Global Analysis of the Methyl-CpG Binding Protein MeCP2

Peter J. Skene

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2010

Declaration

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

Peter J. Skene
April 2010

Acknowledgements

I first met Adrian whilst working in the Mens department of Debenhams, from where he hoped my delicate touch would extend to performing experiments. Since then, working in the Bird laboratory has been a great experience, with the sharing of scientific ideas, vast quantities of cake and the general banter that defines this lab. I would particularly like to thank Rob Illingworth, Shaun Webb and Alastair Kerr for bioinformatic support. Jim Selfridge has provided support, especially in the organisation of the mouse lines, along with Dina De Sousa and Jacky Guy. Through collaboration with the Wellcome Trust Sanger Institute we have been fortunate to take advantage of high-throughput sequencing methods, which have been instrumental in these studies.

I am very grateful to people in the 'real world' that have got me to this point. My parents have been fantastic, with my mother always wanting me to become a doctor and my father without whom I'd still be stuck up to my neck in a cattle shed. My friends have provided welcome relief from the scientific bubble, with whom I have shared many great adventures mountain biking around the world. Most importantly to Suz, for all the fun and general monkeying around that she brings to our lives.

Abstract

MeCP2 was initially identified as an abundant protein in the brain, with an affinity for methylated DNA *in vitro*. Interestingly, both deficiency and excess of the protein leads to severe neurological problems, such as Rett syndrome, which is the result of mutations in the *MECP2* gene. Subsequent transfection experiments showed that MeCP2 can recruit co-repressor complexes and inhibit gene expression *in vivo*. MeCP2 was therefore thought to repress specific gene targets and the aetiology of Rett syndrome was proposed to result from aberrant gene expression in the MeCP2-deficient brain. Although gene expression is perturbed in the *Mecp2*-null mouse brain, few specific targets have been verified and alternative hypotheses for MeCP2 function have been put forward. Previous binding studies have also failed to clearly identify MeCP2 targets. To shed light on these matters, a novel technique was generated to isolate neuronal and glial nuclei and established that the amount of MeCP2 is unexpectedly high in neurons, with an abundance approaching that of the histone octamer. Chromatin immunoprecipitation experiments on mature mouse brain showed widespread binding of MeCP2, consistent with its high abundance, tracking the methyl-CpG density of the genome. MeCP2 deficiency results in global changes in neuronal chromatin structure, including elevated histone acetylation and a doubling of histone H1. The mutant brain also shows elevated transcription of repetitive elements, which are distributed throughout the mouse genome. Based on this data, we propose that MeCP2 binds genome wide and suppresses spurious transcription through binding in a DNA methylation dependent manner.

Table of Contents

List of Figures and Tables

Abbreviations

BSA	Bovine serum albumin
ATRX	Alpha thalassemia/mental retardation syndrome X-linked
bp	Base pair
ChIP	Chromatin immunoprecipitation
ChIP-chip	Microarray based analysis of DNA recovered from ChIP analysis
ChIP-qPCR	Quantitative PCR based analysis of DNA recovered from ChIP analysis
ChIP-seq	High-throughput sequencing of DNA recovered from ChIP
CREB	cAMP response element binding protein
CTCF	CCCTC-binding factor
CXXC	Zinc finger domain present in a number of proteins including MBD1 and Cfp1
DDM1	Deficient in DNA methylation 1
DMR	Differentially methylated region
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTPs	deoxynucleotide triphosphates
EDTA	Ethylenediaminetetraacetic acid
EED	Embryonic Ectoderm Development protein
EGTA	Ethylene glycol tetraacetic acid
ENCODE	ENCyclopedia Of DNA Elements Project
ES cells	Embryonic stem cells
EZH2	Enhancer of zeste homolog 2
FACS	Fluorescence-activated cell sorting
H3Ac	Acetylated histone H3
H3K9me3	Trimethylated lysine 9 of histone H3
H4Ac	Acetylated histone H4
HDACs	Histone deacetylases
HP1	Heterochromatin protein 1
HRP	Horseradish peroxidase
IAP	Intracisternal A particles
ICR	Imprinting control region
IP	Immunprecipitation
IR	Infra red
kb	Kilobase
kDa	Kilodalton
KO	Knockout
KS-test	Kolmogorov-Smirnov test
LSH	Lymphoid specific helicase
Mb	Megabase
MBD	Methyl binding domain
MBD-seq	High-throughput sequencing of DNA recovered from MBD-affinity chromatography

MeCP1	Methyl DNA binding protein 1
MeCP2	Methyl CpG binding protein 2
MeDIP	Methylated DNA immunoprecipitation
MNase	*Micrococcal* nuclease
ncRNA	Non-coding RNA
NEB	New England Biolabs
NeuN	Neuronal nuclear marker protein
NTPs	Nucleotide triphosphates
PAGE	Polyacrylamide gel electrophoresis
PCM1	Protein containing methyl-CpG binding domain 1
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PHD	Plant homeo domain
PI	Pre-immune serum
RNA PolII	RNA polymerase 2
PRC2	Polycomb repressive complex 2
RIP	Repeat induced point mutation
RNA	Ribonucleic acid
RNaseA	Ribonuclease A
RT	Room temperature
SDS	Sodium dodecyl sulphate
SELEX	Systematic evolution of ligands by exponential enrichment
SEM	Standard error of the mean
siRNA	Small interfering RNA
Sp1	Specificity Protein/Krüppel-like Factor transcription factor 1
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TBS	Tris-buffered saline
TEMED	Tetramethylethylenediamine
TRD	Transcriptional repressor domain
Tris	Tris(hydroxymethyl)aminomethane
TSA	Trichostatin A
TSS	Transcription start site
UTR	Untranslated region
UV	Ultraviolet
WB	Western blot
WT	Wildtype
Xa	Active X chromosome
Xi	Inactive X chromosome
YB-1	Y box binding protein 1

Chapter 1 Introduction

1.1 Chromatin

1.1.1 Evolution of the nucleosome

The transition between prokaryotes and eukaryotes is marked by a significant increase in the number of genes and the overall size of the genome, for example, *Escherichia coli* have ~4500 genes encoded within 4.5 Mb, whereas *Homo sapiens* have ~23,000 genes within 3300 Mb. The number of distinct genes and the ability to regulate their expression can be considered as a measure of an organism's complexity. However, genome expansion cannot increase without the ability to control it, namely, to block inappropriate expression of genes and of the bulk intervening DNA (Bird and Tweedie, 1995; Bird, 1995).

There are two stark differences in the packaging of the genome between prokaryotes and eukaryotes, implying that this may be crucial in the accommodation of a larger genome. With the evolution of eukaryotes, the DNA became compartmentalised into the nucleus, with the nuclear envelope now separating the DNA from the cytosolic compartment. This had the effect of controlling which factors had access to the DNA and thereby limiting inappropriate interactions. But importantly, it also spatially uncoupled transcription and translation, therefore providing a window of opportunity for quality control before the cell commits to the translation of a messenger RNA. Indeed, it appears that a lot of transcripts are degraded by the nuclear exosome and never make it to the translation machinery, thereby reducing the deleterious effect of any spurious transcripts (Fasken and Corbett, 2009; Lewin, 1980).

Secondly, as opposed to having largely 'naked' DNA in prokaryotes, the DNA is packaged around proteins to form chromatin. The nucleosome represents the fundamental unit of this structure, consisting of 146 bp of DNA wrapped around an octamer of the four core histones (H2A, H2B, H3 and H4) (Luger et al., 1997). This repeating structure takes the form of 'beads on a string' and also is known as the 10 nm filament. This core particle can further associate with linker histones, such as histone H1. Studies using various nucleosomal positioning sequences have proposed different modes of binding for histone H1 (for review see (Thomas, 1999)), suggesting that H1 may be able to bind in various configurations depending on the local protein and sequence context. The binding of histone H1 to the linker DNA via its basic C-terminal tail has the effect of reducing histone octamer dynamics and is

thought to promote the formation of higher order structures, such as the 30 nm filament (Sera and Wolffe, 1998; Travers, 1999).

This evolution of packaging the DNA into chromatin enables compaction of the large eukaryotic genomes by up to 10,000 fold. However, chromatin is more than just a compaction tool. Nucleosomes, especially in combination with histone H1, repress spurious transcriptional initiation and therefore create a basal level of repression across the entire eukaryotic genome (Laybourn and Kadonaga, 1991). This may be due either alteration of DNA conformation or the simple occlusion of transcription factor binding sites (Almer et al., 1986; Luger et al., 1997). They appear to have been particularly adapted for reducing transcriptional noise, as they inhibit transcription from inappropriate sites without significantly affecting activator-driven initiation from authentic promoters (Laybourn and Kadonaga, 1991).

1.1.2 Histone modifications

Against this backdrop of weak basal genomic repression imposed by chromatin (Laybourn and Kadonaga, 1991), the cell needs to identify genes that need to be expressed and also those that need to be fully silenced. This is largely achieved through the covalent modification of the N-terminal tails of the core histones that protrude from the body of the nucleosome, a few specific examples are detailed below. So far numerous modifications have been identified including methylation, phosphorylation, acetylation, ADP-ribosylation, ubiquitinylation and sumoylation. This is a complex field with the interplay between histone modifications still being determined, only a brief summary will be included here (for review see (Kouzarides, 2007; Li, 2002)). Generally, histone modifications can be grouped into those that are permissive to the transcription of the underlying DNA and those that reinforce the basal repression.

There are two characterised mechanisms as to how these modifications alter the transcriptional potential. The first is through the intrinsic alteration of the chromatin structure. For example, histones can be acetylated on various lysine residues. This has the effect of neutralising the basic charge of the amino group. In the case of histone H4 acetylation this has been shown to directly disrupt inter-nucleosomal electrostatic interactions and therefore destabilises higher-order chromatin structures (Luger et al., 1997; Shogren-Knaak et al., 2006). The second and more widely understood mechanism is through

the recruitment of trans-acting factors that bind via specific domains. For example methylation can be recognised by chromo-domains and PHD-domains, whereas acetylation can be recognised by bromo-domains. The protein containing these domains often also has an associated enzymatic activity, which subsequently determines how the modification is interpreted. For example, the repressive mark, H3K27me3, is bound by the WD-40 repeats of EED (Margueron et al., 2009). This is one of the core components of the polycomb PRC2 complex, which also contains EZH2. This protein contains a SET domain which catalyses the methylation of H3K27 (Cao et al., 2002). Therefore, the recruitment of EED with the associated PRC2 complex to sites of H3K27me3 leads to the self-reinforcing propagation of this repressive mark (Margueron et al., 2009).

Generally speaking, chromatin forms two distinct global environments, either "active" euchromatin or "silent" heterochromatin, and was originally identified on the basis of differential staining (for historical perspective see (Zacharias, 1995)). Heterochromatin is typified by low levels of histone acetylation and high levels of H3K9 and H3K27 methylation, whereas euchromatin is more varied depending on the transcriptional status of the surrounding genes (for review see (Kouzarides, 2007)). By creating domains of active or repressive histone modifications, this may have enabled a simplification or 'bar-coding' of the eukaryotic genome (Barski et al., 2007) For example, the majority of mammalian promoters are marked H3K4me3, in part at least by virtue of underlying DNA sequence ((Thomson et al., 2010); also see section 1.3.4). By highlighting promoters with this modification and the subsequent recruitment of trans-activators (for review see (Ruthenburg et al., 2007)), this may reduce the 'genomic space' that the basal transcriptional machinery has to search and aid in the fidelity of transcriptional initiation.

The large number of proteins that remove these histone modifications indicate that this is a dynamic system, which allows the cell to regulate the accessibility to the underlying DNA. For example, to facilitate the passage of RNA polymerase II (RNA PolII) through the gene body during transcription, histone acetyltransferases are recruited via the C-terminal domain of RNA PolII and result in chromatin destabilisation within gene body. If acetylation was left unchecked, it would result in spurious transcription from within the now accessible gene body. However, gene bodies are also marked by H3K36me3, which recruits histone deacetylases (HDACs) to the coding region in the wake of RNA PolII (Carrozza et al., 2005). This "closing-up" of the chromatin following transcription, in part through HDAC

recruitment (Carrozza et al., 2005), indicates how eukaryotes have evolved to use histones to suppress parts of their genomes whilst still allowing the genes to be read.

As well as providing a highly dynamic system of genomic accessibility, histone modifications are also used to generate stable systems of chromatin silencing. For example, in *Drosophila melanogaster*, HP1 binds to H3K9me3 via its chromodomain (Lachner et al., 2001) and in turn creates a repressive chromatin environment by recruiting a number of proteins including HDACs (Yamada et al., 2005). But importantly, HP1 also recruits SU(VAR)3-9, which is the methyltransferase responsible for H3K9 methylation and therefore creates a self-reinforcing system of chromatin metabolism (Schotta et al., 2002). This propagation of chromatin marks is also exemplified by the polycomb protein EED as discussed above.

1.2 DNA methylation

1.2.1 Methyl-Cytosine

DNA is composed of four basic deoxyribonucleic acid building blocks (deoxyadenosine, deoxythymidine, deoxycytosine, deoxyguanosine) arranged in two anti-parallel helices. Through specific base pairing, these helices act as a mirror image of each other (Watson and Crick, 1953). The semi-conservative replication of the two DNA strands explains how the information encoded in the DNA strands is inherited between cell divisions (Meselson and Stahl, 1958) and has subsequently been dubbed the 'secret of life' (Watson and Berry, 2003).

Biochemical characterisation of DNA identified the existence of a cytosine variant, which has a methyl group on the 5 position of the pyrimidine ring (Johnson, 1925). The methyl group is transferred from S-adenosyl methionine in a reaction catalysed by the family of DNA methyltransferases (DNMTs; see section 1.3). The structure of these bases is shown in figure 1.1. Since the discovery of 5-methylcytosine, it has been found in a wide variety of organisms and is intrinsic to the 'secret of life.' However, the evolution of 5-methylcytosine as component of the hereditary material is somewhat surprising, as this modified base can deaminate to thymidine and adds a heavy mutational load on the cell (Figure 1.1; (Bird, 1980; Cooper and Youssoufian, 1988)). Despite this potential lethal feature of DNA methylation, it has been tolerated across kingdoms and therefore must overall provide a

selective advantage. One suggestion is that DNA methylation protects the genome in various ways as described below.

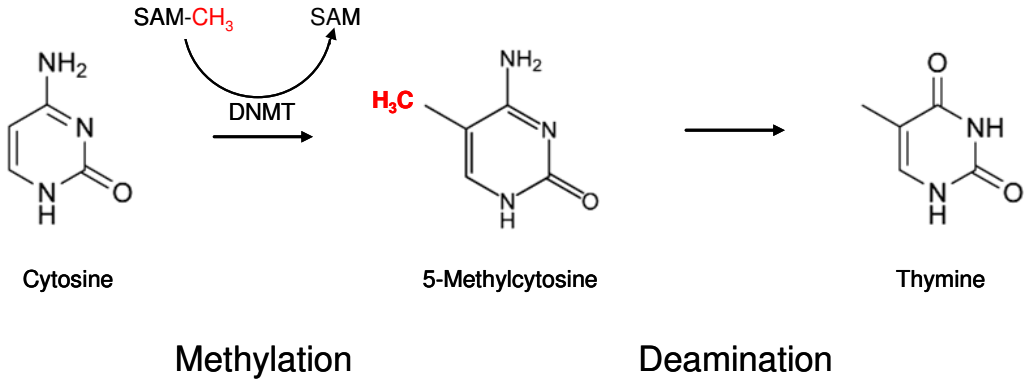

Figure 1.1 – Cytosine can become methylated on the 5-position of pyrimidine ring in a reaction catalysed by DNMT enzymes using S-adenosyl methionine as a substrate. This can spontaneous deaminate to produce a thymine and in the absence of a repair will ultimately result in a mutation. The structure of these bases is shown.

1.2.2 DNA Methylation in prokaryotes

In prokaryotes, DNA can be methylated on various bases giving rise to 5-methylcytosine, *N*6-methyladenine and *N*4-methylcytosine. The production of these bases is catalysed by various DNMTs, which bind to specific recognition sequences that contain the base to be methylated. The methylation of DNA forms part of the host restriction-modification system (for review see (Wilson and Murray, 1991)). Briefly, the bacterium expresses DNMTs that methylate its own genome whilst also expressing methyl-sensitive restriction endonucleases, which cleaves these sequences but only in the unmodified form. In the event of invasion by viruses, the endonucleases digest the unmethylated viral genome, whilst the methylated bacterial genome is protected. Other functions of DNA methylation in bacteria have also been proposed, including DNA replication initiation, DNA mismatch repair and regulation of gene expression (for review see (Palmer and Marinus, 1994)).

In contrast to prokaryotes, DNA methylation in eukaryotes occurs solely in the form of 5-methylcytosine. Surprisingly, the sequence context, distribution and function of this

modification greatly varies between and within phylogenetic groups. DNA methylation in mammals will be discussed in depth, with only a brief description of other eukaryotes.

1.2.3 DNA Methylation in fungi

The occurrence of DNA methylation in fungi is sporadic, with no clear evolutionary pattern between the species where it can be found. The *Saccharomyces cerevisiae* genome encodes no proteins related to known DNMTs and accordingly, contains no detectable cytosine methylation (Proffitt et al., 1984). In *Saccharomyces pombe* there is also no detectable methylation (Antequera et al., 1984) despite the initial identification of a putative methyltransferase, *pmt1*$^{+}$, which contains the 10 conserved domains common to eukaryotic DNMTs (Wilkinson et al., 1995). Further analysis of *pmt1*$^{+}$, identified a key point mutation in the catalytic site rendering this enzyme non-functional (Pinarbasi et al., 1996) thereby explaining the lack of cytosine methylation observed in fission yeast.

In *Neurospora crassa*, 2-3% of all cytosines are methylated, which is comparable with mammalian levels (Bird, 1980; Selker et al., 2003). However, the role of methylation is different, as methyl-cytosine is not found exclusively in a CpG context and the bulk of the genome is unmethylated in contrast to mammals. In *Neurospora crassa* DNA methylation forms part of a system known as repeat induced point mutation (RIP). Duplication of sequences during the sexual phase of the life cycle activates the RIP pathway leading to mutation of these sequences and typically leaves the altered sequences methylated ((Selker et al., 2002) and references therein). Whole genome analysis of these mutated sequences suggested that they were almost exclusively relics of transposons indicating that RIP is a genome defence system (Selker et al., 2003). The loss of DNA methylation by treatment with 5-azacytidine results in the reactivation of transposons (Zhou et al., 2001), however, the mechanism for this inhibition is currently unknown. The DNA methylation is thought propagate through a self-reinforcing system in concert with histone modifications to create stable suppression of these transposons (Selker et al., 2002). Therefore, in some fungi, DNA methylation has evolved to be used in genome dampener through the mutation of transposons.

1.2.4 DNA methylation in plants

Plant genomes contain relatively high levels of 5-methylcytosine, ranging from 6% in *Arabidopsis thalania* to 25% in maize, with cytosine methylation content broadly correlating positively with genome size and complexity (for review see (Rangwala and Richards, 2004)). DNA methylation in plant genomes is found in three nucleotide-sequence contexts: CpG, CNG and CHH sites (where N is any nucleotide and H is A, C or T). Methylation is concentrated at heterochromatic regions such as silenced ribosomal RNA genes, centromeres and transposable elements. In *Arabidopsis*, MET1 was identified as the DNMT responsible for replication dependent propagation of CpG methylation (Finnegan and Dennis, 1993). Upon deletion of MET1, an increase in pseudogene and transposon expression was observed, suggesting a role of DNA methylation in genome suppression within the bulk genome (Zhang et al., 2006).

There are a number of mechanisms for targeting DNA methylation. Transcription of intergenic regions by RNA PolII and two plant specific polymerases (Pol IV and V) have been shown to be involved in the directing DNA methylation through the siRNA machinery and the recruitment of the *de novo* methyltransferase DRM2 ((Wierzbicki et al., 2008; Zheng et al., 2009); for review see (Law and Jacobsen, 2010)). In plants, CHG methylation is thought to be maintained through a reinforcing loop involving histone and DNA methylation. Genome-wide profiling revealed a strong correlation between H3K9 and DNA methylation (Bernatavichute et al., 2008). Loss of SUVH4 (also known as KRYPTONITE), which is the histone methyltransferase largely responsible for H3K9 dimethylation, resulted in a dramatic reduction in DNA methylation (Jackson et al., 2002). Furthermore, the SRA (SET and RING associated domain) of SUVH4 can bind directly to methylated domain and suggest a self-reinforcing system between histone and DNA methylation (Johnson et al., 2007). Recent work has identified another factor that can bind to methylated DNA: RDM1, which binds to single stranded methylated DNA and associates with RNA PolII and the RNAi machinery, through this interaction it has been suggested to link siRNA production with pre-existing DNA methylation (Gao et al., 2010).

As in vertebrates, methylation is generally excluded from the promoter regions of genes (Cokus et al., 2008; Zilberman et al., 2007). However, only 20% of plant gene bodies are typically methylated, compared to the majority of vertebrate gene bodies, (Cokus et al., 2008; Zilberman et al., 2007). A strong correlation was observed with both high and low expressing genes being nonmethylated in the gene body, whereas moderately expressed

genes tended to be highly methylated in the gene body (Zilberman et al., 2007). This led to the suggestion that DNA methylation within the gene body of these moderately expressed genes was preventing spurious transcription in the wake of open chromatin after a transiting RNA polymerase 2 (RNA PolII). However, in low and high expressed genes, stable nucleosomes or a high density of RNA PolII, respectively, block this spurious transcription initiation (Zilberman et al., 2007). This putative use of DNA methylation in suppressing spurious transcription was has also been proposed for invertebrates ((Bird and Tweedie, 1995; Simmen et al., 1999; Suzuki et al., 2007); see section 1.2.5).

1.2.5 DNA methylation in invertebrates

Invertebrates also contain a highly variable amount of CpG methylation (meCpG), generally forming into two categories. Firstly, some show absence or very low levels of methylation, such as *Caenorhabditis elegans* which does not contain any 5-methyl cytosine and contains no putative methyltransferase encoding genes (Simpson et al., 1986). Also the *Drosophila melanogaster* genome contains virtually no 5-methylcytosine, with only 0.3% 5-methylcytosine detected during early embryogenesis (Lyko et al., 2000a; Lyko et al., 2000b). This low level of methylation has recently been shown to be important in the silencing of retrotransposons in somatic cells, with the methylation of retrotransposons initiating histone H4K20 trimethylation (Phalke et al., 2009).

Secondly, some invertebrates display a mosaic pattern of methylation but still with overall levels lower (10-40% meCpG) than their vertebrate counterparts (60-90% meCpG) (Tweedie et al., 1997). Species such as the sea urchin *Echinus esculentus* (Bird et al., 1979) and the sea squirt *Ciona intestinalis* (Simmen et al., 1999) exhibit long tracts of stably methylated DNA and equally long tracts of nonmethylated DNA. Detailed analysis of the sea squirt genome indicated that the methylated regions were not specifically enriched for transposable elements (Suzuki et al., 2007), suggesting that the key role of DNA methylation was not in suppression of transposons as in *Neurospora crassa.* However, the methylated regions showed a strong preference for housekeeping genes, which typically exhibit a moderate expression level, in contrast, highly expressed genes were typically found in unmethylated domains (Suzuki et al., 2007). This led to the suggestion that methylation was inhibiting spurious transcription from within moderately transcribed regions as also suggested in *Arabidopsis* (Suzuki et al., 2007; Zilberman et al., 2007).

1.2.6 DNA methylation in vertebrates

The invertebrate-vertebrate boundary marks a distinctive evolutionary shift from a fractional to a global methylation pattern, with 60-90% of CpG dinucleotides being methylated in vertebrates (Tweedie et al., 1997). Whilst invertebrates have limited gene body methylation, which is typically restricted to housekeeping genes, almost all gene bodies are methylated in vertebrates (Suzuki et al., 2007; Tweedie et al., 1997). Indeed, amphioxus and sea squirts, invertebrates close to the evolutionary boundary, have moderate levels of gene body methylation in housekeeping genes (Suzuki et al., 2007; Tweedie et al., 1997).

The distribution of methylation is likely to give clues to its function. Methylcytosine is broadly distributed throughout the mammalian genome, with ~70% of CpG sites being methylated (Ehrlich et al., 1982; Lister et al., 2009). However, the absolute amount of methylcytosine is surprisingly low (2-3%), due to CpG sites occurring at one-fifth of the expected frequency within the bulk genome, which is largely methylated and forms ~98% of the genome (reviewed in (Bird, 1986)). This apparent lack of CpG sites is due to spontaneous deamination of methylcytosine to become thymine, resulting in a T:G mismatch. If this mismatch is not repaired before replication, it results in a C:G to T:A transition mutation in one of the daughter strands. Indeed, a third of all point mutations that lead to human disease are the result of a deamination event (Cooper and Youssoufian, 1988), with mutations at CpG sites resulting from deamination events occurring at a 10-fold higher rate than at other sites (Nachman and Crowell, 2000). Therefore, cytosine methylation drives the loss of CpG dinucleotides from the genome, resulting in an average only 1 CpG per 100 bp within the bulk genome. However, the remaining ~2% of the mammalian genome does not typically become methylated and therefore shows little CpG deficiency. These sites were first identified by their ability to be cut by methyl-sensitive restriction enzymes, such as *Hpa*II and were therefore originally known as '*Hpa*II Tiny Fragments' (Cooper et al., 1983). These stretches of nonmethylated DNA are typically 1 kb in length, with a high GC and CpG content (>50% and >0.6 o/e[1] respectively) and are now termed CpG islands (CGIs). CGIs have long been associated with promoter regions (Bird, 1986). Recent studies suggest up to 80% of all CGIs may indeed associate with a promoter (R. Illingworth; manuscript in preparation). Despite this clear association, the role and maintenance of CGIs remains elusive. Nonmethylated CGIs are generally considered to be non-repressive to transcription

1 o/e (observed/expected): is the number of CpG dinucleotides in a given length of DNA divided by that predicted by GC content

(Bird et al., 1985). However, CGIs normally remain nonmethylated in somatic tissues where the gene is not expressed (Antequera et al., 1990; Illingworth et al., 2008; Weber et al., 2007). Grossly the dichotomy of mammalian DNA methylation may be summarised as the bulk genome exhibiting a low CpG density but methylated, versus CGIs having a high CpG density and nonmethylated. This is reminiscent of the formation of chromatin domains with histone modifications, and therefore DNA methylation and CGIs may act to simplify the large mammalian genome through the highlighting of specific regions (Illingworth and Bird, 2009; Thomson et al., 2010).

Early work alluded to DNA methylation imposing transcriptional repression, with artificially methylated transgenes inserted into mouse cells being transcriptionally silenced compared to their nonmethylated counterparts (Stein et al., 1982). It is against the backdrop of repression that mammalian DNA methylation will be discussed in detail in the following sections.

Recently it has become clear that there are other forms of cytosine methylation in mammals: (1) 5-methylcytosine in a non-CpG context and (2) 5-hydroxymethylcytosine. First, high-throughput sequencing of bisulphite treated DNA showed that in human ES cells 25% of 5-methylcytosine was in a non-CpG context, whereas in a somatic cell line (fetal lung fibroblasts) only 0.02% of methylcytosine was in a non-CpG context, despite overall amounts of CpG methylation being similar (Lister et al., 2009). A study using a similar technique reported somewhat higher levels of non-CpG methylation in differentiated cell lines (approximately 10-15% of methylcytosine was found in a non-CpG context) (Laurent et al., 2010). Cytosine methylation in non-CpG contexts showed enrichment in gene bodies, with the level of methylation positively correlating with gene expression level (Lister et al., 2009). The functional role of non-CpG methylation and how this mark is interpreted are not clear, with the abundance in stem cells probably as a result of high levels of *de novo* methyltransferase activity (for discussion see (Laurent et al., 2010; Lister et al., 2009)). Second, cytosine has also been recently shown to exist in a 5-hydroxymethylcytosine form (Kriaucionis and Heintz, 2009). 5-hydroxymethylcytosine is highly abundant in the brain (specifically the purkinje cells where it is present at 0.6% of total nucleotides), suggesting a role in epigenetic control of neuronal function (Kriaucionis and Heintz, 2009). The production of 5-hydroxymethylcytosine from 5-methylcytosine has been shown to be catalysed by the TET1 (Tahiliani et al., 2009). The functional role of 5-hydroxymethylcytosine is currently unclear, but it should be noted that MeCP2 has been shown to be unable to bind to this modified base (Valinluck et al., 2004) and can cause

alterations in the specificity of DNMT1 (see section 1.3.1 and (Valinluck and Sowers, 2007)). As such, 5-hydroxymethylcytosine may act to recruit or exclude factors or indeed cause passive demethylation by excluding DNMT1 (Tahiliani et al., 2009).

1.3 Mammalian DNA methyltransferases

Methylation is incorporated into the DNA post-synthetically (Burdon and Adams, 1969; Scarano et al., 1965) and is catalysed by a family of enzymes known as DNA methyltransferases (DNMTs). So far, five mammalian DNMTs have been identified, with their role in establishment and propagation of the methylation signal being largely determined by their substrate specificity (for review see (Hermann et al., 2004)). All known mammalian DNMTs share a common C-terminal domain structure that contains the catalytic domain (Figure 1.2). This domain resembles the prokaryotic counterparts and is characterised by the 10 conserved amino acid motifs required for methyltransferase activity (Klimasauskas et al., 1994; Reinisch et al., 1995).

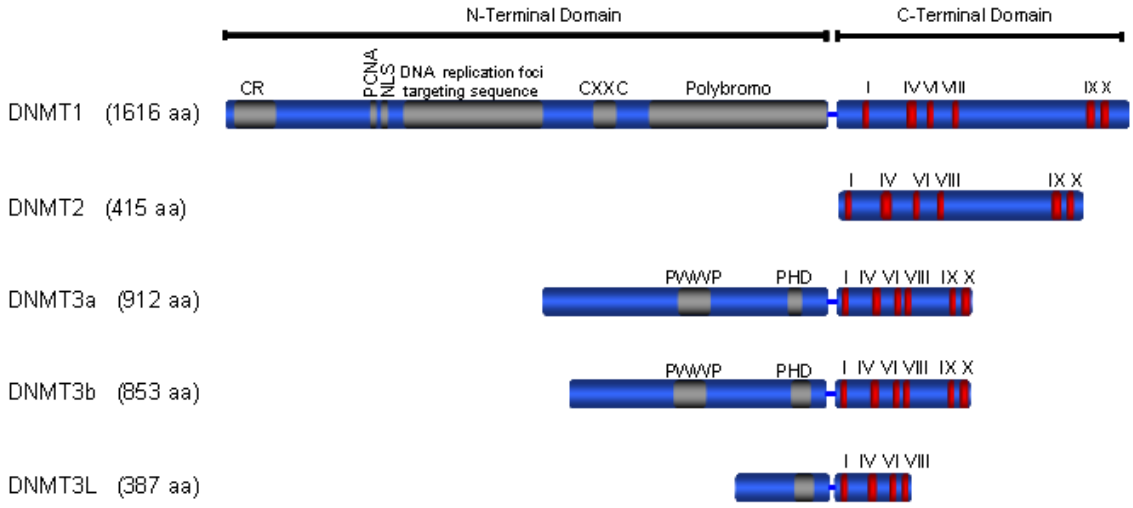

Figure 1.2 – Domain organisation of the mammalian DNMTs. The C-terminal domain contains the 10 conserved amino acid motifs (indicated by with red bars), which are required for catalytic activity and show homology to the prokaryotic DNMTs. The N-terminal domain contains various motifs which are required for localisation and regulation (indicated with grey bars). For details of the regulatory regions see (Hermann et al., 2004). Briefly they are as follows: CR: charge-rich region; PCNA: PCNA binding site; NLS: nuclear localisation site; CXXC: CXXC type Zinc finger; Polybromo domain containing two BAH domains; PWWP: domain containing conserved proline and tryptophan residues; PHD: plant homeodomain . DNMT3L is discussed in section 1.3.4. Figure adapted from (Hermann et al., 2004).

The catalytic domain forms a characteristic six-stranded β-sheet flanked by 1 or 2 α-helices to create the binding pocket for the substrate *S*-adenosyl methionine (SAM) and the target DNA. The donor methyl group is bound to a sulphonium atom, which increases the reactivity of the relatively inert methyl group, leaving it liable to nucleophilic attack. The DNMT binds to the DNA and disrupts the C/G base pairing and flips out the cytosine base into the catalytic site. The now accessible cytosine undergoes a series of reactions resulting in the formation of dihydrocysteine covalently coupled DNA-enzyme intermediate. Nucleophilic attack on the activated SAM results in transfer of the methyl group onto the carbon 5 position of cytosine and the ultimate release of *S*-adeonsyl-L-homocysteine as a byproduct (for review of catalytic mechanism see (Jeltsch, 2002)). A brief description of these enzymes will be given below.

1.3.1 DNMT1

A regulatory domain in the large N-terminal region ensures that the preferred substrate for DNMT1 is hemi-methylated DNA, as deletion of this motif removes this specificity (Bestor, 1992). Therefore, DNMT1 has the ability to copy the methylation pattern from one strand to another and has led to DNMT1 being also known as the maintenance methyltransferase. The N-terminus of DNMT1 also mediates an interaction with PCNA (Chuang et al., 1997), thereby associating DNMT1 with replication forks (Leonhardt et al., 1992) and was thought to allow the faithful replication of the 5-methylcytosine marks onto the newly synthesized strand. However, the interaction with PCNA is only transient and only increases the methylation efficiency by 2-fold (Schermelleh et al., 2007) suggesting that other factors must aid the targeting of DNMT1. Hemi-methylated sites that are missed during the passage of the replication fork can 'picked up' through the binding of UHRF1 to these sites and the subsequent recruitment of DNMT1 (for review see (Jeltsch, 2008)). Mouse cells that are ablated for the *Uhrf1* gene exhibit mislocalisation of DNMT1 and a resulting loss in methylation, indicating that this is an essential backup system for the targeting of Dnmt1 (Bostick et al., 2007; Sharif et al., 2007).

As such DNMT1 is largely considered to be responsible for the heritable nature of DNA methylation and is universally expressed within the nucleus of replicating somatic cells by virtue of a nuclear localisation signal. However, during the initial rounds of DNA replication in fertilised eggs, DNMT1 is sequestered into the cytoplasm and leads to passive demethylation of the female genome within the pre-implantation embryo (Carlson et al.,

1992). During this early phase of development an alternative splice variant maintenance methyltransferase, DNMT1o, is expressed within the nucleus and ensures the faithful propagation of certain imprinted loci (Ratnam et al., 2002).

The importance of DNA methylation and its maintenance can be inferred from the lethality of the *Dnmt1* knock-out mouse (Li et al., 1992). Both the knock-out mouse and embryonic stem (ES) cells show approximately 70% reduction in 5-methylcytosine. However, despite the mutant ES cells being viable, the mouse was embryonic lethal, with the stunted embryos terminating at midgestation (Li et al., 1992). This indicates that DNMT1 is essential for the completion of development. However, the faithful propagation of methylation is not that simple. Early studies using artificially methylated constructs inserted into terminally differentiated cell lines suggested that methylation was not maintained indefinitely, with a failure rate of ~5% per CpG per cell division (Pollack et al., 1980; Wigler et al., 1981). This has also been confirmed for an endogenous locus (Riggs et al., 1998), suggesting that maintenance methylation through replication by DNMT1 is not sufficient for the long term propagation of these marks and that there must be other factors involved (Bird, 2002). There are potential roles of DNMT1 other than cytosine methylation, with DNMT1 been shown to interact with HDAC1 (Myant and Stancheva, 2008), raising the possibility that DNMT recruit factors to influence the local chromatin state. It has, however, been shown that mutation of a key catalytic residue in DNMT1 failed to rescue any of the phenotypes of the *Dnmt1*-null ES cells, suggesting that its key functions are a consequence of its methyltransferase activity (Damelin et al., 2007).

1.3.2 DNMT2

Analysis of expressed sequence tag (EST) databases for sequences with homology to DNMT1 led to the identification of a candidate DNMT, named DNMT2, which was expected to be functionally active (Yoder and Bestor, 1998). However, recent work suggests that its major role is in the cytosine methylation of a transfer RNA for aspartic acid (Goll et al., 2006). This is consistent with its cytoplasmic localisation in zebrafish and the lack of change in methylation levels in null mouse ES cells (Okano et al., 1998b; Rai et al., 2007).

1.3.3 DNMT3a and 3b

Despite the global demethylation observed during pre-implantation development, methylation reaches global levels later in development (for review see (Reik and Walter,

2001)). This coupled with the recorded failure of DNMT1 to maintain methylation (Riggs et al., 1998) and the observation that nonmethylated retroviral DNA inserted into *Dnmt1*-deficient ES cells can become methylated (Lei et al., 1996) suggested that some *de novo* methyltransferase activity must exist. The differentiation of these *Dnmt1*-deficient ES cells into the embryoid bodies resulted in the loss of methylation from endogenous retroviral sequences (Lei et al., 1996). This suggested that the *de novo* methyltransferase activity was most abundant within ES cells and was predicted to be required for the re-setting of the methylation patterns during development (Jahner et al., 1982; Stewart et al., 1982).

By searching for human proteins that showed homology to bacterial methyltransferases, two putative enzymes were identified and termed DNMT3a and 3b (Okano et al., 1998a). Expression analysis indicated that these enzymes were most highly expressed within ES cells suggesting that these factors may indeed explain this missing activity (Okano et al., 1998a). Furthermore, *in vitro* methylation studies suggested that these recombinant enzymes could act on nonmethylated DNA and showed no preference for hemi-methylated DNA (Okano et al., 1998a). The targeted disruption of the genes for *Dnmt3a* and *3b* in mouse ES cells indicated that in the absence of both proteins there was a loss of *de novo* methylation, but single gene disruptions maintained *de novo* activity, suggesting that they may have redundant functions (Okano et al., 1999). Analysis of the corresponding mouse phenotypes showed that *Dnmt3a* knockout mice survived to term but then typically died at 4 weeks of age, whilst *Dnmt3b* knockout mice were not viable and failed to develop normally past embryonic day 9.5 (Okano et al., 1999). The double knockout mouse, presented with a more severe phenotype, indicating that they have partially overlapping functions (Okano et al., 1999), suggesting that differences in function may lie in the N-terminal regulatory domain (Figure 1.2; for review see (Hermann et al., 2004)). Methylation analysis of the genomic DNA from these ES cells revealed that loss of *Dnmt3a* resulted in hypomethylation of major satellite repeats and a loss of some maternal imprints, whereas loss of *Dnmt3b* led to hypomethylation of minor satellite repeats (Chen et al., 2003a; Okano et al., 1999), suggestive of alternative targeting mechanisms.

Extended culture of ES cells lacking both *Dnmt3a* and *3b* led to a progressive loss of methylation to levels approaching that of *Dnmt1* knock out ES cells (Chen et al., 2003a). Overall, this suggests that both DNMT3a and 3b are required in concert with DNMT1 for the faithful propagation of DNA methylation, arguing against the early simplistic labels of 'maintenance' and '*de novo*' methyltransferases (for review see (Bird, 2002)). Indeed, there

is some evidence from exogenously expressed proteins indicating that they may interact (Kim et al., 2002). A more realistic model is one where all three enzymes work in concert to maintain methylation patterns with DNMT3a and 3b helping to fill in gaps left by DNMT1. This may go some way to explaining the lack of specific homogeneous methylation patterns with respect to individual CpG sites, but instead that the overall methylation content of domains is maintained (Bird, 2002; Silva et al., 1993; Stoger et al., 1997).

1.3.4 Targeting DNA methylation: cross-talk with chromatin

As previously discussed the mammalian genome can largely be divided into (1) the bulk genome, which has a low CpG density and is typically methylated, and (2) the CGIs, which have a high CpG density and are usually nonmethylated. This leads to three potential models for the formation of this methylation pattern. Firstly, methylation is the default and CGIs somehow are actively protected from this machinery. Secondly, the CGIs are innocent bystanders, whilst the bulk genome is actively targeted. Thirdly, there is some middle ground between these models, with active repulsion and recruitment of the DNA methylation machinery. There are indeed some special cases for the role of DNA methylation and these will be discussed in section 1.4, whereas this section will focus on how this global pattern is set up. There is also the potential for active demethylation to occur, with the possibility that this could be involved in the maintenance of CGIs as methylation free. In zebrafish, active demethylation has been shown to occur through a deamination step followed by the subsequent repair by MBD4 (see section 1.7.7; (Rai et al., 2008)).

Recent work provides an intriguing link between chromatin and CGIs to suggest how these sequences may be protected from *de novo* methylation. DNMT3L was identified through homology searches with other members of the DNMT3 family (Aapola et al., 2000). The N-terminal region contains a PHD domain, whilst the truncated C-terminal domain only contains 8 of the 10 conserved motifs common to methyltransferases and importantly lacks the key catalytic residues. The expression pattern of DNMT3L was shown to be reminiscent of DNMT3a and 3b (Hata et al., 2002). Indeed, DNMT3L could interact with both these active enzymes (Hata et al., 2002), and *in* vitro methylation assays suggested that Dnmt3L could stimulate their methyltransferase activities (Suetake et al., 2004). However, recombinant DNMT3L was shown to be incapable of binding DNA *in vitro* and was therefore assumed to not being involved in the targeting of the de novo methyltransferases but instead in the modulation of their activity (Suetake et al., 2004). More recently, peptide

interaction assays indicated that DNMT3L bound to the N-terminus of histone H3, but that this interaction was abolished by the specific methylation of lysine 4 (Ooi et al., 2007). More recently CGIs have been shown to recruit Cfp1 (a CXXC CpG binding protein), by virtue of their density of nonmethylated CpGs, which in turn recruits a H3K4 methyltransferase, with the overall effect of 75% of CGIs being bound by Cfp1 and containing H3K4me3 modified chromatin ((Thomson et al., 2010); see Appendix B). Therefore, through specific chromatin modification, as directed by the CGI sequence alone, it is proposed that they are maintained refractory to DNA methylation.

Other studies have also looked at the maintenance of nonmethylated CGIs. Using a transgenic mouse assay, the *Aprt* CGI became methylated upon removal of binding sites for Sp1, a transcriptional transactivator (Macleod et al., 1994). This led to the suggestion that active transcription was required for the protection of CGIs. Given the GC-rich consensus sequence for Sp1 binding (Letovsky and Dynan, 1989) and the requirement for Sp1 binding to maintain the methylation-free status of the *Aprt* CGI, it was hypothesised that there may have been a general requirement for Sp1 for the maintenance of CGIs (for discussion see (Marin et al., 1997)). The generation of *Sp1*-null ES cells, however, showed no gross change in the methylation status of CGIs, suggesting that Sp1 is not generally required the maintenance of CGIs as methylation-free (Marin et al., 1997). More recently it has been shown that another factor, VEZF1, with a similar but distinct GC-rich consensus binding site to Sp1, could bind to one of the "Sp1-sites" at the wild-type *Aprt* promoter (Dickson et al., 2010). This study went on to show that VEZF1 binding in the absence of Sp1 binding was sufficient for the protection of the CGI from methylation (Dickson et al., 2010). It remains possible that VEZF1 may play a global role in the maintenance of CGIs as methylation-free.

Given the early suggestion of a role of transcription in the protection of CGIs, it should be noted that it has long been known that the majority of CGIs remain unmethylated irrespective of their transcriptional activity, such as the human α-globin gene (Bird et al., 1987) and more recent genome-wide studies (Weber et al., 2007). Perhaps this apparent discordance can be explained by a study revealing that H3K4me3 was enriched at both active and inactive promoters, but unexpectedly even 'inactive' promoters were bound by RNA PolII and associated with non-productive transcriptional initiation (Guenther et al., 2007). This raises questions over what constitutes an active or an inactive promoter, with more sensitive techniques required. This may suggest that the Sp1-depleted *Aprt* promoter in the transgenic mouse was no longer able to recruit RNA PolII as per typical 'inactive'

promoters. But this does not explain why the *Aprt* CGI sequence alone did not direct its methylation status through the recruitment of Cfp1, as suggested ((Thomson et al., 2010), Appendix B). Speculatively, transcriptional initiation and Cfp1 may work synergistically to maintain CGIs.

The role of transcription in defining the methylation pattern has also been proposed in the context of transcriptional elongation. Using methyl-sensitive restriction of single nucleotide polymorphisms, gene bodies were shown to be specifically methylated on the active X chromosome (Xa) relative to the inactive X chromosome (Xi) (Hellman and Chess, 2007). One hypothesis would be that methylation of gene bodies on the Xa is required to prevent spurious initiation within the unfolded chromatin in the wake of transiting RNA PolII, reminiscent of the suggestions in *Arabidopsis* and sea squirt (Suzuki et al., 2007; Zilberman et al., 2007). The mechanism of recruiting the methylation machinery to gene bodies is currently unknown. In plants, transcription of repetitive elements has been shown to result in DNA methylation through the RNAi pathway (for review see (Chan et al., 2005)) and was proposed to have a role in plant gene body methylation (Zilberman et al., 2007). However, there is no evidence for a similar mechanism in mammals.

Potential targeting mechanisms can be elucidated through studying the interacting partners of the DNMTs. A genetic study in *Arabidopsis* identified a requirement for DDM1, a SWI2/SNF2-like chromatin remodeler, in the methylation of heterochromatic loci (Jeddeloh et al., 1999). Targeted disruption of LSH, the mammalian homolog, resulted in ~70% loss of methylation throughout the genome and is in complex with the DNMTs (Dennis et al., 2001; Myant and Stancheva, 2008). Similarly, mutation of ATRX, another SWI2/SNF2-like chromatin remodeler, resulted in hypomethylation of ribosomal DNA arrays and specific repeats (Gibbons et al., 2000). However, it is not clear if these interacting partners are involved in the targeting of DNMTs or in facilitating access to the DNA once targeted. Analysis of non-vertebrate species suggests that DNMTs might be recruited by chromatin modifications as well as repelled, as already discussed for H3K4me3. In the fungus *Neurospora crassa*, histone H3K9 methylation by a SET-domain containing protein has been shown to be required for DNA methylation (Tamaru and Selker, 2001). Similarly, in *Arabidopsis*, the H3K9 methyltransferase, Kryptonite, has been shown to be required for CNG methylation, with some suggestion that the DNA methyltransferase CMT3 may be directly recruited by this modification (for review see (Chan et al., 2005)). Currently, in mammals, there is only some evidence to suggest heterochromatic modifications can direct

DNA methylation. A study suggested that H3K27me3 can recruit DNMT activity (Vire et al., 2005). A large amount of contention, however, surrounded this study, due to the use of image manipulation. Despite this, another study reported the role of polycomb complexes in the methylation of the *HOXA9* promoter (Reynolds et al., 2006). Further analysis of mammalian DNMT interacting partners has shown that HDAC complexes can associate with DNMT3a (Fuks et al., 2001). However, it is likely that this interaction is not targeting methylation, but providing a link between DNA-based and chromatin-based repressive systems, as has been suggested for the interaction between G9a, a H3K9 methyltransferase, and DNMT1 (Esteve et al., 2006). This leaves open the question of how, and if, methylation is targeted to the bulk genome.

Finally, analysis of the timecourse of methylation may shed light onto its targeting. Expression of retroviral elements inserted into carcinoma cells has shown to be silenced ~2 days after infection, whereas *de novo* methylation is delayed until ~ 15 days (Gautsch and Wilson, 1983). This is consistent with observations that retroviruses can be shutdown in the absence of the *de novo* methyltransferases (Pannell et al., 2000). This is compatible with the hypothesis that transcriptional initiation maybe required to maintain the DNA in an unmodified state, and in the absence of initiation that the default is to methylate. Potentially against this hypothesis is the observation that insertion of multiple copies of a transgene into the mammalian genome correlated with gene silencing and DNA methylation, whilst a single copy remained active and unmethylated (Garrick et al., 1998), somewhat reminiscent of RIP in *Neurospora*. However, the timecourse of silencing compared to the onset of methylation was not evaluated, leaving open the possibility that these transgenes were silenced before the default status of methylation encroached. Despite this suggestion that genomic methylation is the default status, it may remain that some compartments of the genome are actively targeted, for example all gene bodies may be specifically targeted, as suggested for the gene bodies on the Xa (see above; (Hellman and Chess, 2007)). Details of these regions and the mechanism of targeting are likely to be spurred on by the recent high-throughput bisulphite sequencing projects (Laurent et al.; Lister et al., 2009).

1.4 Role of mammalian DNA methylation

After fertilisation the DNA methylation pattern is generally wiped clean. The male pro-nucleus is thought to undergo active demethylation, through a currently unknown demethylase activity, whilst the female pro-nucleus divides and passively loses DNA

methylation (Monk et al., 1991; Santos et al., 2002). From this point the genome becomes largely methylated and this pattern is relatively stably inherited throughout the organism. In this manner DNA methylation is unusual compared to histone modifications, as it is largely considered to be an irreversible mark. The mouse phenotypes associated with the loss of the DNMTs indicates the essential nature of these methylation patterns (Li et al., 1992; Okano et al., 1999). This essential requirement is mirrored by mutations in the catalytic domain of DNMT3B resulting in the human condition ICF (immunodeficiency, centromere instability and facial anomalies), associated with hypomethylation of satellite DNA (Jeanpierre et al., 1993; Xu et al., 1999). Despite the necessity of DNA methylation for mammalian development, the role it plays has however yet to be fully elucidated. Several hypotheses will be discussed.

1.4.1 Genome stability and integrity

Early observations indicated that treatment of human cells with the demethylating agent, 5-azadeoxcytidine, resulted in an increase in chromosome associations by satellites, secondary constrictions, and telomeric regions (Viegas-Pequignot and Dutrillaux, 1976). Similar abnormalities were observed with lymphocytes derived from ICF patients. In these cells, the juxtacentromeric heterochromatin was abnormally hypomethylated and displayed greatly elongated and thread-like structures in metaphase chromosomes, which is associated with the formation of complex multiradiate chromosomes (Xu et al., 1999). It is however, not clear that a causal relationship exists between the loss of DNA methylation and chromosomal abnormalities, as 5-azadeoxcytidine, can also result in the crosslinking of proteins to DNA (Juttermann et al., 1994) and the chromosome rearrangements observed in these lymphocytes derived from ICF-patients had been cultured for sometime (Xu et al., 1999).

However, genetic studies have added further weight to a role of DNA methylation in genome integrity, with ES cells deficient for *Dnmt1* exhibiting an increased mutation rate suggested to be the result of increased mitotic recombination and associated chromosomal loss (Chen et al., 1998). Furthermore, mice that are hypomorphic for *Dnmt1*, which express 10% of protein compared to wildtype, resulted in a global loss of methylation and developed aggressive T-cell lymphomas at 4-8 months. These tumours often displayed trisomy of chromosome 15, indicating that demethylation may be linked with carcinogenesis through promoting chromosome instability (Eden et al., 2003; Gaudet et al., 2003).

1.4.2 Genome defence

Another suggestion for the role of mammalian genome-wide methylation is that of silencing transposable elements, as previously discussed for the fungus *Neurospora crassa* (see section 1.2.3). However, whether this is true in mammalian organisms is debatable. Mouse cells normally repress the transcription of transposable elements, such intracisternal A particle (IAP) elements. In mouse embryos lacking DNMT1, the transcription of such sites is upregulated 50-fold, suggesting a key role of methylation in their silencing (Walsh et al., 1998). The purpose of this methylation-dependent silencing is debated, and is thought to be either required for the specific suppression of transposition (Yoder et al., 1997) or in the general reduction of transcriptional noise imposed by genome-wide methylation (Bird, 1995). Currently, there no evidence that increased transcription of these elements, resulting from demethylation, leads to in increased transposition. Transposition would only display a heritable phenotype if it occurred within the germ cell lineage, therefore suggesting that transposon methylation at this timepoint would be imperative for in the genome defence model. However, at this stage in development, transposons within germ cells are hypomethylated (for review see (Reik and Walter, 2001)) and IAP elements are often transcriptionally active (for review see (Bird, 1997). It is only later in development where these elements become methylated and silenced, and furthermore, methylation has been shown to be secondary to transcriptional silencing of inserted retroviral sequences (Pannell et al., 2000). Phylogenetic studies may also argue against this model of mammalian genome defence, as methylation analysis of the sea squirt *Ciona intestinalis* showed no indication of targeted transposon methylation, with the methylation status determined by the insertion site (Suzuki et al., 2007). Indeed, mammalian transposons do not appear to be anymore targeted for methylation than the rest of the genome (Rabinowicz et al., 2003). However, despite the limited evidence for a role of DNA methylation in genome defence, there is also currently little evidence to suggest that there is an increase in transcriptional noise associated with of loss of genome-wide methylation.

1.4.3 Transcriptional repression

One of the best characterised functions of DNA methylation is transcriptional repression and this may provide clues as to the reason for the evolution of global DNA methylation in vertebrates. Vertebrates have an increased complexity, with more cell types and a moderate

increase in the number of genes, which require careful temporal and spatial regulation and also an increased bulk genome size. It has been suggested that genomic expansion and complexity of expression patterns is limited by the ability of the cell to faithfully repress and express genes as required (Bird and Tweedie, 1995). As eukaryotes evolved to use nucleosomes to suppress their increased genome size, vertebrates may have evolved to use global DNA methylation in concert with chromatin. A study using a variety of transfected methylated constructs confirmed the repressive nature of methylation as seen in earlier studies, but discovered that the density of meCpG sites correlated with its repressive activity, with low density methylation only sufficient to repress weak promoters (Boyes and Bird, 1992), in a manner reminiscent of chromatin (Laybourn and Kadonaga, 1991). Analysis of the average methylation density within the vertebrate bulk genome indicated a density of ~1 meCpG per 100 bp, which is sufficient to suppress spurious initiation whilst not affecting authentic promoters, and therefore has the characteristics of a noise reduction system for the bulk genome. In contrast to this portion of the genome, the majority of mammalian promoters are embedded within CGIs (Illingworth et al., 2008), and are therefore typically devoid of repressive DNA methylation. This stark contrast between promoter regions and the surrounding bulk genome may facilitate in faithful transcriptional control by acting as landing lights for transcription factors (for review see (Illingworth and Bird, 2009)), whilst suppressing spurious activity outside of these key sites.

Potentially contrary to this model of DNA methylation as a noise reduction system are recent genome-wide transcription profiling studies, including the ENCODE project. These suggest that up to 93% of the human genome is capable of being transcribed (Birney et al., 2007). If transcription is everywhere, can mammalian DNA methylation really be suppressing transcriptional noise? The relevance of these previously unidentified transcripts is still being investigated, but there is increasing evidence that these transcripts are functional, such as anti-sense transcripts and inter-gene splice variants (for review see (Kapranov et al., 2007)). This suggests that the initial model of transcription units proposed by Jacob and Monod with the *lac* operon is far too simple for higher organisms and what may have been initially taken for as noise is actually functional (for review see (Kapranov et al., 2007).

A life-line to this noise reduction model is that global hypomethylation associated with human neoplasia has been shown lead to the expression of certain transposable elements (Rodriguez et al., 2006). In line with the increased expression of IAP elements in *Dnmt1* knock-out mouse embryos (Walsh et al., 1998). These data suggest that methylation within

the mammalian bulk genome has a suppressive effect. However, the mechanism by which this low density of methylation of ~1 meCpG per 100 bp leads to transcriptional suppression is not clear.

Whilst the mammalian genome is highly methylated, the actual density of methylation is low, due to the depleted CpG content of the bulk genome (Bird, 1980). However, there are some specific cases where the typically nonmethylated CGIs become methylated and are associated with transcriptional silencing of the gene. It seems likely that these particular sites evolved to harness the repressive ability of DNA methylation that already existed within the bulk genome. The best understood examples of CGI methylation are involved in the inactivation of the X chromosome and in imprinting, as discussed below. However, recent advances in microarrays and high-throughput sequencing have suggested that methylated CGIs are more prevalent than initially expected. A microarray study focusing on the majority of human promoters indicated that 3% of CGI-containing promoters were methylated in somatic tissues and primary cell lines (Weber et al., 2007). Subsequently, a study focusing on all CGIs, irrespective of their association with annotated promoters, suggested that 6%-8% of CGIs were methylated in primary tissues (Illingworth et al., 2008). Furthermore, the CGIs that displayed tissue-specific methylation were over-represented at numerous genetic loci that are essential for development, including HOX and PAX family members (Illingworth et al., 2008). The role and targeting of methylation to these CGIs is still being investigated. Recently, the chromatin state and the level of RNA PolII association were shown to correlate with the propensity for methylation of CGIs in cancer (Takeshima et al., 2009), suggesting that transcriptional inactivity could lead to methylation.

1.4.4 X inactivation

In mammals, sex determination is mediated by a pair of heteromorphic sex chromosomes, with females harbouring two X chromosomes, compared to males with only one. In order to equalise the expression levels of X-linked genes between the genders, females silence one copy of the X chromosome in each cell during early development (for review see (Heard et al., 2004)). Interestingly, during the first five cell divisions of the cell divisions of female embryogenesis, the paternal X chromosome is inactivated. Subsequently, the marks of this early allele-specific inactivation are removed in the cells of the inner cell mass and then random X inactivation proceeds in the epiblast (Okamoto et al., 2004).

The mechanism of random X inactivation is complex and will only briefly be discussed. There is an initially a reversible phase which ultimately leads to the irreversible repression of a single X chromosome (Wutz and Jaenisch, 2000). The expression of *Xist*, a large non-coding RNA, from the Xi initiates chromosome-wide repression in *cis* (Wutz and Jaenisch, 2000). In humans, Xi heterochromatin appears to relatively distinct forms: (1) the *Xist* RNA physically coats the Xi and associates with polycomb complexes and macro H2A; (2) the chromatin is marked with H3K9me3 and bound by HP1 (for review see (Valley and Willard, 2006)).

DNA methylation has been shown to have a central role in X inactivation. Treatment with 5-azadeoxcytidine of hybrid mammalian cells resulted in transient demethylation and reactivation of the PGK-1 gene located on the inactive X chromosome (Xi), with the heritability of the reactivation dependent on the length of the demethylated region (Hansen and Gartler, 1990). CGIs on the Xi have been shown to specifically acquire methylation during development, consistent with a role of methylation in X inactivation (Heard et al., 2004; Lock et al., 1986). This link is enhanced with the observation that CGIs of genes that escape X-inactivation are infrequently methylated when compared to those that are silenced (Carrel and Willard, 2005; Illingworth et al., 2008).

The role of DNA methylation seems to be in the maintenance rather than in establishment of silencing. The targeted disruption of *Dnmt1* in mouse ES cells led to aberrant X inactivation upon differentiation, with silencing of both X chromosomes thought to be the result of aberrant expression on *Xist* from Xa, indicating a role of methylation in the maintenance of silencing rather than the initiation (Panning and Jaenisch, 1996). Furthermore, methylation of the *Hprt* CGI has shown to occur after the gene has been silenced (Lock et al., 1987), consistent with DNA methylation acting to maintain silencing.

Surprisingly, an early study using methyl-sensitive restriction of human metaphase chromosomes followed by labelling with biotinylated nucleotide and visualisation by immunofluorescence indicated that the Xi was actually overall hypomethylated compared to the Xa (Viegas-Pequignot et al., 1988). More recent studies have indicated that the distribution of this increased Xa methylation is within the intergenic DNA (Weber et al., 2005) and within exons (Hellman and Chess, 2007). The role of Xa specific methylation has yet to be determined, but speculatively may be to suppress spurious transcription of the Xa, whereas the Xi is transcriptionally inert and not prone to inappropriate expression.

1.4.5 Imprinting

Genomic imprinting is where a particular gene is only expressed from a single allele determined by the parental origin and is controlled by 'imprints' that are laid down in the parental germ cells. Imprinting was first discovered in the 1980s where nuclear transplantation of uniparental embryos, which contain only one set of the parental chromosomes, resulted in differing phenotypes dependent on whether the maternal or paternal chromosomes were used (for review see (Reik and Walter, 2001)). Furthermore, loss of control of these imprinted sites often results in human disease, highlighting their importance. There are in region of 100 imprinted genes identified in humans, with the majority appearing within clusters in the genome, with each cluster typically containing several protein coding genes and at least one non-coding RNA (ncRNA). Each cluster is under the control of a single major *cis*-acting element, the imprinting control region (ICR) (for review see (Edwards and Ferguson-Smith, 2007)). The paradigm for this phenomenon comes from the first imprinted site to be identified: the *H19* and *Igf-2* genes, which are transcribed from the maternal and paternal alleles respectively, regulated by an ICR situated between these two genes. In *Igf2* -/+ mice, passage of the null allele through the maternal line resulted in normal offspring, but passage of this allele through the male line resulted in growth deficiency (DeChiara et al., 1990; DeChiara et al., 1991).

DNA methylation has been shown to have a determining role in genomic imprinting. The *H19* locus has been shown to be specifically methylated on the paternal allele (Ferguson-Smith et al., 1993); sites that shown parent-specific methylation are now known are differentially methylated regions (DMRs). Mouse embryos null for *Dnmt1* exhibited a loss of methylation at DMRs and resulted in the loss of imprinting of this locus, with the normally paternally silent *H19* gene now showing activation and the paternal activated *Igf-2* gene showing repression (Li et al., 1993). The reciprocal regulation of the *H19/Igf-2* locus is controlled by the binding of CTCF (CCCTC-binding factor) to four CpG-rich repeats specifically to the nonmethylated maternal *H19* DMR, which is located 2 kb upstream of the *H19* transcriptional start site (TSS). The binding of CTCF on the maternal allele acts as an insulator and blocks the enhancers from driving the *Igf-2* promoters (Arney, 2003; Bell and Felsenfeld, 2000; Edwards and Ferguson-Smith, 2007).

1.5 Epigenetics

The prefix Epi- describes an event which occurs 'upon' and therefore epigenetic literally translates into 'upon the gene'. The term epigenetics has evolved since its introduction by Conrad Waddington, where epigenesis was the study of how genotypes give rise to phenotypes during development (Waddington, 1957).Through the work of Arthur Riggs and colleagues, epigenetics was defined as "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence (Russo, 1996)." More recently, however, the term epigenetics has been used to describe a wide array of chromatin modifications and metabolism with a reduced emphasis on the inheritability of this mark (for discussion see (Bird, 2007)). Overall, this has given rise to a recent definition stating: "the structural adaptation of chromosomal regions so as to register, signal or perpetuate altered activity states" (Bird, 2007). This includes various contemporary features of cell biology such as the Polycomb and Trithorax system, histone modifications and DNA methylation. By having this system of epigenetic memory it allows the cell to maintain or perpetuate the transcriptional status as determined by sequence specific transcription factor binding events, for example in the recruitment of histone acetyltransferases to promoter following transcription and therefore priming the promoter for continued activity (for review see (Clayton et al., 2006) and references therein).

The recent advances in sequencing technology have given rise to a huge amount of genome-wide data regarding the positions of these epigenetic marks. A full review of this field is out of the scope of this discussion. Briefly, an analysis of promoters marked by H3K4me3 and H3K27me3 revealed a strong correlation with the expression level of the associated gene (Mikkelsen et al., 2007). The so-called bivalent genes, those with promoters marked by both H3K4me3 and H3K27me3 and are thought to represent genes that are developmentally poised for activation, typically showed lineage specific resolution of these marks to give either activation or silencing (Bernstein et al., 2006; Mikkelsen et al., 2007). DNA methylation patterns have also been examined using these high throughput sequencing techniques. An initial reduced representation DNA methylation analysis based on an *Msp*1 digest coupled with bisulphite sequencing indicated that as expected the CpG sites within CGIs were predominantly non-methylated and CpG sites within the bulk genome were mainly methylated (Meissner et al., 2008). However, this study identified that this pattern of CpG methylation was not universally true and that histone methylation was a better predictor of the CpG methylation status, with H3K4me3 and H3K4me2 being associated with non-methylated CpG sites, whilst H3K9me3 was associated with methylated CpG sites (Meissner

et al., 2008). The pattern of DNA methylation at CGIs and the bulk genome was shown to change upon differentiation of ES cells into neuronal progenitor cells, with corresponding changes in histone methylation (Meissner et al., 2008). The impact of changes in DNA methylation during development on gene expression and the involvement of proteins that bind meCpG sites, such as MeCP2, is likely to be further elucidated by high-throughput techniques to identify changes in this epigenetic mark.

1.6 DNA methylation and chromatin controlling landscape of the genome

Although eukaryotes and more specifically vertebrates may have evolved to use chromatin and global DNA methylation as apparently distinct means of compartmentalising their larger genomes, there are a number of parallels that can be drawn between these two mechanisms and it is clear that this evolution has been of cost to the organism. Firstly, it appears that chromatin and DNA methylation has impacted the very DNA sequence that they are trying to package. Recent nucleosomal modelling has shown that the entire genome has evolved its sequence to allow the positioning of nucleosomes, due to the manner of tightly wrapping the DNA around the histone octamer leading to sequence constraints (Segal et al., 2006). This is paralleled with DNA methylation altering the DNA sequence through deamination leading to a deficit of CpG sites in the genome. In sea squirt this has occurred to such an extent that it is possible to accurately predict the methylation status solely by looking at the CpG density (Suzuki et al., 2007). How these chromatin and DNA methylation dependent changes in DNA sequence have affected the genes is still being elucidated. Studies have shown that exons exhibit codon bias against sequences that results in the formation of overly stable chromatin, as this would impede transcription (Cohanim and Haran, 2009), whether this effect has been harnessed for the regulation of gene expression is unknown. The effect of DNA methylation on codon bias is currently unknown. From the suggestions of gene body methylation as a mechanism of preventing spurious initiation (Simmen et al., 1999; Suzuki et al., 2007; Zilberman et al., 2007), it may be hypothesised that exons from these methylated genes may preferentially include CpG-containing codons. Chromatin and DNA methylation have both impacted all levels of genomic metabolism. Through CGIs and H3K4me3 domains, promoters become conspicuous within the genome and may act to attract the transcriptional machinery by highlighting key regions of the genome. There are suggestions that gene body chromatin and methylation have reduced the efficiency of transcriptional elongation (Belotserkovskaya et al., 2003; Rountree and Selker, 1997). However, it has

nevertheless been tolerated, perhaps through a need to reduce spurious transcription in wake of transiting RNA PolII (Carrozza et al., 2005; Suzuki et al., 2007; Zilberman et al., 2007). Limited analysis of recent high throughput sequencing of bisulphite treated human DNA suggested that highly transcribed genes show increased gene body methylation (Laurent et al., 2010), potentially providing evidence that this suppression of spurious intragenic transcription also exists in mammals. Indeed, this slowing of transcriptional elongation within the methylated gene body may have resulted in genes having to become longer in plants to avoid stalling of the polymerase early in the transcription cycle at promoter proximal regions (Zilberman et al., 2007). This slowing of elongation may have been harnessed to increase the fidelity of exon definition, with a marked increase in the DNA methylation density being observed at the 5' splice site (Laurent et al., 2010). Splicing also seems to be effected by the surrounding chromatin environment, with specific histone modifications associated with splice sites (Sims et al., 2007). The role of these systems within the bulk genome seems to be create a transcriptionally inert environment, with suggestions that this may be a noise reduction system (Bird, 1995; Boyes and Bird, 1992; Laybourn and Kadonaga, 1991). Overall, it appears that both chromatin and DNA methylation have become highly entrenched within the cell affecting how the genome is utilised at almost every level.

1.7 Interpreting DNA methylation

1.7.1 Identification of methyl-CpG binding proteins

The previous discussion has highlighted the key roles of mammalian DNA methylation. However, the mechanism by which DNA methylation is interpreted to exert these functions is still somewhat enigmatic. The increased level and dependence of DNA methylation in vertebrates may have necessitated an increased ability of the cell to interpret this signal (Hendrich and Tweedie, 2003). There are three potential mechanisms by which DNA methylation could be read by the cell. First, the meCpG moiety could intrinsically alter the chromatin or DNA structure and thereby alter its metabolism and accessibility. Currently, there is no evidence for this as a potential mechanism, with early experiments involving microinjection of the herpes simplex thymidine kinase gene showing that initially the methylated form was equally as active as the non-methylated form, with the methylated copy only becoming silenced after 100 hours, suggesting an indirect mechanism of methylation causing repression (Buschhausen et al., 1985). Second, DNA methylation of a specific site

could actively prevent the binding of transcription factors and this has been shown for several proteins including CTCF (see section 1.4.5; (Arney, 2003)), Cfp1 (see Appendix B; (Thomson et al., 2010)), Sp1 (Clark et al., 1997; Mancini et al., 1999), CREB (Mancini et al., 1999). A recent study using high-throughput bisulphite sequencing demonstrated the use of CpG methylation as a mechanism for controlling transcription factor accessibility during development for p300, which binds to enhancer regions outside of CGIs (Lister et al., 2009). Third and the focus of this discussion is that meCpG could directly recruit factors which in turn interpret the DNA methylation signal. One of the first studies suggesting the existence of meCpG binding factors came from restriction digest of mammalian nuclei. Digestion of naked DNA with *Msp*1 (cut site: CCGG) resulted in extensive cleavage, however, digestion of intact nuclei only resulted in cleavage at CGIs, suggested that the meCpG sites were somehow specifically protected (Antequera et al., 1989). Two meCpG binding activities were subsequently observed. An activity termed MeCP1 (methyl-CpG binding protein 1) was initially identified by bandshift experiments of a methylated probe using nuclear extracts (Meehan et al., 1989). The protein responsible for this activity was not indentified until later as MBD2 (see section 1.6.6;(Feng and Zhang, 2001)). MeCP2 (methyl-CpG binding protein 2) was subsequently shown to possess meCpG binding activity by south-western assay and to be distinct from MeCP1 by virtue of different chromatographic properties (Lewis et al., 1992). Purification of this protein and sequencing by Edman degradation allowed identification of the MeCP2 gene (Lewis et al., 1992).

1.7.2 The MBD protein family

Deletion analysis of MeCP2 in combination with a south western assay identified a sequence 85 amino acids long as the minimal fragment required for methyl-specific DNA binding and was termed the methyl-CpG binding domain (MBD) (Nan et al., 1993). The sequence of the MBD was used to search cDNA databases and led to the identification of PCM1 (protein containing methyl-CpG binding domain 1) (Cross et al., 1997). Further bioinformatic based approaches led to the identification of three more mammalian MBD-containing proteins named MBD2, MBD3 and MBD4 (Hendrich and Bird, 1998). These MBD-containing proteins were shown *in vitro* to specifically bind a methylated probe with the exception of MBD3, which was later shown to contain amino acid substitutions that prevent binding to methyl-CpG (Hendrich and Bird, 1998). In line with the repressive nature of DNA methylation, all of the MBD proteins have been shown to repress transcription, but mouse

genetic studies have indicated that they have non-overlapping roles (Martin Caballero et al., 2009). A schematic of the MBD protein family is shown in figure 1.3.

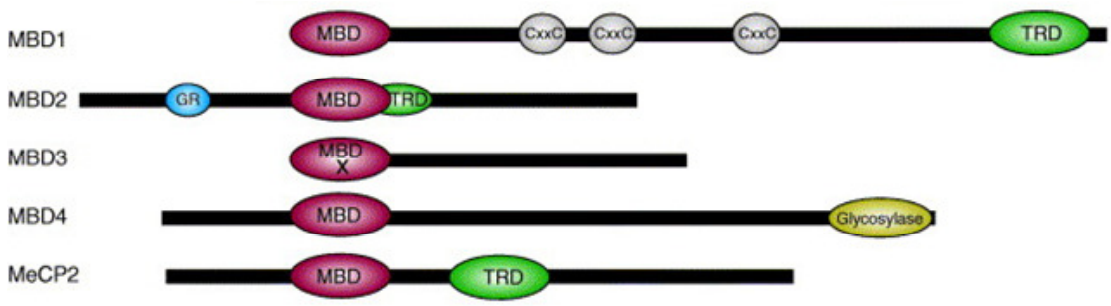

Figure 1.3 – Schematic representation of the MBD protein family, all of which bind meCpG through the methyl binding domain (MBD), with the exception of MBD3. In addition, MBD1 has three zinc finger domains (CXXC), one of which specifically binds non-methylated DNA. The transcriptional repression domain (TRD) interacts with various co-repressors. The GR repeats of MBD2 have been reported to interact with RNA (Jeffery and Nakielny, 2004). The C-terminal glycosylase domain of MBD4 functions in DNA repair. Adapted from (Klose and Bird, 2006).

1.7.3 The structure of the MBD domain

Solution structures of unliganded MBDs of MBD1 (Ohki et al., 1999) and MeCP2 (Wakefield et al., 1999) have been determined. The structure of the MBD domain of MeCP2 indicated the presence of numerous arginine and lysine side chains on the DNA binding surface, suggested to provide non-sequence specific binding to aid association with a methyl-CpG moiety irrespective of sequence context (Wakefield et al., 1999). These structural studies suggested that specificity for methylated DNA was due to a conserved hydrophobic patch on the MBD interacting directly with the methyl group (Ohki et al., 2001; Wakefield et al., 1999). However, a more recent high resolution structure of the MBD domain of MeCP2 bound to methylated DNA indicated that methyl-specific binding was largely through a hydrophilic surface, which included co-ordinated water molecules in the major groove of the DNA (Figure 1.4; (Ho et al., 2008)).

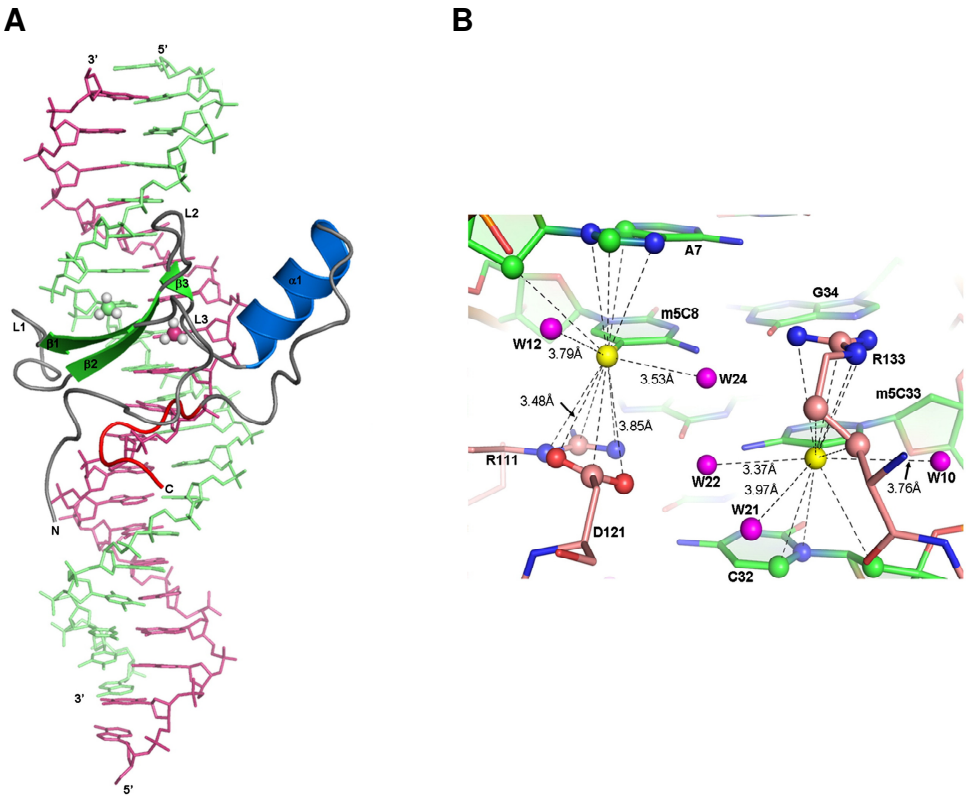

Figure 1.4 – Structure of the MBD domain of MeCP2 bound to a methylated DNA ligand. A polypeptide corresponding to amino acids 77-167 of the human MeCP2 MBD domain was expressed and co-crystallised with a 20 bp DNA fragment from a promoter of the mouse brain derived neurotrophic factor gene. (A) The overall structure is composed of a four stranded anti-parallel beta sheet linked to a single short alpha helix, which is similar to the unliganded structure as determined by NMR (Wakefield et al., 1999). The methyl groups of the meCpG are shown as spheres (B) The methyl group of DNA (shown in yellow) is contacted by co-ordinated water molecules (shown in purple) and hydrophilic interactions with the protein. Adapted from (Ho et al., 2008).

1.7.4 The ancestral MBD protein

Bioinformatic analysis has successfully mapped the evolution of MBD containing proteins with the aid of available genome sequence information (Hendrich and Tweedie, 2003). MBD2 and MBD3 are likely the ancestral MBD proteins as they are the only vertebrate for which homologues can be identified in invertebrate genomes (Hendrich and Tweedie, 2003). The MBD2/3 protein is encoded by a single gene in invertebrate genomes but diverges into

two similar proteins in vertebrate genomes (Hendrich and Tweedie, 2003). It has been suggested that the increase in DNA methylation at the invertebrate-vertebrate transition and the increased dependence on DNA methylation as a means of controlling the genome, possibly by acting to reduce transcriptional noise, would have required an increased number of proteins that could interpret this DNA mark (Hendrich and Tweedie, 2003). Accordingly, vertebrates have evolved a number of other MBD proteins: MeCP2, MBD1 and MeCP2 (Hendrich and Tweedie, 2003).

1.7.5 MeCP2

MeCP2 will be discussed in detail in section 1.7.

1.7.6 MBD1

The MBD domain of MBD1 can bind to symmetrically methylated DNA and more weakly to hemi-methylated DNA *in vitro* (Hendrich and Bird, 1998). MBD1 has been shown to bind to the densely methylated heterochromatic foci *in vivo*, but perhaps surprisingly, this localisation remains in methylation deficient cells (Hendrich and Bird, 1998; Jorgensen et al., 2004). As well as containing a functional MBD domain, MBD1 contains a transcriptional repressor domain (TRD) and three cysteine rich CXXC domains. Transfection studies suggested that the CXXC-3 domain selectively binds nonmethylated DNA *in vitro* and *in vivo* and may explain the methylation independent localisation (Jorgensen et al., 2004). The explanation for the protein having the ability to bind both methylated and nonmethylated DNA is currently not understood. However, accordingly MBD1 has been shown to be able to repress transcription form both methylated and nonmethylated DNA, with this repression dependent on the TRD (Jorgensen et al., 2004). The role and targets of MBD1 *in vivo* is currently unknown with the null mice exhibiting a mild neurological phenotype (Zhao et al., 2003). MBD1 was originally known as PCM1 (Protein containing methyl-CpG-binding domain 1) (Cross et al., 1997).

1.7.7 MBD2 and MBD3

Initially, the methyl binding activity of the MeCP1 complex was thought to be due to the MBD1 protein (Hendrich and Bird, 1998), but it has been later shown that MBD2 is the protein responsible and forms part of the MBD2/NURD chromatin remodelling and histone deacetylase complex (Feng and Zhang, 2001). Despite MBD2 and MBD3 being the most

closely related MBD proteins (~70% amino acid identity), MBD3 cannot specifically bind methylated DNA due to tyrosine to phenylalanine substitution (Fraga et al., 2003; Hendrich and Bird, 1998). The core mammalian NURD co-repressor complex is thought to contain: MBD2, MBD3, HDAC 1/2, MTA2, Mi-2, RbAP46/48 and p66/68; consistent with the repressive nature of DNA methylation. Genetic analysis indicated that MBD2 and MBD3 are not functionally redundant (Hendrich et al., 2001), which led to the biochemical purification of MBD2 and MBD3 specific NURD complexes (Le Guezennec et al., 2006).

The role of MBD2 and MBD3 in this complex is unclear. *Mbd2*-deficient mice are viable, fertile and present only a mild maternal nurturing phenotype (Hendrich et al., 2001). MBD2 has been shown to be required in the colon for the repression of pancreas and duodenum specific genes (Berger et al., 2007). A gene candidate based approach rationalised that genes with methylated CGIs were likely to be targets for MBD proteins. This led to the observation that MBD2 is involved in the repression of the methylated X*ist* promoter, with loss of MBD2 resulting in a low-level reactivation of *Xist* in an HDAC dependent manner (Barr et al., 2007). In contrast, loss of MBD1 and MeCP2 did not show the same effect, suggesting that the MBD proteins do not have overlapping functions (Barr et al., 2007). However, evidence of a generalised role of MBD2 in the repression of methylated promoters has not been forthcoming.

1.7.8 MBD4

The MBD of MBD4 is most similar to that of MeCP2 in primary sequence and was shown to bind to both symmetrically and hemi-methylated DNA *in vitro*, consistent with the observation that a GFP fusion protein was shown to be localised to the pericentromeric heterochromatin in a methylation dependent manner (Hendrich and Bird, 1998). Further work identified that MBD4 preferentially binds to m^5CpG x TpG mismatches, the primary product of deamination of methyl-CpG (Hendrich et al., 1999). MBD4 was shown to efficiently remove thymine or uracil from a mismatched CpG site *in vitro* (Hendrich et al., 1999). In humans, the mutation rate from 5-methylcytosine to thymine is 10-50-fold higher than other transitions, with the consequence that the methylated CpG sequence is under-represented in the genome. Over one-third of germline point mutations associated with human genetic disease and many somatic mutations leading to cancer involve loss of CpG (Cooper and Youssoufian, 1988). Genetic studies using *Mbd4*-/- mice found that the frequency of C > T transitions at CpG sites was increased by a factor of three (Millar et al., 2002). Furthermore, on the cancer susceptible *Apc*(Min/+) background, *Mbd4*-/- mice

showed accelerated tumour formation with CpG > TpG mutations in the wildtype *Apc* allele (Millar et al., 2002). Therefore, MBD4 suppresses CpG mutability and tumorigenesis *in vivo*. MBD4 has also been shown to act as a transcriptional repressor through the recruitment of HDAC complexes and be present at the hypermethylated promoters of the *p16(INK4a)* and *hMLH1* genes (Kondo et al., 2005). However, the siRNA knockdown of MBD4 did result in an upregulation of these genes, therefore the role of MBD4 in gene regulation at these sites and elsewhere is currently unknown.

1.7.9 Kaiso and Kaiso-like proteins

Aside from the MBD family there is a second group of proteins that specifically bind to methylated CpGs. The founder member of this group, Kaiso, has an N-terminal POZ domain and C-terminal zinc fingers; it was originally purified from K652 cells as an activity that could bind specifically to methylated DNA (Prokhortchouk et al., 2001). As well as having affinity for methylated DNA, the zinc fingers of Kaiso can bind to a consensus sequence that lacks CpG dinucleotides (Daniel et al., 2002). However, in reporter assays, transfected Kaiso can still repress transcription in a methylation dependent manner (Prokhortchouk et al., 2001). There is also evidence that Kaiso can function to silence endogenous genes by recruiting histone deacetylase activity to methylated promoters (Yoon et al., 2003). In *Xenopus*, Kaiso deficiency allows zygotic transcription to commence before the mid blastula transition with subsequent phenotypes including developmental arrest and apoptosis (Ruzov et al., 2004). However, *Kaiso*-null mice show no developmental abnormalities or changes in gene expression profiles (Prokhortchouk et al., 2006).

Two more proteins, ZBTB4 and ZBTB38, were identified on the basis of homology with Kaiso (Filion et al., 2006). These proteins were also shown to be able to bind to methylated CpGs via their zinc finger domains and to repress transcription in reporter assays. The localization of these proteins to pericentromeric heterochromatin is also consistent with their binding to methyl-CpG (Filion et al., 2006). Furthermore, ZBTB4 and ZBTB38, as well as Kaiso, were shown by chromatin immunoprecipitation to be present exclusively at the methylated paternal allele of the *H19/Igf2* differentially methylated region (Filion et al., 2006). However, no causal role in imprinting was demonstrated.

1.8 MeCP2

1.8.1 The structure of the *Mecp2* gene

The MeCP2 gene was initially identified from the results of Edman degradation probing a cDNA library (Lewis et al., 1992). The mouse gene was mapped to Xq28 and was shown to be subject to X inactivation (Quaderi et al., 1994). Examination of EST libraries identified alternative polyadenylation sites giving rise to multiple mRNAs with different 3' untranslated regions (UTRs): 2 kb and 10 kb, amongst others (Coy et al., 1999). The sequence of the 10 kb 3'UTR in mouse and human shows only 52% homology, but the distribution of free energy along the length of the 3'UTR is very similar (Coy et al., 1999). Coy *et al.* speculated that the 3'UTR of MeCP2 has key regulatory roles in the expression of MeCP2, with differential expression of the two isoforms between tissues and during development, with the 10 kb isoform prevalent in the mature brain. The exact role of the 3'UTR has, however, remained mysterious. Recent studies have increased interest, with the identification of a binding site within the long 3'UTR for a brain-enriched microRNA: miR132 (Klein et al., 2007). Using cultured cortical neurons, miR132 was shown to be induced upon neuronal depolarisation via the CREB pathway and resulted in a decrease in MeCP2 protein levels through post-transcriptional regulation, thereby allowing immediate response of MeCP2 protein levels to neuronal activity (Klein et al., 2007).

A bioinformatic search for conserved elements throughout the human *MECP2* gene and its neighboring regions identified four enhancer elements, two silencer elements and a conserved ~1 kb region immediately upstream of exon 1 with multiple regulatory elements (Liu and Francke, 2006). Some of these elements have recently been shown to exhibit DNase1 hypersensitivity and be marked by various histone modifications, implying a role in the regulation of the gene (Singh et al., 2008). Additionally, a study taking advantage of the recent explosion of high-throughput data from the ENCODE project, identified that the second intron in the human *MECP2* gene, which is atypically long at almost 60 kb, contains several conserved regions (Singh et al., 2008), suggesting that the regulation of this gene may be highly complicated.

1.8.2 The MeCP2 protein and expression pattern

Initially, the MeCP2 protein was purified from rat brain as a single 84 kDa protein (Lewis et al., 1992). Mapping of the *Mecp2* gene led to its annotation as a three-exon gene, with all

exons contributing to the protein (Coy et al., 1999). More detailed bioinformatic analysis revealed an additional upstream non-coding exon embedded in a CGI (Reichwald et al., 2000). MeCP2 was recently shown to be subject to alternative splicing giving rise to two isoforms (α and β), generating different N-termini with the α isoform containing a polyalanine and polyglycine tract (Figure 1.5A;(Kriaucionis and Bird, 2004)). MeCP2α was shown to be the significantly more abundant isoform in the mouse brain and alignment suggested that MeCP2α more closely resembled the ancestral form of MeCP2 (Kriaucionis and Bird, 2004). The functional differences between these proteins are currently unknown, with both isoforms showing localisation to methylated pericentromeric heterochromatin (Kriaucionis and Bird, 2004). Note, however, that in human brain, both isoforms are more equally expressed, with this study observing that the previously unidentified N-terminus containing the polyA/polyG tract is also present in the signaling protein ERK1, but there is no indication of a functional relationship (Mnatzakanian et al., 2004). Therefore it remains possible that these isoforms have distinct functions.

The distribution of MeCP2 may be key to understanding its role in cells. MeCP2 is ubiquitously expressed throughout human and mouse tissues (Su et al., 2002), but was immediately shown to be particularly abundant within the brain (Figure 1.5B; (Lewis et al., 1992)). The brain is broadly composed of two cell types: neuronal and glial cells. Immunofluorescence based experiments using mouse brain suggested that MeCP2 was primarily expressed within the neuronal nuclei but not in the glial nuclei (Kishi and Macklis, 2004). MeCP2 expression within development has also been examined with the finding that levels increased with neuronal maturation (Balmer et al., 2003; Kishi and Macklis, 2004; Shahbazian et al., 2002b). Despite this being somewhat reminiscent of the increase in DNA methylation during development, MeCP2 upregulation occurs much later.

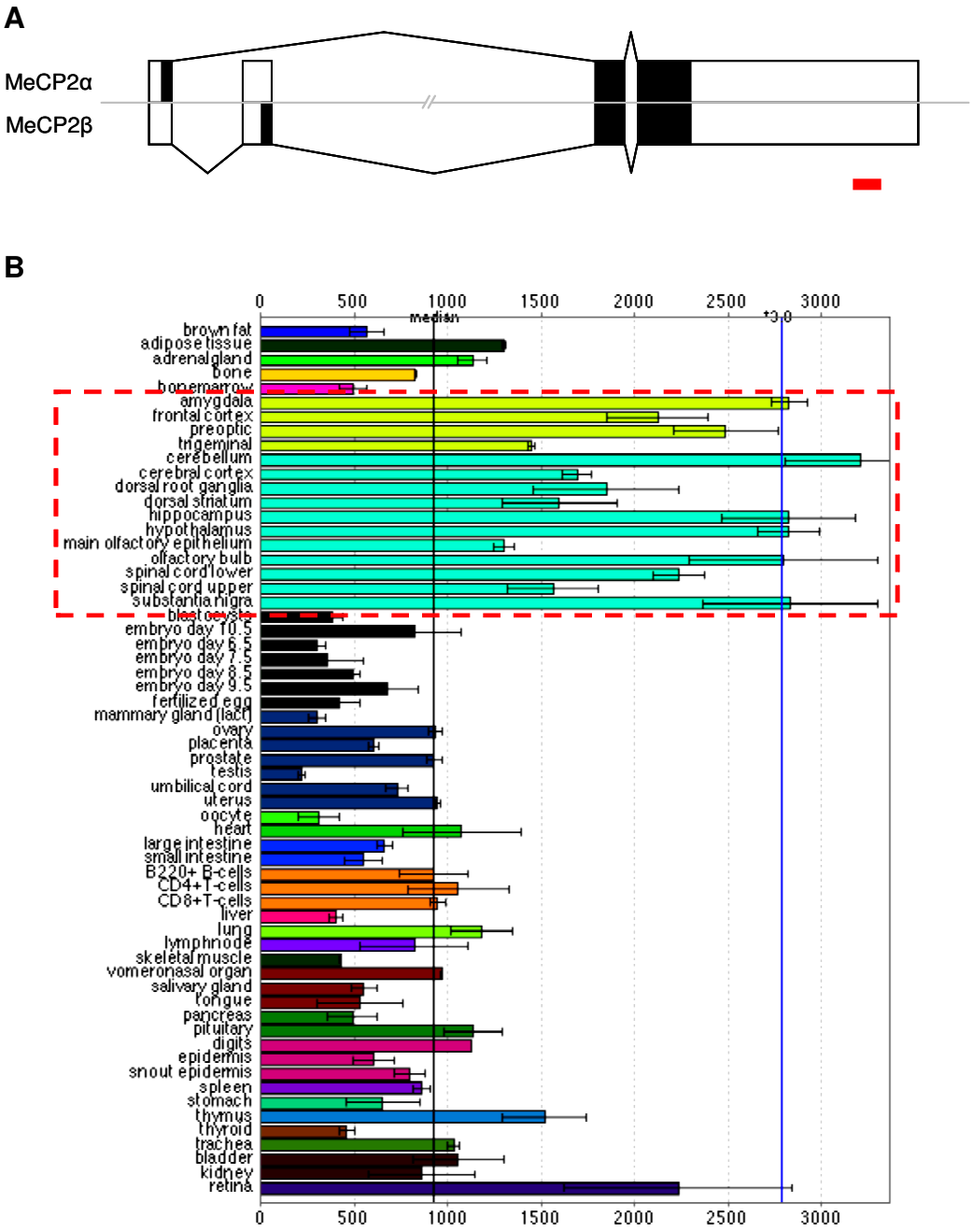

Figure 1.5 – *Mecp2* gene structure and expression pattern (A) Exon structure of MeCP2 indicating the alternative splice forms. Shaded boxes are protein coding and open boxes represent untranslated regions. The large second intron spans ~60 kb and ~43 kb in human and mouse respectively; the intron is cropped in this schematic. The β isoform is thought to be subject to translational interference due to the presence of a AUG codon within exon 1 and is therefore inefficiently translated (Kriaucionis and Bird, 2004). The red bar indicates the position of the probe used in the expression analysis shown in the panel below. (B) MeCP2 gene expression profiling across a wide range of mouse tissues, sourced from BioGPS (Su et al., 2002). Brain regions are highlighted within a red box. Expression values from Affymetrix chips relate to fluorescence intensity.

1.8.3 MeCP2 and Rett Syndrome

Rett syndrome is a relatively frequent form of mental retardation and occurs sporadically once every 10,000-22,000 female births (Kriaucionis and Bird, 2003). In Rett patients, apparently normal development gives way to regression after 6-12 months, with loss of acquired skills, including speech and mobility (Armstrong, 2002). Many patients survive into adulthood, but few procreate. Rare familial cases of Rett syndrome allowed the mapping of the disease region to Xq28, with mutations in the X-linked *MECP2* gene shown to be the primary cause of the disease (Amir et al., 1999). Studies have shown that after the initial crisis, there is no further neurodegeneration and therefore Rett syndrome is classed as a neurodevelopmental disorder. The neurological phenotype resulting from loss MeCP2 is in line with the highest expression of MeCP2 being observed in the brain.

Comprehensive databases of disease-causing *MECP2* mutations have been compiled. Most missense mutations that cause Rett syndrome are tightly clustered within the MBD domain, suggesting the key importance of this domain (Figure 1.6; see section 1.6.5 for discussion). Non-sense and frame shift mutations are distributed throughout the protein (for review see (Kriaucionis and Bird, 2003)). With almost some sense of cruel irony, approximately half of the missense mutations affect an arginine residue, which contains a CpG dinucleotide in its codon. It is likely that these gene body CpG sites are methylated and that the Rett syndrome causing mutation is the result of unrepaired deamination.

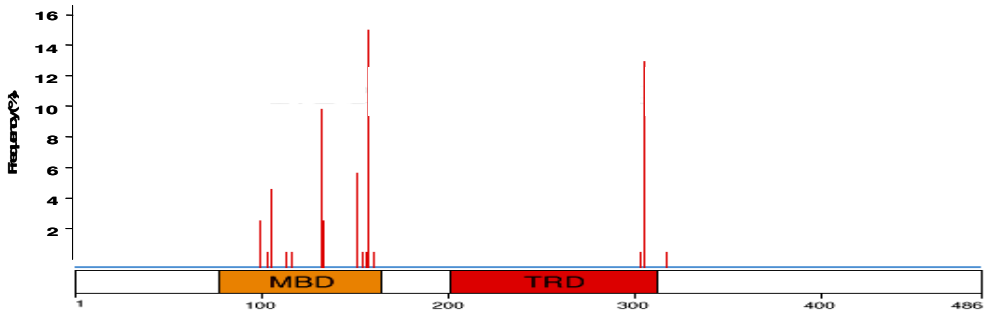

Figure 1.6 – Map indicating the position of the missense mutations that result in Rett syndrome (adapted from www.MeCP2.org.uk)

Rett syndrome is almost exclusively a disease of females, as the *MECP2* gene is X linked and patients are heterozygous for the mutated allele. Following random X-inactivation, the MeCP2 expression forms a mosaic pattern with typically half of the cells expressing the wildtype allele and half expressing the mutated form. Various studies have tried to correlate Rett symptoms with the causative mutation, with the hope of attributing specific functions of the protein to the various protein domains ((Temudo et al., 2010) and references therein). However, there have been no clear links. This may result from a number of reasons, including the possibility of skewed X inactivation, which could lead to some regions of the brain being over-represented in mutant cells and the overall proportion of mutant cells in the whole brain varying. Therefore this heterogeneity may prevent clear genotype-phenotype correlations.

1.8.4 Mouse models of Rett syndrome

Two groups independently produced *Mecp2* null animals (Guy et al., 2001) (Chen et al., 2001). The null mice showed no phenotype until 3-8 weeks of age, at which point they developed a stiff uncoordinated gait, hind limb clasping, irregular breathing and on average died at 54 days after birth (Guy et al., 2001) (Chen et al., 2001). In line with the human condition, the female heterozygote mice do not die prematurely, but display a similar neurological phenotype as the null males, but with onset of symptoms occurring between three and nine months (Guy et al., 2001) (Chen et al., 2001). The parallels between Rett patients and these mice, with the late onset neurological phenotype, suggest that this is a reasonable model for the study of Rett syndrome.

Brain specific knockout of MeCP2 produced mice indistinguishable from that of the null animals (Guy et al., 2001), suggesting that the key function of MeCP2 is within the brain, where it is most abundant. As discussed, immunofluorescence has shown that MeCP2 is expressed primarily from neurons and is absent from the glia (Kishi and Macklis, 2004). This is also mirrored in the mouse genetic studies where neuron specific expression of MeCP2 rescues the null phenotype (Luikenhuis et al., 2004). This suggests the major function of MeCP2 is in the brain and more specifically the neurons. However, this has recently been contested from cell line based experiments, which suggests that MeCP2 expression in the neurons alone is not sufficient for wildtype growth patterns, with the absence of MeCP2 in glial cells acting as a dominant negative to the normal dendritic

morphology of co-cultured neurons (Ballas et al., 2009). These assertions as to the pivotal role of MeCP2 within glia will likely be elucidated from the phonotype a glial-specific knockout mouse.

As well as absence of MeCP2 resulting in a severe neurological phenotype, surprisingly only a mild overexpression of the protein also results in a neurological symptoms. Overexpression of MeCP2 via the *Tau* promoter in mouse post-mitotic neurons led to neurological phenotype distinct from those caused by the loss of MeCP2 (Luikenhuis et al., 2004). This overexpression was only ~2-4 fold and did not result in an alteration in protein localisation as measured by immunofluorescence (Luikenhuis et al., 2004), suggesting that the levels had not reached a saturation point where localisation becomes uniform. This observation has been recapitulated with a transgenic mouse that also expresses the human form of MeCP2 from a large human genomic clone, resulting in two-fold overexpression of MeCP2 and an associated neurological phenotype (Collins et al., 2004). Symptoms associated with the overexpression of MeCP2 in mice have also been observed in humans, with *MECP2* gene duplications events giving rise to developmental delay and mental retardation in a manner distinct from Rett syndrome (Lubs et al., 1999). Overall, these mouse models suggest that too much MeCP2 is just as bad as too little.

1.8.5 MeCP2 as a methyl-binding activity

MeCP2 was originally identified form rat brain extract on the basis of methyl-DNA specific binding as determined by a south-western assay using a probe containing multiple methylated CpG sites (Lewis et al., 1992). Subsequent deletion analysis identified the amino acid sequence between 89 and 162 as the minimal region required for binding to methylated DNA: the MBD domain (Nan et al., 1993). Using a band shift assay with bacterially expressed recombinant proteins, the MBD was shown to interact with a single symmetrically methylated CpG site with a dissociation constant in the order of 1×10^{-9} M (Nan et al., 1993). The *in vivo* binding characteristics have been further analysed. In mouse cells, the predominant DNA sequence in the pericentromeric heterochromatin is the major satellite repeat, which accounts for ~7% of the total genomic DNA but ~40% of the 5-methylcytosine content of the cell. A *lacZ* fusion protein with full length rat MeCP2 was transiently expressed in mouse cells, with microscopy indicating a characteristic punctate pattern co-localising with the pericentromeric heterochromatin (Nan et al., 1996). The methyl-dependence of this binding was interrogated by transfection into *Dnmt1*-deficient cells,

which contain approximately 5% of the wildtype levels of DNA methylation. Here, the fusion MeCP2 protein did not largely display a punctate distribution, but instead was mostly diffuse throughout the nucleus (Nan et al., 1996). More recent work addressed the question of whether the different MBD proteins have different targets due to a requirement for additional DNA sequences at the CpG site (Klose et al., 2005). A methyl-SELEX (Systematic Evolution of Ligands by Exponential Enrichment) assay using the MBD of MeCP2 alone, indicated that MeCP2 binding was enhanced by the presence of a $[A/T\}_{\geq 4}$ run situated one to three base pairs or six to nine base pairs form the meCpG site, whereas the MBD of MBD2 showed no such requirement (Klose et al., 2005).

The validity of MeCP2 as a methyl-specific binding protein has recently been questioned in numerous studies; this will be discussed in section 1.8.8

1.8.6 Transcriptional repression and binding partners

Early experiments involving the introduction of methylated reporter constructs into cells suggested that DNA was a repressive mark (Stein et al., 1982). However, initial experiments did not show that MeCP2 repressed transcription in a methylation specific manner; this was thought to be due to impurities (Meehan et al., 1992). Using cleaner preparations of MeCP2, it was shown to repress *in vitro* in a methylation specific manner and deletion analysis identified the transcriptional repressor domain (TRD) (Nan et al., 1997). The nature of this repression had several interesting features. Firstly, a targeted Gal4-TRD fusion was capable of effectively repressing transcription of a reporter from a distance of at least ~2 kb *in vivo* (Nan et al., 1997). This may suggest that MeCP2 is more suited to setting up repressive domains, rather than targeting key sites or simply working through steric occlusion of promoter bound transcriptional activators. Secondly, the impact of methylation density on repression was determined by using various bacterial methyltransferases. The data clearly showed that repression was non-linear with the methylation density of the construct, but repression appeared suddenly at ~1 meCpG/100 bp, with no further increase in repression with increasing methylation (Nan et al., 1997). This coincides with the methylation density of the bulk mammalian genome with the observation that low density methylation of ~1 meCpG/100 bp is only sufficient to repress weak promoter activity (Laybourn and Kadonaga, 1991). Overall this may suggest that the repressive activity of MeCP2 has been tailored to this low density and may provide a link between genome-wide methylation and noise reduction.

Much work has focussed on which interacting factors are responsible for this methyl-dependent repression. By co-immunoprecipitation, the TRD was shown to interact with a co-repressor complex containing mSin3a and HDACs (Nan et al., 1998). Accordingly, transfection studies using a Gal4-targeted TRD indicated that repression could be partially alleviated using the HDAC inhibitor, trichostatin A (TSA), thereby confirming the role of histone deacetylation in MeCP2 dependent repression (Nan et al., 1998). This has been further corroborated by the observation of increased histone acetylation in mice expressing a truncated form of MeCP2 lacking the TRD (Shahbazian et al., 2002a). This study hypothesised that the increase in acetylation would result in the altered expression of specific genes in the brain (Shahbazian et al., 2002a).

The only partial relief of repression with TSA treatment was suggestive of other mechanisms of repression. MeCP2 was shown to associate with a H3K9 methyltransferase activity (Fuks et al., 2003). However the factor involved is unknown, with no clear evidence of a change in H3K9 methylation levels in the various null mouse models. Deletion analysis showed this activity associated with a region overlapping with the MBD of MeCP2 (Fuks et al., 2003).

Recently the N terminus of MeCP2 has been shown to associate with HP1 in an *in vitro* myogenesis system (Agarwal et al., 2007). This study showed localisation of HP1 to the pericentromeric heterochromatin, but only in MeCP2 positive nuclei and went on to show a physical association using tagged constructs in co-immunoprecipitation experiments. However, the co-immunoprecipitation of endogenous proteins only gave a very weak signal, questioning the validity of this interaction (Agarwal et al., 2007).

Using *in vitro* transcription assays, MeCP2 binding was shown to prevent the assembly of pre-initiation complexes due to interaction of the TRD with the basal transcription factor TFIIB (Kaludov and Wolffe, 2000). However, the *in vivo* significance of this interaction has not been investigated.

MeCP2 has also been shown to interact with DNMT1 through the TRD in a complex that is mutually exclusive with the complex formed between MeCP2 and HDAC1 (Kimura and Shiota, 2003). This study postulated that this interaction had a role in the faithful replication of DNA methylation marks, with MeCP2 recruiting DNMT1 to hemi-methylated sites missed during replication in a manner akin to UHRF1 (Jeltsch, 2008). However, earlier

bandshift experiments did not support the observation of MeCP2 binding to hemi-methylated DNA (Nan et al., 1993).

MeCP2 has been shown to interact with various other protein factors including; ATRX, c-Ski, PU.1, RNA splicing factors (Nan et al., 2007), (Kokura et al., 2001) (Suzuki et al., 2003) (Young et al., 2005). However, the functional nature of these interactions remains poorly understood. The identification of key interacting partners has likely been hampered by the observation that MeCP2 does not form stable complexes (Klose and Bird, 2004).

1.8.7 MeCP2 gene targets

MeCP2 was quickly identified as a transcriptional repressor and considering the global distribution of methyl-cytosine, it was therefore expected that MeCP2 would regulate the activity of a large number of genes. Indeed, misregulation of gene expression was identified in lymphoblast cell lines derived from Rett patients (Ballestar et al., 2005; Traynor et al., 2002). However, cell lines are known to have aberrant high levels of CGI methylation not seen in primary tissues (Antequera et al., 1990), suggesting that cell lines may not be the appropriate model system for identifying target genes. Numerous studies have utilised the *Mecp2*-null mouse to look for changes in gene expression using microarray based experiments, with these changes typically verified by quantitative PCR analysis (Ben-Shachar et al., 2009; Chahrour et al., 2008; Jordan et al., 2007; Nuber et al., 2005; Tudor et al., 2002). Despite several genes being identified as misregulated in the *Mecp2*-null mouse, there is no clear consensus on what are the 'Rett genes.' Furthermore, one study showed that the number of misregulated genes was 30% lower in the mice with exon 3 deletion (*Mecp2*$^{tm1.1Jae}$) than in mice with the larger deletion (*Mecp2*$^{tm1.1Bird}$) suggesting that the manner of the genetic disruption may have an effect, however, this also may be due to the difference in the genetic background between the lines (Jordan et al., 2007). Additionally, these studies may be complicated by the onset of Rett-like symptoms and therefore questioning what can be considered as a true target, as opposed to a secondary change.

Perhaps more success has resulted from using a candidate gene approach for genes known to show activity-dependent neuronal transcription. Brain derived neurotrophic factor (*Bdnf*) has been identified as a direct target for MeCP2 using cultured cortical neurons taken from embryonic animals (Chen et al., 2003b; Martinowich et al., 2003). *Bdnf* plays important roles in neuronal survival, development and plasticity (Aid et al., 2007). *Bdnf* expression is

controlled by at least 8 alternative promoters, which show activity dependent and cell-type specific expression. MeCP2 was shown to bind to the exon IV and modestly silence basal gene expression (Chen et al., 2003b; Martinowich et al., 2003). Upon membrane depolarisation, MeCP2 was shown to be phosphorylated and dissociated from the promoter, accordingly no effect on the expression in the excited state was detected. It was therefore expected that *Bdnf* expression would be up regulated in the MeCP2 null mouse, however, in whole brain lysates of symptomatic animals *Bdnf* mRNA levels were shown to be down at 70% of wild type (Chang et al., 2006). This apparent conflict was suggested to be as a result of different neuronal activities between wild type and null mice (Chang et al., 2006). It may also reflect differences between cultured embryonic neurons and the mature mice, as MeCP2 is known to show a developmental increase in expression (Balmer et al., 2003; Kishi and Macklis, 2004; Shahbazian et al., 2002b).

Horike *et al.* used ChIP-and-clone of 1 day old mice to identify MeCP2 target genes. They identified a putative imprinted gene cluster containing *Dlx5/6* as a MeCP2 target (Horike et al., 2005). They claimed that in wild type mice that *Dlx5/6* was monoallelically expressed, whereas in the *Mecp2*-nulls, biallelic expression was observed. However this work has more recently been refuted, showing that *Dlx5/6* is not imprinted and there was no significant difference in expression between wild type and null mice (Schule et al., 2007). Overall, currently there is no clear consensus as to which genes, if any, MeCP2 specifically regulates.

Work on *Xenopus laevis* embryos showed that MeCP2 interacted with the SMRT co-repressor complex and is involved in the regulation of expression of a neuronal repressor *xHairy2a* in the differentiating neuroectoderm (Stancheva et al., 2003). The MeCP2/SMRT complex was shown to dissociate from the promoter upon *Notch/Delta* signalling (Stancheva et al., 2003).

1.8.8 Current models for MeCP2 function

i Classical methyl-dependent transcriptional repressor

The identification of mutations MeCP2 as primary cause of Rett syndrome led to the expectation that the disorder was the result of aberrant upregulation of genes due to the loss of a methyl-dependent transcriptional repressor. However, the failure to identify clear targets, suggests that MeCP2 is more complicated than originally thought and that MeCP2

does not in fact act as a classical transcriptional repressor by binding to a discrete set of targets.

ii Transcriptional activator

Perhaps the two most surprising MeCP2-related papers in the last few years have suggested that MeCP2 predominantly activates a large number of genes (Ben-Shachar et al., 2009; Chahrour et al., 2008). The first study compared the expression profile of wildtype hypothalamus with that of the *Mecp2*-null hypothalamus and found ~3000 genes whose expression was altered using a modest cut-off of 1.2-fold (Chahrour et al., 2008). Surprisingly, ~85% of these genes were downregulated in the absence of MeCP2, suggesting that they were activated by MeCP2 (Chahrour et al., 2008). The strength of this study came from the use of the *MECP2*-Tg mouse, which over-expresses MeCP2 by two-fold (Collins et al., 2004) and showed upregulation of the majority of these same genes (Chahrour et al., 2008). A closer inspection of the microarray data has questioned the statistical significance of these reciprocal changes (personal communication with J. Guy). ChIP analysis was performed in order to implicate MeCP2 in direct regulation of these genes. The data was however not conclusive, in part due to very poor immunoprecipitation specificity, but moreover due to the limited analysis performed, with only a single primer pair used to investigate MeCP2 binding at the target genes and the lack of MeCP2 binding to non-target genes was not confirmed. Using co-immunoprecipitation experiments, the study went further to suggest that MeCP2 recruits the activator protein CREB to target sites (Chahrour et al., 2008). Further work will be required to test the significance of this interaction, which appears sub-stoichiometric. The methyl-dependence in binding to these sites and acting as a transcriptional activator was not studied in detail, but found that five out of the six genes examined had nonmethylated promoters (Chahrour et al., 2008).

This study was one of the first to use a distinct brain region (Chahrour et al., 2008), which may have gone some way to explain the inconsistency with previous studies that used whole brain and identified very few changes (Nuber et al., 2005; Tudor et al., 2002), suggesting brain region-specific differences in the function of MeCP2. However, a very similar study from the same laboratory indicated that MeCP2 largely activated the same set of genes in the cerebellum (Ben-Shachar et al., 2009). Therefore, suggesting that this role as an activator was not missed in previous studies due to specific brain regions having distinct target genes.

iii Control of alternative splicing

Through co-immunoprecipitation and mass spectrometry analysis, MeCP2 was shown to interact with Y-box binding protein (YB-1) (Young et al., 2005).This interaction was abolished on the treatment of nuclear extracts with RNaseA, suggesting that it was RNA-mediated binding. Comparison of splicing patterns between brains from wildtype and *Mecp2*$^{308/Y}$, which expresses a truncated form of MeCP2, suggested a modest number of genes showed altered splicing patterns (Young et al., 2005). However, the interaction between MeCP2 and YB-1 was not abolished on using MeCP2-R106W, a common Rett mutant form of MeCP2, suggesting that the loss of this interaction is not key in causing Rett syndrome (Young et al., 2005). Further work will be required to examine the role of MeCP2 in controlling splicing.

iv *Nonmethyl-dependent functions*

In contrast to the early studies identifying MeCP2 as a methyl binding activity, there has been a large number of recent studies suggesting that MeCP2 can bind to naked DNA and to chromatin irrespective of the methylation status (Ben-Shachar et al., 2009; Chahrour et al., 2008; Georgel et al., 2003; Nikitina et al., 2006; Yasui et al., 2007). These *in vitro* studies investigated the interaction between recombinant MeCP2 and DNA containing strong nucleosomal positioning sequences. MeCP2 was suggested to bind to nonmethylated nucleosomal arrays and cause extensive compaction of the arrays into a heterogeneous population of structures as visualised by shadowed electron microscopy (Georgel et al., 2003). The second study built upon these findings and used truncated forms of MeCP2 to suggest that the C-terminal domain was responsible for the methylation independent interaction, whilst the MBD maintained a partial methyl-specific interaction (Nikitina et al., 2006). Further electron microscopy suggested that MeCP2 was capable of modulating the structure of nucleosomal arrays by *in cis* compaction of an array and *in trans* association of multiple arrays (Nikitina et al., 2006). However, there have been some concerns that the low salt conditions used in the band shift assays may allow spurious interactions between a positively charged protein such as MeCP2 and DNA. There are no clear *in vivo* examples confirming this structural role for MeCP2 in compacting chromatin. The only indication coming from a limited chromatin conformation capture study suggesting that MeCP2 mediated chromatin looping at the *Dlx5/6* locus (Horike et al., 2005), with this work being subsequently refuted (Schule et al., 2007). Additional support for methylation independent effects has come from recent studies examining the regulation of specific loci, which have

implicated MeCP2 in transcriptional repression through the binding to nonmethylated DNA (Harikrishnan et al., 2010; Kernohan et al., 2010) and expression arrays indicating MeCP2 regulating nonmethylated genes (Chahrour et al., 2008). The validity of these ChIP-based studies has, however, been of question, with only limited analysis using a single primer pair being used to identify MeCP2 binding sites {Harikrishnan, 2010 #454}{Kernohan, 2010 #455}{Chahrour, 2008 #235}.

1.9 Aims of this thesis

There has been a recent flurry of interest in MeCP2, much of which has provided some controversy and intrigue into the role of MeCP2. This study started with a re-analysis of the phenotypes associated with the loss and overexpression of MeCP2. This has led to several key objectives to shed light on the function of MeCP2: (1) determination of the distribution and abundance of MeCP2 in the mature mouse brain; (2) the characterisation of the binding pattern of MeCP2 in the brain; (3) the functional consequences of MeCP2 binding within the genome.

Chapter 2 Materials and Methods

All methods were carried out at room temperature (RT) unless otherwise stated. Some standard techniques will not be outlined in detail.

2.1 Common solutions and reagents

All material and reagents were stored at room temperature (RT) unless otherwise stated

Coomassie blue stain:	40% (v/v) methanol, 10% (v/v) glacial acetic acid, 0.25% (w/v) Coomassie Brilliant Blue R-250. Filtered through a Whatmann number 1 filter.
Coomassie destain solution:	40% (v/v) methanol and 10% (v/v) glacial acetic acid
DNA sequencing buffer (2.5x):	20 mM Tris HCl (pH 8.0) and 5 mM $MgCl_2$
Orange G loading buffer (6x):	0.198% (w/v) orange G, 12% (w/v) Ficoll, 120mM EDTA (pH8.0), 4.2% (w/v) SDS
Phenol for DNA extraction	Phenol: Chloroform: Isoamyl Alcohol (25:24:1). Saturated with 10 mM Tris, (pH 8.0), 1 mM EDTA (Sigma-Aldrich). Stored at 4 °C
Phenol for RNA extraction	Phenol: Chloroform: Isoamyl Alcohol (125:24:1). Saturated with 2 M NaOAc (pH 4.0) (Fluka). Stored at 4 °C
Phosphate buffered saline (PBS):	140 mM NaCl, 3 mM KCl, 2 mM KH_2PO_4, 10 mM Na_2HPO_4
Protein loading buffer (2x):	125 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 20% (v/v) Glycerol, 300 mM β-Mercaptoethanol, 0.2% (w/v) bromophenol blue. Aliquoted and stored at -20 °C
Proteinase K stock solution:	20mg/ml proteinase K, 100mM EDTA pH 7.5, 2% (w/v) SDS. Stored at −20°C
RNaseA stock solution:	20mg/ml RNaseA. Stored at −20°C

Tris buffered saline (TBS):	50 mM Tris-HCl (pH 8.0), 150 mM NaCl
Tris-acetate EDTA (TAE) (1x):	40 mM Tris-acetate, 1 mM EDTA
Tris-borate EDTA (TBE) (1x):	45 mM Tris-borate, 1mM EDTA
Tris-EDTA (TE) buffer (1x):	10 mM Tris-HCl (pH 7.5), 1 mM EDTA
Tris-glycine SDS (1x):	25 mM Tris, 250 mM Glycine, 0.1% (w/v) SDS
Transfer buffer (1x):	25 mM Tris, 192 mM Glyine (supplemented with 20% (v/v) methanol when using PVDF membrane)
Tris-glycine buffer (1x):	50 mM Tris-base, 400 mM Glycine

2.2 List of antibodies

Antibody	Source; Cat no.	Application	Immunogen/Notes
Anti-MeCP2	Sigma-Aldrich; Mec-168	WB (1:1000)	Peptide from human MeCP2; amino acids 471-486
Anti-MeCP2	Upstate; 07-013	FACS	Peptide from mouse MeCP2; amino acids 465-478
Anti-MeCP2	rabbit polyclonal; 674	WB (1:1000); ChIP	Peptide from rat MeCP2; amino acids 1-392 (Nan et al., 1998)
Anti-H4	Abcam; ab7311	WB (1:500)	Proprietary peptide derived from human H4; amino acids 1- 100
Anti-H3	Abcam; ab1791	WB (1:10000)	Proprietary peptide derived from human H3; amino acids 1- 100
Anti-NeuN	Chemicon, MAB377	WB (1:500); FACS	*Unclear:* Cell nuclei from mouse brain
Anti-H3Ac	Millipore; 06-599	WB (1:5000); ChIP	amino acids 1-20 of *Tetrahymena* histone H3 (ARTKQTAR[K*]STGG [K*]APRKQLC) where K* is acetylated
Anti-Histone H1	Abcam; ab1938	WB (1:1000)	Calf thymus intact Histone H1 (MW 23 kDa) complexed to rRNA. No information was available regarding the primary structure of the immunogen, it is however, expected to recognise all H1 isoforms (personal communication with Abcam)
Infra-red anti-rabbit 680 IgGs	Licor; 32223	WB (1:7000)	
Infra-red anti-mouse 680 IgGs	Licor; 32212	WB (1:7000)	
Infra-red anti-goat 680 IgGs	Licor; 32214	WB (1:7000)	
Peroxidase anti-rabbit IgGs	GE Healthcare; NA934	WB (1:5000)	

Peroxidase anti-mouse IgGs	GE Healthcare; NA931	WB (1:5000)
Peroxidase anti-goat IgGs	Sigma-Aldrich; A-3415	WB (1:5000)
Alexa-Fluor 488 anti-rabbit	Invitrogen; A11008	FACS
Biotinylated anti-mouse	Invitrogen; B2763	FACS
Alexa-Fluor 647 streptavidin conjugate	Invitrogen; A11008	FACS

Table 2.1 – List of Antibodies

WB: Western blotting (typical dilutions are indicated)

FACS: Fluorescence activated cell sorting

ChIP: Chromatin immunoprecipitation

2.3 Mouse strains and sample preparation

2.3.1 Mouse strains

C57BL6 Mecp2$^{tm1.1Bird}$ mice were used with wildtype littermates as controls. Unless otherwise stated all mice were 6-8 weeks of age. Mice were typically euthanized by CO_2 asphyxiation, tissues dissected and either frozen in liquid nitrogen for ChIP or used directly for nuclei preparation. Assistance was kindly received from Dr Jim Selfridge and Dina De Sousa in the organisation of the mouse lines and in the setting up of timed matings for the harvesting of embryonic tissues.

2.3.2 Dissection of the striatum granulosum of the dentate gyrus

Assistance in the dissection was gratefully received from Dr S Cobb. Briefly, the whole brain was first removed and rinsed in 1 x PBS at 4 °C. Subsequently, 500 μm sections were cut and the slices corresponding to the hippocampus retained. Using a dissection microscope, the striatum granulosum of the dentate gyrus was conspicuous by virtue of the density of cells. This region was carefully dissected and stored on ice before continuing with further experiments. Typically, from the dissection of a single mouse brain in the order of 1 x 10^6 cells were recovered.

2.3.3 Isolation of nuclei from mouse tissues

Typically nuclei from up to 5 brains or 3 livers were prepared at one time. Freshly taken mouse tissues were placed in a 15 ml dounce (Braun) to which 9 ml (equivalent to approximately 5 volumes of mouse tissue) of ice-cold buffer A was added (10 mM Hepes (pH 7.9), 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 0.5 mM EGTA, 2 M sucrose, 10 % glycerol, 10 mM sodium butyrate and complete protease inhibitors (Roche)). Tissues were homogenised using a Potter S (Braun) motorized homogenizer (7strokes at 1100 rpm). The homogenate was layered onto a 3 ml cushion of buffer A and centrifuged in pre-chilled SW40 rotor (24000 rpm, 40 min, 3 °C) to recover the nuclei. For the analysis of nuclear RNA, the homogenate was supplemented with 200 U of RNasin (Promega).

After centrifugation, the 'fatty plug' was removed and the tube inverted to allow the buffer to drain. The pelleted nuclei were then resuspended gently in typically 1 ml of resuspension buffer (1 x PBS, 20% (v/v) Glycerol, 10 mM sodium butyrate and complete protease inhibitors (Roche)). Nuclei were observed by phase contrast microscopy to ensure they were of good quality. Serial dilutions of nuclei were counted using a haemocytometer to determine concentration. Nuclei were aliquoted as required and flash frozen in liquid nitrogen then stored at -80 °C. For the analysis of nuclear RNA, the resuspension buffer was supplemented with 40 U/ml of RNasin (Promega).

2.3.4 Fluorescence activated cell sorting (FACS)

In order to focus on neurons, a procedure was developed to sort neuronal from glial nuclei using FACS on the basis of expression of the neuronal nuclear marker (NeuN). Frozen total brain nuclei were thawed, pelleted (600 g, 5 min, 4 °C) and resuspended in 1 ml of blocking solution (1 x PBS supplemented with 1.5% (v/v) fetal calf serum (Hyclone), 10 mM sodium butyrate and complete protease inhibitors (Roche)) and incubated for 15 min at RT. At this stage 20 µl was removed and kept at 4 °C as an 'unsorted' population. NeuN antibody was added at 1:500 dilution and MeCP2 antibody (Upstate 07-013) was added at 1:200 dilution. Nuclei were incubated at 4 °C for 2 h rotating on a wheel. Nuclei were washed with 3 x 1 ml blocking solution for 5 min and then incubated with secondary antibodies at 1:1000 dilution for 1 h at 4 °C rotating on a wheel. Nuclei were washed as before and then used for FACS. Recovered nuclei were pelleted as before and then used for future experiments.

This protocol was also modified to include a crosslinking step. Before staining, the pelleted nuclei were resuspended in 1ml 1 x PBS supplemented with 1% formaldehyde and incubated at RT for 10 min. The formaldehyde was quenched by adding glycine to a final concentration of 125 mM. The nuclei were pelleted and resuspended in 1 ml blocking solution and the staining procedure continued as above.

2.3.5 Preparation of mouse samples for protein analysis

For western blot analysis of mouse samples either intact tissues or isolated nuclei were used. In the case of intact tissues, three mouse brains were dissected, pooled and then homogenised in 2 ml NE1 buffer (20 mM HEPES-KOH (pH 7.9), 10 mM KCl, 1 mM

$MgCl_2$, 0.1% (v/v) Triton-X-100, 1 mM DTT, 20% (v/v) Glycerol and complete protease inhibitors (Roche)) using a 5 ml dounce (Braun). The homogenate was added directly to protein loading buffer then mixed vigorously and boiled for 7 min. This was then used directly for gel electrophoresis.

Isolated nuclei in resuspension buffer were added directly to protein loading buffer then mixed vigorously and boiled for 3 min. This was then used for gel electrophoresis.

2.4 DNA manipulation and cloning

2.4.1 Genomic DNA extraction

Nuclei were prepared from the mammalian tissue (see section 2.3.3). Typically 10 x 10^6 nuclei were used per DNA preparation, resulting in approximately 60 µg of DNA. Nuclei were pelleted by centrifugation (330 g, 5 min, 4 °C) and then gently resuspended in a buffer containing 50 mM Hepes (pH 7.9), 10 mM EDTA, 5 mM EGTA, 0.5 µg/µl proteinase K, 0.5 µg/µl RNaseA. After resuspension, NaCl and SDS were added to a concentration of 300 mM and 1% respectively, in a total volume of 300 µl. The extraction was allowed to proceed for 6 h at 55 °C with occasional gentle agitation. To extract the proteins an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added and mixed by shaking. The upper aqueous layer was removed and DNA precipitated by the addition of 3 volumes of ethanol, with the addition of 20 µg of glycogen (Roche) as a carrier. The DNA pellet was subsequently washed in 70% (v/v) ethanol. The DNA pellet was air dried and resuspended in typically 100 µl TE buffer.

2.4.2 Measurement of DNA concentration

DNA solutions were measured at OD_{260nm} and OD_{280nm} using a Nanodrop-1000 spectrophotometer. An automated reading of DNA concentration was calculated using Beer's law (Concentration ng/ul = (Absorbance_OD_{260nm} x Extinction coefficient dsDNA 50 ng/µl/cm) / pathlength cm). DNA purity was determined using OD_{260nm}:OD_{280nm} ratio, with ≥1.8 indicating the absence of residual protein or phenol contaminants from the sample.

2.4.3 Restriction digestion

DNA digest was carried out as per manufacturer's instructions (New England Biolabs). Briefly, DNA was diluted in the appropriate buffer, with 100 µg/ml of bovine serum albumin (BSA) as required and digested with 6 units of enzyme per µg of DNA. Reactions were typically allowed to proceed for 1-2 h at 37 °C.

2.4.4 DNA electrophoresis

DNA was resolved by agarose gel electrophoresis using the Sub-cell system (Bio-Rad). 0.8-2% (w/v) agarose gels were used depending on the DNA fragment(s) size to be resolved. Agarose gels were prepared with TAE or TBE containing 0.5µg/ml ethidium bromide. DNA samples and an appropriate DNA size ladder (New England Biolabs/Fermentas) were prepared in orange G loading buffer and were loaded. Agarose gels were run at constant voltage (75-110V) in electrophoresis buffer and visualised under UV light. Where electrophoresed DNA was to be further manipulated, UV exposure was kept to a minimum to prevent damage.

2.4.5 Gel extraction of DNA

Gel extraction was used to purify a homogeneous population of DNA fragments for cloning. Fragments were resolved by agarose gel electrophoresis in 1 x TAE buffer, visualised under UV and precisely cut out using a scalpel. The gel slice was placed on dry ice for 20 min and then centrifuged (16000 g, 20 min, RT). The resulting liquid was subsequently used for further applications.

2.4.6 Standard polymerase chain reaction (PCR)

The conditions of the PCR were determined for each primer pair and also depended on the aim of the experiment. Typically, reactions were performed in a volume of 20 µl containing ~50 ng of genomic DNA as a template, 250 nM of both forward and reverse primer, 1 x Red Hot PCR buffer (Abgene), 400 nM dNTPs (Abgene) and 1 unit Red Hot Taq (Abgene). Exact composition of the PCR and cycling parameter depended on the primer pair and aim of the experiment. For example, in the testing of primer pairs for real-time PCR, 3 mM $MgCl_2$ was added and typical cycling conditions were as follows: an initial denaturation at 96 °C for 2 min followed by 35 cycles of (1) 96 °C for 30 s; (2) 60 °C for 30 s; (3) 72 °C for 30 s. In PCR amplification from bisulphite-treated DNA, conditions were determined empirically for

each primer pair, however, as a starting point these typical conditions were employed: 2.5 mM $MgCl_2$, 3% (v/v) DMSO and cycling parameters were as follows: an initial denaturation at 94°C for 2min followed by 40 cycles of (1) 94°C for 50 s; (2) annealing temperature for 50 s; (3) 72°C for 50 and then an additional 72°C extension phase for 5 min.

Depending on the purpose of the experiment, some or all of the PCR was resolved by agarose gel electrophoresis and visualised under UV.

2.4.7 Real-time PCR

Each PCR was performed in a 20 µl reaction volume containing (1) 250 nM of both forward and reverse primers; (2) 5 µl of the template diluted in 0.1 x TE; (3) 10 µl of 2 x Quantace Sensimix Plus. Real time PCR was carried out using with a Roche Lightcycler according to manufacturer's instructions with the following cycling parameters: an initial denaturation at 96 °C for 10 min followed by 45 cycles of (1) 96 °C for 10 s; (2) 60 °C for 10 s; (3) 72°C for 15 s (with a single acquisition at 522 nm taken at the end of the extension phase); (4) melt curve analysis by increasing the temperature from 60 °C to 95 °C at a rate of 0.29 °C/s with continuous acquisitions).

Cycle threshold (Ct) values were determined by the second differential maximum method as calculated by the Roche Lightcycler software. An arbitrary measure of DNA quantity (Q) can be calculated using: $Q = 2^{-Ct}$.

2.4.8 Bisulphite sequencing of DNA

This technique was employed to determine the DNA methylation status in two main applications: (1) analysis of genomic DNA and (2) analysis of immunoprecipitated DNA. In the case of genomic DNA analysis, 2-5 µg of DNA was either digested with a restriction enzyme that cleaved outside the region of interest (typically EcoR1) and then subsequently ethanol precipitated. DNA was resuspended in 25µl TE buffer. Otherwise the genomic DNA was lightly sonicated in using a Diagenode Bioruptor (10 s, maximum power) and 2-5 µg of DNA in 25 µl TE buffer was used in the bisulphite treatment.

In the analysis of immunoprecipitated DNA, the DNA was fragmented during the chromatin immunoprecipitation protocol. Typically, 1-2 µg of input DNA and an entire immunoprecipitation was used in a volume of 25 µl 0.1 x TE buffer for bisulphite treatment.

The DNA and denatured at 100 °C for 5 min and then snap cooled on ice. To aid denaturation, 2.5 µl of freshly prepared 3 M NaOH was added to the DNA and incubated at 37 °C for 20 min. Bisulfite modification solution was freshly prepared as follows: 3.8 g sodium hydrogen sulphite ($NaHSO_3$) was mixed gently in 5 ml H_2O and 1.5 ml 2 M NaOH (protected from light). 110 mg hydroquinone was dissolved in 1 ml H_2O at 55 °C for ten minutes. The sodium bisulphite and the hydroquinone solutions were then mixed. 270µl of bisulfite modification solution was added to the denatured DNA and then overlayed with 200µl of mineral oil. The bisulphite treatment was incubated at 55 °C for 5 h in the dark prior to isopropanol precipitation supplemented with 50 µg glycogen (Roche) as carrier. DNA was resuspended in 25 µl and desulphonated by the addition of 2.5 µl of freshly prepared 3 M NaOH and incubation at 37 °C for 15 min. The DNA was ethanol precipitated in the presence of ammonium acetate and resuspended in 30µl of TE.

Typically 2µl bisulfite-treated DNA was used in a standard Red Hot PCR reaction. Bisulphite-specific primers were designed with the aid of Methprimer (www.urogene.org/methprimer/index1.html).

PCRs were resolved by agarose gel electrophoresis (TAE), gel extracted and PCR cloned. The cloning reaction was transformed into StrataClone Solo Pack Competent Cells (Stratagene) according to the manufacturer's instructions. Transformed cells were spread onto blue/white selection ampicillin LB agar plates.

Colonies harbouring plasmids containing inserts were identified by blue/white screening (insertion into the cloning site disrupts the α-fragment of β-galactosidase). Inserts were PCR amplified from single colonies by transferring some cells with a pipette tip into a 20 µl Red Hot PCR reaction containing M13-phage primers flanking the vector insert site. 5µl of each PCR was resolved by agarose gel electrophoresis to confirm that cloned fragments were of the correct size. The remaining 15 µl of the PCRs were incubated with 5 units of Exonuclease I and 5 units of Antartic Phosphatase (New England Biolabs) for 15 min at 37 °C followed by heat inactivation of the enzymes by incubation for 15 min at 80 °C.

Sequencing reactions contained 3.5 ul of the treated PCR product, 4 µl 2.5 x DNA sequencing buffer, 5 pmol M13 sequencing primer (forward) and 2 µl BigDye Terminator v3.1. The reactions were incubated under the following conditions: initial DNA denaturation at 96°C for 10secs, then 24 cycles (1) 96 °C for 30 s; (2) 50 °C for 20 s; (3) 60°C for 4min. Sequencing reactions were cleaned up and were run on an ABI 3730 capillary sequencer by the School of Biological Sciences Sequencing Service. Sequence data was quality controlled and analysed using Chromas Lite and Lasergene software respectively.

Bisulfite genomic sequencing data was analysed using the BiQ_Analyzer software package (Bock et al., 2005).

2.5 Protein manipulation

2.5.1 SDS-Polyacrylamide gel electrophoresis (PAGE)

Samples were diluted in protein loading buffer and boiled for 3 min at 100 °C. 0.75 mm thick gels were assembled in a Bio-rad Mini Protean-3 apparatus according to the manufacturer's instructions. Stacking gels (pH 6.8) contained 4.3% (w/v) acylamide:bis-acrylamide (29:1; Bio-rad) and separating gels (pH 8.8) contained 7-20% (w/v) acylamide:bis-acrylamide (29:1; Bio-rad) depending on the size of the proteins of interest. Both gels contained 0.1 (w/v) SDS and the acrylamide was polymerised with the addition of ammonium persulphate and TEMED as a catalyst. Samples were loaded along with Fermentas pre-stained markers. Electrophoresis was carried out at 120 V in Tris-glycine buffer.

2.5.2 Coomassie blue staining of proteins

For direct visualisation of proteins the SDS-PAGE gel was incubated with fresh Coomassie blue stain for 1 h with gentle mixing. Excess stain was removed with H_20 and then incubated in Coomassie destain solution with mixing until the protein bands were revealed. The gel was then rinsed in water. Quantification of protein bands was performed using the Licor Imaging system with Odyssey software.

2.5.3 Recombinant proteins as standards

Protein	Source; Cat no.	Notes
MeCP2	Gift from R. J. Klose	Full length untagged wildtype MeCP2 (NM_004992.3) was expressed and purified using a baculovirus expression system.
Histone H4	New England Biolabs; M2504S	

Table 2.2 – List of recombinant proteins used as standards

Recombinant MeCP2 and histone H4 was used to quantify the absolute abundance of these proteins in isolated nuclei. SDS-PAGE with coomassie staining was used to determine the concentration of recombinant MeCP2 by comparison to a known concentration of BSA (New England Biolabs).

2.5.4 Wet transfer of proteins to a membrane

Wet transfer was carried out using a Bio-rad Mini trans-blot cell according to the manufacturer's recommendations. Briefly, the transfer cassette was assembled in transfer buffer with 2 layers of 0.3 mm Whatman paper underneath the gel. A single layer of membrane (pre-soaked in transfer buffer) was placed on top of the gel followed by two further layers of Whatman paper. A constant current of 0.3 A for 1.5 h was used. After transfer the membrane was rinsed in H_20 and then subjected to Ponceau S staining (2% Ponceau S (w/v), 30% (w/v) trichloroacetic acid, 30% (w/v) sulfosalicyclic acid) and then washed in PBS.

Both nitrocellulose (0.4 µm pore size, Bio-rad) and PVDF (0.2 µm pore size, Bio-rad) membranes were used. Where PVDF was used, the membrane was first wetted in methanol and the transfer buffer was supplemented with 20% (v/v) methanol.

2.5.5 Western blotting

For most antibodies the following western blot procedure was used. After a 1 hour block in a 5 % milk (w/v), 1 x TBS solution at RT, fresh blocking solution containing a dilution of the primary antibody was applied. The primary antibody was incubated overnight at 4 °C followed by four consecutive 15 min washes in a PBST solution (1 x PBS, 0.1% (v/v) Tween-20). The membrane was then incubated in the blocking solution for 20 min after which the secondary antibody (either HRP-conjugated or IR-dye conjugated) was applied in

a 5 % milk, 1 x TBS solution for one hour at RT. Four consecutive 15 min washes in a PBST solution were followed by four quick rinses in 1 x PBS. For HRP-conjugated secondary antibodies the membrane was then placed in a freshly prepared equal volume mixture of ECL solution 1 (2.5 mM luminol, 0.396 mM p-coumeric acid, and 100 mM Tris-HCl pH 8.5) and solution 2 (5.6 mM H_2O_2, 100 mM Tris-HCl pH 8.5) for one minute. The blot was wrapped in a single layer of saran wrap and exposed to ECL Hyper-film (Amersham Biosciences). For IR-dye conjugated secondary antibodies, the bands were visualised using the Licor Odyssey imaging system. Quantification was performed using NIH ImageJ or Licor Odyssey software. The Kolmogorov-Smirnov (KS) test was used to test for statistical significance between samples, with the statistical software R (version 2.9.1) being used for the analysis. The KS test is a non-parametric test of equality of one-dimensional probability distributions used to compare two samples. Assistance in statistical analysis was provided from Dr A. Kerr.

For typical dilutions of antibodies used in western blotting see section 2.2

2.6 Analysis of chromatin

2.6.1 Standard chromatin immunoprecipitation (ChIP)

Throughout the protocol buffers were supplemented with complete protease inhibitors (Roche Applied Science) and 10 mM sodium butyrate. Frozen mouse tissues were thawed quickly in 2.5 ml 1 x PBS and homogenised using a 5 ml dounce (Braun; 5 strokes). Homogenate was pelleted (1000 g, 3 min, 4 °C) and then crosslinked in 10 ml PBS containing 1% formaldehyde for 10 min at room temperature. Crosslinking was quenched by adding glycine to a final concentration of 125 mM. The homogenate was pelleted as before and then washed in ice-cold PBS. The homogenate was then pelleted and resuspended in 1.4 ml lysis buffer (50 mM Tris-HCl (pH 8.1), 1% (w/v) SDS, 10 mM EDTA) and incubated on ice for 10 min. For each immunoprecipitation, 140 µl (~10 x 10^6 cells) was taken, and 1260 µl dilution buffer added (20 mM Tris-HCl (pH 8.1), 150 mM NaCl, 2 mM EDTA, 1% (v/v) Triton X-100, 3 mM $CaCl_2$, 10 mM $MgCl_2$). Chromatin was incubated with 3 U *Micrococcal* nuclease (Fermentas) at room temperature for 17 min and stopped by adding EDTA and EGTA to a final concentration of 10 mM and 20 mM respectively. Chromatin was further fragmented by sonication for 3 min (Branson digital sonifier). Overall, chromatin was fragmented with an average size of 500 bp. Precipitated debris was pelleted by

centrifugation (16000 g, 10 min, 4 °C) and the chromatin pre-cleared with 50 µl protein A sepharose (GE Healthcare) that had been previously washed and blocked with 1 mg/ml tRNA and BSA.

The beads were pelleted gently and the supernatant was used for immunoprecipitations (1200 µl) and input (200 µl). Antibodies were added to the supernatant (7 µl MeCP2 674 antibody; 5 µl H3Ac Millipore) and rotated at 4 °C overnight. Precipitated debris was pelleted and the supernatant incubated with 25 µl blocked protein A sepharose for 1 h at 4 C. The beads were then washed for 4 min at room temperature using 1 ml of ice-cold buffers as follows: once in buffer 1 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.1); four times in buffer 2 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl, 20 mM Tris-HCl pH 8.1); once in buffer 3 (250 mM LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl pH 8.1); and three times in TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA). Between washes the beads were pelleted by centrifugation for 1 min at 3000 g. Immunocomplexes were then eluted with 200 µl of extraction buffer (1% SDS, 100 mM $NaHCO_3$), by incubation at 55 °C for 20 min and placed on a shaking platform for 20 min. The eluate was collected by passing the slurry through a QIAShredder column (Qiagen) and cross-links were reversed by adding 5 M NaCl to a final concentration of 300 mM, followed by incubation at 65 °C overnight.

DNA was extracted by the supplementing the extraction buffer with 50 mM Tris-HCl (pH 7.5), 8 mM EDTA, 40 µg proteinase K and 40 µg RNaseA. This was further incubated at 55 °C for 1 h. Proteins were removed by the phenol extraction and the DNA precipitated by the addition of 3 columns of ethanol with 20 µg of glycogen as a carrier. DNA pellets were washed in 70% ethanol. Isolated DNA from both immunoprecipitations and inputs were either resuspended in 300 µl 0.1 x TE for real time PCR analysis (5 µl / PCR) or 25 µl for bisulfite treatment.

Real time PCR was carried out as described in section 2.4.7. % IP / Input was calculated using the comparative Ct method ($2^{-Ct(IP)}$ / $2^{-Ct(Input)}$) and factoring in the 6-fold difference in volume of sheared chromatin supernatant used for input and immunoprecipitation and represented as a percentage. Real time amplicons had an average spacing of 1 kb.

2.6.2 Salt wash ChIP protocol

A modified version of the standard ChIP protocol where isolated nuclei were incubated in a salt solution prior to crosslinking was used in order to liberate bound proteins. An aliquot of approximately 60 x 10^6 isolated nuclei was and thawed on ice, split into four tubes and then pelleted by centrifugation (400 g, 5 min, 4 °C). The aliquots of pelleted nuclei were then resuspended in 500 µl extraction buffer (20 mM Hepes (pH 7.9), 0.1 mM EDTA, 1 mM DTT and complete protease inhibitors (Roche)) supplemented with either 150 mM NaCl or 300 mM NaCl. A 30 µl sample was taken and kept for protein analysis. The extraction was allowed to proceed for an various durations at RT. At the end of the allotted time, 500 µl of crosslinking buffer was added (1 x PBS, 2% formaldehyde) and incubated for 10 min at RT. Glycine was added to a concentration of 125 mM to quench the formaldehyde. Nuclei were pelleted by centrifugation (700 g, 5 min, 4 °C) and then resuspended in 470 µl 1 x PBS with complete protease inhibitors (Roche). At this point a second 30 µl sample was taken for protein analysis. The nuclei were pelleted again (700 g, 5 min, 4 °C) and resuspended in 140 µl ChIP lysis buffer and the protocol continued as before (see section 2.6.1).

2.6.3 *Micrococcal* nuclease laddering

The striatum granulosum of the dentate gyrus was dissected as previously described in section 2.3.2. Tissues were homogenised in 150 µl of ice-cold NE1 buffer (20 mM HEPES-KOH (pH 7.9), 10 mM KCl, 1 mM $MgCl_2$, 0.1% (v/v) Triton-X-100, 1 mM DTT, 20% (v/v) Glycerol and complete protease inhibitors (Roche)) using a polypropylene pestle (Sigma-Aldrich Z359947) directly within the eppendorf. Nuclei were isolated by centrifugation (500 g, 5 min, 4 °C) and resuspended in 500 µl MNase buffer (5 mM Tris-HCl (pH 7.5), 80 mM NaCl, 1 mM $CaCl_2$, 0.2 M sucrose and complete protease inhibitors (Roche)). Limited MNase treatment was performed by the addition of 8 units of MNase (Fermentas). A timecourse of digestion was performed by taking 40 µl of the reaction and stopping the reaction with the addition of 260 µl Stop buffer (11 mM Tris-HCl (pH7.5), 11 mM EDTA, 22 mM EGTA, 333 mM NaCl, 1.1% SDS (w/v), 0.5 µg/µl proteinase K, 0.5 µg/µl RNaseA). DNA was phenol extracted and ethanol precipitated. Isolated DNA was resuspended in 20 µl of water. DNA was treated with Antartic phosphatise (NEB) and then subsequently end-labelled with γ-^{32}P-ATP by the addition of T4 polynucleotide Kinase (NEB). Unincorporated ATP was removed by passing the sample through a G-50 column (GE Healthcare). DNA was resolved by agarose gel electrophoresis (1.3% (w/v) in Tris-glycine buffer) and visualised by exposure to a phosphor storage screen.

2.7 High-throughput sequencing

Assistance in high-throughput sequencing and bioinformatic analysis was gratefully received from Dr Rob Illingworth, Shaun Webb, Dr Alastair Kerr. Sequencing was performed in collaboration with the Wellcome Trust Sanger Institute, United Kingdom

2.7.1 Preparation of ChIP DNA for Solexa sequencing

For high-throughput sequencing of MeCP2 bound chromatin (MeCP2 ChIP-seq) six independent MeCP2 immunoprecipitations were performed from a single mouse brain then subsequently pooled. The DNA was end repaired by incubation at 20 °C for 30 min with 3 units of T4 DNA Polymerase (New England Biolabs (NEB)), 10 units of Polynucleotide Kinase (NEB), 2 units DNA Polymerase I Large (Klenow) fragment (NEB), 1x T4 DNA ligase reaction buffer (NEB) and 400 nM dNTPs. The enzymes were then heat inactivated at 75 °C for 20 min, after which the DNA was ethanol precipitated. A tail of 'A' bases was added to the 3' ends of the DNA by incubation with 5 units Klenow Fragment (3'-5' exo-; NEB), 200 nM dATP and 1x buffer 2 (NEB) at 37 °C for 30 min. The enzymes were heat inactivated and cleaned up as before. Illumina paired end adaptors were then ligated to the processed ChIP DNA by incubation with 300 units of T4 DNA ligase (NEB), 1x T4 DNA ligase buffer (NEB), 7.5% PEG-6000 and 2 pmol of annealed Illumina adaptors for 3 h at room temperature. Ligated DNA was purified using MinElute PCR columns (Qiagen) and eluted in 10 µl water. This protocol was gratefully developed by Dr Rob Illingworth.

2.7.2 Methyl-DNA specific chromatography

Dr Rob Illingworth gratefully donated data indicating the genome-wide distribution of meCpG in mouse cerebellum. Only a brief description of the protocol will be described here, for more detailed information see ((Illingworth et al., 2008) and Illingworth et al., manuscript in preparation).

75 µg of mouse cerebellum genomic DNA (C57Bl6) was sonicated to an average length of 400 bp (Branson digital sonifier). The fragmented DNA was end repaired and Illumina paired end adaptors were ligated. The ligated DNA was separated into two 35 µg aliquots and purified independently to enrich for sequences containing methylated CpGs as previously described (Illingworth et al., 2008). Briefly, the ligated DNA was fractionated based on its ability to bind to an immobilized methyl-CpG binding domain derived from

MeCP2. DNA tightly retained by the methyl-affinity matrix (NaCl >700 mM) was passed over the affinity column for a second time increase binding specificity. The volume of purified DNA was reduced to 200 µl using an Amicon Ultra 30 kDa centrifugal filter (30,000 NMWL columns; Millipore) and subsequently ethanol precipitated.

2.7.3 Library preparation and Illumina Solexa sequencing

Illumina Solexa sequencing was used as a technique to produce high-throughput sequence data for the DNA recovered from MeCP2-ChIP and methyl-DNA specific chromatography. Briefly, specific adaptors are ligated on to the ends of the recovered DNA. The ligated DNA is bound to complementary immobilised adaptors on a flow cell, after which amplification results in the generation of clusters of DNA molecules with the same sequence. These clusters are then sequenced.

Ligated DNA was amplified by 10-12 cycles of PCR with primers complementary to the adaptor sequences and Phusion 2x premix (Finnzymes). The DNA was purified using QIAquick PCR Purification columns (Qiagen) and library fragments of >200 bp (insert plus adaptor and PCR primer sequences) were isolated by gel extraction. The purified DNA was captured on an Illumina flow cell for cluster generation. Libraries were sequenced on the Genome Analyzer following the standard Illumina protocol to generate 37 bp reads using a Solexa sequencer. Single-end sequence reads were mapped to the mouse genome (NCBIm37) using MAQ (http://maq.sourceforge.net/). Reads with a mapping score greater or equal to 30 where retained. For MeCP2 ChIP-seq seven independent lanes of sequence were generated and the results from these were combined. For high-throughput sequencing of methylated sequences (MBD-seq) 2 lanes of sequence were generated and the results from these were combined (R. Illingworth, manuscript in preparation).

2.7.4 Analysis of high-throughput sequencing

Assistance in high-throughput sequencing and bioinformatic analysis was gratefully received from Dr Rob Illingworth, Shaun Webb, Dr Alastair Kerr. Sequencing was performed in collaboration with the Wellcome Trust Sanger Institute, United Kingdom. CpG density calculations were performed on the ensEMBL repeat masked version of the mouse NCBI build 37 genome. Sliding window analysis was employed to count observed CpGs, ignoring all windows containing greater than 50% masked repeats. The same sliding window

parameters have been used to produce a measure of MeCP2 hits per window across the mouse genome, according to mapped Solexa sequenced reads. Genome wide levels of methylation have been traced by further applying the sliding window to mouse cerebellum MBD-seq data. An inner join, on relative sets of data, pairs windows with identical co-ordinates to form a relationship between observed CpG, MeCP2 reads and MBD reads across the mappable genome. For visualisation genomic intervals and values have been translated in to wiggle files and displayed via the Integrated Genome Browser.

To determine the level of MeCP2 binding to methylated CpG islands the methylated CpG rich fraction of the mouse genome was defined using MBD-seq data for cerebellum as outlined above. Methylated CGIs were identified by intersecting a repeat masked NCBI CGI-strict dataset with regions of high MBD-seq enrichment (>4 reads per base that span at least 90 bp and with a maximum gap of 250 bp). Sliding window analysis was employed to count MeCP2 reads across 5 kb domains centred upon the midpoint of each methylated CGI (500 bp window with a 100 bp slide).

2.8 RNA manipulation and expression analysis

2.8.1 RNA extraction from whole tissues

Brains were dissected and snap frozen in liquid nitrogen then stored at -80 °C until use. Tri-Reagent (Sigma-Aldrich) was used according to the manufacturer's instructions. Briefly, a single frozen brain was homogenised in 5 ml Tri-Reagent using a Polytron (15 s, maximum power; Janke and Kunkel). The homogenate was spun centrifuged (15 min, 1800 g, RT). The upper fatty layer was removed and the remaining homogenate was incubated at RT for 5 min. 1 ml of chloroform was added, mixed vigourously and incubated for 10 min at RT. The aqueous phase (RNA) was resolved from the organic phase (DNA and proteins) by centrifugation as above. To precipitate the RNA, the aqueous phase was transferred to a tube containing 2.5 ml isopropanol, mixed and allowed to stand for 10min at RT. RNA was pelleted by centrifugation as above, washed with 5 ml 70% (v/v) ethanol, air dried and resuspended in 500 µl of nuclease free water (Ambion). 10 µg RNA was DNase1 treated using RNase-free DNase1 (RQ1 Promega) according to the manufacturer's recommendations with the addition of 40 units RNasin (Promega). RNA was phenol extracted, ethanol precipitated and resuspended in 60 µl Ambion nuclease free water.

2.8.2 RNA extraction from isolated nuclei

Typically 6 x 10^6 isolated nuclei were used per RNA preparation. Nuclei were lysed in extraction buffer (30 mM Tris-HCl (Ph 7.5), 300 mM NaCl, 10 mM EDTA, 1% SDS, 0.5 mg/ml proteinase K, 0.2 units/µl RNAsin). The extraction was allowed to continue for 40 min whilst being incubated at 37 °C, after which RNA was extracted using 500 µl acid phenol (pH) and precipitated using iso-propanol. The isolated RNA was resuspended in 100 µl nuclease free water (Ambion) supplemented with 40 units of RNasin (Promega). The RNA was DNase1 treated using the Ambion DNA-free system according to the manufacturer's recommendations.

2.8.3 RNA electrophoresis

In a crude assessment of RNA integrity, the purified RNA was resolved by gel electrophoresis to confirm the presence of sharp bands corresponding to the ribosomal RNAs. RNA was resolved by agarose gel electrophoresis using the Sub-cell system (Bio-Rad). In an attempt to reduce the presence of ribonucleases the apparatus was cleaned using RNase-Zap (Ambion). 1.8% (w/v) agarose gels were prepared with 1 x TBE containing 0.5µg/ml ethidium bromide. RNA samples and an appropriate size ladder (New England Biolabs/Fermentas) were prepared in orange G loading buffer and were loaded. Agarose gels were run at 80 V voltage in electrophoresis buffer and visualised under UV light

2.8.4 cDNA synthesis

Typcially 2 µg of the RNA was synthesized into cDNA in a total volume of 50 µl. Initially the purified RNA was incubated with 5 nM random hexamers (Roche) at 70 °C for 5 min. The RNA was then snap cooled on ice and 13.5 µl of RT-mix added (RT buffer (Promega), 1.25 mM dNTPs (Abgene), 40 units RNAsin (Promega)) and then incubated at RT for 5 min. Subsequently, 200 units of m-Moloney murine leukemia virus reverse transcriptase (Promega) were added. In addition a minus (-)RT reaction was set up in parallel containing all the components except for the reverse transcriptase to control for DNA contamination by PCR. The reactions were incubated as follows: (1) 25 °C, 10 min; (2) 37 °C, 60 min; (3) 70 °C, 10 min. The 50 µl reactions were then typically diluted 4-fold in H_20 and stored at -20 °C. Typcially 3µl of cDNA was used in each PCR in parallel with the respective (-)RT control.

Chapter 3 MeCP2 Abundance and Distribution

3.1 Introduction

To my mind, research into the function of MeCP2 has to be guided by the disease phenotypes associated with alteration in the expression of MeCP2. In humans, a large range of mutations in the *MECP2* gene are associated with the neurodevelopmental disorder Rett syndrome, which is generally considered to be the result of loss of function of MeCP2 (Amir et al., 1999). However, chromosomal rearrangements which result in the duplication of the region encoding the *MECP2* gene also result in a distinct neurological phenotype (Lubs et al., 1999). These observations have been closely mirrored in mouse models, where loss of *Mecp2* expression or indeed expression of a truncated form, result in a neurological phenotype resembling Rett syndrome (Chen et al., 2001; Guy et al., 2001; Shahbazian et al., 2002a). Additionally, only mild two-fold overexpression of MeCP2 results in a neurological condition (Collins et al., 2004; Luikenhuis et al., 2004). These experiments suggest that the precise expression level of MeCP2 is critical to its function. This is born out by recent studies suggesting that the regulation of the *MECP2* gene is under the control of a great number of factors (for review see (Singh et al., 2008). Additionally, the levels of MeCP2 protein appear to be also modulated through post-transcriptional control via microRNAs in response to neuronal activity (Klein et al., 2007).

From these observations, a greater understanding of the expression pattern of MeCP2 may provide insight into its function. The neurological phenotypes associated with the perturbation of both the human and mouse genes is consistent with the early observations that MeCP2 is particularly abundant in the brain (Lewis et al., 1992; Nan et al., 1997). This is again indicated by mouse genetic studies that revealed a brain-specific knockout of MeCP2 was equivalent to the null animals (Chen et al., 2001; Guy et al., 2001). The late onset of these phenotypes is paralleled with a developmental increase in MeCP2 expression in the brain (Balmer et al., 2003; Kishi and Macklis, 2004; Shahbazian et al., 2002b), suggesting that failure to express MeCP2 to the high levels later in development is key to the symptoms. Immunofluorescence experiments performed on mature mice suggested that MeCP2 was primarily expressed within the nucleus of neuronal cells and absent from glial nuclei (Kishi and Macklis, 2004). Accordingly, a mouse model expressing MeCP2 under the *Tau* promoter, which is primarily neuron-specific, rescued the null animals (Luikenhuis et al., 2004) and a glial specific knockout had only a mild phenotype (Ballas et al., 2009).

These observations highlight that the precise high abundant level of expression within mature neurons is key to the function of MeCP2, correspondingly, failure of this results in late-onset neurological symptoms. From this standpoint, a precise understanding the absolute abundance of MeCP2 in neurons of mature mice was required.

3.2 Timecourse of MeCP2 expression

Numerous studies have reported a developmental increase in MeCP2 expression in the mouse at both the RNA level (Shahbazian et al., 2002b) and the protein level as measured by immunofluorescence (Kishi and Macklis, 2004) and also in humans as measured by laser scanning ctyometry (Balmer et al., 2003). Despite these studies and the importance of MeCP2 expression in these associated late onset phenotypes, there is no clear quantitative data revealing a timecourse analysis of MeCP2 expression. Therefore, quantitative western blotting analysis was used to determine when during postnatal mouse development MeCP2 reaches its maximal expression (Figure 3.1). Licor Odyssey near-infrared imaging was used, whereby the secondary antibody is conjugated to a proprietary fluorophore. This reportedly allows more accurate quantitation over a wider range than chemiluminescence based westerns (for more information please see company website). The majority of western blots were initially performed using chemiluminescence based detection, ensuring that the scope of intensities measured was within a small range and extrapolation was not required. Similar results were obtained using both techniques.

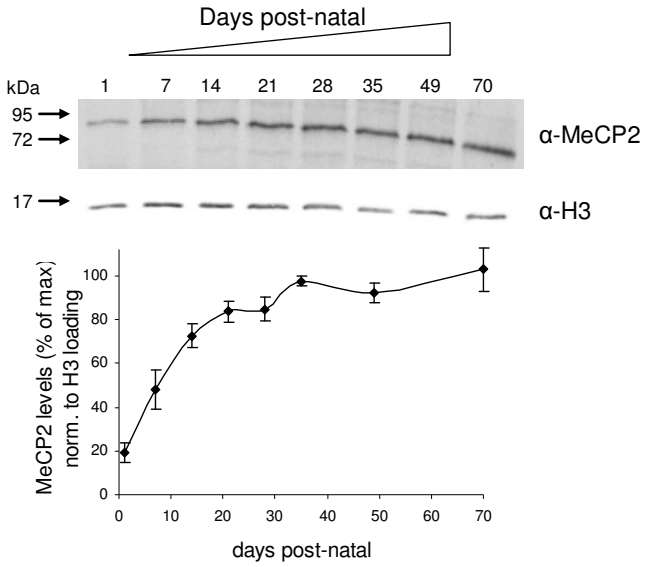

Figure 3.1 – MeCP2 expression in the mouse brain increases dramatically after birth, reaching maximal protein levels at ~5 weeks of age. For each timepoint, three mouse brains were pooled and homogenised. The homogenate was added directly to Laemmli buffer and boiled for 7 min and used directly for gel electrophoresis. Samples were separated by SDS-PAGE and quantitative western blotting was performed for MeCP2 (Sigma Aldrich, Mec-168) and histone H3 as a loading control using infra-red imaging (Licor Odyssey). Molecular weight markers (kDa) are arrowed. Westerns were quantified and the data represented as % of maximal MeCP2 expression, normalised for differences in loading based on H3 quantification; error bars indicate +/- standard error of the mean (SEM).

MeCP2 was weakly detectable in neonatal brain, but increased rapidly to reach a maximum at ~5 weeks of age, after which the total levels remained approximately constant. This is consistent with the mouse phenotypes associated with loss of MeCP2 expression displaying symptoms at 5-6 weeks, which ultimately leads to death between 6 and 12 weeks (Chen et al., 2001; Guy et al., 2001). From this analysis, an understanding of the function of MeCP2 would more likely be gained by the use of mature mice, therefore for all subsequent experiments 6- to 8-week-old mice were used unless otherwise indicated. The developmental expression pattern of other MBD proteins was not investigated.

3.3 FACS purification of neuronal nuclei

Immunofluorescence experiments have indicated that MeCP2 is expressed predominantly in the neuronal nuclei, with much less present in glial nuclei (Kishi and Macklis, 2004). This is consistent with a mouse genetic model suggesting that the primary function is within neurons (Luikenhuis et al., 2004). Therefore, in order to focus on neurons, a procedure was

developed to sort neurons from glia using fluorescence-activated cell sorting (FACS). It is imperative for FACS that a single particle suspension is produced. However, the ability to disaggregate all the cells from a mature mouse brain was questionable, as this is one of the reasons embryonic or neonatal mice are used for neuronal culture (Eide and McMurray, 2005). As a result, the use of purified nuclei, of which a single particle suspension is easily obtainable, was investigated. Despite the use of nuclei limiting the number of available markers, the vast majority of neuronal nuclei are fortunately positive for staining by the neuronal nuclear marker NeuN, whereas glial nuclei are invariably negative (Mullen et al., 1992). Due to the absence of a degenerate marker for glial nuclei, they can only identified by the absence of NeuN staining and therefore can not be formally identified as glia. As expected, NeuN staining of total brain nuclei gave a bimodal distribution comprising of 50% NeuN-positive nuclei (neurons) and 50% negative nuclei (predominantly glia) (Figure 3.2A). Staining of total brain nuclei for MeCP2 gave a similar bimodal distribution: 56% high and 44% low MeCP2 staining, with an approximately 6-fold difference in staining intensity (Figure 3.2B). Sorting of nuclei on the basis of the NeuN staining purified a neuronal population and a predominantly glial population of nuclei (Figure 3.2C). As expected from published immunofluorescence experiments (Kishi and Macklis, 2004), NeuN-positive nuclei co-sorted with the population of nuclei expressing high levels of MeCP2 (Figure 3.2C). This protocol was successfully used with both non-crosslinked and crosslinked nuclei (data not shown).

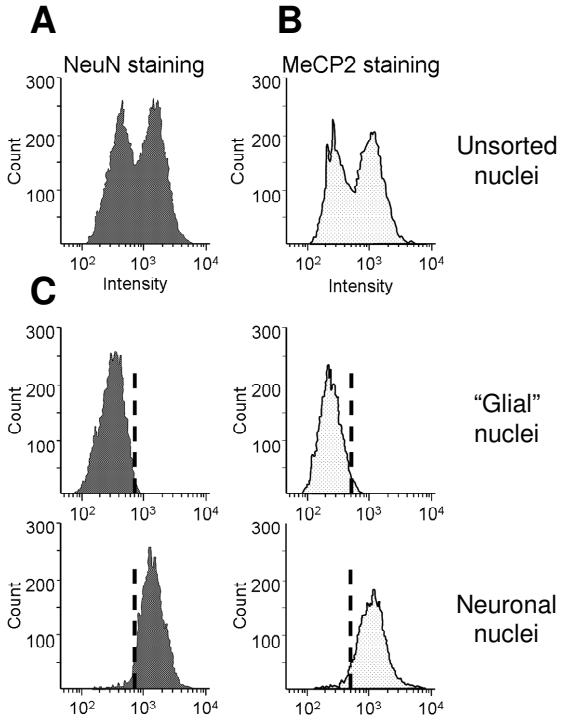

Figure 3.2 – FACS purification of isolated brain nuclei can be used to purify neuronal nuclei from glial nuclei on the basis of NeuN. Nuclei were isolated from wildtype whole mouse brain by homogenisation followed by centrifugation through a sucrose cushion. The single particle suspension of nuclei was FACS-stained for (A) NeuN and (B) MeCP2. C) Nuclei sorted for NeuN-negative and NeuN-positive staining co-sorted with low-MeCP2 and high-MeCP2 stained nuclei, respectively. The NeuN-negative nuclei are marked as "glial" nuclei, as no positive identification was used. Assistance in FACS was gratefully received from Andrew Sanderson and Martin Waterfall.

FACS analysis also allows an analysis of size, as measured by forward scatter, and internal complexity, as measured by side scatter. Isolated total brain nuclei displayed a diverse range of both size and internal complexity (Figure 3.3). Sorting of nuclei on the basis of NeuN staining identified that whilst the neuronal nuclei maintained this heterogeneous range in nuclear morphology they were on average larger and more complex than the predominantly glial population which exhibited a homogeneous morphology with small, less complex nuclei.

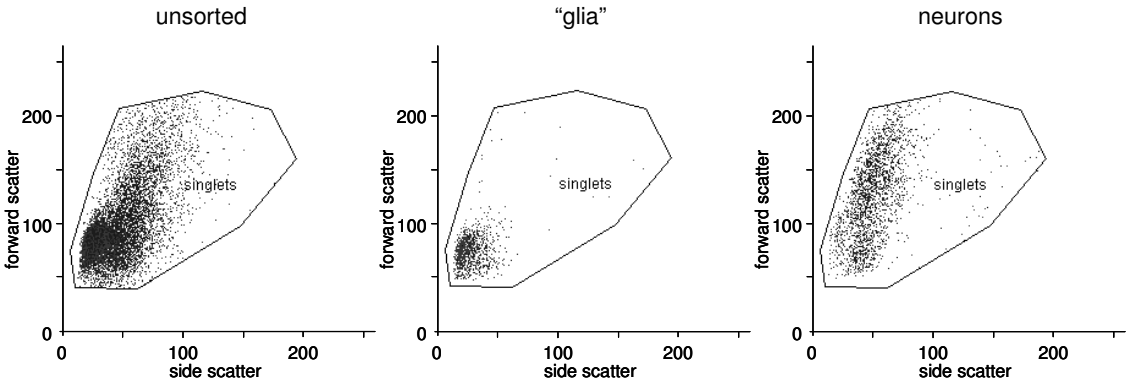

Figure 3.3 – Neuronal nuclei display a heterogeneous morphology, whereas the glial nuclei had a homogenous morphology. FACS staining of isolated brain nuclei and purification was performed as described in Figure 3.2. The morphology of the nuclei populations was investigated for size, as indicated by forward scatter (y-axis) and internal complexity side scatter (x-axis). Assistance in FACS was gratefully received from Andrew Sanderson and Martin Waterfall.

The FACS sorted nuclei were isolated and the purification verified by quantitative infra-red western blotting (Figure 3.4A). Analysis of NeuN levels confirmed that the neuronal nuclei were greatly enriched for this marker, whilst it was almost absent in the "glial" population. Analysis of MeCP2 levels indicated 7-fold more MeCP2 in the neuronal population than in the predominantly glial population (Figure 3.4A and B), similar to the difference in staining intensities recorded by FACS (about 6-fold).

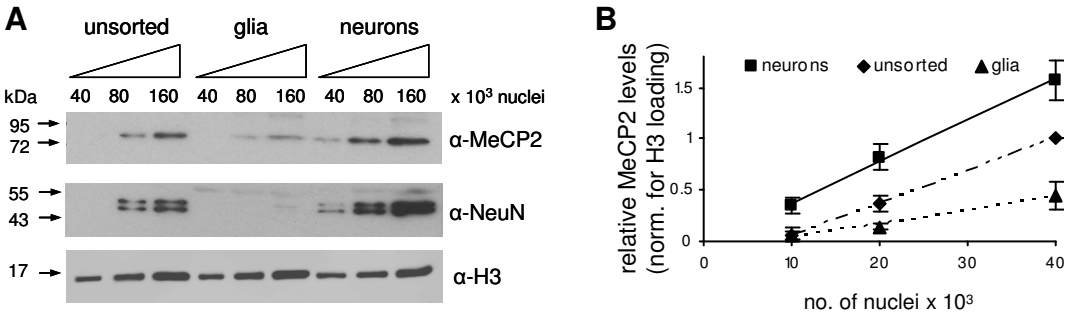

Figure 3.4 – Neuronal nuclei are enriched for both NeuN and MeCP2 relative to glial nuclei. Nuclei isolated from total brain were FACS sorted on the basis of NeuN staining (A) Infra-red western blotting for NeuN, MeCP2 (Sigma Aldrich, Mec-168) and histone H3 (as a loading control) was performed on unsorted nuclei and nuclei recovered from FACS sorting. (B) The graph indicates quantification of signal from MeCP2 westerns normalised for H3 loading. Error bars indicate mean ± SEM from 3 experiments.

As a further verification of the FACS based purification technique and its further potential uses, sorted crosslinked nuclei were used for MeCP2 ChIP analysis. However, the choice of a known binding target was limited, due to the lack a of clear consensus within the literature (for example (Horike et al., 2005; Schule et al., 2007)). The major satellite repeat can be considered as a known binding target as localisation studies have indicated the methylation dependent binding of MeCP2 to the pericentromeric heterochromatin (Nan et al., 1996). As a result, the unsorted and sorted populations were investigated for MeCP2 binding to this repeat (Figure 3.5). Consistent with the western blot analysis, neuronal nuclei showed elevated levels of MeCP2 binding relative to unsorted nuclei, whilst the predominantly glial fraction was depleted for binding. The magnitude of the difference did not however strictly correlate between techniques, with the ChIP data indicating ~20-fold greater binding in neurons and than glia (Figure 3.5), compared to a 7-fold difference in MeCP2 protein levels (Figure 3.4). This may suggest that the distribution of MeCP2 does not strictly correlate with its abundance. However, no additional work has been done to investigate this further.

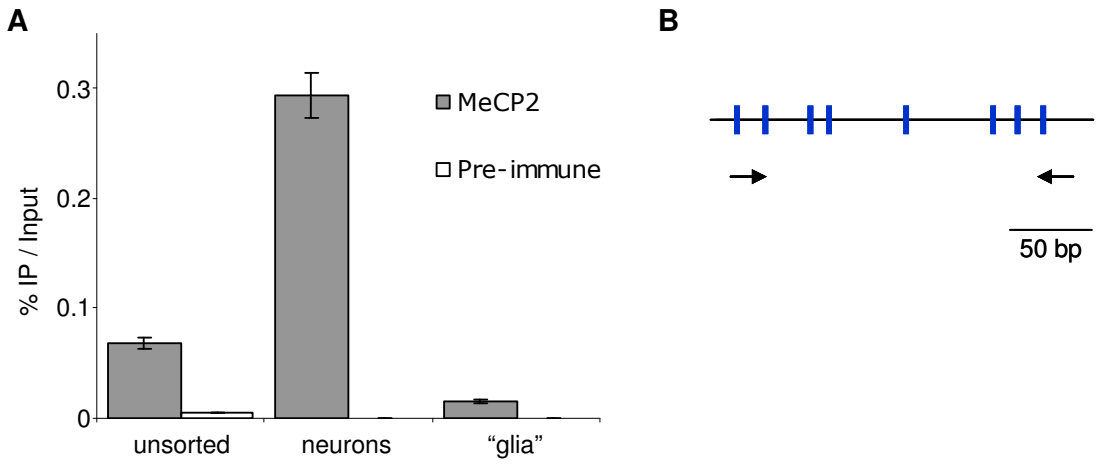

Figure 3.5 – FACS sorted neuronal nuclei show increased binding of MeCP2 to the major satellite repeat compared to glial nuclei. (A) Total wildtype brain nuclei were crosslinked using formaldehyde and then FACS sorted on the basis of NeuN staining into neuronal and predominantly glial fractions. Unsorted and sorted populations were used for ChIP with an antibody against MeCP2 (674) or with pre-immune serum as a negative control (see section 4.2 for verification of the antibody and discussion of the protocol). Immunoprecipitated DNA was analysed by real-time PCR using primers against the major satellite repeat. ChIP data is represented as % immunoprecipitation (IP) / Input. Error bars indicate mean ± SEM. Primers were designed by Xinsheng Nan. (B) A schematic of the 234 bp mouse major satellite repeat is indicated: the blue vertical lines mark the position of CpG sites; black arrows mark the position of the primers used for real-time PCR analysis; a scale bar is indicated. A melting curve analysis performed after the PCR amplification indicated the production of a single species (data not shown).

3.4 Quantification of absolute MeCP2 levels

The precise abundance of MeCP2 seems to be tightly regulated and key to its function (see section 3.1 for discussion). As a result, the absolute abundance of MeCP2 was determined. The concentration of recombinant MeCP2 was determined by titration against a bovine serum albumin (BSA) standard (Figure 3.6)

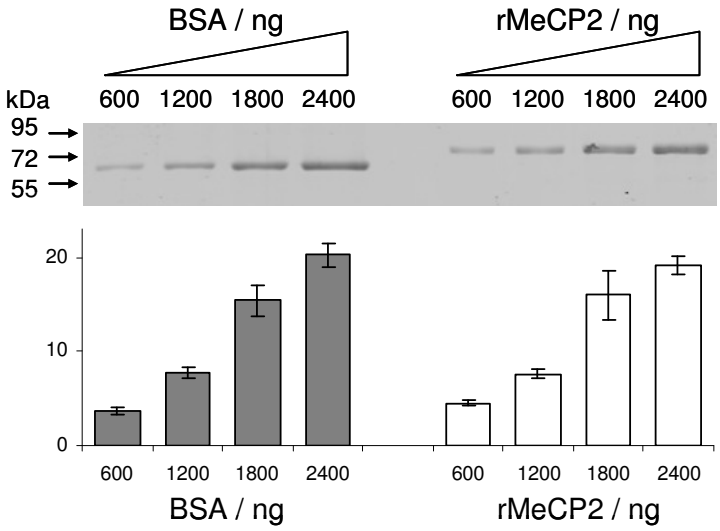

Figure 3.6 – The concentration of recombinant MeCP2 was determined by quantification against a BSA standard. A titration of a known concentration of BSA was compared to recombinant MeCP2 protein by SDS-PAGE and subsequent coomassie staining. The graph depicts densitometry analysis using Licor Odyssey imaging. The positions of molecular weight markers (kDa) are indicated. The recombinant MeCP2 was a kind gift from R. J. Klose. Error bars indicate mean +/- SEM from two SDS-PAGE experiments. Linear co-efficients (R^2) for BSA and recombinant MeCP2 standard curves were calculated as 0.99 and 0.96 respectively.

Quantitative western blotting allowed the determination of the absolute abundance of MeCP2 in unsorted nuclei isolated from mature mouse brain as 6 x 10^6 molecules per nucleus (Figure 3.7A and see Table 3.1 for details of the calculations). A similar analysis was used to calculate the amount of MeCP2 in sorted neuronal nuclei (NeuN positive) as 16 x 10^6 molecules per nucleus, compared with ~2 x 10^6 molecules per nucleus in the predominantly glial nuclear fraction (NeuN negative) (Figure 3.7B and Table 3.1). The level of MeCP2 within the glial nuclei is likely to represent an overestimate due to a proportion of neuronal nuclei staining negative for NeuN, such as cerebellar purkinje cells, olfactory bulb mitral cells, and retinal photoreceptor cells (Mullen et al., 1992). FACS staining suggested there may be some heterogeneity within the neuronal population (Figure 3.2) as previously

suggested (LaSalle et al., 2001), indicating that these average values may not apply to each cell. A previous preliminary analysis of MeCP2 levels in whole brain performed by Xinsheng Nan indicated that there were 6 x 10^6 molecules per nucleus, in agreement with the figure calculated above (Nan et al., 1996).

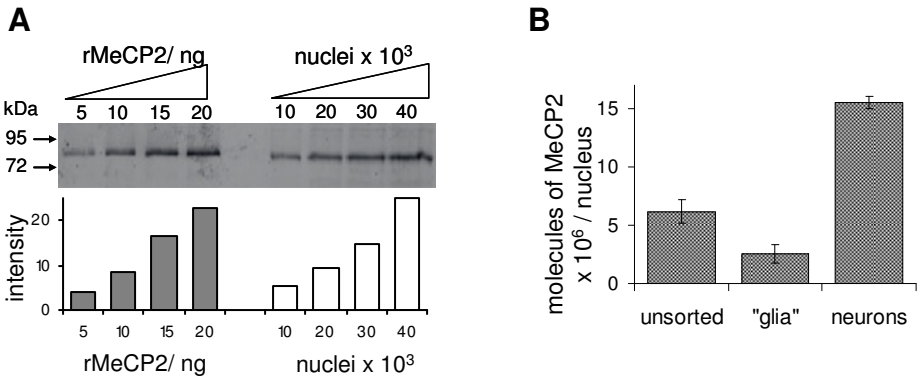

Figure 3.7 – The abundance of MeCP2 in purified neuronal nuclei approaches that of the histone octamer. (A) Nuclei were isolated from mature mouse brain, the concentration of which was determined by counting multiple dilutions with a haemocytometer. The level of MeCP2 in unsorted brain nuclei was quantified against recombinant MeCP2 by infra-red western blotting. The graph below indicates densitometric analysis using Licor Odyssey software. (B) Densitometric analysis of western blots to determine absolute abundance of MeCP2 in the different nuclei populations. The positions of molecular weight markers (kDa) are indicated. Error bars indicate the mean +/- SEM from at least 3 experiments.

Due to the fact that MeCP2 diseases are associated with neurological symptoms, consistent with its neuronal abundance, the absolute abundance of MeCP2 was examined in a known low expressing tissue. The abundance of MeCP2 in liver nuclei was calculated as 0.5 x 10^6 molecules per nucleus (Figure 3.8 and Table 3.1). To relate MeCP2 levels to the core components of chromatin, the abundance of histone H4 was determined using an equivalent technique as 64 x 10^6 molecules per brain nucleus (Figure 3.9 and Table 3.1), which corresponds to 32 x 10^6 nucleosomes in total.

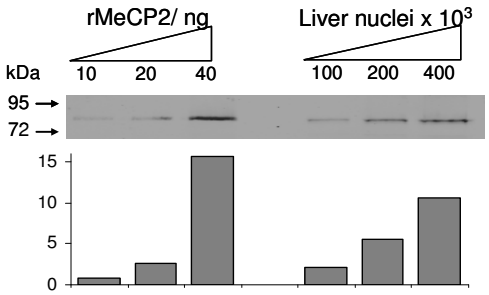

Figure 3.8 – MeCP2 is expressed at much lower amounts in liver nuclei. A titration of a known concentration of recombinant MeCP2 was compared to nuclei isolated from mouse liver by infra-red western blotting. The graph below indicates densitometric analysis using Licor Odyssey. The positions of molecular weight markers (kDa) are indicated.

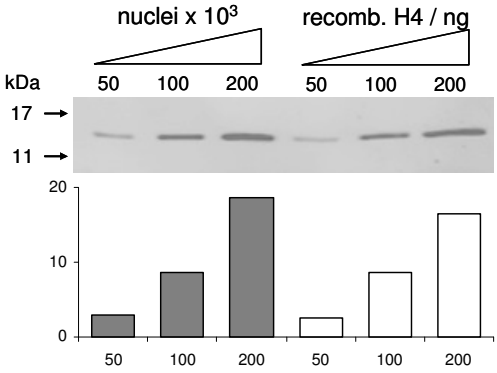

Figure 3.9 – Quantification of the absolute abundance of histone H4 in total brain nuclei. A titration of a known concentration of recombinant H4 and unsorted brain nuclei were compared using quantitative infra-red western blotting. The graph to the left indicates densitometric analysis of the western blot using Licor Odyssey software. This shows that in total brain nuclei i that there are 64 x 10^6 molecules of histone H4 per nucleus, which is equivalent to 32 x 10^6 nucleosomes.

	Tissue	Sub-population	ng / 1000 nuclei	fmol / nucleus	molecules / nucleus
(A) MeCP2	Brain	Unsorted	0.53	0.010	6×10^6
		Glia"	0.22	0.0042	2.5×10^6
		Neuron	1.35	0.026	15.5×10^6
	Liver	-	0.045	0.00086	0.5×10^6
(B) H4	Brain	Unsorted	1.2	0.11	64×10^6

Table 3.1 – Summary of the quantification of (A) MeCP2 and (B) Histone H4 absolute abundance in brain and liver nuclei. Using infra-red western blotting (Licor Odyssey), a titration of a known concentration of recombinant protein was compared to isolated nuclei and used to calculate the abundance (expressed as ng / 1000 nuclei). The number of mols per nucleus was determined from the molecular weight of the protein (MeCP2: 52.3 kDa; H4: 11.4 kDa). From this value the absolute number of molecules per nucleus was calculated using Avogadro's number (6×10^{23} molecules / mol).

3.5 Discussion

This work provides a method for the isolation of nuclei corresponding to a neuronal and a predominantly glial population. Due to heterogeneous distribution of MeCP2 between these populations, this technique allows purification of the neuronal fraction for the study of MeCP2 in the relevant cell type. There are however, a significant number of limitations to this technique as presented here. Firstly, despite NeuN often being considered to be a degenerate neuronal marker and sufficient for their identification (Herculano-Houzel and Lent, 2005), not all neuronal subtypes are positive for this marker (Mullen et al., 1992). Therefore, the purified neuronal nuclei will not truly represent an average of all neuronal subtypes, and correspondingly the glial population will be contaminated with NeuN-negative neuronal nuclei. The exact level of contamination within the glial population is difficult to quantify, due to lack of knowledge of the absolute numbers of neuronal subtypes. However, studies looking at isolated cerebellar purkinje neurons, which are NeuN-negative, suggest that they are sufficiently low in abundance to not significantly contaminate the glial population (Kriaucionis and Heintz, 2009). The second limitation is related to the requirement to isolate nuclei prior to FACS, as this reduces the number of available markers for sorting. Therefore the future possibilities to sort specific neuronal subtypes would be

greatly limited, as the majority of characterised cell markers are cytoplasmic or cell membrane-associated. Thirdly, again due to the loss of the cytoplasmic component, the ability to investigate this cellular compartment, for example its RNA levels, is lost. Fourthly, the overall purification efficiency is very low, with the protocol typically starting with 50-100 x 10^6 nuclei but a final recovery of only 1-2 x 10^6 nuclei, limiting the use of the protocol in downstream applications. This final issue is the greatest problem from a practical point of view. The protocol for the isolation of neuronal nuclei would likely be improved by the physical separation of nuclei based NeuN staining, as opposed to the FACS based procedure as it stands. This could possibly be achieved, through the labelling of NeuN antibody with biotin, followed by the physical purification using streptavidin magnetic beads. However, current bead sizes are significantly above that of the nuclear pore complex.

The use of FACS purified neuronal nuclei allowed the absolute abundance of neuronal MeCP2 to be determined as 16 x 10^6 molecules of MeCP2 per neuronal nucleus. Through the quantification of the histone H4, the number of nucleosomes was calculated as 32 x 10^6 per nucleus, which corresponds to one nucleosome per 165 base pairs of genomic DNA. The estimated number of methyl CpGs per diploid nucleus is 40 x 10^6 methylated CpG sites (based on 1 meCpG every 125 bp). Therefore, the number of MeCP2 molecules in the nucleus of a mature neuron approaches the number of histone octamers and methyl-CpG sites and may be sufficient to almost “saturate” the genome. This huge abundance of MeCP2, which the associated phenotypes indciate is imperative for its function, suggests that role of MeCP2 must be re-considered in the light of these observations.

Chapter 4 Analysis of MeCP2 Binding

4.1 Introduction

The general view amongst the majority of reports is that MeCP2 binds to discrete sites within the genome and regulates the gene at that specific locus, with the methylation dependence of the binding varying between studies (for example see (Ben-Shachar et al., 2009; Chahrour et al., 2008; Chen et al., 2003b; Jordan et al., 2007; Klose et al., 2005; Shahbazian et al., 2002a)). Multiple studies have identified putative target genes through expression analysis of wildtype and *Mecp2*-null brain. To further indicate the role of MeCP2 in the direct regulation of these genes, ChIP analysis at these loci was performed (Ben-Shachar et al., 2009; Chahrour et al., 2008; Jordan et al., 2007; Nuber et al., 2005; Urdinguio et al., 2008). However, these studies have typically looked at single site which confirmed the presence of MeCP2 binding, but did not confirm that genes unaffected by the loss of MeCP2 were negative for MeCP2 binding, as would be predicted by this simple causal mechanism.

Potentially the most characterised MeCP2 target gene is *Bdnf* (Chen et al., 2003b; Martinowich et al., 2003). These studies used cultured neurons taken from immature embryonic day 14 mouse brains (Martinowich et al., 2003) and from embryonic day 18 rat brains (Chen et al., 2003b). ChIP analysis revealed enriched binding of MeCP2 to exon 4 (Chen et al., 2003b; Martinowich et al., 2003). Note a change in the nomenclature of the alternative promoters (Aid et al., 2007). However, MeCP2 binding, albeit at a reduced level, was also observed at the other alternative promoters (Chen et al., 2003b). In the *Mecp2*-null neuronal cultures, the basal level of *Bdnf* expression was moderately elevated by two-fold, but no effect was observed on the stimulated activity, which was 100-fold higher than basal (Chen et al., 2003b). This lack of an effect on the stimulated state was suggested to be the result of a loss of MeCP2 binding upon neuronal stimulation following the phosphorylation of serine 421 on MeCP2 (Chen et al., 2003b). The relevance of *Bdnf* as a target gene in the mature mouse has since been questioned (Chang et al., 2006). This apparent contradiction could be the result of a change in function associated with the developmental increase in MeCP2 levels, as only 20% of MeCP2 is expressed in neonatal mice compared to the mature mouse (Figure 3.1). An independent study cloned 100 DNA fragments isolated by MeCP2 ChIP using neonatal mouse brains and identified the putatively imprinted *Dlx5/6* locus (Horike et al., 2005). Additionally, this study suggested that imprinting was lost in *Mecp2*-null mice and that MeCP2 mediated chromatin looping as indicated by limited chromosome

conformation capture analysis (Horike et al., 2005). The findings of this study have since been refuted (Schule et al., 2007) and indeed this locus has not been identified in expression studies using *Mecp2*-null mice. Analysis of binding targets in immature mice or neuronal cultures may not give a clear picture of its true function, due to the late onset of the MeCP2 associated phenotypes and the developmental increase in expression.

Suggestions that MeCP2 does not specifically bind methylated DNA and may actually act as an transcriptional activator (Chahrour et al., 2008; Georgel et al., 2003; Nikitina et al., 2006) have received some support from ChIP-microarray experiments using the neuronal cultured cell line SH-SY5Y (Yasui et al., 2007). The report claimed that MeCP2 does not selectively bind methylated promoters, but instead is bound predominately at nonmethylated promoters and has a similar binding pattern to RNA polymerase II (Yasui et al., 2007). The validity of this study is questioned by evidence that the MeCP2 antibody used in the ChIP analysis displayed only two-fold greater enrichment in wildtype brain than in the *Mecp2*-null brain (Yasui et al., 2007).

Overall, the understanding of the MeCP2 binding pattern in mature mouse brain is poorly understood. The observation that in the mature neurons, MeCP2 is theoretically abundant enough to coat the genome needs to be considered in the light of published ChIP data. To this end a ChIP assay using mature mouse brain was developed to identify the MeCP2 distribution across the genome.

4.2 Development of a ChIP assay

To allow high resolution mapping of binding sites, chromatin is fragmented prior to immunoprecipitation. This is typically achieved through the either sonication or *micrococcal* nuclease (MNase) digestion. Due to the putative binding of MeCP2 to the linker DNA (Chandler et al., 1999; Nikitina et al., 2007), extensive MNase treatment could therefore result in the selective degradation of MeCP2 binding sites. As a result, sonication of crosslinked brain material was initially attempted (Figure 4.1A). Despite extensive sonication, which is associated with heating of the sample, a significant fraction of chromatin remained refractory to fragmentation. Approximately 40% of the DNA by weight was shown to be greater than 1 kb in size by sucrose gradient centrifugation (data not shown). Accordingly, the fragmentation obtained by sonication alone was not deemed sufficient for the resolution required for ChIP analysis. The use of native ChIP was considered, as it is

likely that chromatin fragmentation would be easier in the absence of crosslinks, but it was thought to limit future experiments, as only stably bound proteins could be studied. Consequently, limited MNase digestion was used to 'loosen' the chromatin prior to sonication (Figure 4.1B). Using a combination of the two, sufficient fragmentation was obtained without extensive digestion of the linker DNA, as indicated by the absence of clear nucleosomal laddering. Figure 4.1C shows typical inputs for wildtype and *Mecp2*-null used in the subsequent ChIP experiments. Due to relatively precise amount of MNase activity required to sufficiently digest without excessive laddering, there is typically some variation associated with the degree of fragmentation between experiments. However, there is no clear indication that this affects the results obtained in ChIP analysis.

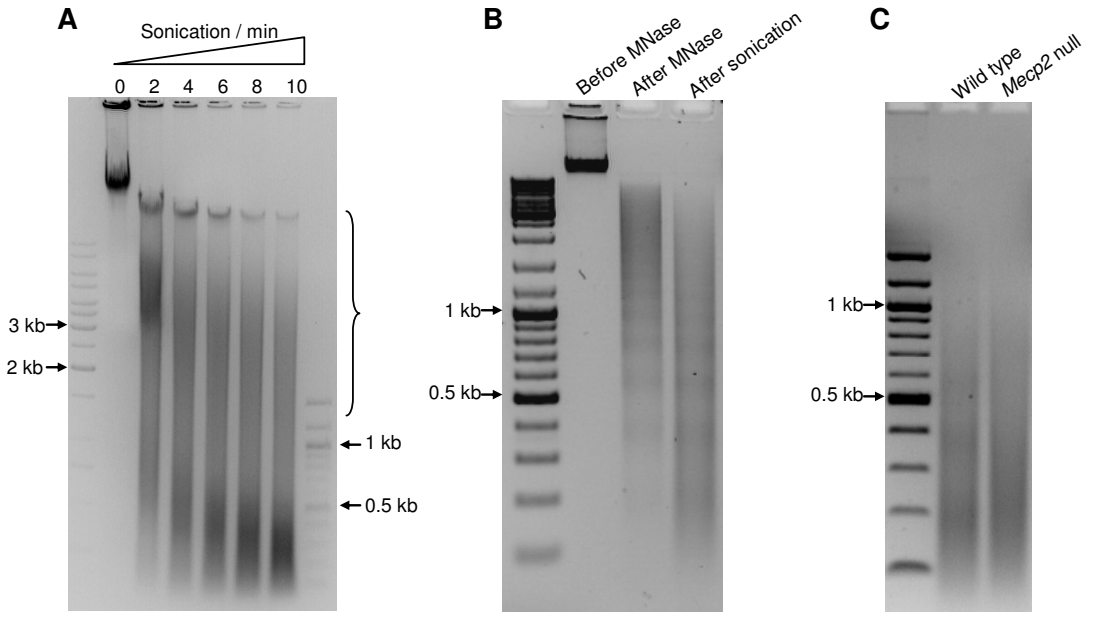

Figure 4.1 – Optimisation of chromatin fragmentation for use in ChIP. Mouse brain tissue was crosslinked using 1% formaldehyde and then subjected to (A) sonication alone; (B) a combination of limited MNase digestion followed by sonication. (C) Represents typical inputs used in ChIP analysis. Samples were taken at the indicated timepoints, after the crosslinks were reversed the DNA was isolated and subjected to gel electrophoresis. The smear of high molecular weight DNA that is refractory to sonication is indicated by a bracket. The positions of molecular weight markers (kb) are indicated.

Relatively standard crosslinking conditions were used (1% formaldehyde for 10 min at room temperature) with no specific optimisation of these conditions. For a discussion on how this might have affected the results, see section 4.7.

Given the poor efficiency of FACS purification of neuronal nuclei and concerns that length of protocol for nuclear isolation in combination with the FACS protocol, it was conceivable that this could lead to artefacts. It was decided therefore to use whole brain for ChIP analysis. Given that ~89% of brain MeCP2 is derived from the neuronal nuclei, the resulting pattern will be dominated by the neuronal population. It is important that the specificity of the antibody is verified, especially in the light of the recent publication suggesting the distribution of MeCP2 correlates with the RNA PolII occupancy (Yasui et al., 2007). The specificity of the 674 MeCP2 antibody (Nan et al., 1998) was tested by using the 234 bp major satellite repeat as a probe, as it is the only clear MeCP2 binding target ((Nan et al., 1996); for discussion see section 3.3). Despite the MeCP2 antibody displaying relatively low efficiency (1.7% IP/Input), it was highly specific, exhibiting a 130-fold enrichment in wildtype compared to *Mecp2*-null mouse brain (Figure 4.2). It is possible that the low the immunoprecipitation efficiency is a result of the formaldehyde crosslinking leading to the epitope becoming inaccessible; for a discussion of the possible effects of over-crosslinking see section 4.7. As a result, it was concluded that this high degree of specificity was sufficient for analysis of MeCP2 binding in the genome.

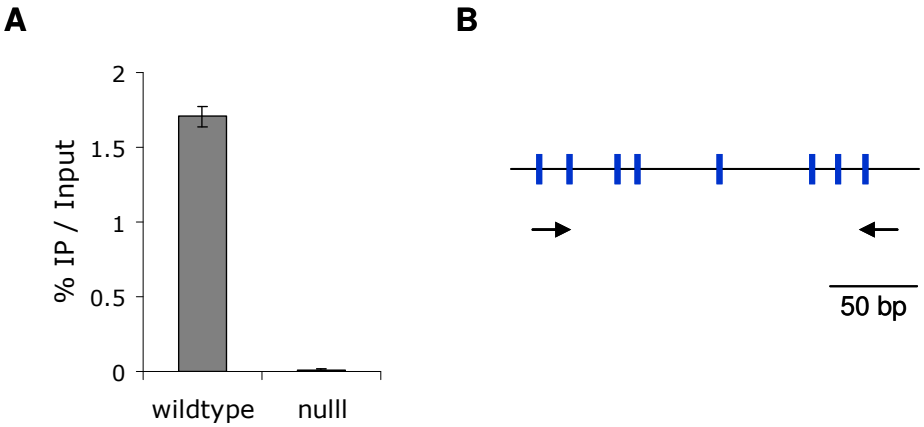

Figure 4.2 – Verification of the MeCP2 antibody for use in ChIP analysis. MeCP2 ChIP for the mouse major satellite repeat shows 130-fold enrichment in wildtype brain compared to *Mecp2*-null brain, indicating that the antibody is highly specific. Whole mouse brain was dissected and ChIP performed using an antibody against MeCP2. Immunoprecipitated DNA was analysed by real time PCR using primers specific for the major satellite repeat. (A) ChIP data is represented as % IP/Input. Error bars indicate mean ± SEM. For a schematic of the major satellite repeat see figure 3.5. (B) A schematic of the 234 bp mouse major satellite repeat is indicated: the blue vertical lines mark the position of CpG sites; black arrows mark the position of the primers used for real-time PCR analysis; a scale bar is indicated.

4.3 MeCP2 shows widespread binding across gene loci in mature mouse brain tracking the meCpG density

Previous reports using cultured immature neurons have suggested that MeCP2 binds to a discrete site within the *Bdnf* locus corresponding to exon 4 (Chen et al., 2003b; Martinowich et al., 2003). As a result, initially MeCP2 binding to the *Bdnf* locus was investigated. The mouse *Bdnf* locus has eight alternative first exons, most of which are located in two large CGIs, with a large intervening exon. A panel of 22 ChIP primers was designed across the a 39 kb region spanning the CGIs and the MeCP2 binding profile was determined in mature mouse brain using ChIP and quantitative PCR (ChIP-qPCR; Figure 4.3A; blue line). Significant enrichment across the entire region was observed when compared with the *Mecp2*-null brain (Figure 4.3A; red line), suggesting that MeCP2 is broadly distributed at the locus, in line with the almost-saturating abundance of MeCP2 in the mature brain (see section 3.4). Interestingly, MeCP2 binding over the *Bdnf* CGIs is depleted relative to the surrounding regions, raising the possibility that these active promoters fail to attract significant amounts of MeCP2 due to their lack of DNA methylation. The intervening regions are CpG deficient by comparison with the CGIs, and as part of the bulk genome would be expected to be largely methylated. To confirm that the methylation status follows these predictions, bisulphite DNA sequencing of a limited number of regions was attempted focusing around the first CGI. Despite, all bisulphite PCR primers giving successful amplification, not all the amplicons were possible to clone even though multiple strategies were attempted. However, the data indicates that the flanking portions are predominantly methylated (Figure 4.3B). The methylation status of the CGIs could not adequately be determined by bisulphite sequencing (Figure 4.3B), but CXXC-affinity chromatography has confirmed that they are largely nonmethylated (personal communication from R. Illingworth).

MeCP2 ChIP over the same region in wildtype liver gave a profile indistinguishable from the *Mecp2*-null brain, suggesting that the wildtype brain profile is a result of the high abundance in neurons (Figure 4.3A). The failure to detect signal for MeCP2 binding across the *Bdnf* locus in liver compared to null brain could be the result of lack of binding or reflect the limit of sensitivity of the experiment. Further work would be required to investigate this, such as confirmation of MeCP2 binding to sites such as the major satellite repeat in liver. It is possible that the MeCP2 binding pattern in the whole brain is an aggregate of a large number of distinct patterns from individual regions and neuronal subtypes. To address this, the ChIP was repeated using a specific brain region - the hippocampus, which represents a relatively

restricted set of neuronal subtypes. The resulting MeCP2 binding profile was very similar to that of the wildtype whole brain, with MeCP2 bound over the entire locus but reduced over the CGIs (Figure 4.3A). This finding makes it likely that the ChIP results represent a consistent pattern across many types of neurons.

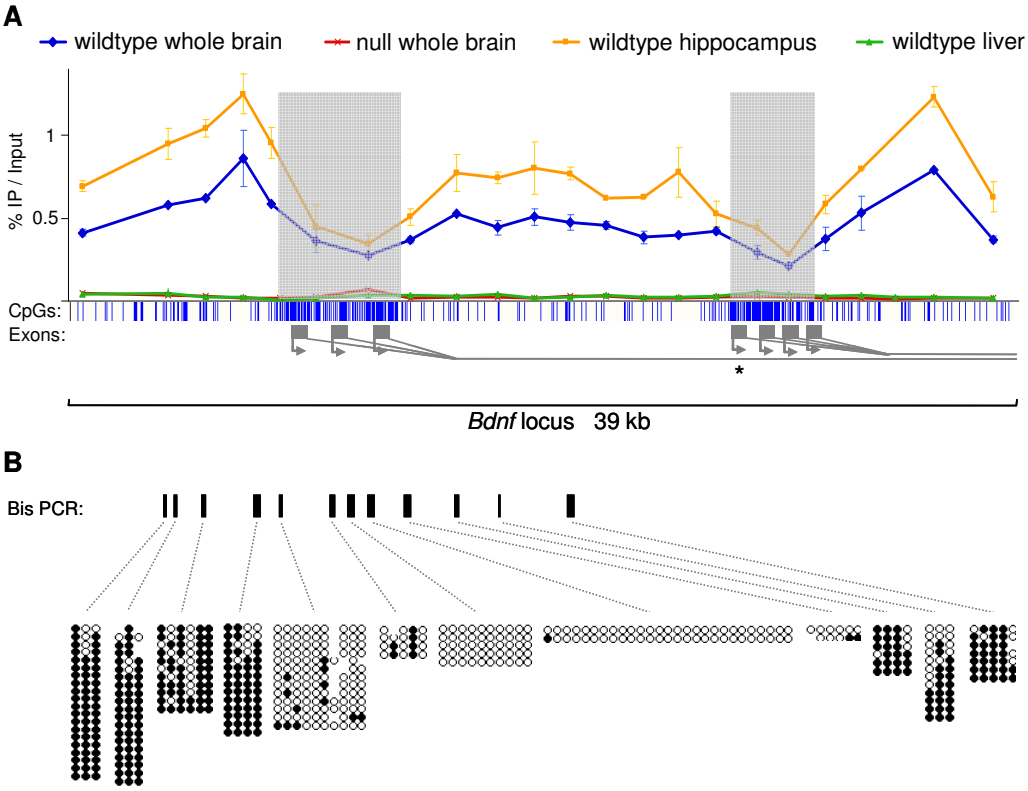

Figure 4.3 - MeCP2 shows widespread binding across 39 kb of the mouse *Bdnf* locus in mature brain. Mouse tissues (brain, hippocampus, liver) were dissected and ChIP performed using an antibody against MeCP2. (A) Immunoprecipitated DNA was analysed by real time PCR using a panel of primers and ChIP signal plotted as % IP/Input. Error bars indicate mean ± SEM from the replicates of the real time PCR analysis. Multiple biological replicates were performed, with this representing a typical result obtained. ChIP was performed on various tissues as indicated. The blue vertical lines below the graph indicate CpG sites, with CGIs shaded in light grey. The gene goes left to right; with alternative *Bdnf* first exons 1-7 indicated with dark grey rectangles, arrows indicate the transcriptional start sites. Bisulphite PCR amplicons are indicated with black bars. The asterisk marks the discreet binding site identified using cultured embryonic cortical neurons (Chen et al., 2003b). The region shown corresponds to chromosome 2: 109505000 – 109544220; NCBI build NCBIM37 (B) Total brain genomic DNA was subjected to bisulfite sequencing at specific sites as indicated above by black bars. Each row represents a single clone. Open and filled circles indicate nonmethylated and methylated CpG sites, respectively. Gaps indicate uncharacterised CpG sites. Due to unforeseen difficulties in cloning the amplicons, only a limited number of clones were retrieved in a number of cases.

A similar approach was used to examine MeCP2 occupancy over two housekeeping genes, *c-Myc* and *Actb*, which have nonmethylated CGIs that is are flanked by comparatively CpG deficient DNA, but which is methylated. ChIP across the *c-Myc* locus revealed MeCP2 bound to the methylated regions flanking the large CGI, but, as in the case of the *Bdnf* locus, MeCP2 binding was depleted across the CGI (Figure 4.4A). Analysis of 9.5 kb surrounding the *Actb* locus, showed a similar pattern with higher MeCP2 occupancy across the flanks, but depleted binding over the CGI (Figure 4.4B). In order to relate this result to methyl-CpG density, bisulphite sequencing was performed to unambiguously determine modification of all 262 CpG sites across the *Actb* region (Figure 4.4B). The results confirm that the CGI is DNA methylation-free, whereas the flanking region is predominantly methylated. Due to the variable CpG density, the methylation level varies dramatically across the locus. Significantly, the MeCP2 binding profile mirrors the methylation density throughout the region. A complete methylation density map could not be produced for either *Bdnf* or *c-Myc* due to the abundance of repetitive elements within these loci. Limited bisulphite analysis at the *Bdnf* locus and CXXC-affinity purification (personal communication with R. Illingworth) confirm that as expected these CGIs are non-methylated whereas the flanks are largely methylated. Overall, this analysis is compatible with the conclusion that MeCP2 shows widespread binding tracking the meCpG density. The observation of 'residual' binding across the CGIs is discussed in sections 4.6 and 4.7.

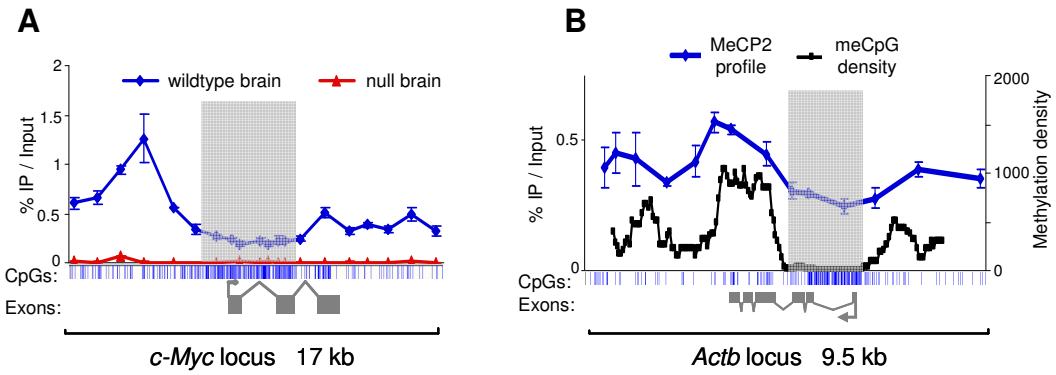

Figure 4.4 – MeCP2 binding is reduced over nonmethylated CGIs of housekeeping genes. Mouse brains were dissected and ChIP performed using an antibody against MeCP2. Immunoprecipitated DNA was analysed by real time PCR using a panel of primers across the (A) *c-Myc* locus and (B) the *Actb* locus. The blue vertical lines below the graph indicate CpG sites, with CGIs shaded in light grey. Exonic structure is indicated with dark grey rectangles. Total brain genomic DNA was subjected to bisulfite sequencing for a continuous run of 262 CpG sites across ~9 kb of the *Actb* locus. From this bisulphite analysis, the % methylation of each CpG site was calculated. The methylation density is plotted based on a window size of 650 bp and step of 50 bp with the cumulative % methylation of the CpG sites within each window calculated. Error bars indicate mean +/- SEM

4.4 MeCP2 specifically binds methylated DNA

Taken together, the results of the above ChIP experiments with the abundance measurements are consistent with the hypothesis that MeCP2 binds globally and the tracks the methyl-CpG density of the neuronal genome. To test whether MeCP2 specifically binds methylated DNA in neurons or binds regardless of the methylation status, as proposed by *in vitro* chromatin assembly experiments (Georgel et al., 2003; Nikitina et al., 2006) and other ChIP studies (Chahrour et al., 2008), genomic regions were selected that occur in both methylated and nonmethylated states within each cell. First, MeCP2 binding over 12 kb of the X-linked *Xist* locus was characterised, comparing male to female mouse brains (Figure 4.5A). In male mice there is a single methylated copy of the *Xist* gene, whereas in females there are two copies, one methylated and one nonmethylated (Hendrich et al., 1993). In the male brain, MeCP2 occupies the entire region, peaking over the densely methylated CGI. The finding that MeCP2 occupancy drops over non-methylated CGIs (Figure 4.3 and 4.4), but peaks over methylated CGIs (Figure 4.5) implies that MeCP2 is tracking the density of methyl-CpG.

In females, MeCP2 occupancy flanking the *Xist* CGI is indistinguishable from the male profile. As part of the bulk genome, it is likely that these flanking regions are methylated on both the Xa and Xi, however, this has not formally been shown. Over the *Xist* CGI, however, MeCP2 binding in female brain is reduced by half relative to the male brain (Figure 4.5A). This result is consistent with the notion that MeCP2 is only bound to one of the two *Xist* CGI alleles in the female. To test this more rigorously, the ChIP protocol was combined with bisulphite sequencing to determine the methylation status of the immunoprecipitated DNA. Input DNA from female brain gave approximately equal numbers of nonmethylated and methylated clones, as expected, confirming unbiased PCR amplification (Figure 4.5B). As a control, bisulphite sequencing of *Xist* DNA immunoprecipitated by an antibody against acetylated histone H3 was performed. This histone mark is specific for the active *Xist* allele and duly recovered the nonmethylated allele in 97% of clones (n=39 clones). Following ChIP with anti-MeCP2 antibody, however, 88% of the recovered clones were methylated (n=58 clones).

A similar result was obtained by analysing 9 kb of the imprinted *Snrpn* locus, which has a paternal nonmethylated allele and a maternal methylated allele (Sutcliffe et al., 1994). MeCP2 is bound through the entire region, peaking over the imprinted CGI (Figure 4.5C). Using the ChIP-bisulfite protocol, the input DNA gave approximately equal numbers of non-methylated and methylated clones (Figure 4.5D). As before, acetylated histone H3 ChIP recovered 100% non-methylated clones (n=23), whereas 89% of the clones obtained by MeCP2 ChIP were methylated (n=19). The conclusion is that MeCP2 preferentially binds to the methylated allele in each case.

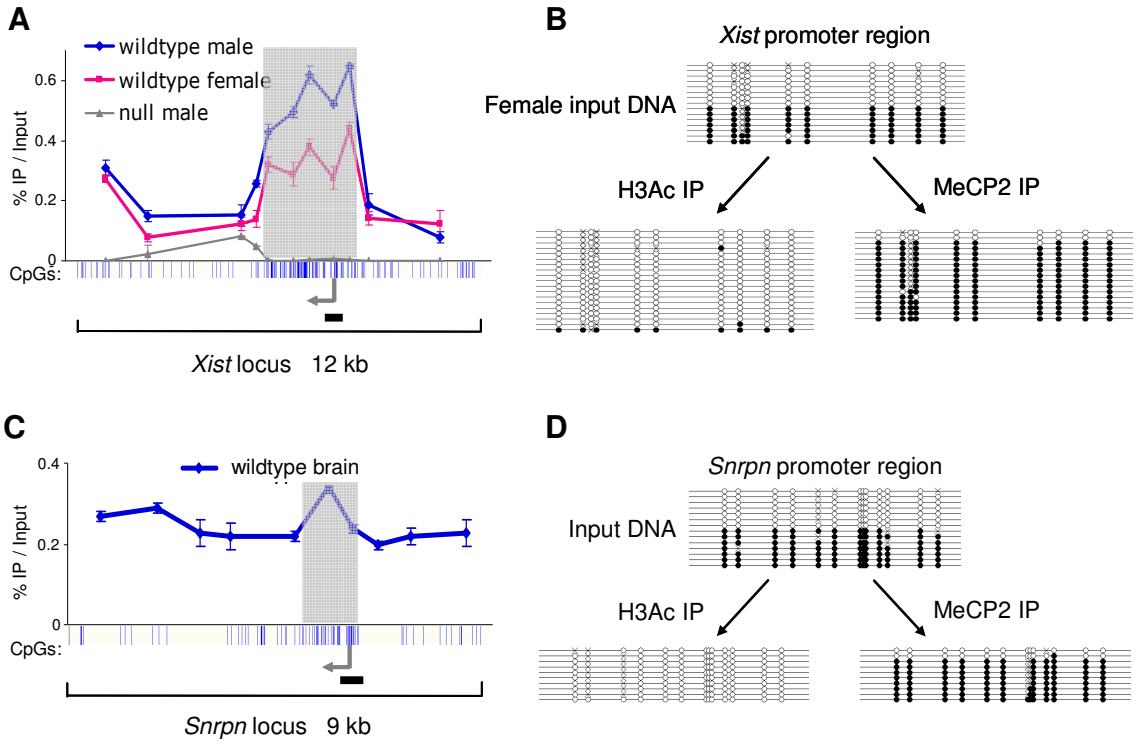

Figure 4.5 – Brain MeCP2 binds selectively to methylated DNA *in vivo*. Whole mouse brains were dissected and ChIP performed using antibodies against MeCP2 and acetylated histone H3. Input DNA and immunoprecipitated DNA were analysed by real time PCR and subjected to bisulfite sequencing. (A) MeCP2 binding profile across 12 kb of the *Xist* locus in wildtype male brain (blue), wildtype female brain (pink) and *Mecp2*-null brain (grey). The blue vertical lines below the graph indicate CpG sites, with the CGI shaded in light grey. The transcription start site is indicated with an arrow. The horizontal black line marks the region amplified for bisulfite sequencing. (B) ChIP was performed on the wildtype female brain for MeCP2 and histone H3 acetylation. Recovered DNA was used for bisulfite sequencing of a region of the *Xist* promoter. A representative number of clones chosen at random are indicated (see main text for details of total numbers of clones); each line represents a single clone. Open and filled circles indicate non-methylated and methylated CpG sites, respectively. Crosses indicate uncharacterised CpG sites. (C) MeCP2 ChIP profile across 9 kb of the imprinted *Snrpn* locus in wildtype brain. (D) ChIP was performed on the wildtype brain for MeCP2 and histone H3 acetylation. The resulting DNA was used for bisulfite sequencing of a region of the *Snrpn* promoter. Error bars indicate mean +/- SEM.

4.5 MeCP2 ChIP-seq analysis of global distribution

Analysis of candidate loci across ~150 kb of the mouse brain genome suggests that MeCP2 is widely distributed. To view MeCP2 occupancy genome-wide, high-throughput DNA sequencing of total MeCP2-bound chromatin using Illumina Solexa sequencing (ChIP-seq) was performed. In all, 2.9 x 10^9 bases were sequenced, of which 1.3 x 10^9 were uniquely

mappable. Despite relatively extensive sequencing, clear peaks corresponding to discrete MeCP2 binding sites were not identified, even with a total of seven lanes of sequencing. For comparison, a typical transcription factor usually requires one lane. Instead, sequencing reads were dispersed throughout the genome, covering 56% of the mappable mouse genome (the haploid mouse genome is 2.5 x 10^9 bp, of which 1.8 x 10^9 bp is not removed by repeat masking). This lack of clear binding sites fits with our analysis of candidate regions by ChIP-PCR. Since MeCP2 binding is genome-wide, it would be necessary to sequence the entire genome several-fold to obtain a comprehensive high-resolution profile. Due to the lack of depth of sequencing, it was generally not possible to directly compare the sequencing data with the results obtained by ChIP across candidate regions (ChIP-qPCR; see section 4.3 and 4.4). However, it is clear that for *c-Myc,* the sequencing data correlates with the previous results (Figure 4.6). For other regions, however, the correlation is less clear (Figure 4.7)

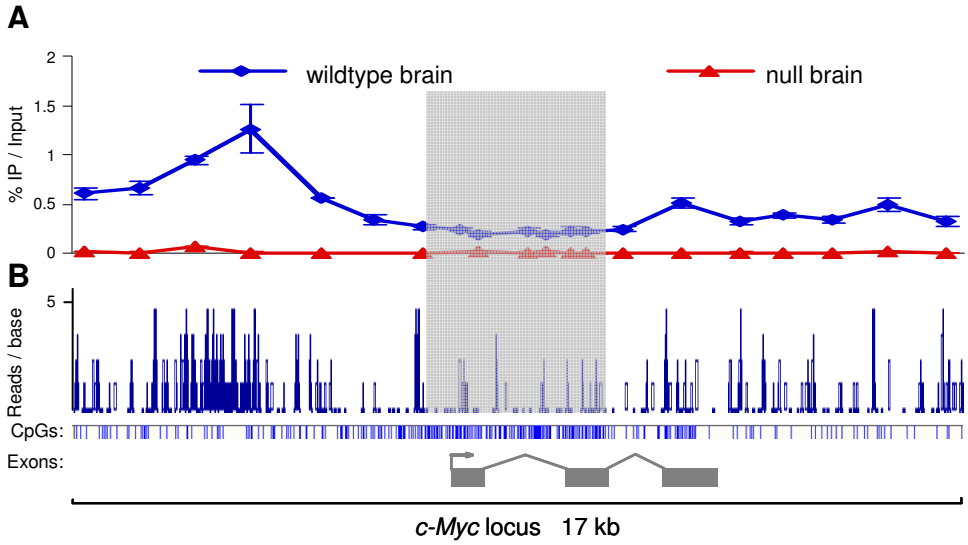

Figure 4.6 – Comparison of the brain MeCP2 ChIP-qPCR with the brain MeCP2 ChIP-seq data across the *c-Myc* locus verifies the high-throughput sequencing data. MeCP2 ChIP was performed using whole mouse brain, and immunoprecipitated DNA was end-repaired and linkers ligated for Illumina Solexa sequencing. Prior to sequencing, a standard PCR using the linkers as primers was performed to confirm successful ligation of the linkers (data not shown). Libraries were sequenced on the Solexa Genome Analyzer to generate 37 bp reads. Single-end sequence reads were mapped to the mouse genome. (A) MeCP2 ChIP-qPCR analysis (see section 4.3) aligned with the (B) ChIP-seq data across 17 kb of the *c-Myc* locus. Vertical axis represents reads per base.

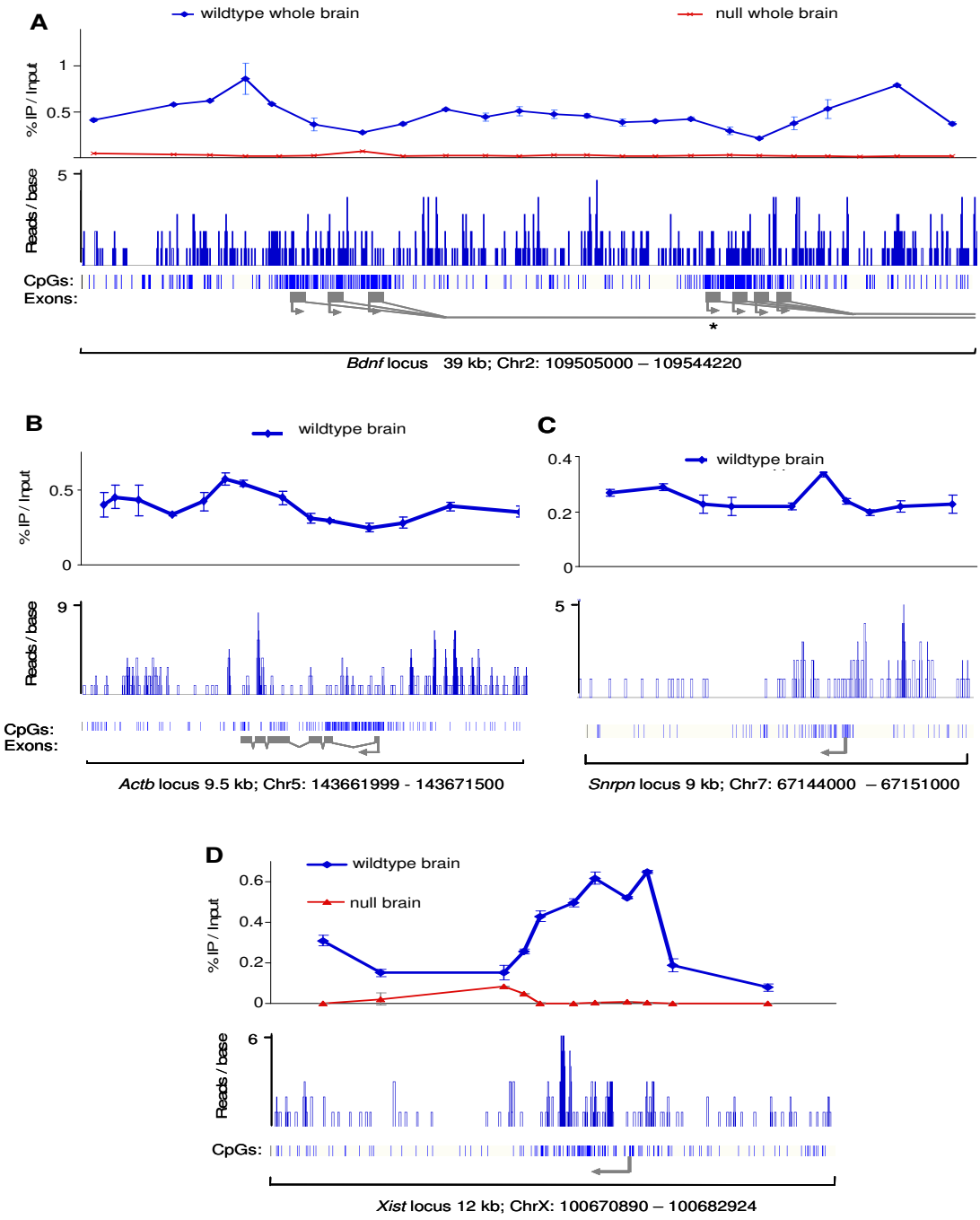
A
wildtype whole brain
null whole brain
% IP / Input
1
0.5
Reads / base
5
CpGs:
Exons:
*
Bdnf locus 39 kb; Chr2: 109505000 – 109544220
B
wildtype brain
% IP / Input
0.5
0
Reads / base
9
CpGs:
Exons:
Actb locus 9.5 kb; Chr5: 143661999 - 143671500
C
wildtype brain
0.4
0.2
0
5
Snrpn locus 9 kb; Chr7: 67144000 – 67151000
D
wildtype brain
null brain
% IP / Input
0.6
0.4
0.2
0
Reads / base
6
CpGs:
Xist locus 12 kb; ChrX: 100670890 – 100682924

Figure 4.7 - Comparison of MeCP2 ChIP-qPCR data with the b MeCP2 ChIP-seq data across (A) *Bdnf* (B) *Actb* (C) *Snrpn* and (D) *Xist* loci. MeCP2 ChIP was performed using whole mouse brain, and immunoprecipitated DNA was sequenced on a Solexa Genome Analyzer to generate 37 bp reads. Single-end sequence reads were mapped to the mouse genome. For each loci, the upper panel indicates the MeCP2 occupancy determined by ChIP-qPCR (vertical axis represents ChIP signal plotted as % IP/Input), whilst the lower panel indicates the binding pattern determined by ChIP-seq (vertical axis represents reads per base). Co-ordinates of loci are indicated according to genome build NCBIM37.

Despite the inability to clearly look at individual loci, we were able to robustly test for a relationship with DNA methylation by comparing the profile of the CpG density in a sliding 5 kb window with that of MeCP2 bound chromatin. CGIs only represent ~2% of the mouse genome and typically have an approximate length of 1 kb, whilst the remaining ~98% is predominantly methylated (Ehrlich et al., 1982). Therefore, by using a 5 kb window size this primarily looks for fluctuations in CpG density within the bulk genome. The match between MeCP2 ChIP-seq and CpG density was striking (Figure 4.8), suggesting that MeCP2 binding coincides with the CpG dinucleotide sequence. As a more direct test of the methyl-dependence of the distribution of MeCP2, the distribution of meCpG was determined by fractionating whole brain genomic DNA based on its affinity for an immobilized methyl-CpG binding domain (Cross et al., 1994; Illingworth et al., 2008). Retained sequences were analyzed by high-throughput sequencing (MBD-seq; this data was kindly provided by R. Illingworth). Despite an apparent circularity in comparing the binding specificity of a MBD-tagged protein *in vitro* and the distribution of a MBD protein *in vivo*, an earlier comparison between this method of meCpG mapping and use of an anti-meC antibody (MeDIP) indicated that both are equally effective at measuring meCpG density (Zhang et al., 2006). Comparison of MeCP2-bound chromatin and MBD-seq revealed that *in vivo* bound MeCP2 mirrors meCpG over long regions of chromosomal DNA (Figure 4.8), consistent with the candidate analysis performed using ChIP-qPCR.

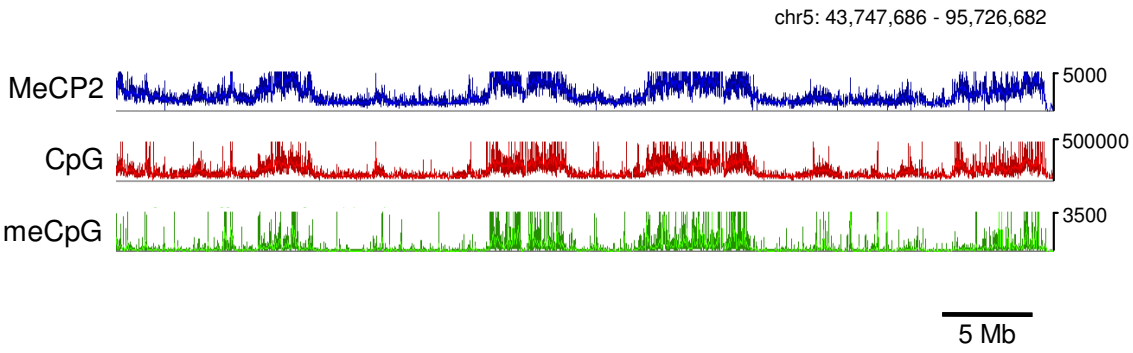

Figure 4.8 – High throughput sequencing of immunoprecipitated chromatin shows MeCP2 globally distributed and tracking the meCpG density. Profiles of MeCP2-bound sequences (blue), CpG density (red) and sequencing from methyl-CpG-rich sequences (green) were analysed using a sliding window (5 kb window; 1 kb step). A 52 Mb region of chromosome 5 is shown. The vertical axis represents sequencing hits or number of CpGs per window. MBD-seq data was provided by R. Illingworth. Bioinformatic support was gratefully received from R. Illingworth, S. Webb and A. Kerr.

ChIP-qPCR analysis indicated that MeCP2 binding was reduced over the nonmethylated CGIs associated with the *c-Myc*, *Actb* and *Bdnf* loci. To determine if this applied genome-wide, the mappable mouse genome sequence was binned according to CpG density in 500 bp windows and the level of MeCP2 binding in each bin determined (Figure 4.9A). The results show that MeCP2 binding initially increases with CpG composition as predicted by Figure 4.7, as these sequences are part of the largely methylated bulk genome. However, MeCP2 occupancy drops at densities above 25 CpGs per 500 bp. This density predominantly reflects CGIs, which are largely nonmethylated (Figure 4.9B)

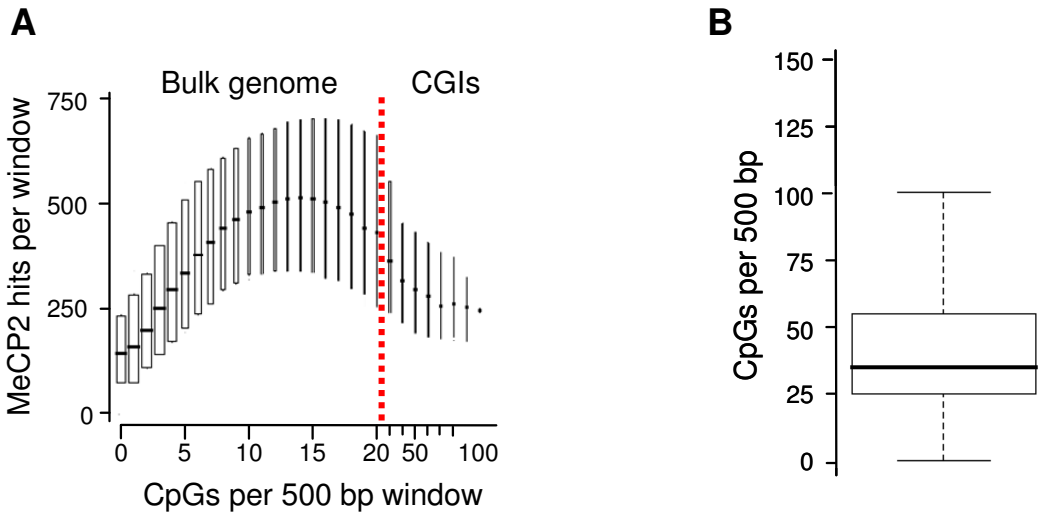

Figure 4.9 – MeCP2 occupancy increases with CpG density within the bulk genome, but is depleted over CGIs. (A) The genome was scanned using a 500 bp window (100 bp step) and the number of MeCP2 hits per window plotted against the number of CpGs per window. Boxplots are shown with the horizontal line indicating the median and the surrounding box showing the interquartile range. The width of the box is proportional to the fraction of the genome that corresponds with that CpG density. (B) Boxplot showing the number of CpGs per 100 bp calculated using the NCBI strict algorithm for CGIs. (www.ncbi.nlm.nih.gov/mapview/static/humansearch.html#cpg). The majority of CGIs contain greater than 5 CpG sites/100 bp as indicated by red line (A), whereas the bulk genome has a lower CpG density.

Furthermore, ChIP-qPCR indicated increased MeCP2 occupancy over the methylated CGIs of the *Xist* and *Snrpn* genes (Figure 4.5). The genome-wide binding to methylated CGIs, as identified by MBD-seq, was therefore determined. The results reveal a peak of MeCP2 binding centred over the methylated CGIs, further confirming the role of DNA methylation in targeting MeCP2 (Figure 4.10). This observation of increased binding over methylated G/C rich sequences confirms that the depleted binding observed over the largely nonmethylated G/C rich sequences (Figure 4.9) is not the result of bias against G/C rich DNA in the sequencing reaction. This bias has previously been suggested at extreme G/C contents, (Hillier et al., 2008). However, this observed bias is outside the typical G/C content range for mouse CGIs.

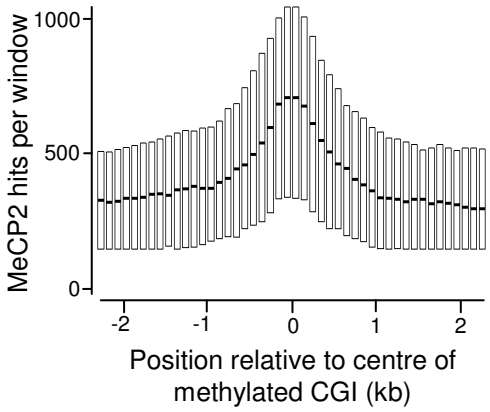

Figure 4.10 – MeCP2 ChIP-seq identifies increased occupancy over methylated CGIs. Methylated CGIs (n = 1273) were indentified by methyl-CpG affinity chromatography (data kindly donated by R. Illingworth). The surrounding 5 kb of genomic DNA was analyzed for MeCP2 binding using a 500 bp window (100 bp step). Data presented as the number of MeCP2 hits per window plotted against the number of CpGs per window. Boxplots are shown with the horizontal line indicating the median and the surrounding box showing the interquartile range

Overall, this first genome-wide analysis of the distribution of MeCP2 in mouse brain suggests that MeCP2 binds globally, consistent with its near-histone octamer abundance, tracking the meCpG density of the genome. The inability to robustly look at individual gene loci, as indicated by some inconsistency between the ChIP-qPCR and the ChIP-seq results (figures 4.6 and 4.7), would hopefully be improved by an increased depth of sequencing.

4.6 Are there high and low affinity binding sites?

A previous *in vitro* study used methyl-SELEX in order to determine any additional sequence requirements for the binding of MeCP2 and identified the frequent presence of an A/T run of four or more bases close to the methylated CpG site (Klose et al., 2005). Alignment of the selected sequences identified the consensus sequence, with the A/T run motif appearing in two vertical stripes, as indicated in Figure 4.11A. The relevance of this consensus for the targeting of MeCP2 was not robustly tested *in vivo* (Klose et al., 2005), but the preferential binding site in the *Bdnf* locus in cultured neurons was noted to contain this A/T run (Chen et al., 2003b; Ho et al., 2008). This suggests that there may be additional levels of targeting, despite the data presented here indicating that the distribution of MeCP2 is simply defined by the meCpG density (Figure 4.9). To determine if the presence of an A/T run impacted the probability of MeCP2 occupancy, isolated CpG sites, with no other CpG site within 40 bp

either side were identified within the genome. This criterion for identifying isolated CpG sites excludes CGIs and focuses on the largely methylated bulk genome. Isolated CpG sites were scored for the presence or absence of this A/T run motif (Figure 4.11A; (Klose et al., 2005)) and then the binding of MeCP2 within this region quantified as determined by ChIP-seq. No preference was observed for CpG sites associated with this motif, questioning the relevance of this consensus sequence *in vivo* for the targeting of MeCP2 (Figure 4.11B).

A

-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
N	N	N	N	N	N	N	N	N	N	N	N	N	**C**	**G**	**W**	**W**	**W**	**W**	N	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N	N	N	**C**	**G**	N	**W**	**W**	**W**	**W**	N	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N	N	N	**C**	**G**	N	N	**W**	**W**	**W**	**W**	N	N	N	N	N	N
N	N	N	N	N	N	N	N	N	N	N	N	N	**C**	**G**	N	N	N	N	N	**W**	**W**	**W**	**W**	N	N	N
N	N	N	N	N	N	N	N	N	N	N	N	N	**C**	**G**	N	N	N	N	N	N	**W**	**W**	**W**	**W**	N	N
N	N	N	N	N	N	N	N	N	N	N	N	N	**C**	**G**	N	N	N	N	N	N	N	**W**	**W**	**W**	**W**	N
N	N	N	N	N	N	N	N	N	N	N	N	N	**C**	**G**	N	N	N	N	N	N	N	N	**W**	**W**	**W**	**W**

W: A or T; N: any base

B

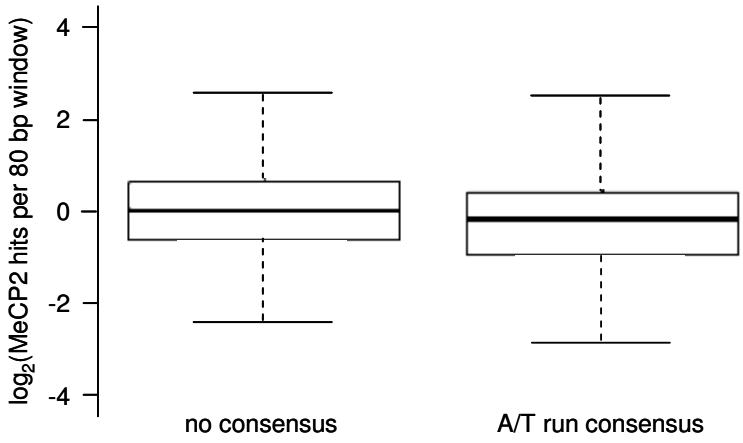

Figure 4.11 – MeCP2 ChIP-seq indicates no clear preference for an A/T run adjacent to the CpG site *in vivo*. (A) methyl-SELEX identified a broad consensus motif that contained an A/T run in two alternative positions after the CpG dinucleotide as indicated in red and green. Due to the palindromic nature of CpG, it is also possible for the A/T run to be situated in front of the CpG site (these possible consensus sequences are not shown). A or T is represented as W; any base is represented as N. (B) Isolated CpG sites within the bulk genome were identified as containing no other CpG site within 40 bp either side. Subsequently, these sites were scored for the presence or absence of the above consensus motif and MeCP2 binding was then assigned within this window using the ChIP-seq data. Data presented as boxplots for $\log_2$(MeCP2 hits per 80 bp window); outliers have been removed.

Increasing salt concentration can be used to selectively solubilise nuclear proteins. An early studying characterising the properties of MeCP2 monitored its release from rat brain nuclei with increasing salt concentrations (Meehan et al., 1992). This indicated that MeCP2 was released over a broad salt range. However, with careful quantification, it was noted that the release of MeCP2 appears to follow a biphasic pattern in comparison with the release of histone H1, suggesting the possibility of high and low affinity binding sites for MeCP2 (Figure 4.12A; (Meehan et al., 1992)). To investigate the possibility of differential binding of MeCP2, the ChIP protocol was combined with an initial 300 mM salt-wash to remove this putative low affinity bound component. This necessitated using isolated nuclei for the ChIP analysis. Preliminary analysis indicated that 60 minute incubation with 300 mM NaCl released 75% of the nuclear MeCP2, whereas only 25% of MeCP2 was released with 150 mM NaCl, which is approximately physiological (Figure 4.12B). ChIP analysis was performed across the mouse *Bdnf* locus in salt treated brain nuclei (Figure 4.12C). For some unknown reason the use of isolated nuclei led to a minor alteration of the MeCP2 binding pattern as observed in whole brain, perhaps suggesting that the time lag between nuclei preparation and crosslinking can induce artefacts. Despite these concerns, the binding pattern observed between the 150 mM-washed sample and 300 mM-washed sample appears largely the same, with on average 60% of the occupancy in the 300 mM sample compared to the 150 mM sample, consistent with the western blot analysis. To test if there are any key differences in the binding pattern between these samples, the signal for each primer pair was normalised to the average of that ChIP sample, thereby setting the average of each sample to 1 (Figure 4.12D). This analysis indicates that there are no clear differences, thereby providing no evidence of differential binding, at least at this locus.

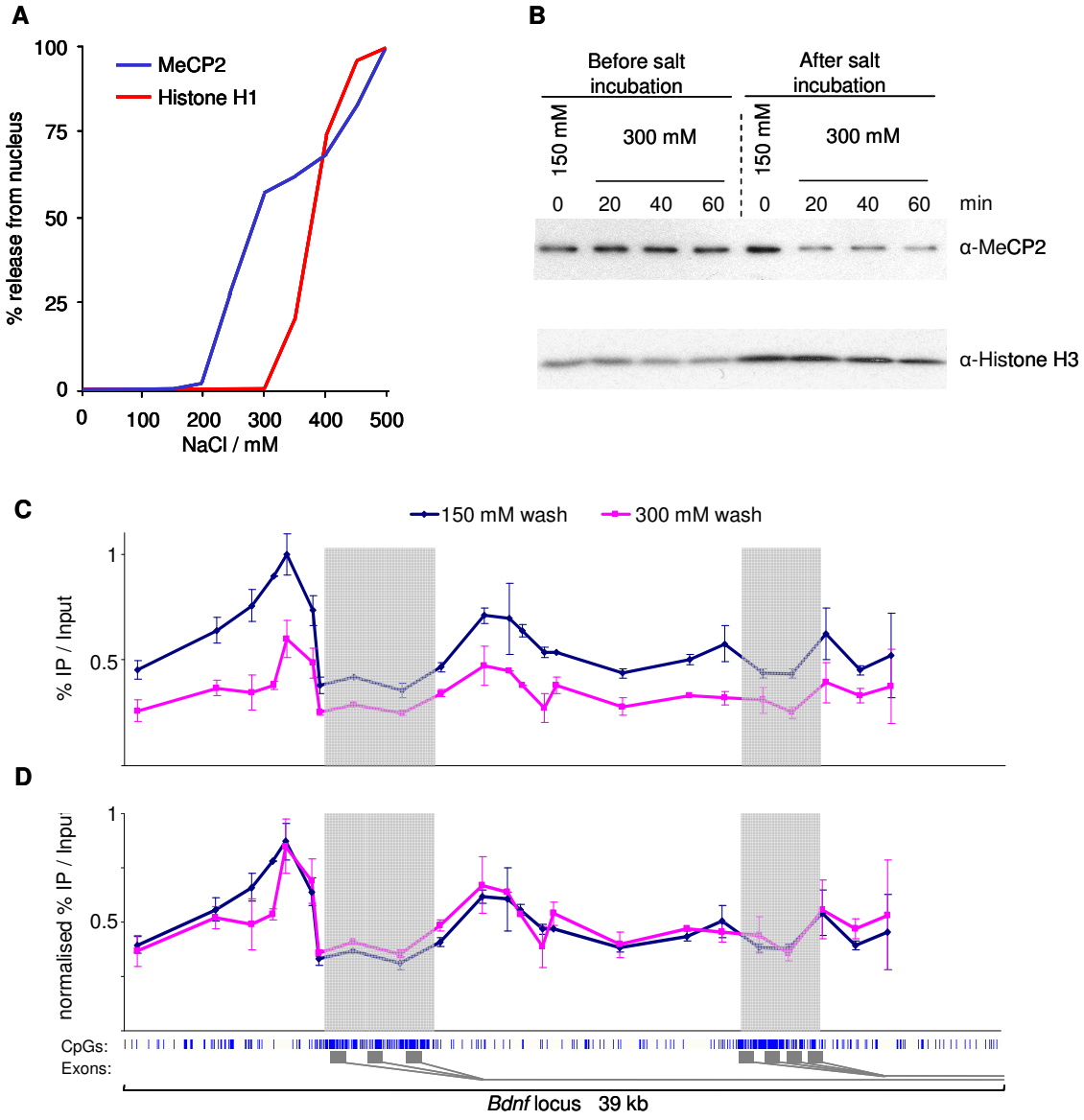

Figure 4.12 – Partial salt-dependent release of MeCP2 indicates no differential binding across the *Bdnf* locus. (A) Salt extraction of proteins from rat brain nuclei indicated a biphasic release of MeCP2 (blue line) compared to histone H1 (red line), proposing putative high and low affinity bound populations of MeCP2 (figure adapted from (Meehan et al., 1992)). (B) Isolated mouse brain nuclei were incubated with extraction buffer supplemented with either 150 mM or 300 mM NaCl for the indicated time. Western blot analysis of the salt-washed pelleted nuclei was performed for MeCP2 and histone H3 (as a loading control) before and after the incubation. Densitometry was used to quantify the western blot, data was normalised to histone H3. (C) Isolated mouse brain nuclei were incubated with extraction buffer supplemented with either 150 mM or 300 mM NaCl for 20 min. Salt-washed nuclei were then crosslinked and used for MeCP2 ChIP analysis across the *Bdnf* locus. ChIP signal plotted as % IP/Input. (D) The ChIP results were normalised to the average of each sample, therefore averaging each dataset to 1. The blue vertical lines below the graph indicate CpG sites, with CGIs shaded in light grey. Alternative *Bdnf* first exons are indicated with dark grey rectangles. Error bars indicate mean +/- SEM.

This data provides clear evidence for methylation dependent binding in neurons, contrary to the reports of focusing on methylation independent binding *in vitro* (Georgel et al., 2003; Nikitina et al., 2006) and *in vivo* (Chahrour et al., 2008). Early quantitative *in vitro* analysis for the methyl-specific affinity suggested that methylation independent binding may be as little as 10-fold less (Nan et al., 1993). However, measurements from *in vitro* studies are likely to be highly dependent on the experimental conditions. It is conceivable that a highly basic protein, such as MeCP2 which binds to a short dinucleotide consensus sequence, will also exhibit some non-specific binding to the negatively charged DNA. The ChIP analysis presented here may suggest that there is a component of non-specific binding. Firstly, it is clear that the MeCP2 ChIP signal for wildtype brain is never as low as the level observed in *Mecp2*-null brain, even at the *Actb* CGI where the methylation level was determined to be zero, suggesting that MeCP2 is bound. This could be due to immunoprecipitation of flanking methylated regions and therefore represents the resolution restraints of the ChIP analysis. However, the *c-Myc* CGI is very large, almost 3 kb, and therefore the residual MeCP2 ChIP signal is larger than would be reasonably expected from this flanking contamination. Secondly, the ChIP-bisulphite analysis for the *Xist* locus reproducibly identified only 88% of clones as methylated, even though the ChIP signal in wildtype female mice was ~150-fold greater than the *Mecp2*-null, consistent with a non-specific binding component. Serendipitously, the binding across the *Dlx5/6* locus had been examined, as previous reports indicated a discrete binding pattern of MeCP2 in 1-day old mice (Horike et al., 2005). Despite the relatively high CpG density across the 40 kb region examined, with 1 CpG per 45 bp compared to a genome average of 1 CpG per ~100 bp, the MeCP2 ChIP signal was surprisingly low (Figure 4.13A and B) with no correlation with CpG density as previously indicated (Figure 4.9). Limited bisulphite sequencing indicated that for the most part the DNA was nonmethylated, with only the peak of MeCP2 binding associating with methylated DNA (Figure 4.13C).

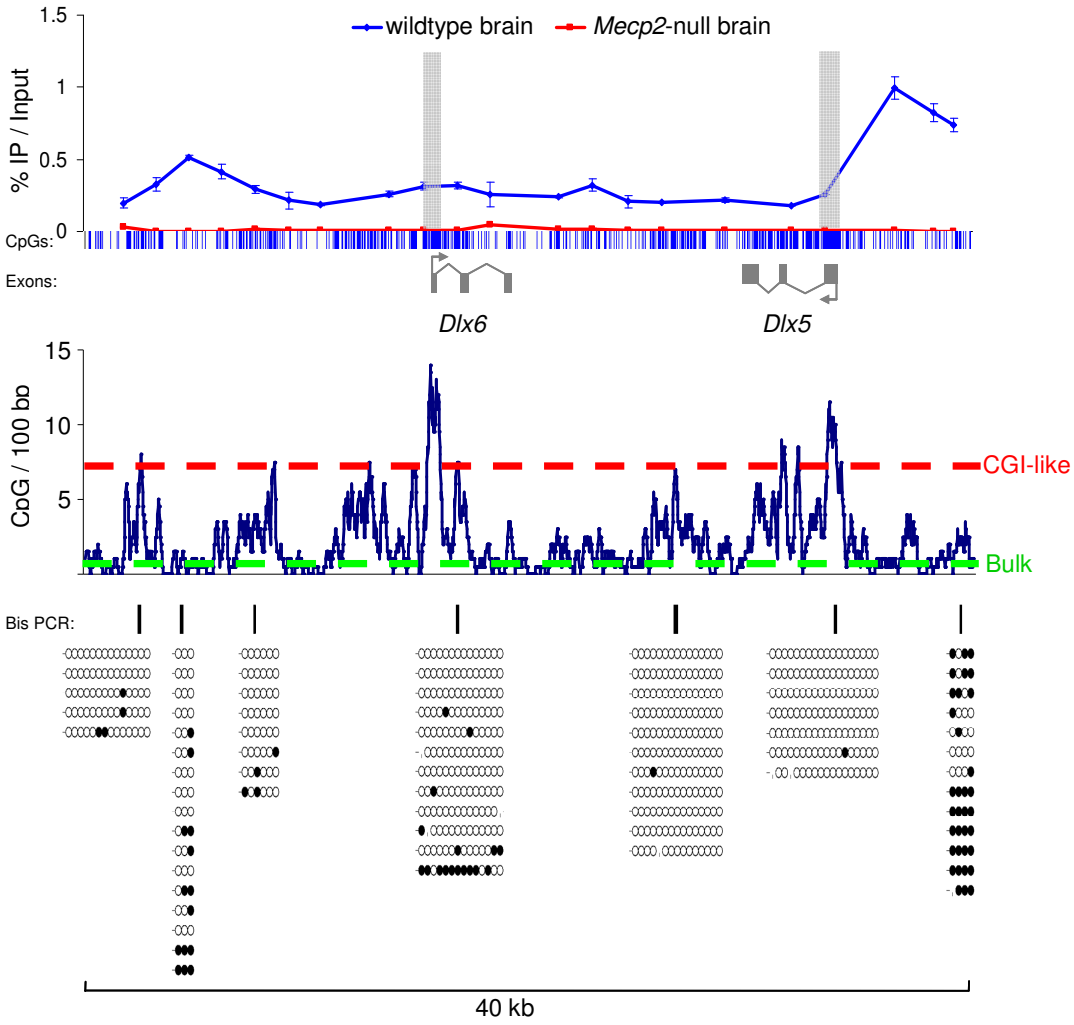

Figure 4.13 – MeCP2 shows a background level of non-specific DNA binding. (A) Mouse brains were dissected and ChIP performed using an antibody against MeCP2. Immunoprecipitated DNA was analysed by real time PCR using a panel of primers across 40 kb of the *Dlx5/6* locus. The blue vertical lines below the graph indicate CpG sites, with nonmethylated CGIs shaded in light grey as determined by CXXC affinity chromatography (data not shown; R Illingoworth). Exonic structure is indicated with dark grey rectangles. (B) CpG density across the locus was calculated using a 200 bp window; 10 bp step, plotted as number of CpGs per 100 bp (blue line). Typical density of mouse CGIs is indicated by a red line; typical density of the bulk genome is indicated by a blue line. (C) Total brain genomic DNA was subjected to bisulfite sequencing at specific sites as indicated above by black bars. Each line represents a single clone. Open and filled circles indicate nonmethylated and methylated CpG sites, respectively. Gaps indicate uncharacterised CpG sites.

Despite the lack of DNA methylation, there is again residual MeCP2 binding. This residual binding is significantly above the signal observed from the *Mecp2*-null brain, suggesting that this background is due to lack of specificity of the antibody. This background binding could,

however, be artefact due to over-crosslinking with formaldehyde (for discussion see section 4.7). Despite questions surrounding the crosslinking producing aretfacts, interestingly, this non-methyl dependent association showed no correlation with CpG density. Speculatively suggesting that this non-sequence dependent binding may be the result of non-specific interactions between the negatively charged sugar-phosphate DNA backbone and some part of the basic MeCP2, as suggested from the crystal structure (Ho et al., 2008; Wakefield et al., 1999).

4.7 Discussion

Previous studies have focused on the concept of MeCP2 binding to discrete sites, perhaps due to the historical model of MeCP2 functioning as a methyl-dependent transcriptional repressor for discrete target genes. However, the finding that MeCP2 is almost as abundant as the histone octamer in the neurons of the mature brain suggests that this discrete binding pattern cannot be true. Indeed, the above ChIP analysis has confirmed that in neurons, MeCP2 binds globally tracking the meCpG density of the neuronal genome.

Indeed, a re-evaluation of the literature supports this view. Early experiments indicated the particularly high abundance of MeCP2 within the brain (Lewis et al., 1992; Meehan et al., 1992; Nan et al., 1997). Despite this high abundance, the release of the majority of MeCP2 has repeatedly been shown to require high salt extraction even in the brain (Klose and Bird, 2004; Meehan et al., 1992; Nan et al., 1997). Therefore, suggesting that there is no clear unbound fraction of MeCP2 within the nucleus (Nan et al., 1997). The distribution of MeCP2 within the genome has also been examined by immunofluorescence. The use of mouse cells for studying MeCP2 localisation is dominated by the punctate staining to the pericentromeric heterochromatin (Lewis et al., 1992). However, rat nuclei exhibit a more even DNA distribution as indicated by staining for DNA with agents such as Hoescht 33258 (Lewis et al., 1992). Immunofluorescence of MeCP2 localisation in rat liver fibroblasts implied a global distribution along the length of the chromosomes (Lewis et al., 1992). This global distribution in rat cells has been more recently confirmed and extended with the analysis of monkey kidney COS7 cells and human cells (Nan et al., 1997). Therefore, despite our initial surprise as to the global binding pattern observed in neurons, where there is huge abundance of MeCP2, there is precedent for these findings within the literature. Furthermore, the conclusion that the majority of MeCP2 resides within the bulk genome is perhaps consistent with the observation that the repressive activity of MeCP2 has been tailored to

work at the low methylation density observed within the bulk genome (Nan et al., 1997). Also, despite the identification of a consensus binding site containing an A/T run being originally thought to target MeCP2 to specific site (Klose et al., 2005), this may actually reflect the fact that MeCP2 has evolved to bind to the bulk genome, where CpGs containing this motif are common, rather than in the rare methylated CGIs. However, MeCP2 ChIP-seq provided no support to the role of the A/T run in defining MeCP2 distribution *in vivo*. This difference may be due to a number of reasons. Firstly, this extra specificity in targeting provided by a relatively weak consensus may be lost in neurons, where the abundance is so high that almost all meCpG sites are occupied. Secondly, this could be the result of an *in vitro* artefact, for example, A/T runs are known to induce a bend in the DNA, similarly to wrapping of the DNA around a nucleosome. Therefore, the identification of an A/T run may have simply been a reflection of the fact that the true binding target of MeCP2 is meCpG within the context of chromatin (for a discussion of MeCP2 binding to chromatin see section 5.2). More work is required to formally disprove the role of an A/T run in the targeting of MeCP2, for example, analysing the ChIP-seq data for regions that show higher than expected MeCP2 occupancy given the CpG density and determining if these regions contain CpGs with A/T runs, or indeed other elements. This analysis would rely on the assumption that the bulk genome is entirely methylated, or may indeed require bisulphite sequencing of the brain genomic DNA. This study has elucidated the genome-wide binding of MeCP2 and is suggestive of a global function of MeCP2, it does not however, preclude the possibility that MeCP2 also has some site specific roles. Evidence for site specific roles could be found from the analysis of the ChIP-seq data for sites that are more enriched than would be predicted for the DNA methylation density.

As discussed there have been various reports proposing that MeCP2 binds independent of methylation (Chahrour et al., 2008; Georgel et al., 2003; Harikrishnan et al.; Kernohan et al.). The genome-wide ChIP data presented here reveals that MeCP2 occupancy is primarily determined by DNA methylation. The observation that there is some residual methylation-independent binding may be due to the almost-saturating levels of MeCP2 within neurons, as suggested by the failure to observe increased binding to the A/T run consensus motif. This hypothesis could be followed up by repeating the ChIP-bisulphite analysis in tissues expressing moderate levels of MeCP2, such as kidney, and low levels, such as liver, and monitoring any changes in the methylation status of the recovered clones. Previous work from this laboratory has shown that transient interactions are not efficiently crosslinked and that this can impact the results obtained by ChIP ((Schmiedeberg et al., 2009); see Appendix

B). Despite this study showing that wildtype MeCP2 is efficiently crosslinked; it perhaps raises the possibility that the residual signal in ChIP analysis could be the result of artefactual non-crosslinked interactions occurring after cell lysis. The mixing of crosslinked wildtype female extracts with crosslinked *Mecp2*-null male extracts and with the subsequent ChIP analysis for MeCP2 occupancy on the Y chromosome would confirm whether indeed there is some non-crosslinked reshuffling of the MeCP2 binding pattern. However, even if this was true, it would not fully explain the lack of methyl-dependence of such spurious interactions. It is also possible that the formaldehyde crosslinking conditions used (1% formaldehyde for 10 min at room temperature) could have 'over-crosslinked', whereby neighbouring DNA that was not specifically bound by MeCP2 became crosslinked and therefore was interpreted as being bound by MeCP2. In light of this possibility it would be appropriate for further work to vary the crosslinking conditions used and investigate whether this 'background' binding to non-methylated DNA is reduced, as would be predicted if this was an effect of over-crosslinking.

These considerations withstanding, it seems most likely that there is some weak non-specific DNA binding activity, as suggested from a solution structure of MBD of MeCP2 (Wakefield et al., 1999). Previous analysis of the localisation of MeCP2 in mouse cells indicated that the punctate staining observed was largely dependent on the presence of an intact MBD, identifying the importance of methylation in binding (Nan et al., 1993). However, it was noted in a minority of cases (~20%), that cells expressing a truncated form of MeCP2 lacking the MBD, retained punctate staining (Nan et al., 1993). It has been hypothesised that this residual binding may reside with the AT-hook present in MeCP2 (Nan et al., 1993). Therefore, the non-specific binding observed in ChIP analysis presented here, may be a result of this alternative weaker binding activity and therefore there is scope for a role of MeCP2 in binding to nonmethylated sequences as indicated by various studies (Chahrour et al., 2008; Georgel et al., 2003; Harikrishnan et al.; Kernohan et al.). However, as the majority of missense mutations causing Rett syndrome are within the MBD domain (Figure 1.6) and the fact that methylation is the primary determinant for this observed binding pattern, then the methyl-CpG binding activity is likely to be key to its function. Also, the ChIP analysis perfomed to indicate binding to non-methylated target sites was very limited, with only a single primer pair used to analyse the binding (Chahrour et al., 2008; Georgel et al., 2003; Harikrishnan et al., 2010; Kernohan et al., 2010), raising questions as to the thoroughness of the work.

Chapter 5 Impact of Global MeCP2 Binding

5.1 Global changes in chromatin modifications

Analysis of the effect of MeCP2 has typically been performed on an individual locus basis, in line with previous thinking that MeCP2 functioned as a classical transcription factor. Consistent with earlier transfection experiments indentifying the role of HDACs in MeCP2 transcriptional repression (Nan et al., 1998), elevated levels of histone acetylation have been observed at the promoter of the human multidrug resistance gene (*MDR1*) upon the loss of MeCP2 due to 5-azadeoxcytidine treatment in cancer cells (El-Osta et al., 2002). Elevated histone acetylation has also been observed at the *p16* tumour suppressor gene promoter in the absence of MeCP2 binding in cancer cells (Nguyen et al., 2001). However, the role of MeCP2 in regulating these specific genes has not been confirmed in expression studies using the various *Mecp2* mouse models (Jordan et al., 2007; Nuber et al., 2005; Tudor et al., 2002; Urdinguio et al., 2008). These findings question the relevance of these genes in the aetiology of Rett syndrome. This inconsistency may be a reflection of the atypical methylation patterns observed in immortalised cell lines (Antequera et al., 1990). Despite questions over the significance of these genes, elevated histone acetylation in the absence of MeCP2 binding is consistent with a mouse model expressing a truncated version of MeCP2 lacking the TRD, where moderately increased histone H3 acetylation (H3Ac) was observed (Shahbazian et al., 2002a). As these mice exhibit a Rett-like phenotype similar to the *Mecp2*-null animals (Shahbazian et al., 2002a), it suggests that this ability to recruit HDAC complexes may be imperative to the function of MeCP2.

The effect of MeCP2 has to be re-evaluated in light of the findings that in neuronal nuclei MeCP2 is expressed at near-nucleosomal levels and binds across the entire genome. These considerations raise the possibility that MeCP2 may globally influence the chromatin state. As approximately 89% of MeCP2 is within neurons, any global changes in histone acetylation between wildtype and *Mecp2*-null brain would likely be restricted to the neurons, but be absent in glia. To test this, H3Ac levels were quantified by western blotting in unsorted nuclei, FACS-purified neuronal and glial nuclei from both wildtype and *Mecp2*-null brain (Figure 5.1). Unsorted *Mecp2*-null brain nuclei displayed a 1.5-fold increase in H3Ac compared to wildtype. This difference increased significantly to 2.6-fold in sorted neuronal nuclei (Kolmogorov-Smirnov [KS] test $p \leq 0.01$), whereas glial nuclei consistently showed no significant increase in H3Ac (0.9 fold decrease).

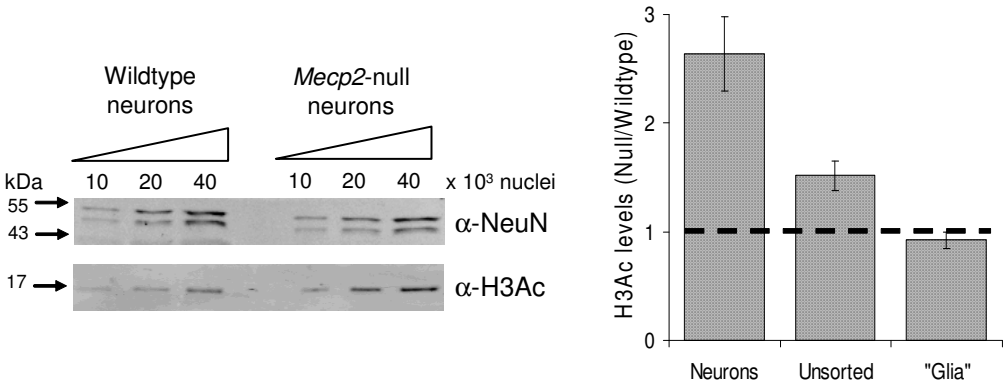

Figure 5.1 – MeCP2-deficiency affects the global chromatin state by elevating levels of histone H3 acetylation (H3Ac). H3Ac levels were determined by western blotting of unsorted brain nuclei and FACS purified nuclei from both wildtype and *Mecp2*-null brain, the antibody specifically recognises H3K9AcK14Ac. The western blot shows H3Ac levels and NeuN (as a loading control) in the purified neuronal nuclei. The graph indicates densitometric analysis of H3Ac levels in the different nuclei populations, normalised for differences in loading. The horizontal line represents no change between wildtype and *Mecp2*-null. Error bars indicate mean ± SEM from at least two biological samples; the KS test was used to determine statistical significance.

As might be predicted this suggests that the small increase in H3Ac observed in whole brain is entirely due to the effect in neurons, where MeCP2 is highly abundant. This finding does not corroborate a recently published study indicating that H3Ac levels were increased in *Mecp2*-null *in vitro* cultured glial cells (Ballas et al., 2009). Additionally, the initial study showing elevated H3Ac levels in the truncated *Mecp2*-null mouse, suggesting that the effect was, surprisingly, largest in the spleen, which only expresses moderate levels of MeCP2 ((Shahbazian et al., 2002a); see also Figure 1.5B). This result may reflect the limited western blotting analysis performed in this by a non-quantitative technique.

The widespread chromosome binding by MeCP2 observed in neurons might be expected to modulate chromatin structure globally rather than at specific sites. ChIP using an anti-acetyl H3 antibody was used to detect alterations across specific chromatin domains in the MeCP2-deficient brain. Due to low recovery of neuronal nuclei via FACS (see section 3.5 for discussion), whole brain was used, although this probably leads to an under-estimate of changes in neuronal chromatin. Across the 39 kb *Bdnf* region, as expected H3Ac levels peaked over the active promoter CGI regions in both wildtype and null (Figure 5.2A), but were low in the flanking regions. Comparing wildtype and mutant brain, levels of anti-H3Ac immunoprecipitated DNA in the mutant were elevated by ~2-fold relative to input throughout the region (KS-test $p<0.004$). Interestingly, the magnitude of the effect mirrored

the pattern of MeCP2 binding (Figure 5.2B), being lowest where MeCP2 binds least (the CGIs), but highest where MeCP2 is relatively concentrated in the bulk genome. This is consistent with a simple causal relationship between binding of MeCP2 and the recruitment of HDAC complexes, as suggested from earlier immunoprecipitation experiments (Nan et al., 1998). A 1.4-fold elevation in histone acetylation was also seen in regions of the male *Xist* locus flanking the CGI (Figure 5.2C; KS-test p=0.03). This difference was notably absent over the heavily methylated *Xist* CGI itself, implying that repression of histone acetylation at this DNA promoter sequence is not solely dependent on MeCP2. Altogether 100 loci were examined covering 90 kb of genomic DNA by quantitative PCR and with an average of 1.4-fold elevation in H3Ac (Figure 5.2D; KS-test: $p<0.002$), comparable to the 1.5-fold increase observed by western blotting for unsorted brain nuclei (Figure 5.1).

These data indicate that the global chromosomal association of MeCP2 imposes a reduction in histone acetylation levels across the genome. This increase appears to be primarily in the bulk chromatin as opposed to CGIs, both nonmethylated CGIs, such as *Bdnf*, and methylated CGIs, such as *Xist*, which are likely to have redundant control mechanisms. However, H3Ac is predominantly associated with CGIs, as shown for the *Bdnf* locus and other regions (data not shown) and confirmed by western blotting analysis of CGI chromatin ((Thomson et al., 2010); see Appendix B). Therefore, the relatively modest 2.6-fold difference in H3Ac levels between wildtype and null neurons (Figure 5.1) is likely to be diluted by this effect, as confirmed by the ChIP analysis (Figure 5.2B). This could be further verified by the use of methyl-sensitive restriction enzymes to selectively remove nonmethylated CGI chromatin ((Thomson et al., 2010); see Appendix B), with subsequent analysis of the remaining bulk chromatin.

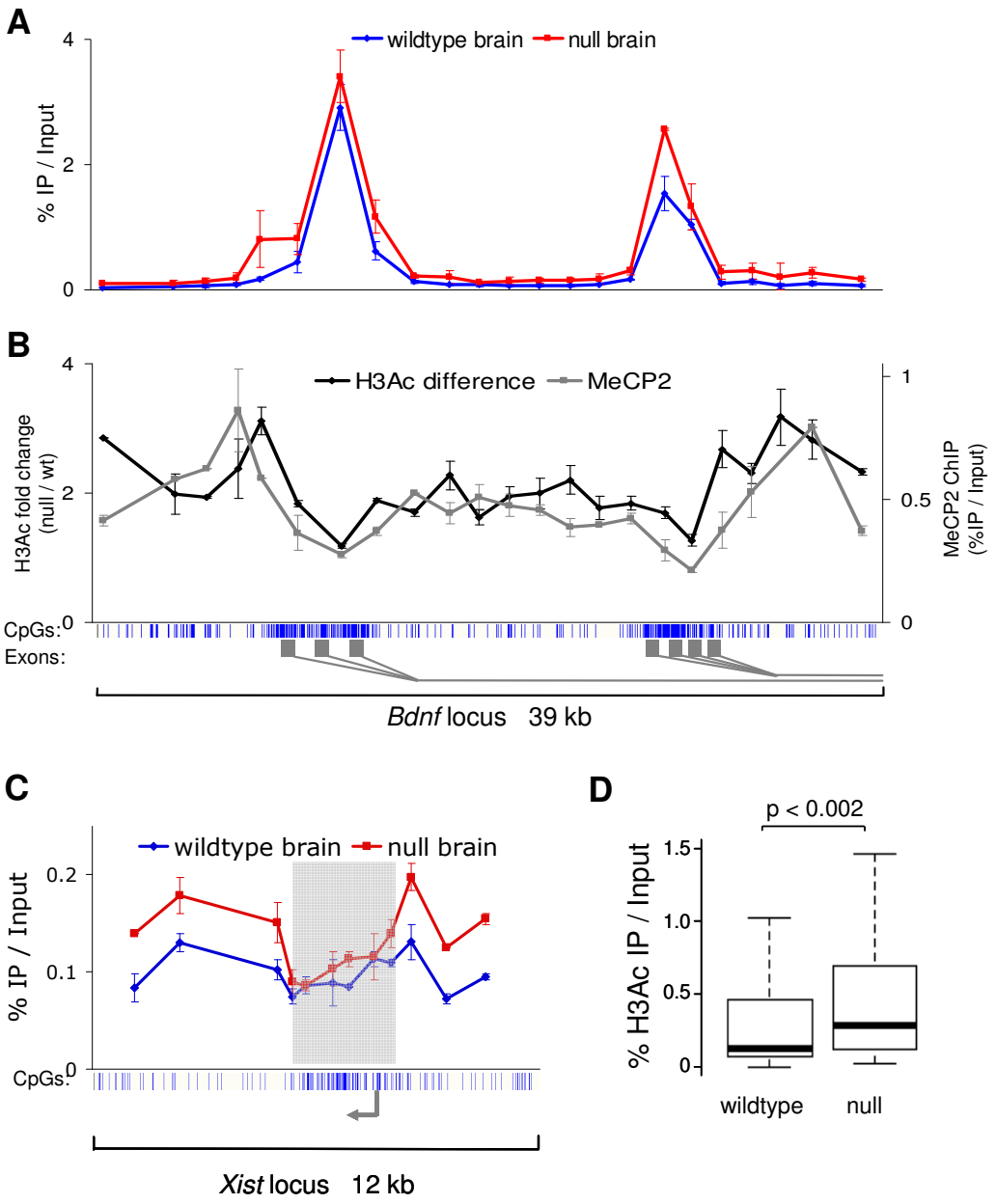

Figure 5.2 – MeCP2 deficiency results in elevated histone H3 acetylation primarily within the bulk genome. Mouse brains were dissected and ChIP performed using an antibody against H3Ac. Immunoprecipitated DNA was analysed by real time PCR using a panel of primers and ChIP signal plotted as % IP/Input. (A) H3Ac profile across 39 kb of the *Bdnf* locus in wildtype (blue line) and *Mecp2*-null (red line) brains. (B) The H3Ac fold difference (null divided by wildtype) across the *Bdnf* locus between wildtype and *Mecp2*-null brain (black line); the MeCP2 ChIP profile is also shown (grey line). (C) . H3Ac ChIP profile across the promoter region of *Xist* showing wildtype brain (blue) and *Mecp2*-null brain (red). The blue vertical lines below the graphs indicate CpG sites, with CGIs shaded in light grey. Gene structure of *Bdnf* is indicated with dark grey rectangles; transcription start site of *Xist* is indicated by an arrow. Error bars indicate mean +/- SEM and the KS test was used to determine statistical significance. (D) Boxplot showing all H3Ac ChIP results for wildtype and *Mecp2*-null brain (n=100).

A preliminary analysis of the impact of MeCP2 on other histone modifications has also been performed. A previous study suggested that MeCP2 can also associate with a histone H3 methyltransferase activity specific for lysine 9, with H3K9 methylation considered as a repressive mark (Fuks et al., 2003). Therefore, as when investigating the effect of loss of MeCP2 on H3Ac levels, H3K9me3 levels were determined in FACS-purified neuronal nuclei by western blotting (Figure 5.3). However, this analysis did not suggest a significant change. This may question the relevance of this interaction in the brain, or alternatively, the lack of a global effect may suggest that MeCP2 only recruits a H3K9-methyltransferase to specific targets and therefore global effects are not observed, in contrast to H3Ac where the effect is genome-wide.

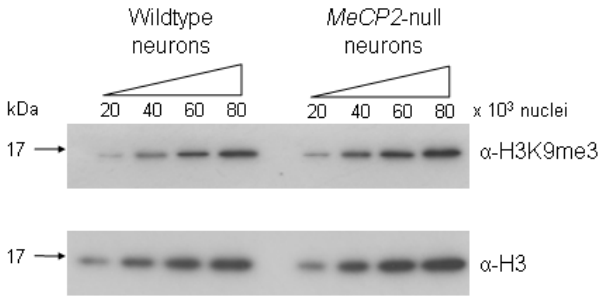

Figure 5.3 – Preliminary analysis indicates no detectable change in the global levels of H3K9me3 in *Mecp2*-null brain. H3K9me3 levels were determined by western blotting of FACS-purified neuronal nuclei from both wildtype and *Mecp2*-null brain; histone H3 was used as a loading control.

Additionally, changes in the level of histone H4 acetylation (H4Ac) have been investigated. Preliminary quantitative western blotting using unsorted nuclei suggested a modest 1.3-fold increase in the *Mecp2*-null brain (Figure 5.4A). However, preliminary ChIP analysis across the *Bdnf* locus failed to indentify any increase in H4Ac (Figure 5.4B). More work would be needed to see if any specific loci exhibit altered levels of H3K9me3 or H4Ac.

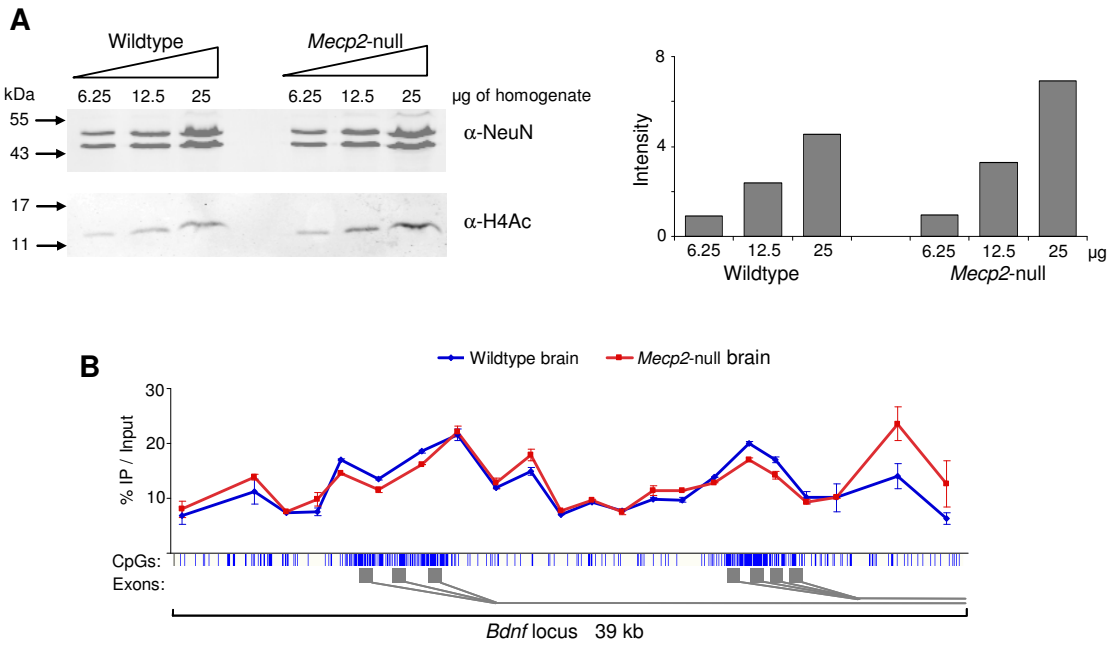

Figure 5.4 – Preliminary analysis of H4Ac levels in wildtype and *Mecp2*-null brain. Mouse brains were dissected and homogenised. (A) Quantitative western blotting was used to determine the level of H4Ac in wildtype and *Mecp2*-null brain homogenate; NeuN was used as a loading control. The positions of molecular weight markers (kDa) are indicated. The graph to the right depicts densitometry analysis, normalised for NeuN loading, using Licor Odyssey imaging software. (B) The homogenate was crosslinked using formaldehyde and ChIP performed using an antibody against H4Ac. Immunoprecipitated DNA was analysed by real time PCR using a panel of primers across 39 kb of the *Bdnf* locus in wildtype (blue line) and *Mecp2*-null (red line) brains.

5.2 Global changes in chromatin composition

Histone H1 is present in most cell types at an approximate stoichiometry of one molecule per nucleosome (Woodcock et al., 2006), but uniquely in neurons this is reduced to one molecule every two nucleosomes, with no comparable reduction in glia (Allan et al., 1984; Pearson et al., 1984). Chromatin reconstitution experiments showed that MeCP2 can compete with histone H1 for binding to methylated chromatin and may function as a substitute linker histone (Nan et al., 1997). Subsequently, various reports have investigated the mode of MeCP2 binding to chromatin. Using the 5S rRNA nucleosomal positioning sequence, MeCP2 was shown to bind to meCpG groups exposed in the major groove on the surface of the nucleosome, as determined by DNase1 footprinting, therefore binding in a manner distinct from histone H1 (Chandler et al., 1999). A more recent study, again based on the 5S rRNA sequence, used positioned meCpG sites along the length of the exposed linker DNA and reported that MeCP2 preferentially binds to the entry and exit sites from the octamer, in

a manner more similar to histone H1 (Ishibashi et al., 2008). This study noted that the DNA at the entry and exit sites forms a structure similar to cruciform DNA, which MeCP2 has also been shown to bind (Galvao and Thomas, 2005). A study using the 601 nucleosomal positioning sequence suggested that MeCP2 bound to the nucleosome asymmetrically, specifically protecting only one side of the linker DNA, in contrast to the observed symmetrical binding of histone H1 (Nikitina et al., 2007). However, this study failed to consider that the distribution of meCpG sites within the linker DNA is highly asymmetrical and this therefore may have influenced the asymmetrical binding of MeCP2. Overall, these various *in vitro* studies may suggest that there are various modes of MeCP2 binding to nucleosomes and that this may depend upon the sequence context, as has also been observed for histone H1 (Travers, 1999).

Given that MeCP2 can function as a putative linker histone and occurs at approximately one molecule per two nucleosomes in neurons, where earlier reports have shown histone H1 is specifically depleted (Allan et al., 1984; Pearson et al., 1984), this provides tantalising evidence for a functional relationship between these two proteins. Therefore, whether the abundance of histone H1 is affected by MeCP2 depletion was investigated. Using FACS purified neuronal nuclei, histone H1 was found to be elevated ~2-fold in the *Mecp2*-null compared to the wildtype (Figure 5.5A; KS-test $p<0.05$). No difference was seen in unsorted nuclei from wildtype and mutant whole brain, suggesting that this effect is specific to neurons where MeCP2 is highly abundant. The absence of an observed difference in the unsorted population is consistent with the studies reporting histone H1 is enriched in the glial nuclei (Allan et al., 1984; Pearson et al., 1984). As further confirmation for an increase in histone H1 in the *Mecp2*-null brain, the levels of histone H1 mRNA were investigated. There are currently six known somatic isoforms of histone H1. Isoforms H1.1 through H1.5 are highly homologous and expressed in a wide variety of tissues, whereas histone H1.0 shows greater sequence diversity and is predominantly expressed in terminally differentiated cells (Sancho et al., 2008). Expression analysis for H1.0 (Figure 5.5B) and H1.1-H1.5 (Figure 5.5C; differences in the individual isoforms could not be easily determined due to the high degree of sequence identity) was consistent with an increase in the *Mecp2*-null brain. However, expression analysis indicated only a mild 1.2-fold upregulation. This may be the result of having to use whole brain for expression analysis due to the limitations of FACS sorting of isolated nuclei, or indeed implicates the role of post-transcriptional mechanisms in the control of histone H1 levels. Overall, the level of histone H1 present in MeCP2-deficient neurons appears to be close to one molecule per nucleosome, as seen in other cell types,

suggesting that the reduced level in neurons may reflect substitution of histone H1 by MeCP2. It should be noted that it is not clear which H1 isoforms this antibody recognises in western blot analysis (see Table 2.1); personal communication with Abcam suggests that all H1 isoforms should be recognised. It will be important, however, that future work addresses this by using antibodies specific to certain isoforms.

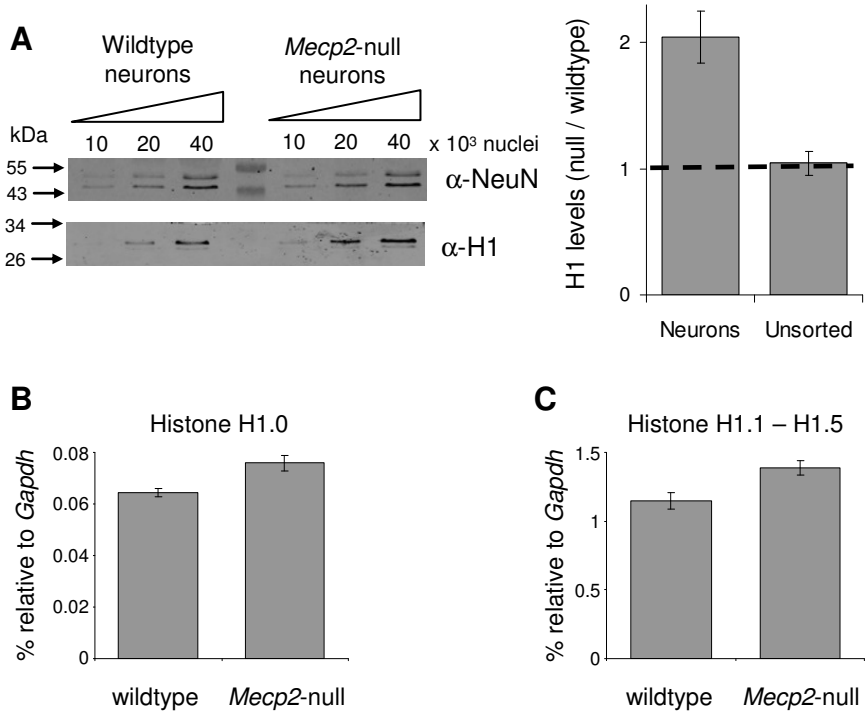

Figure 5.5 – Histone H1 levels are doubled in *Mecp2*-null neurons compared to wildtype neurons. (A) Quantitative western blotting for histone H1 and NeuN (as a loading control) using FACS-purified neuronal nuclei from wildtype and *Mecp2*-null mice. The graph indicates densitometric analysis of H1 levels in the different nuclei populations. RNA was extracted from whole mouse brain and cDNA prepared, with and without reverse transcriptase. Quantitative PCR was used to determine levels of expression of (B) Histone H1.0 and (C) variants Histone H1.1 through H1.5. The data was normalised and shown as % relative to GAPDH. Error bars indicate mean +/- SEM.

Eukaryotes exhibit a wide range of nucleosomal repeat lengths, with a robust positive correlation between histone H1 content and repeat length being observed (for review see (Woodcock et al., 2006)). Interestingly, glia cells display a nucleosomal repeat length of 200 bp in line with their H1 content, whilst neurons have been shown to exhibit a shorter nucleosomal repeat length of ~165-168 bp which is shorter than predicted by their H1 content, despite neuronal chromatin still retaining the ability to fold into a higher order

structure typical of H1-containing chromatin (Allan et al., 1984; Pearson et al., 1984). Limited MNase digestion was used to compare the nucleosomal laddering pattern from wildtype and *Mecp2*-null brains. Initially, FACS-purified neuronal nuclei were used to avoid any diluting effects of the glial nuclei. However, preliminary analysis indicated that the laddering pattern observed was too diffuse to allow accurate comparison between the samples (data not shown). Therefore, in order to isolate a pure population of neurons from an intact brain, the striatum granulosum of the dentate gyrus was dissected from both wildtype and *Mecp2*-null mice (assistance in the dissection was gratefully received from Dr S. Cobb). Limited MNase digestion did not indicate any clear difference in the laddering pattern obtained (Figure 5.6). However, as the bands are not particularly sharp it is possible that this reflects heterogeneity within the repeat length. This may be overcome by future work looking at discrete loci by southern blot. For example, the tandemly arrayed major satellite repeat may be more likely to have phased nucleosomes that would produce sharper bands and therefore allow a more precise comparison between wildtype and *Mecp2*-null neurons. Currently, however, the *in vivo* role of MeCP2 functioning as a linker histone remains enigmatic. More precise measurements of nucleosomal repeat length would be required to conclude if there is any difference or not.

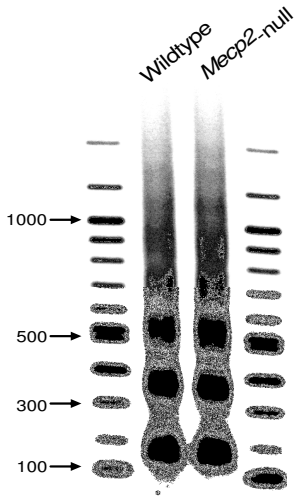

Figure 5.6 – There is no difference in the nucleosomal laddering pattern observed between the striatum granulosum of the dentate gyrus between wildtype and *Mecp2*-null mice. Nuclei from the striatum granulosum were prepared using a hypotonic buffer and then subjected to limited MNase digestion. Subsequently, the DNA was extracted and end labelled with γ ^{32}P-ATP and resolved by agarose gel electrophoresis. DNA was visualised by phosphor storage screen.

5.3 Transcriptional noise

Previous gene expression studies using microarrays have failed to identify clear MeCP2-gene targets. However, this study indicates that MeCP2 functions as a global regulator of chromatin structure, thereby questioning the concept of MeCP2 gene targets. Histone deacetylation is thought to aid the formation of repressive chromatin, in part through the formation of higher order structures (Shogren-Knaak et al., 2006). Linker histones have been shown to specifically repress non-promoter driven transcription, with limited impact in authentic promoter-driven transcription (Laybourn and Kadonaga, 1991). Therefore, through the methyl-dependent binding of MeCP2 throughout the genome and the recruitment of HDAC complexes coupled with the putative role as a linker histone, MeCP2 may act to create a transcriptionally inert environment primarily within the bulk neuronal genome. This is consistent with previous models suggesting a role of low density DNA methylation as a mechanism for suppression of spurious transcription ((Bird and Tweedie, 1995; Boyes and Bird, 1992); see section 1.4.3 for discussion). To test this possibility, the expression of L1 retrotransposons, intracisternal A particles (IAP) and the tandem repetitive units of the mouse major satellite, which are methylated and bound by MeCP2 (Figure 5.7) was investigated.

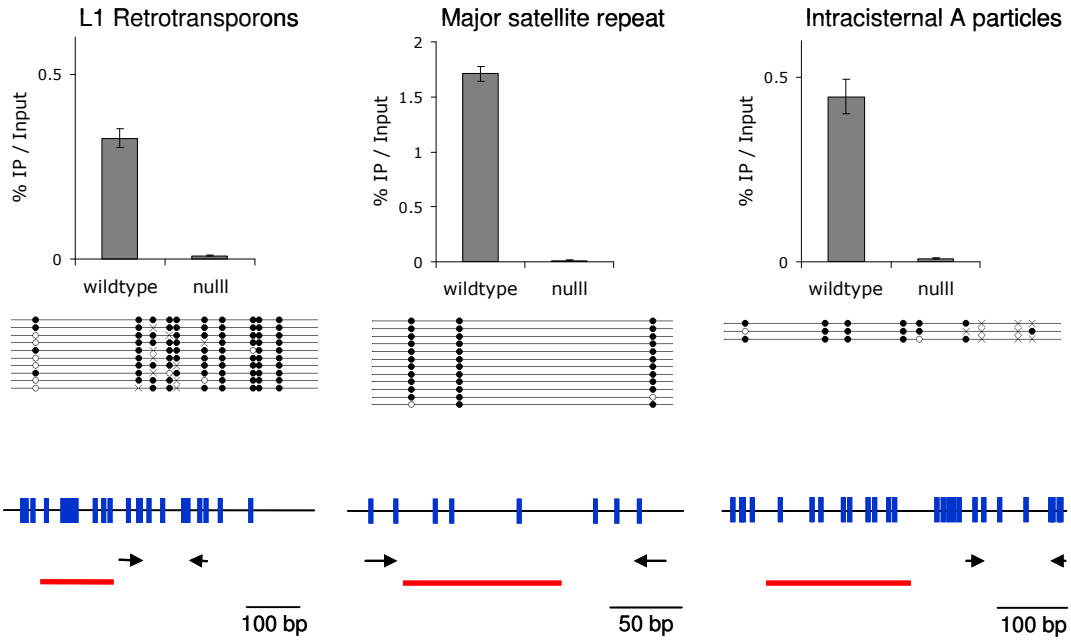

Figure 5.7 – Repetitive elements are distributed throughout the genome and are largely methylated and bound by MeCP2. MeCP2 ChIP followed by quantitative PCR using wildtype mouse brain shows significant binding to 3 classes of repeats as indicated. *MeCP2*-null brain was used as a negative control. Bisulphite sequencing of mouse genomic DNA indicates that these repeats are largely methylated. Each line represents a single clone. Open and filled circles indicate nonmethylated and methylated CpG sites, respectively. Crosses indicate uncharacterised CpG sites. A schematic of the region is indicated: the blue vertical lines mark the position of CpG sites; black arrows mark the position of the primers used for real-time PCR analysis; the red bar indicates the bisulphite PCR amplicon; a scale bar is indicated. It should be noted that these repetitive elements are distributed throughout the genome and it is possible that there is some variation in the level of methylation that is not fully represented by this limited number of clones.

RNA was extracted from whole brain and used to measure the expression levels of these repeats. No significant difference was observed, however, between wildtype and *Mecp2*-null brains (Figure 5.8).

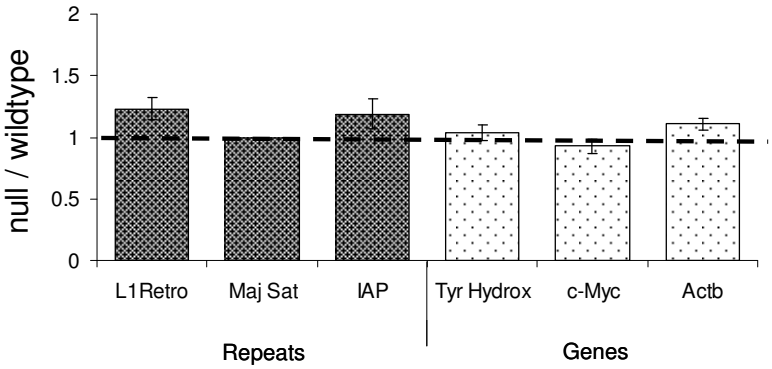

Figure 5.8 – Using RNA extracted from a whole brain there is no difference in the expression of repetitive elements between wildtype and *Mecp2*-null mice. RNA was extracted from whole mouse brain and cDNA prepared, with and without reverse transcriptase. Quantitative PCR was used to determine levels of expression of repetitive regions and gene regions. The data was normalised to *Gapdh* and shown as a ratio between *Mecp2*-null and wildtype mice. The horizontal line marks no change between wildtype and *Mecp2*-null mice. Error bars indicate SEM +/- mean.

It is possible that spurious transcripts might be degraded by the exosome and therefore not survive as stable components of cytoplasmic RNA. To reduce the opportunities for transcript degradation, RNA was extracted directly from isolated nuclei and analysed by RT-PCR. The same repetitive regions now showed significantly increased levels of expression in the *Mecp2*-null brain compared to the wildtype. In contrast, no differences in the expression levels of *Actb, c-Myc* or *tyrosine hydroxylase* mRNAs were observed (Figure 5.9A). On average, repetitive regions showed 1.6-fold overexpression in the *MeCP2*-null brain (4 biological replicates; KS-test $p<0.0002$), whereas the expression of these *bona fide* genes was unaffected (Figure 5.9B). This effect was not seen in embryonic brains from E18.5 mice (Figure 5.9C), suggesting that increased transcriptional noise is confined to the mature brain where MeCP2 is abundant enough to impact the whole genome.

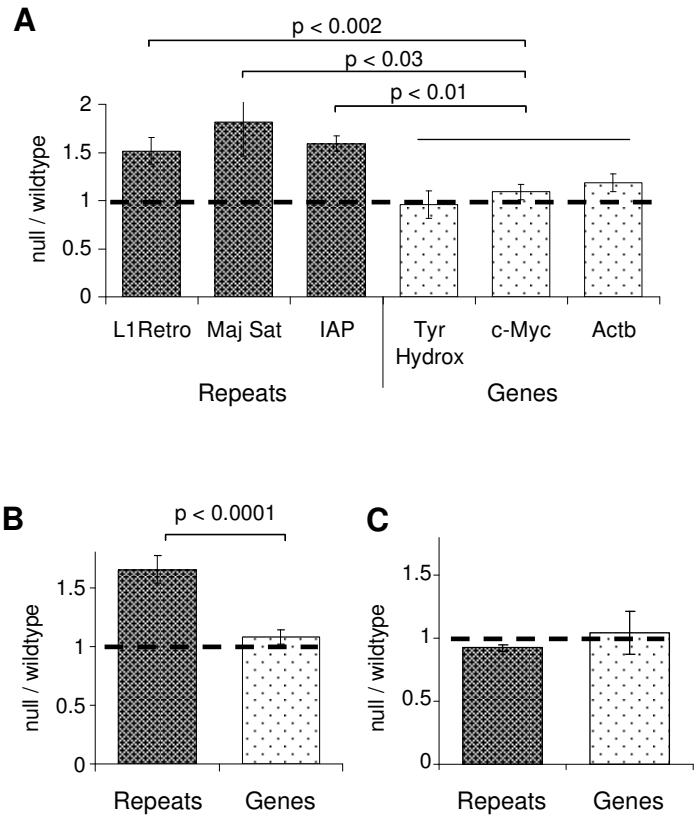

Figure 5.9 – MeCP2 suppresses transcription from repetitive elements distributed throughout the genome. Nuclei were isolated from wildtype and *Mecp2*-null brains. RNA was extracted and cDNA was prepared, with and without reverse transcriptase. (A) Quantitative PCR was used to determine the expression levels of repetitive regions and genic regions in the brains of mature mice. The data was normalised to *Gapdh* and shown as a ratio between *Mecp2*-null and wildtype nuclei. The horizontal line represents no change between wildtype and *Mecp2*-null mice. (B) Shows the pooled data from (A), grouping repetitive regions and genic regions. (C) Elevated expression of repetitive elements is only apparent in mature mice, where MeCP2 is highly abundant. Nuclei were prepared from wildtype and *Mecp2*-null E18.5 embryonic brains. Data presented as for (B). Error bars indicate SEM +/- mean. The KS test was used to determine statistical significance.

5.4 Discussion

In line with the global distribution of MeCP2 within mature neurons, there are global changes in histone modifications and chromatin composition. Elevated histone acetylation in MeCP2-deficient whole brain samples has been reported previously (Shahbazian et al., 2002a). By sorting neuronal from glial nuclei, this difference is shown to be entirely due to a large change in neurons, consistent with the fact that 89% of MeCP2 is expressed within these cells. ChIP analysis indicates this increase is distributed throughout the genome, rather

than at distinct sites. Early evidence suggested that MeCP2 can recruit the HDAC-containing complexes (Nan et al., 1998). A plausible hypothesis is therefore that MeCP2 attracts co-repressors throughout the neuronal genome, which depress histone acetylation according to the local density of meCpG sites. The finding that the ratio of mutant to wildtype H3 acetylation matches the density profile of MeCP2 provides support for this model (Figure 5.2B). Previous findings have indicated that histone H1 is surprisingly depleted within neuronal nuclei (Allan et al., 1984; Pearson et al., 1984). The data presented here shows that in the absence of MeCP2, neuronal histone H1 approximately doubles, with no comparable effect in whole brain and therefore presumably the glia level of H1. Previous studies have suggested a putative linker histone function for MeCP2, by virtue of its ability to compete with histone H1 binding to chromatin (Nan et al., 1997) and in its mode of binding to nucleosomes (Chandler et al., 1999; Ishibashi et al., 2008; Nikitina et al., 2007). It is tempting to propose that competition for chromatin binding sites between neuronal MeCP2 and H1 displaces the linker histone from half of its potential sites, leading to a reduction in the requirement for H1. In the *Mecp2*-null brain, where this competition does not occur, H1 is apparently restored to its conventional nucleosomal stoichiometry of 1:1 and may functionally compensate in part for MeCP2-deficiency. Further work is needed to clarify the role of MeCP2 as a putative alternative linker histone. However, it is possible that the TSA-independent transcriptional repression imposed by MeCP2 in targeted transfection studies could be due to its action as a linker histone (Nan et al., 1998).

Through the use of FACS purified nuclei, it has been possible to propose that global effects of MeCP2-deficiency on both H1 abundance and histone acetylation are seen exclusively in neurons because only in these cells is MeCP2 sufficiently abundant to coat the genome. In other cell types, the low density of MeCP2 occupancy is insufficient to impact bulk chromatin. However, given the duration of the FACS protocol and inherent RNA instability, it is not possible to determine if the observed increase in transcriptional noise is only within neurons. The observation that nuclei from embryonic brains do not show this effect, coupled with the developmental increase in MeCP2 expression, suggests that this increased transcriptional noise will also be specific to neurons. Currently, the analysis on the effect of transcriptional noise has been limited to repetitive regions. Despite this limited analysis, it seems feasible that as MeCP2 is bound globally, repetitive regions are no more silenced by MeCP2 than other methylated unique sites throughout the bulk genome. However, there is currently no evidence for increased transcriptional noise from single copy sites. As the effect only became apparent on isolation of nuclear RNA, with the suggestion that this was due to

removing the cytoplasmic component of the exosome, it is possible that the nuclear exosome is still suppressing the scale of this effect. Therefore by visualising nascent RNA directly through transcription run-on assays it may be possible to increase the sensitivity of this assay and indeed investigate transcriptional noise arising from throughout the genome.

Chapter 6 Conclusions

6.1 Global roles of MeCP2

The identification of MeCP2 as a transcriptional repressor led to the prevailing hypothesis that Rett syndrome was the result of failure to regulate specific gene targets. However, both mouse models and human diseases indicate that the precise abundance of MeCP2 is imperative to its correct function, as both deficiency and mild two-fold overexpression result in a neurological phenotype (Amir et al., 1999; Chen et al., 2001; Collins et al., 2004; Guy et al., 2001; Lubs et al., 1999; Luikenhuis et al., 2004). It is unlikely that such moderate overexpression of a canonical transcription factor, which binds to discrete targets, would lead to a phenotype similar to the loss of the factor. The observation that MeCP2 is almost as abundant as the histone octamer and through genome-wide methylation dependent binding it globally effects chromatin modifications and composition may explain why the abundance of MeCP2 is crucial. It is conceivable that even two-fold overexpression of global regulator, such as MeCP2, could have severe consequences, as implicated by the precise regulation of histone H1 stoichiometry (Woodcock et al., 2006).

The requirement for DNA methylation in the binding of MeCP2 has been of somewhat contentious over recent years. Various studies have observed methyl-specific binding of MeCP2 (El-Osta et al., 2002; Lorincz et al., 2001; Nan et al., 1996; Nguyen et al., 2001), whereas other publications reported binding to nonmethylated DNA (Chahrour et al., 2008; Georgel et al., 2003; Harikrishnan et al., 2010; Kernohan et al., 2010; Nikitina et al., 2006; Yasui et al., 2007). Using ChIP-seq with a highly specific antibody in a biologically relevant model system, we have shown that the global binding pattern of MeCP2 is determined by the meCpG density of the underlying DNA. This methylation-dependent binding may account for the remarkable reversibility of severe neurological symptoms in mice when MeCP2 is restored following development under MeCP2-deficient conditions (Guy et al., 2007). If, as seems likely, DNA methylation patterns are laid down normally in the absence of MeCP2, then the restored protein is expected to distribute correctly according to these genomic marks and take up its usual function.

The global distribution of MeCP2 casts new light on gene expression studies which have been performed in order to identify gene targets. These studies initially detected few significant changes between *Mecp2*-null and wildtype mouse brain (Jordan et al., 2007;

Nuber et al., 2005; Tudor et al., 2002). But with the development of more sensitive microarray-based technologies, large numbers of subtle alterations have been observed upon the loss of MeCP2, comprising both of increases and decreases in gene expression (Ben-Shachar et al., 2009; Chahrour et al., 2008). These studies have suggested the surprising conclusion that MeCP2 is primarily a transcriptional activator, in part through interaction with CREB. These studies however focused on the idea that MeCP2 binds to discrete sites, with the view that when a transcriptional regulator affects a restricted number of genes, coherent expression changes at those genes are expected when the protein is deficient. If, however, a large proportion of all genes are affected due to the loss of a genome-wide factor, it is more difficult to predict how deficiency will impact expression. The histone deacetylase inhibitor trichostatin A, for example, causes dramatic hyperacetylation of histones, but, despite the clear association between histone hyperacetylation and transcriptional activity, the drug usually causes equal numbers of genes to be up- and down-regulated (Peart et al., 2005). A potential explanation is that the nucleus contains limiting supplies of transcriptional machinery, so that increased expression at some loci must be matched by decreases at others, even if all genes are potentially activated by hyper-acetylation. By analogy, the high abundance of MeCP2 may mean that many genes are potentially affected, but analysis of stable gene expression patterns following long term absence of MeCP2 may not accurately reflect these effects.

These findings question the concept of a MeCP2-target gene. The global distribution and effects of MeCP2 do not preclude the possibility that some genes are specifically regulated by MeCP2. However, the failure to identify a consistent list of genes showing altered expression in the *Mecp2*-null brain (Chahrour et al., 2008; Jordan et al., 2007; Nuber et al., 2005; Tudor et al., 2002), coupled with the observation that histone acetylation is primarily elevated outside of the CGIs, suggests that aetiology of Rett syndrome is not the result of misregulation of a limited number of genes. In line with a global role of MeCP2, this work has shown elevated transcriptional noise from repetitive elements distributed throughout the genome. This role of genome wide suppression of transcription is consistent with the earlier *in vitro* characterisation of MeCP2, where the repressive activity of MeCP2 was shown to be tailored to the density of the bulk genome (Nan et al., 1997). Furthermore, MeCP2 was shown to be able to repress transcription from distance of a least 2 kb (Nan et al., 1997), suggesting that MeCP2 has evolved to repress domains rather than discrete sites. This property would be consistent with the heterogeneous nature of the maintenance of domains of DNA methylation, whereby the precise pattern is not faithfully reproduced (for review see

(Bird, 2002)). Additionally, this work has provided preliminary suggestions that MeCP2 may act as a linker histone, as suggested from previous *in vitro* studies (see section 5.2 for discussion). It is premature to suggest that MeCP2 may function in a manner analogous to histone H1, but it is interesting that histone H1 is not primarily involved in the repression of promoter-driven transcription, but in the suppression spurious transcriptional initiation (Laybourn and Kadonaga, 1991). Further work will be required to fully evaluate the extent of genome-wide transcriptional noise in the absence of MeCP2, as indicated from this limited analysis of repetitive sequences. This does not preclude the possibility that MeCP2 is also involved in the regulation of specific genes, in line with the concept of 'MeCP2 target genes.' Given the importance of the precise high level of MeCP2 expression, however, it is likely that the genome-wide binding is most relevant to the aetiology. For example, this analysis has shown that MeCP2 is enriched at methylated CGIs, however, it is likely that there are redundant levels of control at these highly regulated genes. Indeed, MeCP2 has been shown to not be required for the silencing of *Xist*, at least in mouse tail fibroblasts (Barr et al., 2007).

6.2 Rett syndrome: failure of the quiescent or active state?

The near-nucleosomal abundance of MeCP2 in neurons raises questions as to why these cells have evolved a requirement for a genome-wide methyl-dependent transcriptional repressor, whereas other somatic cell types have not. One possibility is that the neurons are particularly sensitive to the levels of transcriptional noise from the bulk genome and therefore have a special need for MeCP2 to quench this. To my mind, a more likely explanation is that neuronal plasticity and homeostasis may depend upon the specific ability of the neuronal nucleus to allow gross changes in its genome metabolism in response to activity. Neuronal firing leads to a burst of synaptic protein synthesis, nuclear transcription and chromatin modification, which is required for development of synaptic systems (Bading, 2000; Gupta et al., 2010). MeCP2 undergoes site-specific phosphorylation at serine 421 following synaptic firing, which is reported to alter its affinity for DNA and its nuclear distribution (Chen et al., 2003b; Zhou et al., 2006). A further report indicates that post-translational modification of MeCP2 also occurs in the quiescent state, including phosphorylation of serine 80 (Tao et al., 2009), raising the possibility that MeCP2 is a signalling hub within neurons and as such may play a crucial role in neuronal maintenance. The functional role of these modifications is yet to be fully elucidated. It should however be noted that a knock-in mouse study has indicated that alanine substitution of serine 80, produces a less severe

phenotype than the null mouse, suggesting that other factors must contribute to the Rett phenotype (Tao et al., 2009). These neuronal specific effects on chromatin structure indicate that it is imperative that MeCP2 is investigated within this context, as opposed to a large number studies which have used cell lines. Additionally, the effect of neuronal stimulation needs to be considered, with an analysis of changes in MeCP2 distribution/modification, transcriptional activity and chromatin modification. Given the genome-wide distribution of MeCP2, it is likely that these analysis need to be performed on a global scale, rather than focusing on gene loci. These analyses will provide evidence as the role of MeCP2 in both the active and quiescent states of the brain.

6.3 The CpG dinucleotide as a signalling module

The mammalian genome exhibits a dimorphic distribution of CpG and meCpG dinucleotides, with CpG islands being conspicuous by their high density of CpG sites and being primarily nonmethylated, whilst the bulk genome shows CpG suppression and is largely methylated. It is clear that both genomic compartments have specific roles to play, as indicated by existence of proteins that can interpret these signals *in vivo*: MBD proteins binding to meCpG sites and CXXC proteins binding to nonmethylated CpG sites. This study has provided evidence that, specifically in neurons, MeCP2 is sufficiently abundant to bind almost every meCpG dinucleotide within the genome and in turn regulates the chromatin state. How meCpG sites are interpreted within other tissues and whether a transient sub-stoichiometric interaction with MeCP2 or the other MBD proteins is sufficient is currently unknown. Nonmethylated CGIs, of which 80% are potentially associated with a promoter (R. Illingworth; manuscript in preparation), are specifically bound by Cfp1 by virtue of the underlying DNA sequence, through which histone H3K4 becomes trimethylated ((Thomson et al., 2010); Appendix B). Overall, both meCpG and nonmethylated CpG sites recruit factors which specifically modify the chromatin state to create a transcriptionally inert or permissive state respectively. In this way, the global distribution of these marks may be crucial in determining the transcriptional capacity of the genome. DNA methylation has historically been considered as an additional level of silencing with a rather static role, rather than being strategic in the initiation of silencing, in part through the assignment of maintenance methylation to DNMT1 (see sections 1.3 and 1.4). It is however now clear that DNA methylation perhaps is more dynamic than initially thought, with the observation that in transformed cells, multiple signalling and transcriptional factors are required for the faithful propagation of methylation marks (Gazin et al., 2007). The dynamics of the CpG

dinucleotide as a signalling module and how it impacts the transcriptional landscape is likely to be of great interest in the coming years.

Appendix A: List of Primers

Primers were resuspended in water at a concentration of 100 μM and stored at -20 °C.

Primers for real time PCR analysis

Locus	*Amplicon name*	*Forward Primer*	*Reverse Primer*
Major satellite	Maj Sat	GGCGAGAAAACTGAAAATCACG	AGGTCCTTCAGTGTGCATTTC
Bdnf	-3	GGAACATCTCCACAGTAATTTGC	CCTAGAGGTTAAACAACTTCAGC
	-1	GGAGTCAGCCAGAGTTACCC	CACCACTTTTCTGCCTCATCCA
	0	GCAGCTGTGAAATCTTCCCTC	CCATGGAGTATGGAACTAGAGC
	1	GCTGACTAGAGTTGTGTGAGG	CCTGGTTCATTATCAGCTTTCC
	1.2	CGCAACACTGGATTTACTGCT	GCTGTGTTTGAAGGGAATTCG
	3	CGTTGAGAAAGCTGCTTCAGG	CCTTCGCAATATCCGCAAAGC
	4.2	CCTTTCTATCATCCCTCCCCG	GGCTCTTCGATCTAGAAAGGAC
	5	CGTCATCCTCTTACACCCAAG	GGTTAAATGTCTCTCTCTCTCC
	6	CCAACACTGTAAAAAGAATCCTG	CAGACCGTCCACTGTGGTATG
	7	GTGTCTTGTTCTTATAAATAATGC	GCAGGGAATGGCTTGCAGTC
	8	GCAAATACTACTGGGTAAGATCC	GCCTTTTACACACTGTGTAGCT
	9	CCAGAAATAAATATCTAACCCCAC	GGAAAGGCAAACTTATGAAATGG
	10	CACATGTTGAGTGTAGGACTTC	CCAGCCTTACTGAAGATGAAGG
	11	GCCTAGGCAGGACCAGGTC	CCATGTTTTAATGTGTGATAGTCG
	12	GTGGTAGAGAAAACATTGGCATC	GTGTCACTATTTAGCTCAAACTC
	13	CCACTCTGGTGACAACAGACA	GGCATATGTGACTCAGAAAGG
	14	CCACAGAACTTGGGTGCTGG	GCTGCTTAAACCCAAAGCTCTG
	15	GGTCCAAGGTCAACGTTTAAC	CATCCCTGGAAATCAAAACTAGC
	16	CCATTATTGGAATTCCAGAATCAT	GCACAACAGTCTTGTTTATCTGG
	17	GGAATCAGAAGAGAGTTTTGCAG	GCATTACTCCATACTAGACAGAA
	19	CCTAGACCAAAGGCTCCAAGG	CCACACACAGAAGTTAATTACAC
	20	CCTTAAACTAGAAGCCATGTAAG	CAGGAATATAGGCTTTGATCTTTG
c-Myc	1	CCAGAAGCTTTCCCAGCAAGC	CAGCTCAGCCTTGCTTGCTC
	2	GCATGCCATGGCTAGCTTGG	GAGAAGCAGGAGACCCTAAGG
	3	AGTGACTGAATGAATGCTGACC	TCTCATACTAACAGTCATGATGC
	4	GTGATGTCATCAGGCTGGGG	CACCAACAGATATGACAGTGGG
	5	AGGCAATAATAAGCTAATGCTCC	GTGAAGGAGATCTAACAGAAACG
	6	TGGTTAATAAGCTAGATTATCGTG	CCTTCGTATGTGTGTGTTAAGC
	7	CCTAGATAACTCATTCGTTCGTC	CCCTGCGTATATCAGTCACCG
	7.1	GACTCGCTGTAGTAATTCCAGC	GCAAAGCCCCTCTCACTCCA
	8	GCTGTTTGAAGGCTGGATTTCC	CAACTACTCTTGAGAAAAGTGTC
	9	CAGTGCTGAATCGCTGCAGG	CCGATTGCTGACTTGGAGGAG
	9.1	GCTCTTAGCAGACTGTATTCCC	GTCGGACTAGCAGCTGCTCC
	10	GGAAGAGAATTTCTATCACCAGC	ACATAGGATGGAGAGCAGAGC

	10.1	CCAGAGCTTCATCTGCGATCC	TCAGGCTGGTGCTGTCTTTGC
	11	GGCTTATCTTTCAGCTCCATCC	TGGGTCTTAGACAAACGTATCC
	12	GACACACAACGTCTTGGAACG	GTGAGCTTGTGCTCGTCTGC
	13	CTCAACCCAAGGACTCTGCC	CCAGGATCAACTTAGCAGTGG
	14	GCCTCTGAAAAGCCACAATCC	GAATGATGGTGAATTAAAGTCCC
	15	CTGGGATGGTGCTCTTCAGG	CTACATACGTGTTCAAGCTTTGG
	16	GGTCTATATGGCTAGGAAAGAG	GCTTATGTGACTTCAGGCCAG
	17	GGTTTCTCTGTGTCTGAGAACC	GTGGTAGGTAGTAAGTGTATCG
Actb	-1	CCATTTGACCTTCTTGTGTTGC	CCAGATCTTCCTTCATGGAACC
	-1.1	GATCCAGTGACCTGAAAAGACC	CCACTCTTGATTCTTTGCTTTCC
	0	CCAGTTTGCCTGCTGGACCC	CCATTTGAGAAATGGACACACC
	1	CCAAGCTAACCTCAGCCTTGC	CCAGTATCACTGTACATTGAGCA
	1.1	CCAGTTATGCAGATGGAGGC	CCTTGCTGATGGTATCTAGTGG
	1.2	CAGTACATAATTTACACAGAAGC	CCAAGTATCCATGAAATAAGTGG
	1.3	AAGCCATGCCAATGTTGTCTC	AGCAGATGTGGATCAGCAAGC
	2.1	TGAAGCTGTAGCCACGCTCG	CTGTCCCTGTATGCCTCTGG
	3	GCAGAAACTGCAAAGATCCAAG	CCACACCTTCTACAATGAGCTG
	3.1	CCAAAGTAACAGGTCACTTACC	GTGTCTTGATAGTTCGCCATGG
	4	CCTAATACGGCTTTTAACACCC	CCTGAGGATCACTCAGAACGG
	4.2	CGAGCACTTAAGTGGATGAGG	GCTTTCCGGCTATTGCTAGC
	5	CCTTGTCTGGAAGAGGTGACC	CTGAGCAACTGAGAAATACAAGG
	6	CCTAAGCTGCACATTTTCAAGTG	GCATTATGGCTGGATCTTAAAGC
Xist	-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
	7	GCTACTGCTCATAGGTAGGC	CATGATCTTTGGTAGATTGATTTC
Snrpn	-3	TCAGTCTAGGCTAAATGAATCC	GTTTTGTGGAGCAGGTTCTCTC
	-2	CATGTTAGGAACTGCAGAAACC	CAACTGTTAACTGGTCAGGTGC
	-1	AGGAGAAAGCCTCATCAAAGAG	CACACTGCAATTGCCAGACTTC
	1	CAGTCTTAAATCATGTACAATCG	CAGGACCTATTATTGGTGATGG
	3	GGACACACAATCACTTCTCTG	CTCTGTTGCTGCTTCATGTGG
	4	CACAGTAACAGTTACAAAATCCC	TGGAAGTCAGAGCTGTGTTGC
	5	CCTCAGAACCAAGCGTCTGG	TGCATTGCGGCAAAAATGTGC
	6	TCCTGATTCCTAATTCTACATTC	CCCTGTCTCTAAAACCAACAAC
	7	GTAACCAAAACCTGAACTTTCG	TGAGCTACACTCCACCATTCC
	8	CGTTAGATTTAGGAAGATTACAG G	CAGGAATAGAATGATCTCTCCC

Dlx 5/6	-2	TCCATGATCTCCTAGGTATTCC	TGAAAGCTATTCAAATGAATGTGC
	-1	GCACAAGGCAGCTGGTGAGC	GGCTCACGTGAATACTCTTGC
	0	TAGGTTCTGTCTGCAACCACC	TAGGATGGTAGACTGCTTGTCC
	1	CCTACCATGAAATTTACTTCCAC	AGGGTGCCAAGGTTCACTGG
	2	CGCTAACACCTATTTCACAACC	CCCTCTCCTTGAAAATCTTCCA
	3	CAAAGATGACACAGTCAAATGC	CTCACTCCTTTCCACCTCTGG
	4	CAGGTACCTATTCTTGGAAACC	CTCTAGCAATAAAACTGCAAAGC
	6	GGTTCTACCACCTGGATGTGC	CTCTAATTATACACCAGTGTCGC
	7	GAGCTAAGGTGGCTGCAAAGG	GGATTTGGACGAGTCCTGGC
	8	CTTACAGCGCCACGGACTTC	GCTATACCACTGTGGGCACG
	9	GTAATGCTACATTGTAGGTTTCC	CTACCTCCTATGTTGCTTAAGC
	11	CGATCAGTCTTGTCATTTTCTAGC	CCCAATGTCTGCTTCAAATTGG
	12	GAGACTACAGCTCTATGCAAGG	GCAGTCAGCTATGGGGATGG
	13	GCTCTACTGCGTCCTAGAGG	CGAATTGTGACATAGAAGCAACC
	14	GGTCAAAAATGTGCAAGGTCTG	CTCCACTCTCAAGCACCTTCC
	16	GCATAGGCTAGATATCTACATGC	CTCTTTCAGGATCTCCTGTGC
	18	CCATCCTCAGATCATACATGGA	GAAACACACAAGGTGAATACAGG
	19	GTGCTGTCCTGTTTGCAGTCC	CCCCTACCACCAGTACGGC
	21	CCTGAAGTGCTGAAAGACTGG	ACTCGAGGCTGTCTATAGTGG
	22	GGTCATGCGTTCTGTGAAGGC	CCATAGAACTGCAGTCACATGG
	23	CTTTGCAAACCAAAATCGTGGC	CACCTGACAGTTCCGTTCTCC

Primers for bisulphite sequencing

Bdnf	BDNF Bis F1R1	TGATAGAAGATTAGAGTTTTTTGATAA	ATTAATATTTTTATATATACAATCCCAACA
	BDNF Bis F2R2	AATAGAGTTTTTGAGTTTGGGGAA	CCTAACAAAAAATAAAAACCAAAAAAC
	BDNF Bis F3R3	GAGTGAATTTATTGTTTGGAAGAATA	AAAAACCAATAAAAAAAACATAATAAAC
	BDNF Bis F9R9	TTAATATAGTTGTTAGGGTTTTAGA	ACAATTAAAATACCTTTACTTTTAC
	BDNF Bis F11R11	AGTTTGTAGAAGGTGTGTGGGTATTTG	AAACTACTACAACAATCCAACCAATC
	BDNF Bis A	GGAAATATTTTTAGTATAATTTTGTGTTGT	TTCATCATACTATTAATAAATTCTCTTACC
	BDNF Bis B	GGTTTATTGGTGGGAATTTTAGAG	AAATCTTCCTTTACCTACCCTCAAC
	BDNF Bis G	GGTTGTTTAAAGTTGATGTTTGTTT	CACATATTAATAAAACCCTTTCCAC
	BDNF Bis H	GGTTGGAATAGATTTTTGGTAAGTTT	CAAAAAAATTAAACTTCCTCCACC
	BDNF Bis I	ATTTAGGTGAGAATTTGGGGTAAAT	AATACAAAATCAAATCTTAAAAAATC
	BDNF Bis J	GGTTTTTGTTTTTTAAGGGAAGG	AAACTACCAAACCACCCAAATAAT

			A
	BDNF Bis F10R10	TTTTAATATTGTAAAAAGAATTTTG	CCTTAAAACTAAAAAAAACAAACT
		TAAGG	CC
	BDNF Bis F14R14	AGTTGTTATAATTGTTGTTATGTT	AAAATATTCTCCATCAAAACATCT
		TTAGAT	ATAATA
Actb	Act Bis F1R1	TTGGTTGTTTTGGAATTTATTTTG	CCTCCCAATACATACTACATTTAA
		T	CC
	Act Bis F3R3	TTGTGAGGTTGAGTAGATGTTTTT	CTACATAATAAATTAAAAACCAAC
		G	CTAAAC
	Act Bis F4R4	TGTTGGAGTTTGGTTTTGAATAGT	TAACAAAAACCCAAATAACACAAA
			C
	Act Bis F5R5	ATAGTGTTTAGTTTGATGGGGAA	CCAAAAAAAAACAAAAACCTTACTA
		AA	AAA
	Act Bis F6R6	TGTTTTTTGTATGTGTGAATAAAT	CCAAAAACCTAAACCAAATTAACC
		GATT	TA
	Act Bis F7R7	AGGGAGGTTTTAGATTTGGGTTA	TCCCCCAAAATTCTACAAATATAA
		T	CTA
	Act Bis F8R8	GGATGTTTGTTTTAATTAATTGTT	CTTCAAATCATTACTCCTCCTAAA
		GT	C
	Act Bis F9R9	TTAGTAATGTTTGGGTATATGGTG	CATCCTCTTCCTCCCTAAAAAAA
		GTAT	
	Act Bis F10R10	AAGGAAGGTTGGAAAAGAGTTTT	AATCACCCACACTATACCCATCTA
		AG	C
	Act Bis F11R11	TTTAGTTGTGGTGGTGAAGTTGT	ACACTCCTTACATATCTCAAATCT
		AG	ATCC
	Act Bis F12R12	GTTTAGTGTGTTGGGAGTTTTAG	CCTTCTCTTTAACCAACTTCTCAA
		GA	C
	Act Bis F13R13	TTGGGAAAGAGTAGAAATTGTAA	ATCCTAACCCTAAAATACCCCATT
		AGAT	A
	Act Bis F14R14	GTGTGGTGTTAGATTTTTTTTATG	AACCTTCTTTTATATCTTAATAATT
		T	C
	Act Bis F15R15	TTTAAAGTAATAGGTTATTTATTT	AAATCTAACTTCCTACCCTAAATC
		GGTGTT	C
	Act Bis F16R16	GTTTAAGTTTAGGGGATAAAGGA	TAACCTCTAAATCTTTATCCAAAC
		AG	C
	Act Bis F17R17	GGATAAAGATTTAGAGGTTATTGA	AACACAACTTCTTTACAACTCCTT
		GG	C
	Act Bis F18R18	GAAGGAGTTGTAAAGAAGTTGTG	TAACCCCCAAAATACAAACCTAAT
		TT	A
	Act Bis F19R19	ATTATTTTAGGATTTTTTGGGTGT	CAACACCTCCAATTATTAAACCAC
		G	TA
	Act Bis F20R20	GTGGTTTAATAATTGGAGGTGTT	CACCAACTTTTCCCTAAACAACTT
		GTA	A
	Act Bis F21R21	TGATAAGTTGTTTAGGGAAAAGTT	AAAACCTTCTTTAATTTCAACACC
		G	TAC
	Act Bis F22R22	AAGTTTTTTGTAGTTATTAATTTTG	CAAACAAAACAAACCAAAATTATT
		AGAGA	TC

	Act Bis F23R23	TTTTAGTTTTGGGATTGGTTAAGG	AAAAACCACTATTTCAAACAAAAAAAC
	Act Bis F24R24	TGTTTTTTTTGTTTTAGTTTTTTGTGT	AACCAACTTCACTTCCTCTACCTT
	Act Bis F25R25	TTTGGTTTAGAGTATGGGTGATATTT	AAAAACATTATAACTAAATCTTAAAACACT
	Act Bis F26R26	GTATTGGGAGGGTTATTTTTTTATT	CCTACAAACCACCCAAAACTTATAC
	Act Bis F27R27	TTAATTGGGGTAGAATTTGTAAAGG	CCAACCTAAACATTAAAAACATTATCTC
	Act Bis F28R28	TTTGGGATTTTTTAATGGTAAAGAT	AATAACACAAACACCAAAACAAAAC
	Act Bis F29R29	TAGTGATAGGTTTTGTTGGAGTTTG	AATTACCATTTAAAAAATAAACACACCA
	Act Bis F30R30	GGTTAGGATATTTTTTTTGTTTTGG	CAAACCTAAAATTTTCTTAAAAATC
	Act Bis F31R31	AATTTTGGTTTGTTTTGTTTGTAATG	AAAACCCCTCAAAAAACTAACTTATTT
	Act Bis F32R32	GGAGGTTTTAGATTTGGGTTATTTAG	AATAATTACAAAAAATCCCTCACCC
	Act Bis F33R33	GTTATGTTAATGTTGTTTTTTATTTTT	TAACTCCATCCTAACCTCACTATCC
	Act Bis F35R35	TATATTTGTTGGAAGGTGGATAGTG	TAAATTTAAAAAAAATCCCAACACC
	Act Bis F36R36	ATATATTGAGGGGTTTTGTTTTGTG	AAACCAACTTCACTTCCTCTACCTT
	Act Bis F37R37	TTGGTTTAGAGTATGGGTGATATTTG	CTATAACAAAAAATTCCCCTCTTCC
	Act Bis F38R38	AGGAAGAGGGGAATTTTTTGTTATA	TCAAACCCATCTCAATTTTCTAAAC
	Act Bis F39R39	GAAAATTGAGATGGGTTTGATAGAG	AACCAAAAATTCCAACTTTAAAACA
	Act Bis F40R40	GGTTTTGAATAGTTAGTTTGTTTGTTG	TCTTTACCATTAAAAAATCCCAAAC
	Act Bis F41R41	TGTAATGTAGTTTGTGTGATTAAAAA	AAAAATAACCCTCCCAATACATACTAC
	Act Bis F42R42	AGAATTGTTTTATGGGGGATTTAGT	AAAAACATTATCTCAAAACAAAAACAAA
Xist	Xist BS 5/6	TGTAARTTTTGTGGTTATTTTTTTT	ATATTCCCCCAAAACTCCTTAAATA
Snrpn	Snrpn Bis F2R2	TTTAGAATGTTTTGGTTAAATAGGATGTAT	CAAAAAAACAAAAACCCCTACATTA
Dlx 5/6	Dlx Bis 1F1R	TTTTGGTAGGAAAATAAGAATTTATTG	ACACATCCCTAAATAAAAAAACCAC

	Dlx Bis 2F2R	GGATTAAGTTGTAGATAAATATTGTTAGGA	ATAAACAAATAAAACCCAAAACCAC
	Dlx Bis 3F3R	TTTTAGGGGTGGAAAAATATTTTAA	CCCAAACTTTAATTTAACTAACC
	Dlx Bis 4F4R	GTTGGGAAAAGTTAGGTTAAGAAAT	TACCCCCAAAAACTATAAAAACTTC
	Dlx Bis 5F5R	AGGGTAATTTAAGTTAGAGAGTGTT	CTTTAACCCACACAAAAATATACAC
	Dlx Bis 6F6R	TTGTTATAAGAAGTAGAGGTAGGAGAGTAG	ATTTAACAAAAAAATCCCAAACATC
	Dlx Bis 8F8F	ATGTTTTTGGTGTAGTTATGGAGAA	CCCAACCTCAAAAATTTAAAATCTA

L1 retro-transposons	Line1BS F2R2	TAGGAAATTAGTTTGAATAGGTGAGAGGT	TCAAACACTATATTACTTTAACAATTCCCA
Major satellite	MajSat BisF1R1	GGAAAATGAGAAATATATATTTTAGGA	TCAATTTTCTTACCATATTCCAC
Intracisternal A particles	IAPBS F2R2	TTGTGTTTTAAGTGGTAAATAAATAATTTG	CAAAAAAAACACACAAACCAAAAT

Primers for expression analysis

Histone H1	H1.0	ATGACCGAGAACTCCACCTC	CGATCATGTCTGAATACTTG
	H1.2-H1.5	CGGTGTCCGAGCTCATCACC	GCTGTTGTTCTTCTCCACATCG

L1 retro-transposons	L1R1	CAATCGCGTGGAACTTGAGAC	GACTCAGCTGGCAAGGTAGC
Major satellite	Maj Sat	GGCGAGAAAACTGAAAATCACG	AGGTCCTTCAGTGTGCATTTC
Intracisternal A particles	IAP 1	GCTTTCGTTTTTGGGGCTTGG	CTTACTCCGCGTTCTCACGAC

Tyrosine hydroxylase	Tyrhy Ex2	GCAGAGTCTCATCGAGGATG	CTCAAACACTTTCAAAGCCC
c-Myc	c-Myc 10.1	CCAGAGCTTCATCTGCGATCC	TCAGGCTGGTGCTGTCTTTGC
Gapdh	Gapdh P3P4	TACCCCCAATGTGTCCGTCG	CCTGCTTCACCACCTTCTTG
Actb	Actb 2.1	TGAAGCTGTAGCCACGCTCG	CTGTCCCTGTATGCCTCTGG

Primers for colony PCR

	M13F/R	GTAAAACGACGGCCAGTG	CAGGAAACAGCTATGAC

Appendix B: Publications

1) CpG islands influence chromatin structure via the CpG-binding protein Cfp1

2) Neuronal MeCP2 Is Expressed at Near Histone-Octamer Levels and Globally Alters the Chromatin State

3) A Temporal Threshold for Formaldehyde Crosslinking and Fixation

References

Aapola, U., Kawasaki, K., Scott, H.S., Ollila, J., Vihinen, M., Heino, M., Shintani, A., Minoshima, S., Krohn, K., Antonarakis, S.E.*, et al.* (2000). Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics *65*, 293-298.

Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H., and Cardoso, M.C. (2007). MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. Nucleic Acids Res.

Aid, T., Kazantseva, A., Piirsoo, M., Palm, K., and Timmusk, T. (2007). Mouse and rat BDNF gene structure and expression revisited. J Neurosci Res *85*, 525-535.

Allan, J., Rau, D.C., Harborne, N., and Gould, H. (1984). Higher order structure in a short repeat length chromatin. J Cell Biol *98*, 1320-1327.

Almer, A., Rudolph, H., Hinnen, A., and Horz, W. (1986). Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. Embo J *5*, 2689-2696.

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., Boyes, J., and Bird, A. (1990). High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. Cell *62*, 503-514.

Antequera, F., Macleod, D., and Bird, A.P. (1989). Specific protection of methylated CpGs in mammalian nuclei. Cell *58*, 509-517.

Antequera, F., Tamame, M., Villanueva, J.R., and Santos, T. (1984). DNA methylation in the fungi. J Biol Chem *259*, 8033-8036.

Armstrong, D.D. (2002). Neuropathology of Rett syndrome. Ment Retard Dev Disabil Res Rev *8*, 72-76.

Arney, K.L. (2003). H19 and Igf2--enhancing the confusion? Trends Genet *19*, 17-23.

Bading, H. (2000). Transcription-dependent neuronal plasticity the nuclear calcium hypothesis. Eur J Biochem *267*, 5280-5283.

Ballas, N., Lioy, D.T., Grunseich, C., and Mandel, G. (2009). Non-cell autonomous influence of MeCP2-deficient glia on neuronal dendritic morphology. Nat Neurosci *12*, 311-317.

Ballestar, E., Ropero, S., Alaminos, M., Armstrong, J., Setien, F., Agrelo, R., Fraga, M.F., Herranz, M., Avila, S., Pineda, M.*, et al.* (2005). The impact of MECP2 mutations in the expression patterns of Rett syndrome patients. Hum Genet *116*, 91-104.

Balmer, D., Goldstine, J., Rao, Y.M., and LaSalle, J.M. (2003). Elevated methyl-CpG-binding protein 2 expression is acquired during postnatal human brain development and is correlated with alternative polyadenylation. J Mol Med *81*, 61-68.

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.

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

Belotserkovskaya, R., Oh, S., Bondarenko, V.A., Orphanides, G., Studitsky, V.M., and Reinberg, D. (2003). FACT facilitates transcription-dependent nucleosome alteration. Science *301*, 1090-1093.

Ben-Shachar, S., Chahrour, M., Thaller, C., Shaw, C.A., and Zoghbi, H.Y. (2009). Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. Hum Mol Genet *18*, 2431-2442.

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.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One *3*, e3156.

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.

Bestor, T.H. (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. EMBO J *11*, 2611-2617.

Bird, A. (1997). Does DNA methylation control transposition of selfish elements in the germline? Trends Genet *13*, 469-472.

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

Bird, A. (2007). Perceptions of epigenetics. Nature *447*, 396-398.

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. Philos Trans R Soc Lond B Biol Sci *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. (1986). CpG-rich islands and the function of DNA methylation. Nature *321*, 209-213.

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

Bird, A.P., Taggart, M.H., Nicholls, R.D., and Higgs, D.R. (1987). Non-methylated CpG-rich islands at the human alpha-globin locus: implications for evolution of the alpha-globin pseudogene. EMBO J *6*, 999-1004.

Bird, A.P., Taggart, M.H., and Smith, B.A. (1979). Methylated and unmethylated DNA compartments in the sea urchin genome. Cell *17*, 889-901.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E*., et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799-816.

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 *21*, 4067-4068.

Bostick, M., Kim, J.K., Esteve, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760-1764.

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. EMBO J *11*, 327-333.

Burdon, R.H., and Adams, R.L. (1969). The in vivo methylation of DNA in mouse fibroblasts. Biochim Biophys Acta *174*, 322-329.

Buschhausen, G., Graessmann, M., and Graessmann, A. (1985). Inhibition of herpes simplex thymidine kinase gene expression by DNA methylation is an indirect effect. Nucleic Acids Res *13*, 5503-5513.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S., and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science *298*, 1039-1043.

Carlson, L.L., Page, A.W., and Bestor, T.H. (1992). Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting. Genes Dev *6*, 2536-2541.

Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400-404.

Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., and Workman, J.L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell *123*, 581-592.

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.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat Rev Genet *6*, 351-360.

Chandler, S.P., Guschin, D., Landsberger, N., and Wolffe, A.P. (1999). The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. Biochemistry *38*, 7008-7018.

Chang, Q., Khare, G., Dani, V., Nelson, S., and Jaenisch, R. (2006). The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. Neuron *49*, 341-348.

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, R.Z., Pettersson, U., Beard, C., Jackson-Grusby, L., and Jaenisch, R. (1998). DNA hypomethylation leads to elevated mutation rates. Nature *395*, 89-93.

Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003a). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol *23*, 5594-5605.

Chen, W.G., Chang, Q., Lin, Y., Meissner, A., West, A.E., Griffith, E.C., Jaenisch, R., and Greenberg, M.E. (2003b). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science *302*, 885-889.

Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G., and Li, B.F. (1997). Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science *277*, 1996-2000.

Clark, S.J., Harrison, J., and Molloy, P.L. (1997). Sp1 binding is inhibited by (m)Cp(m)CpG methylation. Gene *195*, 67-71.

Clayton, A.L., Hazzalin, C.A., and Mahadevan, L.C. (2006). Enhanced histone acetylation and transcription: a dynamic perspective. Mol Cell *23*, 289-296.

Cohanim, A.B., and Haran, T.E. (2009). The coexistence of the nucleosome positioning code with the genetic code on eukaryotic genomes. Nucleic Acids Res *37*, 6466-6476.

Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C.D., Pradhan, S., Nelson, S.F., Pellegrini, M., and Jacobsen, S.E. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. Nature *452*, 215-219.

Collins, A.L., Levenson, J.M., Vilaythong, A.P., Richman, R., Armstrong, D.L., Noebels, J.L., David Sweatt, J., and Zoghbi, H.Y. (2004). Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. Hum Mol Genet *13*, 2679-2689.

Cooper, D.N., Taggart, M.H., and Bird, A.P. (1983). Unmethylated domains in vertebrate DNA. Nucleic Acids Res *11*, 647-658.

Cooper, D.N., and Youssoufian, H. (1988). The CpG dinucleotide and human genetic disease. Hum Genet *78*, 151-155.

Coy, J.F., Sedlacek, Z., Bachner, D., Delius, H., and Poustka, A. (1999). A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3"-untranslated region of the methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene expression. Hum Mol Genet *8*, 1253-1262.

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.

Daniel, J.M., Spring, C.M., Crawford, H.C., Reynolds, A.B., and Baig, A. (2002). The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. Nucleic Acids Res *30*, 2911-2919.

DeChiara, T.M., Efstratiadis, A., and Robertson, E.J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature *345*, 78-80.

DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell *64*, 849-859.

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 Dev *15*, 2940-2944.

Dickson, J., Gowher, H., Strogantsev, R., Gaszner, M., Hair, A., Felsenfeld, G., and West, A.G. (2010). VEZF1 elements mediate protection from DNA methylation. PLoS Genet *6*, e1000804.

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

Edwards, C.A., and Ferguson-Smith, A.C. (2007). Mechanisms regulating imprinted genes in clusters. Curr Opin Cell Biol *19*, 281-289.

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.

Eide, L., and McMurray, C.T. (2005). Culture of adult mouse neurons. Biotechniques *38*, 99-104.

El-Osta, A., Kantharidis, P., Zalcberg, J.R., and Wolffe, A.P. (2002). Precipitous release of methyl-CpG binding protein 2 and histone deacetylase 1 from the methylated human multidrug resistance gene (MDR1) on activation. Mol Cell Biol *22*, 1844-1857.

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 Dev *20*, 3089-3103.

Fasken, M.B., and Corbett, A.H. (2009). Mechanisms of nuclear mRNA quality control. RNA Biol *6*, 237-241.

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

Ferguson-Smith, A.C., Sasaki, H., Cattanach, B.M., and Surani, M.A. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. Nature *362*, 751-755.

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.

Finnegan, E.J., and Dennis, E.S. (1993). Isolation and identification by sequence homology of a putative cytosine methyltransferase from Arabidopsis thaliana. Nucleic Acids Res *21*, 2383-2388.

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.

Fuks, F., Burgers, W.A., Godin, N., Kasai, M., and Kouzarides, T. (2001). Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. Embo J *20*, 2536-2544.

Fuks, F., Hurd, P.J., Wolf, D., Nan, X., Bird, A.P., and Kouzarides, T. (2003). The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem *278*, 4035-4040.

Galvao, T.C., and Thomas, J.O. (2005). Structure-specific binding of MeCP2 to four-way junction DNA through its methyl CpG-binding domain. Nucleic Acids Res *33*, 6603-6609.

Gao, Z., Liu, H.L., Daxinger, L., Pontes, O., He, X., Qian, W., Lin, H., Xie, M., Lorkovic, Z.J., Zhang, S.*, et al.* An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. Nature *465*, 106-109.

Garrick, D., Fiering, S., Martin, D.I., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. Nat Genet *18*, 56-59.

Gaudet, F., Hodgson, J.G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J.W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. Science *300*, 489-492.

Gautsch, J.W., and Wilson, M.C. (1983). Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature *301*, 32-37.

Gazin, C., Wajapeyee, N., Gobeil, S., Virbasius, C.M., and Green, M.R. (2007). An elaborate pathway required for Ras-mediated epigenetic silencing. Nature *449*, 1073-1077.

Georgel, P.T., Horowitz-Scherer, R.A., Adkins, N., Woodcock, C.L., Wade, P.A., and Hansen, J.C. (2003). Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. J Biol Chem *278*, 32181-32188.

Gibbons, R.J., McDowell, T.L., Raman, S., O'Rourke, D.M., Garrick, D., Ayyub, H., and Higgs, D.R. (2000). Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet *24*, 368-371.

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.

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.

Gupta, S., Kim, S.Y., Artis, S., Molfese, D.L., Schumacher, A., Sweatt, J.D., Paylor, R.E., and Lubin, F.D. (2010). Histone methylation regulates memory formation. J Neurosci *30*, 3589-3599.

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.

Hansen, R.S., and Gartler, S.M. (1990). 5-Azacytidine-induced reactivation of the human X chromosome-linked PGK1 gene is associated with a large region of cytosine demethylation in the 5' CpG island. Proc Natl Acad Sci U S A *87*, 4174-4178.

Harikrishnan, K.N., Bayles, R., Ciccotosto, G.D., Maxwell, S., Cappai, R., Pelka, G.J., Tam, P.P., Christodoulou, J., and El-Osta, A. (2010). Alleviating transcriptional inhibition of the norepinephrine slc6a2 transporter gene in depolarized neurons. J Neurosci *30*, 1494-1501.

Hata, K., Okano, M., Lei, H., and Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. Development *129*, 1983-1993.

Heard, E., Chaumeil, J., Masui, O., and Okamoto, I. (2004). Mammalian X-chromosome inactivation: an epigenetics paradigm. Cold Spring Harb Symp Quant Biol *69*, 89-102.

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

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 Dev *15*, 710-723.

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

Hendrich, B., and Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. Trends Genet *19*, 269-277.

Hendrich, B.D., Brown, C.J., and Willard, H.F. (1993). Evolutionary conservation of possible functional domains of the human and murine XIST genes. Hum Mol Genet *2*, 663-672.

Herculano-Houzel, S., and Lent, R. (2005). Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. J Neurosci *25*, 2518-2521.

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

Hillier, L.W., Marth, G.T., Quinlan, A.R., Dooling, D., Fewell, G., Barnett, D., Fox, P., Glasscock, J.I., Hickenbotham, M., Huang, W*., et al.* (2008). Whole-genome sequencing and variant discovery in C. elegans. Nat Methods *5*, 183-188.

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. Mol Cell *29*, 525-531.

Horike, S., Cai, S., Miyano, M., Cheng, J.F., and Kohwi-Shigematsu, T. (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet *37*, 31-40.

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 Biol *6*, e22.

Illingworth, R.S., and Bird, A.P. (2009). CpG islands--'a rough guide'. FEBS Lett *583*, 1713-1720.

Ishibashi, T., Thambirajah, A.A., and Ausio, J. (2008). MeCP2 preferentially binds to methylated linker DNA in the absence of the terminal tail of histone H3 and independently of histone acetylation. FEBS Lett *582*, 1157-1162.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

Jahner, D., Stuhlmann, H., Stewart, C.L., Harbers, K., Lohler, J., Simon, I., and Jaenisch, R. (1982). De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature *298*, 623-628.

Jeanpierre, M., Turleau, C., Aurias, A., Prieur, M., Ledeist, F., Fischer, A., and Viegas-Pequignot, E. (1993). An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. Hum Mol Genet *2*, 731-735.

Jeddeloh, J.A., Stokes, T.L., and Richards, E.J. (1999). Maintenance of genomic methylation requires a SWI2/SNF2-like protein. Nat Genet *22*, 94-97.

Jeffery, L., and Nakielny, S. (2004). Components of the DNA methylation system of chromatin control are RNA-binding proteins. J Biol Chem *279*, 49479-49487.

Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. Chembiochem *3*, 274-293.

Jeltsch, A. (2008). Reading and writing DNA methylation. Nat Struct Mol Biol *15*, 1003-1004.

Johnson, L.M., Bostick, M., Zhang, X., Kraft, E., Henderson, I., Callis, J., and Jacobsen, S.E. (2007). The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol *17*, 379-384.

Johnson, T.B., and Coghill, R. D. (1925). Research on pyrimidines. C111. The discovery of 5-methylcytosine in tuberculinic acid, the nucleic acid of tubercule bacillus. J Amer Chem Soc, 2838-2845.

Jordan, C., Li, H.H., Kwan, H.C., and Francke, U. (2007). Cerebellar gene expression profiles of mouse models for Rett syndrome reveal novel MeCP2 targets. BMC Med Genet *8*, 36.

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.

Juttermann, R., Li, E., and Jaenisch, R. (1994). Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. Proc Natl Acad Sci U S A *91*, 11797-11801.

Kaludov, N.K., and Wolffe, A.P. (2000). MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. Nucleic Acids Res *28*, 1921-1928.

Kapranov, P., Willingham, A.T., and Gingeras, T.R. (2007). Genome-wide transcription and the implications for genomic organization. Nat Rev Genet *8*, 413-423.

Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R., and Berube, N.G. (2010). ATRX Partners with Cohesin and MeCP2 and Contributes to Developmental Silencing of Imprinted Genes in the Brain. Dev Cell *18*, 191-202.

Kim, G.D., Ni, J., Kelesoglu, N., Roberts, R.J., and Pradhan, S. (2002). Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. EMBO J *21*, 4183-4195.

Kimura, H., and Shiota, K. (2003). Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. J Biol Chem *278*, 4806-4812.

Kishi, N., and Macklis, J.D. (2004). MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. Mol Cell Neurosci *27*, 306-321.

Klein, M.E., Lioy, D.T., Ma, L., Impey, S., Mandel, G., and Goodman, R.H. (2007). Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. Nat Neurosci *10*, 1513-1514.

Klimasauskas, S., Kumar, S., Roberts, R.J., and Cheng, X. (1994). HhaI methyltransferase flips its target base out of the DNA helix. Cell *76*, 357-369.

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. J Biol Chem *279*, 46490-46496.

Klose, R.J., and Bird, A.P. (2006). Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci *31*, 89-97.

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. Mol Cell *19*, 667-678.

Kokura, K., Kaul, S.C., Wadhwa, R., Nomura, T., Khan, M.M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001). The Ski protein family is required for MeCP2-mediated transcriptional repression. J Biol Chem *276*, 34115-34121.

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005). The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. Mol Cell Biol *25*, 4388-4396.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kriaucionis, S., and Bird, A. (2003). DNA methylation and Rett syndrome. Hum Mol Genet *12 Spec No 2*, R221-227.

Kriaucionis, S., and Bird, A. (2004). The major form of MeCP2 has a novel N-terminus generated by alternative splicing. Nucleic Acids Res *32*, 1818-1823.

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

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.

LaSalle, J.M., Goldstine, J., Balmer, D., and Greco, C.M. (2001). Quantitative localization of heterogeneous methyl-CpG-binding protein 2 (MeCP2) expression phenotypes in normal and Rett syndrome brain by laser scanning cytometry. Hum Mol Genet *10*, 1729-1740.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., and Wei, C.L. (2010). Dynamic changes in the human methylome during differentiation. Genome Res.

Law, J.A., and Jacobsen, S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Laybourn, P.J., and Kadonaga, J.T. (1991). Role of nucleosomal cores and histone H1 in regulation of transcription by RNA polymerase II. Science *254*, 238-245.

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.

Lei, H., Oh, S.P., Okano, M., Juttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195-3205.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865-873.

Letovsky, J., and Dynan, W.S. (1989). Measurement of the binding of transcription factor Sp1 to a single GC box recognition sequence. Nucleic Acids Res *17*, 2639-2653.

Lewin, B. (1980). Gene Expression: Eucaryotic Chromosomes (Vol. 2) (2nd edn.). John Wiley & Sons.

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, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet *3*, 662-673.

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.

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 *462*, 315-322.

Liu, J., and Francke, U. (2006). Identification of cis-regulatory elements for MECP2 expression. Hum Mol Genet *15*, 1769-1782.

Lock, L.F., Melton, D.W., Caskey, C.T., and Martin, G.R. (1986). Methylation of the mouse hprt gene differs on the active and inactive X chromosomes. Mol Cell Biol *6*, 914-924.

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

Lorincz, M.C., Schubeler, D., and Groudine, M. (2001). Methylation-mediated proviral silencing is associated with MeCP2 recruitment and localized histone H3 deacetylation. Mol Cell Biol *21*, 7913-7922.

Lubs, H., Abidi, F., Bier, J.A., Abuelo, D., Ouzts, L., Voeller, K., Fennell, E., Stevenson, R.E., Schwartz, C.E., and Arena, F. (1999). XLMR syndrome characterized by multiple respiratory infections, hypertelorism, severe CNS deterioration and early death localizes to distal Xq28. Am J Med Genet *85*, 243-248.

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.

Luikenhuis, S., Giacometti, E., Beard, C.F., and Jaenisch, R. (2004). Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. Proc Natl Acad Sci U S A *101*, 6033-6038.

Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000a). DNA methylation in Drosophila melanogaster. Nature *408*, 538-540.

Lyko, F., Whittaker, A.J., Orr-Weaver, T.L., and Jaenisch, R. (2000b). The putative Drosophila methyltransferase gene dDnmt2 is contained in a transposon-like element and is expressed specifically in ovaries. Mech Dev *95*, 215-217.

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 Dev *8*, 2282-2292.

Mancini, D.N., Singh, S.M., Archer, T.K., and Rodenhiser, D.I. (1999). Site-specific DNA methylation in the neurofibromatosis (NF1) promoter interferes with binding of CREB and SP1 transcription factors. Oncogene *18*, 4108-4119.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V.*, et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762-767.

Marin, M., Karis, A., Visser, P., Grosveld, F., and Philipsen, S. (1997). Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. Cell *89*, 619-628.

Martin Caballero, I., Hansen, J., Leaford, D., Pollard, S., and Hendrich, B.D. (2009). The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. PLoS One *4*, e4315.

Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G., and Sun, Y.E. (2003). DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science *302*, 890-893.

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.

Meselson, M., and Stahl, F.W. (1958). The Replication of DNA in Escherichia Coli. Proc Natl Acad Sci U S A *44*, 671-682.

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.

Mnatzakanian, G.N., Lohi, H., Munteanu, I., Alfred, S.E., Yamada, T., MacLeod, P.J., Jones, J.R., Scherer, S.W., Schanen, N.C., Friez, M.J.*, et al.* (2004). A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. Nat Genet *36*, 339-341.

Monk, M., Adams, R.L., and Rinaldi, A. (1991). Decrease in DNA methylase activity during preimplantation development in the mouse. Development *112*, 189-192.

Mullen, R.J., Buck, C.R., and Smith, A.M. (1992). NeuN, a neuronal specific nuclear protein in vertebrates. Development *116*, 201-211.

Myant, K., and Stancheva, I. (2008). LSH cooperates with DNA methyltransferases to repress transcription. Mol Cell Biol *28*, 215-226.

Nachman, M.W., and Crowell, S.L. (2000). Estimate of the mutation rate per nucleotide in humans. Genetics *156*, 297-304.

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., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S., and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proc Natl Acad Sci U S A *104*, 2709-2714.

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.

Nan, X., Tate, P., Li, E., and Bird, A. (1996). DNA methylation specifies chromosomal localization of MeCP2. Mol Cell Biol *16*, 414-421.

Nguyen, C.T., Gonzales, F.A., and Jones, P.A. (2001). Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation. Nucleic Acids Res *29*, 4598-4606.

Nikitina, T., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., Grigoryev, S.A., and Woodcock, C.L. (2007). MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. J Biol Chem.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2006). Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. Mol Cell Biol.

Nuber, U.A., Kriaucionis, S., Roloff, T.C., Guy, J., Selfridge, J., Steinhoff, C., Schulz, R., Lipkowitz, B., Ropers, H.H., Holmes, M.C., and Bird, A. (2005). Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. Hum Mol Genet *14*, 2247-2256.

Ohki, I., Shimotake, N., Fujita, N., Jee, J., Ikegami, T., Nakao, M., and Shirakawa, M. (2001). Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. Cell *105*, 487-497.

Ohki, I., Shimotake, N., Fujita, N., Nakao, M., and Shirakawa, M. (1999). Solution structure of the methyl-CpG-binding domain of the methylation-dependent transcriptional repressor MBD1. EMBO J *18*, 6653-6661.

Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. Science *303*, 644-649.

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.

Okano, M., Xie, S., and Li, E. (1998a). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. Nat Genet *19*, 219-220.

Okano, M., Xie, S., and Li, E. (1998b). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res *26*, 2536-2540.

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.

Palmer, B.R., and Marinus, M.G. (1994). The dam and dcm strains of Escherichia coli--a review. Gene *143*, 1-12.

Pannell, D., Osborne, C.S., Yao, S., Sukonnik, T., Pasceri, P., Karaiskakis, A., Okano, M., Li, E., Lipshitz, H.D., and Ellis, J. (2000). Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. EMBO J *19*, 5884-5894.

Panning, B., and Jaenisch, R. (1996). DNA hypomethylation can activate Xist expression and silence X-linked genes. Genes Dev *10*, 1991-2002.

Pearson, E.C., Bates, D.L., Prospero, T.D., and Thomas, J.O. (1984). Neuronal nuclei and glial nuclei from mammalian cerebral cortex. Nucleosome repeat lengths, DNA contents and H1 contents. Eur J Biochem *144*, 353-360.

Peart, M.J., Smyth, G.K., van Laar, R.K., Bowtell, D.D., Richon, V.M., Marks, P.A., Holloway, A.J., and Johnstone, R.W. (2005). Identification and functional significance of genes regulated by structurally different histone deacetylase inhibitors. Proc Natl Acad Sci U S A *102*, 3697-3702.

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M.C., and Reuter, G. (2009). Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. Nat Genet *41*, 696-702.

Pinarbasi, E., Elliott, J., and Hornby, D.P. (1996). Activation of a yeast pseudo DNA methyltransferase by deletion of a single amino acid. J Mol Biol *257*, 804-813.

Pollack, Y., Stein, R., Razin, A., and Cedar, H. (1980). Methylation of foreign DNA sequences in eukaryotic cells. Proc Natl Acad Sci U S A *77*, 6463-6467.

Proffitt, J.H., Davie, J.R., Swinton, D., and Hattman, S. (1984). 5-Methylcytosine is not detectable in Saccharomyces cerevisiae DNA. Mol Cell Biol *4*, 985-988.

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 Dev *15*, 1613-1618.

Prokhortchouk, A., Sansom, O., Selfridge, J., Caballero, I.M., Salozhin, S., Aithozhina, D., Cerchietti, L., Meng, F.G., Augenlicht, L.H., Mariadason, J.M.*, et al.* (2006). Kaiso-deficient mice show resistance to intestinal cancer. Mol Cell Biol *26*, 199-208.

Quaderi, N.A., Meehan, R.R., Tate, P.H., Cross, S.H., Bird, A.P., Chatterjee, A., Herman, G.E., and Brown, S.D. (1994). Genetic and physical mapping of a gene encoding a methyl CpG binding protein, Mecp2, to the mouse X chromosome. Genomics *22*, 648-651.

Rabinowicz, P.D., Palmer, L.E., May, B.P., Hemann, M.T., Lowe, S.W., McCombie, W.R., and Martienssen, R.A. (2003). Genes and transposons are differentially methylated in plants, but not in mammals. Genome Res *13*, 2658-2664.

Rai, K., Chidester, S., Zavala, C.V., Manos, E.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2007). Dnmt2 functions in the cytoplasm to promote liver, brain, and retina development in zebrafish. Genes Dev *21*, 261-266.

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.

Rangwala, S.H., and Richards, E.J. (2004). The value-added genome: building and maintaining genomic cytosine methylation landscapes. Curr Opin Genet Dev *14*, 686-691.

Ratnam, S., Mertineit, C., Ding, F., Howell, C.Y., Clarke, H.J., Bestor, T.H., Chaillet, J.R., and Trasler, J.M. (2002). Dynamics of Dnmt1 methyltransferase expression and intracellular localization during oogenesis and preimplantation development. Dev Biol *245*, 304-314.

Reichwald, K., Thiesen, J., Wiehe, T., Weitzel, J., Poustka, W.A., Rosenthal, A., Platzer, M., Stratling, W.H., and Kioschis, P. (2000). Comparative sequence analysis of the MECP2-locus in human and mouse reveals new transcribed regions. Mamm Genome *11*, 182-190.

Reik, W., and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nat Rev Genet *2*, 21-32.

Reinisch, K.M., Chen, L., Verdine, G.L., and Lipscomb, W.N. (1995). The crystal structure of HaeIII methyltransferase convalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. Cell *82*, 143-153.

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. J Biol Chem *281*, 24790-24802.

Riggs, A.D., Xiong, Z., Wang, L., and LeBon, J.M. (1998). Methylation dynamics, epigenetic fidelity and X chromosome structure. Novartis Found Symp *214*, 214-225; discussion 225-232.

Rodriguez, J., Frigola, J., Vendrell, E., Risques, R.A., Fraga, M.F., Morales, C., Moreno, V., Esteller, M., Capella, G., Ribas, M., and Peinado, M.A. (2006). Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. Cancer Res *66*, 8462-9468.

Rountree, M.R., and Selker, E.U. (1997). DNA methylation inhibits elongation but not initiation of transcription in Neurospora crassa. Genes Dev *11*, 2383-2395.

Russo, V.E.A., Martienssen, R.A. & Riggs, A.D. (eds) (1996). Epigenetic Mechanisms of Gene Regulation. Cold Spring Harbour Laboratory Press, Woodbury.

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. Mol Cell *25*, 15-30.

Ruzov, A., Dunican, D.S., Prokhortchouk, A., Pennings, S., Stancheva, I., Prokhortchouk, E., and Meehan, R.R. (2004). Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. Development *131*, 6185-6194.

Sancho, M., Diani, E., Beato, M., and Jordan, A. (2008). Depletion of human histone H1 variants uncovers specific roles in gene expression and cell growth. PLoS Genet *4*, e1000227.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol *241*, 172-182.

Scarano, E., Iaccarino, M., Grippo, P., and Winckelmans, D. (1965). On methylation of DNA during development of the sea urchin embryo. J Mol Biol *14*, 603-607.

Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C., and Leonhardt, H. (2007). Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. Nucleic Acids Res *35*, 4301-4312.

Schmiedeberg, L., Skene, P., Deaton, A., and Bird, A. (2009). A temporal threshold for formaldehyde crosslinking and fixation. PLoS One *4*, e4636.

Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J *21*, 1121-1131.

Schule, B., Li, H.H., Fisch-Kohl, C., Purmann, C., and Francke, U. (2007). DLX5 and DLX6 Expression Is Biallelic and Not Modulated by MeCP2 Deficiency. Am J Hum Genet *81*, 492-506.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772-778.

Selker, E.U., Freitag, M., Kothe, G.O., Margolin, B.S., Rountree, M.R., Allis, C.D., and Tamaru, H. (2002). Induction and maintenance of nonsymmetrical DNA methylation in Neurospora. Proc Natl Acad Sci U S A *99 Suppl 4*, 16485-16490.

Selker, E.U., Tountas, N.A., Cross, S.H., Margolin, B.S., Murphy, J.G., Bird, A.P., and Freitag, M. (2003). The methylated component of the Neurospora crassa genome. Nature *422*, 893-897.

Sera, T., and Wolffe, A.P. (1998). Role of histone H1 as an architectural determinant of chromatin structure and as a specific repressor of transcription on Xenopus oocyte 5S rRNA genes. Mol Cell Biol *18*, 3668-3680.

Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R., and Zoghbi, H. (2002a). Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. Neuron *35*, 243-254.

Shahbazian, M.D., Antalffy, B., Armstrong, D.L., and Zoghbi, H.Y. (2002b). Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. Hum Mol Genet *11*, 115-124.

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K.*, et al.* (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature *450*, 908-912.

Shogren-Knaak, M., Ishii, H., Sun, J.M., Pazin, M.J., Davie, J.R., and Peterson, C.L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. Science *311*, 844-847.

Silva, A.J., Ward, K., and White, R. (1993). Mosaic methylation in clonal tissue. Dev Biol *156*, 391-398.

Simmen, M.W., Leitgeb, S., Charlton, J., Jones, S.J., Harris, B.R., Clark, V.H., and Bird, A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. Science *283*, 1164-1167.

Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). Caenorhabditis elegans DNA does not contain 5-methylcytosine at any time during development or aging. Nucleic Acids Res *14*, 6711-6719.

Sims, R.J., 3rd, Millhouse, S., Chen, C.F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol Cell *28*, 665-676.

Singh, J., Saxena, A., Christodoulou, J., and Ravine, D. (2008). MECP2 genomic structure and function: insights from ENCODE. Nucleic Acids Res *36*, 6035-6047.

Stancheva, I., Collins, A.L., Van den Veyver, I.B., Zoghbi, H., and Meehan, R.R. (2003). A mutant form of MeCP2 protein associated with human Rett syndrome cannot be displaced from methylated DNA by notch in Xenopus embryos. Mol Cell *12*, 425-435.

Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proc Natl Acad Sci U S A *79*, 3418-3422.

Stewart, C.L., Stuhlmann, H., Jahner, D., and Jaenisch, R. (1982). De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc Natl Acad Sci U S A *79*, 4098-4102.

Stoger, R., Kajimura, T.M., Brown, W.T., and Laird, C.D. (1997). Epigenetic variation illustrated by DNA methylation patterns of the fragile-X gene FMR1. Hum Mol Genet *6*, 1791-1801.

Su, A.I., Cooke, M.P., Ching, K.A., Hakak, Y., Walker, J.R., Wiltshire, T., Orth, A.P., Vega, R.G., Sapinoso, L.M., Moqrich, A.*, et al.* (2002). Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A *99*, 4465-4470.

Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. J Biol Chem *279*, 27816-27823.

Sutcliffe, J.S., Nakao, M., Christian, S., Orstavik, K.H., Tommerup, N., Ledbetter, D.H., and Beaudet, A.L. (1994). Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. Nat Genet *8*, 52-58.

Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., and Oikawa, T. (2003). Direct association between PU.1 and MeCP2 that recruits mSin3A-HDAC complex for PU.1-mediated transcriptional repression. Oncogene *22*, 8688-8698.

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 Res *17*, 625-631.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.

Takeshima, H., Yamashita, S., Shimazu, T., Niwa, T., and Ushijima, T. (2009). The presence of RNA polymerase II, active or stalled, predicts epigenetic fate of promoter CpG islands. Genome Res *19*, 1974-1982.

Tamaru, H., and Selker, E.U. (2001). A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. Nature *414*, 277-283.

Tao, J., Hu, K., Chang, Q., Wu, H., Sherman, N.E., Martinowich, K., Klose, R.J., Schanen, C., Jaenisch, R., Wang, W., and Sun, Y.E. (2009). Phosphorylation of MeCP2 at Serine 80 regulates its chromatin association and neurological function. Proc Natl Acad Sci U S A *106*, 4882-4887.

Temudo, T., Santos, M., Ramos, E., Dias, K., Vieira, J.P., Moreira, A., Calado, E., Carrilho, I., Oliveira, G., Levy, A.*, et al.* Rett syndrome with and without detected MECP2 mutations: An attempt to redefine phenotypes. Brain Dev.

Thomas, J.O. (1999). Histone H1: location and role. Curr Opin Cell Biol *11*, 312-317.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., 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.

Travers, A. (1999). The location of the linker histone on the nucleosome. Trends Biochem Sci *24*, 4-7.

Traynor, J., Agarwal, P., Lazzeroni, L., and Francke, U. (2002). Gene expression patterns vary in clonal cell cultures from Rett syndrome females with eight different MECP2 mutations. BMC Med Genet *3*, 12.

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. Proc Natl Acad Sci U S A *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.

Urdinguio, R.G., Lopez-Serra, L., Lopez-Nieva, P., Alaminos, M., Diaz-Uriarte, R., Fernandez, A.F., and Esteller, M. (2008). Mecp2-null mice provide new neuronal targets for Rett syndrome. PLoS One *3*, e3669.

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

Valinluck, V., Tsai, H.H., Rogstad, D.K., Burdzy, A., Bird, A., and Sowers, L.C. (2004). Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). Nucleic Acids Res *32*, 4100-4108.

Valley, C.M., and Willard, H.F. (2006). Genomic and epigenomic approaches to the study of X chromosome inactivation. Curr Opin Genet Dev *16*, 240-245.

Viegas-Pequignot, E., and Dutrillaux, B. (1976). Segmentation of human chromosomes induced by 5-ACR (5-azacytidine). Hum Genet *34*, 247-254.

Viegas-Pequignot, E., Dutrillaux, B., and Thomas, G. (1988). Inactive X chromosome has the highest concentration of unmethylated Hha I sites. Proc Natl Acad Sci U S A *85*, 7657-7660.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M.*, et al.* (2005). The Polycomb group protein EZH2 directly controls DNA methylation. Nature.

Waddington, C.H. (1957). The Strategy of the Genes Allen & Unwin, London.

Wakefield, R.I., Smith, B.O., Nan, X., Free, A., Soteriou, A., Uhrin, D., Bird, A.P., and Barlow, P.N. (1999). The solution structure of the domain from MeCP2 that binds to methylated DNA. J Mol Biol *291*, 1055-1065.

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.

Watson, J.D., and Crick, F.H. (1953). The structure of DNA. Cold Spring Harb Symp Quant Biol *18*, 123-131.

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.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell *135*, 635-648.

Wigler, M., Levy, D., and Perucho, M. (1981). The somatic replication of DNA methylation. Cell *24*, 33-40.

Wilkinson, C.R., Bartlett, R., Nurse, P., and Bird, A.P. (1995). The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. Nucleic Acids Res *23*, 203-210.

Wilson, G.G., and Murray, N.E. (1991). Restriction and modification systems. Annu Rev Genet *25*, 585-627.

Woodcock, C.L., Skoultchi, A.I., and Fan, Y. (2006). Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. Chromosome Res *14*, 17-25.

Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. Mol Cell *5*, 695-705.

Xu, G.L., Bestor, T.H., Bourc'his, D., Hsieh, C.L., Tommerup, N., Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E. (1999). Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature *402*, 187-191.

Yamada, T., Fischle, W., Sugiyama, T., Allis, C.D., and Grewal, S.I. (2005). The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. Mol Cell *20*, 173-185.

Yasui, D.H., Peddada, S., Bieda, M.C., Vallero, R.O., Hogart, A., Nagarajan, R.P., Thatcher, K.N., Farnham, P.J., and Lasalle, J.M. (2007). Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A *104*, 19416-19421.

Yoder, J.A., and Bestor, T.H. (1998). A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. Hum Mol Genet *7*, 279-284.

Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. Trends Genet *13*, 335-340.

Yoon, H.G., Chan, D.W., Reynolds, A.B., Qin, J., and Wong, J. (2003). N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. Mol Cell *12*, 723-734.

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., and Zoghbi, H.Y. (2005). Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. Proc Natl Acad Sci U S A *102*, 17551-17558.

Zacharias, H. (1995). Emil Heitz (1892-1965): chloroplasts, heterochromatin, and polytene chromosomes. Genetics *141*, 7-14.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., and Ecker, J.R. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189-1201.

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. Proc Natl Acad Sci U S A *100*, 6777-6782.

Zheng, B., Wang, Z., Li, S., Yu, B., Liu, J.Y., and Chen, X. (2009). Intergenic transcription by RNA polymerase II coordinates Pol IV and Pol V in siRNA-directed transcriptional gene silencing in Arabidopsis. Genes Dev *23*, 2850-2860.

Zhou, Y., Cambareri, E.B., and Kinsey, J.A. (2001). DNA methylation inhibits expression and transposition of the Neurospora Tad retrotransposon. Mol Genet Genomics *265*, 748-754.

Zhou, Z., Hong, E.J., Cohen, S., Zhao, W.N., Ho, H.Y., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C.*, et al.* (2006). Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. Neuron *52*, 255-269.

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.