cells of the brain. hmC could represent a distinct modification state of DNA with its own specialized functions thus adding further complexity to epigenetic regulation in mammals. An alternative is that hmC could be an intermediate in DNA demethylation and be preferentially excised from DNA and replaced by unmodified cytosine. Support for this comes from the observation that TET1 is essential for maintaining the *Nanog* promoter in an unmethylated state in ES cells (Ito et al., 2010)

1.1.5 Role of DNA methylation

DNA methylation is repressive at promoters and CGIs

The methylation of promoters has been shown to result in gene repression (Siegfried et al., 1999; Stein et al., 1982; Vardimon et al., 1982). Genome-wide methylation studies also support an association between promoter methylation and transcriptional inactivity, but this is strongest for CGI promoters (Mohn et al., 2008). Promoter DNA methylation is traditionally thought to repress transcription in one of two ways: 1) Directly through inhibition of protein binding to methyl-cytosine (Figure 1.1.2A) or 2) Indirectly via proteins that specifically bind to methylated DNA and recruit repressive chromatin-modifying activities (Figure 1.1.2B).

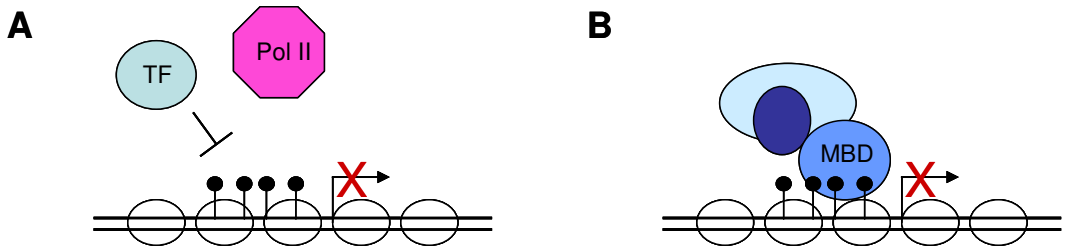


Figure 1.1.2 Mechanisms of DNA methylation-mediated transcriptional repression

A) DNA methylation (black lollipops) can directly prevent binding of transcription factors (TF) and other proteins to the promoter preventing transcription.

B) Methylation can recruit repressive MBD proteins that cause chromatin modification and gene silencing.

The model whereby methylation represses transcription by disrupting protein binding was initially proposed for transcription factors. Indeed, methylation-dependent inhibition of binding has been demonstrated for the ETS, MLTF, E2F, CREB and ATF/c-Jun activators (Campanero et al., 2000; Gaston and Fried, 1995; Iguchi-Ariga and Schaffner, 1989; Jones and Chen, 2006; Watt and Molloy, 1988) but binding of other transcription factors, such as Sp1, is not disrupted by methylation (Harrington et al., 1988). Further support for the direct inhibition model comes from the

observation that DNA methylation is depleted at active enhancers and at protein interaction sites (Lister et al., 2009). In addition, promoter DNA methylation may be refractory to binding of RNA Polymerase II (Pol II) itself. Support that this is the case for CGI-associated promoters comes from genome-wide studies profiling DNA methylation and Pol II binding showing that methylated promoters are not occupied by Pol II (Mohn et al., 2008). Therefore the binding of proteins, other than transcriptional activators, is likely to depend on the methylation status of cytosine. One example of this is the binding of the chromatin insulator CTCF which can separate promoters from the effect of distal enhancer elements. CTCF binding is inhibited by DNA methylation and this inhibition of CTCF binding can activate transcription as is the case at the imprinted *Igf2/H19* locus (discussed in section 1.5.5) (Bell and Felsenfeld, 2000; Bell et al., 1999).

Transcriptional repression can also be accomplished by proteins with specific affinity for methylated DNA. Members of the methyl-CpG binding domain (MBD) family of proteins are involved in methylation-dependent transcriptional repression (Hendrich and Bird, 1998; Nan et al., 1997) and act through recruitment of chromatin modifying activities (see sections 1.2 and 1.3). For example, methyl-CpG binding domain protein 2 (MBD2) is part of the NuRD chromatin remodelling complex which contains nucleosome remodelling and histone deacetylase activity (Ng et al., 1999). The unrelated Kaiso proteins can also bind methylated DNA and repress transcription (Filion et al., 2006; Prokhortchouk et al., 2001).

Gene body methylation occurs at active genes

Much attention in recent years has focused on the role of intragenic or gene body methylation following reports that it is associated with actively transcribed genes in *Arabidopsis* (Zhang et al., 2006; Zilberman et al., 2007). In keeping with this, invertebrates such as *C. intestinalis* show gene body methylation associating with the most highly expressed housekeeping genes (Suzuki et al., 2007) .The transcriptionally active X chromosome in humans was found to show monoallelic gene body methylation that was absent from the inactive X chromosome (Hellman

and Chess, 2007). It has been suggested that this may contribute to dosage compensation through up-regulation of X-linked genes in order to balance the gene dosage with that of the autosomes (Payer and Lee, 2008). A number of recent studies in human have shown a more general correlation between DNA methylation in gene bodies and gene expression, with highly expressed genes tending to have more intragenic methylation (Ball et al., 2009; Deng et al., 2009; Rauch et al., 2009). Comprehensive analyses of methylation using whole genome bisulfite sequencing have also investigated the relationship between gene body methylation and transcriptional activity. Lister et al. (2009) found that in stem cells there was no correlation between the level of gene body methylation and gene expression but did observe a correlation for foetal fibroblasts where the most highly expressed genes had the highest level of methylation. Laurent and colleagues also showed that gene body methylation was positively correlated with gene expression but this effect was very small (Laurent et al., 2010). One mechanism by which gene body methylation is proposed to contribute to efficient transcription is through repression of spurious transcription initiating within the gene body. Alternative hypotheses are that gene body methylation is reinforced by transcriptional elongation or is targeted by RNA produced by antisense transcription [these possibilities are discussed in (Suzuki and Bird, 2008)]. A sharp change in methylation levels at exon-intron boundaries has been observed suggesting a role for DNA methylation in splicing or a role for the splicing machinery in directing methylation (Laurent et al., 2010).

Role of DNA methylation in the bulk genome

Outwith its role at genes, what is the function of genome-wide methylation? It has been proposed that DNA methylation could be related to the high incidence of repeats in the mammalian genome. Many of these repeats derive from transposable elements and are methylated and repressed although levels of methylation are often lower than those found in gene bodies (Laurent et al., 2010). In Dnmt1-deficient mice, transcription of IAP transposable elements is massively induced and human LINE and SINE promoters are also de-repressed when DNA methylation is depleted (Liu et al., 1994; Walsh et al., 1998; Woodcock et al., 1997). One theory is that

methylation of transposable elements is involved in "genome defence" and prevents unconstrained transposition which would be damaging to the genome (Yoder et al., 1997). However, there is no evidence that depleting methylation leads to increased transposition (Wilson et al., 2007). Another hypothesis is that methylation of these repeats is involved in suppressing transcription from cryptic promoter elements as this could interfere with normal gene expression (Bird and Tweedie, 1995). Indeed, one of the most common repeat classes in the genome, the Alu elements, are usually transpositionally inert but contain numerous promoter sequences (Bird, 2002).

Evidence suggests that DNA methylation is necessary to maintain genomic integrity. Cancer cells show global hypomethylation and this is associated with genome instability and chromosomal abnormalities (Jones and Baylin, 2007). In the inherited syndrome ICF (immunodeficiency, centromeric heterochromatin and facial anomalies) Dnmt3b mutation leads to hypomethylation of pericentromeric repeats and chromosome segregation defects in lymphoblastoid cells (Ehrlich, 2003). In normal cells DNA methylation is enriched at sub-telomeric regions of chromosomes and is important for regulating telomere length and recombination (Gonzalo et al., 2006; Lister et al., 2009; Steinert et al., 2004).

1.2 DNA is packaged into chromatin

In the eukaryotic nucleus DNA is packaged into a structure called chromatin. Its core subunit is the nucleosome which consists of approximately 150bp of DNA wrapped around an octamer of the four core histone proteins (H2A, H2B, H3 and H4) (Luger et al., 1997). Nucleosomes are separated by linker DNA that interacts with histone H1. Histone H1 can facilitate compaction of nucleosomes into higher order chromatin structures, such as the 30nm chromatin fibre (Happel and Doenecke, 2009; Tremethick, 2007) with the ultimate level of organisation being the chromosome (Figure 1.2.1). The remodelling of chromatin, through the covalent modification of histones or the action of ATP-dependent complexes, allows modulation of DNA accessibility and thus influences gene expression.

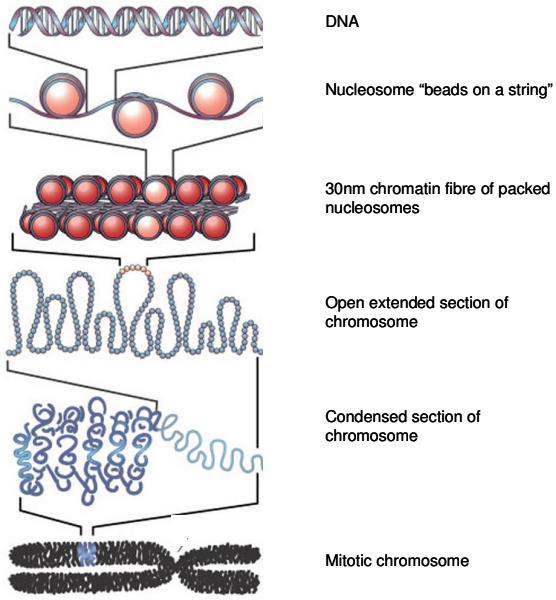


Figure 1.2.1 Packaging of DNA into chromatin

DNA is wrapped around a histone octamer to form the nucleosome and strings of nucleosomes can be packaged into a 30nm chromatin fibre, the structure of which is disputed. Chromatin fibres are further organised to form the chromosome. Within the chromosome chromatin can adopt an open transcriptionally permissive conformation or exist as a tightly packed condensed structure. Modulation of each level of chromatin structure can help regulate gene expression.

Diagram adapted from (Felsenfeld and Groudine, 2003).

1.2.1 Chromatin can be altered by covalent modification of histones

Histones can be post-translationally modified. This occurs most extensively on the histone tails which are more accessible to modifying enzymes. Histone modifications include: acetylation (ac), mono-, di- or tri-methylation (me1, me2, me3), phosphorylation, ubiquitination and sumoylation. Acetylation and methylation of lysine residues (K) in the histone tails, in particular, have been the focus of much research. Histone modifications are often classified as “activating” or “repressive” depending on whether they are associated with transcriptional activity or transcriptional silencing. The most well-defined of these is the association of histone acetylation with gene activity. Histone modifications have been profiled extensively in mammalian genomes using chromatin immunoprecipitation (ChIP) combined with

either microarrays (ChIP-chip) or high-throughput sequencing (ChIP-seq) and related to transcriptional status. Modifications associated with transcriptional activity identified by these studies include H3K4me2, H3K4me3, and H3/H4ac, which occur at gene promoters, and H3K36me3, which is found in the gene body of actively transcribed genes. The H3K9me3 and H3K27me3 marks, amongst others, are associated with transcriptional silencing (Barski et al., 2007; Bernstein et al., 2005; Mikkelsen et al., 2007; Roh et al., 2005; Wang et al., 2008). A schematic illustrating some of these data is shown in Figure 1.2.2.

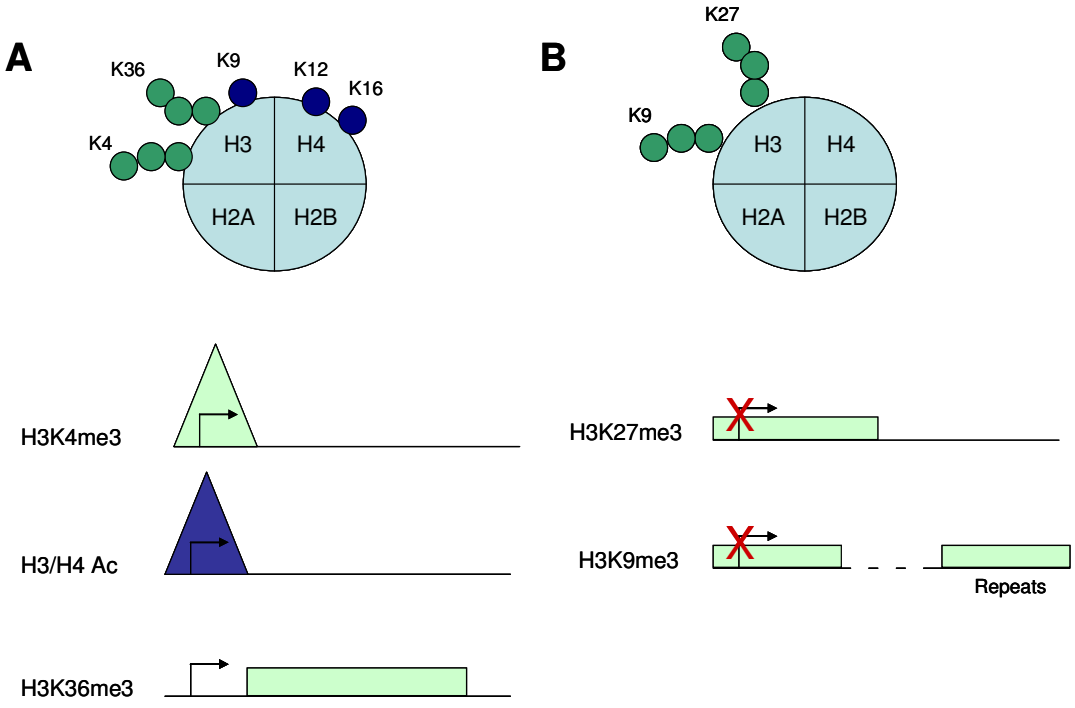


Figure 1.2.2 Histone modifications and transcriptional activity

A) Modifications of histone tails that have been associated with gene activity. **B)** Histone modifications associated with gene silencing. Top: nucleosome bearing me (green) or ac (blue) marks on lysine (K) residues. Bottom: location of histone modification relative to the gene promoter (arrow).

The ability of histone modifications to be activating or repressive with respect to transcription can be modulated by direct influence of the histone modification on nucleosome structure or by proteins that recognize modified histones themselves. In the case of histone acetylation, the acetyl-group neutralizes positively charged lysines at histone tails leading to a conformational change that renders nucleosomal

DNA more accessible to the transcriptional machinery (Hong et al., 1993; Lee et al., 1993). H3K9me3 recruits Heterochromatin Protein 1 (HP1) and H3K27me3 can recruit Polycomb Repressive Complex 1 (PRC1) both of which lead to formation of a repressive chromatin environment (Bannister et al., 2001; Cao and Zhang, 2004; Lachner et al., 2001; Simon and Kingston, 2009). H3K4me3 has been shown to interact with the NuRF chromatin remodelling complex amongst others (Li et al., 2006; Ruthenburg et al., 2007; Wysocka et al., 2006) and *in vitro* recruitment of the core transcription factor TFIID by H3K4me3 has also been reported (Vermeulen et al., 2007), both of which may help to promote transcription.

It has been proposed that a "histone code" could predict transcriptional activity on the basis of post-translational modification of histones (Jenuwein and Allis, 2001; Strahl and Allis, 2000). Several studies have identified bivalent chromatin domains marked by overlapping regions of the activating mark H3K4me3 and the repressive mark H3K27me3 (Barski et al., 2007; Bernstein et al., 2006; Roh et al., 2006). Such findings demonstrate that chromatin regulation is more complex than simple "on" or "off" marks and may involve many layers of complexity. Many genes have been shown to be "poised" for activation bearing active chromatin modifications even when the gene is silent (Barski et al., 2009; Guenther et al., 2007; Roh et al., 2006) and, consistent with this, the so-called active mark H3K4me3 is present at nearly all CGI promoters regardless of gene activity (Illingworth et al., 2010; Mikkelsen et al., 2007). This evidence weakens the hypothesis that histone modifications alone are sufficient to predict transcriptional activity. It is important to bear in mind that the appropriate transcription factors must be present to activate transcription, even if the chromatin environment is permissive for gene expression.

1.2.2 Histone modifications can be actively removed

Although an enzyme which actively demethylates DNA has remained elusive there are several activities that deacetylate and demethylate histones. Histone deacetylases (HDACs) are recruited to DNA via direct interaction with transcriptional repressors or as part of multi-subunit complexes such as the NuRD (nucleosome-remodelling

and histone deacetylase) complex. The deacetylation of histones correlates with gene repression and, in general, HDACs can be viewed as having repressive activity (HDACs are reviewed in (Haberland et al., 2009). More recently, enzymes which actively demethylate histones have been identified. These enzymes are specialised to remove specific methylation marks and, so far, demethylases that act on H3K9me1/2/3, H3K36me1/2/3, H3K4me1/2/3 and H3K27me2/3 have been discovered. The vast majority of these belong to the Jumonji-C (JmjC) domain-containing family of proteins [reviewed in (Agger et al., 2008; Klose and Zhang, 2007)]. Significantly a H3K36 demethylase has recently been shown to associate with unmethylated CGIs, highlighting the link between histone modification and the DNA methylation state (Blackledge et al., 2010).

1.2.3 ATP-dependent chromatin remodelling activities

The packaging of DNA into nucleosomes can occlude access of the transcriptional machinery and other proteins to the DNA sequence. Therefore transcriptional activity can be regulated by altering the histone-DNA contacts, potentially changing the position of nucleosomes and the chromatin architecture at specific loci. There are several families of chromatin remodelling complexes that physically disrupt nucleosome interactions through the use of ATP hydrolysis. They can be divided into four main classes of complex; SWI/SNF (Switch/Sucrose non-fermentable), CHD (chromodomain and helicase domain), ISWI (imitation SWI) and INO80. Significantly, these complexes contain domains that recognise specific histone modification states and have been shown to be important for differentiation and development. For example, the SWI/SNF family members brahma (BRM) and brahma-like 1 (BRG-1) contain a bromodomain that recognises acetylated histones, while CHD family members possess chromodomains which bind to methylated histone tails [for reviews see (de la Serna et al., 2006; Ho and Crabtree, 2010)].

1.2.4 The polycomb and trithorax proteins form an epigenetic system based on chromatin modification

One of the first epigenetic systems to be identified was the Polycomb (PcG)/Trithorax (TrxG) system which maintains expression patterns of developmental *Hox* genes in *Drosophila* even after removal of initial transcription factors. This system is conserved in vertebrates where it also has a role in regulating *Hox* gene expression as well as roles in stem cell identity (Pietersen and van Lohuizen, 2008), lineage specification (Schwartz and Pirrotta, 2007), imprinting (Wu and Bernstein, 2008) and X chromosome inactivation (Payer and Lee, 2008). PcG proteins promote transcriptional repression through two multimeric complexes – Polycomb Repressive Complex 2 (PRC2) which di- and tri-methylates H3K27 and Polycomb Repressive Complex 1 (PRC1) which can be recruited by H3K27me3, ubiquitylates histone H2A and is thought to mediate gene repression by blocking transcriptional elongation. TrxG proteins maintain transcriptional activity and can have H3K4 trimethylase activity (MLL1-3) as well as the ability to remodel chromatin (SWI/SNF and NuRF complexes) [reviewed in (Schuettengruber et al., 2007; Simon and Kingston, 2009)]. The TrxG protein UTX (Ultratrithorax) has recently been shown to be an H3K27me2/3 demethylase, demonstrating how TrxG proteins can directly antagonize polycomb silencing (Agger et al., 2007). There have been reports that PcG proteins can be recruited by ncRNAs (Rinn et al., 2007; Zhao et al., 2008) further demonstrating the complexity of mammalian gene regulation.

There have been several studies focusing on the interaction between the polycomb and DNA methylation systems. It has been claimed that in cancer cells polycomb-mediated H3K27me3 marks genes for aberrant methylation (Schlesinger et al., 2007). Physical interaction between polycomb components and the Dnmts have been reported (Reynolds et al., 2006; Vire et al., 2006) although this has not been substantiated by *in vivo* findings. In normal gene expression, DNA methylation and polycomb (as assessed by presence of H3K27me3) are mutually exclusive silencing mechanisms (Fouse et al., 2008; Mohn et al., 2008) and studies have also shown this to be the case in cancer (Coolen et al., 2010; Kondo et al., 2008). However, the two

systems do show interdependence. Many polycomb targets in ES cells acquire DNA methylation during neuronal differentiation (Mohn et al., 2008) and there are several studies showing that CGI promoters becoming aberrantly methylated in cancer tend to be marked by H3K27me3 in ES cells (Keshet et al., 2006; Ohm et al., 2007; Widschwendter et al., 2007) (Illingworth et al., 2010)). In X inactivation H3K27me3 precedes methylation of CGIs suggesting that H3K27me3 may be able to "instruct" DNA methylation (Payer and Lee, 2008).

1.2.5 Functional interplay between DNA methylation, histone modification and ATP-dependent chromatin remodelling

DNA methylation state, histone modification and the action of ATP-dependent chromatin remodellers all co-operate to regulate gene expression. As mentioned previously, particular histone modifications can be recognised by different classes of chromatin remodelling complexes. DNA methylation also has a function in regulating the chromatin state. DNA methylation state correlates with specific histone modification states, for example, unmethylated CGIs are marked by H3K4me3 (Illingworth et al., 2010; Mikkelsen et al., 2007) whilst in ES cells distinct sets of promoters are marked by H3K27me3 and DNA methylation (Fouse et al., 2008). Furthermore, DNA methylation or its absence can directly recruit chromatin and histone modifying activities. There is new evidence suggesting that unmethylated CGIs recruit the H3K4me3 methyltransferase complex SETD1 via CxxC finger protein 1 (CFP1) (Thomson et al., 2010) and as well as the H3K36 demethylase KDM2A (Blackledge et al., 2010). A protein that binds specifically to methylated DNA, MBD2, is part of the NuRD complex which possesses histone deacetylase and chromatin remodelling activity (Feng and Zhang, 2001; Ng et al., 1999) as well as recently reported histone demethylase activity (Wang et al., 2009).

The enzymes responsible for DNA methylation, the Dnmts, are also associated with chromatin modifying activities. There have been a number of reports describing Dnmt interaction with HDACs (Fuks et al., 2000; Robertson et al., 2000; Rountree et al., 2000) and histone methyltransferases (Epsztejn-Litman et al., 2008; Esteve et al.,

2006; Fuks et al., 2003; Lehnertz et al., 2003; Vire et al., 2006). Interaction between Dnmts and the chromatin remodeller hSNF2H has also been observed (Geiman et al., 2004; Robertson et al., 2004) as has an interaction with a related protein, LSH (Myant and Stancheva, 2008). Knock out of LSH in mice leads to genomic hypomethylation and premature death (Dennis et al., 2001) and one possibility is that chromatin remodelling is necessary in order for the Dnmts to gain access to DNA.

1.3 The MBD proteins

1.3.1 The MBD proteins specifically bind to methylated DNA

One way in which the DNA methylation pattern can be interpreted is through the binding of factors with specific affinity for methylated DNA which can, in turn, affect the chromatin state. The first specific methyl-CpG binding protein, MeCP (later renamed MeCP1) was isolated based on its ability to bind probes containing multiple methylated CpGs in bandshift assays. MeCP1 complexes were detected in a range of tissues but were not formed when probes were unmethylated (Meehan et al., 1989). The MeCP1 complex was shown to exert a repressive effect on transcription (Boyes and Bird, 1991, 1992) and was later identified as being the MBD2/NuRD complex (Feng and Zhang, 2001; Ng et al., 1999). A second methyl-binding activity, MeCP2, was subsequently identified and was also shown to repress transcription (Lewis et al., 1992; Meehan et al., 1992). Deletion analysis was used to characterise the domain from MeCP2 required for binding to methylated DNA and this was termed the methyl-CpG binding domain (MBD) (Nan et al., 1993). A search for proteins containing an MBD homologous to that of MeCP2 led to the identification of the MBD family of proteins (Cross et al., 1997; Hendrich and Bird, 1998). The MBD proteins consist of MeCP2, MBD1, MBD2, MBD3 and MBD4 (Figure 1.3.1). MeCP2, MBD1 and MBD2 specifically bind methylated DNA and repress transcription. The effect of knocking out these proteins in mouse is a lot less severe than the effect of disrupting DNA methylation arguing that DNA methylation exerts its effects independently of MBD proteins in many cases. MBD3 is a transcriptional

repressor but has no methyl-specific binding while MBD4 is involved in DNA repair. In addition, an unrelated family of Kaiso-like proteins also have binding specificity for methylated DNA (Filion et al., 2006; Prokhortchouk et al., 2001).

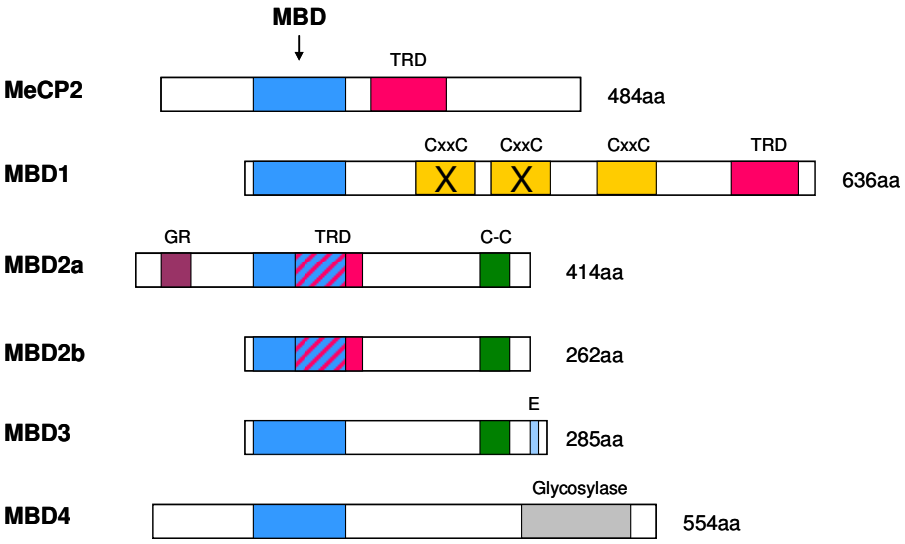


Figure 1.3.1 The MBD proteins

The MBD proteins were identified on the basis of their MBD domain (blue). MeCP2, MBD1 and MBD2 contain a transcriptional repression domain (TRD; pink) that interacts with chromatin modifying activities. MBD1 contains three CxxC domains, the third of which is functional. MBD2 exists in a long isoform, MBD2a which contains a run of glycine arginine residues (GR) and a short isoform, MBD2b. MBD3 has 75% homology to MBD2b and contains a coiled-coil domain (C-C) similar to MBD2 but has a unique glutamatic acid repeat (E) at the C-terminus. MBD4 has a glycosylase domain at its C-terminus which is involved in DNA repair. All amino acid lengths are given for the mouse MBD proteins.

1.3.2 MeCP2

MeCP2 represses transcription in a methylation-dependent manner and this repression is mediated by its transcriptional repression domain (TRD) in reporter assays (Nan et al., 1997). MeCP2 has been reported to exert its repressive effect by interacting with the Sin3a HDAC co-repressor complex (Nan et al., 1998) although this association must be transient as MeCP2 usually exists as a monomer (Klose and Bird, 2004). *In vitro* studies have shown that, in addition to a binding requirement for methyl-CpG, the MeCP2 MBD prefers to bind DNA sequences containing a run of A/T residues (Klose et al., 2005).

MeCP2 is particularly abundant in the brain, specifically neurons, and mutations in MeCP2 are responsible for the neurological disorder Rett syndrome (Amir et al., 1999). Rett syndrome is an autism spectrum disorder characterised by a period of normal development followed by regression in terms of speech, motor function and acquired skills around 6-18 months of age. Knock out of the *MeCP2* gene in mice recapitulates many features of the Rett syndrome phenotype and results in the premature death of null animals (Chen et al., 2001; Guy et al., 2001). Significantly, this phenotype can be reversed by reintroduction of a functional copy of *MeCP2* even after the onset of symptoms (Guy et al., 2007). The majority of Rett mutations occur in the MBD of MeCP2 highlighting the importance of its methyl-CpG binding function. The structure of the MeCP2 MBD bound to methylated DNA has recently been solved showing how many of these mutations interfere with binding (Ho et al., 2008).

Despite the severity of the Rett phenotype, changes in gene expression in the *MeCP2* null mouse brain are rather subtle (Nuber et al., 2005; Tudor et al., 2002). This, along with the abundance of MeCP2 in neurons, has led to the proposal that MeCP2 has a global role in regulating neuronal chromatin rather than one as a classical transcriptional repressor controlling expression of a defined set of genes. A recent study showed that in neurons, there is one molecule of MeCP2 for every two nucleosomes and that MeCP2 binds throughout the genome tracking methyl-CpG density. A global increase in the levels of histone acetylation and histone H1 in neurons lacking MeCP2 was observed, consistent with a role for MeCP2 in regulating the chromatin state (Skene et al., 2010). There have also been suggestions that MeCP2 can act as a transcriptional activator (Chahrour et al., 2008) and that it functions in alternative splicing (Young et al., 2005).

1.3.3 MBD1

MBD1 was the first MBD to be identified on the basis of homology to MeCP2 and was initially believed to be the methyl-binding activity of MeCP1 (Cross et al., 1997). This 70kDa protein has an MBD and a TRD as well as two or three CxxC

domains depending on alternative splicing. One of these CxxC domains, CxxC3, is functional and facilitates binding of MBD1 to unmethylated CpGs. In DNA methylation-deficient cells localisation of MBD2 and MeCP2 to CpG-rich major satellite sequence is lost due to lack of methylation. However, MBD1 remains correctly localised in these cells and this localisation is dependent on its CxxC3 domain (Hendrich and Bird, 1998; Jorgensen et al., 2004). A recent study has shown that the MBD of MBD1 prefers to bind methylated CpGs in a TGCGCA or TCGCA context while the CxxC3 domain requires a single unmethylated CpG for binding. Several MBD1 target genes were identified and binding to these targets was shown to be dependent on the MBD but not on the CxxC3 domain suggesting that, *in vivo*, MBD1 acts primarily as a methyl-binding protein (Clouaire et al., 2010).

MBD1 can act as a transcriptional repressor on methylated and unmethylated templates (Fujita et al., 1999; Jorgensen et al., 2004) and this repression seems to involve histone deacetylation (Ng et al., 2000) but, contrast to MeCP2 and MBD2, MBD1 does not directly recruit HDACs (Fujita et al., 2003). MBD1 can mediate transcriptional repression via the H3K9 methyltransferase Suvar39h1 and HP1α and this seems to be partially dependent on interactions between Suvar39h1 and HDACs (Fujita et al., 2003). MBD1 has also been reported to form a complex with another H3K9 methyltransferase SETDB1 and interact with chromatin assembly factor (CAF-1) and the DNA replication machinery during S phase. This could represent a mechanism by which histone and DNA methylation could be propagated together through cell division (Sarraf and Stancheva, 2004).

MBD1 is expressed in many tissues and has high expression in brain, but is not detectable in ES cells (Hendrich and Bird, 1998). Disruption of the *Mbd1* gene in mouse results in a mild phenotype. Animals develop normally but show decreased neurogenesis, impaired learning and autism-like behavioural characteristics (Allan et al., 2008; Zhao et al., 2003). Similar to MeCP2, disruption of MBD1 does not result in many changes in gene expression detectable by microarray analysis. Expression

microarrays on *Mbd1-/-* hippocampus showed the only transcripts with increased expression were from IAP endogenous retroviruses (Zhao et al., 2003).

1.3.4 MBD3

MBD3 shows strong homology to MBD2b and the two are thought to have evolved from duplication of an ancestral *Mbd2/3* gene (Hendrich et al., 1999a). However, due to an amino acid substitution in the MBD, MBD3 does not have the ability to bind methylated DNA (Fraga et al., 2003; Hendrich and Bird, 1998; Saito and Ishikawa, 2002). Like MBD2, MBD3 is a component of the NuRD chromatin remodelling complex (Zhang et al., 1999) although MBD2 and MBD3 form separate NuRD complexes (Le Guezennec et al., 2006). Consistent with this, MBD2 and MBD3 have non-redundant functions. *Mbd2-/-* mice are viable and fertile whereas *Mbd3-/-* embryos die before implantation indicating that MBD3 is essential for normal development, (Hendrich et al., 2001). This lethality is thought to be due to a failure to repress pluripotency markers and undergo proper differentiation. *Mbd3-/-* ES cells fail to form a functional NuRD complex and cannot differentiate or down-regulate Oct4 (Kaji et al., 2006). Similarly, *in vivo*, lack of MBD3 results in a failure of the inner cell mass (ICM) of the blastocyst to develop further after implantation (Kaji et al., 2007).

1.3.5 MBD4

Although MBD4 has specific binding affinity for methylated DNA it does not act as a transcriptional repressor but rather functions in DNA mismatch repair (Hendrich and Bird, 1998; Hendrich et al., 1999b). As discussed, mC is mutagenic because it can be deaminated to a thymine residue (T) resulting in a methyl-CpG (mCpG) dinucleotide which is paired with a TpG dinucleotide. MBD4 can specifically recognise these sites and excise the mismatched nucleotide using its glycosylase domain which has homology to bacterial repair proteins (Hendrich et al., 1999b). In *Mbd4-/-* mice, CpG to TpG transitions are enhanced 3-fold supporting a function for MBD4 in limiting the mutability of mCpG. In addition, tumour prone Apc^{Min} mice

show increased tumour formation when MBD4 is deleted due to C to T mutations in the *Apc* gene (Millar et al., 2002).

1.3.6 MBD2

MBD2 has binding specificity for methylated DNA and stably associates with the NuRD co-repressor complex (Feng and Zhang, 2001; Le Guezennec et al., 2006; Ng et al., 1999). Despite a relatively mild phenotype in the absence of MBD2 (Hendrich et al., 2001), significant gene deregulation has been observed in all systems examined in the *Mbd2-/-* mouse (Barr et al., 2007; Berger et al., 2007; Hutchins et al., 2002). This points to a role for MBD2 as a methylation-dependent transcriptional repressor. One example of gene deregulation in *Mbd2-/-* mice occurs in the immune system and it is tempting to speculate that many more phenotypes may be observed upon careful probing of immune system function. A main focus of this thesis is to explore the role of MBD2 in gene expression using the immune system as a model.

MBD2 is widely expressed and exists in two isoforms

MBD2 is expressed in all tissues with the notable exception of ES cells where MBD3 exerts an important role instead. MBD2 exists in 2 isoforms; a full-length isoform, MBD2a, and a shorter isoform, MBD2b, thought to result from use of an alternative translation start site (Hendrich and Bird, 1998). The functional significance of these two isoforms is still unclear. The most significant difference is that MBD2a contains a run of glycine-arginine (GR) repeats that can be subject to arginine methylation. This has been reported to impair MBD2's function as a transcriptional repressor (Le Guezennec et al., 2006; Tan and Nakielny, 2006). MBD2 does not saturate the genome as MeCP2 does. Estimates suggest that there are approximately 4×10^5 molecules of MBD2 per cell which is much less than the number of potential binding sites (there are approximately 2×10^7 mCpGs in the haploid genome). This suggests that MBD2 must have specific targets in the genome. (R Ekiert, unpublished results).

MBD2 is part of the NuRD complex

MBD2 is part of the NuRD chromatin remodelling and HDAC complex and this was found to represent the MeCP1 methyl-binding activity previously described (Feng and Zhang, 2001; Ng et al., 1999; Zhang et al., 1999). The NuRD complex contains the SWI/SNF chromatin remodelling protein Mi-2α/β, HDAC 1 and HDAC2, along with other factors including the MTA (metastasis-associated) proteins (different members of which may be involved in functional specialisation of NuRD complexes), p66α/β and the histone binding proteins RbAp48 and RbAp46 (Bowen et al., 2004; Brackertz et al., 2006; Xue et al., 1998; Zhang et al., 1998). MBD2 always associates with the NuRD complex (Feng and Zhang, 2001; Le Guezennec et al., 2006) but the NuRD complex can exist without MBD2 as the related protein, MBD3, can also interact with NuRD components. MBD2/NuRD preferentially remodels and deacetylates methylated nucleosomes *in vitro*. Consistent with this, MBD2/NuRD has been shown to repress transcription in reporter assays in a methylation-dependent manner which requires nucleosome remodelling and, to some extent, deacetylation (Feng and Zhang, 2001; Ng et al., 1999).

In addition to members of the NuRD complex a number of interactors have been reported for MBD2 which include the arginine methylase PRMT5, the zinc-finger protein MIZF, a novel protein MBD2in and focal adhesion kinase (FAK) (Lembo et al., 2003; Luo et al., 2009; Sekimata et al., 2001; Tan and Nakielny, 2006). A report claiming that MBD2b is a DNA demethylase (Bhattacharya et al., 1999) has not been substantiated by other studies (Hendrich et al., 2001; Ng et al., 1999; Santos et al., 2002).

Mbd2-/- mice are viable and fertile

Knockout of the *Mbd2* gene in mice has a relatively mild phenotype and mice are viable and fertile. *Mbd2-/-* mice exhibit a maternal behaviour defect but imprinting and repression of retroviral elements tested is normal. *Mbd2-/-* cells from these mice fail to fully repress methylated reporter constructs and this can be rescued by reintroduction of MBD2a or MBD2b but not MBD3 (Hendrich et al., 2001).

Strikingly, absence of MBD2 results in resistance to tumours in a mouse model of intestinal cancer. When *Mbd2-/-* mice are crossed onto a tumour-prone Apc^{Min} background they show reduced tumour burden and increased life-span (Sansom et al., 2003). This suggests that MBD2 is necessary for development or maintenance of tumours in these mice. This could be due to a role for MBD2 regulating a repressor of the Wnt signalling pathway which is constitutively active in these tumours (Phesse et al., 2008) or due to increased immune surveillance of cancer cells in the absence of MBD2.

Removal of MBD2 results in gene expression changes

Although *Mbd2-/-* mice are viable and fertile the three systems that our lab has studied in detail in this mouse show significant gene de-regulation; interleukin-4 (IL-4) and interferon-γ (IFN-γ) are aberrantly expressed during T helper cell differentiation leading to a compromised immune response, genes coding for pancreatic enzymes are mis-expressed in the colon and a fraction of *Xist* is de-repressed on the active X chromosome (Barr et al., 2007; Berger et al., 2007; Hutchins et al., 2005; Hutchins et al., 2002). In addition, MBD2 is required for silencing of particular tumour-suppressor genes in human colorectal carcinoma cells and other cancer cell lines (Auriol et al., 2005; Bakker et al., 2002; Lin and Nelson, 2003; Martin et al., 2008; Pulukuri and Rao, 2006). Other reported targets for MBD2 include the gene for the pluripotency factor Oct4, the telomerase reverse transcriptase gene *hTERT* and the estrogen responsive gene *pS2/TFF1* (Chatagnon et al., 2010; Chatagnon et al., 2009; Gu et al., 2006). This suggests that it is worth comprehensively examining systems in the *Mbd2* null mouse to look for gene expression changes.

It is worth noting that many of these MBD2 targets were confirmed to be methylated and, for a number of these, disruption of DNA methylation has been shown to relieve gene repression. In several cases, removal of MBD2 increased histone acetylation of

target genes or rendered them more sensitive to inhibition of HDACs (Barr et al., 2007; Berger et al., 2007; Hutchins et al., 2002).

MBD2 is important for immune system function

MBD2 has been implicated in a number of aspects of immune system function. Reported examples include *in vitro* T helper cell differentiation, response to infection *in vivo* and memory T cell development (Hutchins et al., 2005; Hutchins et al., 2002; Kersh, 2006).

T helper (CD4) cells of the immune system, upon stimulation with certain antigens, can differentiate into Th1 or Th2 effector cells and this can be recapitulated in a cell culture system. These cells are characterised by different repertoires of cytokine expression; Th1 cells express cytokine IFN-γ while Th2 cells express IL-4 and this is accompanied by DNA methylation changes at the relevant gene loci. When *Mbd2-/-* T helper cells are differentiated *in vitro*, IFN-γ and IL-4 expression, and consequently cell identity, is perturbed. Th1 cells from *Mbd2-/-* mice aberrantly express IL-4 while Th2 cells mis-express IFN-γ. MBD2 was found to bind two regions of the *Il4* locus in Th1 cells where presumably it contributes to gene silencing (Hutchins et al., 2002). Th1 cells mediate the immune response to intracellular bacteria and viruses while Th2 cells coordinate the response to extracellular parasites. When infected with a Th2-inducing intestinal worm, *Mbd2-/-* mice show an impaired Th2 response compared to wild-type animals but show enhanced resistance to Th1-inducing *Leishmania*. These phenotypes are reportedly due to excess IFN-γ production (Hutchins et al., 2005).

Memory T cells are produced during an infection and help protect the organism if the same pathogen is encountered again. During viral infection, cytotoxic CD8 T cells are produced in order to fight infection and some of these differentiate into memory CD8 cells. Infection of *Mbd2-/-* mice with lymphocytic choriomeningitis virus showed a reduction in the numbers of CD8 memory T cells produced compared to

wild-type animals. In addition, the CD8 memory cells that were formed were found to be defective in protecting against re-infection (Kersh, 2006) further highlighting a role for MBD2 in the immune system.

1.4 The CxxC proteins

Much as the MBDs link methylated CpG to the chromatin state, the CxxC domain proteins have been implicated in linking non-methylated CpG to chromatin modification.

1.4.1 The CxxC proteins bind unmethylated CpGs and are involved in epigenetic regulation

The CxxC domain proteins are characterised by the presence of a cysteine-rich zinc finger domain (CxxC amino acid motif) that has specific affinity for non-methylated CpGs (Voo et al., 2000). CxxC proteins have been implicated in linking non-methylated CpGs to specific chromatin states and also in catalysing and recognising DNA methylation. The CxxC domain family of proteins consists of CFP1, MLL 1 and 2, KDM2A, Dnmt1, TET1 and MBD1 as well as several proteins with uncharacterised function. Several CxxC proteins are involved in histone methylation; CFP1 is part of the SETD1 H3K4 methyltransferase complex (Lee and Skalnik, 2005), the trithorax group proteins MLL 1 and 2 are also H3K4 methyltransferases (Birke et al., 2002) whereas KDM2A has H3K36 demethylase activity (Tsukada et al., 2006). Some enzymes involved in the chemical modification of DNA also possess CxxC domains, for example Dnmt1 the maintenance DNA methyltransferase, and TET1 which catalyses conversion of mC to hmC, a potential DNA demethylation intermediate (Pradhan et al., 2008; Tahiliani et al., 2009) (Dnmt1 and TET1 are discussed in section 1.1). The methyl-binding protein MBD1 which was identified on the basis of its MBD domain (Cross et al., 1997) also has a CxxC domain with specific affinity for unmethylated DNA (Jorgensen et al., 2004) (MBD1 is discussed in section 1.3.2). The CxxC protein family is shown in Fig 1.4.1.

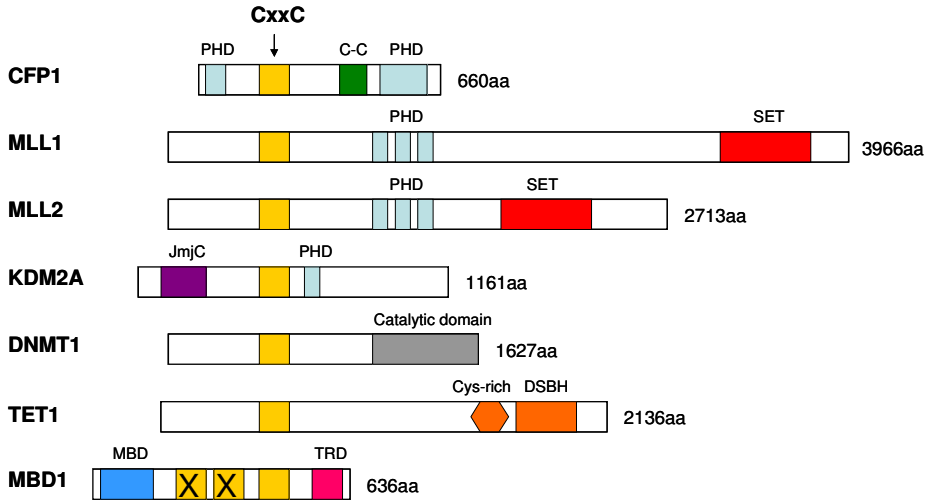


Figure 1.4.1 The CxxC domain proteins

Members of this protein family contain a CxxC domain (yellow) capable of binding unmethylated CpG. MLL1 and 2 contain a SET domain (red) which catalyses H3K4 methylation while CFP1 associates with a complex that has H3K4 methylation activity. CFP1, MLL1 and 2 and KDM2A contain chromatin binding PHD fingers. KDM2A contains the JmjC domain characteristic of this family of histone demethylases and Dnmt1 contains a C-terminal catalytic domain. TET1 possesses cysteine-rich (Cys-rich) and double stranded beta-helix (DSBH) domains that are responsible for its catalytic activity. MBD1 has an MBD domain and two non-functional CxxC domains as well as the characteristic CxxC3 domain. Sizes are given for the mouse proteins.

1.4.2 CFP1 and KDM2A have a role at CGIs

Recently the CxxC proteins CFP1 and KDM2A have been shown to direct chromatin structure at unmethylated CGIs. CFP1 (CxxC finger protein 1; also known as CGBP or CxxC1) contains two plant homeodomains (PHD) along with its CxxC domain and a set interaction domain (Voo et al., 2000). CFP1 interacts with the SETD1 methyltransferase complex (Lee and Skalnik, 2005) and is an orthologue of the protein SPP1 a component of the only H3K4 histone methyltransferase complex in yeast, SET1 (Miller et al., 2001). Deletion of *Cfp1* results in early embryonic lethality consistent with an essential role for the protein (Carlone and Skalnik, 2001) while *Cfp1* null ES cells show increased apoptosis and failure to differentiate (Carlone et al., 2005). Based on its ability to bind to unmethylated CpGs, Thomson and colleagues investigated whether CFP1 was found at CGIs and discovered that it could direct H3K4me3 to unmethylated CGIs (Thomson et al., 2010).

KDM2A (Lysine specific demethylase 2a) was characterised a number of years ago based on its ability to demethylate H3K36me2 (Tsukada et al., 2006). Recently a study by Blackledge and colleagues (2010) demonstrated specific affinity of KDM2A's CxxC domain for unmethylated CpG and showed that KDM2A specifically binds the vast majority of CGIs. Knockdown of KDM2A led to an increase in H3K36me2 specifically over CGIs (Blackledge et al., 2010). The CxxC domain has also been utilised to biochemically purify CGIs, a technique known as "CxxC affinity purification" (CAP) which uses the CxxC3 domain of MBD1 (Illingworth et al., 2008). The relationship between CxxC proteins and CGIs will be discussed further in section 1.5.

1.5 CpG islands

The genome is punctuated by CpG-rich unmethylated CGIs that frequently associate with gene promoters. CGIs are permissive to transcription except in the rare cases when they become silenced by DNA methylation. Methylation of CGIs is well-documented in X inactivation, imprinting and cancer but also occurs in normal development.

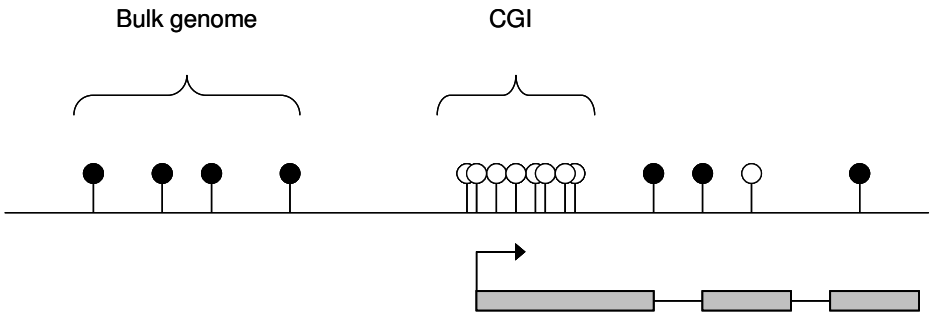


Figure 1.5.1 CGIs represent a distinct fraction of the genome

The genome can be divided into two fractions with respect to DNA methylation and CpG content. The bulk genome is CpG deficient and methylated (~70% of CpGs) while CGIs are generally unmethylated with little CpG suppression. CGIs often associate with gene promoters. Filled circles = methylated CpG, empty circles = unmethylated CpG. The arrow indicates a TSS and the direction of transcription.

1.5.1 Characterisation of CGIs

CGIs represent a distinct and functionally important fraction of the genome. They show elevated G+C content and little CpG depletion due to the fact that, unlike the bulk genome, CGIs are usually unmethylated. Methylation is mutagenic and leads to CpG deficiency, therefore maintaining an unmethylated state prevents loss of CpG (section 1.1.2). CGIs were first identified by digestion of genomic DNA with the methyl-sensitive restriction enzyme HpaII (recognition site CCGG) which left bulk genomic DNA intact whilst liberating small fragments of CpG-rich unmethylated DNA initially known as HpaII tiny fragments (Bird et al., 1985). Based on initial characterisation of these fragments CGIs were defined as being regions longer than 200bp, with a G+C content of 50% and a CpG observed/expected ratio (CpGo/e)[i] of at least 0.6 (Gardiner-Garden and Frommer, 1987; Larsen et al., 1992). These criteria are often used to bioinformatically predict the location of CGIs but can be sensitive to the presence of CpG-rich repeats such as Alu elements. More stringent criteria for prediction of CGIs were employed by Takai and Jones which, whilst eliminating repetitive Alu elements, are also likely to exclude some genuine CGIs (Takai and Jones, 2002). Initial analysis of the human genome sequence identified over 50,000 CGIs of which, just over half were unique. Building repeat-masking into CGI prediction algorithms can help eliminate contaminating repeats from CGI sets. However, all prediction algorithms fail to take into account methylation status when predicting CGIs and, by necessity they employ arbitrary thresholds, alteration of which dramatically changes the number of CGIs predicted (Illingworth et al., 2008; Illingworth and Bird, 2009).

Biological identification of CGIs is preferable to prediction algorithms especially as it takes methylation status into account. This has been accomplished using a CxxC domain that specifically binds unmethylated CpG-rich sequences to isolate CGIs from genomic DNA. The technique has been termed “CxxC affinity purification” (CAP) after the protein domain used. An initial study purifying CGIs from human

blood combined with traditional Sanger sequencing identified approximately 17,000 CGIs (Illingworth et al., 2008). This number does not represent the full complement of CGIs in the genome because of a high rate of sequence failure due to the GC-rich nature of the sequence and because any CGIs that are fully methylated in blood were excluded as they are not bound by the CxxC domain. Germline CGIs are hypomethylated (Weber et al., 2007) making tissues such as sperm more suitable for comprehensive identification of CGIs.

Recently the CGI complement of human and mouse has been comprehensively characterised using CAP combined with high throughput sequencing technology (CAP-seq) (Illingworth et al., 2010). A panel of tissues including sperm was used so practically all CGIs should be included. CGI numbers in human and mouse were very similar, at 25,495 and 23,021 respectively, despite the fact that algorithms predict a reduction in mouse CGI numbers (Waterston et al., 2002), further emphasising the importance of using biological purification to identify CGIs.

1.5.2 CGIs are associated with promoter function

CGIs are traditionally associated with the promoters of housekeeping genes and because of their promoter association localise to gene-rich regions of the genome (Lander et al., 2001). Approximately 60% of human genes have a CGI promoter including all housekeeping genes as well as approximately 40% of genes with tissue-specific expression patterns (Larsen et al., 1992). Approximately half of CGIs associate with a known TSS with the remainder occurring within gene bodies (intragenic) or between genes (intergenic) in approximately equal proportions (Illingworth et al., 2010; Illingworth et al., 2008). Many of these "orphan" CGIs show evidence for promoter function and it is hypothesised that all CGIs act as promoters at some stage in development (Illingworth et al., 2010; Macleod et al., 1998; Maunakea et al., 2010). Indeed, the presence of orphan CGIs has often led to

[i] CpG o/e is the actual number of CpG sites present in a sequence compared to the number expected from its G+C content.

the identification of novel transcripts and alternative gene promoters (Gardiner-Garden and Frommer, 1994; Macleod et al., 1998; Maunakea et al., 2010; Rauch et al., 2009). Most studies examining chromatin modification and DNA methylation of CGIs focus only on promoter CGIs so a lot remains to be uncovered about the role and properties of CGIs occurring outwith annotated TSSs.

1.5.3 CGIs form transcriptionally permissive chromatin environments

Consistent with their promoter function CGIs form a chromatin environment permissive for transcription (Tazi and Bird, 1990). Genome-wide studies have shown promoter CGIs to be frequently associated with the active chromatin mark H3K4me3 regardless of transcriptional activity (Mikkelsen et al., 2007). This also holds true for intra- and intergenic islands although the association is weaker (Illingworth et al., 2010; Maunakea et al., 2010). The H3K4me3 mark is directed to CGIs via the CxxC domain protein CFP1 (CxxC finger protein 1) (Thomson et al., 2010), a component of the H3K4 methyltransferase complex SETD1 (Lee and Skalnik, 2005). CFP1 marks the vast majority of CGIs and knockdown of CFP1 leads to depletion of H3K4me3 from CGIs. Insertion of an artificial CGI into the genome recruits CFP1 and creates a novel peak of H3K4me3 in the absence of Pol II (Thomson et al., 2010). It is likely, however, that at many CGIs the presence of Pol II contributes to H3K4 tri-methylation as the core transcription complex itself can recruit H3K4 methyltransferases (Ruthenburg et al., 2007).

Another CxxC protein, KDM2A an H3K36me2 demethylase (Tsukada et al., 2006), also influences chromatin structure at CGIs. Recently a study by Blackledge et al (2010) showed that KDM2A specifically binds over 90% of CGIs which are depleted for H3K36me2 compared to non-CGI promoters and the bulk genome. The requirement for H3K36me2 depletion at CGIs is, as yet, unclear but H3K36me2 has been reported to inhibit transcription initiation through HDAC recruitment in both yeast and human so this could be related to the promoter function of CGIs (Li et al., 2009; Strahl et al., 2002; Youdell et al., 2008).

In addition to being marked by H3K4me3 and depleted for H3K36me2, CGIs have a low density of nucleosomes and transcription from CGI promoters can initiate without a requirement for chromatin remodelling complexes (Ramirez-Carrozzi et al., 2009). This evidence suggests that the function of CGIs may be to provide environments permissive for transcription initiation within the complex mammalian genome. Most of the genome is not transcribed yet contains abundant transcription factor binding sites. Therefore chromatin structure and DNA methylation state must be important in directing transcription factors to their appropriate targets. To this end CGIs can be thought of as beacons directing transcription in the genome (Bird, 1995; Illingworth and Bird, 2009).

1.5.4 Origin of CpG islands

CGIs are distinct from the rest of the genome. One major question is how, in the face of genome-wide methylation, CGIs are protected from methylation and thus prevented from becoming CpG deficient. This may be directly related to their function as promoters. Deletion or mutation of SP1 transcription factor binding sites in the promoter of the mouse *Aprt* gene was shown to lead to methylation of the *Aprt* CGI suggesting that transcription keeps it free of methylation (Brandeis et al., 1994; Macleod et al., 1994). One way in which this could be achieved is if the presence of the transcriptional machinery at a CGI inhibits recruitment of Dnmts. However, CGIs tend to be larger than the area occupied by the basal transcription machinery. Transcriptionally active CGIs often act as origins of replication and an alternative hypothesis is that initiation of replication could help prevent Dnmt action, thus maintaining CGIs (Antequera, 2003; Delgado et al., 1998). At CGI promoters, transcription initiation occurs broadly over hundreds of base pairs (Carninci et al., 2006) and, in addition, divergent initiation occurring in both orientations has been described at CGI promoters (Core et al., 2008). The combination of these phenomena could help to increase the methylation-free footprint maintained by transcription

If transcription from a CGI is necessary to keep it free of methylation all CGIs would need to be transcribed during early development when genome-wide waves of

methylation occur. Genome-wide *de novo* methylation occurs at the blastocyst stage of the embryo and later in the developing germ cells (Morgan et al., 2005). An early example supporting this hypothesis is that of the tissue-specific gene α-globin which is transcribed in the embryo and possesses a CGI while the related β-globin gene is silent and lacks a CGI (Daniels et al., 1997). Several other tissue-specific genes possessing CGIs have also been found to be expressed in early development (Daniels et al., 1995; Gardiner-Garden and Frommer, 1994). More compelling evidence comes form a study showing that 93% of a panel of genes expressed in the early embryo have a promoter CGI (Ponger et al., 2001). Embryonic stem cells have, more recently, been shown to be transcriptionally hyperactive with tissue-specific genes seen to be expressed at low levels (Efroni et al., 2008). If this is reflective of the situation in the early embryo such promiscuous transcription could help keep CGIs free of methylation. A recent genome-wide study showed that 90% of CGI-associated genes were expressed in early embryo or testis and that 83% of annotated promoter CGIs functioned as origins of replication in ES cells (Sequeira-Mendes et al., 2009). This adds support to the notion that transcription, perhaps with the aid of replication initiation, acts to keep CGIs free of methylation.

Large-scale promoter analysis using a method such as cap analysis of gene expression (CAGE) or global run-on sequencing (GRO-Seq) (Core et al., 2008; Faulkner et al., 2009) in the early embryo and in developing germ cells would allow promoter activity to be mapped so that it could be compared to the location of CGIs and, in particular, non-TSS-associated CGIs, in the genome. A major limitation, however, is likely to be the limiting amounts of material available from developing germ cells.

CGIs are generally free of methylation in all cell types even when they are not expressed, despite the fact that *de novo* methylation can occur outside of early development (Meissner et al., 2008; Mohn et al., 2008). How then, are CGIs kept methylation free during other developmental stages with more restrictive expression patterns? One hypothesis is that *de novo* DNA methylation has to be targeted to

specific sequences in these contexts and does not promiscuously affect CGIs. This could be accomplished by recruitment of Dnmts to areas marked by repressive histone modifications as occurs in X inactivation where H3K27 tri-methylation precedes CGI methylation [reviewed in (Payer and Lee, 2008) and (Okamoto and Heard, 2009)].

Another possibility is that the presence of CFP1 and H3K4me3 renders CGIs refractory to methylation. This suggestion initially came from evidence showing that binding of Dnmt3L, the partner protein of the *de novo* Dnmts is inhibited by H3K4 tri-methylation (Ooi et al., 2007). A more recent report demonstrated that Dnmt3a complexes methylate chromatin modified by H3K4me3 less efficiently than unmodified H3 or H3K9me3 (Zhang et al., 2010). However, Thomson and colleagues (2010) see acquisition of methylation at an artificially inserted CGI despite the presence of CFP1 and H3K4me3.

1.5.5 CGI methylation

It was previously assumed that all CGIs were unmethylated except in special cases, such as X inactivation and imprinting, and aberrant situations such as cancer. Many studies have now shown that a small percentage of CGIs become methylated during normal development and that there are tissue-specific patterns of CGI methylation. CGI methylation is clearly associated with gene silencing although whether it initiates transcriptional silencing or "locks-down" already repressed genes is less certain.

CGI methylation occurs in X inactivation

Dosage compensation in female mammals occurs by inactivating of one of the X chromosomes through formation of a repressive chromatin environment. This is directed by the non-coding RNA (ncRNA) Xist which coats the future inactive X. Xist coating is followed by establishment of a repressive chromatin state involving polycomb recruitment and subsequent H3K27 tri-methylation of the X chromosome.

A later event is the acquisition of DNA methylation at X-linked CGIs. This is accomplished by the action of Dnmt3a and 3b and occurs after gene silencing [X inactivation is reviewed in (Payer and Lee, 2008) and (Okamoto and Heard, 2009)]. On the active X chromosome the CGI of the *Xist* gene becomes methylated ensuring that *Xist* is not transcribed and the chromosome is not subject to spurious inactivation. Depletion of methylation, via mutation of Dnmt1 or treatment with the chemical 5-azacytidine (5-azaC), leads to reactivation of X-linked genes (Csankovszki et al., 2001). The order of events in X inactivation suggests that CGI methylation is targeted to areas of already silent chromatin and is not the initiating event in transcriptional shut-down.

Control of imprinting involves CGI methylation

Imprinted genes are monoallelically expressed and this is determined by parental origin of the allele. Imprinting is controlled by imprinting control regions and often involves differential methylation of CGIs within these regions. CGI methylation can control imprinting in a number of ways. The CGI can be a promoter for a long ncRNA that induces silencing of the imprinted gene. One well-characterised example is that of the maternally expressed cluster of genes *Igf2r/Slc22a2/Slc22a3*. Expression of this cluster is inhibited by a paternally imprinted ncRNA called *Air* which might act by targeting repressive histone modifications to the locus. The *Air* CGI is methylated on the maternal allele which prevents *Air* expression and thus allows *Igf2r/Slc22a2/Slc22a3* to be transcribed (Nagano et al., 2008; Sleutels et al., 2002). Differentially methylated CGIs can also act as binding sites for proteins with insulator function such as CTCF. Binding of CTCF to the CGI can prevent the action of enhancers on the imprinted gene. In these cases methylation interferes with insulator function by preventing CTCF binding, thereby promoting gene expression. A CGI in the imprinted *Igf2/H19* locus acts as an insulator controlling expression of *Igf2* in a methylation-dependent manner. In addition, many differentially methylated CGIs are found at gene promoters where methylation can directly inhibit expression of the methylated allele [a more comprehensive discussion of imprinting can be found in (Edwards and Ferguson-Smith, 2007)].

Abnormal CGI methylation is a hallmark of cancer

Aberrant acquisition of CGI methylation in cancer has been the focus of much research. The CGIs of several tumour suppressor genes have been shown to acquire cancer-specific methylation and many genes involved in familial forms of cancer undergo methylation-associated silencing in sporadic cancers (Jones and Baylin, 2002). This is associated with gene silencing and is thought to contribute to uncontrolled proliferation and thus tumour development. Methylation in tumours has been found to affect specific groups of CGIs and this has been proposed to be a distinct phenotype known as the CGI methylator phenotype (Toyota et al., 1999). Genome-wide studies have confirmed that a small proportion of CGIs acquire methylation in cancer (Illingworth et al., 2010; Ruike et al., 2010; Weber et al., 2005). Tumour-specific CGI methylation is now known to affect genes with a variety of functions and not just tumour-suppressors (Keshet et al., 2006), although for some of these repression does not seem to be dependent on methylation (Costello et al., 2000). Whether cancer CGI methylation mirrors normal developmental methylation or chromatin state has received much recent attention. Acquisition of methylation in cancer has been found to preferentially occur at genes marked by H3K27me3 in ES cells (Ohm et al., 2007; Widschwendter et al., 2007). A recent report identified methylation occurring at the edges of CGIs in cancer and this was found to be the same as normal tissue-specific methylation (Irizarry et al., 2009). However, a subsequent study examining bone fide CGI methylation found the pattern of cancer-specific methylation to be distinct to that of normal CGI methylation (Illingworth et al., 2010). Paradoxically, increased CGI methylation in cancer is accompanied by a genome-wide decrease in methylation associated with genome instability (Jones and Baylin, 2007).

CGI methylation occurs during normal development

There has been a plethora of studies in recent years profiling CGI methylation in normal tissues and it is now clear that CGI methylation occurs during development outwith X inactivation and imprinting. An initial study profiling methylation of three human chromosomes identified approximately 10% of CGIs as being extensively

methylated (Eckhardt et al., 2006). Weber and colleagues (2007) later surveyed methylation at the majority of promoters in human sperm and fibroblast cells. Sperm CGI promoters were found to be hypomethylated, confirming previous reports. They found that approximately 3% of strong CGI promoters were methylated in somatic cells compared to sperm but more methylation was seen at a class of weak CGI promoter (21% methylated). A significant proportion of CGIs that acquired methylation in fibroblasts (17%) were associated with germline-specific genes (Weber et al., 2007). Illingworth et al found 6-8% of CGIs to be methylated in any one human tissue and showed that CGIs outwith annotated gene promoters preferentially acquire methylation (Illingworth et al., 2008). *In vitro* differentiation systems have been used to track the extent of methylation changes during development. A gain in CGI methylation occurs during differentiation of ES cells to neurons, with the majority occurring at the early stages of differentiation and being maintained in all somatic tissues (Mohn et al., 2008)[ii]. There have been numerous other reports describing CGI methylation in normal tissues [for example (Meissner et al., 2008; Schilling and Rehli, 2007; Shen et al., 2007)]. Such methylation studies have shown that a small percentage of CGIs become methylated in somatic tissues and that there are CGI methylation differences between cell types, although these occur less frequently (Meissner et al., 2008).

CGIs that can acquire methylation tend to be less CpG-dense than CGIs refractory to methylation (Eckhardt et al., 2006; Illingworth et al., 2010; Meissner et al., 2008; Weber et al., 2007). A likely explanation for this is the mutagenic effect of methylation which depletes CpGs from the genome. Alternatively, CpG-density could correlate with CFP1 occupancy and therefore the H3K4me3 level which may protect against methylation. Indeed, Illingworth and colleagues (2010) have shown H3K4me3 to decrease with CpG density.

[ii] Whilst sperm CGIs are devoid of methylation (except at paternally imprinted CGIs) a very small percentage of CGIs are methylated in ES cells (Mohn et al., 2008; Meissner et al., 2008; Fouse et al., 2008).

Cell type-specific CGI methylation and its relationship to gene expression

A major question remaining involves the significance of somatic CGI methylation and its relationship to gene expression. CGI methylation is clearly associated with transcriptional silencing but does differential methylation correspond to differential gene expression? A study using restriction landmark genomic scanning (RLGS) to identify tissue-specific CGI methylation found that for several genes, differential CGI methylation, corresponded to differential gene expression (Song et al., 2005) and this was also the case in a study by Schilling and Rehli (2007). However, Eckhardt and colleagues observed that a large proportion of CGIs specifically unmethylated in sperm but methylated in somatic cells did not show germline-specific expression of the associated gene (Eckhardt et al., 2006) and Illingworth et al did not find any obvious correlation between differential CGI methylation and gene expression (Illingworth et al., 2008).

Another factor making the relationship between differential CGI methylation and gene expression less clear is the existence of large numbers of non-promoter-associated CGIs. These intra- and intergenic CGIs represent the majority of methylated CGIs (Illingworth et al., 2010; Illingworth et al., 2008; Maunakea et al., 2010; Rauch et al., 2009) and two recent studies have identified many of these as being novel promoters (Illingworth et al., 2010; Maunakea et al., 2010). Could CGIs within gene bodies be promoters for ncRNAs that regulate expression of the associated gene? Or might they represent an alternative TSS for that gene as suggested by Maunakea et al. (2010)? Alternatively, can methylation of an intragenic CGI affect transcription of the associated gene by another mechanism?

1.5.6 CGI methylation – cause or consequence?

Whether CGI methylation is an initiating event in gene silencing, a maintenance mechanism or a mere consequence of a transcriptionally repressed state remains unclear. CGI methylation is often preceded by gene silencing and repressive histone modifications, for example in X inactivation, arguing that DNA methylation is

targeted to already silent chromatin (Csankovszki et al., 2001). Mohn and colleagues (2008) found that more than half of CGIs that acquired methylation upon differentiation were already silent in ES cells prior to *de novo* methylation. Many of these were marked by H3K27me3 in ES cells but lost H3K27me3 and acquired DNA methylation upon differentiation. Studies showing gene reactivation as a result of depleting methylation argue that CGI methylation is essential for maintaining gene silencing (De Smet et al., 1999; Li et al., 1993; Sado et al., 2000). Most evidence seems to argue for targeting of CGI methylation to silent genes where it acts to maintain gene silencing in a heritable manner. It is also possible that CGI methylation can be a stochastic event occurring in the absence of transcriptional activity. This is supported by the occurrence of differential CGI methylation in tissues where the associated gene is constitutively repressed (Illingworth et al., 2008; Schilling and Rehli, 2007). However, a role for CGI methylation in initiation of gene silencing cannot be ruled out. Silencing of the trophectoderm-specific *Rhox* genes in the embryo proper has been shown to be highly dependent on *de novo* methylation mediated by Dnmt3a and 3b (Oda et al., 2006).

1.6 Epigenetics and the immune system

Immune system cells represent a distinct developmental lineage all derived from a common progenitor. Mature immune cells can further differentiate into specialised effector cells designed to deal with particular pathogens and some of these become memory cells which are poised to respond to re-infection. These factors make the immune system an appealing model in which to study epigenetic mechanisms. Epigenetic modification at the key cytokine loci, *Ifng* and *Il4* has been extensively documented during T cell differentiation and, in addition, recent genome-wide studies have helped to shed light on the role of histone modifications in regulating T cell gene expression.

1.6.1 The *Ifng* and *Il4* loci undergo epigenetic changes during T helper cell differentiation

The first evidence for involvement of epigenetic mechanisms in T helper cell differentiation came from the observation that several rounds of cell division are required in order to achieve signature cytokine expression during this process. Naïve CD4 T cells can differentiate into Th1 cells expressing IFN-γ and repressing IL-4 or Th2 cells expressing IL-4 and silencing IFN-γ. Treatment of differentiating cells with agents that promote histone acetylation or prevent DNA methylation allowed earlier and increased expression of these cytokine genes in their appropriate lineages (Bird et al., 1998). Since then, histone acetylation (Avni et al., 2002; Baguet and Bix, 2004; Fields et al., 2002; Fields et al., 2004; Grogan et al., 2003; Morinobu et al., 2004) and DNA methylation (Jones and Chen, 2006; Kim et al., 2007; Lee, 2002; Makar and Wilson, 2004; Santangelo et al., 2002; Schoenborn et al., 2007; Winders et al., 2004; Yano et al., 2003) changes at the *Ifng* and *Il4/Il5/Il13* loci have been profiled extensively during T cell differentiation as well as the appearance of DNase I hypersensitive sites, an indicator of open chromatin conformation (Agarwal and Rao, 1998; Fields et al., 2004; Santangelo et al., 2002; Schoenborn et al., 2007; Takemoto et al., 2000) and, more recently, histone methylation marks (Chang, 2007; Schoenborn et al., 2007; Wei et al., 2009). In general, the *Ifng* locus shows increased histone acetylation and reduced DNA methylation in Th1 cells while the *Il4/Il5/Il13* locus acquires acetylation and looses DNA methylation in Th2 cells consistent with gene activity. These epigenetic changes occur not only at gene promoters but also at conserved regulatory elements throughout the cytokine gene loci and are dependent on lineage-specific transcription factors and signalling events. Appropriate cytokine expression during T cell differentiation has been found to depend not only on DNA methylation but on MBD2, a reader of this mark [(Hutchins et al., 2002) and section 1.3.6].

In many studies, DNA methylation changes at the *Ifng* and *Il4/Il5/Il13* loci have been shown to require cell division indicating a gradual accumulation of methylation (Bird et al., 1998; Lee, 2002). Specific CpG sites, however, have been shown to undergo

rapid methylation changes. A CpG site in the *Ifng* promoter is rapidly methylated during Th2 differentiation and this interferes with transcription factor binding (Jones and Chen, 2006) while a CpG dinucleotide in the *Il4/Il5/Il13* locus control region undergoes active demethylation under Th2 conditions (Kim et al., 2007) . Upon T cell activation a site in the *Il2* promoter shows active demethylation concomitant with recruitment of the activator Oct-1 to this site (Bruniquel and Schwartz, 2003; Murayama et al., 2006). However, none of these genes is associated with a CGI.

1.6.2 Dnmts are required in immune system cells

Insights into the role of DNA methylation in the immune system have been gained from analysis of the phenotypes of immune-specific Dnmt knockouts. A T cell-specific knockout (CD4 and CD8 cells) of the maintenance methyltransferase Dnmt1 showed reduced proliferation of naïve CD4 cells upon stimulation but increased expression of *Il2*, *Il4* and *Ifng* consistent with a role for DNA methylation in regulating appropriate expression of these genes (Lee et al., 2001). Dnmt1 and Dnmt3b were found to be recruited to regulatory regions of the *Il4* locus in resting thymocytes (Makar et al., 2003) and, intriguingly, these regions are reported to be bound by MBD2 in Th1 cells (Hutchins et al., 2002). Removal of Dnmt1 specifically in cytotoxic CD8 T cells leads to reduced immunity to viral infections due to a massive proliferation defect as well as impaired differentiation into effector and memory cells (Chappell et al., 2006).

Dnmt3a has recently been shown to be upregulated following T cell activation and demonstrated to be necessary for establishment of DNA methylation at the *Ifng* locus to ensure silencing under Th2 conditions. When *Dnmt3a-/-* T cells were stimulated and cultured over time, they simultaneously expressed *Ifng* and *Il4* and, consistent with this, the genes were demethylated. Intriguingly, these cells also co-expressed *Tbx21* and *Gata3* suggesting that these genes might also be regulated by methylation (Gamper et al., 2009).

The effect of depleting Dnmt1 earlier in development of the hematopoietic lineage has recently been described (Broske et al., 2009). Immune cells develop from a hematopoietic stem cell (HSC) which gives rise to either; a myeloid progenitor cell that is the precursor of monocytes, granulocytes and dendritic cells or a lymphoid progenitor from which B and T cells are formed. Conditional knockout of Dnmt1 resulted in complete loss of HSCs and progenitor cells. Reducing Dnmt1 levels using a hypomorphic allele impaired HSC self-renewal and, in addition, HSCs were unable to suppress key myeloid genes. These mice could not make lymphoid progenitors due to a failure to silence the myeloid differentiation programme and as a consequence had drastically reduced numbers of B and T cells. In contrast, reducing DNA methylation after commitment to the B cell lineage had no effect on the number of mature B cells produced. This indicates that DNA methylation may be more important in specifying progenitor cell identity rather than maintaining the identity of differentiated cells and suggests that methylation patterns may be set up early in hematopoiesis.

1.6.3 Global changes in histone modification during T cell activation and differentiation

The genome-wide distribution of various histone modifications has been profiled in human and mouse T cells both with a view to understanding the role of histone modification in gene regulation generally and to elucidating T cell-specific roles. Some of the first studies profiling histone acetylation and methylation genome-wide by ChIP-Seq were carried out in resting human CD4 T cells. These drew general conclusions such as the association of specific histone methylation and acetylation marks with the promoters of active genes and also characterised bivalent promoters marked by both H3K27me3 and H3K4me3 in these cells (Barski et al., 2007; Wang et al., 2008). A subsequent study compared the pattern of histone modifications in resting CD4 cells to that of activated CD4 cells (i.e. those stimulated to respond to infection) and related this to gene expression. Genes that were induced upon activation were found to be poised for expression in resting cells and bore active chromatin modifications and together with Pol II at their promoters even though the genes were silent (Barski et al., 2009). Another study examining T cell stimulation

also found that the chromatin state was often set prior to gene activation (Smith et al., 2009). This suggests that few epigenetic changes occur rapidly upon T cell activation and that a pre-existing chromatin state defines the ability of T cells to respond to stimuli in the short-term.

However, longer-term differentiation into effector T cells is associated with some epigenetic changes. A study profiling the histone modifications H3K4me3 and H3K27me3 in naïve, Th1, Th2 and two recently discovered T cell types, Th17 and Treg cells, identified a small number of genes that change their histone modification state upon differentiation. This demonstrates that genes apart from *Il4* and *Ifng* undergo epigenetic changes during T cell differentiation. Interestingly, the genes encoding two key regulators of T helper cell fate, *Tbx21* (the gene for T-bet, a key Th1 transcription factor) and *Gata3* (the key Th2 transcription factor) were found to be bivalent in cells when they were silent (Wei et al., 2009).

1.6.4 DNA methylation changes in immune system cells

In addition to the well-documented DNA methylation changes that take place at the *Ifng* and *Il4* loci during T cell differentiation other genes have been reported to show methylation differences in immune system cells. Tregs show lineage-specific DNA demethylation at the gene for their master regulator Foxp3 (Floess et al., 2007; Kim and Leonard, 2007; Lal et al., 2009; Polansky et al., 2008). The gene encoding IL18-Rα which is expressed in Th1 but not Th2 cells is differentially methylated in the two cell types and this involves a mechanism dependent on lineage-specific transcription factors (Yu et al., 2007).

DNA methylation of key immune system genes has also been investigated in contexts outwith T cell differentiation. For example, the B cell-specific gene *Cd19* shows demethylation of an upstream enhancer in all immune cell progenitors regardless of gene expression. At the *Cd19* promoter, however, the level of methylation inversely correlates with gene expression. Demethylation coincides with

gene expression in early B cell stages and is further enhanced during B cell differentiation (Walter et al., 2008). The DNA methylation state of 13 immune-specific genes has been profiled in human naïve and memory CD4 T cells, CD8 T cells and B cells as well as in hematopoietic progenitors. Regions of the *Spi1*, *Gata3*, *Tcf-7*, *Etv5*, *c-maf* and *Tbx21* promoters showed methylation differences between cell types. At the majority of these genes, methylation differed between naïve and memory cells of the same cell type (e.g. naïve and memory CD4 T cells) rather than between different cell types such as B and T cells (Ivascu et al., 2007). A study using targeted microarrays identified regions differentially methylated between Treg cells and conventional CD4 T cells and extended this analysis to include other immune cell types including B cells, monocytes, hematopoietic progenitor cells as well as the Jurkat T cell line. Most of the reported methylation differences occurred distal to gene promoters (Schmidl et al., 2009).

These observations make it appealing to look globally for DNA methylation changes between immune system cells, although only a few of the examples described occur at CGIs. Two studies have characterised the "methylome" in CD4 T cells but have not compared between cell types or investigated methylation with regards to immune system function (Hughes et al., 2010; Smith et al., 2009). Studies in T cells describing DNA methylation and histone modifications such as H3K4me3, which is intimately associated with CGIs, do not differentiate between CGI and non-CGI promoters (Barski et al., 2009; Hughes et al., 2010; Smith et al., 2009). This may complicate some of the correlations between epigenetic modification and gene expression observed. Therefore it is worthwhile examining the methylation status of features with a clear function such as CGIs with respect to their behaviour in different immune cell lineages.

1.7 PhD Objectives

The aim of this PhD is to use an immune system model to assess the role of CGI methylation and the role of the methylation reader MBD2 in the regulation of gene

expression. CGI methylation will be profiled globally and compared to gene expression as well as H3K4me3 distribution in different cells of the hematopoietic lineage. After CGI methylation in wild-type cells is assessed, cells lacking MBD2 will be investigated, as depletion of MBD2 has been shown to give immune system phenotypes. Gene expression changes in *Mbd2-/-* immune cells will be examined and related to histone acetylation and DNA methylation. It is hoped this study will provide insights into the role of CGI methylation and MBD2 generally as well as assessing their importance in immune cells.

Chapter 2 Materials and Methods

The main reagents and experimental methods used will be outlined in this chapter. Some standard molecular biology reagents and protocols will not be described and details of these can be found in standard text books such as "*Molecular Cloning: A laboratory manual*" (Sambrook and Russell, 2001). All experiments were carried out at room temperature (R/T) unless otherwise stated.

2.1 Experimental methods

2.1.1 Immune cell purification and FACS staining

Immune cell isolation and purification

Immune cells were isolated from the spleens of C57/BL6 mice unless otherwise stated. Spleens were physically disrupted and cells were passed through a cell strainer (70µm; BD Falcon) into Hanks media and pelleted by centrifugation at 300g for 4min at 4°C. For isolation of splenic dendritic cells, spleens were chopped into small pieces and incubated in Hanks supplemented with DNase at 37°C for 15min with agitation prior to disruption in order to facilitate recovery. Red blood cells were lysed using Red Blood Cell Lysing Buffer (Sigma) according to manufacturers' instructions. Splenocytes were then counted using a haemocytometer combined with trypan blue staining to check cell viability and the cells purified using magnetic cell sorting (MACS, Miltenyi Biotec) or fluorescence activated cell sorting (FACS).

CD4 T cells were purified using the MACS system (Miltenyi Biotec). Hanks media was used as a buffer throughout. For all MACS purifications, buffers and reagents were kept at 4°C. Splenocytes were diluted in the appropriate volume of buffer (90µl beads per 10^7 cells) and incubated with magnetic CD4 microbeads (10µl beads per 10^7 cells) at 4°C for 15min. Cells were then washed and CD4+ cells selected using

LS cell separation columns (Miltenyi Biotec) according to manufacturers' in instructions. To isolate naïve CD4 T cells, CD4+ cells recovered by MACS were stained for the cell surface markers CD44 and CD62L. Naïve cells expressing low levels of CD44 and high levels of CD62L (Swain et al., 1991) were sorted by FACS. Dendritic cells were also isolated using the MACS system and MACS buffer was used for all purification steps. First, plasmacytoid dendritic cells which are lymphoid in origin were depleted using α-PDCA microbeads (400µl buffer and 100µl beads per 10^8 cells) and LD columns according to manufacturers' instructions. Then dendritic cells were blocked for 5min with 10µg/ml α-FcR, selected using CD11c microbeads (400µl buffer and 100µl beads per 10^8 cells) and purified by two passes over an LS column. Selection of CD11c+ cells and depletion of PDCA+ cells was confirmed by FACS analysis. To isolate B cells, splenocytes were stained for the B cell marker CD19 and cells positive for this marker were sorted by FACS.

T cell differentiation

T cell differentiation was carried out using CD4+ cells purified by MACS as described above. For Th1 and Th2 differentiation α-CD3 and α-CD28 (1.6µg of each/well in 250µl PBS) were bound to 24-well tissue-culture plates for 1-2 hours prior to plating cells and any excess removed by washing each well three times with PBS. Cells were seeded in "T cell media" at a density of $1.5x10^6$ cells/well (*Mbd2-/-* cells were seeded at $2x10^6$ cells/well to compensate for a reduced rate of proliferation) along with recombinant murine IL-2 (20ng/ml; Peprotech) and cytokines and antibodies to skew towards Th1 [IL-12 (5ng/ml; R&D) and α-IL-4 (5µg/ml; BD Pharmingen)] or Th2 [IL-4 (10ng/ml; Peprotech) and α-IL-12 (5µg/ml; BioLegend)] cell fate in a final volume of 1ml. Cells were cultured at 37°C, 5% CO_2 and split approximately every two days replenishing cytokines and antibodies but without re-stimulating with α-CD3/α-CD28. Th1 and Th2 cells were cultured for 7 days and then stimulated with PMA at 50ng/ml and ionomycin at 500ng/ml for 4 hours prior to further analysis. For cells that were to be analysed by intracellular FACS staining a 1:2000 dilution of GolgiStop (BD Pharmingen) was added 2 hours prior to harvesting to help retain cytokines in the cell.

For *in vitro* generation of Treg cells culture conditions were similar to those described for Th1 and Th2 cells except that cells were cultured in 96-well plates and volumes scaled down accordingly. Cells were treated with plate-bound α-CD3 and α-CD28 (0.4μg/ml) along with IL-2 (20ng/ml) and varying concentrations of recombinant human TGF-β1 (0.01-10ng/ml; gift from R. Maziels lab). Tregs were differentiated for 72 hours and then stained for Foxp3 expression.

For proliferation studies ^{3}H-thymidine was added to cultures (0.5miC/well) 24 hours after seeding and incorporation analysed after an additional 24 hours by scintillation counting. Alternatively, cells were labelled with CFSE (final concentration 5μM) prior to seeding, then stimulated as normal. After 72 hours the CFSE signal was visualised by FACS.

Note: Treg generation and proliferation experiments were carried out as part of a collaboration with John Grainger.

Isolation of T cells from schistosome-infected mice

Six spleens were harvested from mice that had been infected with *Schistosoma mansoni* for 8 weeks (infection carried out by P. Cook and L. Jones). Splenocytes were isolated as described above and CD4 cells purified by MACS. Cells were plated in "Ex vivo media" at a density of $4x10^{6}$ cells per well in a 24-well plate. The CD4 cells were stimulated for 4 hours with 10ng/ml PMA and 1μg/ml ionomycin at 37°C, 5% CO_2 and 2 hours prior to harvest 1:1000 GolgiStop was added. Cells were stained for IL-4 and IFN-γ production and IL-4+ and IL-4+ IFN-γ+ cell populations were FACS sorted.

FACS staining and sorting

All FACS procedures were carried out at 4°C. FACS staining for cell surface markers (i.e. CD4, CD44, CD62L, CD11c, PDCA, CD19) was carried out as follows.

Cells were washed once in FACS buffer and then stained with the appropriate fluorescent antibodies diluted in FACS buffer for 20min protected from light (fluorophores and antibody concentrations are given in Table 2.2.6). Cells were then washed three times in FACS buffer to remove excess antibody and resuspended in an appropriate volume of buffer for FACS analysis or sorting.

Intracellular staining (i.e. for IFN-γ and IL-4) was carried out as follows. Cells were washed in PBS and fixed in 4% paraformaldehyde for 15min. Cells were then washed three times in Perm buffer and stained for cytokine production using fluorescent antibodies diluted in Perm buffer (see Table 2.2.6). Staining was carried out for 30min protected from light. After staining, cells were washed twice in Perm buffer and resuspended in PBS for FACS analysis or sorting. Foxp3 staining was carried out in a similar manner except that a commercial fixation and permeabilisation kit was used (Foxp3 Fixation/Permeabilization Buffer; eBioscience).

Unstained and single colour controls were prepared for each FACS experiment. FACS analysis was carried out using a BD FACS Canto machine and FACS sorting was carried out with the help of Martin Waterfall using a BD FACS Aria.

2.1.2 DNA manipulation

Genomic DNA extraction

To extract genomic DNA from immune cells, cells were washed in PBS then resuspended in an appropriate volume of genomic DNA extraction buffer (500µl per $3x10^6$ cells) supplemented with proteinase K (40µg/ml) and incubated at 55°C overnight. RNase A (20µg/ml) was then added and the solution was incubated for a further 1 hour at 37°C. DNA was extracted by adding an equal volume of phenol:choloroform:isoamylalcohol (25:24:1, Sigma), vortexing and centrifuging for 15min at 12000g. The aqueous phase was recovered and DNA precipitated by

addition of an equal volume of isopropanol. The DNA pellet was washed in 70% ethanol, dried and resuspended in an appropriate volume of dH_2O. DNA concentration was determined by measuring the absorbance of the solution at 260nm (OD_{260nm}) using a Nanodrop-1000 spectrophotometer. DNA concentration was automatically calculated according to Beer's law (concentration = OD_{260nm} x molar absorptivity constant/path length) whereby an OD_{260nm} value of one indicates a DNA concentration of 50ng/µl.

Polymerase Chain reaction (PCR)

PCR was used to amplify DNA molecules of interest from a starting template and was used primarily to test primers for qPCR, verify enrichment of methylated sequences in MAP fractions and in bisulfite genomic sequencing. Reactions were generally carried out in a 20µl volume and consisted of DNA template (0.1-50ng), 400nM forward and reverse primers, 400µM dNTPs (Abgene), 2.5mM $MgCl_2$, Red Hot Taq reaction buffer (Abgene) and 1.5U Red Hot Taq (Abgene). PCR amplification of bisulfite-treated DNA was carried out in the presence of 3% dimethyl sulfoxide (DMSO). PCR reactions were set up on ice. PCR cycling was carried out on a G-Storm thermal cycler and typical conditions were as follows: initial denaturation at 94°C for 2min followed by 30 cycles of denaturation at 94°C for 30sec, primer annealing at T^{anneal} for 30sec and primer extension at 72°C for 30sec. An additional 72°C primer extension phase for 5min was carried out at the end of the 30 cycles to amplify any incomplete DNA molecules. In general, to determine the optimal annealing temperature (T^{anneal}) for a particular primer pair a range of temperatures was tested (using the gradient feature of the thermal cycler over a range of 53-64°C). PCR amplification of bisulfite-treated DNA was carried out for 40 cycles with longer step times: 94°C for 40sec, T^{anneal} for 50sec and 72°C for 50sec. PCR products were resolved by agarose gel electrophoresis.

Quantitative PCR (qPCR)

Quantitative PCR (qPCR) was used to accurately quantify a target region in a DNA sample of interest and was used to assess enrichment of specific regions in ChIP samples and to measure relative expression of genes by qRT-PCR (see 2.1.3). To carry out qPCR I used SYBR Green technology. SYBR Green I is a fluorescent dye that binds to DNA and, when included in a PCR reaction, allows fluorescence to be used as a read-out for the amount of DNA synthesised in real-time. qPCR reactions (25µl) contained SYBR Green SensiMix (Quantace), 250nM primers and 0.1-10ng template DNA. PCR was carried out using a Roche Lightcycler and cycling conditions were as follows; initial denaturation at 94°C for 10min followed by 45 cycles of denaturation at 94°C for 10sec, primer annealing at T^{anneal} for 10sec and primer extension at 72°C for 15sec. Data were collected at the end of each amplification cycle. After amplification, melting curves for PCR products were generated by denaturing the DNA at 94°C for 1min and then increasing the temperature from 35°C to 95°C in increments of 0.1°C. A single melt curve indicates the presence of a single PCR product whilst multiple melt curves could indicate the presence of multiple PCR products or primer dimers that would interfere with DNA quantification.

Using the Roche Lightcycler software, SYBR Green fluorescence measurements were plotted relative to cycle number to generate curves for each PCR reaction and the 2^{nd} derivative maximum method was used to determine the cycle threshold values (Ct) for each sample. An arbitrary measure of DNA quantity was calculated using the formula 2^{-Ct}. This value was then compared to that of a reference sample (for example input DNA in a ChIP reaction or expression of a reference gene in qRT-PCR). qPCR reactions were carried out in duplicate or triplicate and the standard deviation calculated for each biological sample.

DNA gel electrophoresis

DNA was resolved by gel electrophoresis using the Bio-Rad Sub-Gel system. Agarose gels (1-2%; depending on the size of the DNA fragment to be resolved) were prepared with TAE containing 0.5µg/ml ethidium bromide, a DNA intercalating agent that fluoresces in ultraviolet (UV) light. DNA samples and appropriate size markers (100bp or 1kb ladder; NEB) prepared in orange G loading buffer were loaded into the wells of the gel. Gels were run at constant voltage (80-110V) in TAE electrophoresis buffer and visualised under UV light.

Bisulfite genomic sequencing

Genomic DNA was sonicated using a Diagenode Bioruptor for 10sec on high power in order to break up DNA and facilitate bisulfite treatment. 2µg of this DNA (25µl volume) was used in each bisulfite reaction. DNA was denatured by adding 2.5µl of freshly prepared 3M NaOH and incubating at 37°C for 20mins. 270µl of bisulfite solution was added to the DNA and the reaction was overlaid with 200µl of mineral oil to prevent evaporation. The bisulfite reaction was incubated at 55°C for 5 hours protected from light. DNA was then isopropanol precipitated (including 50µg glycogen carrier; Roche) and resuspended in 25µl TE buffer. DNA was desulfonated by the addition of 2.5µl 3M NaOH and incubation at 37°C for 15min. The DNA was then precipitated with three volumes of 100% ethanol and 1/10 volume 5M ammonium acetate pH 7, washed in 70% ethanol and resuspended in 50µl of TE. Typically 0.5µl of this DNA was used in a PCR reaction containing primers specific for bisulfite modified DNA and other reaction components as detailed above. PCR products were resolved by agarose gel electrophoresis and gel extracted as follows. The product of interest was excised from the gel, placed at -80°C and subject to a number of freeze-thaw cycles. After this the gel fragment was centrifuged for 15min at 12000g to liberate DNA in solution from the gel.

Bisulfite PCR products were then cloned using the StrataClone PCR Cloning Kit (Stratagene) and transformed into StrataClone Solo Pack Competent Cells

(Stratagene) according to manufacturers' instructions. Transformed cells were spread onto ampicillin LB agar plates containing X-gal for blue/white selection. Colonies positive for inserts appeared white in colour due to disruption of the β-galactosidase gene by the cloned insert. These colonies were selected and the cloned insert amplified by PCR using the T7 and M13 reverse primers designed against the cloning vector backbone. A proportion (5µl) of each PCR reaction was analysed by gel electrophoresis to verify that cloned inserts were of the correct size. To prepare PCR products for sequencing, exonuclease I (5U; NEB) and Antarctic phosphatase (5U; NEB) were added to the remaining 15µl along with 11.75µl of dH_2O. The reaction was incubated at 37°C for 15min followed by heat inactivation of the enzymes. This step degrades single-stranded primers and removes the phosphate groups from dNTPs thus preventing residual PCR components from interfering with the sequencing reaction. DNA sequencing of these PCR products was carried out as outlined below using the T7 primer. Bisulfite sequencing results were analysed using the BiQ Analyzer software package (Bock et al., 2005).

DNA sequencing

DNA sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). Sequencing reactions were carried out in a volume of 10µl and contained 4µl of 2.5x sequencing buffer, 2µl BigDye Terminator, 500nM primer and 3.5µl DNA (prepared for sequencing as above). Reaction conditions consisted of initial DNA denaturation at 94°C for 10sec followed by 24 cycles of DNA denaturation at 94°C for 30sec, primer annealing at 50°C for 20sec and extension at 60°C for 4min. Sequencing reactions were then passed onto the Gene Pool Sequencing Service (School of Biological Sciences) where they were cleaned up and run on an ABI 3730 capillary sequencer.

Preparation of DNA for MAP-seq and ChIP-seq

Prior to purification by MAP, genomic DNA was sonicated to an average size of 500bp using a Diagenode Bioruptor on medium setting for approximately 2 minutes.

Fragmentation was verified by gel electrophoresis and Illumina sequencing adaptors were attached to the DNA as follows. To repair DNA ends after sonication DNA (25µg) was incubated with 400µM dNTPs, T4 DNA Polymerase (30U; NEB), DNA Polymerase I Large (Klenow) fragment (17U; NEB), T4 polynucleotide kinase (10U; NEB) and T4 DNA ligase buffer (NEB) in a volume of 150µl for 30min at 20°C. Enzymes were heat-inactivated, DNA ethanol precipitated (using 50µg glycogen as a carrier) and resuspended in 85µl dH_2O. A tail of 'A' bases was added to the end of repaired fragments to facilitate attachment of adaptors. This was carried out by incubating DNA with 200µM dATP and Klenow 3' to 5' exo- fragment (30U; NEB) in the presence of NEB buffer 2 for 30min at 37°C. DNA was ethanol precipitated and resuspended in 32µl dH_2O. Illumina paired end adaptors were annealed to each other in a reaction containing 28µl of each adaptor (100µM stock), 7µl 10x Red Hot Taq buffer and 7µl dH_2O. The reaction was incubated at 94°C for 2min and the temperature was then decreased from 94°C to 20°C over a period of 25min to allow the adaptors to anneal. These adaptors were ligated to DNA in a 200µl reaction containing; 32µl DNA, 70µl annealed adaptors, 50µl PEG 6000 (final concentration 7.5%), T4 DNA ligase (300U) and T4 DNA ligase buffer (NEB). The reaction was incubated for 15min at 20°C, enzymes heat-inactivated and reactions stored at -20°C until required for MAP.

For ChIP-seq experiments adaptors were attached to the small quantities of DNA recovered after ChIP. Reactions were carried out in a similar manner to the protocol described above except that enzyme concentrations and reaction volumes were adjusted as follows to account for the small amount of DNA present. End repair: 14.8µl ChIP DNA (10-20ng), 400µM dNTPs, T4 DNA Polymerase (3U; NEB), Klenow (2U; NEB), T4 polynucleotide kinase (1U; NEB) and T4 DNA ligase buffer (NEB) in a volume of 20µl. Addition of 'A' bases: 200µM dATP, Klenow 3' to 5' exo- fragment (5U; NEB) and NEB buffer 2 in a final volume of 20µl. Ligation: 0.5µl of annealed adaptors (1:10 dilution), 5µl PEG 6000, T4 DNA ligase (30U) and T4 DNA ligase buffer (NEB) in a volume of 20µl. The ligation reaction was carried out for 2 hours at R/T with gentle agitation. Reactions were cleaned up after adaptor

ligation using the Qiagen MinElute kit and DNA eluted in 10µl dH_2O. All reactions were carried out in siliconised PCR tubes or low-bind tubes (Eppendorf).

2.1.3 RNA manipulation

RNA extraction

RNA was extracted using TRI Reagent (Sigma) according to manufacturer's instructions. Cell pellets were resuspended in 1ml TRI Reagent (for 1-6 x 10^6 cells) and incubated for 5min to disrupt nucleoprotein complexes. Chloroform (200µl) was added and the solution was mixed vigorously and incubated for 10min. The mixture was centrifuged for 15min at 12,000g at 4°C and the aqueous phase containing the RNA recovered. RNA was precipitated using isopropanol, pelleted by centrifugation for 10min at 12,000g at 4°C, washed with 70% ethanol and resuspended in 25-50µl of nuclease free water (Ambion). To extract RNA from oesophagus tissue was harvested and immediately snap frozen in liquid nitrogen. Tissue was then homogenised in TRI Reagent using a Polytron homogeniser and extraction carried out as detailed above. RNA concentration was determined using a Nanodrop-1000 spectrophotometer. An OD_{260nm} value of one indicated an RNA concentration of 40ng/µl.

RNA gel electrophoresis

Integrity of the recovered RNA was determined by agarose gel electrophoresis. Electrophoresis was carried out as for DNA except that TBE buffer was used instead of TAE and was made using DEPC-treated dH_2O. RNA loading buffer (NEB) was added to samples which were then denatured at 65°C for 5min and snap chilled on ice. Samples and marker (RNA ladder; NEB) were loaded onto the gel which was run at 100V. RNA integrity was assessed by visualisation of the 28S and 18S rRNA bands.

DNase treatment of RNA

For cDNA synthesis and NimbleGen microarrays RNA was DNase treated prior to use in order to remove contaminating genomic DNA. This was carried out using the DNA-free kit (Ambion). Briefly, up to 2µg RNA was treated in a 50µl volume containing 4U DNase I and DNase I reaction buffer for 30min at 37°C. DNase was then inactivated by treating with DNase inactivation reagent according to manufacturers' instructions.

cDNA synthesis (reverse transcription)

Prior to reverse transcription (RT) RNA was denatured at 65°C for 5min and snap chilled on ice. RNA (250ng) was added to a 25µl reaction containing M-MLV reaction buffer (Promega), 5µM random hexamers (Roche), 1mM dNTPs, 40U RNasin (RNase inhibitor; Promega) and 200U of M-MLV reverse transcriptase (RNase H minus; Promega). A minus (-)RT reaction was set up in parallel containing all the reaction components except for the reverse transcriptase to control for contaminating DNA. cDNA was synthesised by carrying out 4 cycles of 20°C for 8min, 25°C for 8min and 37°C for 30min followed by heat inactivation of the enzyme at 70°C for 15min.

Quantitative Reverse Transcription-PCR (qRT-PCR)

qRT-PCR was used to determine the expression levels of particular genes. cDNA (1-2µl, synthesised as outlined above) from a sample of interest was added to each qRT-PCR reaction and additional reactions were carried out with the (-) RT controls. qRT-PCR was carried out using Quantace Sensimix as described above and gene expression was expressed relative to that of a control gene (usually *Eef1A1*). Three PCR replicates were carried out for each biological sample and use to calculate the standard deviation.

2.1.4 Protein extraction, Western blotting and chromatin immunoprecipitation

Protein extraction

To prepare protein extracts for SDS-PAGE and Western blot, cells were harvested, washed in PBS and counted. They were then resuspended in Laemmli buffer (100µl per $1x10^6$ cells) and stored at -20°C until needed. Immediately prior to use, SDS-loading buffer was added, the samples boiled for 10min and snap chilled on ice.

SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to separate proteins based on size and was employed for analysis of purified recombinant protein and Western blotting. The Mini-PROTEAN 3 system (Bio-Rad) was utilized for SDS-PAGE. Stacking gels contained 5% acrylamide and resolving gels contained 8 to 15% acrylamide depending on the molecular weight of the proteins of interest. Protein samples (prepared as above) were loaded into the wells of the stacking gel alongside pre-stained protein ladder (Fermentas). Electrophoresis was carried out at 110V in Tris-glycine buffer.

Coomassie staining

To visualise proteins on an SDS-PAGE gel, gels were stained with Coomassie Blue solution for 20min with agitation. Gels were de-stained in dH_2O by boiling for 3min. The dH_2O was then replaced and the procedure repeated twice more until the last traces of dye had been removed. Gels were scanned and dried between two sheets of cellulose film (Bio-Rad).

Western Blotting

For Western blot analysis protein from SDS-PAGE gels was transferred to nitrocellulose membrane (Protran, Schleier & Schull) or PVDF membrane (Bio-Rad). PVDF membrane was used for analysis of histone proteins due to its smaller

pore size while nitrocellulose was used when blotting for MBD2. PVDF membrane was pre-activated prior to use by soaking in 100% methanol. Semi-dry transfer was carried out in transfer buffer for 60min at 250mA using the Trans-Blot SD system (Bio-Rad). Membranes were then Ponceau stained to verify that protein had been transferred, de-stained in dH_2O and blocked in 5% milk solution for 2 hours. Primary antibody diluted in 1% milk was then added (for antibody information see Table 2.2.7) and the membrane incubated overnight at 4°C with agitation. The membrane was washed in TBS-T three times for 10min before addition of peroxidise-conjugated secondary antibody. Incubation with secondary antibody was carried out for one hour at R/T, the membrane washed in TBS-T as before and ECL solution added for 2min. Signal was then captured by exposing the membrane to X-ray film. Western blots were quantified using Scion Image software.

2.1.5 Chromatin immunoprecipitation (ChIP) and cross-linked IP

Cells were harvested, washed and cross-linked in 1% formaldehyde for 10min at R/T. Cross-linking was stopped by adding 125mM glycine. The cells were then pelleted by centrifugation at 300g for 4min at 4°C and washed in cold PBS containing protease inhibitors. 4-6x10^6 cells were used per ChIP reaction (4x10^6 for histone modifications and 6x10^6 for MBD2). Cells were resuspended in 150µl lysis buffer per reaction and incubated on ice for 10min. Lysates were sonicated to give an average DNA fragment size of 250bp using a Bransen digital sonifier (25% power for 3min – 2.5sec, 5sec off) and then centrifuged at 12,000g for 10min at 4°C. The sonicated lysates were diluted 1:10 in IP-dilution buffer and chromatin pre-cleared by incubating with protein A (rabbit antibodies) or protein G (sheep antibodies) sepharose beads (200µl slurry for up to five reactions) for 2 hours at 4°C. Chromatin was then filtered using a 0.8µm syringe filter. An input sample (5%) was retained and 1300µl of pre-cleared chromatin was used per ChIP reaction. The appropriate antibody was added to each sample (see Table 2.2.7) and incubated overnight at 4°C with rotation. Protein A or protein G beads (100µl slurry) were then added for 2 hours at 4°C to capture antibody-protein complexes. Beads were washed as follows:

1x TSE I, 4x TSE II, 1x LiCl buffer, 3x TE. Washes were carried out for 4min at R/T with rotation and all buffers were kept at 4°C.

For cross-linked IP, protein was recovered by boiling beads in SDS-loading buffer for 10min and then analysed by Western blot.

To recover DNA, ChIP extraction buffer (200µl) was added to the beads and incubated at 55°C for 15min followed by 15min on a shaking platform at R/T. Beads were removed by passing the mixture through a QIA-Shredder column (Qiagen). Cross-links were reversed by treating input and ChIP samples with 300mM NaCl overnight at 65°C. DNA was purified by proteinase K and RNase A treatment at 37°C for 1 hour, phenol:chloroform extraction and ethanol precipitation. Typically DNA was resuspended in 100µl (histone modification ChIP) or 50µl (MBD2 ChIP) 0.1x TE and analysed using qPCR (2µl of DNA per reaction). For ChIP-seq and ChIP-chip analysis DNA was resuspended in 25µl dH_2O.

FLAG ChIP and cross-linked IP were carried out as described above except that pre-cleared chromatin extracts were incubated with α-FLAG M2 agarose beads (20µl of a 50% slurry; Sigma) for 2 hours at 4°C and then washed as usual.

2.1.6 Recombinant protein production

Plasmid constructs

The pet30b-MBD construct was a gift from R. Illingworth and codes for the MBD of human MeCP2 (amino acids 77-167) tagged with 6x histidine residues.

The pGex-MBD2 construct was a gift from H.H. Ng and codes for glutathione-S-transferase (GST)-tagged human MBD2 (amino acids 1-151).

Bacterial expression of recombinant proteins

The BL21 *E. coli* strain was used for over-expression of target genes under the control of a T7 promoter regulated by the lacO system. BL21 cells and cloning vectors encode the lac repressor protein which prevents gene expression. Addition of IPTG to the culture media prevents binding of the lac repressor and allows expression of recombinant protein.

An overnight culture of BL21 cells transformed with the desired construct was diluted 1:100 into LB media and the appropriate antibiotics. The culture was incubated at 37°C until reaching an OD_{600nm} of 0.45-0.6. 1mM IPTG was added to the media to induce recombinant protein expression and cultures were incubated for a further 2-3 hours.

Preparation of bacterial whole cell lysate

Bacterial cells were centrifuged at 4200rpm for 30min at 4°C and washed with cold PBS. The bacterial pellets were resuspended in cold bacterial lysis buffer containing 1mg/ml lysozyme and incubated on ice for 30min. The lysates were sonicated (Branson sonifier 250) at a duty cycle of 30% and an output setting of 5 for 5min on ice. Lysates were then centrifuged at 17,000g for 40min at 4°C and the supernatant retained.

Nickel purification of MBD protein for MAP experiments

All stages of this protocol were carried out at 4°C unless otherwise stated. Nickel sepharose beads (made in-house) were washed with two volumes of N10 buffer and added to the bacterial lysate (0.5ml of beads per litre of initial bacterial culture). The mixture was incubated for 2 hours and then pelleted at 500g for 5min. Beads were washed twice with 10 bed volumes of N20 buffer and then transferred to a disposable chromatography column (GE Healthcare) and washed once more. His-tagged protein was eluted from the column with imidazole as follows. One bed volume of N250

buffer was added to the beads, equilibrated for 10min and then allowed to flow from the column. This process was repeated five times and fractions containing appreciable quantities of purified protein were pooled and dialyzed overnight in N20 buffer to remove imidazole.

Note: MBD protein production was primarily carried out by U. Gruenewald-Schneider.

Glutathione purification of MBD2-GST for antibody affinity purification

Purification of GST-tagged MBD2 protein was carried out in a similar manner to the nickel purification outlined above except that glutathione-sepharose beads (GE Healthcare) were used for purification. Equilibration and washing steps were carried out using GST purification buffer and elution was carried out with GST elution buffer containing reduced glutathione. Purified protein was dialysed into GST dialysis buffer overnight prior to use for antibody affinity purification.

2.1.7 Antibody purification

All solutions were kept at 4°C and all procedures were carried out at 4°C.

Affinity purification

As MBD2 has a high isoelectric point (pI ~ 10) Affi-Gel 10 (Bio-Rad) was selected as a matrix to which to bind recombinant MBD2. Affi-Gel (600µl) was washed in coupling buffer three times. GST-MBD2 (amino acids 1-151) was then added to give a 1:1 slurry with the Affi-Gel beads and a final protein concentration of 1.3mg/ml. Protein was coupled to Affi-Gel for 1 hour at 4°C and the coupling reaction was stopped by adding 50mM ethanolamine pH 7.5. MBD2-coupled beads were then washed with five bed volumes of the following solutions: 0.1M $NaHCO_3$, 1M $NaHCO_3$, dH_2O, 0.2M glycine, 150mM NaCl and TBS. S923 serum (2ml) was then

diluted 1:1 in TBS+0.01% azide and bound to the beads for 2 hours at 4°C with rotation. Beads were washed as follows: 1x TBS, 2x column wash buffer, 4x TBS and then loaded onto a disposable chromatography column (GE Healthcare). Antibody was eluted using low pH buffer and immediately neutralised in 1/10 volume 1M Tris-HCl pH 8. Antibody was dialysed against TBS+0.01% azide and antibody concentration assessed by carrying out SDS-PAGE alongside BSA standards of known protein concentration. Glycerol (50%) was added and the antibody stored at -20°C in aliquots.

Protein G purification

Protein G sepharose beads were washed twice in 100mM Tris-HCl pH 8. The pH of the S923 serum was adjusted by adding 1/10 volume of 1M Tris-HCl pH 8. Serum (2ml), protein G sepharose (2ml) and 100mM Tris-HCl pH 8 (36ml) were mixed and incubated at 4°C for 2 hours with rotation. The beads were then washed three times with 10 bed volumes of binding buffer, loaded onto a disposable chromatography column and washed three more times with 100mM Tris-HCl pH 8 and then three times with 10mM Tris-HCl. Antibody was eluted with 1M glycine pH 2.7 and neutralized with 1/10 volume 1M Tris-HCl pH 8. Antibody was dialysed against TBS+ 0.01% azide, 50% glycerol added and the antibody was stored at -20°C in aliquots.

2.1.8 MAP chromatography

Preparation of MBD chromatography columns

MBD columns were prepared in batches in order to minimize variation between different columns. Recombinant MBD protein was bound to nickel charged sepharose (50mg of protein per 0.5ml beads) in N20 buffer at 4°C for 2 hours. The beads were washed with 10 bed volumes of wash buffer 1, 10 bed volumes of wash buffer 2 and then 10 bed volumes of wash buffer 1 again. The beads were packed onto a 0.5ml chromatography column (Tricorn 5/20; GE Healthcare) at a constant

1ml/min linear flow rate. Once assembled, the chromatography column was equilibrated with a 200ml linear salt gradient (0.1-1M NaCl; gradient formed between column buffers A and B) at a 1ml/min flow rate on an AKTA purifier (GE Healthcare). This equilibration served to pack the beads and remove any bacterial DNA from the matrix.

MAP Chromatography

Chromatography steps were carried out on an AKTA purifier (GE Healthcare) at a linear flow rate of 1ml/min and a maximum back pressure of 0.3MPa. Column buffers A and B were mixed to get appropriate salt concentrations. Fragmented DNA (25µg) was bound to the column in 10ml 0.1M NaCl (90% buffer A and 10% buffer B). The salt concentration was then increased to 0.75M NaCl and non-specifically bound DNA eluted. The column was washed (18 column volumes) with 0.75M NaCl before increasing the salt concentration to 1M NaCl and eluting specifically bound DNA. 3ml fractions were collected and 300µl of each fraction was ethanol precipitated, resuspended in 40µl TE buffer and used for PCR analysis (2µl per PCR; for primer sequences see Table 2.2.1). Fractions containing methylated CGIs (i.e. those eluted between 0.75M-1M NaCl) were pooled, concentrated using a Amicon Ultra-15 centrifugal filter unit (30kDa cut off; Millipore) and subject to a second round of purification. After PCR analysis, methylated fractions from the second purification were pooled and concentrated. DNA was then precipitated and passed onto the Wellcome Trust Sanger Institute for Illumina library production and sequencing. Two biological replicates were carried out for each cell type of interest.

CAP chromatography

CAP-seq was carried out for CD4 cells using a CxxC chromatography column gifted by R. Illingworth. This column consists of recombinant CxxC3 domain from MBD1 bound to a sepharose matrix and was prepared in a similar manner to the MBD columns described above. CAP chromatography was carried out in a similar way to MAP except that non-specifically bound DNA was eluted at 0.1-0.6M NaCl, the

wash was carried out at 0.6M NaCl and specifically bound unmethylated CGIs were eluted over 0.6-1M NaCl. Only one round of CAP was necessary to sufficiently enrich unmethylated CGIs prior to sequencing.

2.1.9 Illumina library production and sequencing

Illumina library production and sequencing were carried out by the Wellcome Trust Sanger Institute, Hinxton, Cambridge. The project was co-ordinated by K. Auger and mapping done by R. Andrews. To produce illumine libraries for sequencing, ligated DNA was amplified by 10-12 cycles of PCR with primers complementary to the adaptor sequences using Phusion DNA polymerase mastermix (Finnzymes) and the DNA was purified using QIAquick PCR Purification columns (Qiagen). The purified DNA was attached to an Illumina flow cell for cluster generation. Libraries were sequenced on the Illumina Genome Analyzer following the standard protocol to generate single-end 37bp reads. Reads were mapped to the mouse genome (mm9 – NCBI37; repeat-masked) using Maq software (http://maq.sourceforge.net/) and reads with a mapping ≥ 30 were retained. Replicate samples were combined to give two lanes of sequencing for each sample of interest.

2.1.10 Bioinformatic analysis (MAP-seq and ChIP-seq)

Sequencing data were received as .wig files and analysed using custom bioinformatics tools developed in conjunction with R. Illingworth and S. Webb. Tools were accessed via an in-house version of the Galaxy application (Blankenberg et al., 2007; Taylor et al., 2007) which can be found at http://bifx-core.bio.ed.ac.uk:8080/. Raw sequencing data were normalised to the average read number across all samples of a given purification type in order to account for variable sequence depth. Peakfinding was then carried out to identify regions of prominent enrichment using the following parameters; read height (H), length in bp (L) and gap permitted in the length parameter (G). These were adjusted for each purification type so that regions of known methylation or histone modification status were isolated. Details of parameters used for the various purification types (e.g. MAP, ChIP-seq) can be found in the relevant results sections.

To identify regions showing differential enrichment for a particular modification, peaks for each sample were combined to give a set of regions of interest over which a sliding window analysis was performed. For each region the average number of hits per base was calculated in 100bp windows with a 20bp slide. Values were then compared between samples giving a ratio for each window. If both windows being compared contained less than 4 reads this ratio was set to 1 in order to remove spurious hits attributed to small fluctuations at low read depth. Differentially enriched regions were identified according to parameters listed in the results sections but in general were defined as those containing 9 out of 13 contiguous windows with a log2 ratio ≥1.7, 1.8 or 2 (depending on the purification) between samples. The X and Y chromosomes were excluded from analyses in order to avoid spurious results due to fluctuations in the proportion of male and female DNA in samples.

Peaks and differentially modified regions were examined for association with CGIs, RefSeq genes, TSSs and other genomic features using standard tools available on Galaxy. An association was defined as direct overlap with a particular feature by at least 1bp. The CGIs examined were from the mouse CGI set characterised by Illingworth and colleagues (2010) and RefSeq annotated genes were used as the gene set. Brain methylation data was obtained from MAP-seq carried out on cerebellum DNA from wild-type C57/BL6 mice (Skene et al., 2010) and was analysed in parallel with the other MAP-seq samples. Brain H3K4me3 data was from Thomson et al. (2010) and was analysed in the same manner as the immune cell H3K4me3 samples. ES cell Pol II ChIP-seq data was from Illingworth et al. (2010) and CAGE data was from a study by Faulkner et al. (2009). Pol II peaks, as defined by Illingworth and colleagues, or regions with more than five CAGE tags were taken as evidence of a TSS.

2.1.11 Microarray procedures and analysis

NimbleGen expression arrays (oesophagus expression)

Oesophagus RNA samples were DNase treated and then purified and concentrated using an RNAeasy column (Qiagen). Total RNA (20µg) was sent to NimbleGen for hybridisation to whole genome mouse expression arrays (mm8_60mer_expr). Three biological replicates were carried out for each condition of interest (in this case wild-type and *Mbd2-/-* oesophagus). Arrays were RMA normalized and the results analysed using GeneSpring GX8 (Agilent Technologies). Briefly, genes with constitutively low expression were eliminated, genes showing a statistically significant change in gene expression (student's t-test, $p \leq 0.05$) retained and genes showing at least a two-fold change in expression selected as differentially expressed.

Illumina BeadChip arrays (immune cells)

RNA (300ng) was amplified and labelled using the TotalPrep RNA Amplification kit (Ambion) and passed to the Wellcome Trust Clinical Research Facility, Edinburgh for hybridisation to Illumina MouseWG-6 BeadChips. Three biological replicates were carried out for each cell type. Bead level data were summarised using Illumina BeadStudio and data were normalised using the average normalisation method in BeadStudio. This scales each array based on a scaling factor derived from the average intensity across all arrays. Subsequent analysis was carried out using GeneSpring GX10 (Agilent technologies). Genes with signal below the 50^{th} percentile of signal values on a particular array were categorised as "not expressed" and this was consistent with the expression values for silent tissue-specific genes and the background signal level of the array. In order to identify changes in gene expression between cell types, genes with low expression (Flags absent) across all cell types were removed and statistical analysis was carried out to find genes that were differentially expressed between any of the cell types (one-way ANOVA, $p \leq 0.5$). Pair-wise comparisons were performed on this gene set and genes changing at least two-fold between cell types were deemed to be differentially expressed.

Brain expression data was obtained from BeadChip arrays carried out on whole brain from wild-type C57/BL6 mice (J. Guy; unpublished) and were analysed as outlined above.

NimbleGen CpG island plus promoter arrays (MAP verification and ChIP-chip)

DNA labelling

Input DNA and MAP/ChIP DNA were fluorescently labelled prior to microarray hybridisation using the BioPrime labelling kit (Invitrogen). 2.5x random primer mix was added to DNA samples (100ng input, entire MAP/ChIP sample) and made up to a final volume of 130.5µl, denatured at 100°C for 10min and immediately chilled on ice. On ice, 15µl of dNTP mix (1 mM dCTP, 2 mM dATP, dTTP and dGTP), 1.5µl Cyanine 3 or 5-dCTP (Cy3 for MAP/ChIP and Cy5 for input DNA; GE Healthcare) and 3µl of Klenow (120U; Invitrogen) was added to each reaction. These were incubated at 37°C for 4 hours protected from light. Labelled DNA was purified using the Purelink PCR cleanup kit (Invitrogen) according to manufacturers' instructions and labelled DNA eluted in 100µl elution buffer. Cyanine incorporation was calculated using a Nanodrop spectrophotometer and labelled samples were stored short-term at -20°C.

Hybridisation and washing

Labelled DNA and 100µg of mouse Cot-1 DNA (Invitrogen) were ethanol precipitated and resuspended in 60µl of hybridisation buffer. Samples were denatured at 100°C for 10mins and snap chilled on ice protected from light. A prehybridization step was carried out whereby 60µl hybridisation buffer was incubated at 70°C for 10mins, applied to the microarray (NimbleGen mouse CpG island plus promoter arrays) and incubated in a humidified chamber for 1 hour at

37°C. The array was washed once in MWA buffer for 2min, once in MWB buffer for 2min and rinsed twice in dH_2O. The hybridisation mix was incubated for 1 hour at 37°C and then applied to the microarray. Hybridisation was carried out for 24 hours at 37°C in a humidified chamber protected from light. The array was then washed 4x 2min in 37°C MWA buffer, 3x 2min in 52°C MWB buffer, 2x 2min in R/T MWA buffer and twice in dH_2O. Hybridised microarrays were then dried and scanned using a 2 laser GenePix Autoloader 4200AL (Axon) and processed using GenePix Pro 6.0 (Axon).

Microarray analysis

Microarray analysis was carried out by R. Illingworth and J. de las Heras using the Limma package in the R statistical environment. Cy3 and Cy5 signals were transformed into M values (log2 [Cy5/Cy3]) and quantile normalisation was carried out. Male and female samples (MAP) or wild-type and *Mbd2-/-* samples (ChIP) were then compared to find differences.

2.2 Materials and reagents

All reagents were stored at R/T unless otherwise stated.

2.2.1 Immune cell purification and FACS staining

Hanks media: Hanks Balanced Salt Solution (Sigma) supplemented with 2% (v/v) foetal bovine serum (FBS) and penicillin/streptomycin (P/S; Gibco). Stored at 4°C.

Hanks with DNase: Hanks Balanced Salt Solution (Sigma) supplemented with P/S, 1mM $MgCl_2$, 1mM $CaCl_2$ and 8U/ml DNase (Roche). DNase added immediately prior to use.

Phosphate buffered saline (PBS 1x): 140mM NaCl, 3mM KCl, 2mM KH_2PO_4, 10mM Na_2HPO_4.

T cell media: Iscove's media (Sigma), 10% (v/v) FBS, 1x glutamine (Gibco), 1x β-mercaptoethanol (Gibco).

Ex vivo media: X-Vivo 15 medium (Lonza, BioWhittaker), 1x glutamine (Gibco), 1x β-mercaptoethanol (Gibco).

PMA (phorbol 12-myristate 13-acetate) stock solution: PMA was diluted to 1mg/ml in ethanol and then diluted to a working stock of 100µg/ml in T cell media. Stored at -20°C.

Ionomycin stock solution: Ionomycin was diluted to 1mg/ml in ethanol. Stored at -20°C.

CFSE stock solution: 5mM 5-carboxyfluorescein diacetate succinimidyl ester (CFSE; Molecular Probes) in DMSO at a final concentration of 5mM. Stored at -20°C.

FACS buffer: PBS, 0.5% (v/v) FBS. Stored at 4°C.

4% paraformaldehyde: 4% (w/v) paraformaldehyde in PBS, pH adjusted to 7.3. Stored at -20°C.

Permeabilisation buffer: 0.1% (w/v) saponin (Fluka) and 1% (v/v) FBS in PBS, pH adjusted to 7.0. Stored at 4°C.

2.2.2 DNA manipulation

Genomic DNA extraction buffer: 10mM Tris-HCl pH 8.0, 400mM NaCl, 3mM EDTA pH 8.0, 1% (w/v) SDS.

Proteinase K stock solution: 20mg/ml proteinase K, 100mM EDTA pH 7.5, 2% (w/v) SDS. Stored at −20°C.

RNase A stock solution: 10mg/ml RNase A in dH_2O. Stored at −20°C.

TAE buffer (1x): 40mM Tris-acetate, 1mM EDTA

Orange G loading buffer (6×): 0.2% (w/v) orange G, 12% (w/v) Ficoll, 120mM EDTA pH 8.0, 4.2% (w/v) SDS. Stored at −20°C (long-term) or 4°C (short-term).

Bisulfite solution: 3.8g sodium bisulfite ($NaHSO_3$) was dissolved in 5ml dH_2O and 1.5ml 2M NaOH (protected from light). 110mg hydroquinone was dissolved in 1ml dH_2O at 55°C for ten minutes and added to the sodium bisulfite solution. Bisulfite solution was prepared immediately prior to use.

TE buffer: 10mM Tris HCl pH 7.5, 1mM EDTA.

DNA Sequencing Buffer (2.5x): 20mM Tris-HCl pH8.0, 5mM $MgCl_2$.

Polyethylene glycol (PEG) 6000: 30% (w/v) PEG 6000 in dH_2O. Stored in aliquots at -20°C.

2.2.3 RNA manipulation

TBE buffer (1x): 45mM Tris-borate, 1mM EDTA.

DEPC-dH_2O: 1ml diethylpyrocarbonate (DEPC) added to 1L dH_2O, treated overnight and autoclaved to inactivate DEPC.

2.2.4 Protein extraction, Western blotting and chromatin immunoprecipitation

Laemmli buffer: 60mM Tris pH 6.8, 100mM DTT, 10% (v/v) glycerol. Stored at 4°C.

SDS-PAGE loading buffer (6x): 300mM Tris-HCl pH 6.8, 12% (w/v) SDS, 20% (v/v) glycerol, 0.2% (w/v) bromophenol blue. Stored at −20°C.

SDS-PAGE stacking gel: 5% (w/v) 29:1 acrylamide:bis-acrylamide (Bio-Rad), 0.1% (w/v) SDS, 129mM Tris-HCl pH 6.8, 0.1% (v/v) TEMED, 0.1% (w/v) ammonium persulfate.

SDS-PAGE resolving gel: 8-15% (w/v) 29:1 acrylamide:bis-acrylamide (Bio-Rad), 0.1% (w/v) SDS, 390mM Tris-HCl pH8.8, 0.08% (v/v) TEMED, 0.1% (w/v) ammonium persulfate.

Tris-glycine electrophoresis buffer: 25mM Tris, 250mM glycine, 0.1% (w/v) SDS.

Coomassie Blue staining solution: 50% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.1% (w/v) Coomassie Brilliant Blue R-250. Filtered prior to use.

Transfer Buffer: 25mM Tris, 250mM glycine, 20% (v/v) methanol (for nitrocellulose membranes only).

TBS-T (1x): 50mM Tris-HCl pH 7.4, 150mM NaCl, 0.1% (v/v) s 20.

Milk solution: 1% or 5% (w/v) dried milk powder (Marvel) in TBS-T. Stored at 4°C.

ECL: Equal volumes of solution I [2.5mM (w/v) luminol, 396μM ρ-coumaric acid, 100mM Tris-HCl pH 8.5] and solution II [0.2% (v/v) hydrogen peroxide, 100mM Tris-HCl pH 8.5] were mixed immediately prior to use. Solutions were stored at 4°C.

2.2.5 ChIP

Protease inhibitors (Protease inhibitors, complete EDTA-free; Roche) were added to lysis and IP dilution buffers before use. All ChIP solutions were kept at 4°C.

Lysis buffer: 2% (w/v) SDS, 20mM EDTA, 100mM Tris-HCl pH 8.

IP dilution buffer: 1% (v/v) Triton-X 100, 2mM EDTA, 150mM NaCl, 20mM Tris-HCl pH 8.

Protein A/G sepharose: Protein A or G sepharose (GE Healthcare) was washed three times in IP dilution buffer and blocked by addition of one-fifth the bed volume of tRNA (stock solution 20mg/ml) and BSA (stock solution 10mg/ml). IP dilution buffer was added to give a 50% slurry and the mixture incubated for 2 hours at 4°C with rotation. Beads were then washed into IP dilution buffer and stored as 50% slurry at 4°C until needed.

TSE I: 0.1% (w/v) SDS, 1% (v/v) Triton-X 100, 2mM EDTA, 150mM NaCl, 20mM Tris-HCl pH 8.

TSE II: 0.1% (w/v) SDS, 1% (v/v) Triton-X 100, 2mM EDTA, 500mM NaCl, 20mM Tris-HCl pH 8.

LiCl buffer: 0.25 M LiCl, 1% (v/v) NP-40, 1% deoxycholate (w/v), 1mM EDTA, 10mM Tris-HCl pH8.

ChIP extraction buffer: 1% (w/v) SDS, 0.1M $NaHCO_3$.

2.2.6 Bacterial reagents (for cloning and protein expression)

LB medium: Bacto tryptone (10g/l; Difco), Bacto yeast extract (5g/l Difco), NaCl (10g/l), pH adjusted to 7.0 with NaOH. Bacto agar (20g/l; Difco) was added if making LB agar. Solutions were autoclaved prior to use. LB agar plates were stored inverted at 4°C and LB broth was stored at R/T.

Ampicillin stock solution: 50mg/ml ampicillin in dH_2O. Filter sterilised (0.2µm filter) and stored at −20°C. Added to LB medium at a final concentration of 50µg/ml.

Kanamycin stock solution: 50mg/ml kanamycin in dH_2O. Filter sterilised (0.2µm filter) and stored at −20°C. Added to LB medium at a final concentration of 50µg/ml.

Choramphenicol stock solution: 34mg/ml chloramphenicol in 96% ethanol. Stored at −20°C protected from light. Added to LB medium at a final concentration of 34µg/ml.

Blue/white selection plates: LB plates containing the appropriate antibiotic were spread with 40µl 100mM IPTG and 40µl X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; 40mg/ml) and dried at 37°C. Prepared on day of use.

IPTG stock solution: 1M IPTG (isopropyl-β-D-thiogalactoside) in dH_2O. Filter sterilised (0.2µm filter) and stored at −20°C.

2.2.7 Recombinant protein purification

All solutions for recombinant protein purification were kept at 4°C.

N10 buffer: 50mM sodium phosphate buffer pH8.0, 300mM NaCl, 10% (v/v) glycerol, 10mM imidazole, 15mM β-mercaptoethanol and 0.5mM PMSF**.** β-mercaptoethanol and PMSF were added immediately prior to use.

N20 buffer: As per N10 buffer but with 20mM imidazole.

N250 buffer: As per N10 buffer but with 250mM imidazole.

GST purification buffer: 500mM KCl, 50mM Hepes pH 8.0, 1mM DTT, 0.1% (v/v) Triton X-100.

GST elution buffer: As above with 100mM reduced glutathione, pH adjusted to 8.0.

GST dialysis buffer: 50mM Hepes pH 7.6, 100mM KCl, 30% (v/v) glycerol.

2.2.8 Antibody purification

All solutions for antibody purification were kept at 4°C.

Coupling buffer: 50mM Hepes pH 7.6, 200mM KCl.

Column wash buffer: 500mM NaCl, 0.2% (v/v) Triton X-100, 20mM Tris pH 7.5.

TBS (1x): 50mM Tris-HCl pH 7.4, 150mM NaCl.

Low pH buffer: 2M glycine pH 2, 150mM NaCl.

2.2.9 MAP chromatography

For chromatography all solutions were filtered prior to use (0.2µm filter) and stored at 4°C.

Wash buffer 1: 20mM Hepes pH 7.9, 10% (v/v) Glycerol, 0.1% (v/v) Triton X-100, 100mM NaCl, 0.5mM PMSF and 10mM β-mercaptoethanol. β-mercaptoethanol and PMSF were added immediately prior to use.

Wash buffer 2: As for buffer 1 + 10mM imidazole.

Column buffer A: 20mM Hepes pH 7.9, 10% (v/v) Glycerol, 0.1% (v/v) TritonX-100, 0.5mM PMSF and 10mM β -mercaptoethanol. β-mercaptoethanol and PMSF were added immediately prior to use.

Column buffer B: As for buffer A + 1M NaCl.

2.2.10 Microarray Reagents

SSC (20x): 3M NaCl, 0.3M Na citrate, pH adjusted to 7.0.

Hybridisation buffer: 2x SSC, 50% (v/v) deionised formamide, 10mM Tris-HCl pH 7.5, 0.1% (v/v) Tween 20. Filtered (0.2µm) prior to use.

MWA buffer (Microarray Wash A): 1x PBS, 0.05% (v/v) Tween 20. Filtered (0.2µm) prior to use.

MWB buffer (Microarray Wash B): 1x SSC. Filtered (0.2µm) prior to use.

2.2.11 Mouse lines

Wild-type mice:

Wild-type C57/BL6 male and female mice aged 6-12 weeks were used for all experiments in Chapter 3. Wild-type BALB/c male and female mice were used in the Treg experiments (aged 6-12 weeks) and oesophagus experiments (aged 85 days on average) in Chapter 4.

Mbd2-/- mice:

Mbd2-/- C57/BL6 mice were produced from a heterozygous cross and, where possible, wild-type littermates were used as controls. *Mbd2-/-* C57/BL6 male and female mice aged 6-12 weeks were used for all experiments in Chapter 4 except where indicated. *Mbd2-/-* BALB/c mice were maintained as a homozygous line and used for the Treg and oesophagus experiments.

All mouse work was carried out in accordance with Home Office regulations.

2.2.12 Oligonucleotides

Table 2.2.1 MAP PCR primers

Name	Genomic location	Sequence forward	Sequence reverse
meCGI1	chr1:188786630 - 188793725	GGATGCCTACAGCAGCTTCG	CGACCAGTCATCCTCGTCTG
Bulk1	chr1: 188786630 - 188793725	CCTAGCATGTTTAAAGCCCTTG	GCCCTGTCTATAGCTGTATCAT
meCGI2	chr13: 8929225 - 8933895	CGTTTAAAACATTGTCTATACAGC	CGTCTCTCCATCTCTCCGTC
Bulk2	chr13: 8929225 - 8933895	GCTACTTAGAGCATGTTCTCAG	CCACTTCTACATTTTCCCTGAC
meCGI6	chr17: 26565389 - 26572892	GCTTCCATGCGCAGGCTAAG	GGATATCCTGTGCGCTCATGA
Bulk6	chr:17: 26565389 - 26572892	GGACCTTGGTCTTTGGGCAC	CCTAAATGCCAAGCCAAGAGG

Table 2.2.2 Bisulfite Primers

Name	Genomic location	Sequence forward	Sequence reverse
Npb (11_12)	chr11: 120469741 - 120470220	GTTTGTTTTTGTTATAGTTTTATGGT	CCTACTTACACTCACAAAACTCCTTA
Tnfaip2 (12_11)	chr12: 112683701 - 112684060	TTGGGTTTTAGTGGAGGAGTTTA	ACAAAACCTCCACCTTACTCTACC
Kcnn4 (7_25)	chr7: 25161961 - 25162900	TTGTTGTTTTTGGTTATGTTTTTG	AATATCTCCTCCCTATATTCCAAAC

Gata3 (2_97)	chr2: 9795943 - 9796614	GGGAGTAAAAGATTTTTAGGGGTAG	TAACCACCCTAAATAAAACCTCATC
Sema4a (3_882)	chr3: 88240621 - 88241160	GGAGGAGAGTGAGTATTTTTAGGAGTA	ATTAAATAAACAACCAAAACCAACC

Table 2.2.3 qRT-PCR primers

Name	Sequence forward	Sequence reverse
Eef1A1	TTGGTTCAAGGGATGGAAAG	AGCAAAGGTAACCACCATGC
Kctd17	GAATTTCCTCAGGCATGGAA	GTGTGAGCTCCTCCTCTTGG
Klf2	GCCTGTGGGTTCGCTATAAA	AAGGAATGGTCAGCCACATC
Kcnn4	AAGCACACTCGAAGGAAGGA	CCGTCGATTCTCTTCTCCAG
Gata3	GTGGTCACACTCGGATTCCT	GCAAAAAGGAGGGTTTAGGG
Sema4a	CCCTGACTCTGAGCCTGTTC	AGGCTTGGTAGAGCCAGACA
Il4	TCAACCCCCAGCTAGTTGTC	TGTTCTTCGTTGCTGTGAGG
Ifng	ACTGGCAAAAGGATGGTGAC	TGAGCTCATTGAATGCTTGG
Ccl11 (eotaxin-1)	TCCACAGCGCTTCTATTCCT	CTATGGCTTTCAGGGTGCAT
Gapdh	TACCCCCAATGTGTCCGTCG	CCTGCTTCACCACCTTCTTG
Ppic	TGATTGGCCTCTTTGGAAAC	CCTCCAGTGCCATCTCTAGC
Il4i1	AAGCCCCAGAAGGTGGTAGT	GCTTGTGCAAGATCCTGTGA
Il8ra	CCGTCATGGATGTCTACGTG	CAGCAGCAGGATACCACTGA
Ptpn14	GACTCCTCGCCTGATACTGC	CTGGGTGGACAAAGTGACCT
Socs3	CCTTTGACAAGCGGACTCTC	GCCAGCATAAAAACCCTTCA
Tnfaip2	AAAAAGGACCAGCCCAGATT	TACAGAGCCTCCACCTTGCT

Table 2.2.4 ChIP qPCR primers

Name	Genomic location	Amplicon name	Sequence forward	Sequence reverse

Kcnn4	**chr7: 25155333 - 25170220**	ex2	GTGCTGTACCTGCTCCTGGT	TGGCATGGAAGACCACAATA
		int2	GTCGGGAGAAGTGAAAATGG	ACAATAGTGCCGACCCCTAA
		ex3 1	ATGCTCCTGCGTCTCTACCT	ATGTACAGCTTGGCCACGAA
		ex3 2	ACACAGGGAGGAGACACCAG	TTCTGGGTGGCTCAGAAGTT
		int3	CTCTGGATGGTTCCTGCTTC	GTTGCGCTAGAGCCTGTTTG
		int4	GGCCCACTGCTGATACTCAT	TCAGAAATCCACCTGCCTCT
Gata3	**chr2: 9,778,705 - 9,816,661**	5' 1a	GTCTTTGGCTCCTGCAGTTC	CTTACCCTTTCCCCAAGAGG
		5' 1	CTTAATCCCGGGCCTTTTAG	TTTGACCAACCCATCTCTCC
		ex1	AGCTGTCTGCGAACACTGAG	GCGTAGAGGAGGAGGAGGAG
		int1	TGTTTGGGGGTTTGTTTGTT	AGCCCAGGACTGACTAAGCA
		ex2	GTGGTCACACTCGGATTCCT	GCAAAAAGGAGGGTTTAGGG
		int2 3	TTGTGTGGTCGAGACAGGAG	GCGGCTCTAGGGTAGAGGAT
		int2 2	ACCTTCTGGAAGCAGCTTTG	TTTCAATGGCCCTGATTCTC
		ex3 2	CGGCTTCATCCTCTTCTCTG	GCAGCTGCACCTGATACTTG
		int3 1	TTCCAAGACCAAATGGAAGG	CCCAGTGACCCTCTGTGAGT
		int3 2	CGGCTGGAAACATTTCATTC	TGGCCCTGTAGTACCCTCTG
		ex4	AAGGCAGGGAGTGTGTGAAC	AACTTACCAGCCTTCGCTTG
c-myc	**chr15: 61,809,896 - 61,824,907**	1	CCAGAAGCTTTCCCAGCAAGC	CAGCTCAGCCTTGCTTGCTC
		5	AGGCAATAATAAGCTAATGCTCC	GTGAAGGAGATCTAACAGAAACG
		9	CAGTGCTGAATCGCTGCAGG	CCGATTGCTGACTTGGAGGAG
		10	GGAAGAGAATTTCTATCACCAGC	ACATAGGATGGAGAGCAGAGC
		13	CTCAACCCAAGGACTCTGCC	CCAGGATCAACTTAGCAGTGG
		17	GGTTTCTCTGTGTCTGAGAACC	GTGGTAGGTAGTAAGTGTATCG

Sema4a	**chr3: 88239884 - 88261661**	ex11	TGACAAAGCCTTGACCTTCA	TACATGACCACATGGCTGCT
		int11	TAAAAAGACGGCTGGGAATG	CACTGCTGGATTTGCAGGTA
		ex13	TCTGGAGGCATCTGGAGAGT	CCAGACAGAAGGCTGCAGAG
		int13	GTGAGTGCTGACCCTGAATG	GCTGGAATCTGGGAGTGAGA
		int14	GGAGGAGGTAACCAGGCACT	CCACAGTCCACTTCACATCG
		ex15	TCTGCCAGTGACGTAGATGC	GCAGAAGAGAGTCCCACAGG
		3' 1	CTCTTTGAGCCCCTGTGGTA	ACTCAGGTCAAATGGGCTTG
		3' 2	CCTGGGGCTGATAACAAAGA	TACATGCAGGTCAGGGTGTC
Xist	**chrX: 100,675,536 - 100,685,072**	-3	GACAGCCTTATCCAGTGTCC	GAACAGCGTAAAACTGTAATGG
		-2	GTGGTCTCATTGGTTGGCAC	GGAACATTTTATGTGGATAATCC
		1.21	GTCTTGAGGAGAATCTAGATGC	TTCTATACCAGTTCAGGCTTTGC
		2	GAATTCAACAAGTAAGCAAATCG	GCCAGAGTCATAGTGGATCAC
		2.1	GCACTGTAAGAGACTATGAACG	CGCATGCTTGCAATTCTAACAC
		3	CCTGTACGACCTAAATGTCC	GTATTAGTGTGCGGTGTTGC
		3.1	CTCAGTTTAAGAGCAAAGTCGT	GCTTGGTGGATGGAAATATGG
		4	CAAAAAGTATGGAGGACATGTC	CGTGCAACGGCTTGCTCCAG
		4.1	AGGTCACACACCTGTCTATGC	CCAAGGAGCCATTTTGTGAGG
		5	GTCTCGTTGATTCACGCTGAC	GTTTATTCAGTCTGTGTGCATC
		6	CCTCCATTCCTGTACACTTAAC	CTTAATGTTAAGAATATGCAATG
		7	GCTACTGCTCATAGGTAGGC	CATGATCTTTGGTAGATTGATTTC
		FR2	GTGTCCAAGACGCGGAGATAC	CAAACATGGCTGGAGCAAGCC
		FR3	CGTGATACGGCTATTCTCGAG	CCATTGCTACACACCAGAACA
TFF2	**chr21: 43,771,237 - 43,772,237 (human)**	TFF2A	GATGGAGCCCTGGAGAATG	GAAAAGTGTGGTGGGAAGGAC
		TFF2B	AGCACATCCTACCCCAGGAAG	GTACAGGCCAGCAGATAGCAT

		TFF2C	GGTCCCTGCCCACTCTATCA	AATCTCCCCTCTGTTTGTCCAA
Il4	**chr11: 53,416,484 - 53,444,171**	CNS-1	GGCAGCTGGTCAGTGGGTACC AGGTCTG	CAGTTGATCTGGGAAAGTTCGTTGC
		2IE	GAGCTGTCCTCTGGCTGGAGT CTG	GCAGGCAGTCTGGAGACCACTGG
Major satellite	**repetitive**	Maj Sat	GGCGAGAAAACTGAAAATCACG	AGGTCCTTCAGTGTGCATTTC
Fcgr3	**chr1: 172994070 - 172996091**	FR2	AAGGCTGTCTTGCGTCAAAT	GAACGGCAGAAGATGAAAGC
Eef2k	**chr7: 127985096 - 127987217**	FR2	GTCTACCACCATGCCCAAGT	AGGAGGCAGACACAGAAGGA
Nkx6-2	**chr7: 146766819 - 146769992**	FR1	AGGGGGCAAATTTACAGTCC	GGATGGACGCACTGATAACC
ActB	**chr5: 143,665,420 - 143,669,404**	4	CCTAATACGGCTTTTAACACCC	CCTGAGGATCACTCAGAACGG

Table 2.2.5 Other primers

Name	**Sequence forward**	**Sequence reverse**
Illumina paired-end adaptors	5' Phos-GATCGGAAGAGCGGTTCAGCAGGAAT GCCGAG	ACACTCTTTCCCTACACGACGCTCTTCC GATC*T (*phosphorothioate)
T7/M13 R	TAATACGACTCACTATAGGG	CAGGAAACAGCTATGAC

2.2.13 Antibodies

Table 2.2.6 Antibodies for immune cell isolation and FACS

Antibody	Fluorophore	Source	Cat no.	Concentration
α-CD3	N/A; purified for culture	BD Pharmingen	553057	1.6µg per well
α-CD28	N/A; purified for culture	BD Pharmingen	553294	1.6µg per well
α-IL-4	N/A; purified for culture	BD Pharmingen	554432	5µg/ml
α-IL-12	N/A; purified for culture	Biolegend	505304	5µg/ml
α-FcR	N/A	Gift from A. MacDonald	N/A	10µg/ml
α-CD4	FITC	Biolegend	100405	1:100
α-CD4	PerCP	BD Pharmingen	553052	1:100
α-CD44	PE	Biolegend	103007	1:250
α-CD62L	APC	Biolegend	104411	1:100
α-CD11c	Alexa Fluor 647	eBioscience	51-0114-82	1:200
α-PDCA	PE	Miltenyi Biotec	130-091-962	1:25
α-CD19	PE	BD Pharmingen	557399	1:200
α-Ifng	APC	Biolegend	505809	1:100
α-IL-4	PE	BD Pharmingen	554435	1:100
α-Foxp3	APC	eBioscience	17-5773	1:200
α-CD103	PE	BD Pharmingen	557495	1:200

Table 2.2.7 Antibodies for ChIP and Western blot

Antibody	Application	Source	Cat no./Ref	Concentration
α-H3K4me3	ChIP	Millipore	07-473	5µg/IP
α-H3K27me3	ChIP	Millipore	07-499	5µg/IP
α-H3Ac	ChIP, Western blot	Millipore	06-599	10µg/IP, 1:10000 for Western

α-H3	ChIP, Western blot	Abcam	ab1791	10µg/IP, 1:20000 for Western
α-MBD2, S923	ChIP	Self-made, sheep polyclonal. IgG purified where indicated	Ng et al. 1999	15µl/IP (serum), 25µg/IP (purified)
Preimmune serum	ChIP	Sheep, isolated prior to S923 production.	Gift from H. H. Ng.	10µl/IP (or equivalent to antibody concentration)
α-MBD2, R593	ChIP, Western blot	Self-made, rabbit polyclonal	Ng et al. 1999	10µl /IP, 1:1000 for Western
α-MBD2	ChIP	Abcam	ab3754	20µl/IP
α-MBD2	ChIP	Imgenex	IMG-147	10µl/IP
α-FLAG M2 agarose	ChIP	Sigma	A2220	10µl beads/IP
Rabbit control IgG	ChIP	Abcam	ab46540	10µg/IP (or equivalent to antibody concentration)
Anti-goat/sheep IgG peroxidase conjugate	Western Blot	Sigma	A9452	1:2500
Anti-rabbit IgG peroxidase conjugate	Western Blot	GE Healthcare	NA934	1:10000

Chapter 3 Differential CGI methylation in an immune system model

3.1 Introduction

The functional implications of cell-type specific CGI methylation are still unclear. The well-characterised changes in DNA methylation that take place during T cell differentiation make immune cells an attractive system in which to study this phenomenon and its relationship to gene expression. This investigation will be facilitated by the fact that a comprehensive CGI set for mouse now exists which includes non-promoter CGIs. In addition, the advent of high throughput sequencing technology means that DNA methylation in the genome can be more comprehensively investigated as analysis is no longer limited to the range of features included on a microarray.

3.1.1 Differential CGI methylation and its relationship to gene expression

Differences in CGI methylation between somatic cells and tissues have been well-documented although these are small in number [section 1.5 and (Eckhardt et al., 2006; Illingworth et al., 2008; Meissner et al., 2008; Schilling and Rehli, 2007)]. Less clear is their functional significance and how they relate to gene expression. Differential methylation in different cells and tissues is not always associated with differential gene expression (Illingworth et al., 2008; Schilling and Rehli, 2007). This could be due to the fact that differential methylation is a relic of previous development (i.e. differential expression earlier in the developmental lineage), an indication of transcription potential or stochastically acquired as a result of a transcriptionally silent state. The situation is further complicated by the existence of non-promoter-associated "orphan" CGIs many of which seem to be unannotated promoters (Illingworth et al., 2010; Maunakea et al., 2010). A large proportion of these reside in gene bodies where they have the potential to affect expression of the

associated gene. In addition, analysis in complex tissues such as brain and blood is complicated by the presence of multiple cell types which may have distinct methylation and gene expression profiles. Illingworth and colleagues found composite methylation patterns at many CGIs in human blood, brain, muscle and spleen and in some cases this was shown to be due to distinct methylation patterns in different cell types (Illingworth et al., 2008). To avoid this complication, homogenous cell cultures could be used. However, methylation profiling of cultured cell lines is not ideal because cultured cells frequently have abnormal methylation patterns (Jones et al., 1990; Meissner et al., 2008; Smiraglia et al., 2001). To address these concerns I will use pure primary cells from the mouse hematopoietic lineage for methylation analysis. Some of these cell types have previously reported DNA methylation differences although many occur outside CGIs (section 1.6). The extent of CGI methylation differences between these developmentally related cell types and how this relates to gene expression will be investigated. Such analysis has been facilitated by the coupling of affinity chromatography methods for isolating unmethylated and methylated CGIs to next generation sequencing technology.

3.1.2 Existence of a comprehensive CGI set

Accurate identification of CGIs is of paramount importance for elucidating the extent and function of CGI methylation. As discussed in section 1.5, algorithms for predicting CGI position on the basis of sequence composition show high variability and fail to take into account methylation status. Early studies aimed to biochemically purify human CGIs using an affinity column containing the MBD from MeCP2 which binds methylated CpG-rich sequence. Methylated CpG-rich sequences bind specifically to this column and thus can be separated from unmethylated CGIs and bulk methylated genomic DNA which elute from the column at low salt concentrations. The fractions containing unmethylated CGIs and bulk genome were then artificially methylated and reapplied to the column. Previously unmethylated CGIs were converted to methylated CGIs and thus bound specifically to the MBD column. These could be now be separated from CpG-poor bulk DNA and cloned giving a human CGI library (Cross et al., 1994). This technique for identifying CGIs

was superseded by CxxC affinity purification (CAP) which takes advantage of the CxxC protein domain which has specific affinity for unmethylated CpGs. In CAP, genomic DNA is applied to a CxxC domain-containing affinity column and unmethylated CpG-rich sequences (CGIs) specifically bind and can be eluted at high salt concentrations. An initial CGI library was generated from human blood using this method and sequenced using conventional Sanger sequencing but a high rate of sequence failure meant that all CGIs were not included (Illingworth et al., 2008). In addition, any CGIs constitutively methylated in blood were excluded because of the specific affinity of the CxxC column for unmethylated CpGs. Our lab has recently generated comprehensive CGI sets for both human and mouse by coupling CAP to high throughput sequencing (CAP-seq). This was carried out in a range of tissues including sperm where CGIs are hypomethylated (Weber et al., 2007) so that somatically methylated CGIs would not be excluded. These sets are thought to represent the full CGI complement for human and mouse and contain 25,495 and 23,021 CGIs respectively. Significantly, many of these CGIs do not associate with annotated gene promoters. In both human and mouse approximately half of CGIs occur at annotated transcription start sites (TSS) with the remainder located between genes (intergenic) or within gene bodies (intragenic) (Illingworth et al., 2010). The generation of a mouse CGI set is of particular advantage since mouse is a genetically tractable and widely used model organism, mouse cells can be acquired more easily and in greater numbers than human cells and are not subject to the same ethical constraints. In addition, mouse CGIs are less CpG-dense than human CGIs so many are missed by prediction algorithms (Illingworth et al., 2010). Now that the location and sequence composition of mouse CGIs is known their methylation status can be comprehensively investigated.

3.1.3 Methods of determining DNA methylation state

A number of methods can been used to assess DNA methylation. These include bisulfite conversion of DNA, methyl-sensitive restriction digestion, antibody pull-down of methyl-cytosine and affinity purification using a methyl-binding domain. Such techniques can be used to assess the methylation status of candidate loci or be

applied genome-wide through the use of microarrays or next generation sequencing. Each of these techniques differs with respect to resolution, methylation context (CpG or non-CpG) and CpG density observable. Significantly, when using high throughput sequencing as a read-out, the depth of sequence necessary to obtain good coverage over regions of interest is of paramount importance.

Sodium bisulfite treatment of DNA converts unmodified cytosine to uracil but does not affect methyl-cytosine. When bisulfite-treated DNA is amplified uracil is read as thymine and methylation can be assessed by sequencing and examining which cytosines remain unconverted and which have been converted to thymine (Frommer et al., 1992). Bisulfite-sequencing allows detection of methylated cytosine at single base-pair resolution in any sequence context (i.e. CpG and non-CpG), at any region of the genome. Whole-genome bisulfite sequencing (BS-seq) has recently provided valuable information about the extent of non-CpG methylation, the role of gene body methylation and DNA methylation differences between cell types as well as uncovering a potential function for DNA methylation in splicing (Laurent et al., 2010; Lister et al., 2009). However, when using bisulfite sequencing to examine methylation genome-wide, a huge amount of sequencing is necessary in order to obtain adequate coverage over the whole genome (Laurent et al., 2010; Lister et al., 2009). Restriction digestion and size selection that preferentially enriches for CGIs has been used in combination with bisulfite-sequencing to examine methylation at approximately half the CGIs in mouse as well as some other genomic regions and this approach required a much smaller amount of sequencing [a technique know as reduced representation bisulfite sequencing or RRBS (Meissner et al., 2008)].

Restriction enzymes that are sensitive to methylation can be used to assess DNA methylation status at regions containing the relevant restriction site. One technique taking advantage of this is restriction landmark genome scanning (RLGS) which involves digesting DNA with the methyl-sensitive enzyme NotI[iii] and separating

[iii] Methyl-sensitive restriction enzymes do not cut methylated DNA.

restriction products by two-dimensional electrophoresis (Song et al., 2005). The HELP (HpaII tiny fragment enrichment by ligation-mediated PCR) assay digests DNA with the methyl-sensitive enzyme HpaII, and co-hybridises this along with DNA digested by the methyl-insensitive enzyme MspI to a microarray (Khulan et al., 2006). In contrast, CHARM (comprehensive high-throughput array-based relative methylation) analysis is a technique which uses an enzyme, McrBC, which specifically cuts methylated DNA. Unmethylated DNA, resistant to digestion is hybridised to a microarray along with mock-digested DNA (Irizarry et al., 2008). CHARM has recently been used to identify tissue- and cancer-specific methylation occurring at the edges of CGIs (Irizarry et al., 2009). A major limitation when using digestion-based methods to analyse methylation is that they are limited to genomic regions containing the restriction site of interest. For example, only 20% of mouse and human CGIs contain the NotI sites that are used in RLGS and only 3.9% of non-repetitive DNA contains sites for HpaII (Fazzari and Greally, 2004). McrBC cuts more promiscuously in the genome and cleaves all methylated CGIs and half of all methylated CpG sites (Irizarry et al., 2008). However, detection of the methylation status of multiple CpG sites is desirable in order to be able to quantitatively assess CGI methylation and this would require several restriction sites in each CGI.

Methylated DNA immunoprecipitation (MeDIP) uses an antibody directed against methyl-cytosine to immunoprecipitate single-stranded methylated DNA and mC in any sequence context can be recognised. MeDIP has been effectively used in combination with microarrays to look at genome-wide methylation in cancer cells (Weber et al., 2005), at promoter methylation in normal cells (Weber et al., 2007) and at methylation of promoters occurring during ES cell differentiation (Mohn et al., 2008). MeDIP is well suited to characterising methylation at CGIs because methylated CGIs are efficiently immunoprecipitated by the mC-antibody due to a high density of mCpG. However, when applying next generation sequencing methods to immunoprecipitated DNA, more sequencing is necessary than for techniques that only enrich for CpG-dense methylation because MeDIP also identifies methylation in the bulk genome. In principle, the methylation status of the

entire genome could be determined by MeDIP-seq; however, to accurately quantify methylation in CpG-poor regions a lot of sequencing would be necessary. Two recent studies coupled MeDIP-seq to high throughput sequencing and reported good coverage of CpG residues. However, both studies carried out a large amount of sequencing and, in addition, artificially extended their reads to the average fragment size of DNA used for immunoprecipitation which increases CpG coverage (Down et al., 2008; Ruike et al., 2010).

Methylated DNA can also be isolated through its affinity for the MBD domain of a methyl-binding protein. The MBD of MeCP2 used as an early means of identifying CGIs in the genome has been utilised to profile CGI methylation, a technique known as MBD affinity purification (MAP). A chromatography column containing recombinant MBD protein is prepared and when genomic DNA is applied, CpG-rich methylated sequences bind with high affinity to the MBD. This methylated CpG-rich fraction has been used to probe human CGI arrays (Illingworth et al., 2008) and, more recently, been sequenced directly to profile CGI methylation in human and mouse [MAP-seq; (Illingworth et al., 2010)]. Other MBD-affinity techniques use an immunoprecipitation-like method which involves incubating genomic DNA with MBD-coupled beads, washing and eluting specifically bound DNA. This can be applied to microarrays or directly sequenced (Gebhard et al., 2006; Rauch et al., 2009; Serre et al., 2009). MBD affinity methods are best suited to the isolation of methylated CGIs and other CpG-rich methylated DNA because the MBD binds clusters of methyl-CpG with high specificity. For a CGI-centric study, MAP provides good coverage of CGIs for a relatively small amount of sequencing (Illingworth et al., 2010). Therefore, I have chosen to use MAP-seq to analyse CGI methylation in immune cells.

3.1.4 Aim

I will use immune cells as a model system in which to investigate the extent of cell type-specific DNA methylation in the immune system and how this relates to gene expression and chromatin modification. This system is particularly suitable as it is

possible to study homogenous primary cell populations which have already been shown to have sites of differential methylation (section 1.6). Although not all of these sites are located at CGIs, CGIs will be focused on in this chapter because they represent a distinct and functionally important fraction of the genome.

3.2 Results: Isolation of immune cell subtypes

3.2.1 The immune cell lineage

Immune cells arise from a self-renewing progenitor, the hematopoietic stem cell (HSC), which gives rise to myeloid and lymphoid progenitor cells that can differentiate into a variety of mature cell types. Cells of the myeloid lineage include granulocytes and macrophages which are involved in destroying and engulfing invading bacteria as well as specialised antigen presenting dendritic cells[iv]. The lymphoid lineage primarily consists of B and T cells. B cells are involved mainly in antibody production as well as in antigen presentation. T cells consist of two subtypes; cytotoxic CD8 T cells which destroy virally infected cells and CD4 T helper cells which coordinate the immune response (Elgert, 2009). T helper cells further specialise into effector cells designed to deal with particular types of infection. Th1 cells are produced in response to intracellular bacteria and viruses while Th2 cells mediate the immune response to extracellular parasites. Additional varieties of specialised T cell have been characterised in recent years including Th17 cells which have been implicated in autoimmunity and Treg cells involved in immune modulation [T helper cell subsets are reviewed in (Reiner, 2007; Romagnani, 2006)]. After infection, some B and T cells persist as memory cells, and these can respond rapidly if the same antigen is encountered again. The cell types analysed in this study are from both the myeloid and lymphoid lineage, they are; dendritic cells, B cells, CD4 T cells, Th1 and Th2 cells (Figure 3.2.1).

[iv] Some kinds of dendritic cell can also develop from the lymphoid lineage and these express the cell surface marker plasmacytoid dendritic cell antigen (PDCA) (Shortman and Liu, 2002).

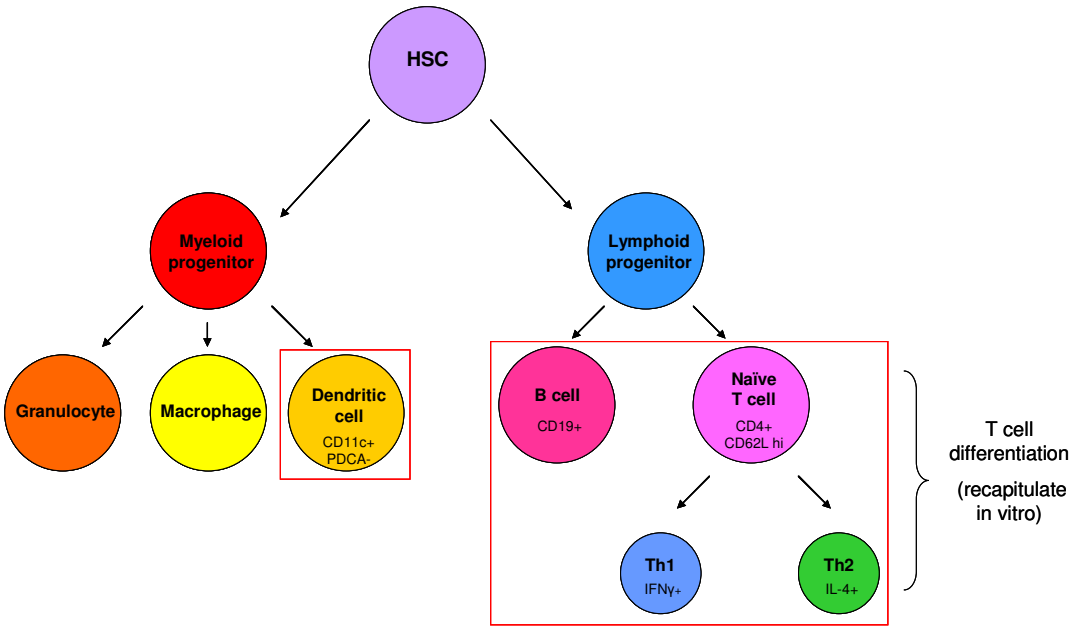


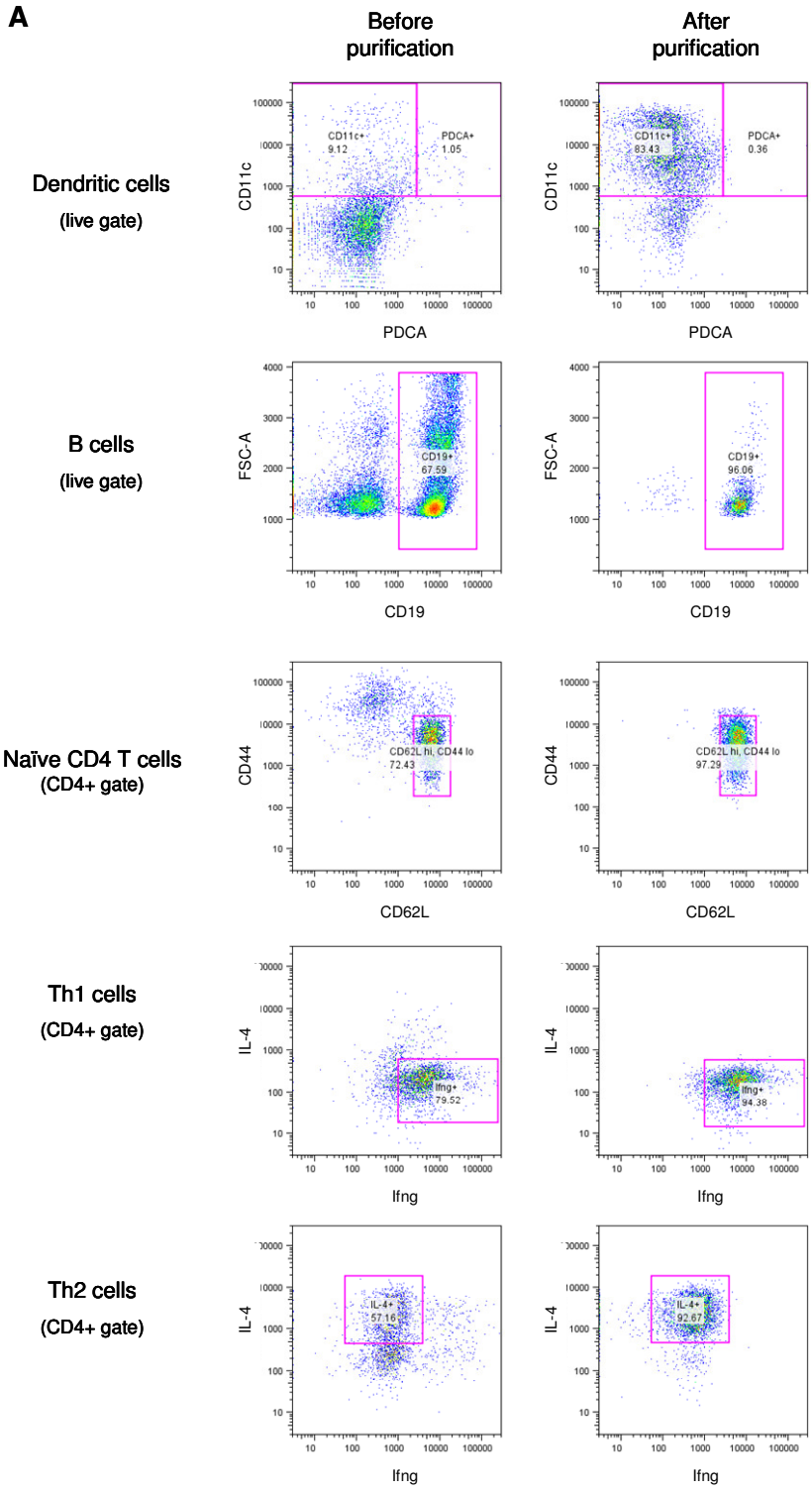
Figure 3.2.1 The immune cell lineage

Immune cells derive from a hematopoietic stem cell (HSC) which can develop into myeloid and lymphoid progenitors. Myeloid cells include granulocytes, macrophages and dendritic cells whilst the lymphoid lineage consists mainly of B and T cells. CD4 T cells can further differentiate into Th1 and Th2 cells and this process can be recapitulated using an *in vitro* culture system. Cells examined in this study are marked with red boxes and markers used to characterise them are indicated below the cell name.

3.2.2 Isolation of immune cells

A major advantage of immune cells is that they represent primary cell types which can be easily purified. Dendritic cells, B cells and CD4 T cells were purified from mouse spleen on the basis of cell surface markers using fluorescence activated cell sorting (FACS) and magnetic cell sorting (MACS). The following markers were used for selection; myeloid dendritic cells CD11c+ PDCA-, B cells CD19+ and T cells CD4+. For gene expression analysis naïve CD4 T cells (CD44 low, CD62L high) were specifically selected (Swain et al., 1991). T helper cell differentiation is of particular interest as DNA methylation changes occur during this process (see section 1.6). I have taken advantage of an *in vitro* differentiation system for T cells to look for DNA methylation changes globally with a focus on CGIs. CD4 T cells were activated by mimicking TCR signalling and cytokines and antibodies were added that skewed cells towards Th1 (IL-12, α-IL-4) or Th2 (IL-4, α-IL-12) cell fate. This was

characterised by staining for cytokine production; Th1 cells produce IFN-γ and Th2 cells produce IL-4. For methylation analysis, Th1 and Th2 cells were sorted on the basis of signature cytokine production using FACS. Purification of all immune cell types was assessed by FACS analysis (Figure 3.2.2).



B	DNA methylation		Gene expression	
cell type	**markers**	**purity**	**markers**	**purity**
DC	CD11c+ PDCA-	85-90%	CD11c+ PDCA-	73-85%
B cell	CD19+	97-98%	CD19+	93-97%
CD4 T cell	CD4+	87-97%	CD4+ CD44 hi, CD62L lo	93-97%
Th1	CD4+ IFN-γ+	90-96%	CD4+ IFN-γ+	70-89%
Th2	CD4+ IL-4+	92-95%	CD4+ IL-4+	40-52%

Figure 3.2.2 Purification of immune cells

A) Representative purifications of each cell type are shown. Dendritic cells (DCs) were purified using MACS by depleting for PDCA+ cells and selecting for CD11c+. B cells were FACS sorted on the basis of CD19+. CD4 cells were purified by MACS and then stained and sorted for naïve cell markers (CD44 lo, CD62L hi). Th1 and Th2 cells were generated in culture from CD4 cells, stained and FACS sorted for cytokine production.

B) Table detailing the purity of cells used for DNA methylation and gene expression analysis.

3.3 Results: Profiling CGI methylation in immune cells by MAP-seq

MAP was used to isolate methylated CpG-rich DNA from dendritic cells, B cells, CD4 T cells, Th1 and Th2 cells. Genomic DNA from each cell type of interest was applied to an MBD affinity column, weakly bound DNA washed off and specifically bound fractions eluted at high salt concentrations. These methylated CpG-rich fractions were then sequenced and differential methylation assessed using custom bioinformatic tools.

3.3.1 Preparation and testing of the MBD column

MBD affinity columns were prepared as per Illingworth et al using the MBD domain (amino acids 77-167) of human MeCP2 (Illingworth et al., 2008)[v]. Histidine-tagged MBD protein was expressed in bacteria and isolated using nickel affinity

purification. MBD protein was then adsorbed onto nickel-charged sepharose via the histidine tag and packed onto 0.5ml chromatography columns for use on an AKTA purifier system (GE Healthcare). Prior to use, MBD columns were washed with 1M NaCl to remove bacterial DNA and other contaminants.

MAP was carried out as follows. Genomic DNA (25µg) was sonicated to an average size of 500bp and bound to the MBD column at a low salt concentration (100mM NaCl). The salt concentration was then increased to 750mM NaCl and unmethylated CGIs and bulk genomic DNA were eluted and retained for analysis. The column was then washed at 750mM NaCl to ensure separation between bulk and specifically-bound DNA. Methylated CGIs were eluted by increasing the salt concentration to 1M NaCl. A schematic outlining the MAP procedure is shown in Figure 3.3.1. For each new batch of columns small adjustments, based on PCR analysis of purified DNA, had to be made to the salt concentration at which the wash step was carried out in order to account for variation in column packing and density of recombinant MBD protein. Where practical, MBD protein from the same large-scale purification was used and was bound to nickel sepharose beads in batches in order to minimise variation between columns. A maximum of 8 runs was carried out on each MBD column to avoid purification differences due to column degeneration.

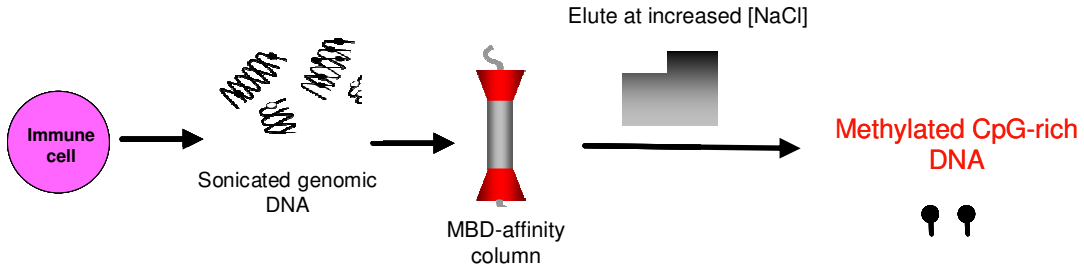


Figure 3.3.1 MBD affinity purification (MAP)

Genomic DNA from a cell type of interest is sonicated and applied to an MBD affinity chromatography column. DNA is eluted by increasing the salt concentration as methylated CpG-rich fractions elute at high salt concentrations.

[v] The majority of the MBD protein production was carried out by U. Gruenewald-Schneider.

Specific enrichment of methylated sequences was tested by carrying out PCR for regions of known methylation status and CpG-density. Unmethylated CGIs were eluted at low salt concentrations along with bulk CpG-deficient genomic DNA while methylated CGIs eluted at high salt concentrations. Three somatically methylated CGIs (meCGI 1, meCGI 2, meCGI 6) and associated regions of bulk genomic DNA (bulk 1, bulk 2, bulk 6) were used for PCR testing of MAP fractions (identified by R. Illingworth, see "Materials and Methods" for genomic location). Using associated CGI and bulk regions allowed us to assess whether DNA fragmentation was sufficient to separate nearby regions. As a further verification of the MAP technique methylated CGI fractions isolated from male and female mice were labelled and hybridised to microarrays containing all mouse promoters and CGIs (NimbleGen mouse promoter+CpG island arras). X-linked CGIs which are specifically methylated on the female inactive X chromosome were, as expected, enriched in the female sample compared to the male. There was no such sex-specific enrichment of autosomal CGIs (Figure 3.3.2; microarray analysis carried out by R. Illingworth and J. de las Heras).

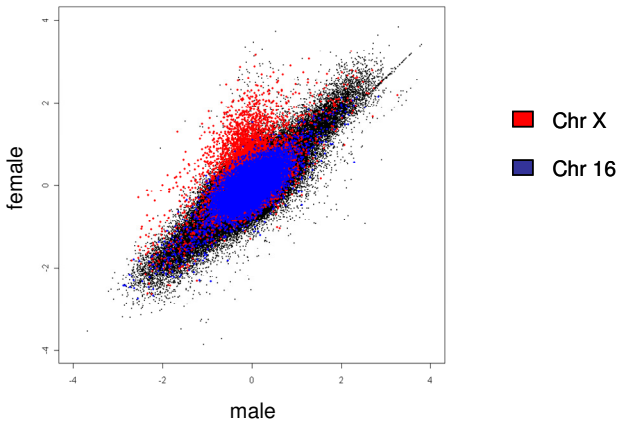


Figure 3.3.2 MAP identifies methylated X-linked CGIs in females

Genomic DNA from male and female mice was purified by MAP, labelled and hybridised to NimbleGen mouse promoter+CpG island arrays. Chromosome X shows specific enrichment for methylation in females due to the presence of methylated CGIs on the inactive X chromosome. Chromosome 16 shows no difference in methylation between males and females.

3.3.2 Optimisation of MAP purification for high throughput sequencing

Although the MAP protocol detailed above was sufficient to detect differential methylation using microarrays additional optimisation was necessary in order to analyse methylated fractions using Illumina sequencing technology. Initial sequencing of methylated CpG-rich fractions gave a high level of background (female CD4 cells, data not shown). Sequencing is more sensitive to background than microarrays as the entire genome is analysed rather than a defined subset of features. Background was substantially reduced by carrying out a second MBD column purification. Specifically bound fractions from the first affinity purification were reapplied to the column and washing and elution of specifically-bound DNA was carried out as before. PCR testing of column fractions confirmed a complete absence of signal for bulk genomic DNA and enrichment for methylated CGIs after the second purification (Figure 3.3.3A). AKTA run traces show the amount of DNA eluted at each purification stage (Figure 3.3.3B). For high throughput sequencing adaptors are attached to fragmented DNA to allow amplification and attachment to the flow cell. The small amount of DNA obtained from MAP[vi] made attachment of adaptors after purification problematic. To circumvent this issue, adaptors were attached to total input DNA before it was applied to the column.

[vi] Nanogram quantities of DNA are recovered as CGIs make up a small fraction (0.56%) of the genome.

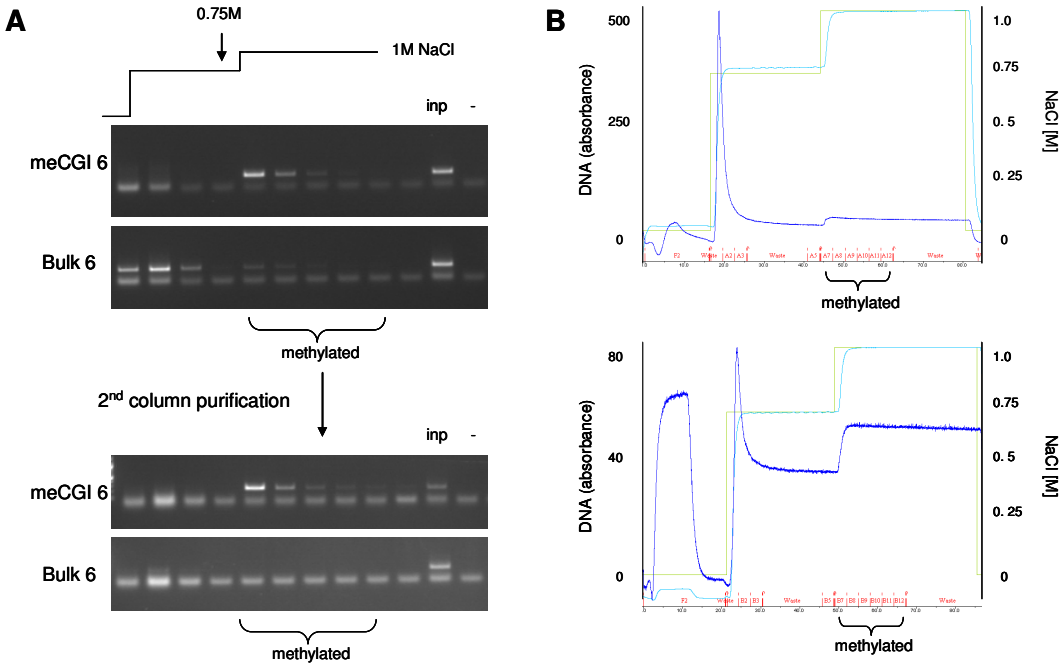


Figure 3.3.3 Two rounds of MAP are carried out prior to high throughput sequencing

A) Fractions eluted from the MBD column over different salt concentrations were tested by PCR to verify enrichment for methylated CGIs over bulk DNA. Here PCR for meCGI 6 is shown along with PCR for an associated region of bulk genome (bulk6). Methylated fractions were pooled and subject to a 2nd round of purification followed by PCR testing.

B) Run traces from MAP carried out on an AKTA purifier. The upper panel depicts the first column purification where UV absorbance (dark blue trace) shows that the majority of DNA elutes at less than 0.75M [NaCl] (green line) and a small amount of methylated DNA elutes at higher concentrations. The lower panel is representative of a second round of purification where a smaller amount of non-specific material elutes at low salt. For A) and B) brackets indicate methylated CpG-

3.3.3 Generation of methylation profiles using MAP-seq

Genomic DNA was isolated from dendritic cells, B cells, CD4 T cells, Th1 and Th2 cells. For each cell type two biological replicates were generated and purified separately using MAP as outlined above. Methylated CpG-rich fractions from the second column purification were pooled, precipitated and passed on to the Wellcome Trust Sanger Institute for Illumina library production and sequencing. 37bp single-end reads were generated using an Illumina Genome Analyzer according to the standard Illumina protocol. Reads were mapped to the mouse genome (mm9 – NCBI37; repeat masked) using Maq (http://maq.sourceforge.net) and only high

quality read placements (Maq score ≥ 30) were retained (mapping carried out by R. Andrews, Wellcome Trust Sanger Institute). For each replicate one lane of sequencing was carried out and each lane gave an average of 16.2 million sequencing reads. Data were received as .wig files containing the number of sequencing hits for each location in the genome. These were visualised using the Integrative Genome Browser (IGB; Affymetrix) (Nicol et al., 2009) and each lane was inspected to ensure comparability between samples in terms of background reads and signal over CGIs of known methylation status. The two lanes of sequencing carried out for each sample were then combined and all analysis performed on the combined data. An average of 429 million base pairs of sequence was generated per cell type. Sequencing data generated for each cell type is summarised in Table 3.3.1.

Sample	Number of reads generated (Maq ≥ 30)	Number of bp mapped
DC	32,699,330	448,7184,25
B	48,087,285	613,253,355
CD4 1	25,005,336	427,865,743
CD4 2	39,376,768	426,508,805
Th1	21,903,336	291,476,713
Th2	24,343,600	364,400,346

Table 3.3.1 Summary of MAP-seq data generated

Number of 37bp sequencing reads generated for each cell type and the number of base pairs (bp) of mapped sequence obtained. For CD4 cells two samples were generated (see section 3.3.4).

Regions highly enriched using MAP included methylated CGIs, as expected, as well as other CpG-dense methylated regions. Methylated CGIs were identified by comparing the MAP-seq data to the mouse CGI set generated by CAP-seq (Illingworth et al., 2010). Figure 3.3.4 shows representative profiles of MAP-seq reads for CD4 cells, Th1 cells and dendritic cells along with the CAP-seq profile for sperm which indicates the position of CGIs. Plotting the CpG frequency (in 300bp windows with a 10bp slide) shows that regions identified by MAP-seq are CpG-rich compared to the bulk genome. It is also worth noting that MAP-seq profiles for the

different cell types are strikingly similar indicating that, in general, the methylation pattern is conserved between these cell types.

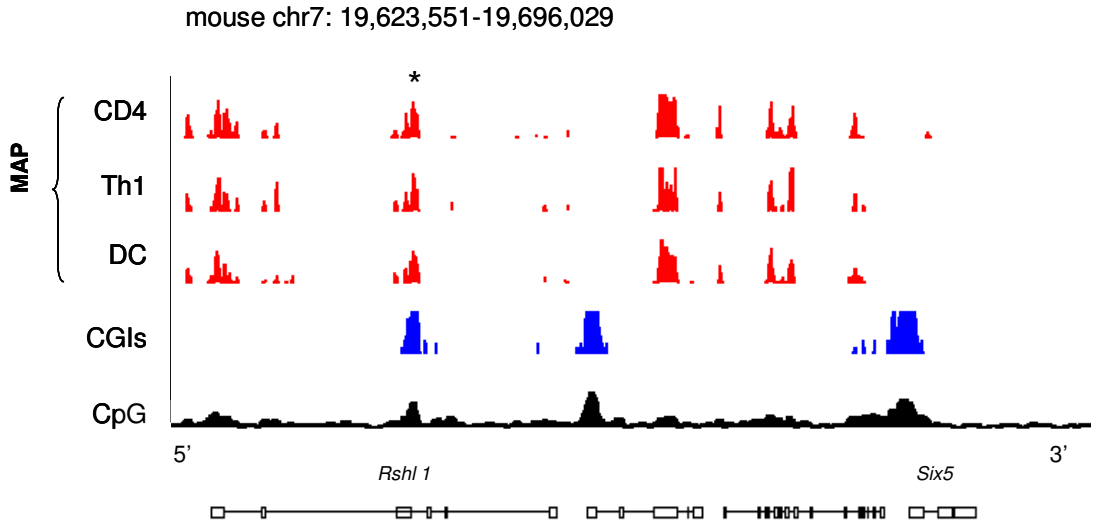


Figure 3.3.4 MAP-seq identifies methylated CGIs and CpG-rich methylated regions
Typical MAP-seq read density profiles (red) are shown for CD4 T cells (CD4), Th1 cells (Th1) and dendritic cells (DC). The CAP-seq profile for sperm is shown to indicate the position of CGIs (CGIs, blue) and the CpG density of the genome calculated in 300bp windows with a 10bp slide is also shown (CpG; scale 0-64 CpGs per 300bp). RefSeq genes are annotated below the sequence profiles. The asterisk denotes a methylated CGI. Unmethylated CGIs are also evident (in CGI profile) along with non-CGI-associated MAP-seq peaks (in MAP-seq profile).

3.3.4 Analysis of MAP-seq data using custom bioinformatic tools

In order to analyse the MAP-seq data, a number of custom bioinformatic tools were developed. These were designed in conjunction with R. Illingworth and S. Webb and accessed via an in-house version of the Galaxy application (Blankenberg et al., 2007; Taylor et al., 2007). Analysis was performed on .wig files. Initial analysis involved normalisation of .wig files in order to make read depth between samples comparable and peak-finding to identify regions enriched by MAP-seq. The parameters used in each analysis step are outlined in Table 3.3.2. They are; read height (H), length in bp (L) and gap permitted in the length parameter (G). The gap parameter allows small gaps in read coverage occurring in otherwise contiguous regions.

Analysis step	Parameters
Background removal	H: 2 L: 90 G: 20
Normalisation	Scale to 250 million reads
Peak-finding	H: 4 L: 90 G: 250
Sliding window	100bp window, 20bp slide
Identification of differences	log2 difference > 1.8 in 9/13 contiguous windows

Table 3.3.2 Parameters for analysis of MAP-seq data

Parameters for each analysis step are shown (described in section 3.3.4). H = height in number of reads, L = length in bp, G = gap in bp.

Data normalisation

Raw sequencing data were normalised in order to make different samples directly comparable. MAP data for each cell type were scaled to the average read number across all the MAP samples generated in the lab (approximately 250 million after background removal) in order to account for variable sequence depth. Background was removed prior to normalisation as fluctuations in the number of background reads could skew the normalisation procedure.

Peak-finding

Regions of prominent enrichment, or "peaks", were identified using the parameters listed in Table 3.3.2. Regions with a height of at least 4 reads for a length of at least 90bp were deemed to be enriched. This was verified by identification of regions of known methylation state as well as by bisulfite sequencing. Peaks of enrichment were then examined to assess the CpG density and gene association of regions identified by MAP-seq (section 3.3.5). MAP peaks from all of the mouse MAP samples generated in the lab were combined with the mouse CGI set to give a set of regions of interest over which to perform further analysis.

Identification of differentially methylated regions

In order to accurately identify differentially methylated CGIs, a sensitive "sliding window" analysis was carried out over each of these regions of interest. For each region the average number of hits per base was calculated in 100bp windows with a 20bp slide. Values for each window were then compared between samples. This gave a ratio for each window. If both windows being compared contained less than 4 reads this ratio was set to 1 in order to remove spurious hits attributed to small fluctuations at low read depth. Differentially methylated regions were defined as those containing 9 out of 13 contiguous windows with a log2 ratio of >1.8 (i.e. 3.5-fold). Contiguous differentially methylated regions were knitted together. These parameters were verified by manual scoring of false positives and false negatives and also by correct identification of CGIs on the X chromosome which are differentially methylated between males and females (tested by comparing MAP-seq data for male and female blood). Individual examples were also tested by bisulfite sequencing and a selection of these is shown in Figure 3.3.7 as well as in section 3.6. An entire analysis scheme comparing Th1 and dendritic cells methylation data visualised in IGB is shown in Figure 3.3.5.

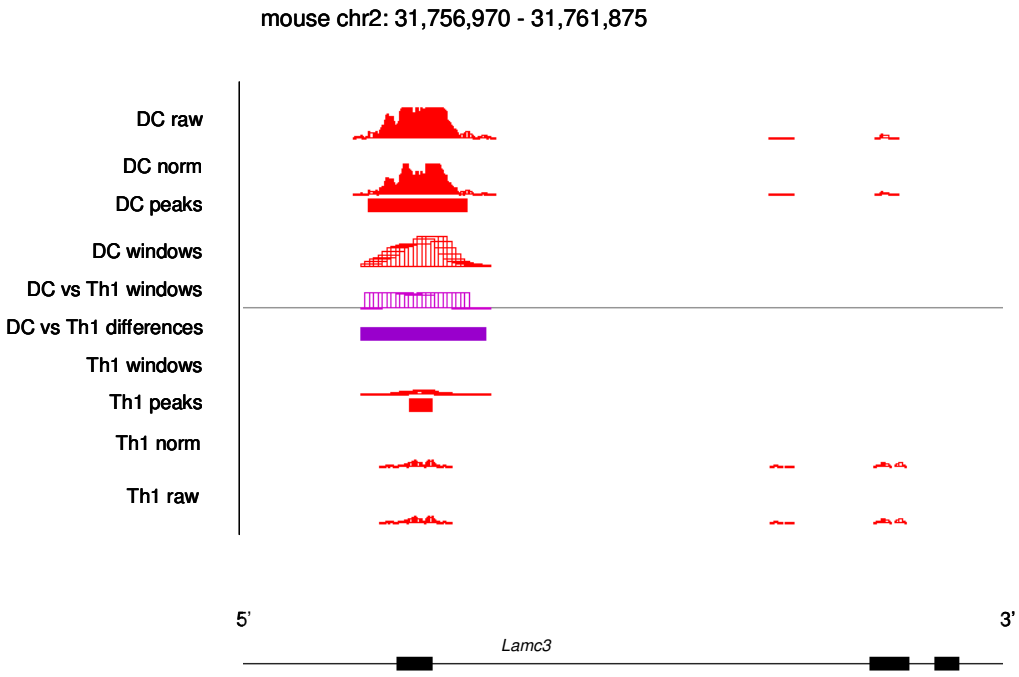


Figure 3.3.5 Analysis scheme for MAP-seq data

Example of a region differentially methylated in dendritic cells (DC) and Th1 cells (Th1) identified by analysis of MAP-seq data. Read density profiles are shown for raw (raw) and normalised (norm) data along with peaks of enrichment (peaks). Sliding window analysis on each sample (100bp window, 20bp slide) calculated the average number of hits per bp in a given window (DC windows and Th1 windows) and was carried out for all regions of interest. Values for these windows were then compared between samples (DC vs Th1 windows) and used to identify differential methylation (DC vs Th1 differences). RefSeq genes are plotted below the profiles. Data were visualised in IGB.

3.3.5 MAP-seq identifies methylated CGIs and other CpG-rich methylated DNA

After the analysis method had been developed it was necessary to characterise the regions that could be identified by MAP. It was evident from the read density profiles (Figure 3.3.4) and peak-finding analysis that, in addition to methylated CGIs, MAP also enriches for other CpG-rich methylated DNA. All peaks of enrichment identified in CD4 T cells were analysed for genomic location and CpG density to give an idea of what a typical MAP purification identifies. MAP-seq in CD4 cells identified 50,778 regions; 9.3% (4,738) of which were located at CGIs and 90.7% (46,040) which were located elsewhere in the genome (non-CGI) (Figure 3.3.6A).

Methylated CGI peaks, as expected, showed strong association with genes (66.3% gene-associated – defined as location anywhere in a transcription unit) and non-CGI peaks were also preferentially gene-associated (66.9% gene-associated) compared to the whole genome (27%) (Figure 3.3.6B). The gene-association of non-CGI MAP-seq peaks is most likely due to elevated CpG content in exons due to coding restraints although this could also indicate a functional importance for DNA methylation in these regions. The CpG density of MAP-seq peaks in CD4 cells was assessed by calculating the number of CpG sites per 100bp in each peak. The tenth percentile of these data was 3.3 CpGs per 100bp and this CpG density was taken to represent the minimum CpG density necessary for binding to the MBD column (Figure 3.3.6C). To put these figures in context mouse CGIs contain, on average, 6.7 CpGs per 100bp while the bulk genome has less than 1 CpG per 100bp. The fraction of the non-repetitive genome interrogated using MAP-seq was also calculated. For a single MAP-seq analysis in CD4 cells, peaks of enrichment covered 0.6% of the repeat-masked mouse genome. This corresponds to 674,000 CpG sites or 3.2% of all CpGs in mouse genome that occur outside repeats.

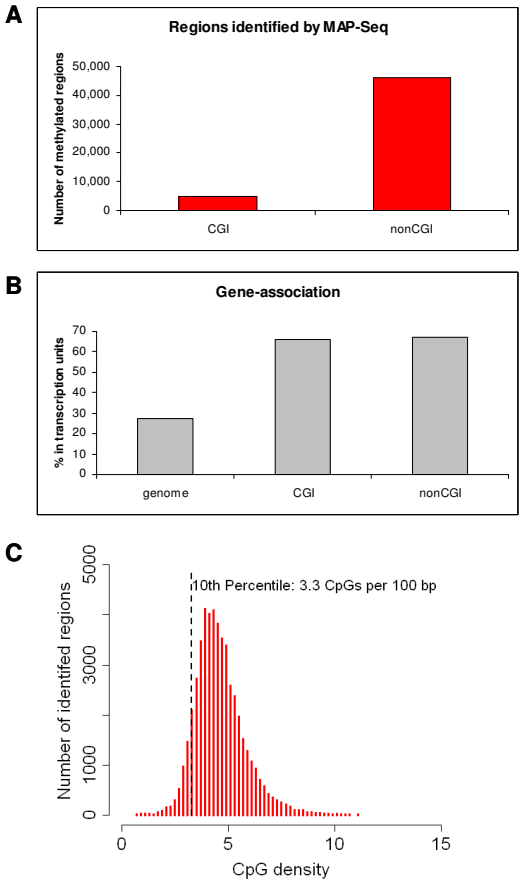


Figure 3.3.6 Regions identified by MAP-seq

A) Number of CGI-associated (CGI) and non-CGI-associated (nonCGI) methylated peaks identified by MAP-seq in CD4 cells. **B)** Gene-association of MAP-seq peaks. The percentage of the genome that occurs within transcription units (genome) is indicated along with the percentage of CGI and non-CGI peaks in CD4 cells that overlap genes. **C)** CpG density of regions identified by MAP-seq in CD4 cells. The 10th percentile is taken to represent the CpG density necessary for detection of a region by MAP-seq.

3.3.6 MAP-seq identifies methylation differences between cell types

Differential methylation was assessed between cell types as described (section 3.3.4). Cell types were compared pair-wise and differential methylation was categorised as CGI (overlapping a CGI) or non-CGI (not overlapping a CGI). Some variation between MBD columns was observed meaning that some of the samples run on different batches of column were too variable to be compared using my quantitative analysis method. In practise this meant that B cell and CD4 cell MAP samples run on the same MBD column were only comparable to each other and CD4, Th1, Th2 and

dendritic cells samples run at a separate time were comparable. As most of the samples contained pools of male and female DNA the sex chromosomes were excluded from the analysis in case fluctuations in the proportion of male and female DNA between samples skewed the results. CD4, Th1 and Th2 cells which are all highly developmentally related showed a very small number of methylation differences between them. In particular, Th1 and Th2 cells had almost identical methylation patterns with just one CGI showing a methylation difference. More differences were seen between T cells and less related B and dendritic cells (Table 3.3.3).

Comparison	CGI changes	Non-CGI changes
DC vs Th1	102	249
DC vs Th2	59	175
DC vs CD4	37	79
B cell vs CD4	83	187
Th1 vs CD4	16	75
Th2 vs CD4	5	42
Th1 vs Th2	1	0

Table 3.3.3 Methylation differences between cell-types as identified by MAP-seq
Cell types were compared pair-wise and for each comparison the number of methylation differences at CGIs and non-CGI regions is indicated. Differences were calculated using sliding window analysis on normalised .wig files as outlined in section 3.3.4.

3.3.7 Verification of MAP-seq results by bisulfite sequencing

The methylation status of 11 CGIs showing differential methylation between cell types by MAP-seq was checked using bisulfite sequencing and all regions tested confirmed MAP-seq results (summarised in Table 3.3.4).

Gene	Genomic region	MAP-seq (cell types showing increased methylation listed first)	Bisulfite (% methylation)
Gata3	chr2: 9795943 -9796614	Th1 (vs Th2)	72% Th1, 54% Th2
Dtx1	chr5: 121160061-121160460	Th1 (vs CD4/DC)	55% Th1, 30% CD4, 17% DC
Kcnn4	chr7: 25161961 - 25162900	DC, B cell, CD4 (vs Th1)	66% DC, 70% B, 32% CD4, 20% Th1
Npb	chr11: 120469741 - 120470220	Th1 (vs DC)	54% Th1, 9% DC
Sema4a	chr3: 88240621 - 88241160	CD4, Th1, Th2 (vs DC)	80% Th1, 29% DC
Tnfaip2	chr12: 112683701 - 112684060	DC (vs Th1)	66% DC, 26% Th1
Csf2ra	chr19: 61304341-61304760	Th1 (vs DC)	81% Th1, 60% DC
Nr4a3	chr4: 48064301-48065140	Th1 (vs DC)	73% Th1, 16% DC
Hic1	chr11:74979221-74979660	Th1 (vs DC)	33% Th1, 7% DC
Sgms2	chr3:131033081-131033520	DC (vs Th1)	44% DC, 25% Th1
Klf2	chr8: 74844581-74845040	DC (vs Th1)	80% DC, 47% Th1

Table 3.3.4 Bisulfite verification of methylation differences

The CGI methylation differences between cell types identified by MAP-seq that were verified by bisulfite genomic sequencing.

Two examples of bisulfite sequencing carried out on regions which are differentially methylated in dendritic cells compared to Th1 cells are shown in Figure 3.3.7. Additional MAP-seq verification by bisulfite sequencing is shown for the examples discussed in section 3.6.

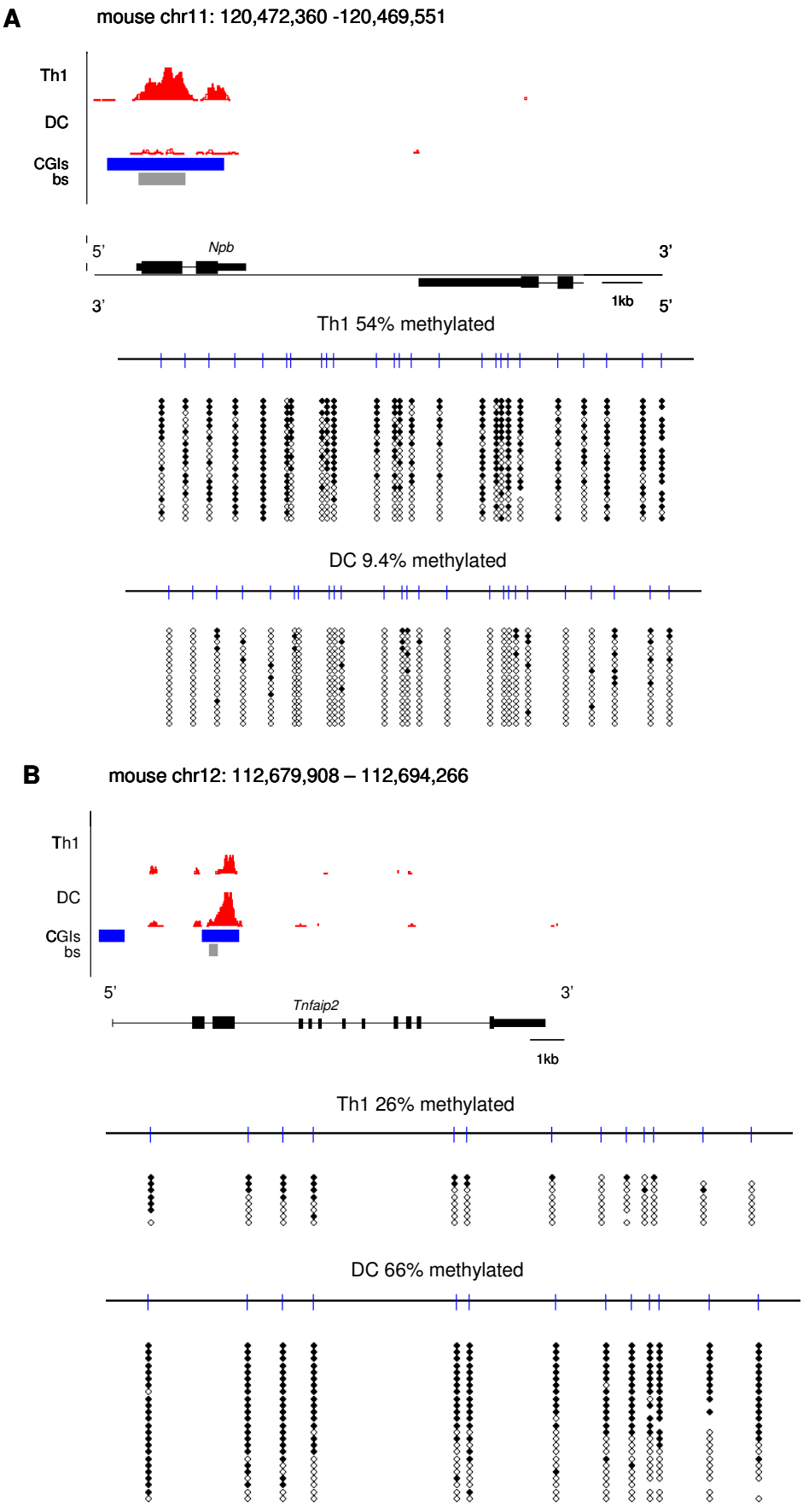

A
mouse chr11: 120,472,360 -120,469,551
Th1
DC
CGIs
bs
5'
Npb
3'
3'
1kb
5'
Th1 54% methylated
DC 9.4% methylated
B
mouse chr12: 112,679,908 – 112,694,266
Th1
DC
CGIs
bs
5'
Tnfaip2
3'
1kb
Th1 26% methylated
DC 66% methylated

Figure 3.3.7 Verification of differential methylation using bisulfite sequencing

A) The *Npb* CGI is hypermethylated and **B)** the *Tnfaip* CGI is hypomethylated in Th1 cells compared to dendritic cells (DC). For each example the upper panel shows read density profiles generated by MAP-seq (red), CGIs (blue) and the region interrogated by bisulfite sequencing (bs, grey bars). The lower panel shows the bisulfite sequencing result; a blue dash indicates a CpG site, filled circle = methylated CpG, empty circle = unmethylated CpG.

3.4 Results: Profiling gene expression in immune cells

Gene expression was profiled in dendritic cells, B cells, naïve CD4 cells, Th1 and Th2 cells using Illumina Whole-Genome Expression BeadChips and used to identify genes showing differential expression between cell types.

3.4.1 Analysis of gene expression data

RNA was extracted from cells of interest, labelled and passed on to the Wellcome Trust Clinical Research Facility, Edinburgh (WTCRF) for hybridisation to Illumina MouseWG-6 BeadChips. These chips contain probes for all RefSeq transcripts as well as probes for some other transcripts from additional databases. BeadChip arrays consist of beads to which probes for each transcript have been covalently attached. Approximately 30 different beads are present for every probe sequence giving a high level of redundancy. Six RNA samples can be hybridised to a chip in parallel which gives high reproducibility and makes this a cost effective way of interrogating gene expression. For each cell type three biological replicates were carried out. Data for each identical probe sequence were summarised using Illumina BeadStudio. Data were normalised using the average normalisation method in BeadStudio which scales each array based on a scaling factor derived from the average intensity across all arrays. Background was not removed from samples during normalisation as this has been shown to artificially bias BeadChip gene expression results (Dunning et al., 2008). Subsequent analysis was carried out using GeneSpring GX10 (Agilent technologies). For analysis of differential gene expression, genes with low expression (Flags absent) across all cell types were removed. Statistical analysis was carried out to find genes that were differentially expressed between any of the cell

types (one-way ANOVA, $p \leq 0.05$). Pair-wise comparisons were performed on this gene set and genes changing at least two-fold between cell types were deemed to be differentially expressed. Where multiple probes for the same gene indicated differential expression this was counted as one change. Only RefSeq genes from the array were examined further as they represent a well-defined gene set which is easy to overlay with CGIs and other genomic features.

3.4.2 There are large gene expression differences between immune cell types.

Many changes in gene expression were observed between different immune cell types. For example, over 2,000 differences of at least two-fold were observed between most cell types. The exception to this was Th1 and Th2 cells which showed more similar gene expression profiles having just 362 expression differences between them. Strikingly, naïve CD4 cells had a similar number of gene expression differences compared to developmentally-related Th1 cells as they did compared to more developmentally and functionally distinct dendritic cells (Table 3.4.1 and Figure 3.4.1).

Comparison	RefSeq genes changed
DC vs Th1	2782
DC vs Th2	2677
DC vs CD4	2059
B cell vs CD4	1352
Th1 vs CD4	2811
Th2 vs CD4	2518
Th1 vs Th2	362

Table 3.4.1 Gene expression difference between cell types

Cell types were compared pair-wise and genes showing at least a two-fold expression change were classed as differentially expressed. The number of RefSeq genes that change is indicated for each comparison.

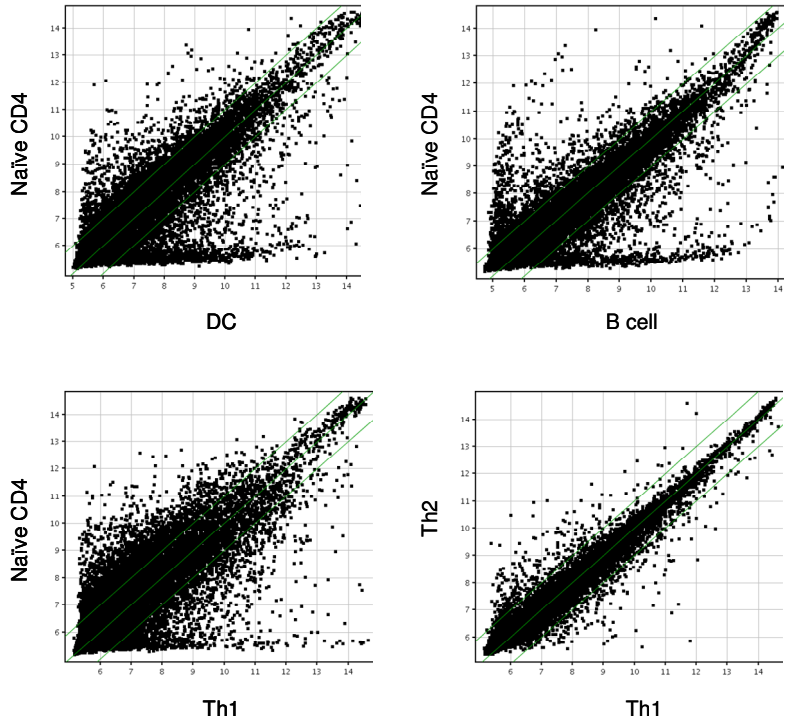


Figure 3.4.1 Pair-wise plots comparing gene expression between cell types

Scatter plots showing pair-wise comparisons of gene expression between cell types. Each dot represents a unique probe on the array and probes located outside of the green lines show at least a two-fold difference in expression. The scale is log2 of intensity values.

3.4.3 Validation of expression arrays

A number of the gene expression differences detected by the microarray analysis were verified by quantitative reverse transcription PCR (qRT-PCR). In qRT-PCR, expression of the gene being tested is compared to that of a reference gene that is not expected to change its expression under the conditions tested. This accounts for variability in RNA quality or in the efficiency of the reverse transcription reaction. Common references include housekeeping genes such as *Gapdh* and *ActB*, however, these genes change expression upon T cell activation (Hamalainen et al., 2001) making them an unsuitable choice for analyses involving T cells. The gene for elongation factor 1 α-1 (*Eef1A1*) was chosen as a reference based on its use in a published study (Hamalainen et al., 2001) as well as relatively consistent expression across different cells in the microarray study and in T cell types when tested by qRT-PCR.

A total of 12 genes showing differential expression in the microarray experiment were verified by qRT-PCR and results were found to be consistent between the two techniques. qRT-PCR and microarray results for two of the genes tested; *Kctd17* and *Klf2* are shown in Figure 3.4.2. Additional qRT-PCR verification is shown for particular examples discussed in section 3.6.

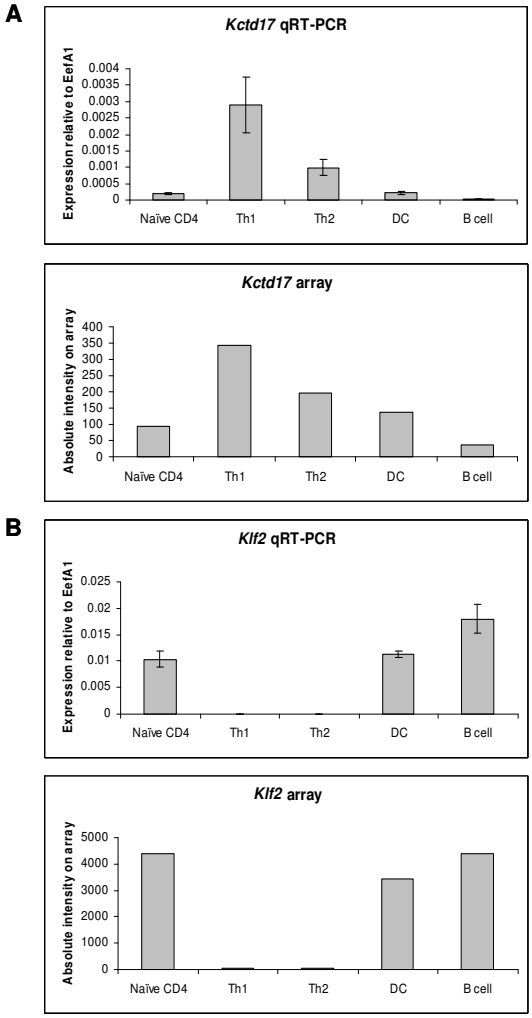


Figure 3.4.2 Verification of expression array data by qRT-PCR

A) *Kcd17* and **B)** *Klf2*. For each, the top panel shows qRT-PCR relative to *Eef1A1* (error bars = standard deviation from PCR replicates) and the bottom panel shows the absolute intensity value from the BeadChip arrays averaged across replicates.

3.5 Results: Analysis and comparison of DNA methylation and gene expression

Methylation results were analysed in detail with respect to location of changes, function of associated genes and categories of CGI that showed differential methylation. These were then related to changes in gene expression between cell types. MAP-seq and gene expression data for brain have been generated in our lab and this allowed brain to be used as a distantly related tissue with which to compare data for immune cells.

3.5.1 Methylation differences between immune cells are small in number and show an association with developmental similarity

The number of methylation differences between immune cells was relatively small ranging from 351 changes between dendritic cells and Th1 cells to just one change between Th1 and Th2 cells (Table 3.3.3). In general, the number of methylation differences between cells was associated with developmental similarity as the most related cell types, CD4, Th1 and Th2 cells, showed the fewest differences. In addition, these changes were smaller in magnitude than those between more distantly related immune cells. For example, regions differentially methylated in Th1 and CD4 cells showed, on average, a 3.1-fold difference in the number of sequencing hits while differentially methylated regions in more distantly related Th1 and dendritic cells showed, on average, a 4-fold difference. In terms of the number of methylation differences that were identified between particular cell types, CGIs showed the same trend with respect to developmental similarity as non-CGI-associated regions although the number of methylation changes at CGIs was smaller (Figure 3.5.1A and B).

MAP-seq data for immune cells were compared to data generated for brain in our lab (Skene et al., 2010). In contrast to the similar patterns of methylation seen between different immune cells, MAP-seq identified large numbers of methylation differences between immune cells and brain. Over 2,000 CGIs and 20,000 non-CGI regions

showed differential methylation between CD4 cells and brain further illustrating the correlation between DNA methylation and developmental proximity (Figure 3.5.1A and B). Interestingly, brain was hypomethylated at both CGI and non-CGI regions compared to CD4 cells.

Gene expression differences between immune cell-types were much more numerous than methylation differences. Up to 2811 genes were differentially expressed between immune cell types and changes had little correlation with developmental similarity; Th1 and CD4 cells showed the same number of gene expression differences as Th1 and dendritic cells (Table 3.4.1 and Figure 3.5.1C). The exception to this was Th1 and Th2 cells which showed the most similar gene expression profiles (362 changes) indicating functional similarity (Figure 3.5.1C). Although naïve CD4 and Th1/Th2 cells are very close developmentally and show similar DNA methylation profiles naïve CD4 cells are quiescent while Th1 and Th2 cells have been stimulated to respond to infection. Evidently stimulation elicits large changes in the transcriptional program and therefore seems to play a larger part than developmental relationship in determining gene expression. Like non-activated CD4 cells, the majority of B cells and dendritic cells examined in this study are quiescent as they have been taken from the spleens of uninfected animals. Non-activated B and CD4 cells profiled here are more similar in terms of gene expression than naïve CD4 cells and activated Th1 or Th2 cells. Methylation on the other hand does not seem to be substantially affected by T cell activation (Table 3.3.3 and Figure 3.5.1A and B) consistent with a report that resting *ex vivo* T cells and T cells activated and expanded in culture retain the same T cell-specific methylation patterns (Schmidl et al., 2009).

Immune cell gene expression data were compared to expression data generated for brain using BeadChip arrays (J. Guy, unpublished). Nearly 6,000 differentially expressed genes were detected between naïve CD4 cells and brain indicative of their drastically different functions (Figure 3.5.1C). The observation that there are fewer gene expression changes (~5,719) than methylation changes (~23,029) when CD4

cells are compared to brain is in contrast to the situation when different immune cells are compared. There are 2,811 gene expression differences but only 91 methylation differences between CD4 and Th1 cells. This suggests a developmental origin for DNA methylation patterns. Cells with a common developmental origin such as cells of the hematopoietic lineage share similar DNA methylation profiles despite large differences in gene expression while developmentally distinct tissues such as brain show a very different methylation pattern when compared to immune cells.

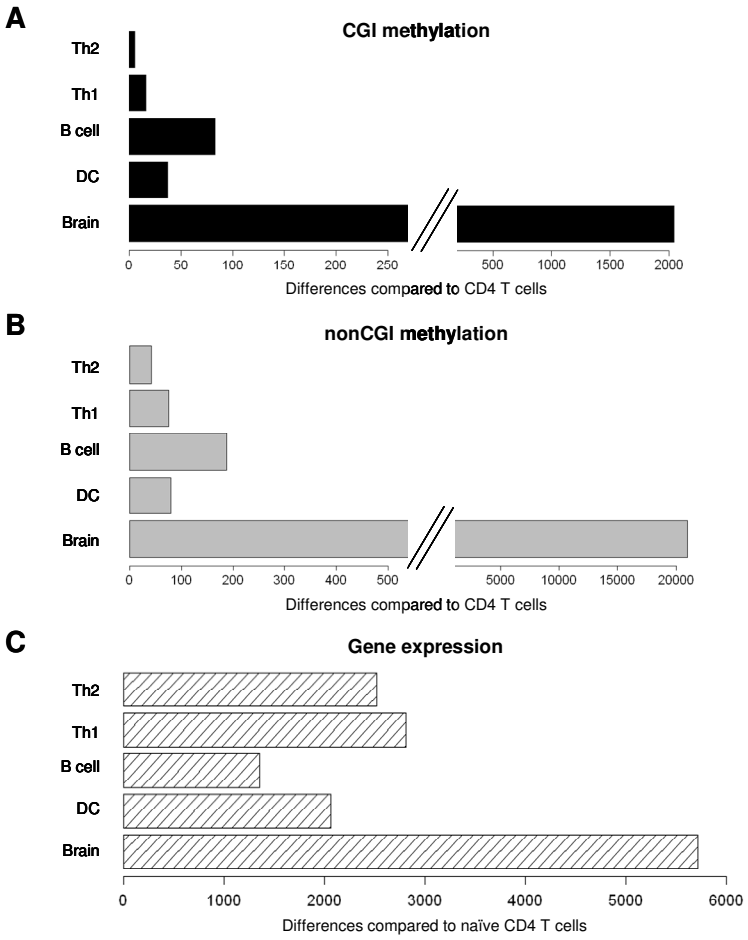


Figure 3.5.1 DNA methylation and gene expression differences between CD4 cells and other cell types

A) The number of differences in CGI methylation detected between CD4 cells and brain, dendritic cells (DC), B cells, Th1 and Th2 cells. **B)** Number of non-CGI methylation differences **C)** Number gene expression differences of at least two-fold.

3.5.2 Methylation differences preferentially occur in genes with immune system function

Although the number of methylation changes between different immune cells is small these changes may be functionally important. I assessed this by examining the function of genes that were associated with cell-specific methylation occurring at both CGI and non-CGI regions (i.e. genes with a methylation change located anywhere within the transcription unit). 78% of CGI and 71% of non-CGI methylation changes occurred within genes. Gene ontology (GO) terms for these genes were investigated and overrepresented categories identified using the Panther Database (http://www.pantherdb.org). Genes showing differential CGI methylation were enriched for functions in immunity and defence, developmental processes and cytokine/chemokine-mediated immunity when all CGI-associated genes were used as a reference (binomial test with multiple testing correction, $p \leq 0.05$ taken to indicate significant enrichment). Genes differentially methylated at non-CGI regions showed an overrepresentation of GO terms associated with various signalling pathways including T cell activation and inflammation when compared to all genes associated with a MAP-seq peak (Table 3.5.1). These results suggest that DNA methylation changes in immune system cells may play a functional role and have not simply occurred by chance.

GO biological process enriched in genes w CGI changes	p-value
Immunity and defence	0.0075
Developmental processes	0.014
Cytokine/chemokine mediated immunity	0.043
GO pathways enriched in genes w non-CGI changes	**p-value**
T cell activation	0.0048
Apoptosis signalling pathway	0.026
EGF receptor signalling pathway	0.035
Inflammation	0.04

Table 3.5.1 Over-represented GO terms for genes showing differential methylation in the immune system

GO term analysis was carried out using tools available on the Panther database. GO terms for genes associated with differentially methylated CGIs were compared to GO terms for all CGI-associated genes. GO terms for genes showing non-CGI differential methylation were compared to all genes containing a peak identified by MAP-seq. P-values ≤0.05 were taken as significant (binomial test with multiple testing correction).

Genes that show differential methylation in immune system cells were then examined individually with respect to function and many key immune system genes were identified. Significantly, the only methylation difference between Th1 and Th2 cells occurred in a CGI in exon 3 of the *Gata3* gene, a key determinant of Th2 cell fate (Nawijn et al., 2001). Although changes between Th1 and Th2 cells at the *Ifng* and *Il4* loci have been extensively documented, it is important to note that these loci are CpG-poor and therefore can not be investigated by MAP-seq. The only gene-associated CGIs to show methylation differences between CD4 cells and both Th1 and Th2 cells are located in *Dtx1*, a gene which suppresses T cell activation and thus is down regulated when T cells are activated (Liu and Lai, 2005), and *Kcnn4*, which codes for a calcium-activated potassium channel involved in T cell activation (Begenisich et al., 2004). Other genes with an immune system role that show differential CGI methylation include *Fgr* which negatively regulates myeloid cell chemokine signalling (Zhang et al., 2005), *Sema4a* which provides a co-stimulatory signal to T cells and is necessary for Th1 cell differentiation (Kumanogoh et al.,

2005), *Bcl11b* which regulates thymocyte development (Wakabayashi et al., 2003), *Runx3* which promotes Th1 cell fate (Djuretic et al., 2007), *Klf2* which regulates the *Il2* promoter (Wu and Lingrel, 2005), interleukin-10 receptor subunit alpha (*Il10ra*) and many more (see Table 3.5.2). In addition, many genes that show cell-specific methylation at non-CGI regions also function in the immune system. Some of these are genes that also show methylation changes at their CGIs such as *Kcnn4* and *Bcl11b*. Other genes showing differential non-CGI methylation include *Tcf4*, a transcription factor with roles in dendritic cell, T cell and B cell development (Cisse et al., 2008; Wikstrom et al., 2006; Wikstrom et al., 2008), the chemokine receptor *Ccr7* (Worbs and Forster, 2007) and the interleukin-17 receptor *Il17ra* (Toy et al., 2006).

Gene	**Cell types showing CGI hypermethylation (comparison in brackets)**	**Function**	**References**
Gata3	Th1 (vs Th2)	Th2-specific transcription factor	(Nawijn et al., 2001)
Dtx1	Th1, Th2 (vs CD4/DC)	Suppression of T cell activation	(Liu and Lai, 2005)
Kcnn4	CD4, DC, B cell (vs Th1/Th2)	Potassium channel involved in T cell activation	(Begenisich et al., 2004)
Fgr	CD4 (vs DC)	Regulation of myeloid cell chemokine signalling	(Zhang et al., 2005)
Sema4a	CD4, Th1, Th2 (vs DC)	Co-stimulation, T cell differentiation	(Kumanogoh et al., 2005)
Bcl11b	Th1 (vs DC)	Thymocyte development	(Wakabayashi et al., 2003)
Runx3	B cell (vs CD4)	Transcription factor, promotes Th1 cell fate	(Djuretic et al., 2007)
Klf2	DC (vs Th1)	Transcription factor regulating IL-2 production early in T cell activation	(Wu and Lingrel, 2005)
Il10ra	CD4 (vs B cell)	Interleukin-10 receptor, regulation of immune response.	(Pils et al., 2010)
Tnfaip2	DC (vs Th1/Th2)	Pro-inflammatory gene	(Mookherjee et al., 2006)

Cxcr5	CD4 (vs B cell)	Chemokine receptor	(Hardtke et al., 2005; Junt et al., 2005)
Notch 2	Th1,Th2,CD4 (vs DC)	Signalling during immune cell differentiation	(Cheng et al., 2003; Fiorini et al., 2009)
Il16	Th1 (vs DC)	T cell chemoattractant. Regulation of T cell growth.	(Cruikshank and Little, 2008)

Table 3.5.2 Many differentially methylated CGIs are associated with immune system genes

A selection of the immune system genes that display differential CGI methylation is shown although this list is by no means exhaustive. For each example the first cell type shows increased methylation (hypermethylation) compared to the cell type(s) in brackets.

3.5.3 Cell type-specific methylation preferentially occurs at CGIs

MAP-seq identified CGI and non-CGI regions showing cell type-specific methylation. A recent study examining tissue- and cancer-specific methylation using the restriction digestion-based CHARM method reported that the majority of methylation differences occurred not at CGIs but at the edges of CGIs, so-called "CpG island shores" (Irizarry et al., 2009). In order to explore this possibility the location of differential methylation was examined in more detail to see if it occurred at the edges of CGIs. I categorised my differentially methylated regions as occurring; in a CGI (any overlap with a CGI), 0-2kb from a CGI (not overlapping a CGI but within 2kb of a CGI, counted as a shore region by Irizarry et al.) or greater than 2kb from a CGI (>2kb). The greatest number of differentially methylated regions belonged to the >2kb category (Figure 3.5.2B) but regions more than 2kb away from a CGI also made up the majority of genomic regions that could be identified by MAP-seq (Figure 3.5.2A). For each category of genomic region, the number of differentially methylated regions was expressed as a percentage of MAP-seq regions interrogated. The reference set of MAP-seq regions interrogated was obtained by combining MAP-seq peaks from all mouse MAP-seq experiments carried out in the

lab[vii] (Figure 3.5.2A). Using this set as a reference ensures that these regions can be identified by MAP and are amenable to detection by Illumina sequencing (i.e. not filtered out by repeat masking).

When all regions showing differential methylation between immune cell types were examined, 3.1% of CGI regions showed a change in methylation while just 0.76% of 0-2kb and 0.49% of >2kb regions showed cell-specific methylation differences (Figure 3.5.2B). To verify that this observation was not biased by the small number of methylation changes investigated and that it was not specific to the immune system, methylation changes occurring between CD4 cells and brain were also categorised. In this comparison CGIs still showed preferential differential methylation although the distinction between the categories was less pronounced; 37% of methylated CGI regions, 30% of 0-2kb regions and 18% of regions distal to CGIs showed differential methylation (Figure 3.5.2C). This result suggests a functional importance for differential CGI methylation as methylation differs more at these regions than elsewhere in the genome. The rest of this study will focus on CGI methylation both because differential methylation in immune cells occurs preferentially at CGIs and because cell type-specific CGI methylation is of particular interest. Importantly, if all CGIs are promoters as has been suggested (Illingworth et al., 2010; Maunakea et al., 2010) then methylation at these sites will have an obvious repressive function.

[vii] MAP-seq has been carried out in mouse brain, blood, sperm and as well as the immune cells discussed here.

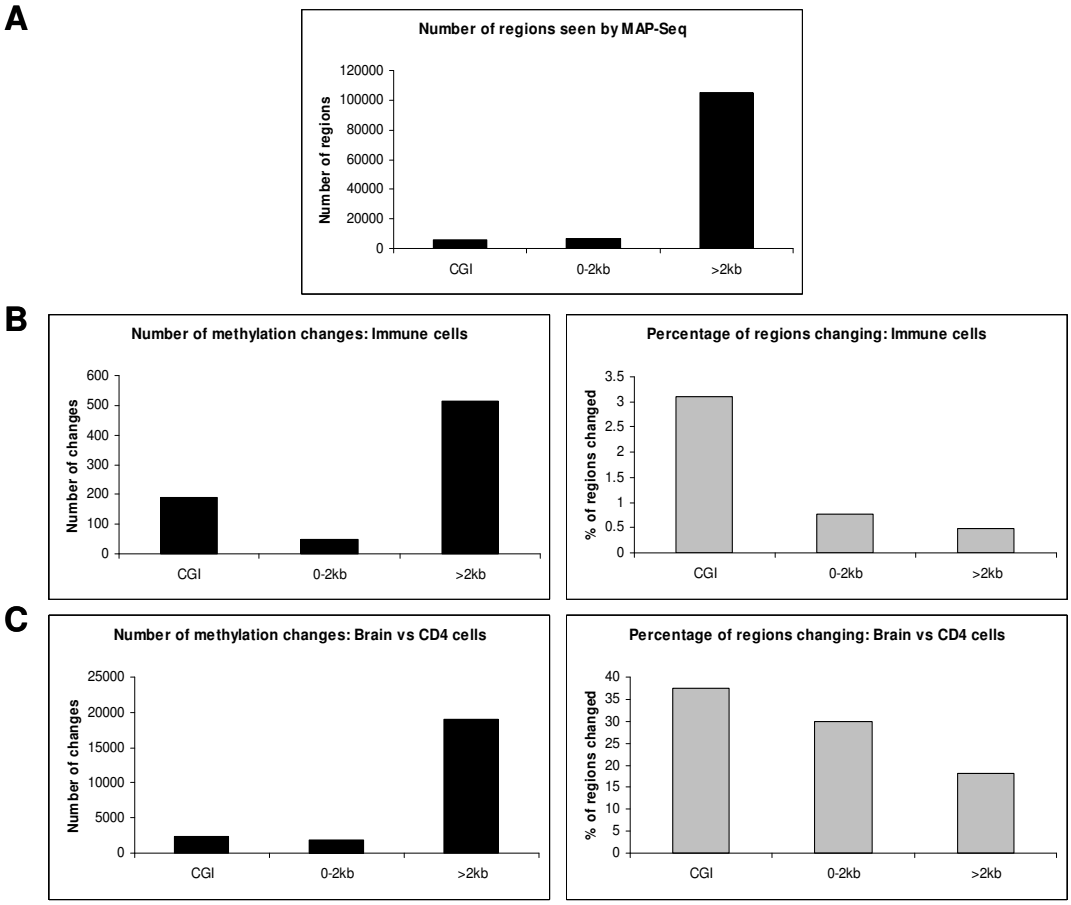


Figure 3.5.2 Cell-specific methylation preferentially occurs at CGIs

A) The number of CGIs, regions 0-2kb and >2kb away from CGIs interrogated by MAP-seq. These figures were used to calculate the percentages in B and C. **B)** Location of differential immune cell methylation with respect to CGIs. **C)** Location of differential methylation between brain and CD4 cells. The left panel shows the absolute number of changes occurring at each category while the right panel shows the percentage of each category showing a methylation change.

3.5.4 Intragenic CGIs show the most methylation changes

CGIs can be located at the transcription start sites of genes (TSS), within gene bodies (intragenic) or outside of annotated genes (intergenic) (Figure 3.5.3A). Approximately half of mouse CGIs are located at a TSS consistent with their function as promoters while the rest are divided equally between intra- and intergenic locations (Figure 3.5.3B). The function of these non-TSS-associated "orphan" CGIs is unclear though many show evidence for promoter function (Illingworth et al., 2010; Maunakea et al., 2010). A recent study comparing CGI methylation in mouse

sperm to that in brain and blood has shown that intra- and intergenic CGIs preferentially acquire methylation in somatic tissues. Approximately 26% of intragenic and 13% of intergenic CGIs became methylated compared to only 2.4% of TSS CGIs (Illingworth et al., 2010). Similar findings have been observed for methylation in human brain (Maunakea et al., 2010). I found that preferential methylation of "orphan" CGIs also occurs in immune cells where 2.8% of TSS CGIs, 33% of intragenic and 16% of intergenic CGIs are methylated in CD4 cells compared to sperm (data not shown). However, I am interested not in CGIs that are constitutively methylated in somatic cells but in those whose methylation varies between cell types. I examined the location of all CGIs that showed cell type-specific methylation in the immune system and found that 63% of these were intragenic despite the fact that only 25% of all CGIs are located in intragenic regions (Figure 3.5.3B). This trend was the same for all immune cell comparisons (Figure 3.5.3C). This makes intragenic CGIs seven times more likely than TSS CGIs and three times more likely than intergenic CGIs to be differentially methylated in the immune system. Therefore, examining intragenic CGIs is crucial with respect to elucidating the function of cell type-specific CGI methylation in the immune system as this is where most of the methylation differences occur. Many of the differentially methylated genes with immune system function mentioned in section 3.5.2 showed changes at intragenic CGIs including *Gata3*, *Kcnn4*, *Sema4a* and *Notch2*.

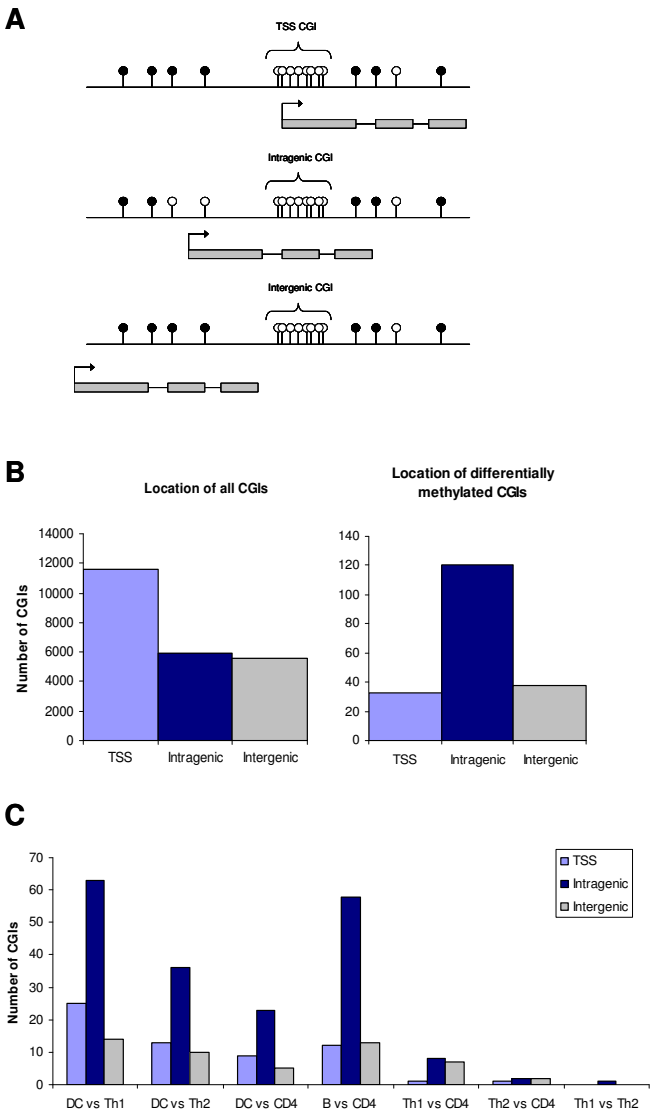


Figure 3.5.3 Most differential CGI methylation in the immune system is intragenic

A) CGIs can be located at transcription start sites (TSS), in gene bodies (intragenic) or between genes (intergenic). CGIs are shown here as unmethylated although intra- and intergenic CGIs are susceptible to methylation. Filled circles = methylated CpGs, empty circles = unmethylated CpGs. **B)** Location of all CGIs in the mouse genome (left; taken from Illingworth et al., 2010) and location of all CGIs that are differentially methylated in immune system cells (right; each CGI was only counted once even if it changed in multiple comparisons. **C)** Pair-wise comparisons of immune cell types showing the number and location of CGIs that are differentially methylated.

It is known that methylation of promoter CGIs is incompatible with transcription yet it is unclear what the effect of CGI methylation in gene bodies might be. As intragenic CGIs have recently been associated with novel TSSs (Illingworth et al.,

2010; Maunakea et al., 2010) it is tempting to speculate that these could represent a distinct class of promoter that is susceptible to regulation by methylation. To try and elucidate the effects of differential intragenic methylation I examined its relationship with gene expression.

3.5.5 Differential intragenic CGI methylation negatively correlates with gene expression

Differential intragenic CGI methylation was associated with differential expression of the associated gene in an average of 26% of cases. For these genes, an increase in methylation tended to correspond to a reduction in gene expression. This was the case for 33 out of 47 (70%) intragenic CGIs showing cell-specific methylation in the immune system (Figure 3.5.4A). Several CGIs did not follow this trend but this is not surprising given that the functional relationship between intragenic CGIs and their associated genes is not clear.

To see if the negative correlation between increased intragenic CGI methylation and gene expression was a general phenomenon, the relationship between differential intragenic CGI methylation and differential gene expression in brain and CD4 cells was examined as a much larger number of CGI methylation changes occur here (2,304). Almost 70% of these changes occurred at intragenic locations and brain showed less methylation than CD4 cells. When a loss of intragenic CGI methylation in brain was accompanied by a gene expression change, expression was increased in brain compared to CD4 cells in the vast majority (82%) of cases (Figure 3.5.4B). Thus, when differential intragenic CGI methylation is associated with a gene expression change, methylation negatively correlates with expression consistent with a repressive effect for intragenic CGI methylation.

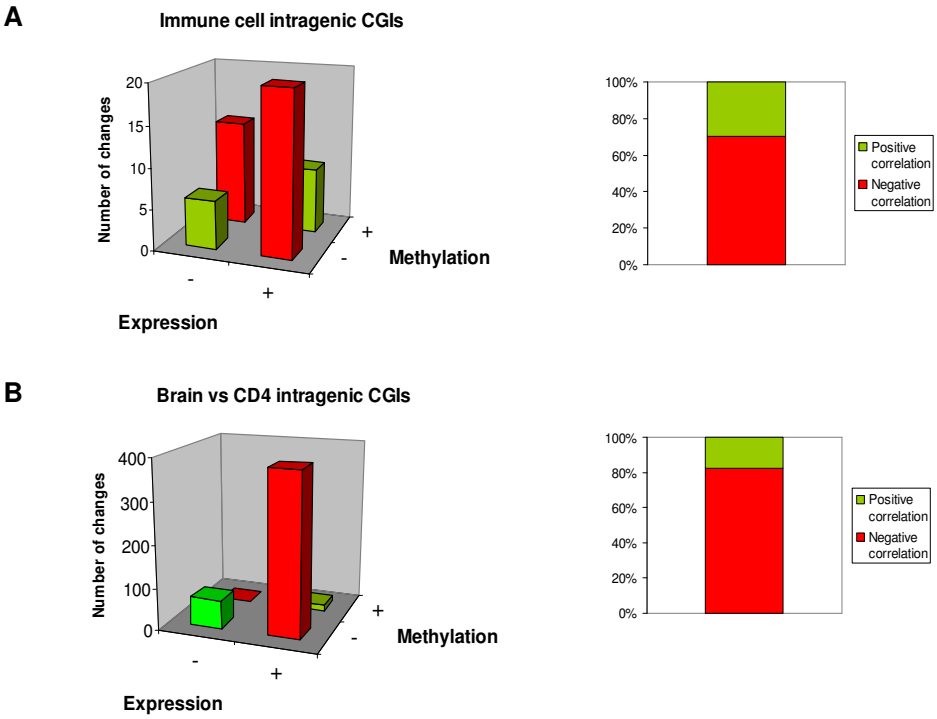


Figure 3.5.4 Cell type-specific intragenic CGI methylation negatively correlates with gene expression

The relationship between differential intragenic CGI methylation and differential gene expression is shown for **A)** All immune cell comparisons **B)** Brain versus CD4 cells. In the left panel the number of changes are shown, "+" denotes an increase and "-"denotes a decrease in expression/methylation. In the right panel the percentage of methylation changes showing a positive or negative correlation with gene expression is shown. Red = negative correlation and green = positive correlation between methylation and expression.

However, most cell-specific intragenic CGI methylation is not associated with a change in gene expression. I examined whether these genes were likely to function in the immune system or whether genes showing both a methylation and expression change showed a greater enrichment for GO terms associated with immune system function. Intragenic CGIs that associated with both cell-specific methylation and cell-specific gene expression were preferentially located in genes with functions in "immunity and defense" and "cytokine/chemokine-mediated immunity" ($p = 0.001$ and $p = 0.00146$ respectively) while differentially methylated genes that did not change their expression showed no enrichment for terms associated with immune system function. This suggests that the CGI methylation changes associated with a gene expression change may be more likely to have a functional role in the immune

system. I then investigated whether the differentially methylated intragenic CGIs not associated with an expression change were located in genes that were expressed. When dendritic cells and Th1 cells were compared, 70% of differentially methylated intragenic CGIs were associated with genes that were silent in both cell types. In contrast, when all genes were examined, 46% were silent in both dendritic cells and Th1 cells. This indicates a tendency for genes associated with differentially methylated intragenic CGIs to be repressed.

As the relationship between intragenic CGIs and the associated gene is unclear it is not known if or how methylation at intragenic sites might impact on gene expression. It is possible that in the instances where cell-specific methylation negatively correlates with gene expression, the intragenic CGI might be involved in transcriptional regulation of the associated gene but in other cases its function may be completely independent of this gene.

3.6 Results: Association between differential methylation and histone modification

One way to explain the negative correlation between intragenic CGI methylation and gene expression is that unmethylated CGIs facilitate transcription through the setting up of a permissive chromatin environment in the gene body which is lost when the CGI becomes methylated. Unmethylated CGIs are associated with the active histone modification H3K4me3 (Illingworth et al., 2010) which is a common feature of promoters and deemed to be a mark of transcriptional competence. H3K4me3 is recruited to CGIs, at least in part, by the CxxC domain protein CFP1 (Thomson et al., 2010). I examined examples of differential intragenic CGI methylation occurring in immune system genes and investigated how this impacted the chromatin state. Results obtained from analysis of candidate genes prompted us to look genome-wide at H3K4me3 patterns in Th1 and dendritic cells.

3.6.1 Differential intragenic CGI methylation coincides with a difference in H3K4me3 at *Kcnn4* and *Gata3*.

I examined methylation, gene expression and H3K4me3 levels for a number of intragenic CGIs that show cell type-specific methylation in the immune system. An intragenic CGI in the *Kcnn4* gene which codes for a potassium channel involved in T cell activation, shows increased expression and decreased methylation in Th1 cells compared to dendritic cells and B cells (Figure 3.6.1A). I verified the MAP-seq methylation data using bisulfite sequencing and found that the *Kcnn4* CGI was 20% methylated in Th1 cells but 66% and 70% methylated in dendritic and B cells respectively (Figure 3.6.1B). The gene expression difference between the cell types was confirmed using qRT-PCR (Figure 3.6.1C). H3K4me3 was then profiled across the *Kcnn4* intragenic CGI using ChIP followed by qPCR. In Th1 cells where the CGI is hypomethylated, there is a peak of H3K4me3 enrichment over the CGI. However, H3K4me3 is not enriched over this CGI in B cells and dendritic cells where it is heavily methylated (Figure 3.6.1D). There is also a methylation difference at the *Kcnn4* intragenic CGI between CD4 cells and Th1/Th2 cells; consistent with the role of *Kcnn4* in T cell activation (20% methylated versus 32% methylated, Figure 3.6.1A and B).

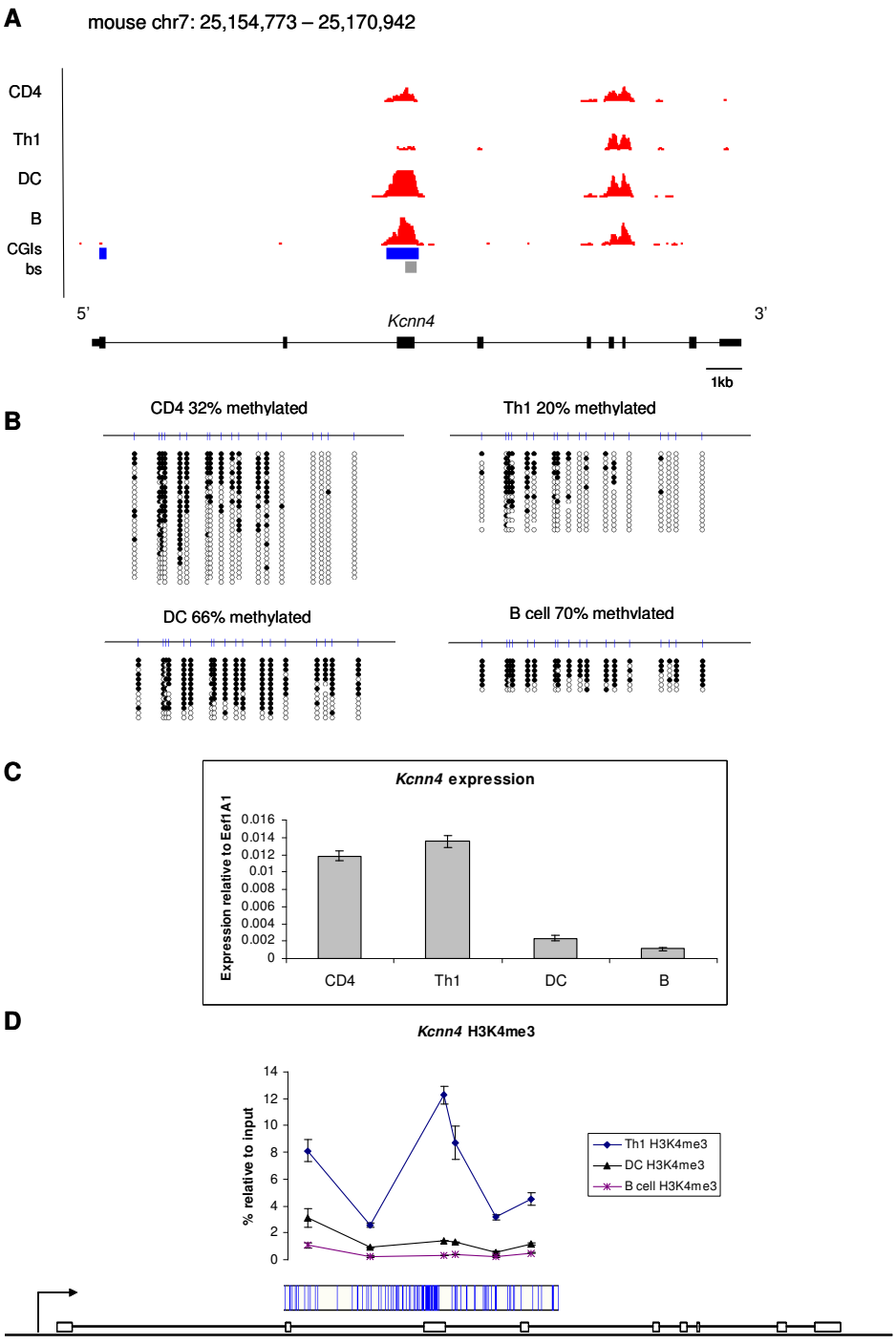
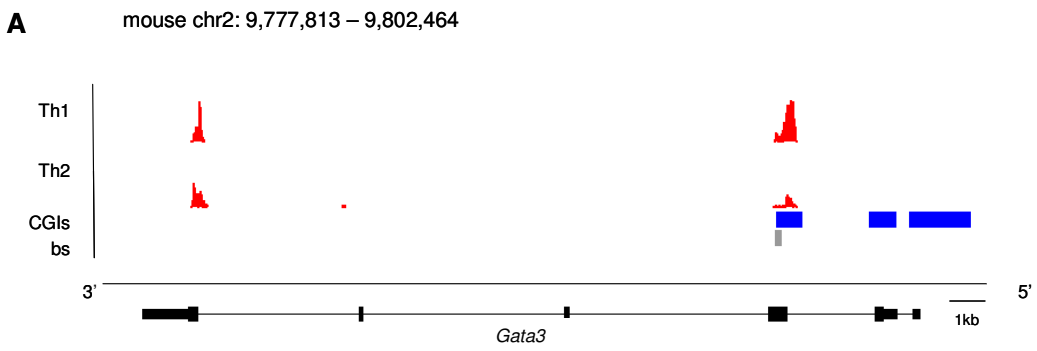
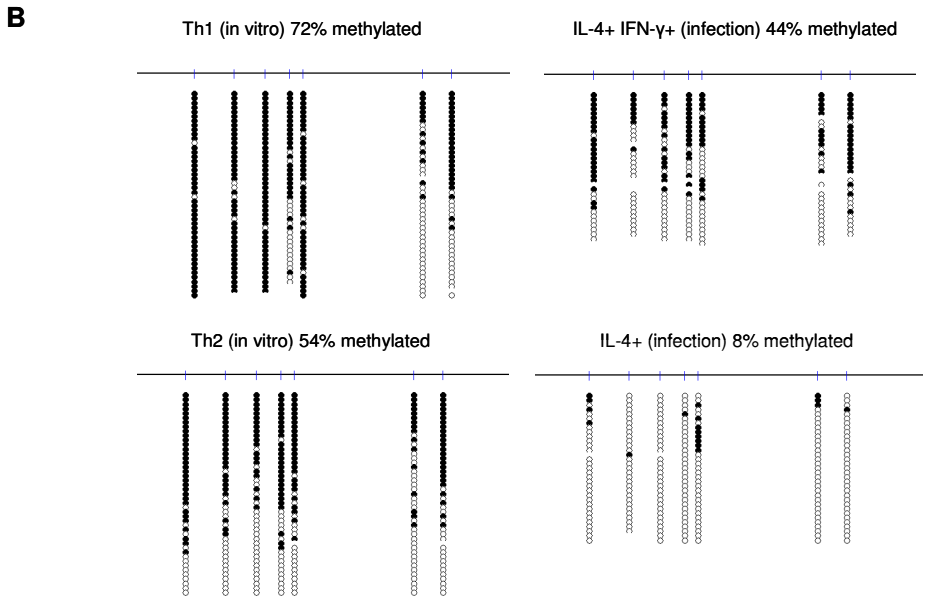


Figure 3.6.1 Decreased CGI methylation corresponds to increased H3K4me3 at the *Kcnn4* intragenic CGI

A) MAP-seq profiles (red) for *Kcnn4* in CD4, Th1, dendritic and B cells, CGIs (blue bars), region examined by bisulfite sequencing (bs, grey bar). **B)** Confirmation of methylation difference by bisulfite genomic sequencing. **C)** Verification of *Kcnn4* gene expression by qRT-PCR. **D)** ChIP for H3K4me3 across the *Kcnn4* intragenic CGI in Th1, dendritic cells and B cells. Blue dashes underneath the profile indicate CpG density and a schematic of the gene is shown with the direction of transcription indicated by the arrow.

A particularly interesting example is that of the *Gata3* intragenic CGI as it shows the only methylation difference detected between Th1 and Th2 cells (Figure 3.6.2A). Bisulfite sequencing was used to more quantitatively assess the extent of this methylation difference. In Th1 cells, the *Gata3* intragenic CGI was 72% methylated and this dropped to 54% in Th2 cells (Figure 3.6.2B; bisulfite sequencing over an additional region of this CGI gave a similar result – 76% methylation in Th1 cells and 57% in Th2 cells) concomitant with a large increase in *Gata3* expression (Figure 3.6.2C). This drop in methylation associated with an increased level of H3K4me3 over the CGI in Th2 cells (Figure 3.6.2D). Interestingly, H3K4me3 over the *Gata3* promoter CGI is unchanged in Th1 and Th2 cells despite a difference in gene expression and this may be explained by reports that the *Gata3* promoter is bivalent (i.e. marked by H3K4me3 and the repressive mark H3K27me3) in Th1 cells (Wei et al., 2009). Although the *Gata3* methylation change in Th2 cells generated *in vitro* is subtle, its biological relevance was confirmed by examining Th2 cells isolated from mice infected with the parasitic worm *Schistosoma mansoni*. CD4+ cells were harvested from the spleens of mice that had been infected for 8 weeks and re-stimulated to produce cytokines. IL-4-producing (i.e. Th2) cells were isolated through intracellular staining and FACS sorting. Bisulfite sequencing of the *Gata3* intragenic CGI in IL-4-producing cells showed complete demethylation (8%) of the region (Figure 3.6.2B). This may indicate that a longer time period is necessary to see methylation changes (the length of the infection is 8 weeks and *in vitro* Th2 cells are differentiated for just one week) which would be consistent with a passive demethylation mechanism. Alternatively, Th2 cells generated in culture may be phenotypically and functionally distinct from *ex vivo* Th2 cells generated in response to a real infection. When cells from the schistosome infection were re-stimulated a population of cells expressing both IL-4 and IFN-γ was generated in addition to IL-4 single positive cells. The function of these cells is unclear although double positive cells have been reported previously (Hegazy et al., 2010). When IL-4+ IFN-γ+ cells were examined for methylation at *Gata3* they were found to be 44% methylated (Figure 3.6.2D) indicating that they represent a distinct population of cells to the IL-4+ cells.


A
mouse chr2: 9,777,813 – 9,802,464
Th1
Th2
CGIs
bs
3'
5'
1kb
Gata3


B
Th1 (in vitro) 72% methylated
IL-4+ IFN-γ+ (infection) 44% methylated
Th2 (in vitro) 54% methylated
IL-4+ (infection) 8% methylated


C
Gata3 expression
Expression relative to EefA1
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
Th1
Th2

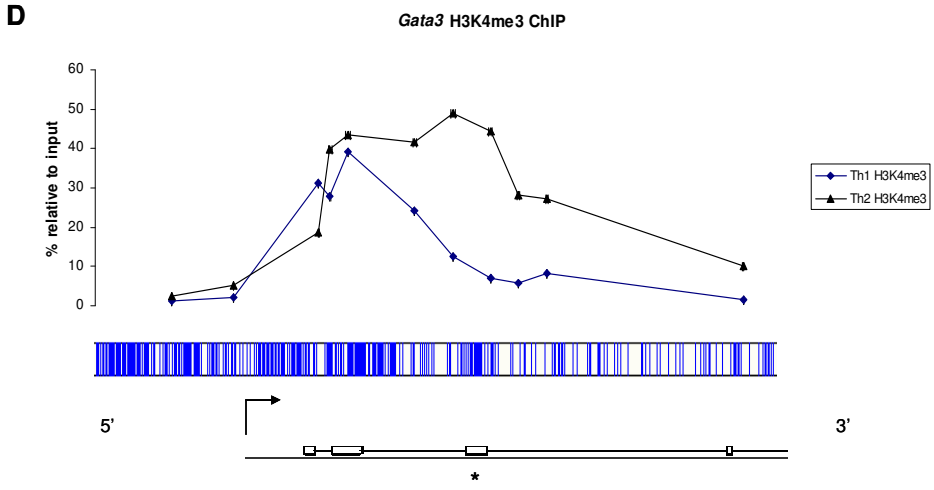

D
Gata3 H3K4me3 ChIP
% relative to input
60
50
40
30
20
10
0
Th1 H3K4me3
Th2 H3K4me3
5'
3'
*

Figure 3.6.2 Demethylation of the *Gata3* intragenic CGI occurs in Th2 cells generated *in vitro*, those isolated from an infection and is associated with increased H3K4me3

A) MAP-seq profiles (red) for *Gata3* in Th1 and Th2 cells. *Gata3* is on the antisense strand running 3' to 5'. **B)** Bisulfite sequencing showing a moderate methylation difference between cultured Th1 and Th2 cells (*in vitro*) and an even greater loss of methylation in IL-4+Ifnγ+ cells and IL-4+ cells from a *Schistosome* infection (infection). **C)** Verification of increased *Gata3* expression in Th2 cells by qRT-PCR. **D)** ChIP for H3K4me3 across the *Gata3* promoter and intragenic CGI in Th1and Th2 cells. Both cell types show high H3K4me3 at the promoter but Th2 cells alone show a peak over the intragenic CGI in exon 3 (asterisk).

3.6.2 Genome-wide analysis of H3K4me3 in dendritic cells and Th1 cells

These observations prompted us to look genome-wide at H3K4me3 in Th1 and dendritic cells to see if increased intragenic CGI methylation always associated with decreased H3K4me3 as is the case at the *Kcnn4* and *Gata3* CGIs. This also allowed us to look more generally at patterns of H3K4me3 at intragenic CGIs. ChIP for H3K4me3 was carried out in both cell types and enrichment was tested by examining the H3K4me3 profile over the *c-myc* locus which contains an unmethylated CGI. The H3K4me3 profile at *c-myc* was virtually identical in Th1 and dendritic cells with minimal variation between experimental replicates (Figure 3.6.3).

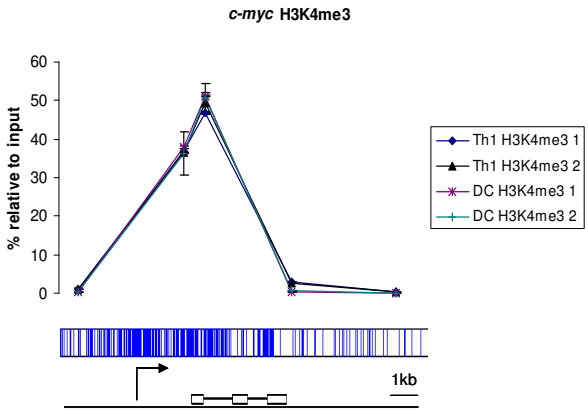


Figure 3.6.3 Verification of H3K4me3 enrichment at *c-myc* prior to high throughput sequencing

ChIP for H3K4me3 in dendritic cells and Th1 cells was tested by examining enrichment over the *c-myc* CGI. Blue dashes underneath the profile indicate CpG density and a schematic of the gene is shown with the direction of transcription indicated by the arrow.

Illumina adaptors were ligated to the ChIP DNA using a protocol optimised in our lab for attachment of adaptors to small amounts of DNA. One lane of sequencing was carried out for each replicate as described previously and, as before, replicate lanes were combined to give two lanes of sequencing for each cell type. Analysis was performed on .wig files in a similar manner to that described for MAP-seq data in section 3.3. In brief, each sample was normalised to the average amount of sequence across all the samples and prominent peaks of enrichment identified. To identify regions containing differential H3K4me3 enrichment a sliding window analysis over regions of interest was performed. Analysis parameters are outlined in Table 3.6.1.

Analysis step	**Parameters**
Normalisation	Scaled to 1.025×10^9 reads
Peak-finding	H: 4 L: 150 G: 250
Sliding window	100bp window, 20bp slide
Identification of differences	log2 difference > 2 in 9/13 contiguous windows

Table 3.6.1 Parameters for analysis of H3K4me3 ChIP-seq data

Parameters for each analysis step are shown. H = height in number of reads, L = length in bp, G = gap in bp.

Intragenic CGIs show tissue-specific H3K4me3

88% of TSS CGIs but just 35% of intragenic CGIs associate with a peak of H3K4 tri-methylation in dendritic cells and a similar percentage associate with H3K4me3 in Th1 cells (Figure 3.6.4A). This is consistent with the fact that intragenic CGIs are more frequently methylated than TSS CGIs as dense DNA methylation is not compatible with H3K4me3 (Illingworth et al., 2010; Thomson et al., 2010) although this is not accounted for by methylation alone. I examined whether intragenic CGIs that are not marked by H3K4me3 in immune cells are marked by H3K4me3 in other cell types by examining H3K4me3 data for brain (Thomson et al., 2010). 13% of intragenic CGIs that were negative for H3K4me3 in dendritic cells were positive for H3K4me3 in brain and, similarly, 11.5% of intragenic CGIs lacking H3K4me3 in brain were marked by H3K4me3 in dendritic cells (Figure 3.6.4B). This

demonstrates that intragenic CGIs tend to show tissue-specific methylation. In contrast, the vast majority of TSS CGIs in both dendritic cells and brain were marked by H3K4me3 (88% and 92% respectively) with most TSS CGIs positive for H3K4me3 in both cell types (Figure 3.6.4B).

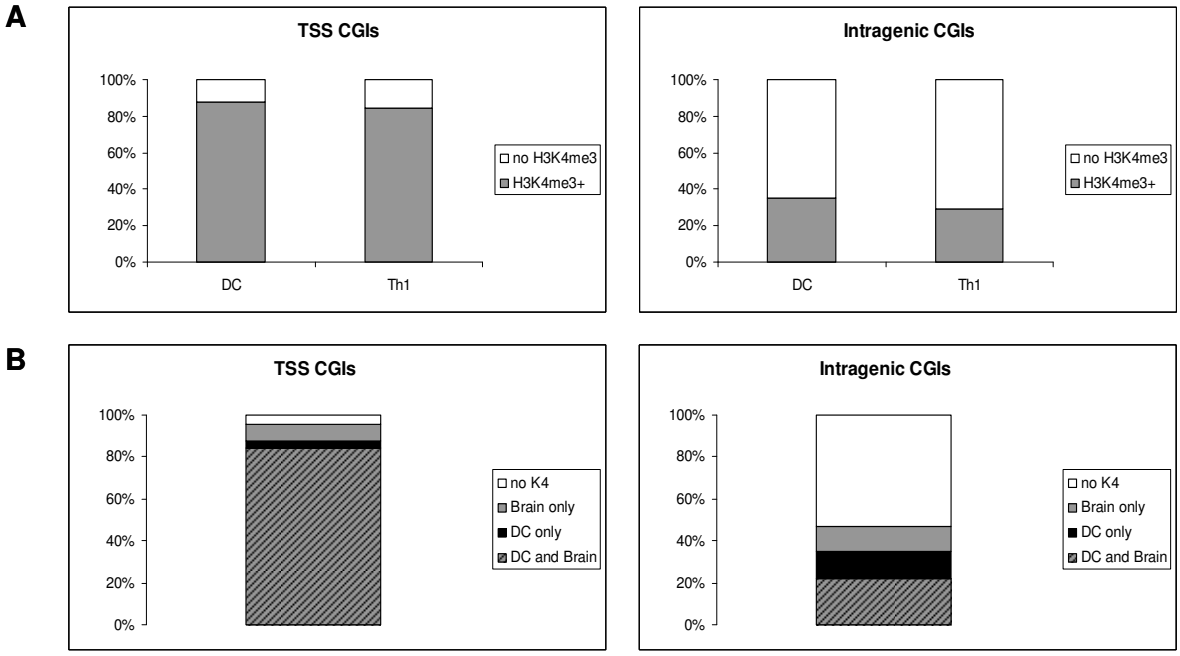


Figure 3.6.4 Intragenic CGIs show tissue-specific H3K4me3

A) Association of TSS CGIs (left) and intragenic CGIs (right) with H3K4me3 in dendritic cells (DC) and Th1 cells. **B)** Most TSS CGIs associate with H3K4me3 in both dendritic cells and brain (left) while intragenic CGIs show more tissue-specific H3K4me3 (right).

Increased intragenic CGI methylation does not always associate with a decrease in H3K4me3

I looked to see if differential methylation at intragenic CGIs tended to associate with changes in H3K4 tri-methylation. 63 intragenic CGIs are differentially methylated in dendritic cells compared to Th1 cells and for 11 of these increased DNA methylation associated with a decreased level of H3K4me3 similar to what was seen at the *Kcnn4* and *Gata3* intragenic CGIs (Figure 3.6.5). Four of these CGIs showed an associated decrease in gene expression, for example *Kcnn4*, but the rest occurred in genes whose expression did not change between cell types. Surprisingly, the majority of differentially methylated intragenic CGIs (55% or 35 CGIs) lacked H3K4me3 in

both Th1 and dendritic cells indicating that these CGIs are not permissive for transcription in either cell type despite a difference in DNA methylation (Figure 3.6.5). In addition, some cell-specific methylation occurred at CGIs positive for H3K4me3 in both immune cell types (Figure 3.6.5). This observation is, at first glance, contradictory to reports that H3K4me3 is not compatible with DNA methylation (Illingworth et al., 2010). However, the majority of these CGIs appear to be compositely methylated (70% based on CAP-seq data for CD4 T cells; data not shown) meaning that they contain unmethylated CpGs that can associate with H3K4me3 and methylated CpGs that cannot.

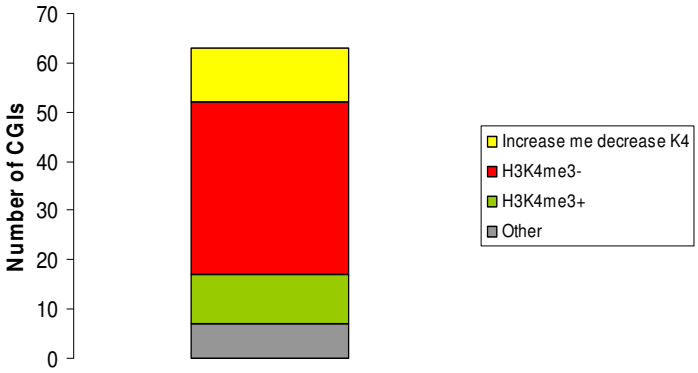


Figure 3.6.5 Increased intragenic CGI methylation does not always associate with decreased H3K4me3

The number of differentially methylated intragenic CGIs where; increased methylation is associated with a decrease in H3K4me3 (yellow), the CGI is negative (red) or positive (green) for H3K4me3 in both cell types or either a non-significant change in H3K4me3 is observed (n=5) or increased methylation is associated with increased H3K4me3 (n=2) (grey).

3.6.3 Differential methylation at intragenic CGIs can reflect a difference in silencing mechanism

Differential methylation frequently occurs at CGIs that lack the "active" chromatin mark H3K4me3. For example, a CGI at the 3' end of the *Sema4a* gene lacks H3K4me3 in both dendritic cells and Th1 cells despite being differentially methylated (Figure 3.6.6A and B). This suggests that this CGI is silenced in both cell types. An alternative mechanism of repression to DNA methylation is polycomb-mediated silencing characterised by the presence of H3K27me3. Thomson and

colleagues found that 58% of CGIs lacking H3K4me3 in mouse brain were marked by H3K27me3 (Thomson et al, 2010). In addition, studies have shown that H3K27me3 and DNA methylation of CGIs tend to be mutually exclusive (Fouse et al., 2008; Mohn et al., 2008).

Therefore, I used ChIP to look for presence of the repressive mark H3K27me3 at the *Sema4a* intragenic CGI. In dendritic cells, H3K27me3 was detected across the *Sema4a* CGI but this was absent in heavily methylated Th1 cells (Figure 3.6.6D). Therefore, this intragenic CGI seems to be repressed by different mechanisms in different cell types – DNA methylation in Th1 cells and polycomb in dendritic cells. The *Sema4a* gene is expressed in both Th1 and dendritic cells although expression is lower in Th1 cells where the CGI is methylated (Figure 3.6.6C). Another differentially methylated CGI lacking H3K4me3 occurs in an intron of the *Bcl11b* gene and this was found to behave like the *Sema4a* CGI with the unmethylated CGI marked by H3K27me3 (data not shown).

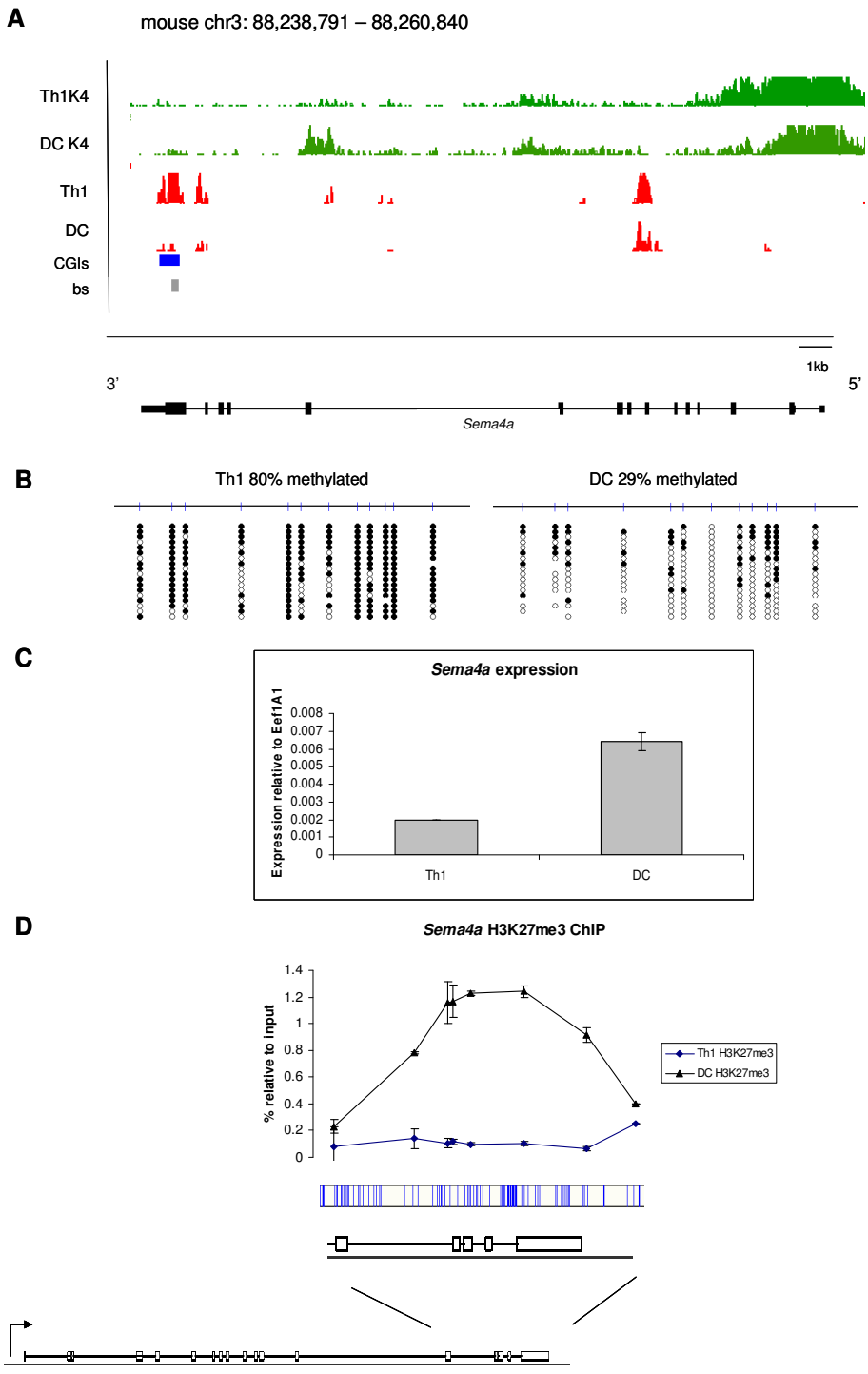


Figure 3.6.6 The *Sema4a* intragenic CGI is repressed by DNA methylation in Th1 cells and H3K27me3 in dendritic cells

A) H3K4me3 ChIP-seq (green) and MAP-seq profiles (red) for *Sema4a* in Th1 and dendritic cells, CGIs (blue bars), bisulfite amplicon (grey bar). *Sema4a* is on the antisense strand running 3' to 5'. **B)** Bisulfite sequencing confirming the methylation difference between Th1 cells and dendritic cells. **C)** *Sema4a* expression in Th1 and dendritic cells quantified using qRT-PCR. **D)** ChIP for H3K27me3 across the CGI.

3.7 Intragenic CGIs showing immune-cell specific methylation may function as promoters in other cell types

It has been proposed that intragenic CGIs act as alternative promoters for the associated annotated gene (Maunakea et al., 2010) and this would explain the observation that cell-specific methylation of these CGI sometimes correlates with gene repression. At some intragenic CGIs, for example those associated with *Kcnn4* and *Gata3*, an association between decreased DNA methylation, increased gene expression and increased H3K4me3 was observed. These CGIs represent good candidates to act as alternative promoters in immune cells. However, such examples represent the minority of cases (4 out of 63 differentially methylated intragenic CGIs when dendritic cells and Th1 cells are compared). Based on the fact that the majority of differentially methylated intragenic CGIs do not possess H3K4me3 it is unlikely that they act as promoters in the immune cells studied. However, these CGIs may act as promoters for genes expressed elsewhere, for example during development or in other tissue types. To investigate this possibility I examined whether there was evidence for promoter function for these intragenic CGIs using Pol II ChIP-seq data for ES cells (Illingworth et al., 2010) and CAGE tag data from a panel of mouse tissues (Faulkner et al., 2009). Nearly half of intragenic CGIs differentially methylated in dendritic cells and Th1 cells (30 out of 63) showed Pol II binding in ES cells (which are reflective of an early developmental stage) or showed a cluster of CAGE tags, indicative of the 5' end of a transcript, in a library compiled from a panel of tissues that did not include the immune cell types investigated here[viii] (Figure 3.7.1).

[viii] The CAGE tag set consisted of data from brain, urogenital tissue, embryonic fibroblasts, muscle, hippocampus, liver, lung, hepatoma, embryonic testis, neuroblastoma, adipose tissue, whole embryo and monocytes.

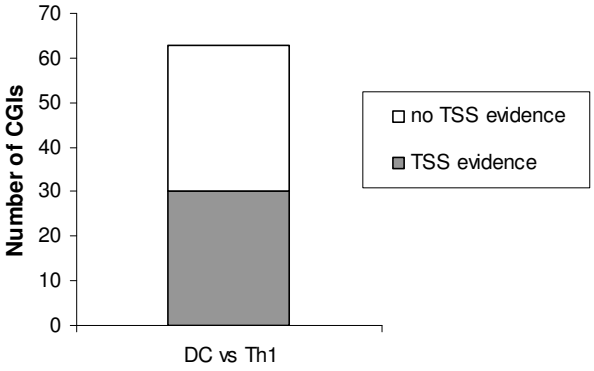


Figure 3.7.1 Intragenic CGIs differentially methylated in the immune system show evidence for promoter function in other tissues

Intragenic CGIs differentially methylated in dendritic cells and Th1 cells (DC vs Th1) that show evidence for function as a TSS based on ES cell Pol II ChIP-seq data or CAGE data from a panel of tissues.

3.8 Discussion

In light of a number of recent reports on tissue-specific CGI methylation (Illingworth et al., 2008; Meissner et al., 2008; Mohn et al., 2008; Schilling and Rehli, 2007) this chapter has examined developmentally related somatic cell types for methylation differences and investigated how these relate to gene expression. Cells of the immune system were used as a model partly because there have been a number of reports of cell-type specific methylation in this lineage (Ivascu et al., 2007; Schmidl et al., 2009) and also because pure immune cell types can easily be isolated.

3.8.1 DNA methylation differences reflect developmental similarity rather than the gene expression programme

MAP-seq gave robust methylation profiles of immune system cells which covered methylated CGIs as well as other CpG-rich methylated regions. Samples were normalised in order to account for variable read depth which allowed quantitative assessment of methylation differences between cell types. Methylation differences between closely related immune cell types were small in number, both at CGIs and non-CGI regions, but may be functionally important given the role of many differentially methylated genes in the immune system. Methylation differences

showed an inverse correlation with developmental similarity. Closely related Th1 and Th2 cells showed just a single methylation difference whilst more differences were observed between more distantly related immune cells types. Examination of a developmentally distinct tissue, brain, revealed thousands of differences compared to immune cells. Gene expression differences, however, did not follow the same trend. Even closely related cells types such as naïve CD4 and Th1 cells showed a considerable difference in their gene expression profiles. Given the sizable number of genes that show differential expression in immune cells and the small number of methylation changes that occur, it is clear that differential methylation at CpG-rich regions does not make a major contribution to regulation of gene expression in immune cells.

It is possible that methylation follows developmental similarity because methylation patterns are established over the extended time period of development. The large gene expression changes associated with immune cell activation are likely to be rapid and dependent on transcription factors whereas such a short time frame is insufficient to allow extensive methylation changes to take place. However, differential methylation of a small number of key immune system genes may be important as indicated by the fact that genes showing cell-specific methylation often have immune system function.

3.8.2 Cell-specific methylation occurs preferentially at CGIs

Given that MAP-seq identifies both CGI and non-CGI-associated methylated regions and in light of a recent report claiming that the majority of tissue-specific methylation occurs at the edges of CGIs (Irizarry et al., 2009), I examined the genomic location of the methylation changes. When the small number of methylation changes occurring in the immune system were examined these preferentially occurred in CGIs rather than at CGI "shores". As this is a much smaller set of changes than those examined by Irizarry et al., I also examined the location of differential methylation between brain and CD4 cells. Here, differential methylation was also found to be preferentially located in CGIs. Differences between my

observations and those of Irizarry and colleagues could be due the fact different techniques were used to assess methylation, that different sets of features were interrogated (i.e. the features present on a microarray compared to all genomic regions amenable to analysis by MAP-seq) or the fact that I have related the data to a biochemically identified CGI set rather than a set defined by prediction algorithms. For the remainder of the study I chose to focus on differential methylation at CGIs both because there is enrichment for methylation changes at CGIs and because CGIs represent a distinct and functionally important fraction of the genome.

3.8.3 Intragenic CGIs show the most cell type-specific methylation

Most cell type-specific CGI methylation in the immune system occurs at intragenic CGIs which are located in gene bodies but not associated with an annotated TSS. The functional implications of methylation at these regions are unclear. Promoter CGI methylation has long been associated with gene repression and two recent studies have suggested that many intra- and intergenic CGIs represent novel promoters (Illingworth et al., 2010; Maunakea et al., 2010). Intragenic and intergenic CGIs seem to be distinct with respect to cell type-specific methylation as intragenic CGIs more frequently show differential methylation. I examined whether differential intragenic methylation affects expression of the associated gene despite the fact that it is not located at the TSS. When cell-specific intragenic CGI methylation associated with a change in gene expression, in general, increased methylation correlated with decreased gene expression suggesting a repressive role for intragenic CGI methylation.

3.8.4 Intragenic CGIs showing differential H3K4me3 may act as alternative promoters or enhancers

Why intragenic CGI methylation associates with transcriptional repression is unclear given that the relationship between intragenic CGIs and their associated genes is not known. I examined if differential methylation at these regions is associated with an altered chromatin state. For a number of intragenic CGIs (17.5%) decreased DNA

methylation was accompanied by an increase in H3K4 tri-methylation. H3K4me3 can be targeted to unmethylated CGIs by CFP1 (Thomson et al., 2010). H3K4me3 is thought to contribute to a transcriptionally permissive chromatin environment in a number of ways including the recruitment of activities involved in histone acetylation and ATP-dependent chromatin remodelling (Kim et al., 2006; Li et al., 2007; Ruthenburg et al., 2007). It has also been reported to target the basal transcription machinery (TFIID) to nucleosomes *in vitro* (Vermeulen et al., 2007). How might this chromatin structure affect transcription if it is not located at an annotated transcription start site? It has been suggested that the majority of intragenic CGIs act as alternative promoters for the associated gene (Maunakea et al., 2010). Indeed, this has been found to be the case for an exonic CGI in the MHC class II I-Aβ gene (Macleod et al., 1998) and intragenic CGIs in the *SHANK3* and *Nfix* genes (Maunakea et al., 2010). It is possible that this may also be the case for intragenic CGIs such as those located in the *Kcnn4* and *Gata3* genes where a decrease in methylation is associated with an increase in H3K4me3 and an increase in gene expression. ChIP for the hypophosphorylated form of Pol II that is recruited to the promoter (Brookes and Pombo, 2009), however, gave low signal over the *Kcnn4* and *Gata3* CGIs (data not shown). ChIP with alternative Pol II antibodies or 5' RACE to map transcription start sites will be necessary to determine if these CGIs do indeed act as promoters.

Alternatively, intragenic CGIs may function as regulatory elements such that an open chromatin conformation helps to promote transcription. Recent studies have classified the chromatin state of enhancers occupied by the co-activator p300 and found them to lack H3K4me3 but be enriched for H3K4me1 (Heintzman et al., 2007). Therefore, if intragenic CGIs have enhancer function they must represent a distinct class of enhancer which is marked by H3K4me3 and located downstream of the transcription start site. Indeed, a study last year identified differentially methylated regions in immune cells located outside of promoters but often associated with differential H3K4me3. Some of these regions showed evidence for enhancer function in reporter assays that was lost upon methylation (Schmidl et al., 2009).

3.8.5 Many intragenic CGIs showing immune cell-specific methylation may act as promoters in other cell types

The majority of differentially methylated intragenic CGIs lack H3K4me3 in immune cells, however, implying that they do not act as promoters (or indeed enhancers) in the immune system. It is possible that these CGIs act as promoters in other tissues or during development. Indeed Pol II binding and CAGE tag data suggest that this is the case for many of these differentially methylated intragenic CGIs, consistent with reports that the majority of intragenic CGIs associate with sites of transcriptional initiation (Illingworth et al., 2010; Maunakea et al., 2010). An attractive hypothesis is that expression in early development could help keep intragenic CGIs free of methylation and thus maintain their CpG density but that a developmentally restricted or highly tissue-specific expression pattern makes them susceptible to methylation in somatic cells. Intragenic CGIs as well as showing cell type-specific DNA methylation in the immune system are also frequently subject to tissue-specific H3K4me3. For example, I have observed that many intragenic CGIs show differential H3K4me3 between dendritic cells and brain. These observations imply that intragenic CGIs represent a distinct category of genomic element more susceptible to tissue-specific regulation than TSS CGIs which frequently associate with ubiquitously expressed genes. It is tempting to speculate that many of these CGIs may act as promoters for ncRNAs or genes with restricted expression patterns which would explain why they have escaped annotation by conventional means.

For intragenic CGIs lacking H3K4me3 differential methylation can reflect the use of alternative silencing mechanisms. For example the *Sema4a* 3' CGI and a CGI in an intron of the *Bcl11b* gene are marked by H3K27me3 in dendritic cells but subject to DNA methylation in Th1 cells. This is consistent with previous reports that at CGIs H3K27me3 and DNA methylation tend to act as mutually exclusive silencing mechanisms (Fouse et al., 2008; Mohn et al., 2008).

3.8.6 Intragenic CGI methylation may affect gene expression by an indirect mechanism

A key question remaining is why cell-specific intragenic methylation often associates with gene repression. For intragenic CGIs acting as alternative promoters the reason for this association is clear, however, most differentially methylated CGIs identified here do not seem to act as promoters in immune cells. Intriguingly, it has been reported that transcription elongation through densely methylated CpGs is less efficient than through unmethylated CpG-rich sequences. Lorincz and colleagues inserted a transgene encoding GFP specifically methylated or unmethylated in the intragenic region into a defined genomic site. Expression of the transgene was reduced by 60% when it was methylated and, consistent with this, levels of elongating Pol II were also reduced. Methylation coincided with a four-fold decrease in histone acetylation and formation of a more compact chromatin structure (Lorincz et al., 2004). This may have been due to the recruitment of MBD proteins to methylated DNA which could bring in HDACs and chromatin remodelling activities and result in formation of a less accessible chromatin environment which inhibits elongation by Pol II. Methylation of intragenic CGIs might also have a similar effect and, in some cases, impair transcription elongation through creation of a repressive chromatin environment. It is also possible that densely methylated DNA physically impairs the passage of Pol II. Such dense intragenic CpG methylation may have different effects to the relatively sparse methylation of the bulk genome where gene body methylation has been reported to associate with gene activity (Ball et al., 2009; Deng et al., 2009; Rauch et al., 2009). Since many intragenic CGIs have evidence for promoter function one hypothesis is that they differentially acquire methylation as a means of repressing a novel promoter which can sometimes influence expression of the associated annotated gene. Mechanistic studies are needed to fully elucidate the role of dense intragenic methylation on transcription.

A speculative model to explain the relationship between intragenic CGI methylation and gene expression is illustrated in Figure 3.8.1. One hypothesis is that intragenic CGIs are promoters for genes expressed only at a particular developmental stage

(Figure 3.8.1A) and that these CGIs are silenced in somatic tissues resulting in a loss of H3K4me3 and a gain in either DNA methylation (Figure 3.8.1B) or H3K27me3 (3.8.1C). Intragenic DNA methylation might then, in certain cases, impair transcription of the associated annotated gene (Figure 3.8.1B) leading to a methylation-dependent reduction in gene expression. However, this model does not take into account the fact that polycomb proteins may also impair transcriptional elongation (Simon and Kingston, 2009).

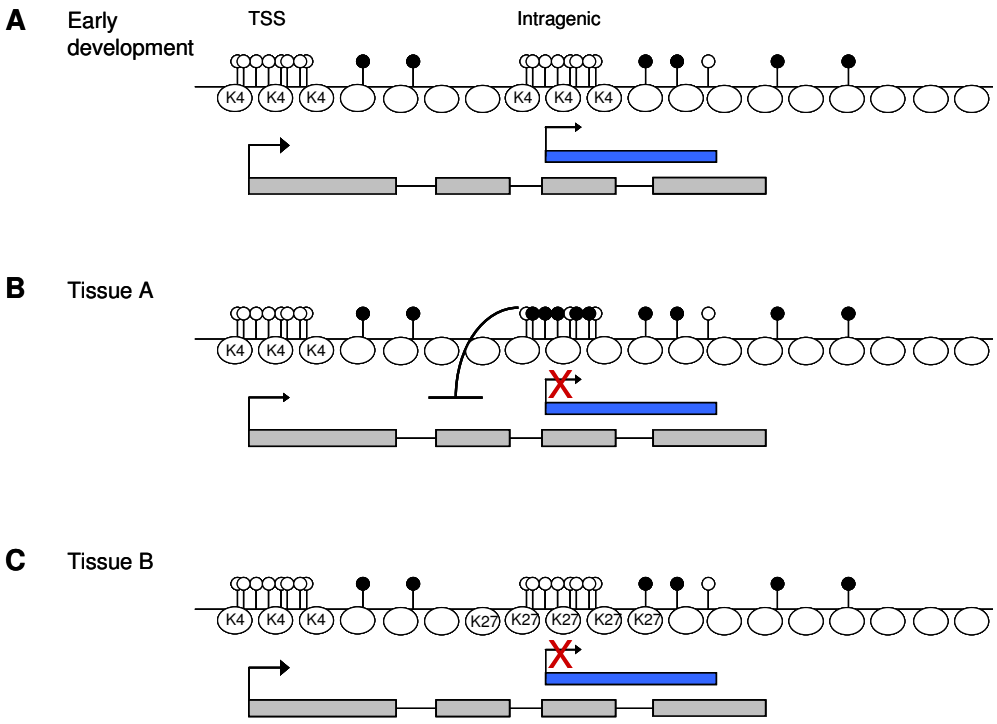


Figure 3.8.1 A possible model to explain differential intragenic CGI methylation

TSS CGIs are promoters for housekeeping genes or genes with broad expression patterns (grey) and are generally unmethylated and marked by H3K4me3 while intragenic CGIs act as promoters for genes expressed in early development or in particular cell-types (blue). **A)** In early development the intragenic CGI is unmethylated, marked by H3K4me3 and is an active promoter for this novel gene. **B)** The intragenic CGI is silenced by DNA methylation in a tissue-specific manner which then interferes with transcription of the associated annotated gene. **C)** Alternatively, the CGI can be silencing by polycomb (H3K27me3 mark). K4 = H3K4me3, K27 = H3K27me3. The arrow indicates a promoter and the direction of transcription.

I have observed that most cell type-specific methylation occurs at CGIs located in intragenic regions and that this is often associated with gene repression. It is likely

that there are a number of reasons for this phenomenon but it is clear that intragenic CGIs represent a distinct class of genomic element particularly susceptible to cell type-specific regulation.

Chapter 4 Regulation of gene expression by MBD2 in immune system cells

4.1 Introduction

4.1.1 MBD2 is involved in gene regulation in the immune system

One of the ways in which methylation exerts its repressive effects is through MBD proteins which specifically recognise methylated DNA and recruit chromatin modifying activities. Surprisingly, depletion of MBD1 or MeCP2 does not lead to many changes in genes expression (Nuber et al., 2005; Tudor et al., 2002; Zhao et al., 2003). Knockout of MBD2 on the other hand results in gene expression changes in all systems tested to date, consistent with a role as a methylation-dependent transcriptional repressor (Barr et al., 2007; Berger et al., 2007; Hutchins et al., 2002). However, *Mbd2-/-* mice are viable and fertile suggesting that MBD2 may regulate genes involved in inducible processes such as the immune response. In keeping with this, a number of immune system defects have been observed in *Mbd2-/-* mice. *In vitro* T helper cell differentiation is disordered with *Mbd2-/-* Th1 and Th2 cells aberrantly expressing both IFN-γ and IL-4. MBD2 has been reported to directly silence the *Il4* gene by competing for binding with the Th2 transcription factor Gata-3 (Hutchins et al., 2002). A follow on study looking at Th1- and Th2-inducing infections *in vivo* showed that *Mbd2-/-* mice were defective in mounting a Th2 response against the intestinal worm *Trichuris muris* but showed enhanced immunity to a Th1-type *Leishmania* infection due to excessive IFN-γ production (Hutchins et al., 2005). Defects in the production of CD8 memory T cells in *Mbd2-/-* mice have been reported (Kersh, 2006) and impaired induction of Th2 responses by *Mbd2-/-* dendritic cells have also been observed (P. Cook; unpublished results).

4.1.2 Elucidating MBD2 function

The extent of MBD2's role in regulating gene expression and the mechanisms governing this remain to be elucidated. A major question is the degree to which gene expression is perturbed when MBD2 is removed from a physiologically relevant cell type and whether this has an effect on chromatin modification. MBD2 exerts its repressive effect via the NuRD chromatin remodelling and HDAC complex so it is anticipated that changes in gene expression in the absence of MBD2 will also be accompanied by changes in the chromatin state. In addition, it is unclear whether different patterns of DNA methylation between cell types instruct MBD2 to regulate distinct sets of target genes. In many studies cancer cell lines have been used to look for MBD2 targets (Auriol et al., 2005; Bakker et al., 2002; Lin and Nelson, 2003; Martin et al., 2008; Pulukuri and Rao, 2006). However, as cancer cells frequently show abnormal DNA methylation (Jones et al., 1990; Meissner et al., 2008; Smiraglia et al., 2001) they are not a suitable system for elucidating the role of a methyl-binding protein in a normal cellular context. It is also an advantage to study pure cell types rather than tissues as heterogeneous cell populations could mask the effects of MBD2 deficiency in a particular cell type. Cells of the immune system are primary cell types that can be isolated as relatively homogenous cell populations. In addition, MBD2 has been shown to be important for immune system function and DNA methylation data already exists for immune cells (chapter 3). For these reasons the immune system was chosen as a model in which to study the function of MBD2.

4.1.3 Aim

The aim of this work is to examine gene expression in *Mbd2-/-* immune system cells, determine histone acetylation patterns genome-wide in *Mbd2-/-* compared to wild-type cells as well as to identify where MBD2 binds in the genome and how this relates to DNA methylation. It is hoped that this study will help elucidate the role of MBD2 in transcriptional regulation and its relationship to DNA methylation and histone acetylation.

4.2 Results: Mbd2-/- mice show T cell differentiation defects

4.2.1 Intracellular FACS staining confirms previously reported T cell differentiation phenotype

Experiments were carried out with CD4 T cells harvested from the spleens of wild-type and *Mbd2-/-* C57/BL6 mice in order to confirm reported differentiation defects in the absence of MBD2 (Hutchins et al., 2002). Cells were treated with α-CD3 and α-CD28 to mimic T cell activation as well as IL-2 which is important for T cell proliferation and either: not treated further (Th0 conditions); treated with IL-12 and α-IL-4 (Th1 conditions); or treated with IL-4 and α-IL-12 (Th2 conditions). Cells were cultured for 7 days, re-stimulated and cytokine production was assessed by intracellular FACS staining. In general, wild-type cells showed production of IFN-γ only under Th1 conditions and expression of IL-4 under Th2 conditions. Under both Th1 and Th2 conditions, however, *Mbd2-/-* cells showed a large number of IFN-γ+ IL-4+ double positive cells (15% and 38% respectively) compared to wild-type cells (1.25% and 7.42%). *Mbd2-/-* cells also showed a larger number of double positive cells under Th0 conditions (Figure 4.2.1). Aberrant expression of the Th1 and Th2 signature cytokines during *in vitro* differentiation of *Mbd2-/-* T helper cells has previously been reported (Hutchins et al., 2002). This confirms previous reports that *Mbd2-/-* T cells fail to appropriately silence the *Ifng* and *Il4* genes during differentiation (Hutchins et al., 2002). It was also noted that proliferation of *Mbd2-/-* cells during T cell differentiation was reduced compared to wild-type cells (see section 4.3.1).

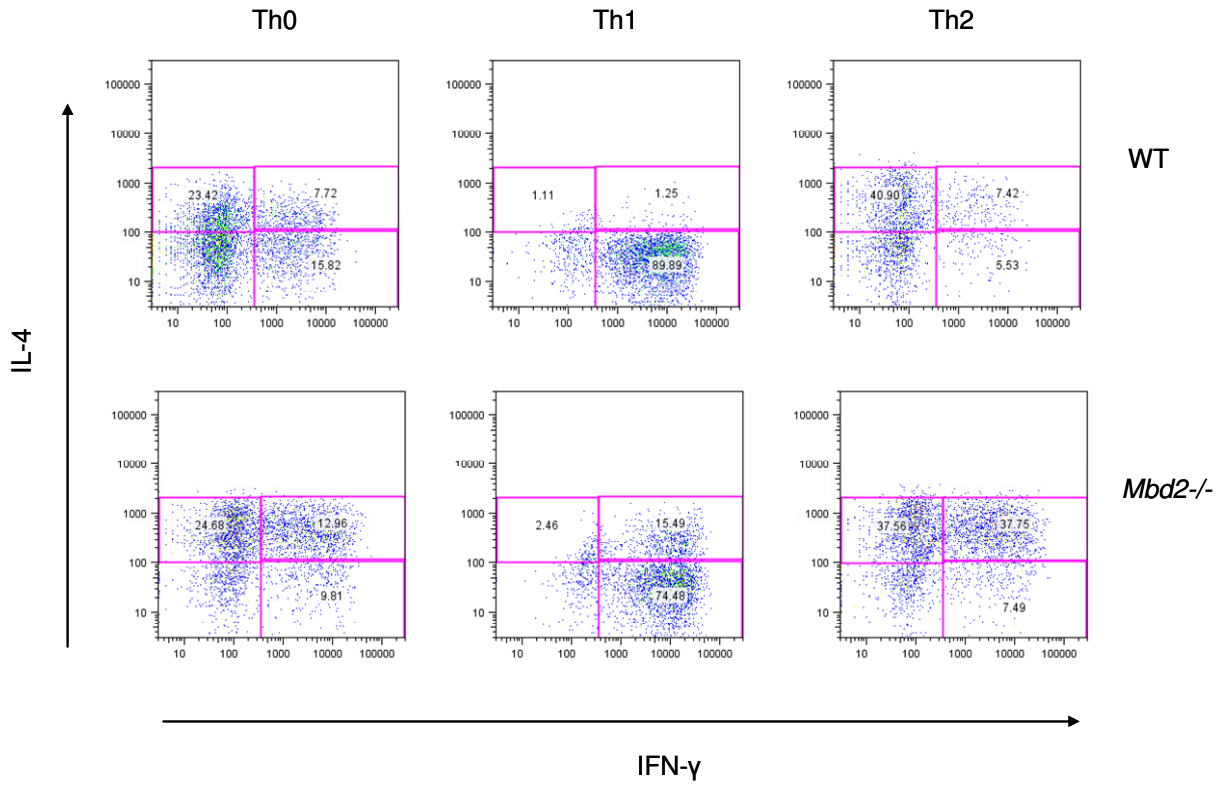


Figure 4.2.1 Recapitulation of T cell differentiation defects in *Mbd2-/-* cells

Results of T cell differentiation under different conditions (Th0, Th1, Th2) are shown for wild-type cells (WT; top) and *Mbd2-/-* cells (*Mbd2-/-*; bottom). Staining for IFN-γ production is shown on the x-axis with IL-4 on the y-axis. IFN-γ+ IL-4+ double positive cells are located in the top right hand quadrant.

4.2.2 qRT-PCR shows that the *Ifng* and *Il4* genes are mis-expressed in the absence of MBD2

RNA was extracted from Th1 and Th2 cell cultures as well as from naïve CD4 cells and the expression levels of *Ifng* and *Il4* in wild-type and MBD2-deficient animals assessed by qRT-PCR. Although this technique can not assess the proportion of cells expressing both IFN-γ and IL-4 it gives a measure of the magnitude of the difference in *Ifng* and *Il4* RNA levels between wild-type and *Mbd2-/-* cells. As expected, *Mbd2-/-* Th1 cells mis-expressed the Th2 cytokine gene *Il4* (8-fold increase compared to wild-type; Figure 4.2.2B) while *Mbd2-/-* Th2 cells expressed twice as much of the Th1 cytokine gene *Ifng* as wild-type Th2 cells (Figure 4.2.2A). Both wild-type and *Mbd2-/-* Th1 and Th2 cells appropriately expressed their own signature cytokine

gene (Figure 4.2.2A and B) but levels were slightly lower in the *Mbd2* nulls, although not to a statistically significant degree. This reduction may be due to inhibition mediated by over-expression of the opposing signature cytokine in the *Mbd2* nulls.

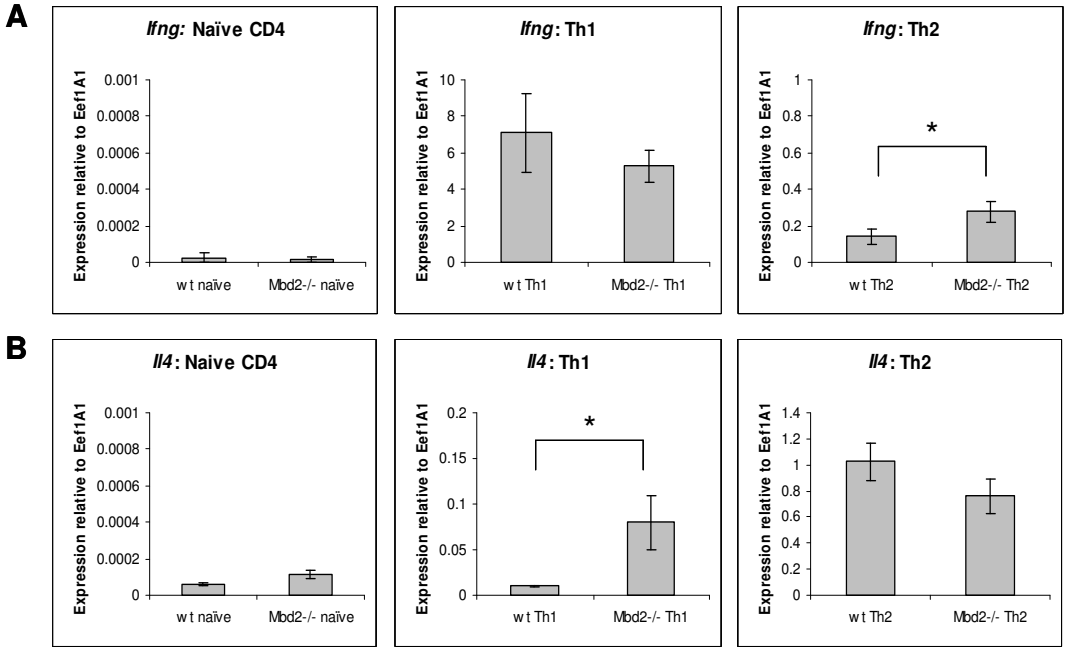


Figure 4.2.2 qRT-PCR shows that mis-expression of *Ifng* and *Il4* in *Mbd2-/-* T cells occurs at the RNA level

A) *Ifng* expression and **B)** *Il4* expression in wt and *Mbd2-/-* naïve CD4, Th1 and Th2 cells relative to *Eef1A1* as assessed by qRT-PCR. Results of Student's t-test: * $p<0.05$. Error bars = standard deviation from PCR replicates.

4.3 Mbd2-/- mice show additional immune system phenotypes

4.3.1 T helper cell proliferation is reduced in the absence of MBD2

During T helper cell differentiation experiments it was observed that *Mbd2-/-* CD4 cells exhibit reduced proliferation in response to stimulation and IL-2 treatment. We therefore examined proliferation of *Mbd2-/-* CD4 cells (C57/BL6 background) more closely by monitoring incorporation of tritiated (^{3}H) thymidine after stimulation

under Th0, Th1, Th2 and Treg conditions[ix]. Regulatory T cells (Tregs) are a distinct class of CD4 T cell which are involved in modulation of the immune response (Tang and Bluestone, 2008). *Mbd2-/-* cells showed reduced incorporation of ^{3}H compared to wild-type and *Mbd2+/-* cells under all four sets of differentiation conditions (Figure 4.3.1).

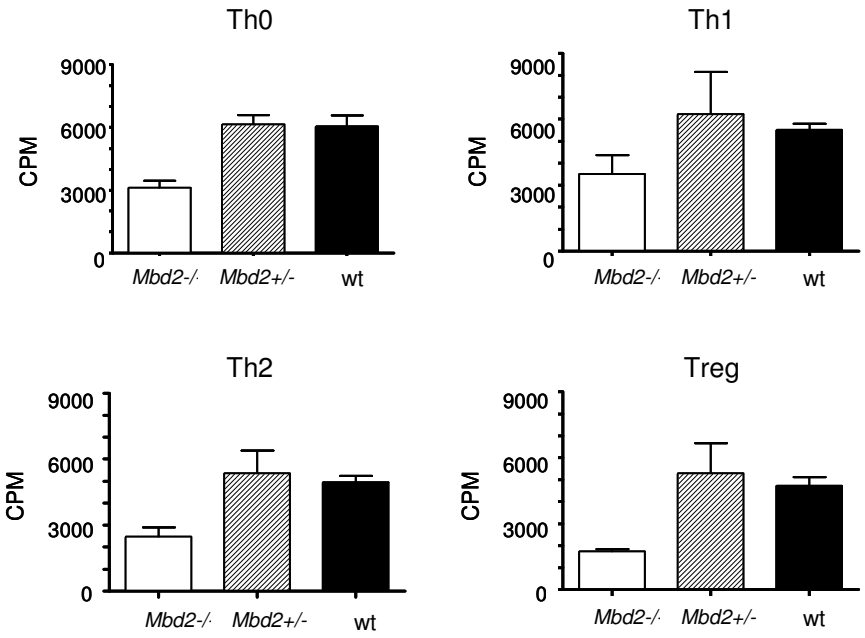


Figure 4.3.1 *Mbd2-/-* CD4 cells show reduced proliferation during T cell differentiation
CD4 cells from wild-type, *Mbd2+/-* and *Mbd2-/-* cells were stimulated, treated with IL-2 and cultured under Th0 (no skewing cytokines), Th1 (IL-12, α-IL-4), Th2 (IL-4, α-IL-12) or Treg (TGF-β1) conditions. ^{3}H-thymidine was added to the media and incorporation assessed after 48 hours by scintillation counting. CPM = counts per minute.

The proliferation defect was examined further under Treg differentiation conditions by labelling CD4 cells with the fluorescent molecule CFSE (carboxyfluorescein diacetate succinimidyl ester) and FACS staining to follow cell division after stimulation and treatment with TGF-β. After three days, *Mbd2-/-* Treg cell cultures contained more cells that had under gone 0, 1 or 2 cell divisions and less that had undergone 3 or more divisions compared to cultures of wild-type cells (Figure 4.3.2).

[ix] Note: All proliferation (section 4.3.1) and Treg (section 4.3.2) experiments were carried out in conjunction with John Grainger (Institute for Immunology and Infection Research, University of

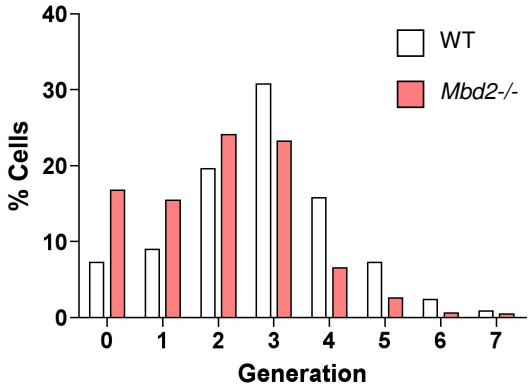


Figure 4.3.2 CFSE labelling shows that *Mbd2-/-* T cells divide less in response to stimulation and TGF-β

CD4 cells from wild-type and *Mbd2-/-* cells were stimulated, treated with IL-2 and cultured under Treg conditions (TGF-β1; 1ng/ml). CFSE was added to the cultures so that cell division could be followed by examining associated reduction of the fluorescent signal by FACS staining. "Generation" represents the number of cell divisions undergone by a population of cells. Wild-type = white bars, *Mbd2-/-* = pink bars.

4.3.2 *Mbd2-/-* mice show altered Treg cell phenotypes both *in vitro* and *in vivo*

Expression of the key transcription factor controlling Treg cell fate, Foxp3 has been reported to be regulated by DNA methylation (Baron et al., 2007; Floess et al., 2007; Josefowicz et al., 2009; Kim and Leonard, 2007; Lal et al., 2009; Nagar et al., 2008; Polansky et al., 2008). This prompted us to examine whether removal of MBD2 would affect generation of Treg cells. It is possible to differentiate Tregs from CD4 cells *in vitro* in a similar manner to Th1 and Th2 cells through activation of cells and treatment with recombinant TGF-β (Chen et al., 2003; Fantini et al., 2004). We assessed Treg generation (on a BALB/c background) at various concentrations of TGF-β1 (0.01-10ng/ml) by FACS staining for Foxp3 expression. *Mbd2-/-* CD4 cells gave rise to more Tregs than wild-type cells with the greatest difference seen at a concentration of 1ng/ml TGF-β1 (69% versus 34%; Figure 4.3.3A). Minimal expression of Foxp3 was seen in the absence of TGF-β1 in both cell types (data not shown). This suggests that either *Mbd2-/-* CD4 cells are hyper-responsive to TGF-β

Edinburgh).

or that they have a propensity to activate Foxp3 perhaps due to ineffective Foxp3 silencing in the absence of MBD2. Increased Treg differentiation in the absence of MBD2 was observed in three separate experiments but did not occur in two subsequent experiments performed. This could perhaps be due to incomplete penetrance of the phenotype (perhaps related to individual variation between mice) or a change in the health status of the *Mbd2-/-* animals.

Although Tregs can be generated from naïve CD4 cells as described above, they are also produced as a distinct cell lineage in the thymus (Tang and Bluestone, 2008). We therefore looked at CD4 T cells isolated from the mesenteric lymph nodes of naïve mice to see if the absence of MBD2 had an effect on the number of Tregs generated *in vivo.* A greater proportion of Foxp3+ Tregs were found in *Mbd2-/-* mice than in wild-type animals (on average 18% versus 14%) but, in particular, more "active" Tregs were produced as indicated by the presence of the homing marker CD103 (Finney et al., 2007) on Foxp3+ cells (71% versus 28% CD103+; Figure 4.3.3B). Infection by the intestinal helminth *Heligmosomoides polygyrus* results in production of Treg cells which are induced by the parasite in order to suppress the immune response and thus evade elimination (Finney et al., 2007). We examined whether the increased number of Foxp3+ CD103+ Tregs in *Mbd2-/-* mice led to an increased susceptibility to *H. polygyrus* infection. A trend towards increased worm burden was seen in the duodenum of *Mbd2-/-* compared to wild-type mice after a 2 week *H. polygyrus* infection (Figure 4.3.3C). However, this difference did not reach statistical significance ($p>0.05$; Student's t-test) and survival and cytokine production in *Mbd2-/-* mice were not compromised (data not shown). Interestingly, the difference in the number of Foxp3+ CD103+ cells between wild-type and *Mbd2-/-* mice did not change appreciably as a result of infection (data not shown).

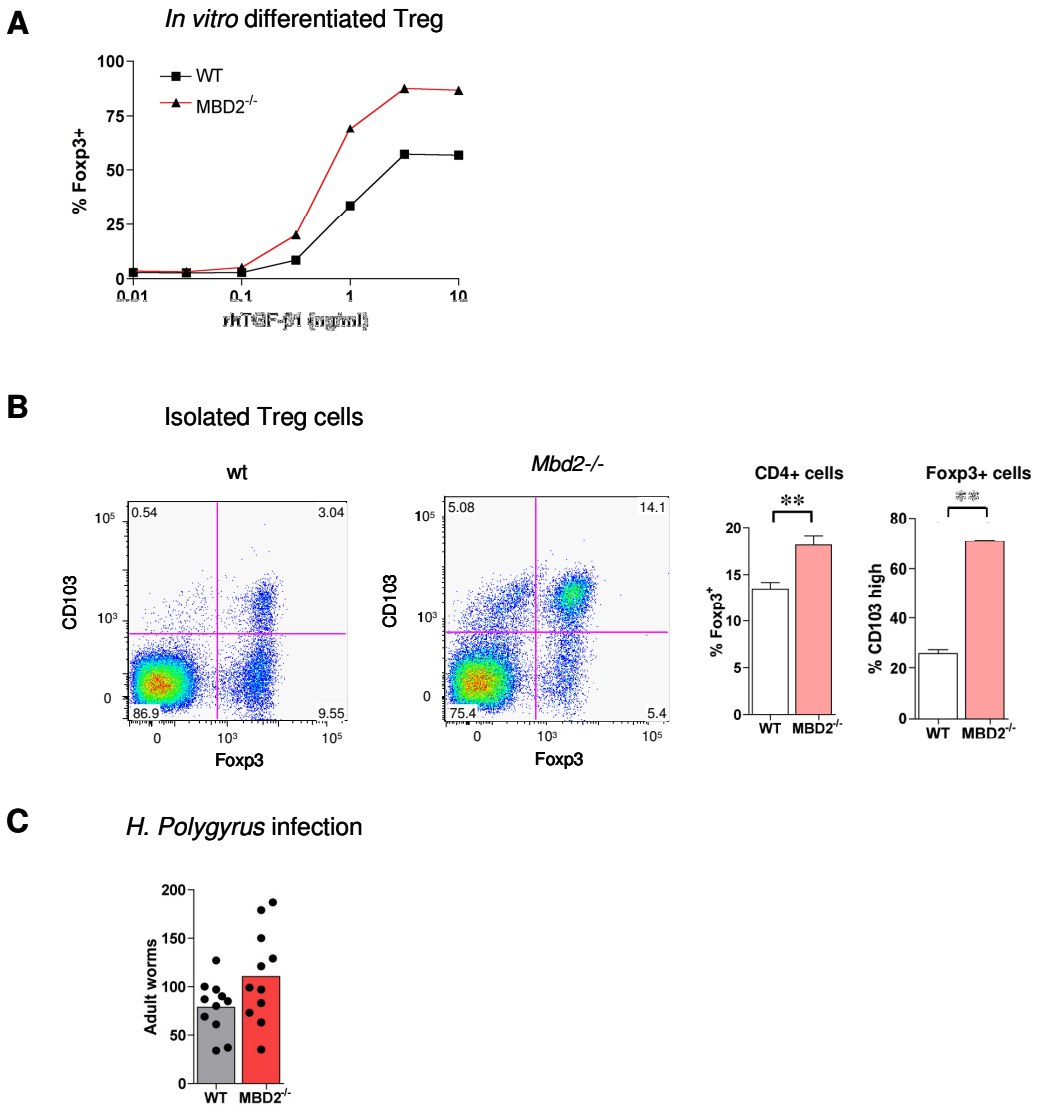


Figure 4.3.3 *Mbd2-/-* mice show altered generation of Tregs *in vitro* and altered Treg phenotypes *in vivo*

A) Generation of Tregs *in vitro* by treatment with TGF-β. CD4 cells from wild-type and *Mbd2-/-* mice were treated with various concentrations of TGF-β (0.01-10ng/ml) and Treg production assessed by intracellular FACS staining for Foxp3. **B)** Staining of CD4 cells isolated from the mesenteric lymph nodes for Foxp3 and CD103. The left hand panels show representative cytokine stains from one wild-type and one *Mbd2-/-* mouse while the graphs on the right summarise data obtained from 10 mice of each genotype in two independent experiments. Results of Student's t-test: * *p<0.01 **C)** Number of worms isolated from wild-type and *Mbd2-/- H. polygyrus*-infected mice. Black dots represent individual animals and the bars represent the mean for each group.

It has been reported that MBD2 binds to a differentially methylated CGI upstream of the *Foxp3* locus and thus helps silence the gene in naïve CD4 cells when it is not needed (Lal et al., 2009). However, this CGI is located more than 5kb upstream of the *Foxp3* gene and is associated with the 5' end of another transcript. Subsequent studies have failed to repeat the differential methylation reported and found this CGI to be constitutively unmethylated (Josefowicz et al., 2009). However, a number of other differentially methylated regions have been identified in the *Foxp3* locus one of which, an intronic CpG-rich region, has been reported to bind MeCP2 (Kim and Leonard, 2007). Unfortunately it is not possible to perform ChIP to test MBD2 binding at this region for reasons discussed in section 4.7. Therefore it remains unclear how MBD2 contributes to gene regulation in Tregs and what the relationship is between the increased number of Tregs generated in response to TGF-β *in vitro* and the increased number of CD103+ Foxp3+ Tregs generated *in vivo*. TGF-β is involved in both *in vitro* Treg induction (Chen et al., 2003; Fantini et al., 2004) and in CD103 expression (El-Asady et al., 2005; Robinson et al., 2001) so it is possible that this provides a link between the two phenotypes. The variability of the *in vitro* Treg induction phenotype in *Mbd2-/-* mice also needs to be addressed. Nevertheless, these Treg phenotypes point to further involvement for MBD2 in the immune system.

4.3.3 *Mbd2-/-* mice show increased susceptibility to oesophageal infection

Histopathological examination of organs from wild-type and *Mbd2-/-* mice on a BALB/c background revealed hyperplasia and infiltration of eosinophils and lymphocytes specifically in the oesophagus of *Mbd2-/-* mice. Five out of six *Mbd2-/-* mice over 60 days of age were affected while younger *Mbd2-/-* mice and wild-type mice were not affected (Figure 4.3.4; D. Brownstein, unpublished).

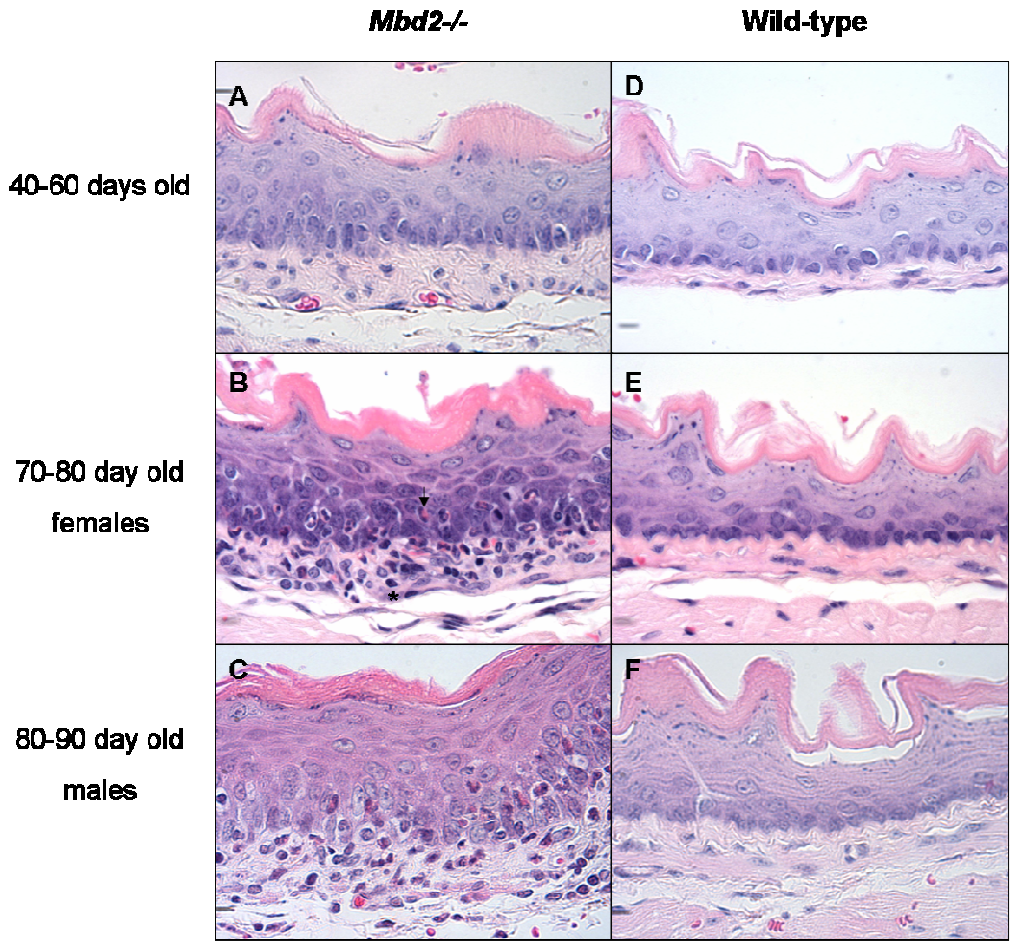
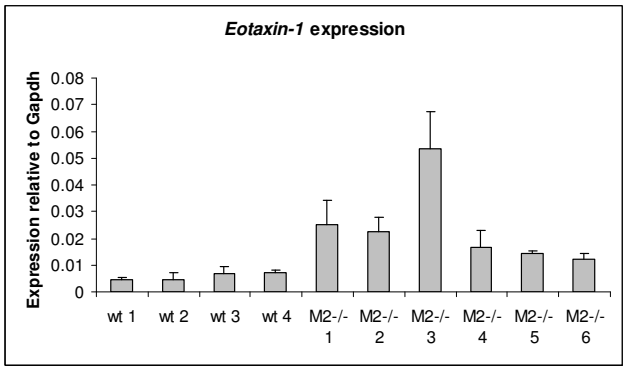


Figure 4.3.4 *Mbd2-/-* BALB/c mice show immune cell infiltration of the oesophagus
H&E stained sections of oesophagus from **A-C)** *Mbd2-/-* and **D-F)** Wild-type mice. Lymphocytes (asterisk) and eosinophils (arrow) were observed in male and female *Mbd2-/-* mice over 60 days of age. [Histopathology by D. Brownstein].

It was initially thought that this represented a spontaneous condition similar to the human condition eosinophilic oesophagitis which is characterised by upregulation of the chemokine eotaxin-3 and is associated with Th2 cytokine production (Blanchard et al., 2006). Since mice lack a homologue of the *eotaxin-3* gene we examined expression of the gene encoding eotaxin-1 (*Ccl11*) which also acts as a eosinophil chemoattractant (Borchers, 2002) to investigate whether it might contribute to the oesophageal phenotype. *Eotaxin-1* was upregulated in all six *Mbd2-/-* mice examined (aged 75-90 days) compared to wild-type mice (Figure 4.3.5A). In order to more thoroughly assess gene expression changes in the *Mbd2-/-* oesophagus, microarray analysis was performed (NimbleGen mm8_60mer_expression array) in triplicate on

pools of male and female animals with an average age of 85 days. A total of 304 transcripts were upregulated and 94 transcripts downregulated in *Mbd2-/-* oesophagus compared to wild-type, including many genes with immune system function (Figure 4.3.5B and C).

A



B

Gene	Description	Fold upregulation
Sprr2e	Small praline-rich protein 2E	162
Uox	Urate oxidase	145
Defb3	Defensin beta 3	52
Ear5	Eosinophil-associated ribonuclease 5	33
IL-19	Interleukin 19	25
Ccl8	Chemokine (C-C motif) ligand 8	11
Cxcl2	Chemokine (C-X-C motif) ligand 2	9

C

GO Biological process	p-value
Immune response	3.59E-22
Inflammatory response	3.69E-08
Antigen presentation	4.95E-05
Immune cell chemotaxis	0.0419

Figure 4.3.5 Gene expression in the *Mbd2-/-* oesophagus

A) *Eotaxin-1* expression in wild-type (wt) and *Mbd2-/-* (M2-/-) oesophagus. Expression is relative to *Gapdh* and error bars represent standard deviation from PCR replicates. **B)** A selection of genes upregulated in *Mbd2-/-* oesophagus compared to wild-type identified using NimbleGen expression arrays. **C)** GO terms over-represented in genes that change their expression in the absence of MBD2.

This analysis of the *Mbd2-/-* oesophagus phenotype was carried out on mice just prior to a rederivation[x] of our transgenic mouse lines due to a pinworm infection in the animal unit. It was decided to re-examine the oesophagus phenotype in the re-derived *Mbd2-/-* mice. *Mbd2-/-* mice on both a C57/BL6 and BALB/c background were examined. C57/BL6 mice are bred as heterozygotes which allowed wild-type littermates to be used for controls. As the *Mbd2-/-* BALB/c mice are maintained as a homozygous line, wild-type controls were co-housed with *Mbd2-/-* animals for two weeks prior to analysis to control for any environmental variation. All five of the *Mbd2-/-* BALB/c mice examined (aged 85-92 days) showed the oesophageal phenotype observed previously but, surprisingly, so did two out of five wild-type animals examined (aged 85 days). This suggests that this is not a spontaneous phenotype but rather reflects increased susceptibility of *Mbd2-/-* mice to an unknown oesophageal infection. None of the wild-type or *Mbd2-/-* C57/BL6 mice examined displayed the oesophagus phenotype indicating either that this genetic background is resistant to the infection or that the C57/BL6 mice were not exposed to the oesophageal pathogen. Background-specific immune responses have been observed previously between BALB/c and C57/BL6 mice (Reiner and Locksley, 1995).

The increased susceptibility of *Mbd2-/-* mice to oesophageal infection is in keeping with the numerous immune system phenotypes observed in these mice including the compromised immune responses observed previously [(Hutchins et al., 2005) and P. Cook, unpublished]. However, it seems likely that the gene expression differences observed in *Mbd2-/-* oesophagus are not due to direct regulation of these genes by MBD2 but can be attributed to infiltrating immune system cells induced by the infection.

[x] The rederivation involved transfer of pre-implantation embryos into pseudopregnant females and was carried out to establish specific pathogen free mice.

4.4 Results: Gene expression changes in the absence of MBD2

The immune system phenotypes observed in *Mbd2-/-* mice prompted us to use immune cells as a model in which to explore gene expression changes occurring in the absence of MBD2.

4.4.1 Wild-type and *Mbd2-/-* immune cells show differences in gene expression

Dendritic cells, B cells, naïve CD4, Th1 and Th2 cells were purified from *Mbd2-/-* C57/BL6 mice as described in section 3.2. RNA was extracted from these cells, labelled and hybridised to Illumina BeadChips (WTCRF) in parallel with the wild-type samples used in chapter 3. Microarray analysis was carried out as described in section 3.4 and genes showing at least a 2-fold change in expression between wild-type and *Mbd2-/-* cells were selected as differentially expressed. In general, modest numbers of gene expression differences were seen between wild-type and MBD2-deficient cells although the number of differences varied between cell types. For example, quiescent naïve CD4 cells showed just 16 changes while activated Th1 cells showed 93 changes between wild-type and *Mbd2-/-* cells. The greatest number of differences were seen between wild-type and *Mbd2-/-* dendritic cells (218 differences) pointing to a prominent role for MBD2 in these cells. In most cases genes showing expression changes in the absence of MBD2 were upregulated, consistent with MBD2's function as a transcriptional repressor (Figure 4.4.1 and Table 4.4.1). These data suggest that the removal of MBD2 does not have a global effect on gene expression but results in upregulation of a moderate number of genes. This could indicate that MBD2 does not have many target genes in the cell or that in the absence of MBD2 other silencing mechanisms can compensate, as has been reported at the *Xist* locus (Barr et al., 2007).

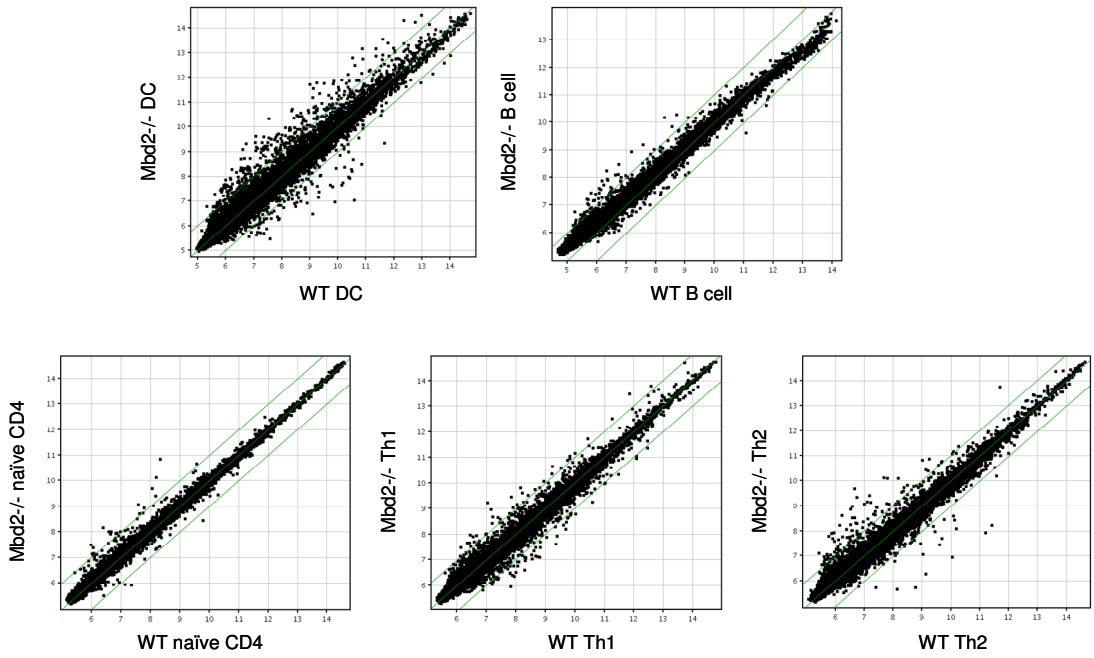


Figure 4.4.1 Comparison of gene expression in wild-type and *Mbd2-/-* immune cells

Pair-wise comparison plots depicting gene expression in wild-type (WT) and *Mbd2-/-* dendritic cells (DC), B cells, naïve CD4, Th1 and Th2 cells. Each dot represents a unique probe on the array and probes located outside of the green lines show at least a 2-fold difference in expression. The scale is log2 of intensity values.

Cell type	**RefSeq genes changed in *Mbd2-/-* vs WT**	**Upregulated genes**
DC	218	208
B cell	32	32
Naïve CD4	16	14
Th1	93	71
Th2	58	31

Table 4.4.1 Gene expression differences between wild-type and *Mbd2-/-* cells

For each cell type the number of RefSeq genes showing at least a 2-fold expression difference between wild-type and *Mbd2-/-* cells is indicated along with the number upregulated in *Mbd2-/-* cells compared to wild-type.

In section 4.2, mis-expression of the key cytokine genes *Ifng* and *Il4* was observed in the absence of MBD2, consistent with the reports of Hutchins and colleagues (2002). These genes were also found to be upregulated by the microarray study (*Ifng* was

upregulated 1.5-1.95-fold in *Mbd2-/-* Th2 cells and *Il4* was upregulated 4-fold in *Mbd2-/-* Th1 cells). Consistent with a perturbed cell identity in these cells, *Mbd2-/-* Th1 cells mis-expressed a number of additional Th2-specific genes (8 genes; including *Lat2, C3* and *Il17f*) while *Mbd2-/-* Th2 cells over-expressed some Th1-specific genes[xi] (6 genes in total; including *Ifitm1*, *Ctsw* and *Fh12*).

4.4.2 Validation of microarray results

The array results for several genes were verified by qRT-PCR using *Eef1A1* expression as a reference (see section 3.4.3). Genes that changed expression in each of the immune cells types were tested and for each, qRT-PCR data confirmed microarray results. Verification was performed for 4 genes that change expression in *Mbd2-/-* naïve CD4 cells, 7 genes changed in both Th1 and Th2 cells, 5 genes changed in dendritic cells and 3 that change expression in *Mbd2-/-* B cells. A selection of these; *Ppic*, *Il4i1*, *Il8ra* and *Ptpn14* is shown in Figure 4.4.2.

[xi] Th1-specific genes are those that are upregulated at least 2-fold in wild-type Th1 cells compared to wild-type Th2 cells. Th2-specific genes are those that are upregulated in Th1 cells compared to Th2 cells.

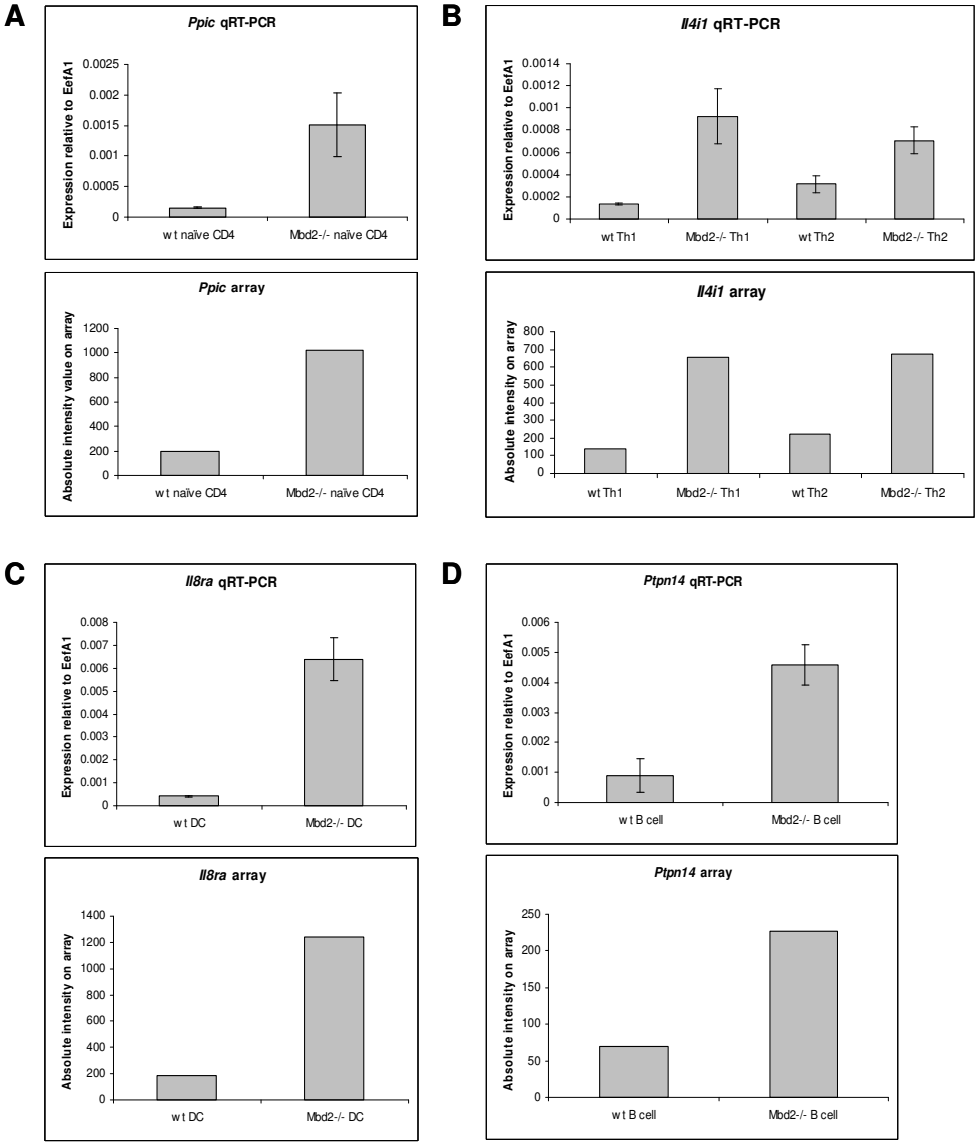


Figure 4.4.2 Validation of microarray results by qRT-PCR

A) *Ppic* expression in wt and *Mbd2-/-* naïve CD4 cells **B)** *Il4i1* in Th1 and Th2 cells **C)** *Il8ra* in dendritic cells **D)** *Ptpn14* in B cells. For each, the top panel shows qRT-PCR relative to *Eef1A1* (error bars = standard deviation from PCR replicates) and the bottom panel shows the absolute intensity value from the BeadChip arrays averaged across replicates.

4.4.3 Absence of MBD2 affects different genes in different immune cell types

We chose to focus on upregulated genes as these may be directly repressed by MBD2 and looked to see if the same genes are upregulated in the absence of MBD2 in different cell types. Th1 and Th2 cells share the greatest number of MBD2 targets

with 16 genes upregulated in both *Mbd2-/-* Th1 and Th2 cells. This is consistent with the similar transcriptional profiles of Th1 and Th2 cells (see section 3.4). However, when more distant cell types were compared, very few of the same genes were changed in the absence of MBD2. For example, dendritic cells show 208 genes upregulated by at least 2-fold in the absence of MBD2 but only 7 of these are also upregulated in *Mbd2-/-* Th1 cells. Therefore, in general, removal of MBD results in different expression changes in different cell types.

4.5 Results: ChIP for MBD2 was not successful

In order to properly assess the number of MBD2 target genes and their relationship to DNA methylation it is necessary to determine where MBD2 binds in the genome not just which genes are mis-regulated in its absence. Once binding sites are known, one can assess whether MBD2 binds different targets in different cell types as the gene expression data suggest and if differential binding is directed by DNA methylation or involves other factors.

4.5.1 Purification and testing of MBD2 antibodies

ChIP involves cross-linking proteins to DNA followed by antibody-mediated isolation of a protein of interest and recovery of bound DNA sequences. In order to successfully carry out ChIP the antibody used must be specific for the protein of interest and be able to recognise its epitope after cross-linking. The α-MBD2 antibody S923 was raised in sheep against an N-terminally truncated form of MBD2a-GST and has been shown to specifically immunoprecipitate MBD2a without cross-reacting with closely related MBD3 (Ng et al., 1999). In addition, it has been used previously in MBD2 ChIP studies (Barr et al., 2007; Berger et al., 2007; Hutchins et al., 2002). Western blot analysis using this antibody to probe cell extracts from mouse tail fibroblasts (MTFs) revealed a number of non-specific bands in addition to MBD2a (Figure 4.5.1). As one of the previously reported MBD2 ChIP experiments used affinity purified antibody (Berger et al., 2007) I decided to affinity purify S923 serum in order to remove antibody activity that was not specific for

MBD2a. Initial attempts involved coupling GST-tagged recombinant human MBD2a (N-terminal fragment, amino acids 1-151) to an Affi-Gel matrix (Bio-Rad), binding the anti-serum and eluting antibody with specificity for MBD2a at low pH. However, coupling of recombinant MBD2 to Affi-Gel was inefficient (approximately 30% of recombinant MBD2 bound) and low yields of purified antibody were obtained (200µg of antibody was recovered from 2ml of serum). In addition, MBD2 signal by Western blot was weaker than when the unpurified serum was used and ChIP for MBD2 at known targets proved unsuccessful (data not shown). Further attempts to affinity purify S923 including using a different fragment of MBD2 (mouse MBD2 amino acids 10-156) and coupling recombinant MBD2 to CNBr beads rather than Affi-Gel were also unsuccessful.

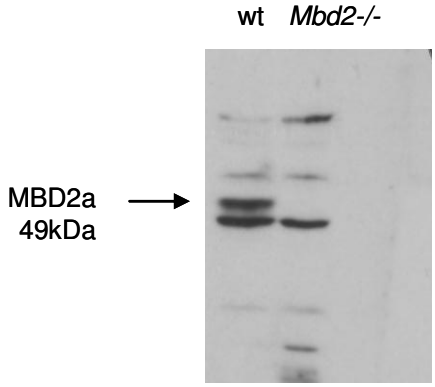


Figure 4.5.1 Western blot with S923 antibody gives non-specific bands in addition to MBD2

Western blots in wild-type and *Mbd2-/-* MTFs using unpurified S923 serum shows that the antibody recognises MBD2 (arrow) but also recognises proteins that are not MBD2 as evidenced by the number of bands in the *Mbd2-/-* sample.

Therefore, as an alternative to affinity purification, it was decided to purify IgG molecules from the serum in order to remove contaminating proteins and to concentrate the antibody activity. This was carried out using protein G which has specific affinity for the Fc domain of IgG molecules from many species including sheep. High yields of concentrated IgG were obtained from S923 serum using this method (concentration = 3.4mg/ml, 5mg IgG recovered from 2ml serum). I tested whether this antibody could immunoprecipitate MBD2a after cross-linking of MTF

cell extracts using 1% formaldehyde. *Mbd2-/-* MTFs were used as a control to ensure that MBD2a specifically was being recognised. Immunoprecipitates were analysed by Western blot by probing with another MBD2 antibody R593 (Ng et al., 1999). Protein G-purified S923 pulled-down approximately 5-10% of MBD2a protein after cross-linking and, as expected, no band was seen for MBD2a in *Mbd2-/-* MTFs (Figure 4.5.2). This confirmed that S923 can immunoprecipitate MBD2. However, efficiency was low possibly due to the fact that some epitopes were obscured by cross-linking of MBD2 to other NuRD components. A similar percentage of MBD2 was precipitated when cross-linked immunoprecipitation (IP) was carried out in human cells (HCT116 cells; data not shown).

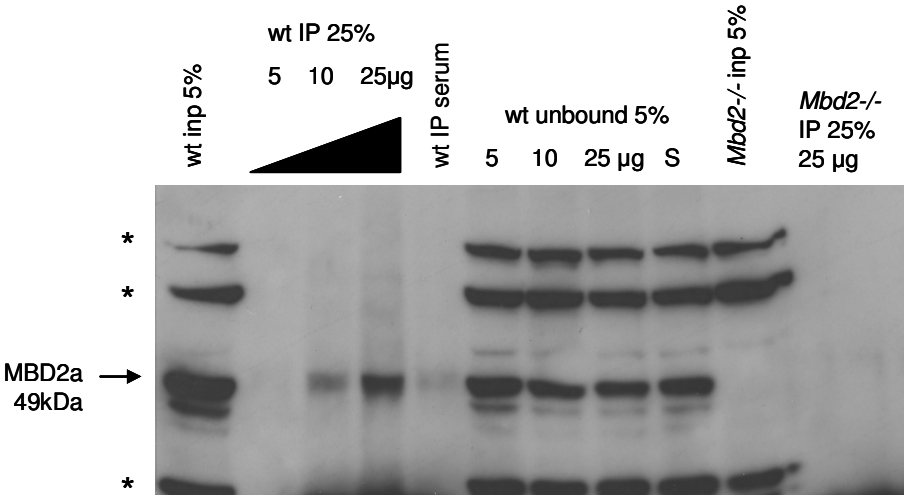


Figure 4.5.2 S923 can immunoprecipitate MBD2 after formaldehyde cross-linking
Western blot for MBD2 (R593 antibody) on input from IP (inp, 5%), the fraction immunoprecipitated by purified S923 (IP, 25%) at various antibody concentrations (5, 10, 25µg) or S923 serum and the unbound fraction from the IP (5, 10, 25µg antibody, S = serum). MBD2a runs at 49kDa and no signal is seen for an IP in *Mbd2-/-* MTFs. Non-specific bands are indicated by the asterisk.

4.5.2 MBD2 ChIP was attempted at known targets

I tried to repeat previously published MBD2 ChIP results using the purified S923 antibody. MBD2 has been reported to bind the promoter of the *TFF2* gene (which codes for a secretory protein expressed in the gastrointestinal mucosa) in the HCT116 human colon cancer cell line (Berger et al., 2007). I carried out ChIP for MBD2 over this region (Figure 4.5.3A) using MBD2-depleted cells (80%

knockdown of MBD2 at the protein level) as a control for antibody-specificity. ChIP carried out in parallel using an antibody against histone H3 (Abcam) acted as a positive control while ChIP with pre-immune serum was used as a negative control. Surprisingly, the purified S923 antibody gave no enrichment for MBD2 over any of the three regions of the *TFF2* promoter interrogated. ChIP signal in wild-type HCT116 cells was low (<0.01% of input DNA) and comparable to the signal seen when pre-immune serum was used for immunoprecipitation. In addition, signal did not decrease in the MBD2 knockdown cells (Figure 4.5.3B). Using crude S923 serum gave similar results (data not shown). However, histone H3 was successfully detected at all three regions interrogated (Figure 4.5.3C).

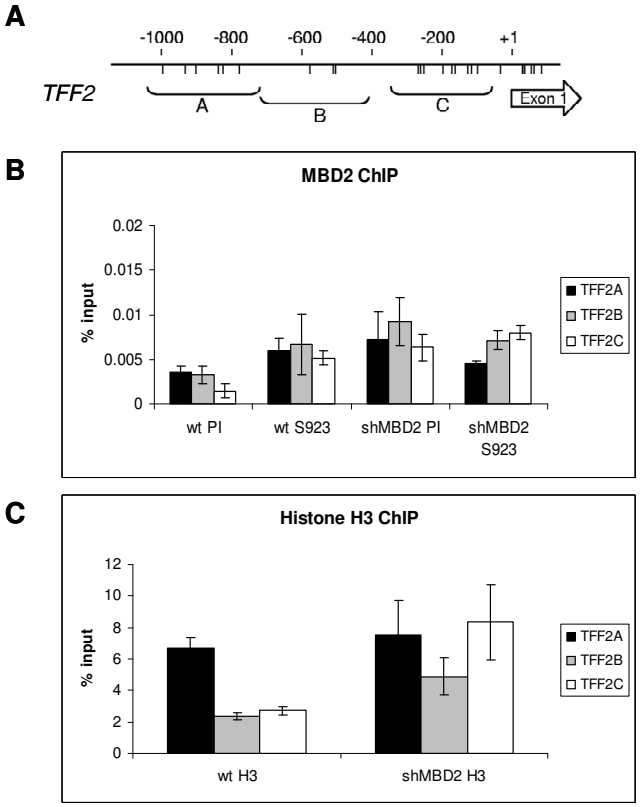


Figure 4.5.3 ChIP for MBD2 at the *TFF2* locus in HCT116 cells

A) Regions of the *TFF2* promoter interrogated by ChIP. Black lines indicate CpG residues. (Taken from Berger et al., 2007) **B)** ChIP using pre-immune serum (PI, negative control) and α-MBD2 (purified S923). **C)** ChIP for histone H3 (positive control). Wt = wild-type cells, shMBD2= cells where MBD2 was depleted by shRNA. Error bars = standard deviation from PCR replicates.

I concluded that the S923 antibody is unable to recognise MBD2 at the *TFF2* locus. This could be due to the fact that the antibody pulls down only a small fraction of MBD2 present in the cell and thus is below the detection limit for ChIP. Alternatively, *TFF2* may not actually be a direct target of MBD2.

To assess whether I am simply looking at the wrong genomic location for MBD2 binding I attempted to repeat ChIP for MBD2 at another published target, the *Xist* gene in male MTF cells (Barr et al., 2007). Two primer pairs spanning the *Xist* promoter (FR2 and FR3) were used to test DNA immunoprecipitated using purified S923. ChIP gave no enrichment for MBD2 at *Xist* in wild-type compared to *Mbd2-/-* cells with signal for MBD2 close to background levels (Figure 4.5.4A). Primers spanning a more extensive 9kb region of the *Xist* locus also failed to detect MBD2 binding (data not shown). I then tested a range of other MBD2 antibodies by performing ChIP at the *Xist* promoter in MTFs. These included crude S923 serum, R593 antibody, ChIP grade α-MBD2 from Abcam (Phesse et al., 2008) and α-MBD2 from Imgenex (Le Guezennec et al., 2006) (Figure 4.5.4B and C) as well as antibodies received from Paul Wade and Irina Stancheva (data not shown). None of the antibodies tested gave specific enrichment in wild-type compared to *Mbd2-/-* MTF cells and in general signal was similar to the levels seen for pre-immune serum and purified S923. Custom antibodies raised against peptides from the MBD2 protein were produced in conjunction with Cambridge Research Biosciences but these failed to recognise MBD2.

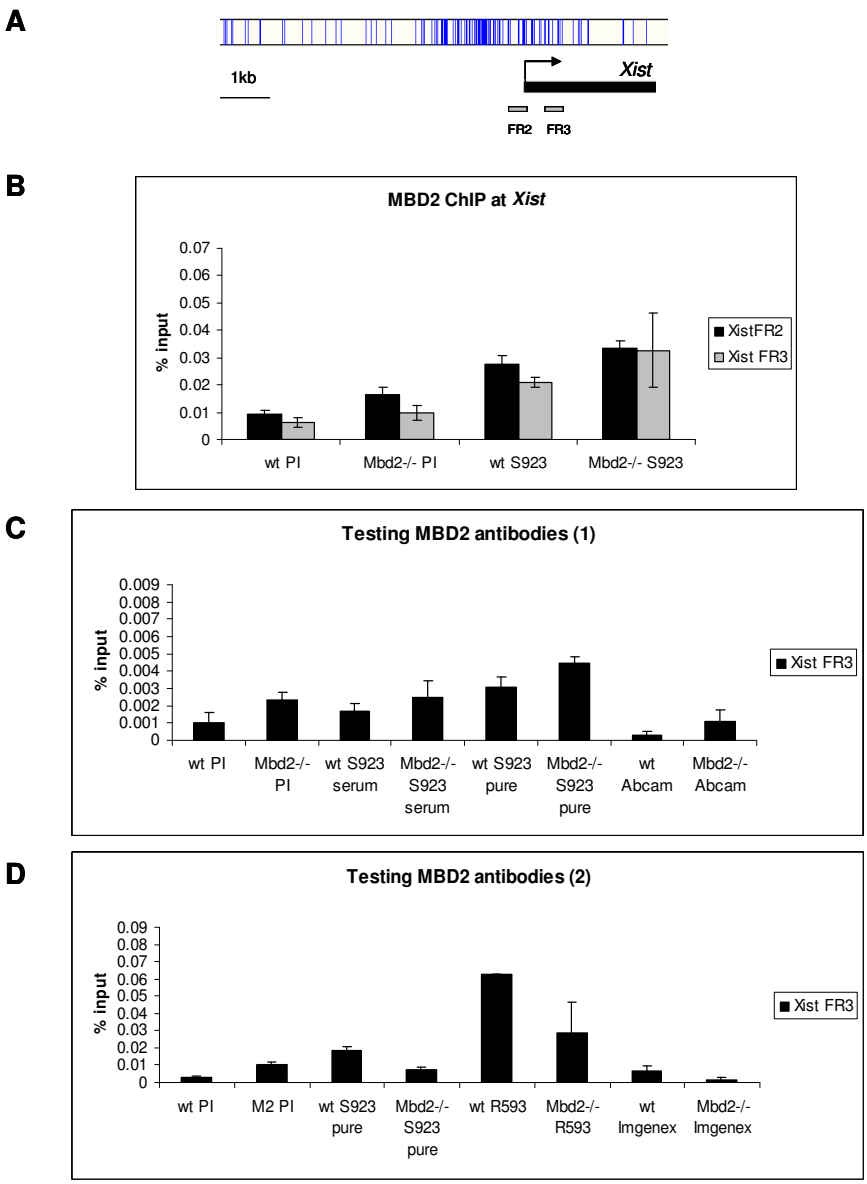


Figure 4.5.4 ChIP for MBD2 at the *Xist* locus in MTFs

A) Schematic of the *Xist* promoter. Grey bars indicate primer positions and the blue lines indicate CpG residues. **B)** ChIP for MBD2 at the *Xist* promoter (FR2 and FR3) carried out with pre-immune serum (PI) and purified S923 (S923) in wild-type and *Mbd2-/-* MTF cells. **C)** ChIP at *Xist* FR3 using pre-immune (PI), unpurified S923 (S923 serum), purified S923 (S923 pure) and Abcam α-MBD2 (Abcam). **D)** ChIP using pre-immune, purified S923, R593 and Imgenex α-MBD2 (Imgenex). Error bars = standard deviation from PCR replicates

The lack of ChIP signal observed for MBD2 at *Xist* prompted us to examine additional genomic locations for MBD2 binding. MBD2 binding has been reported at two regions of the *Il4* locus, conserved non-coding sequence 1 (CNS-1) and the

intronic enhancer (2IE) in mouse Th1 cells (Hutchins et al., 2002). However, I failed to detect MBD2 binding at these regions (Figure 4.5.5A). Major satellite is highly methylated repetitive DNA which is organised into heterochromatic foci in mouse cells. Since, MBD2 has been shown to localise to these foci by immunofluorescence (Hendrich and Bird, 1998), we tested whether we could detect MBD2 binding by ChIP at the major satellite repeat. However, similar signal was seen in wild-type and *Mbd2-/-* MTF cells indicating no specific enrichment for MBD2 (Figure 4.5.5B).

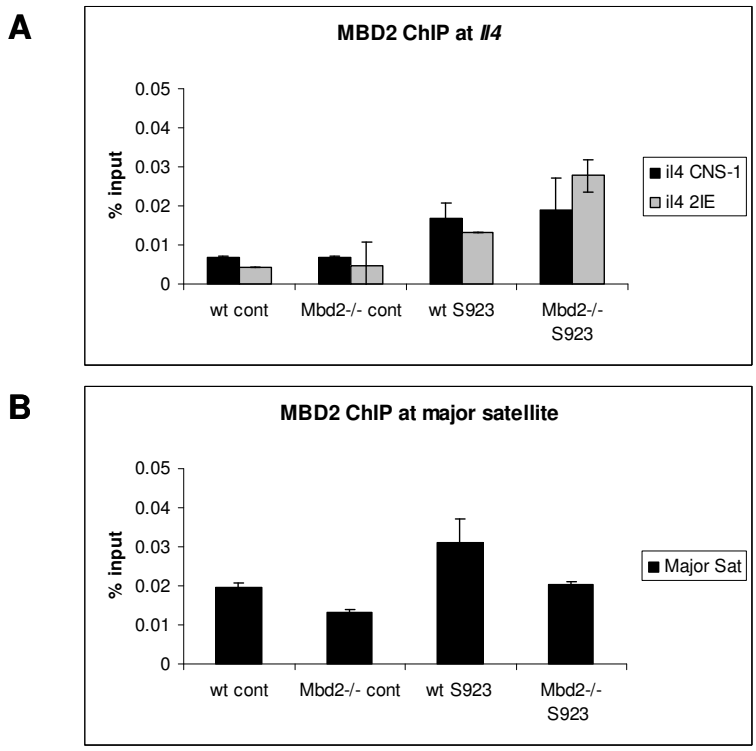


Figure 4.5.5 ChIP for MBD2 at *Il4* and major satellite

A) MBD2 ChIP at CNS-1 and the second intronic enhancer (2IE) of the *Il4* gene carried out in wild-type and *Mbd2-/-* Th1 cells. **B)** ChIP at major satellite in MTFs. Cont = negative control (rabbit IgG) and S923 = purified S923. Error bars = standard deviation from PCR replicates

There are a number of explanations for why MBD2 binding could not be detected at *Xist* or any of the other regions tested. One is that these regions, despite published ChIP data and expression differences in the absence of MBD2, are not actually bound by MBD2. Alternatively, the antibodies tested may not immunoprecipitate

enough MBD2 to be detectable by ChIP or MBD2 may not be efficiently cross-linked to DNA so that despite MBD2 pull-down bound DNA is not recovered.

4.5.3 ChIP-chip for MBD2 did not identify convincing targets

To test the first possibility I hybridised ChIP DNA (obtained using the purified S923 antibody) from wild-type and *Mbd2-/-* CD4 T cells to NimbleGen CpG island plus promoter arrays to see if we could identify any MBD2 target genes. These arrays consist of tiled probes covering 1.3kb upstream and 500bp downstream of all RefSeq promoters and UCSC-annotated CpG islands. Microarray analysis was carried out by R. Illingworth and 7 regions were identified that showed enrichment for consecutive probes in wild-type compared to *Mbd2-/-* CD4 cells. Two of these regions were associated with methylated CGIs based on MAP-seq in CD4 cells (Table 4.5.1).

Region	**Median wt**	**Median *Mbd2-/-***	**Gene**	**CGI**	**Methylation (MAP-seq CD4)**
chr1:121316871-121320125	0.02031	-0.1195	*Inhbb*	yes	no
chr1:172994070-172996091	0.5614	-0.3767	*Fcgr3*	yes	yes
chr11: 59952283-59954663	0.1985	-0.05886	*Rai1*	yes	yes
chr13:37916607-37919761	0.01448	-0.06141	*Rreb1*	yes	no
chr13:63664548-63668128	0.00674	-0.05751	*Ptch1*	yes	no
chr7:127985096-127987217	0.2589	0.01006	*Eef2k*	yes	no

chr7:146766819-146769992	0.10250	-0.02093	*Nkx6-2*	yes	no

Table 4.5.1 Potential MBD2 binding sites identified by ChIP-chip in CD4 cells

Regions enriched for ChIP signal in wild-type compared to *Mbd2-/-* T cells. For each region, genomic location, median signal in wild-type and *Mbd2-/-* cells, associated gene, CGI association and methylation status as determined by MAP-seq in CD4 cells is shown.

PCR primers were designed against some of these regions and MBD2 binding tested by ChIP combined with qPCR in wild-type and *Mbd2-/-* T cells. None of the regions tested (*Fcgr3*, *Ptch1*, *Eef2k*, *Nkx6-2* – two genomic regions for each) showed enrichment for MBD2 in wild-type compared to *Mbd2-/-* CD4 cells implying that the ChIP-chip results are artefacts. A selection of the ChIP-qPCR results are shown in Figure 4.5.6A, B and C.

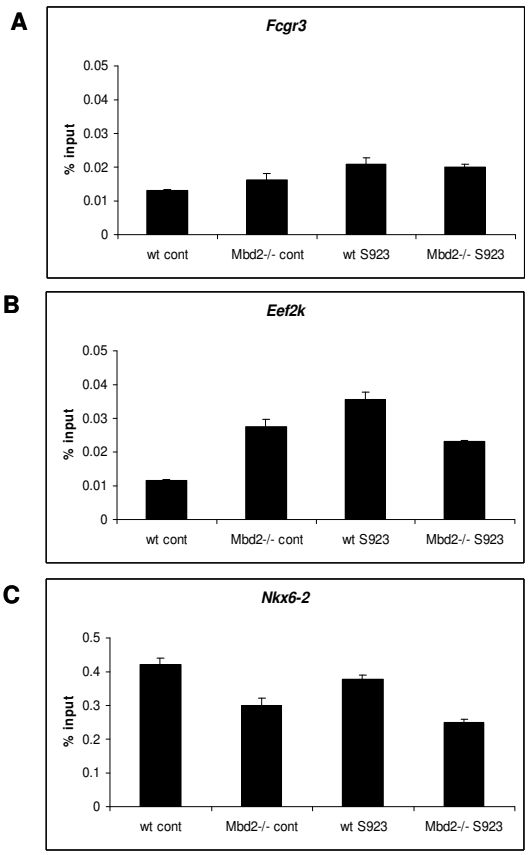


Figure 4.5.6 Testing MBD2 ChIP-chip results by ChIP-qPCR

ChIP was carried out using a negative control (cont; rabbit IgG) and purified S923 and tested using primers located in **A)** *Fcgr3* **B)** *Eef2k* and **C)** *Nkx6-2*. Error bars = standard deviation from PCR replicates

My inability to identify any *bona fide* MBD2 targets using this microarray-based approach suggests that the S923 antibody is not capable of immunoprecipitating enough MBD2 to give a specific signal in ChIP experiments or that cross-linking of MBD2 to DNA is inefficient.

4.5.4 ChIP using FLAG-tagged MBD2 failed to identify specific binding sites

To circumvent the issue of antibody efficiency in MBD2 ChIP (S923 only recovers 5-10% of MBD2 in the cell) I made use of *Mbd2-/-* MTF cells where FLAG-tagged MBD2 has been introduced by retroviral infection (Barr et al., 2007). Cross-linked IP in these cells using α-FLAG-coupled beads (Sigma) pulled down the majority of MBD2 as evidenced from the depletion of MBD2 in the unbound fraction (Figure 4.5.7).

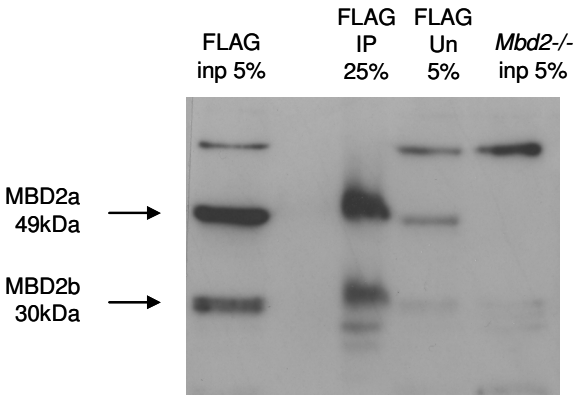


Figure 4.5.7 Cross-linked immunoprecipitation of FLAG-MBD2
α-FLAG beads efficiently immunoprecipitate MBD2 in *Mbd2-/-* MTFs infected with FLAG-MBD2. Both MBD2a and MBD2b are recovered (IP, 25%) and are depleted in the unbound fraction (Un, 5%).

Since α-FLAG antibody was much more efficient at pulling down MBD2 than S923 I tested whether it could detect MBD2 binding over the *Xist* promoter in these cells. ChIP over the *Xist* promoter using primers FR2 and FR3 gave a 2.5 to 3-fold enrichment of signal in the FLAG-MBD2 cells compared to the *Mbd2-/-* cells. Surprisingly a ChIP signal for MBD2 was also seen over the unmethylated CGI of the active gene *ActB* although the enrichment in FLAG-MBD2 cells compared to

Mbd2-/- cells was less than at *Xist* (Figure 4.5.8A). This could indicate non-specific binding of FLAG-MBD2 which is expressed at approximately five times the endogenous level in these cells (data not shown). To test this possibility further I performed ChIP over a 9kb region encompassing the *Xist* CGI in FLAG-MBD2 and *Mbd2-/-* cells. The ChIP signal was higher in the FLAG-MBD2 cells compared to *Mbd2-/-* cells throughout most of the region with a slight increase over the methylated CGI (Figure 4.5.8B).

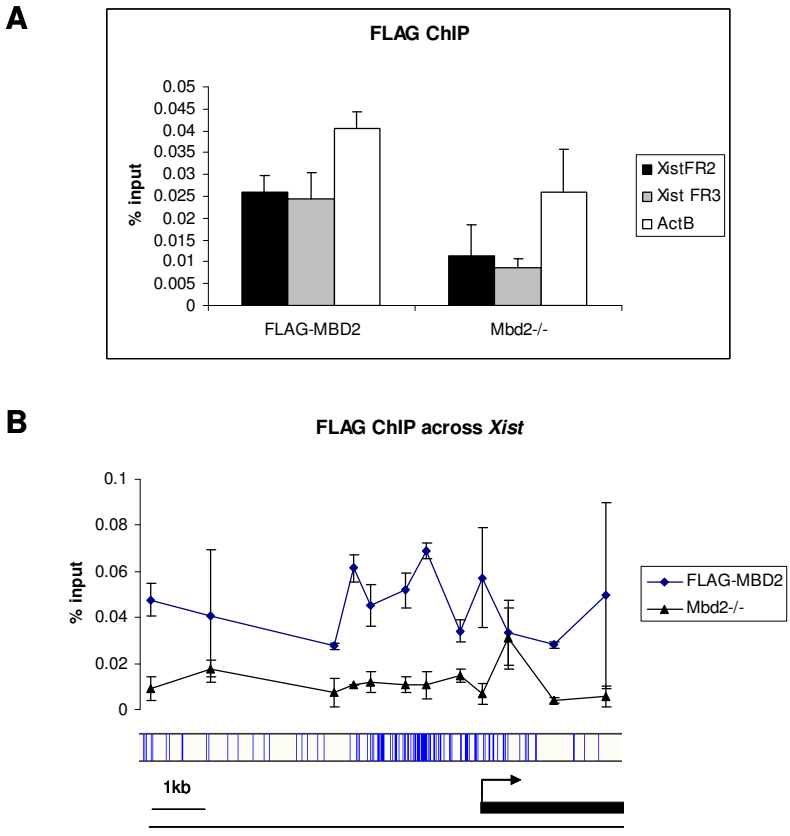


Figure 4.5.8 FLAG ChIP across *Xist*

ChIP using α-FLAG in FLAG-MBD2 and *Mbd2-/-* MTFs at **A)** *Xist* promoter (FR2 and FR3) and the *ActB* promoter **B)** Across a 9kb region of the *Xist* locus. Blue dashes underneath the ChIP profile indicate CpG density, the black box indicates exon 1 of *Xist* and the arrow indicates the direction of transcription.

Subsequent studies in the lab have shown that FLAG-MBD2 binding to the *Xist* CGI can be detected only when the protein is expressed at least five times the endogenous level (H. Owen, unpublished results). One explanation for this result is that, when over-expressed, a large proportion of MBD2 does not associate with the other NuRD

complex components and therefore is easier to immunoprecipitate as it is not obscured by the rest of the complex. Indeed, gel filtration experiments show that in cells over-expressing FLAG-MBD2, the majority of MBD2 does not associate with the NuRD complex (H. Owen, unpublished). It is therefore unclear whether binding of over-expressed FLAG-MBD2 reflects binding of the endogenous protein or is associated with the NuRD complex. For this reason, as well as the fact that the increase in ChIP signal in FLAG-MBD2 cells compared to *Mbd2-/-* cells is relatively low, I decided to discontinue FLAG-MBD2 ChIP experiments.

4.5.5 Formaldehyde cross-linking does not efficiently capture MBD2-DNA interactions

Another explanation for my inability to recover DNA during MBD2 ChIP despite recovery of protein by cross-linked IP is that cross-linking of MBD2 to the DNA is inefficient. Studies using Fluorescence Recovery After Photobleaching (FRAP) have shown that the heterochromatic localisation of proteins with a recovery half-time of 2.5 seconds or less is not captured by formaldehyde cross-linking (Schmiedeberg et al., 2009). Studies carried out in my lab have shown that in NIH3T3 fibroblast cells MBD2b-GFP has a FRAP half-time of 4.5 seconds (L. Schmiedeberg, unpublished). Schmiedeberg and colleagues found that proteins with half-times of this magnitude showed partial mis-localisation when formaldehyde-fixed cells were examined by immunofluorescence but proper localisation in live cells where fixation was not required to capture interactions (Schmiedeberg et al., 2009). Interestingly, this also translated into a reduced efficiency in ChIP experiments. To examine whether this could also be the case for MBD2, I transfected NIH3T3 cells with MBD2b-GFP and examined MBD2 localisation in fixed and live cells. When MBD2b-GFP localisation was examined in cells that had been fixed with formaldehyde approximately half of the cells showed preferential localisation of MBD2 to densely methylated heterochromatic foci (indicated by DAPI bright spots) but the rest of the cells showed diffuse nuclear localisation of MBD2 (Figure 4.5.9A). In contrast, when live cells were examined all of them showed correct heterochromatic localisation of MBD2b-GFP (Figure 4.5.9B). This suggests that formaldehyde may be inefficient at

capturing MBD2-chromatin interactions due to a relatively rapid exchange rate of MBD2 on DNA as indicated by its short recovery half-time in FRAP experiments.

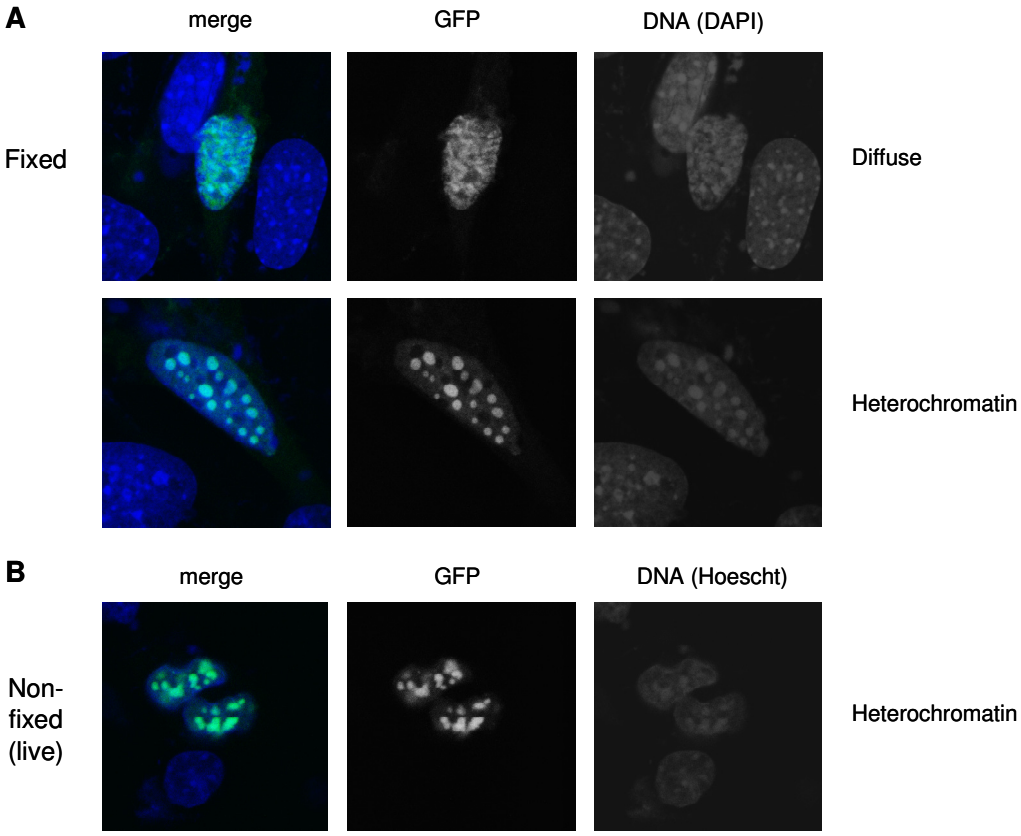


Figure 4.5.9 Formaldehyde cross-linking fails to efficiently capture MBD2-DNA interactions

A) Localisation of MBD2b-GFP in cells fixed with formaldehyde. Cells show either diffuse nuclear localisation of MBD2 or preferential localisation to heterochromatic foci. **B)** In live cells MBD2b-GFP always preferentially associates with heterochromatin. GFP = MBD2b-GFP (green in merge), DAPI/Hoechst = DNA (blue in merge).

Inefficient cross-linking of MBD2 protein to DNA could partially explain my difficulty in performing ChIP for MBD2. To try and overcome this problem I included an additional cross-linker dimethyl 3,3´-dithiobispropionimidate (DTBP; Pierce) in my ChIP experiments which is reported to cross-link proteins at greater distances than formaldehyde (11.9 Å compared to 2Å for formaldehyde) although it is more efficient at capturing protein-protein interactions than protein-DNA interactions (Fujita and Wade, 2004). However, no difference was observed in ChIP

over *Xist* or major satellite using this cross-linker compared to formaldehyde alone (data not shown).

4.6 Results: Analysis of histone H3 acetylation in Mbd2-/- Th1 cells

Histone acetylation is a mark of gene activity and is enriched over the promoters of active genes (Wang et al., 2008). Deacetylation, on the other hand, is associated with gene repression and many of the MBD proteins recruit HDACs in order to facilitate transcriptional silencing (see section 1.3). As MBD2 is a component of the HDAC-containing NuRD co-repressor complex it is expected that absence of MBD2 may lead to increases in histone acetylation due to a failure to recruit HDACs. Therefore, acetylation of histone H3 was examined in *Mbd2-/-* Th1 cells by Western Blot and ChIP-seq analysis.

4.6.1 Global levels of histone H3 acetylation are not detectably changed in the absence of MBD2

I used Western blot analysis to examine whether there was a change in the overall level of histone acetylation in *Mbd2-/-* Th1 cells compared to wild-type Th1 cells. Western blots were probed with an antibody with specificity for histone H3 when it is acetylated at the K9 and K14 residues (anti-histone H3 pan-acetyl; Millipore) and blots against unmodified histone H3 (Abcam) were carried out in parallel to act as a loading control. Signal was detected using chemiluminescence (Figure 4.6.1A) and blots were quantified using densitometry. The level of histone H3 acetylation (H3Ac) was expressed relative to total histone H3 (H3) (Figure 4.6.1B). Levels of H3Ac were found to be similar between wild-type and *Mbd2-/-* cells at least at the sensitivity detectable by Western blot.

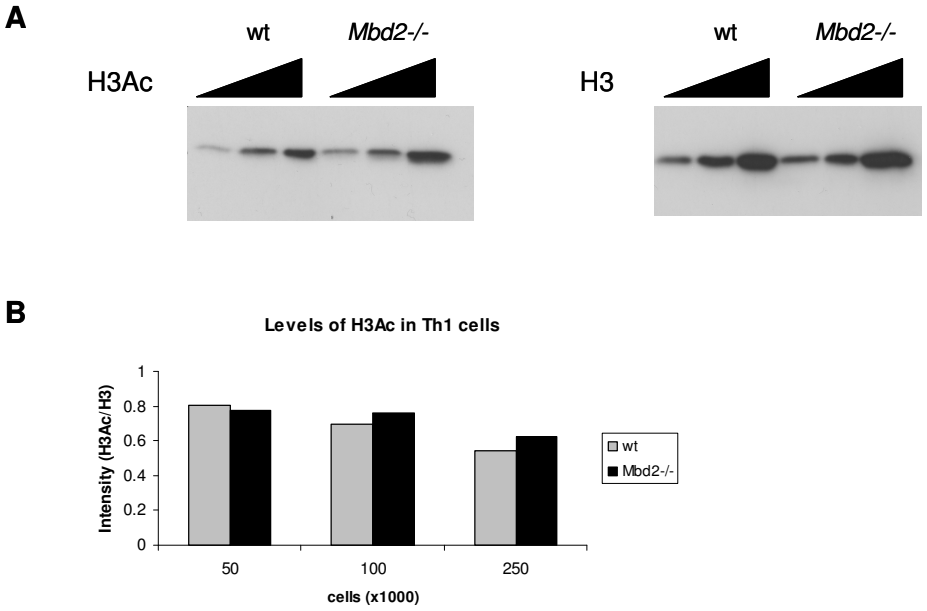


Figure 4.6.1 Western blot analysis of H3Ac levels in wt and *Mbd2-/-* Th1 cells

A) Western blots for acetylated histone H3 (left) and total histone H3 (right). Increasing numbers of cells were loaded (50,000, 100,000, 250,000). **B)** Quantification of Western blots by densitometry. For each cell concentration the level of H3Ac was expressed relative to total H3.

4.6.2 Genome-wide profiling of H3Ac by ChIP shows differences between wild-type and *Mbd2-/-* Th1 cells

In order to see if absence of MBD2 affected histone acetylation at particular genomic locations I profiled H3Ac in wild-type and *Mbd2-/-* Th1 cells using ChIP-seq. ChIP was carried out using the histone H3 pan-acetyl antibody and enrichment for H3Ac was verified by examining the active gene *c-myc*. At *c-myc* the level of H3Ac was comparable between wild-type and *Mbd2-/-* Th1 cells and consistent between experimental replicates (Figure 4.6.2). Illumina adaptors were attached to the remaining ChIP DNA and it was passed to the Wellcome Trust Sanger Centre for sequencing and mapping as previously described (section 3.3.3). Replicate lanes of sequencing were combined giving two lanes of sequencing for each sample.

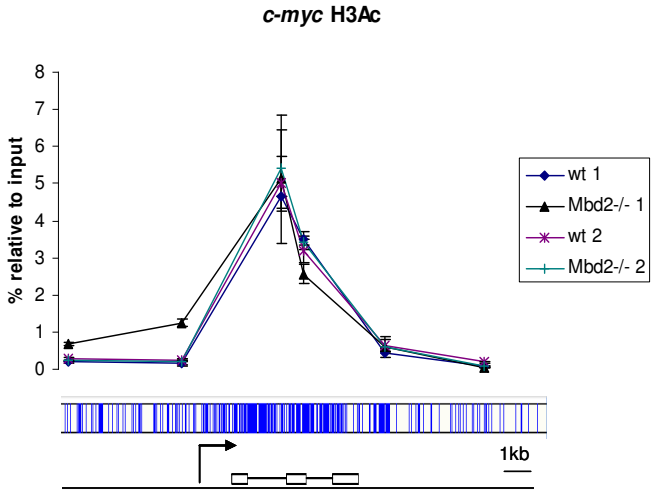


Figure 4.6.2 Verification of H3Ac enrichment at *c-myc* in wt and *Mbd2-/-* Th1 cells
Enrichment for H3Ac was tested over the *c-myc* locus prior to sequencing. Blue dashes underneath the profile indicate CpG density and a schematic of the gene is shown with the direction of transcription indicated by the arrow.

There was a 1.4-fold difference in sequencing coverage between the wild-type and *Mbd2-/-* samples (Table 4.6.1, upper panel). It is possible that this reflects a greater level of H3Ac in *Mbd2-/-* Th1 cells compared to wild-type Th1 cells although analysis of H3Ac levels by Western blot as well as ChIP analysis of the *c-myc* locus argue against this. Alternatively, the difference in read number may simply be reflective of better sequencing depth in the *Mbd2-/-* sample. ChIP-seq data were normalised to the average number of reads for the two samples in order to correct for this and analysis performed using the principles outlined in section 3.3 (Table 4.6.1, lower panel).

Sample	**Number of reads generated (Maq ≥ 30)**	**Number of bp mapped**
WT H3Ac	21,768,054	664,150,962
Mbd2-/- H3Ac	30,477,938	924,657,232

Analysis step	Parameters
Normalisation	Scaled to 794 million reads
Peak-finding	H: 4 L: 150 G: 250
Sliding window	100bp window, 20bp slide
Identification of differences	log2 difference > 1.7 in 9/13 contiguous windows

Table 4.6.1 Sequencing information and parameters for analysis of H3Ac ChIP-seq data

Upper panel; number of 37bp sequencing reads generated for each cell type and the number of base pairs (bp) of mapped sequence obtained. Lower panel; parameters for each analysis step. H = height in number of reads, L = length in bp, G = gap in bp.

Peak-finding followed by sliding window analysis identified 140 regions that showed increased H3Ac and 45 regions that showed decreased H3Ac in the absence of MBD2. I looked to see if any of the genes associated with these regions changed their expression in *Mbd2-/-* Th1 cells. Surprisingly only three genes; *Fam83g*, *Slc15a3* and *Plxnd1* showed increased expression and increased H3 acetylation in *Mbd2-/-* Th1 cells compared to wild-type (Figure 4.6.3). None of the genes downregulated in the absence of MBD2 showed a change in acetylation. *Fam83g* is a novel gene of unknown function, *Slc15a3* codes for a peptide transporter while *Plxnd1* encodes a transmembrane receptor reported to play a role in thymocyte maturation (Choi et al., 2008). None of these genes are master regulators in the immune system so it is unlikely that an increase in their acetylation and gene expression will affect a large number of other genes. The observation that many genes that change expression in *Mbd2-/-* Th1 cells do not show changes in H3Ac could indicate that absence of MBD2 has more subtle effects on histone acetylation, as my analysis parameters identify relatively large differences (3.25-fold). Conversely, many genes that show a change in histone acetylation in *Mbd2-/-* cells do not change their expression. This could suggest that there are multiple mechanisms repressing each gene and that an increase in acetylation due to removal of MBD2 alone is not always sufficient to induce expression. Alternatively it could suggest that the appropriate transcription factors are not present to activate expression of these genes in Th1 cells even if repression is compromised by the absence of MBD2.

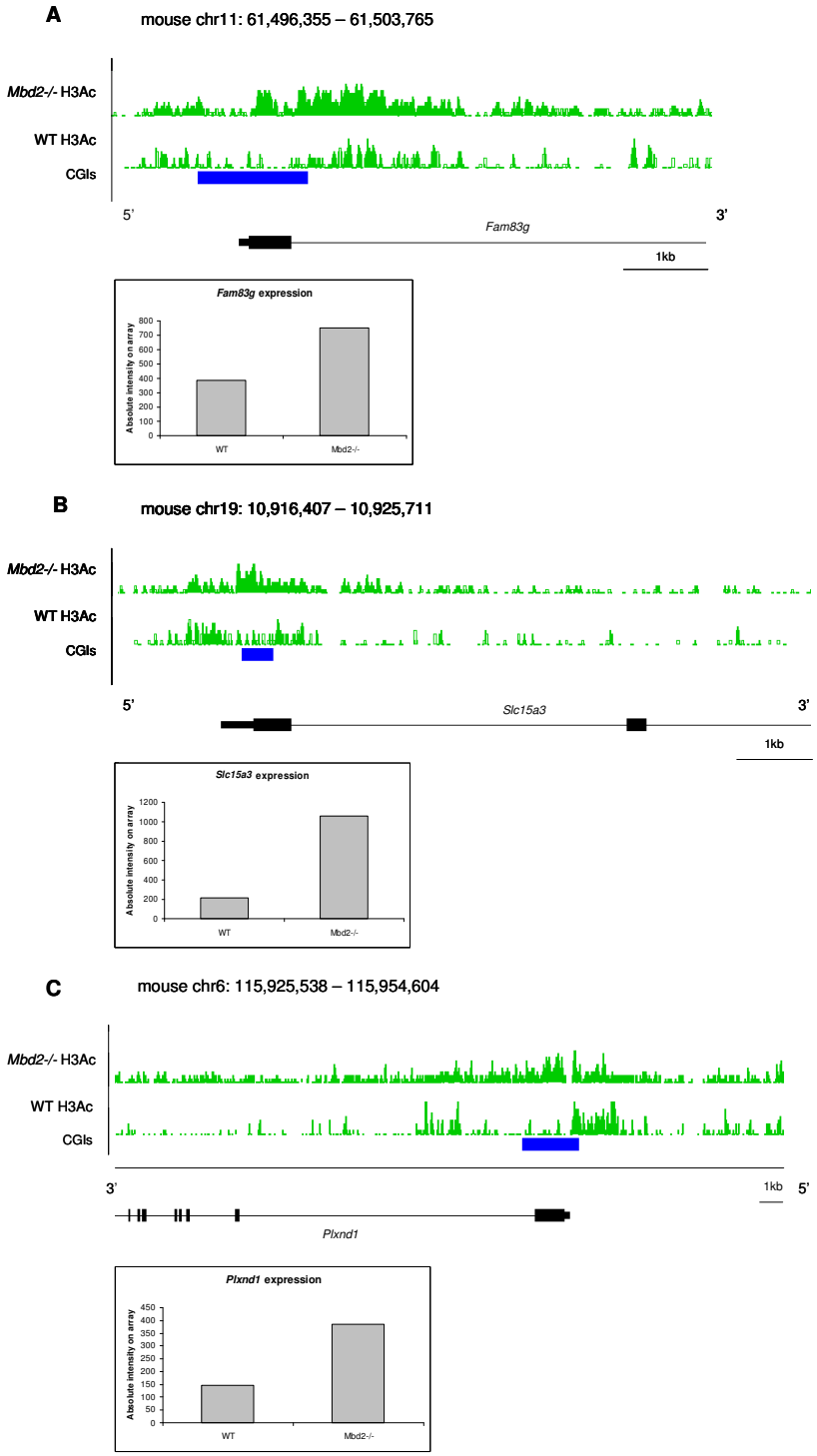


Figure 4.6.3 Three genes have increased H3Ac and increased expression in *Mbd2-/-* Th1 cells compared to wild-type Th1 cells

A) *Fam83g* (also known as *2310040C09Rik*) **B)** *Slc15a3* **C)** *Plxnd1*. For each example the upper panel shows read density profiles for by H3Ac ChIP-seq (green) and CGIs (blue). The lower panel shows the expression of each gene in wild-type (WT) and *Mbd2-/-* Th1 cells as determined by microarray analysis.

I then investigated if genes upregulated in the absence of *Mbd2* showed a more subtle tendency to increase histone H3 acetylation. Using normalised data, I examined the average H3Ac profile for a region 3kb upstream and 3kb downstream of the transcription start site in wild-type and *Mbd2-/-* samples. This was carried out for all RefSeq genes in the mouse genome and the profile compared to that of genes upregulated in *Mbd2-/-* Th1 cells (analysis carried out with assistance from Shaun Webb). For the analysis, only genomic regions without overlapping genes were included to avoid complications due to intersection of multiple TSS regions. When the average profile of H3Ac across all TSSs was examined, as expected, acetylation was enriched 1kb either side of the TSS (Wang et al., 2008) and a dip in H3Ac signal was seen just before the TSS, indicative of the nucleosome free region characteristic of active promoters and promoters poised for transcription (Ozsolak et al., 2007). On average, a 2.25-fold increase in H3Ac was seen in *Mbd2-/-* Th1 cells compared to wild-type Th1 cells when all genes were examined (1kb up- and downstream of the TSS) (Figure 4.6.4A). This could indicate that absence of MBD2 affects histone acetylation at many gene promoters. Although the data were normalised prior to analysis, however, the possibility that this difference could be due to increased sequencing depth in the *Mbd2-/-* sample cannot be ruled out. H3 acetylation was then examined at the set of genes upregulated in *Mbd2-/-* Th1 cells compared to wild-type Th1 cells. At these genes, H3Ac was increased by an average of 2.5-fold over the region 1kb up- and downstream of the TSS in *Mbd2-/-* compared to wild-type Th1 cells, similar to the increase seen at the TSS for all genes (Figure 4.6.4B). However, between 0-500bp downstream of the TSS the difference in acetylation was more pronounced for genes upregulated in the absence of MBD2 that for all genes (2.7-fold compared to 1.7-fold) (Figure 4.6.4A and B).

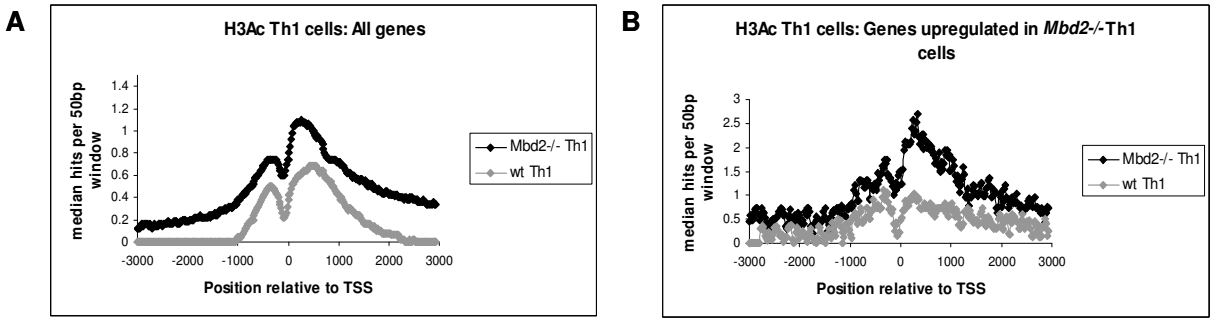


Figure 4.6.4 The average H3Ac ChIP-seq profile at transcription start sites

H3Ac ChIP-seq hits (calculated in 50bp window, 20bp slide, the median number of hits per window is plotted) for a region 3kb upstream and 3kb downstream of the TSS in wild-type (grey) and *Mbd2-/-* (black) Th1 cells in: **A)** all non-overlapping RefSeq genes (n=1700) **B)** non-overlapping genes upregulated in *Mbd2-/-* compared to wt Th1 cells (n=66).

4.7 Results: The relationship between MBD2-regulated gene expression and DNA methylation

I wanted to see whether genes that increased their expression in the absence of MBD2 were methylated and whether differential methylation could be responsible for directing MBD2 to cell type-specific targets. I also examined whether removal of MBD2, a methylation reader, could affect the methylation pattern itself.

4.7.1 The majority of genes upregulated in *Mbd2-/-* cells are methylated

A major question remaining about MBD2 function is whether methylation is responsible for directing MBD2/NuRD binding *in vivo*. To help address this question in the absence of ChIP data for MBD2 we examined whether genes that changed their expression in the absence of MBD2 were methylated. Gene expression data for *Mbd2-/-* dendritic cells and Th1 cells were compared to the methylation data generated for these cells using MAP-seq (chapter 3). These cell types were chosen as they show the greatest number of gene expression changes in the absence of MBD2. In total, 56% of genes upregulated in dendritic cells and 72% of genes upregulated in Th1 cells were associated with a peak of methylation determined by MAP-seq. A modest proportion of these upregulated genes associated with a methylated CGI

(17% in dendritic cells and 27% in Th1 cells) with the remainder of methylation occurring at non-CGI-associated regions. Some of the genes that associated with methylated CGIs were also associated with methylated non-CGI regions (Figure 4.7.1). These methylated CGIs tended to be located in intragenic regions rather than at TSSs consistent with the fact that these CGIs are frequently methylated (Illingworth et al., 2010; Maunakea et al., 2010).

A significant proportion of MBD2-regulated genes were not associated with methylation in the relevant cell type. However, this does not mean that these genes do not contain methylated CpG sites as MAP-seq can only detect methylation at relatively CpG-rich regions and most of the genome is CpG-deficient (see section 3.3). This could be addressed by using a technique, such as bisulfite sequencing, that can detect DNA methylation regardless of CpG density.

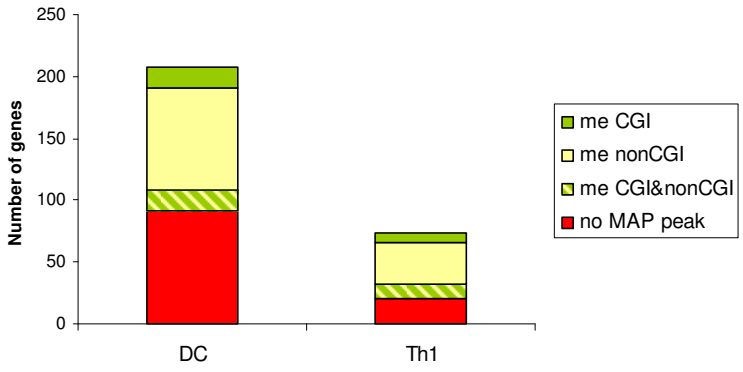


Figure 4.7.1 Association of genes upregulated in the absence of MBD2 with DNA methylation

Genes upregulated in *Mbd2-/-* dendritic cells (DC) or Th1 cells (Th1) were examined for association with methylated peaks identified by MAP-seq in the relevant cell type. Genes were categorised as overlapping a CGI peak (meCGI), a non-CGI-associated peak (me nonCGI), both (meCGI&nonCGI) or as not overlapping methylation detected by MAP-seq (no MAP peak).

4.7.2 MBD2-regulated genes are generally not associated with cell type-specific DNA methylation

Another question regarding the relationship between MBD2 and the methylation pattern is whether cell type-specific DNA methylation directs MBD2 to different

targets in different cell types. I examined whether genes that were upregulated in the absence of MBD2 only in specific immune cells also showed cell-type specific methylation patterns. I investigated 201 genes upregulated in *Mbd2-/-* dendritic cells that are not mis-expressed in *Mbd2-/-* Th1 cells (i.e. dendritic cell-specific targets). Of these, only four genes showed dendritic cell-specific methylation suggesting that differential methylation is not the major mechanism directing MBD2-mediated gene repression. This is in keeping with the observation that methylation patterns are relatively static between immune cell types and are therefore unlikely to be a major mechanism governing gene regulation (section 3.5). Of the genes that do show dendritic cell-specific methylation, the *Socs3*, *Tnfaip2* and *Scnn1a* genes are methylated at intragenic CGIs while *Hk3* is differentially methylated at a non-CGI-associated region (MAP-seq and expression data for *Socs3* and *Tnfaip2* are shown in Figure 4.7.2A and B). *Socs3* codes for suppressor of cytokine signalling 3 and regulates signalling during a number of immune system processes including T cell differentiation (Egwuagu et al., 2002; Kubo et al., 2003; Yu et al., 2003) while *Tnfaip2* is a pro-inflammatory gene [see section 3.5.2 and (Mookherjee et al., 2006)]. It is hypothesised that these four genes are direct targets of MBD2 in dendritic cells but not in Th1 cells where they are hypomethylated. Mis-expression of these genes may have secondary effects on the expression of other genes which could account for some of the other expression changes occurring in *Mbd2-/-* dendritic cells.

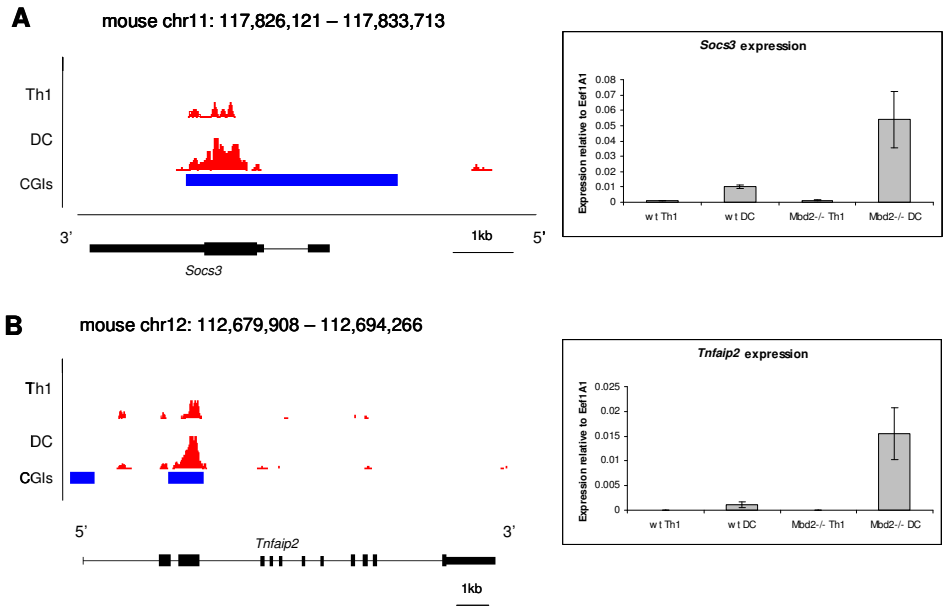


Figure 4.7.2 Genes upregulated in *Mbd2-/-* dendritic cells showing dendritic cell-specific DNA methylation

A) *Socs3* and **B)** *Tnfaip2* which show increased methylation in dendritic cells compared to Th1 cells by MAP-seq and are upregulated in *Mbd2-/-* dendritic cells. For each the left hand panel shows the MAP-seq read density profile (red) and CGIs (blue) while the right hand panel shows expression of the genes (relative to *Eef1A1*) as determined by qRT-PCR.

4.7.3 DNA methylation is not changed in the absence of MBD2

Given that MBD2 is involved in reading the methylation pattern I investigated whether its absence affected DNA methylation itself. One hypothesis is that DNA methylation may not be efficiently maintained at MBD2 target genes in *Mbd2-/-* cells because methylation is not functional without MBD2 present to mediate its effect. Alternatively, if MBD2 can act as a DNA demethylase (Bhattacharya et al., 1999) DNA methylation would be expected to increase in *Mbd2-/-* cells compared to wild-type cells. To address this issue, MAP-seq in wild-type and *Mbd2-/-* CD4 T cells was carried out. Two separate MAP-seq experiments were carried out for wild-type CD4 cells. Each of these wild-type samples was compared to *Mbd2-/-* CD4 MAP-seq data separately and only differential methylation identified in both comparisons was taken as genuine. This compensated for variability between samples run at different times on different batches of MBD column. Differential methylation was identified as before (section 3.3) and sequencing information is outlined in Table 4.7.1.

Sample	Number of reads generated (Maq ≥ 30)	Number of bp mapped
WT CD4 1	25,005,336	427,865,743
WT CD4 2	39,376,768	426,508,805
Mbd2-/- CD4	35,855,841	394,428,880

Table 4.7.1 Sequencing information for MAP in wild-type and *Mbd2-/-* CD4 cells

Number of 37bp reads generated for each cell type and the number of base pairs (bp) of mapped sequence obtained.

In general, DNA methylation detectable by MAP-seq did not change between wild-type and *Mbd2-/-* CD4 cells. Six regions showed an increase in methylation in *Mbd2-/-* cells compared to wild-type while just one region showed a decrease in methylation. Four of the regions are located within genes although none of these genes changed expression in the absence of MBD2 in any of the cell types tested. Two of the regions overlapped a methylated CGI. Details of methylation changes in *Mbd2-/-* compared to wild-type CD4 cells are summarised in Table 4.7.2 and examples of differences are shown in Figure 4.7.3A and B.

Region	Gain/loss of methylation in *Mbd2-/-*	Gene association	Methylated CGI association	Exp change in *Mbd2-/-*
chr2: 33789520 - 33790020	gain	no	no	no
chr5: 28495080 - 28495680	gain	*En2*	yes	no
chr5: 39458460 - 39458900	gain	no	no	no
chr7: 85291980 - 85292460	gain	no	no	no
chr16: 87901340 - 87901800	gain	*Grik1*	no	no
chr17: 85012720 - 85013180	gain	*Plekhh2*	no	no
chr14: 119321880 -119322300	loss	*Dzip1*	yes	no

Table 4.7.2 Regions showing a methylation difference in *Mbd2-/-* CD4 cells compared to wild-type CD4 cells

Chromosomal location, whether the region shows gains or loss of methylation in *Mbd2-/-* cells, gene association and association with methylated CGIs are shown. None of the regions is associated with can expression change in *Mbd2-/-* cells.

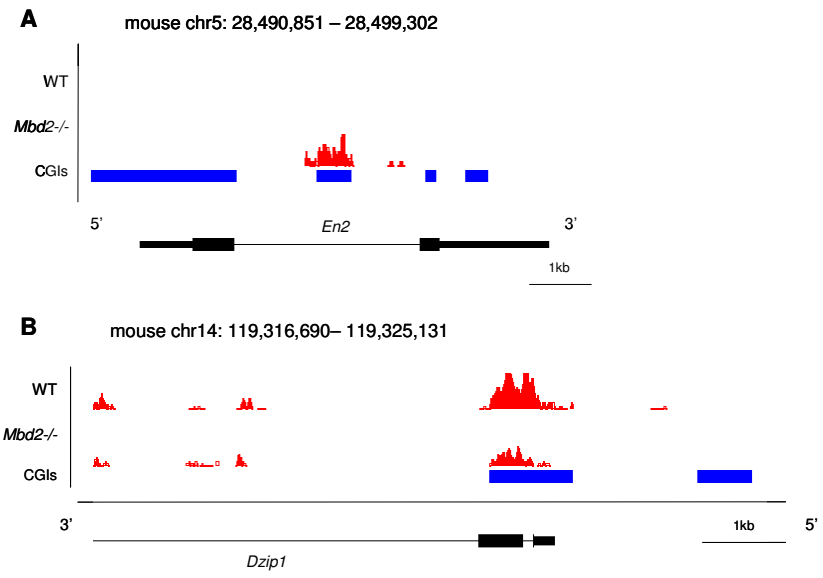


Figure 4.7.3 *En2* and *Dzip1* are two genes associated with a methylation difference in *Mbd2-/-* compared to wild-type CD4 cells

MAP-seq profiles (red) are shown for **A)** *En2* CGI that increases methylation in *Mbd2-/-* cells **B)** *Dzip1* CGI that looses methylation in *Mbd2-/-* cells. CGIs are indicated by blue bars.

4.8 Discussion

As MBD2 has previously reported roles in the immune system (Hutchins et al., 2005; Hutchins et al., 2002; Kersh, 2006) this represented a good system in which to study MBD2's role in gene expression and how this relates to DNA methylation. Reported defects in T helper cell differentiation in *Mbd2-/-* mice were confirmed by FACS staining for cytokine production and by analysis of gene expression using qRT-PCR. Additional immune system defects were observed in *Mbd2-/-* mice including impaired T cell proliferation, increased production of activated Treg cells *in vivo* and increased susceptibility to oesophageal infection. This prompted us to assess the extent to which gene expression is perturbed in immune system cells lacking MBD2. The idea that MBD2 is involved in regulating inducible processes such as the immune response is intriguing as this might explain the relatively mild phenotype of *Mbd2-/-* mice in the absence of immune challenge (Hendrich et al., 2001).

4.8.1 MBD2 regulates gene expression in the immune system, perhaps in a redundant manner

A modest number of gene expression changes were seen in *Mbd2-/-* immune cells compared to wild-type with the greatest number of differences occurring in dendritic cells. This suggests a prominent role for MBD2 in dendritic cells which is supported by collaborators who have observed that dendritic cells from *Mbd2-/-* mice seem to be defective in directing a Th2-type immune response (P. Cook, unpublished results). The majority of genes that change their expression in the absence of MBD2 were upregulated, consistent with MBD2's role as a transcriptional repressor. The number of gene expression changes (between 16 and 218 genes depending on the cell type) occurring in the absence of MBD2 suggests that MBD2 is not a major regulator of gene expression or, at least, that it has a partially redundant function. As the NuRD complex can contain MBD3 instead of MBD2 (Le Guezennec et al., 2006) it is possible that MBD3/NuRD could compensate for a lack of MBD2 although in this case repression would not be directed by methylation as MBD3 does not preferentially bind methylated DNA (Hendrich and Bird, 1998). This possibility could be tested by knocking out MBD2 and MBD3 specifically in the immune

system which would circumvent the issue of embryonic lethality in the absence of MBD3 (Hendrich et al., 2001). Another possibility is that MBD2 is one of a number of redundant repressive mechanisms which are "stacked" at particular loci to ensure efficient gene silencing. This has been suggested to be the case at the *Xist* locus where absence of MBD2 leads to low level reactivation of *Xist* but renders *Xist* hypersensitive to the effects of impairing histone deacetylation and DNA methylation (Barr et al., 2007). A final and related explanation could be that *Mbd2-/-* mice have adapted to compensate for a lack of MBD2 during development. It would be intriguing to examine whether more gene expression changes would occur after sudden removal of MBD2 specifically in adult cells using a conditional knockout strategy. MBD2 stands out from the methyl-binding proteins MBD1 and MeCP2 as significant numbers of genes change expression in *Mbd2-/-* cells and most of these are upregulated consistent with a role as a transcriptional repressor.

4.8.2 Absence of MBD2 results in different gene expression changes in different immune cell types

I have shown that removal of MBD2 results in distinct expression changes in different cell types. It is possible that MBD2 binds the same sites in these cells but that genes are only upregulated in the absence of MBD2 when the appropriate cell-specific transcription factors are present or when the chromatin environment is permissive to transcription. An alternative is that MBD2 binds to different sites in different cells, perhaps due to differential methylation in these cell types. I examined whether genes upregulated specifically in *Mbd2-/-* dendritic-cell were associated with dendritic cell-specific DNA methylation. This was not the case for the majority of genes, consistent with the relatively conserved patterns of methylation observed between immune system cell types (see chapter 3). However, the four genes, *Socs3*, *Tnfaip2*, *Scnn1a* and *Hk3* that do show cell-specific methylation and regulation by MBD2 may be directly bound by MBD2 in a methylation-dependent manner. It is possible that upregulation of these genes in the absence of MBD2 may have secondary effects on other genes and be responsible for many of the gene expression changes observed in *Mbd2-/-* dendritic cells. Mapping MBD2 binding sites in different immune cell types will be necessary to ultimately determine the relationship

between the methylation pattern and MBD2 binding. However, as this is currently not possible (see section 4.7), the four genes showing MBD2-dependent repression and cell-specific methylation could be overexpressed and the effects on gene expression observed to see if any of the genes upregulated in *Mbd2-/-* dendritic cell are also upregulated in this context.

4.8.3 MBD2 binding sites have not yet been determined

To get to the heart of many of the aspects of MBD2 function discussed it is necessary to determine where MBD2 binds *in vivo*. Unfortunately, numerous attempts to carry out ChIP for MBD2 were unsuccessful. My lab's custom made MBD2 antibody, S923, despite immunoprecipitating MBD2 protein, did not detect MBD2 at published targets. Furthermore, use of a microarray-based strategy to identify novel targets also proved unsuccessful. This could be due to the fact that only a small percentage of MBD2 is immunoprecipitated by S923 and this is not enough to give a signal above background level in ChIP experiments. Alternatively cross-linking of MBD2 to the DNA could be inefficient resulting in recovery of protein but compromised recovery of the associated DNA. This has shown to be the case for mutants of MeCP2 which show aberrant localisation in formaldehyde-fixed cells and reduced binding in ChIP experiments (Schmiedeberg et al., 2009). Also, since MBD2 is part of a large protein complex if there are instances where it is not directly bound to DNA it may be more difficult to capture interactions by cross-linking. Immunoprecipitation of FLAG-tagged MBD2 using an α-FLAG antibody recovered the majority of MBD2 in the cell; however, specific association of MBD2 with known targets was not detected. Rather, FLAG-MBD2 cells showed a modest increase in ChIP signal to *Mbd2-/-* cells at all regions tested including the promoter of the active gene *ActB*. Therefore, it seems that MBD2 ChIP is either not efficient enough (using S923 in wild-type cells) or not specific enough (using α-FLAG in FLAG-MBD2 cells) to be used to determine genome-wide patterns of MBD2 binding in immune cells. This represents a major barrier to elucidating the extent and role of MBD2 in regulating gene expression. Strategies for circumventing this problem are discussed in section 5.2.2.

4.8.4 Changes in H3Ac were observed in the absence of MBD2

As MBD2 is part of a complex containing HDAC activity I examined whether removal of MBD2 was associated with an increase in acetylation of histone H3. Total levels of histone acetylation as assessed by Western blot were not affected in *Mbd2-/-* Th1 cells consistent with the abundance of MBD2 in the nucleus. Estimates suggest that there is approximately one molecule of MBD2 for every 36 nucleosomes in the cell (R. Ekiert, unpublished results). This is in contrast to MeCP2 which, in neurons, is nearly as abundant as the histone octamer and, as a consequence, its removal affects histone acetylation levels globally (Skene et al., 2010). However, ChIP-seq analysis of H3Ac distribution in wild-type and *Mbd2-/-* Th1 cells suggested that there may be a global increase in H3 acetylation over gene promoters in *Mbd2-/-* cells. This could be due to a failure to recruit the NuRD complex and associated HDACs to a large number of promoters although it is also possible that the observed difference might be due simply to a discrepancy in the number of sequencing reads between wild-type and *Mbd2-/-* samples which may have been more pronounced over promoters. To distinguish between these possibilities mapping of MBD2 binding sites in Th1 cells would need to be carried out to investigate whether MBD2 and other NuRD components bind a large number of gene promoters. In addition, Western blots using a more sensitive fluorescence detection method rather than chemiluminescence should be carried out to investigate whether a global difference in the level of histone acetylation in *Mbd2-/-* Th1 cells can be detected. It is possible that a 1.4-fold change in acetylation (which would be predicted over the whole genome from the difference in sequencing reads) might be too small to be detectable by Western blot or those changes in H3Ac level occurring specifically at promoters are diluted out by the rest of the H3Ac in the nucleus. This could explain the apparent discrepancy between the Western blot and ChIP-seq results.

If MBD2 has a role in regulating acetylation at a large number of gene promoters, this is in conflict with the fact that only a modest number of genes change their expression in its absence. One explanation for this may be that, at many MBD2 targets, redundant silencing mechanisms act together to prevent transcription. In

support of this, peak-finding analysis carried out on H3Ac ChIP-seq data identified many regions that showed a large increase in H3Ac in *Mbd2-/-* Th1 cells but were not associated with changes in gene expression.

4.8.5 The methylation dependence of MBD2/NuRD binding needs to be established *in vivo*

When genes upregulated in *Mbd2-/-* dendritic cells or Th1 cells were examined for association with DNA methylation in these cell types, although the majority (56% in dendritic cells and 71% in Th1 cells) were associated with DNA methylation, a sizeable proportion of genes did not associate with methylation detectable by MAP-seq. One explanation for this may be that these regions are methylated but CpG-poor and cannot be interrogated by MAP-seq. Another possibility that can't be excluded is that, although MBD2 preferentially binds to methylated CpGs *in vitro* and in reporter assays (Ng et al., 1999), other NuRD complex components or sequence-specific transcription factors (Miccio et al., 2010; Sridharan and Smale, 2007) can direct MBD2/NuRD to unmethylated targets. To assess the relationship between MBD2/NuRD binding and DNA methylation *in vivo* it would be necessary to identify MBD2 binding sites in the cell and assess their methylation status, ideally by a method such as bisulfite sequencing that can assess methylation at CpG-poor regions.

4.8.6 MBD2 does not act as a DNA demethylase

Assessment of DNA methylation by MAP-seq in wild-type and *Mbd2-/-* CD4 T cells revealed no global effect on methylation due to the absence of MBD2. This supports studies claiming that MBD2 does not function as a DNA demethylase (Hendrich et al., 2001; Ng et al., 1999; Santos et al., 2002) as previously reported (Bhattacharya et al., 1999). However, seven regions showed a difference in methylation between wild-type and *Mbd2-/-* cells. One explanation for these differences is that wild-type and *Mbd2-/-* mice, while both being on a C57/BL6 background, were not from the same mouse line and there could be methylation differences between these lines. Bisulfite

sequencing of some of these regions will be necessary in order to determine if these changes are genuine.

Chapter 5 Discussion

This thesis has examined cell type-specific CGI methylation and the role of the methyl-binding protein MBD2 using cells of the immune system as a model. Although somatic CGI methylation in normal cells has been well-documented in recent years (Eckhardt et al., 2006; Illingworth et al., 2008; Mohn et al., 2008; Schilling and Rehli, 2007; Weber et al., 2007) the extent and role of cell type-specific methylation and, in particular its relationship to gene expression was still in need of further investigation. In addition, global gene expression changes in MBD2-deficient primary single cell types had not yet been assessed. This is in spite of the fact that, unlike other MBD proteins, numerous examples of gene mis-regulation have been reported in *Mbd2-/-* mice (Barr et al., 2007; Berger et al., 2007; Hutchins et al., 2002). Immune cells are developmentally related, easily purified cell types that have documented instances of cell-specific DNA methylation and reported phenotypes in the absence of MBD2 making them particularly suitable for this kind of investigation. The main findings on cell-specific CGI methylation and the effect of removing MBD2 in immune system cells are discussed here along with possibilities for future work.

5.1 Discussion

5.1.1 The relationship between CGI methylation and gene expression

CGIs represent a distinct fraction of the genome in terms of their CpG density, usually unmethylated state and association with a transcriptionally permissive chromatin environment. Methylation of CGIs has long been associated with gene repression (Csankovszki et al., 2001; Edwards and Ferguson-Smith, 2007). Many CGIs associate with gene promoters and the remainder have been suggested to represent novel transcription start sites (Illingworth et al., 2010; Maunakea et al., 2010). Therefore, where cell-specific CGI methylation occurs and how this relates to

gene expression and histone modification is of particular interest. DNA methylation in a number of cell types from the hematopoietic lineage (dendritic cells, B cells, CD4 T cells, Th1 and Th2 cells) was assessed using MAP-seq with a particular focus on CGIs. To see if trends observed in these developmentally related cells applied to more distinct tissue types, methylation and expression data on immune cells were compared to those on distantly related brain. This showed that findings using an immune system model relate to tissue-specific CGI methylation in general.

It is clear from my results that between highly related cell types, such as immune cells, the DNA methylation pattern is relatively static which is in stark contrast to the vast number of gene expression differences observed between these cells. Notably, during activation and differentiation of CD4 cells into Th1 or Th2 cells nearly 3,000 genes change their expression by at least two-fold but only a handful of CGIs show a change in methylation. This argues that differential gene expression, especially in rapidly inducible systems such as T cell activation, is dependent on cell-specific transcription factors which can exert their effects more rapidly than epigenetic mechanisms. However, DNA methylation and other epigenetic changes may occur more gradually at key genes in order to facilitate continued expression or re-expression of particular genes (for example upon re-exposure to the same immune stimulus). It seems no coincidence that the only CGI to show differential methylation between Th1 and Th2 cells is located in the gene encoding Gata3, the key Th2 cell transcription factor.

A fundamental question regarding the relationship between CGI methylation and gene silencing is whether DNA methylation is a cause or consequence of gene activity. The fact that CGI methylation is often targeted to already silent genes, for example in X inactivation (Csankovszki et al., 2001) and ES cell differentiation (Mohn et al., 2008), suggests that it often follows initial silencing events to help maintain gene repression in a heritable manner. However, rapid methylation and demethylation events have been reported for individual CpG sites (Bruniquel and Schwartz, 2003; Jones and Chen, 2006). I observed specific demethylation of the

Gata3 intragenic CGI during Th2 cell differentiation. After 7 days of T cell culture this methylation change was relatively subtle despite a large increase in *Gata3* expression. However, after an 8 week Th2-inducing infection, the *Gata3* CGI had become completely demethylated. This is consistent with a progressive loss of *Gata3* methylation occurring after initial gene activation rather than active demethylation driving gene expression.

Evidence suggests a developmental origin for CGI methylation patterns. Small numbers of CGI and non-CGI methylation differences were observed between developmentally related immune cells with the fewest differences occurring between the most related cells (i.e. CD4, Th1 and Th2 cells). In contrast, a developmentally distinct tissue, brain, showed 2,000 CGI methylation differences compared to CD4 cells and nearly 20,000 differences at non-CGI regions. Here, total methylation changes actually out-numbered gene expression changes suggesting that the DNA methylation pattern is more related to developmental history rather than it is to the gene expression program and is probably established early during lineage commitment.

Intriguingly, most cell-specific methylation was found to occur at CGIs within gene bodies consistent with the findings of Maunakea et al. (2010). Despite not occurring at the annotated TSS, intragenic CGI methylation tended to associate with decreased gene expression. It is unclear how intragenic CGI methylation might affect expression of the associated gene. For a number of intragenic CGIs, for example those within the *Kcnn4* and *Gata3* genes, increased DNA methylation associated with a decrease in the "active" chromatin mark H3K4me3 and a decrease in gene expression. Such CGIs may represent alternative promoters for the associated gene as suggested by Maunakea and colleagues (2010) or a class of H3K4me3 positive enhancer. However, most differentially methylated intragenic CGIs were constitutively depleted for H3K4me3 in immune cells implying that they do not act as promoters in these cells. Given that many intragenic CGIs show evidence for promoter function (Illingworth et al., 2010; Maunakea et al., 2010) this implies that

these differentially methylated CGIs must act as promoters for novel genes in other tissues or at particular developmental stages as suggested by CAGE and Pol II binding data. Intragenic methylation might then affect expression of the associated annotated gene by an indirect mechanism such as impairment of Pol II elongation (Lorincz et al., 2004). Differential methylation may be stochastic or reflect the use of alternative silencing mechanisms such as polycomb which could be determined by the developmental origin of the cell type.

It is clear from this work and the work of others that methylation of CGIs at annotated transcription start sites is a relatively rare phenomenon with just 2% of all TSS CGIs reported to be methylated (compared to 26-34% of intragenic CGIs) (Illingworth et al., 2010; Maunakea et al., 2010). For the approximately 60% of TSS CGIs that associate with housekeeping genes the reason for this is clear, these genes are ubiquitously expressed and therefore need to be maintained in an unmethylated state. However, many developmental and tissue-specific genes also possess CGI promoters (Illingworth et al., 2008; Larsen et al., 1992). In these cases perhaps DNA methylation is not a fast enough or dynamic enough mechanism to regulate gene expression. Perhaps the strong association of TSS CGIs with H3K4me3 helps keep them free of methylation and specific mechanisms are needed to disrupt this. It is conceivable that TSS CGI methylation is only employed when permanent, long-term silencing is required such as in the case of X inactivation, imprinting and in repression of some germline-specific and developmental genes in somatic cells (Csankovszki et al., 2001; Edwards and Ferguson-Smith, 2007; Illingworth et al., 2008; Mohn et al., 2008). Intragenic CGI methylation, on the other hand, could be more common because these CGIs control expression of novel genes with very restricted expression patterns.

5.1.2 Regulation of gene expression by MBD2

I then examined the role of a methylation "reader", the protein MBD2, in gene regulation in immune cells. Absence of MBD2 was found to result in a number of previously uncharacterised immune system phenotypes and a modest number of gene

expression changes. The abundance of MBD2 in the nucleus together with the number of genes upregulated in *Mbd2-/-* cells suggests a role for MBD2 in repressing a discrete number of target genes or at least in binding discrete genomic sites. This implies that MBD2 functions more like a classical transcriptional repressor than other MBD proteins. *Mbd1-/-* animals show a mild phenotype and very few gene expression changes (Zhao et al., 2003) while *Mecp2y/-* mice show a gross phenotype which mimics the symptoms of Rett syndrome (Guy et al., 2001) along with global changes in the chromatin state (Skene et al., 2010). My ChIP-seq results suggest that *Mbd2-/-* mice may show genome-wide changes in histone acetylation at gene promoters and as well as showing larger magnitude changes at a number of discrete genomic sites. The regions that increase their histone acetylation in *Mbd2-/-* Th1 cells compared to wild-type Th1 cells do not always associate with upregulated genes. In these cases, gene mis-expression may be prevented by other repressive mechanisms which maintain repression in the absence of MBD2-dependent HDAC recruitment. This seems to be the case at the *Xist* locus where MBD2, DNA methylation and histone deacetylation all contribute to gene repression independently (Barr et al., 2007). This "stacking" of repressive tendencies may help to limit the number of gene expression changes occurring in *Mbd2-/-* mice and explain their relatively mild phenotype.

In order to better understand the relationship between MBD2 and DNA methylation, I examined the association of genes upregulated in the absence of MBD2 with methylation detected by MAP-seq. Up to one-fifth of these associated with methylated intragenic CGIs (in Th1 cells). An intriguing possibility is that MBD2 binding to methylated intragenic CGIs could lead to formation of a repressive chromatin environment and thus modulate transcriptional elongation as suggested by Lorincz and colleagues (2004). However, the fact that the majority of genes regulated by MBD2 do not associate with methylated CGIs suggests that MBD2 often binds regions other than methylated CGIs. Indeed, a large proportion of MBD2-regulated genes associate with methylated non-CGI regions detected by MAP-seq. A related issue is whether MBD2 binds to different targets in different cell

types, as suggested by the gene expression data, and how these targets are recognised. I have observed little evidence that cell-specific DNA methylation directs MBD2 to particular target genes. The location of MBD2 binding sites will need to be determined in order to definitively address these questions, which, unfortunately, has so far proved unsuccessful.

5.1.3 The contribution made by cell-specific CGI methylation and MBD2 to gene expression

CGI methylation and MBD2 are part of the same system of gene repression. The fact that DNA methylation is essential for life (Li et al., 1992) but the phenotype of the *Mbd2-/-* mouse is relatively mild (Hendrich et al., 2001) suggests that DNA methylation exerts many of its effects independently of MBD2 and has a more significant role in development. Although I have focused on CGI methylation in this study, the majority of DNA methylation in the genome occurs outwith CGIs, and knockout of the Dnmts does not allow specific assessment of the contribution CGI methylation makes to development, gene regulation and cell survival. One hypothesis is that CGI methylation is necessary for developmental processes such as X inactivation and imprinting and for regulating a small number of key genes whose sustained repression is of paramount importance, but does not play a major role in gene regulation in general. A role in regulating a discrete number of target genes is also likely for MBD2 judging by the number of mis-regulated genes I have observed in *Mbd2-/-* immune cells.

An important question given the focus of this work is the extent to which methylated CGIs tend to be bound by MBD proteins. Genes upregulated in the absence of MBD2 do not tend to associate with methylated TSS CGIs although a reasonable proportion contains intragenic CGIs. Although genes containing methylated TSS CGIs tend not to be upregulated in the absence of MBD2, this does not necessarily mean that MBD2 does not bind to these regions. It is possible that these genes are not mis-expressed when MBD2 is removed because methylated CGIs are inherently refractory to transcription, perhaps due to a failure to bind components of the

transcriptional machinery. MBD protein binding may simply add an additional layer of silencing to already silent methylated CGIs. Indeed, MBD2 has been reported to bind to the methylated *Xist* TSS CGI in male fibroblast cells but removal of MBD2 only leads to *Xist* mis-expression in ~1% of cells implying that MBD2 is just one part of the silencing mechanism (Barr et al., 2007).

5.2 Future work

5.2.1 Elucidating the function of intragenic CGI methylation

There are a number of experiments that could be done to further elucidate the role of cell type-specific intragenic CGI methylation in immune system cells. The first is to determine whether the intragenic CGIs with differential DNA methylation and H3K4me3 enrichment such as those located in *Kcnn4* and *Gata3* can act as alternative promoters for these genes and in what contexts these promoters are used. This could be determined by examining Pol II occupancy using ChIP (although initial attempts using an antibody against the hypophosphorylated form of Pol II were inconclusive) or by using 5' RACE or CAGE to map transcription start sites in the relevant immune cell types.

The second major task is to determine the role of the intragenic CGIs that are devoid of H3K4me3 and appear to be silenced in immune cells. Additional intragenic CGIs need to be examined for H3K27me3 association in the unmethylated context to determine whether the use of alternative silencing mechanisms is a common mechanism leading to cell-specific methylation. Since these CGIs don't appear to act as promoters in immune cells it would be interesting to know if they act as promoters in other tissues. Overlap with published CAGE and Pol II data already suggests that this is the case for many of these. One place to look for expression is the germline as here CGIs are hypomethylated and a number of germline-specific genes have TSS CGI methylation in somatic tissues (Eckhardt et al., 2006; Weber et al., 2007).

Alternatively, it is possible that some of these intragenic CGIs regulate genes with a role earlier in the immune cell lineage, for example in HSCs.

The remaining outstanding issue is to determine the reason for the association between intragenic CGI methylation and decreased gene expression. If the hypothesis that dense intragenic methylation impairs transcriptional elongation is correct, specifically disrupting methylation at these sites should lead to increased gene expression. Unfortunately targeted disruption of DNA methylation is not possible. A crude method of depleting DNA methylation is to treat cells with the chemical 5azaC which prevents the action of Dnmt1 (Creusot et al., 1982). This can have many secondary effects on the cell but has been used successfully to examine methylation-dependent regulation of alternative promoters (Maunakea et al., 2010). Addition of 5azaC to Th1 and Th2 cell cultures may allow the relationship between *Gata3* intragenic methylation and gene expression to be examined. One could also attempt to transgenically delete a methylated intragenic CGI (provided it does not disrupt the coding potential of the gene) in ES cells to see how this affects gene expression. However, one side effect of this may be to delete the promoter for a novel gene which could have secondary effects on the cell. Although immune system cells have been a good primary cell type in which to examine the role of cell-specific methylation a more tractable and manipulatable system, such as ES cell differentiation, would be more suitable for more mechanistic investigations of some of these questions.

5.2.2 Determining MBD2 binding sites

To confirm many of the aspects of MBD2 function suggested by this work, identification of MBD2 binding sites in immune cells is necessary. Ideally their methylation status would be assessed by a high resolution method such as bisulfite sequencing which could interrogate CpG-poor sites although a lot of information could be gleaned by comparing binding data to the methylation maps of immune cells generated here by MAP-seq. Certainly the relationship between CGI methylation and MBD2 binding could be definitively assessed through genome-wide

mapping of MBD2 binding sites and comparison with MAP-seq data. It is unclear how precisely to overcome some of the difficulties with MBD2 ChIP reported here. One way may be to tag the endogenous MBD2 protein so that a more efficient antibody against the tag could be used for immunoprecipitation. Alternatively, ChIP for other NuRD components could be used as a substitute for MBD2 ChIP, a comparison of wild-type and *Mbd2-/-* cells would then allow specific MBD2/NuRD binding sites to be identified.

5.3 Concluding remarks

This work gives an insight into the role DNA methylation and its mediators play in gene regulation in the immune system. Notably it has identified that intragenic CGIs are particularly susceptible to cell type-specific DNA methylation and further confirmed a role for MBD2 in regulating immune cell gene expression. Additional experiments will be necessary in order to fully elucidate the function of intragenic CGI regions, why they are susceptible to cell-specific methylation and if the effects of this methylation are mediated by methyl-binding proteins such as MBD2.

References

Agarwal, S., and Rao, A. (1998). Modulation of chromatin structure regulates cytokine gene expression during T cell differentiation. Immunity *9*, 765-775.

Agger, K., Christensen, J., Cloos, P.A., and Helin, K. (2008). The emerging functions of histone demethylases. Current opinion in genetics & development *18*, 159-168.

Agger, K., Cloos, P.A., Christensen, J., Pasini, D., Rose, S., Rappsilber, J., Issaeva, I., Canaani, E., Salcini, A.E., and Helin, K. (2007). UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. Nature *449*, 731-734.

Allan, A.M., Liang, X., Luo, Y., Pak, C., Li, X., Szulwach, K.E., Chen, D., Jin, P., and Zhao, X. (2008). The loss of methyl-CpG binding protein 1 leads to autism-like behavioral deficits. Hum Mol Genet *17*, 2047-2057.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet *23*, 185-188.

Antequera, F. (2003). Structure, function and evolution of CpG island promoters. Cell Mol Life Sci *60*, 1647-1658.

Auriol, E., Billard, L.M., Magdinier, F., and Dante, R. (2005). Specific binding of the methyl binding domain protein 2 at the BRCA1-NBR2 locus. Nucleic Acids Res *33*, 4243-4254.

Avni, O., Lee, D., Macian, F., Szabo, S.J., Glimcher, L.H., and Rao, A. (2002). T(H) cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. Nature immunology *3*, 643-651.

Baguet, A., and Bix, M. (2004). Chromatin landscape dynamics of the Il4-Il13 locus during T helper 1 and 2 development. Proceedings of the National Academy of Sciences of the United States of America *101*, 11410-11415.

Bakker, J., Lin, X., and Nelson, W.G. (2002). Methyl-CpG binding domain protein 2 represses transcription from hypermethylated pi-class glutathione S-transferase gene promoters in hepatocellular carcinoma cells. The Journal of biological chemistry *277*, 22573-22580.

Ball, M.P., Li, J.B., Gao, Y., Lee, J.H., LeProust, E.M., Park, I.H., Xie, B., Daley, G.Q., and Church, G.M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nature biotechnology *27*, 361-368.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Baron, U., Floess, S., Wieczorek, G., Baumann, K., Grutzkau, A., Dong, J., Thiel, A., Boeld, T.J., Hoffmann, P., Edinger, M.*, et al.* (2007). DNA demethylation in the human FOXP3 locus discriminates regulatory T cells from activated FOXP3(+) conventional T cells. European journal of immunology *37*, 2378-2389.

Barr, H., Hermann, A., Berger, J., Tsai, H.H., Adie, K., Prokhortchouk, A., Hendrich, B., and Bird, A. (2007). Mbd2 contributes to DNA methylation-directed repression of the Xist gene. Mol Cell Biol *27*, 3750-3757.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Barski, A., Jothi, R., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., and Zhao, K. (2009). Chromatin poises miRNA- and protein-coding genes for expression. Genome research *19*, 1742-1751.

Begenisich, T., Nakamoto, T., Ovitt, C.E., Nehrke, K., Brugnara, C., Alper, S.L., and Melvin, J.E. (2004). Physiological roles of the intermediate conductance, Ca2+-activated potassium channel Kcnn4. The Journal of biological chemistry *279*, 47681-47687.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482-485.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387-396.

Berger, J., Sansom, O., Clarke, A., and Bird, A. (2007). MBD2 is required for correct spatial gene expression in the gut. Mol Cell Biol *27*, 4049-4057.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R.*, et al.* (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169-181.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Bhattacharya, S.K., Ramchandani, S., Cervoni, N., and Szyf, M. (1999). A mammalian protein with specific demethylase activity for mCpG DNA. Nature *397*, 579-583.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2009). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & development *16*, 6-21.

Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.

Bird, A., and Tweedie, S. (1995). Transcriptional noise and the evolution of gene number. Philosophical transactions of the Royal Society of London *349*, 249-253.

Bird, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res *8*, 1499-1504.

Bird, A.P. (1995). Gene number, noise reduction and biological complexity. Trends Genet *11*, 94-100.

Bird, J.J., Brown, D.R., Mullen, A.C., Moskowitz, N.H., Mahowald, M.A., Sider, J.R., Gajewski, T.F., Wang, C.R., and Reiner, S.L. (1998). Helper T cell differentiation is controlled by the cell cycle. Immunity *9*, 229-237.

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res *30*, 958-965.

Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. Molecular cell *38*, 179-190.

Blanchard, C., Wang, N., Stringer, K.F., Mishra, A., Fulkerson, P.C., Abonia, J.P., Jameson, S.C., Kirby, C., Konikoff, M.R., Collins, M.H.*, et al.* (2006). Eotaxin-3 and a uniquely conserved gene-expression profile in eosinophilic esophagitis. The Journal of clinical investigation *116*, 536-547.

Blankenberg, D., Taylor, J., Schenck, I., He, J., Zhang, Y., Ghent, M., Veeraraghavan, N., Albert, I., Miller, W., Makova, K.D.*, et al.* (2007). A framework for collaborative analysis of ENCODE data: making large-scale analyses biologist-friendly. Genome research *17*, 960-964.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics (Oxford, England) *21*, 4067-4068.

Borchers, M., Ansay T, DeSalle R, Daugherty BL, Shen H, Metzger M, Lee NA, Lee JJ (2002). In vitro assessment of chemokine receptor-ligand interactions mediating mouse eosinophil migration. Journal of leukocyte biology *71*, 1033-1041.

Bowen, N.J., Fujita, N., Kajita, M., and Wade, P.A. (2004). Mi-2/NuRD: multiple complexes for many purposes. Biochimica et biophysica acta *1677*, 52-57.

Boyes, J., and Bird, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell *64*, 1123-1134.

Boyes, J., and Bird, A. (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. The EMBO journal *11*, 327-333.

Brackertz, M., Gong, Z., Leers, J., and Renkawitz, R. (2006). p66alpha and p66beta of the Mi-2/NuRD complex mediate MBD2 and histone interaction. Nucleic Acids Res *34*, 397-406.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature *371*, 435-438.

Brookes, E., and Pombo, A. (2009). Modifications of RNA polymerase II are pivotal in regulating gene expression states. EMBO reports *10*, 1213-1219.

Broske, A.M., Vockentanz, L., Kharazi, S., Huska, M.R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R.*, et al.* (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet *41*, 1207-1215.

Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nature immunology *4*, 235-240.

Campanero, M.R., Armstrong, M.I., and Flemington, E.K. (2000). CpG methylation as a mechanism for the regulation of E2F activity. Proceedings of the National Academy of Sciences of the United States of America *97*, 6481-6486.

Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Current opinion in genetics & development *14*, 155-164.

Carlone, D.L., Lee, J.H., Young, S.R., Dobrota, E., Butler, J.S., Ruiz, J., and Skalnik, D.G. (2005). Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. Mol Cell Biol *25*, 4881-4891.

Carlone, D.L., and Skalnik, D.G. (2001). CpG binding protein is crucial for early embryonic development. Mol Cell Biol *21*, 7601-7606.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C.*, et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet *38*, 626-635.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science *320*, 1224-1229.

Chang, S., Aune TM (2007). Dynamic changes in histone-methylation 'marks' across the locus encoding interferon-gamma during the differentiation of T helper type 2 cells. Nature immunology *8*, 723-731.

Chappell, C., Beard, C., Altman, J., Jaenisch, R., and Jacob, J. (2006). DNA methylation by DNA methyltransferase 1 is critical for effector CD8 T cell expansion. J Immunol *176*, 4562-4572.

Chatagnon, A., Ballestar, E., Esteller, M., and Dante, R. (2010). A role for methyl-CpG binding domain protein 2 in the modulation of the estrogen response of pS2/TFF1 gene. PLoS ONE *5*, e9665.

Chatagnon, A., Bougel, S., Perriaud, L., Lachuer, J., Benhattar, J., and Dante, R. (2009). Specific association between the methyl-CpG-binding domain protein 2 and the hypermethylated region of the human telomerase reverse transcriptase promoter in cancer cells. Carcinogenesis *30*, 28-34.

Chen, R.Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001). Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. Nat Genet *27*, 327-331.

Chen, W., Jin, W., Hardegen, N., Lei, K.J., Li, L., Marinos, N., McGrady, G., and Wahl, S.M. (2003). Conversion of peripheral CD4+CD25- naive T cells to

CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. J Exp Med *198*, 1875-1886.

Cheng, P., Nefedova, Y., Miele, L., Osborne, B.A., and Gabrilovich, D. (2003). Notch signaling is necessary but not sufficient for differentiation of dendritic cells. Blood *102*, 3980-3988.

Choi, Y.I., Duke-Cohan, J.S., Ahmed, W.B., Handley, M.A., Mann, F., Epstein, J.A., Clayton, L.K., and Reinherz, E.L. (2008). PlexinD1 glycoprotein controls migration of positively selected thymocytes into the medulla. Immunity *29*, 888-898.

Cisse, B., Caton, M.L., Lehner, M., Maeda, T., Scheu, S., Locksley, R., Holmberg, D., Zweier, C., den Hollander, N.S., Kant, S.G.*, et al.* (2008). Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. Cell *135*, 37-48.

Clouaire, T., de Las Heras, J.I., Merusi, C., and Stancheva, I. (2010). Recruitment of MBD1 to target genes requires sequence-specific interaction of the MBD domain with methylated DNA. Nucleic Acids Res.

Coolen, M.W., Stirzaker, C., Song, J.Z., Statham, A.L., Kassir, Z., Moreno, C.S., Young, A.N., Varma, V., Speed, T.P., Cowley, M.*, et al.* (2010). Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. Nat Cell Biol *12*, 235-246.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C.*, et al.* (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat Genet *24*, 132-138.

Creusot, F., Acs, G., and Christman, J.K. (1982). Inhibition of DNA methyltransferase and induction of Friend erythroleukemia cell differentiation by 5-azacytidine and 5-aza-2'-deoxycytidine. The Journal of biological chemistry *257*, 2041-2048.

Cross, S.H., Charlton, J.A., Nan, X., and Bird, A.P. (1994). Purification of CpG islands using a methylated DNA binding column. Nat Genet *6*, 236-244.

Cross, S.H., Meehan, R.R., Nan, X., and Bird, A. (1997). A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet *16*, 256-259.

Cruikshank, W., and Little, F. (2008). lnterleukin-16: the ins and outs of regulating T-cell activation. Critical reviews in immunology *28*, 467-483.

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. The Journal of cell biology *153*, 773-784.

Daniels, R., Kinis, T., Serhal, P., and Monk, M. (1995). Expression of the myotonin protein kinase gene in preimplantation human embryos. Hum Mol Genet *4*, 389-393.

Daniels, R., Lowell, S., Bolton, V., and Monk, M. (1997). Transcription of tissue-specific genes in human preimplantation embryos. Human reproduction (Oxford, England) *12*, 2251-2256.

de la Serna, I.L., Ohkawa, Y., and Imbalzano, A.N. (2006). Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers. Nature reviews *7*, 461-473.

De Smet, C., Lurquin, C., Lethe, B., Martelange, V., and Boon, T. (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. Mol Cell Biol *19*, 7327-7335.

Delgado, S., Gomez, M., Bird, A., and Antequera, F. (1998). Initiation of DNA replication at CpG islands in mammalian chromosomes. The EMBO journal *17*, 2426-2435.

Deng, J., Shoemaker, R., Xie, B., Gore, A., LeProust, E.M., Antosiewicz-Bourget, J., Egli, D., Maherali, N., Park, I.H., Yu, J.*, et al.* (2009). Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. Nature biotechnology *27*, 353-360.

Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. Genes & development *15*, 2940-2944.

Djuretic, I.M., Levanon, D., Negreanu, V., Groner, Y., Rao, A., and Ansel, K.M. (2007). Transcription factors T-bet and Runx3 cooperate to activate Ifng and silence Il4 in T helper type 1 cells. Nature immunology *8*, 145-153.

Down, T.A., Rakyan, V.K., Turner, D.J., Flicek, P., Li, H., Kulesha, E., Graf, S., Johnson, N., Herrero, J., Tomazou, E.M.*, et al.* (2008). A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. Nature biotechnology *26*, 779-785.

Dunning, M.J., Barbosa-Morais, N.L., Lynch, A.G., Tavare, S., and Ritchie, M.E. (2008). Statistical issues in the analysis of Illumina data. BMC bioinformatics *9*, 85.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A.*, et al.* (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet *38*, 1378-1385.

Edwards, C.A., and Ferguson-Smith, A.C. (2007). Mechanisms regulating imprinted genes in clusters. Current opinion in cell biology *19*, 281-289.

Efroni, S., Duttagupta, R., Cheng, J., Dehghani, H., Hoeppner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D., Buetow, K.H.*, et al.* (2008). Global transcription in pluripotent embryonic stem cells. Cell stem cell *2*, 437-447.

Egwuagu, C.E., Yu, C.R., Zhang, M., Mahdi, R.M., Kim, S.J., and Gery, I. (2002). Suppressors of cytokine signaling proteins are differentially expressed in Th1 and Th2 cells: implications for Th cell lineage commitment and maintenance. J Immunol *168*, 3181-3187.

Ehrlich, M. (2003). The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clinical immunology (Orlando, Fla *109*, 17-28.

Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. Nucleic Acids Res *10*, 2709-2721.

El-Asady, R., Yuan, R., Liu, K., Wang, D., Gress, R.E., Lucas, P.J., Drachenberg, C.B., and Hadley, G.A. (2005). TGF-{beta}-dependent CD103 expression by CD8(+) T cells promotes selective destruction of the host intestinal epithelium during graft-versus-host disease. J Exp Med *201*, 1647-1657.

Elgert, K.D. (2009). Immunology: Understanding the immune system. John Wiley and Sons *2nd edition*.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H.*, et al.* (2008). De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nature structural & molecular biology *15*, 1176-1183.

Esteve, P.O., Chin, H.G., Smallwood, A., Feehery, G.R., Gangisetty, O., Karpf, A.R., Carey, M.F., and Pradhan, S. (2006). Direct interaction between DNMT1 and G9a coordinates DNA and histone methylation during replication. Genes & development *20*, 3089-3103.

Fantini, M.C., Becker, C., Monteleone, G., Pallone, F., Galle, P.R., and Neurath, M.F. (2004). Cutting edge: TGF-beta induces a regulatory phenotype in CD4+CD25- T cells through Foxp3 induction and down-regulation of Smad7. J Immunol *172*, 5149-5153.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T.*, et al.* (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat Genet *41*, 563-571.

Fazzari, M.J., and Greally, J.M. (2004). Epigenomics: beyond CpG islands. Nature reviews *5*, 446-455.

Felsenfeld, G., and Groudine, M. (2003). Controlling the double helix. Nature *421*, 448-453.

Feng, Q., and Zhang, Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes & development *15*, 827-832.

Fields, P.E., Kim, S.T., and Flavell, R.A. (2002). Cutting edge: changes in histone acetylation at the IL-4 and IFN-gamma loci accompany Th1/Th2 differentiation. J Immunol *169*, 647-650.

Fields, P.E., Lee, G.R., Kim, S.T., Bartsevich, V.V., and Flavell, R.A. (2004). Th2-specific chromatin remodeling and enhancer activity in the Th2 cytokine locus control region. Immunity *21*, 865-876.

Filion, G.J., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P.A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. Mol Cell Biol *26*, 169-181.

Finney, C.A., Taylor, M.D., Wilson, M.S., and Maizels, R.M. (2007). Expansion and activation of CD4(+)CD25(+) regulatory T cells in Heligmosomoides polygyrus infection. European journal of immunology *37*, 1874-1886.

Fiorini, E., Merck, E., Wilson, A., Ferrero, I., Jiang, W., Koch, U., Auderset, F., Laurenti, E., Tacchini-Cottier, F., Pierres, M.*, et al.* (2009). Dynamic regulation of notch 1 and notch 2 surface expression during T cell development and activation revealed by novel monoclonal antibodies. J Immunol *183*, 7212-7222.

Floess, S., Freyer, J., Siewert, C., Baron, U., Olek, S., Polansky, J., Schlawe, K., Chang, H.D., Bopp, T., Schmitt, E.*, et al.* (2007). Epigenetic control of the foxp3 locus in regulatory T cells. PLoS biology *5*, e38.

Fouse, S.D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. (2008). Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell stem cell *2*, 160-169.

Fraga, M.F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P.A., and Esteller, M. (2003). The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res *31*, 1765-1774.

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences of the United States of America *89*, 1827-1831.

Fujita, N., Takebayashi, S., Okumura, K., Kudo, S., Chiba, T., Saya, H., and Nakao, M. (1999). Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. Mol Cell Biol *19*, 6415-6426.

Fujita, N., and Wade, P.A. (2004). Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. Methods (San Diego, Calif *33*, 81-85.

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., and Nakao, M. (2003). Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. The Journal of biological chemistry *278*, 24132-24138.

Fuks, F., Burgers, W.A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000). DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet *24*, 88-91.

Fuks, F., Hurd, P.J., Deplus, R., and Kouzarides, T. (2003). The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucleic Acids Res *31*, 2305-2312.

Gamper, C.J., Agoston, A.T., Nelson, W.G., and Powell, J.D. (2009). Identification of DNA methyltransferase 3a as a T cell receptor-induced regulator of Th1 and Th2 differentiation. J Immunol *183*, 2267-2276.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. Journal of molecular biology *196*, 261-282.

Gardiner-Garden, M., and Frommer, M. (1994). Transcripts and CpG islands associated with the pro-opiomelanocortin gene and other neurally expressed genes. Journal of molecular endocrinology *12*, 365-382.

Gaston, K., and Fried, M. (1995). CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes. Nucleic Acids Res *23*, 901-909.

Gebhard, C., Schwarzfischer, L., Pham, T.H., Schilling, E., Klug, M., Andreesen, R., and Rehli, M. (2006). Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer research *66*, 6118-6128.

Geiman, T.M., Sankpal, U.T., Robertson, A.K., Zhao, Y., Zhao, Y., and Robertson, K.D. (2004). DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. Biochemical and biophysical research communications *318*, 544-555.

Goll, M.G., Kirpekar, F., Maggert, K.A., Yoder, J.A., Hsieh, C.L., Zhang, X., Golic, K.G., Jacobsen, S.E., and Bestor, T.H. (2006). Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science *311*, 395-398.

Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldan-Arjona, T., David, L., and Zhu, J.K. (2002). ROS1, a repressor of transcriptional gene silencing in Arabidopsis, encodes a DNA glycosylase/lyase. Cell *111*, 803-814.

Gonzalo, S., Jaco, I., Fraga, M.F., Chen, T., Li, E., Esteller, M., and Blasco, M.A. (2006). DNA methyltransferases control telomere length and telomere recombination in mammalian cells. Nat Cell Biol *8*, 416-424.

Gowher, H., and Jeltsch, A. (2002). Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. The Journal of biological chemistry *277*, 20409-20414.

Gowher, H., Liebert, K., Hermann, A., Xu, G., and Jeltsch, A. (2005). Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. The Journal of biological chemistry *280*, 13341-13348.

Grogan, J.L., Wang, Z.E., Stanley, S., Harmon, B., Loots, G.G., Rubin, E.M., and Locksley, R.M. (2003). Basal chromatin modification at the IL-4 gene in helper T cells. J Immunol *171*, 6672-6679.

Gu, P., Le Menuet, D., Chung, A.C., and Cooney, A.J. (2006). Differential recruitment of methylated CpG binding domains by the orphan receptor GCNF initiates the repression and silencing of Oct4 expression. Mol Cell Biol *26*, 9471-9483.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. Cell *130*, 77-88.

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science *315*, 1143-1147.

Guy, J., Hendrich, B., Holmes, M., Martin, J.E., and Bird, A. (2001). A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet *27*, 322-326.

Haberland, M., Montgomery, R.L., and Olson, E.N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nature reviews *10*, 32-42.

Hamalainen, H.K., Tubman, J.C., Vikman, S., Kyrola, T., Ylikoski, E., Warrington, J.A., and Lahesmaa, R. (2001). Identification and validation of endogenous reference genes for expression profiling of T helper cell differentiation by quantitative real-time RT-PCR. Analytical biochemistry *299*, 63-70.

Happel, N., and Doenecke, D. (2009). Histone H1 and its isoforms: contribution to chromatin structure and function. Gene *431*, 1-12.

Hardtke, S., Ohl, L., and Forster, R. (2005). Balanced expression of CXCR5 and CCR7 on follicular T helper cells determines their transient positioning to lymph node follicles and is essential for efficient B-cell help. Blood *106*, 1924-1931.

Harrington, M.A., Jones, P.A., Imagawa, M., and Karin, M. (1988). Cytosine methylation does not affect binding of transcription factor Sp1. Proceedings of the National Academy of Sciences of the United States of America *85*, 2066-2070.

Hegazy, A.N., Peine, M., Helmstetter, C., Panse, I., Frohlich, A., Bergthaler, A., Flatz, L., Pinschewer, D.D., Radbruch, A., and Lohning, M. (2010). Interferons direct Th2 cell reprogramming to generate a stable GATA-3(+)T-bet(+) cell subset with combined Th2 and Th1 cell functions. Immunity *32*, 116-128.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A.*, et al.* (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311-318.

Hellman, A., and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. Science *315*, 1141-1143.

Hendrich, B., Abbott, C., McQueen, H., Chambers, D., Cross, S., and Bird, A. (1999a). Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. Mamm Genome *10*, 906-912.

Hendrich, B., and Bird, A. (1998). Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol *18*, 6538-6547.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V.A., and Bird, A. (2001). Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes & development *15*, 710-723.

Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999b). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature *401*, 301-304.

Hermann, A., Gowher, H., and Jeltsch, A. (2004). Biochemistry and biology of mammalian DNA methyltransferases. Cell Mol Life Sci *61*, 2571-2587.

Ho, K.L., McNae, I.W., Schmiedeberg, L., Klose, R.J., Bird, A.P., and Walkinshaw, M.D. (2008). MeCP2 binding to DNA depends upon hydration at methyl-CpG. Molecular cell *29*, 525-531.

Ho, L., and Crabtree, G.R. (2010). Chromatin remodelling during development. Nature *463*, 474-484.

Hodgkin, J. (1994). Epigenetics and the maintenance of gene activity states in Caenorhabditis elegans. Developmental genetics *15*, 471-477.

Hong, L., Schroth, G.P., Matthews, H.R., Yau, P., and Bradbury, E.M. (1993). Studies of the DNA binding properties of histone H4 amino terminus. Thermal denaturation studies reveal that acetylation markedly reduces the binding constant of the H4 "tail" to DNA. The Journal of biological chemistry *268*, 305-314.

Hughes, T., Webb, R., Fei, Y., Wren, J.D., and Sawalha, A.H. (2010). DNA methylome in human CD4+ T cells identifies transcriptionally repressive and non-repressive methylation peaks. Genes and immunity.

Hutchins, A.S., Artis, D., Hendrich, B.D., Bird, A.P., Scott, P., and Reiner, S.L. (2005). Cutting edge: a critical role for gene silencing in preventing excessive type 1 immunity. J Immunol *175*, 5606-5610.

Hutchins, A.S., Mullen, A.C., Lee, H.W., Sykes, K.J., High, F.A., Hendrich, B.D., Bird, A.P., and Reiner, S.L. (2002). Gene silencing quantitatively controls the function of a developmental trans-activator. Molecular cell *10*, 81-91.

Iguchi-Ariga, S.M., and Schaffner, W. (1989). CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes & development *3*, 612-619.

Illingworth, R., Gruenewald-Schneider, U., Webb, S., Kerr, A., James, K., Turner, D., Smith, C., Harrison, D., Andrews, R., and Bird, A. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS genetics *Accepted*.

Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J.*, et al.* (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS biology *6*, e22.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS letters *583*, 1713-1720.

Irizarry, R.A., Ladd-Acosta, C., Carvalho, B., Wu, H., Brandenburg, S.A., Jeddeloh, J.A., Wen, B., and Feinberg, A.P. (2008). Comprehensive high-throughput arrays for relative methylation (CHARM). Genome research *18*, 780-790.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M.*, et al.* (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet *41*, 178-186.

Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature.

Ivascu, C., Wasserkort, R., Lesche, R., Dong, J., Stein, H., Thiel, A., and Eckhardt, F. (2007). DNA methylation profiling of transcription factor genes in normal lymphocyte development and lymphomas. The international journal of biochemistry & cell biology *39*, 1523-1538.

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

Jones, B., and Chen, J. (2006). Inhibition of IFN-gamma transcription by site-specific methylation during T helper cell development. The EMBO journal *25*, 2443-2452.

Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. Nature reviews *3*, 415-428.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683-692.

Jones, P.A., Wolkowicz, M.J., Harrington, M.A., and Gonzales, F. (1990). Methylation and expression of the Myo D1 determination gene. Philosophical transactions of the Royal Society of London *326*, 277-284.

Jorgensen, H.F., Ben-Porath, I., and Bird, A.P. (2004). Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol Cell Biol *24*, 3387-3395.

Josefowicz, S.Z., Wilson, C.B., and Rudensky, A.Y. (2009). Cutting edge: TCR stimulation is sufficient for induction of Foxp3 expression in the absence of DNA methyltransferase 1. J Immunol *182*, 6648-6652.

Junt, T., Fink, K., Forster, R., Senn, B., Lipp, M., Muramatsu, M., Zinkernagel, R.M., Ludewig, B., and Hengartner, H. (2005). CXCR5-dependent seeding of follicular niches by B and Th cells augments antiviral B cell responses. J Immunol *175*, 7109-7116.

Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A., and Hendrich, B. (2006). The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol *8*, 285-292.

Kaji, K., Nichols, J., and Hendrich, B. (2007). Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. Development (Cambridge, England) *134*, 1123-1132.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900-903.

Kersh, E.N. (2006). Impaired memory CD8 T cell development in the absence of methyl-CpG-binding domain protein 2. J Immunol *177*, 3821-3826.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H.*, et al.* (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat Genet *38*, 149-153.

Khulan, B., Thompson, R.F., Ye, K., Fazzari, M.J., Suzuki, M., Stasiek, E., Figueroa, M.E., Glass, J.L., Chen, Q., Montagna, C.*, et al.* (2006). Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome research *16*, 1046-1055.

Kim, H.P., and Leonard, W.J. (2007). CREB/ATF-dependent T cell receptor-induced FoxP3 gene expression: a role for DNA methylation. J Exp Med *204*, 1543-1551.

Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO reports *7*, 397-403.

Kim, S.T., Fields, P.E., and Flavell, R.A. (2007). Demethylation of a specific hypersensitive site in the Th2 locus control region. Proceedings of the National Academy of Sciences of the United States of America *104*, 17052-17057.

Klose, R.J., and Bird, A.P. (2004). MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. The Journal of biological chemistry *279*, 46490-46496.

Klose, R.J., Sarraf, S.A., Schmiedeberg, L., McDermott, S.M., Stancheva, I., and Bird, A.P. (2005). DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Molecular cell *19*, 667-678.

Klose, R.J., and Zhang, Y. (2007). Regulation of histone methylation by demethylimination and demethylation. Nat Rev Mol Cell Biol *8*, 307-318.

Kondo, Y., Shen, L., Cheng, A.S., Ahmed, S., Boumber, Y., Charo, C., Yamochi, T., Urano, T., Furukawa, K., Kwabi-Addo, B.*, et al.* (2008). Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. Nat Genet *40*, 741-750.

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

Kubo, M., Hanada, T., and Yoshimura, A. (2003). Suppressors of cytokine signaling and immunity. Nature immunology *4*, 1169-1176.

Kumanogoh, A., Shikina, T., Suzuki, K., Uematsu, S., Yukawa, K., Kashiwamura, S., Tsutsui, H., Yamamoto, M., Takamatsu, H., Ko-Mitamura, E.P.*, et al.* (2005). Nonredundant roles of Sema4A in the immune system: defective T cell priming and Th1/Th2 regulation in Sema4A-deficient mice. Immunity *22*, 305-316.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.

Lal, G., Zhang, N., van der Touw, W., Ding, Y., Ju, W., Bottinger, E.P., Reid, S.P., Levy, D.E., and Bromberg, J.S. (2009). Epigenetic regulation of Foxp3 expression in regulatory T cells by DNA methylation. J Immunol *182*, 259-273.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W.*, et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics *13*, 1095-1107.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J.*, et al.* (2010). Dynamic changes in the human methylome during differentiation. Genome research *20*, 320-331.

Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W.A., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. Mol Cell Biol *26*, 843-851.

Lee, D., Agarwal S, Rao A (2002). Th2 lineage commitment and efficient IL-4 production involves extended demethylation of the IL-4 gene. Immunity *16*, 649-660.

Lee, D.Y., Hayes, J.J., Pruss, D., and Wolffe, A.P. (1993). A positive role for histone acetylation in transcription factor access to nucleosomal DNA. Cell *72*, 73-84.

Lee, J.H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. The Journal of biological chemistry *280*, 41725-41731.

Lee, P.P., Fitzpatrick, D.R., Beard, C., Jessup, H.K., Lehar, S., Makar, K.W., Perez-Melgosa, M., Sweetser, M.T., Schlissel, M.S., Nguyen, S.*, et al.* (2001). A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. Immunity *15*, 763-774.

Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol *13*, 1192-1200.

Lembo, F., Pero, R., Angrisano, T., Vitiello, C., Iuliano, R., Bruni, C.B., and Chiariotti, L. (2003). MBDin, a novel MBD2-interacting protein, relieves MBD2 repression potential and reactivates transcription from methylated promoters. Mol Cell Biol *23*, 1656-1665.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905-914.

Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.

Li, B., Jackson, J., Simon, M.D., Fleharty, B., Gogol, M., Seidel, C., Workman, J.L., and Shilatifard, A. (2009). Histone H3 lysine 36 dimethylation (H3K36me2) is sufficient to recruit the Rpd3s histone deacetylase complex and to repress spurious transcription. The Journal of biological chemistry *284*, 7970-7976.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362-365.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature *442*, 91-95.

Lin, X., and Nelson, W.G. (2003). Methyl-CpG-binding domain protein-2 mediates transcriptional repression associated with hypermethylated GSTP1 CpG islands in MCF-7 breast cancer cells. Cancer research *63*, 498-504.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M.*, et al.* (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature.

Liu, W.H., and Lai, M.Z. (2005). Deltex regulates T-cell activation by targeted degradation of active MEKK1. Mol Cell Biol *25*, 1367-1378.

Liu, W.M., Maraia, R.J., Rubin, C.M., and Schmid, C.W. (1994). Alu transcripts: cytoplasmic localisation and regulation by DNA methylation. Nucleic Acids Res *22*, 1087-1095.

Lorincz, M.C., Dickerson, D.R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. Nature structural & molecular biology *11*, 1068-1075.

Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature *389*, 251-260.

Luo, S.W., Zhang, C., Zhang, B., Kim, C.H., Qiu, Y.Z., Du, Q.S., Mei, L., and Xiong, W.C. (2009). Regulation of heterochromatin remodelling and myogenin expression during muscle differentiation by FAK interaction with MBD2. The EMBO journal.

Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in Drosophila melanogaster. Nature *408*, 538-540.

Macleod, D., Ali, R.R., and Bird, A. (1998). An alternative promoter in the mouse major histocompatibility complex class II I-Abeta gene: implications for the origin of CpG islands. Mol Cell Biol *18*, 4433-4443.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes & development *8*, 2282-2292.

Makar, K.W., Perez-Melgosa, M., Shnyreva, M., Weaver, W.M., Fitzpatrick, D.R., and Wilson, C.B. (2003). Active recruitment of DNA methyltransferases regulates interleukin 4 in thymocytes and T cells. Nature immunology *4*, 1183-1190.

Makar, K.W., and Wilson, C.B. (2004). DNA methylation is a nonredundant repressor of the Th2 effector program. J Immunol *173*, 4402-4406.

Martin, V., Jorgensen, H.F., Chaubert, A.S., Berger, J., Barr, H., Shaw, P., Bird, A., and Chaubert, P. (2008). MBD2-mediated transcriptional repression of the p14ARF tumor suppressor gene in human colon cancer cells. Pathobiology *75*, 281-287.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y.*, et al.* (2010). Conserved role of

intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253-257.

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

Meehan, R.R., Lewis, J.D., and Bird, A.P. (1992). Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res *20*, 5085-5092.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989). Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell *58*, 499-507.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B.*, et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Miccio, A., Wang, Y., Hong, W., Gregory, G.D., Wang, H., Yu, X., Choi, J.K., Shelat, S., Tong, W., Poncz, M.*, et al.* (2010). NuRD mediates activating and repressive functions of GATA-1 and FOG-1 during blood development. The EMBO journal *29*, 442-456.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science *297*, 403-405.

Miller, T., Krogan, N.J., Dover, J., Erdjument-Bromage, H., Tempst, P., Johnston, M., Greenblatt, J.F., and Shilatifard, A. (2001). COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proceedings of the National Academy of Sciences of the United States of America *98*, 12902-12907.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schubeler, D. (2008). Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Molecular cell *30*, 755-766.

Mookherjee, N., Brown, K.L., Bowdish, D.M., Doria, S., Falsafi, R., Hokamp, K., Roche, F.M., Mu, R., Doho, G.H., Pistolic, J.*, et al.* (2006). Modulation of the TLR-mediated inflammatory response by the endogenous human host defense peptide LL-37. J Immunol *176*, 2455-2464.

Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. (2005). Epigenetic reprogramming in mammals. Hum Mol Genet *14 Spec No 1*, R47-58.

Morinobu, A., Kanno, Y., and O'Shea, J.J. (2004). Discrete roles for histone acetylation in human T helper 1 cell-specific gene expression. The Journal of biological chemistry *279*, 40640-40646.

Murayama, A., Sakura, K., Nakama, M., Yasuzawa-Tanaka, K., Fujita, E., Tateishi, Y., Wang, Y., Ushijima, T., Baba, T., Shibuya, K.*, et al.* (2006). A specific CpG site demethylation in the human interleukin 2 gene promoter is an epigenetic memory. The EMBO journal *25*, 1081-1092.

Myant, K., and Stancheva, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. Mol Cell Biol *28*, 215-226.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717-1720.

Nagar, M., Vernitsky, H., Cohen, Y., Dominissini, D., Berkun, Y., Rechavi, G., Amariglio, N., and Goldstein, I. (2008). Epigenetic inheritance of DNA methylation limits activation-induced expression of FOXP3 in conventional human CD25-CD4+ T cells. International immunology.

Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471-481.

Nan, X., Meehan, R.R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res *21*, 4886-4892.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386-389.

Nawijn, M.C., Dingjan, G.M., Ferreira, R., Lambrecht, B.N., Karis, A., Grosveld, F., Savelkoul, H., and Hendriks, R.W. (2001). Enforced expression of GATA-3 in transgenic mice inhibits Th1 differentiation and induces the formation of a T1/ST2-expressing Th2-committed T cell compartment in vivo. J Immunol *167*, 724-732.

Ng, H.H., Jeppesen, P., and Bird, A. (2000). Active repression of methylated genes by the chromosomal protein MBD1. Mol Cell Biol *20*, 1394-1406.

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999). MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. Nat Genet *23*, 58-61.

Nicol, J.W., Helt, G.A., Blanchard, S.G., Jr., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics (Oxford, England) *25*, 2730-2731.

Nuber, U.A., Kriaucionis, S., Roloff, T.C., Guy, J., Selfridge, J., Steinhoff, C., Schulz, R., Lipkowitz, B., Ropers, H.H., Holmes, M.C.*, et al.* (2005). Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. Hum Mol Genet *14*, 2247-2256.

Oda, M., Yamagiwa, A., Yamamoto, S., Nakayama, T., Tsumura, A., Sasaki, H., Nakao, K., Li, E., and Okano, M. (2006). DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. Genes & development *20*, 3382-3394.

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., *et al.* (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet *39*, 237-242.

Okamoto, I., and Heard, E. (2009). Lessons from comparative analysis of X-chromosome inactivation in mammals. Chromosome Res *17*, 659-669.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., *et al.* (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714-717.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol *10*, 475-478.

Ozsolak, F., Song, J.S., Liu, X.S., and Fisher, D.E. (2007). High-throughput mapping of the chromatin structure of human promoters. Nature biotechnology *25*, 244-248.

Payer, B., and Lee, J.T. (2008). X chromosome dosage compensation: how mammals keep the balance. Annual review of genetics *42*, 733-772.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S., and Fischer, R.L. (2007). DNA demethylation in the Arabidopsis genome. Proceedings of the National Academy of Sciences of the United States of America *104*, 6752-6757.

Phesse, T.J., Parry, L., Reed, K.R., Ewan, K.B., Dale, T.C., Sansom, O.J., and Clarke, A.R. (2008). Deficiency of Mbd2 Attenuates Wnt signalling. Mol Cell Biol.

Pietersen, A.M., and van Lohuizen, M. (2008). Stem cell regulation by polycomb repressors: postponing commitment. Current opinion in cell biology *20*, 201-207.

Pils, M.C., Pisano, F., Fasnacht, N., Heinrich, J.M., Groebe, L., Schippers, A., Rozell, B., Jack, R.S., and Muller, W. (2010). Monocytes/macrophages and/or neutrophils are the target of IL-10 in the LPS endotoxemia model. European journal of immunology *40*, 443-448.

Polansky, J.K., Kretschmer, K., Freyer, J., Floess, S., Garbe, A., Baron, U., Olek, S., Hamann, A., von Boehmer, H., and Huehn, J. (2008). DNA methylation controls Foxp3 gene expression. European journal of immunology *38*, 1654-1663.

Ponger, L., Duret, L., and Mouchiroud, D. (2001). Determinants of CpG islands: expression in early embryo and isochore structure. Genome research *11*, 1854-1860.

Pradhan, M., Esteve, P.O., Chin, H.G., Samaranayke, M., Kim, G.D., and Pradhan, S. (2008). CXXC domain of human DNMT1 is essential for enzymatic activity. Biochemistry *47*, 10000-10009.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes & development *15*, 1613-1618.

Pulukuri, S.M., and Rao, J.S. (2006). CpG island promoter methylation and silencing of 14-3-3sigma gene expression in LNCaP and Tramp-C1 prostate cancer cell lines is associated with methyl-CpG-binding protein MBD2. Oncogene *25*, 4559-4572.

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell *135*, 1201-1212.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114-128.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proceedings of the National Academy of Sciences of the United States of America *97*, 5237-5242.

Rauch, T.A., Wu, X., Zhong, X., Riggs, A.D., and Pfeifer, G.P. (2009). A human B cell methylome at 100-base pair resolution. Proceedings of the National Academy of Sciences of the United States of America *106*, 671-678.

Reiner, S.L. (2007). Development in motion: helper T cells at work. Cell *129*, 33-36.

Reiner, S.L., and Locksley, R.M. (1995). The regulation of immunity to Leishmania major. Annu Rev Immunol *13*, 151-177.

Reynolds, P.A., Sigaroudinia, M., Zardo, G., Wilson, M.B., Benton, G.M., Miller, C.J., Hong, C., Fridlyand, J., Costello, J.F., and Tlsty, T.D. (2006). Tumor suppressor p16INK4A regulates polycomb-mediated DNA hypermethylation in human mammary epithelial cells. The Journal of biological chemistry *281*, 24790-24802.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E.*, et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311-1323.

Robertson, A.K., Geiman, T.M., Sankpal, U.T., Hager, G.L., and Robertson, K.D. (2004). Effects of chromatin structure on the enzymatic and DNA binding functions of DNA methyltransferases DNMT1 and Dnmt3a in vitro. Biochemical and biophysical research communications *322*, 110-118.

Robertson, K.D., Ait-Si-Ali, S., Yokochi, T., Wade, P.A., Jones, P.L., and Wolffe, A.P. (2000). DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet *25*, 338-342.

Robinson, P.W., Green, S.J., Carter, C., Coadwell, J., and Kilshaw, P.J. (2001). Studies on transcriptional regulation of the mucosal T-cell integrin alphaEbeta7 (CD103). Immunology *103*, 146-154.

Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proceedings of the National Academy of Sciences of the United States of America *103*, 15782-15787.

Roh, T.Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes & development *19*, 542-552.

Romagnani, S. (2006). Regulation of the T cell response. Clin Exp Allergy *36*, 1357-1366.

Rountree, M.R., Bachman, K.E., and Baylin, S.B. (2000). DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nat Genet *25*, 269-277.

Ruike, Y., Imanaka, Y., Sato, F., Shimizu, K., and Tsujimoto, G. (2010). Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. BMC genomics *11*, 137.

Russo, V.E.A., Martienssen, R.A., and Riggs, A.D. (1996). Epigenetics mechanisms of gene regulation. Cold Spring Harbour Laboratory Press.

Ruthenburg, A.J., Allis, C.D., and Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Molecular cell *25*, 15-30.

Sado, T., Fenner, M.H., Tan, S.S., Tam, P., Shioda, T., and Li, E. (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Developmental biology *225*, 294-303.

Saito, M., and Ishikawa, F. (2002). The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. The Journal of biological chemistry *277*, 35434-35439.

Sambrook, J., and Russell, D., W. (2001). Molecular Cloning: A laboratory manual. Cold Spring Harbour Laboratory Press *3rd edition*.

Sansom, O.J., Berger, J., Bishop, S.M., Hendrich, B., Bird, A., and Clarke, A.R. (2003). Deficiency of Mbd2 suppresses intestinal tumorigenesis. Nat Genet *34*, 145-147.

Santangelo, S., Cousins, D.J., Winkelmann, N.E., and Staynov, D.Z. (2002). DNA methylation changes at human Th2 cytokine genes coincide with DNase I hypersensitive site formation during CD4(+) T cell differentiation. J Immunol *169*, 1893-1903.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Developmental biology *241*, 172-182.

Sarraf, S.A., and Stancheva, I. (2004). Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Molecular cell *15*, 595-605.

Schilling, E., and Rehli, M. (2007). Global, comparative analysis of tissue-specific promoter CpG methylation. Genomics *90*, 314-323.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., *et al.* (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet *39*, 232-236.

Schmidl, C., Klug, M., Boeld, T.J., Andreesen, R., Hoffmann, P., Edinger, M., and Rehli, M. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome research *19*, 1165-1174.

Schmiedeberg, L., Skene, P., Deaton, A., and Bird, A. (2009). A temporal threshold for formaldehyde crosslinking and fixation. PLoS ONE *4*, e4636.

Schoenborn, J.R., Dorschner, M.O., Sekimata, M., Santer, D.M., Shnyreva, M., Fitzpatrick, D.R., Stamatoyannopoulos, J.A., and Wilson, C.B. (2007). Comprehensive epigenetic profiling identifies multiple distal regulatory elements directing transcription of the gene encoding interferon-gamma. Nature immunology *8*, 732-742.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735-745.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nature reviews *8*, 9-22.

Sekimata, M., Takahashi, A., Murakami-Sekimata, A., and Homma, Y. (2001). Involvement of a novel zinc finger protein, MIZF, in transcriptional repression by interacting with a methyl-CpG-binding protein, MBD2. The Journal of biological chemistry *276*, 42632-42638.

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS genetics *5*, e1000446.

Serre, D., Lee, B.H., and Ting, A.H. (2009). MBD-isolated Genome Sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. Nucleic Acids Res.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A., and Issa, J.P. (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS genetics *3*, 2023-2036.

Siegfried, Z., Eden, S., Mendelsohn, M., Feng, X., Tsuberi, B.Z., and Cedar, H. (1999). DNA methylation represses transcription in vivo. Nat Genet *22*, 203-206.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol *10*, 697-708.

Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Molecular cell *37*, 457-468.

Sleutels, F., Zwart, R., and Barlow, D.P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature *415*, 810-813.

Smiraglia, D.J., Rush, L.J., Fruhwald, M.C., Dai, Z., Held, W.A., Costello, J.F., Lang, J.C., Eng, C., Li, B., Wright, F.A., *et al.* (2001). Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. Hum Mol Genet *10*, 1413-1419.

Smith, A.E., Chronis, C., Christodoulakis, M., Orr, S.J., Lea, N.C., Twine, N.A., Bhinge, A., Mufti, G.J., and Thomas, N.S. (2009). Epigenetics of human T cells during the G0-->G1 transition. Genome research *19*, 1325-1337.

Song, F., Smith, J.F., Kimura, M.T., Morrow, A.D., Matsuyama, T., Nagase, H., and Held, W.A. (2005). Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. Proceedings of the National Academy of Sciences of the United States of America *102*, 3336-3341.

Sridharan, R., and Smale, S.T. (2007). Predominant interaction of both Ikaros and Helios with the NuRD complex in immature thymocytes. The Journal of biological chemistry *282*, 30227-30238.

Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proceedings of the National Academy of Sciences of the United States of America *79*, 3418-3422.

Steinert, S., Shay, J.W., and Wright, W.E. (2004). Modification of subtelomeric DNA. Mol Cell Biol *24*, 4571-4580.

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

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., *et al.* (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol Cell Biol *22*, 1298-1306.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. The Journal of biological chemistry *279*, 27816-27823.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nature reviews *9*, 465-476.

Suzuki, M.M., Kerr, A.R., De Sousa, D., and Bird, A. (2007). CpG methylation is targeted to transcription units in an invertebrate genome. Genome research *17*, 625-631.

Swain, S.L., Bradley, L.M., Croft, M., Tonkonogy, S., Atkins, G., Weinberg, A.D., Duncan, D.D., Hedrick, S.M., Dutton, R.W., and Huston, G. (1991). Helper T-cell subsets: phenotype, function and the role of lymphokines in regulating their development. Immunological reviews *123*, 115-144.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., *et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proceedings of the National Academy of Sciences of the United States of America *99*, 3740-3745.

Takemoto, N., Kamogawa, Y., Jun Lee, H., Kurata, H., Arai, K.I., O'Garra, A., Arai, N., and Miyatake, S. (2000). Cutting edge: chromatin remodeling at the IL-4/IL-13 intergenic regulatory region for Th2-specific cytokine gene cluster. J Immunol *165*, 6687-6691.

Tan, C.P., and Nakielny, S. (2006). Control of the DNA methylation system component MBD2 by protein arginine methylation. Mol Cell Biol *26*, 7224-7235.

Tang, Q., and Bluestone, J.A. (2008). The Foxp3+ regulatory T cell: a jack of all trades, master of regulation. Nature immunology *9*, 239-244.

Taylor, J., Schenck, I., Blankenberg, D., and Nekrutenko, A. (2007). Using galaxy to perform large-scale interactive data analyses. Current protocols in bioinformatics *Chapter 10*, Unit 10 15.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell *60*, 909-920.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R., Deaton, A., Andrews, R., James, K.D.*, et al.* (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082-1086.

Toy, D., Kugler, D., Wolfson, M., Vanden Bos, T., Gurgel, J., Derry, J., Tocker, J., and Peschon, J. (2006). Cutting edge: interleukin 17 signals through a heteromeric receptor complex. J Immunol *177*, 36-39.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999). CpG island methylator phenotype in colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America *96*, 8681-8686.

Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. Cell *128*, 651-654.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M.E., Borchers, C.H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of JmjC domain-containing proteins. Nature *439*, 811-816.

Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002). Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. Proceedings of the National Academy of Sciences of the United States of America *99*, 15536-15541.

Tweedie, S., Charlton, J., Clark, V., and Bird, A. (1997). Methylation of genomes and genes at the invertebrate-vertebrate boundary. Mol Cell Biol *17*, 1469-1475.

Vardimon, L., Kressmann, A., Cedar, H., Maechler, M., and Doerfler, W. (1982). Expression of a cloned adenovirus gene is inhibited by in vitro methylation. Proceedings of the National Academy of Sciences of the United States of America *79*, 1073-1077.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T.

(2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58-69.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M.*, et al.* (2006). The Polycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871-874.

Voo, K.S., Carlone, D.L., Jacobsen, B.M., Flodin, A., and Skalnik, D.G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol *20*, 2108-2121.

Wakabayashi, Y., Watanabe, H., Inoue, J., Takeda, N., Sakata, J., Mishima, Y., Hitomi, J., Yamamoto, T., Utsuyama, M., Niwa, O.*, et al.* (2003). Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. Nature immunology *4*, 533-539.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Walter, K., Bonifer, C., and Tagoh, H. (2008). Stem cell-specific epigenetic priming and B cell-specific transcriptional activation at the mouse Cd19 locus. Blood *112*, 1673-1682.

Wang, Y., Zhang, H., Chen, Y., Sun, Y., Yang, F., Yu, W., Liang, J., Sun, L., Yang, X., Shi, L.*, et al.* (2009). LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. Cell *138*, 660-672.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P.*, et al.* (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520-562.

Watt, F., and Molloy, P.L. (1988). Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes & development *2*, 1136-1143.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet *37*, 853-862.

Weber, M., Hellmann, I., Stadler, M.B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet *39*, 457-466.

Webster, K.E., O'Bryan, M.K., Fletcher, S., Crewther, P.E., Aapola, U., Craig, J., Harrison, D.K., Aung, H., Phutikanit, N., Lyle, R.*, et al.* (2005). Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. Proceedings of the National Academy of Sciences of the United States of America *102*, 4068-4073.

Wei, G., Wei, L., Zhu, J., Zang, C., Hu-Li, J., Yao, Z., Cui, K., Kanno, Y., Roh, T.Y., Watford, W.T., *et al.* (2009). Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity *30*, 155-167.

Widschwendter, M., Fiegl, H., Egle, D., Mueller-Holzner, E., Spizzo, G., Marth, C., Weisenberger, D.J., Campan, M., Young, J., Jacobs, I., *et al.* (2007). Epigenetic stem cell signature in cancer. Nat Genet *39*, 157-158.

Wiebauer, K., and Jiricny, J. (1990). Mismatch-specific thymine DNA glycosylase and DNA polymerase beta mediate the correction of G.T mispairs in nuclear extracts from human cells. Proceedings of the National Academy of Sciences of the United States of America *87*, 5842-5845.

Wikstrom, I., Forssell, J., Goncalves, M., Colucci, F., and Holmberg, D. (2006). E2-2 regulates the expansion of pro-B cells and follicular versus marginal zone decisions. J Immunol *177*, 6723-6729.

Wikstrom, I., Forssell, J., Penha-Goncalves, M.N., Bergqvist, I., and Holmberg, D. (2008). A role for E2-2 at the DN3 stage of early thymopoiesis. Molecular immunology *45*, 3302-3311.

Wilson, A.S., Power, B.E., and Molloy, P.L. (2007). DNA hypomethylation and human diseases. Biochimica et biophysica acta *1775*, 138-162.

Winders, B.R., Schwartz, R.H., and Bruniquel, D. (2004). A distinct region of the murine IFN-gamma promoter is hypomethylated from early T cell development through mature naive and Th1 cell differentiation, but is hypermethylated in Th2 cells. J Immunol *173*, 7377-7384.

Woodcock, D.M., Lawler, C.B., Linsenmeyer, M.E., Doherty, J.P., and Warren, W.D. (1997). Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. The Journal of biological chemistry *272*, 7810-7816.

Worbs, T., and Forster, R. (2007). A key role for CCR7 in establishing central and peripheral tolerance. Trends in immunology *28*, 274-280.

Wu, H.A., and Bernstein, E. (2008). Partners in imprinting: noncoding RNA and polycomb group proteins. Developmental cell *15*, 637-638.

Wu, J., and Lingrel, J.B. (2005). Kruppel-like factor 2, a novel immediate-early transcriptional factor, regulates IL-2 expression in T lymphocyte activation. J Immunol *175*, 3060-3066.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P., *et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

Xue, Y., Wong, J., Moreno, G.T., Young, M.K., Cote, J., and Wang, W. (1998). NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. Molecular cell *2*, 851-861.

Yano, S., Ghosh, P., Kusaba, H., Buchholz, M., and Longo, D.L. (2003). Effect of promoter methylation on the regulation of IFN-gamma gene during in vitro differentiation of human peripheral blood T cells into a Th2 population. J Immunol *171*, 2510-2516.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Youdell, M.L., Kizer, K.O., Kisseleva-Romanova, E., Fuchs, S.M., Duro, E., Strahl, B.D., and Mellor, J. (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. Mol Cell Biol *28*, 4915-4926.

Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S.*, et al.* (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. Proceedings of the National Academy of Sciences of the United States of America *102*, 17551-17558.

Yu, C.R., Mahdi, R.M., Ebong, S., Vistica, B.P., Gery, I., and Egwuagu, C.E. (2003). Suppressor of cytokine signaling 3 regulates proliferation and activation of T-helper cells. The Journal of biological chemistry *278*, 29752-29759.

Yu, Q., Thieu, V.T., and Kaplan, M.H. (2007). Stat4 limits DNA methyltransferase recruitment and DNA methylation of the IL-18Ralpha gene during Th1 differentiation. The EMBO journal *26*, 2052-2060.

Zhang, H., Meng, F., Chu, C.L., Takai, T., and Lowell, C.A. (2005). The Src family kinases Hck and Fgr negatively regulate neutrophil and dendritic cell chemokine signaling via PIR-B. Immunity *22*, 235-246.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E.*, et al.* (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189-1201.

Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W.*, et al.* (2010). Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. Nucleic Acids Res.

Zhang, Y., LeRoy, G., Seelig, H.P., Lane, W.S., and Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell *95*, 279-289.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999). Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes & development *13*, 1924-1935.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750-756.

Zhao, X., Ueba, T., Christie, B.R., Barkho, B., McConnell, M.J., Nakashima, K., Lein, E.S., Eadie, B.D., Willhoite, A.R., Muotri, A.R.*, et al.* (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal

function. Proceedings of the National Academy of Sciences of the United States of America *100*, 6777-6782.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T., and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. Nat Genet *39*, 61-69.

Appendix

1. CpG islands influence chromatin structure via the CpG-binding protein Cfp1

2. A temporal threshold for formaldehyde crosslinking and fixation