THE UNIVERSITY
of EDINBURGH

The Molecular Basis of R133C Rett Syndrome

Dr Kyla J Brown

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2016

Declaration

The work presented in this thesis is my own, unless otherwise stated and has not been submitted for any other degree or professional qualification.

Kyla J. Brown
2016

Acknowledgements

The Bird lab's knowledge, patience and teaching have been indispensable in producing this work. I can still remember meeting Adrian and Jim Selfridge after securing this post to discuss potential projects and the looks on their faces as I admitted to never having handled a Gilson pipette before. Particular thanks must go to Jim Selfridge for guiding me through those shaky early days, producing those long sought after chimaeras and for many hours lamenting and navigating the complexity of the project. Jacky Guy shared all her carefully thought out protocols and took me under her cell culture and immunohistochemistry wing, which was bolstered by Dave Kelly. Sabine Lagger provided ChIP tutelage and was always available, no matter how busy, to discuss all matters "non-CpG". John Connelly generously walked me through protein, Martha Koerner and Justyna Cholewa-Waclaw through RNA and Matt Lyst through the reasons why most experiments would not work, along with some that would. Dina De Sousa and Cara Merusi provided a constant source of support and advice. I would also like to thank Alastair Kerr, Shaun Webb and Gabriele Schweikert for bioinformatic support, Christine Struthers and Joanne Ness for administrative support and Alan McClure for help in the animal unit. Mentorship from John Iredale, Brian Walker, Andrew Jackson, Adele Marston and of course, Adrian - which always came with a good dose of humour - was greatly appreciated.

I could not have completed this "basic science" journey without the love and support of my long-suffering husband, Richard, my parents, sister, friends and last but not least, my beautiful daughters, Erin and Julie who provided a much needed (and mostly welcome) source of distraction throughout.

Abstract

Rett syndrome is a debilitating autistic spectrum disorder affecting one in ten thousand girls. Patients develop normally for up to eighteen months before a period of regression involving stagnation in head growth, loss of speech, hand use and mobility. It is almost exclusively caused by mutation in Methyl CpG binding Protein 2 (MeCP2). MeCP2 has traditionally been thought of as a transcriptional repressor, although its exact function remains unknown and it has recently been shown that the protein can also bind to hydroxymethylation and non-CpG methylation, which occurs predominantly at CAC sites in the mature nervous system. Genotype-phenotype studies of the most common Rett-causing mutations in affected patients revealed that a missense mutation, R133C results in a milder form of Rett syndrome. The reasons for this are unclear, as the mutation lies right in the heart of the methylated DNA binding domain. Previous *in vitro* studies of R133C showed a severe deficit in binding to methylated cytosine. A subsequent study found that R133C binding to hydroxymethylated cytosine was specifically impaired, whereas binding to methylated cytosine was indistinguishable from wildtype. Defining the DNA binding impairment of $MeCP2^{R133C}$ would yield important insights into Rett disease pathophysiology and provide an explanation for the phenotypic spectrum seen in patients. To shed light on these matters, a novel mouse model of the R133C mutation was created. The R133C mouse had a phenotype that was less severe than other missense mutant mice, in terms of survival, growth, Rett-like phenotypic score and some behavioural paradigms thus recapitulating the patient data. At the molecular level in adult mouse brain, $MeCP2^{R133C}$ protein abundance was reduced. Immunohistochemistry showed that $MeCP2^{R133C}$ had an abnormal pattern of localisation in the nucleus of neurons. *In vitro* electrophoretic mobility shift assays suggested that $MeCP2^{R133C}$ binding to (hydroxy)methyl-CAC may be reduced to a greater extent than binding to mCpG. Chromatin immunoprecipitation experiments confirmed the deficit in binding to methylated sites and supported a disproportionate reduction in binding to methylation in a CAC sequence context. Analysis of adult mouse cerebellar gene expression revealed a subtle upregulation of long genes and downregulation of short genes. Based on these data, it is proposed that Rett

syndrome caused by the R133C mutation results from a combination of protein instability and defective binding to methylated DNA. Methyl-CAC binding is potentially abolished. The downstream biological consequence of this is a length-dependent deregulation of gene expression in the brain.

Lay Summary

Rett syndrome is an autistic spectrum disorder affecting one in ten thousand girls that results in severe disability. Affected patients develop normally for up to eighteen months before losing hand and language skills they had already achieved. Their mobility and growth are also reduced. It is almost exclusively caused by mutation in a protein called MeCP2 on the X chromosome. This protein binds to methylation in DNA and the result is thought to be that fewer genes are expressed. Recently it has been shown that MeCP2 can also bind to bits of DNA that would result in more genes being expressed (hydroxymethylation). If the clinical severity of Rett syndrome and the mutation causing it are assessed in large numbers of patients, the R133C mutation is reproducibly associated with milder clinical features. For example, patients may still be able to use their hands. This is unexpected as the mutation lies in a key part of the protein, which it uses to bind DNA. Previous studies have shown that $MeCP2^{R133C}$ can't bind methylated DNA. A subsequent study found that only binding to hydroxymethylated DNA and not methylated DNA was affected. If the $MeCP2^{R133C}$ binding deficit can be understood, the differences that we see between patients and the reason they get Rett syndrome can be disentangled. In order to answer this question, a mouse model of $MeCP2^{R133C}$ was made. Like the patients, the mice showed milder symptoms compared to other mutant mice. Reduced levels of mutant protein and an abnormal binding pattern in neurons were found by examining brain tissue. Binding to known methylated target sequences was impaired in the brain. Analysis of gene expression in a specific area of the brain – the cerebellum – showed that genes of long length were slightly more abundant in the mutant and genes of short length were slightly less abundant. The outcome of this study is that Rett syndrome caused by the R133C mutation is milder than Rett syndrome caused by other mutations because the protein is modestly reduced and it binds abnormally to methylated DNA leading to a subtle length dependent change in gene expression.

Table of Contents

Tables and Figures

Abbreviations

5-HIAA	5- hydroxyindoleacetic acid
A	Adenine
aa	Amino acids
BDNF	Brain derived neurotrophic factor
Bisulfite-seq	High throughput sequencing of DNA following bisulfite treatment
bp	base pair
BSA	Bovine serum albumin
C	Cytosine
CA1	Cornu Ammonis area 1
CA3	Cornu Ammonis area 3
CAC	Cytosine-adenine-cytosine base sequence
CD	Circular dichroism
CH	Cytosine followed by any base other than guanine
ChIP	Chromatin immunoprecipitation
ChIP-seq	High throughput sequencing of DNA recovered by ChIP
CMV	Cytomegalovirus
Co-IP	Co-immunoprecipitation
CpG	Cytosine base connected to adjacent guanine base
Cre	Cre recombinase
C-terminus	Carboxyl-terminus
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
Dnmt	DNA (cytosine-5)-methyltransferase
dNTPs	Deoxynucleotide triphosphates
DTT	Dithriothreitol
EB	Embryoid body
EEG	Electroencephalography
EGFP	Enhanced green fluorescent protein
EMSA	Electrophoretic mobility shift assay
EPSC	Excitatory post-synaptic currents

ES cells	Embryonic stem cells
FBS	Fetal bovine serum
FRAP	Fluorescence recovery after photobleaching
G	Guanine
H1	Histone 1
H3K4me1	Monomethylated lysine 4 of histone 3
H3K27ac	Acetylated lysine 27 of histone 3
hm	Hydroxymethylated
IAP elements	Intracisternal-A particle elements
iPSC	Induced pluripotent stem cell
IP	Immunoprecipitate
IQ	Intelligence quotient
L1 LINE1	L1 long interspersed nuclear elements
m	Methylated
Maj sat	Major satellite DNA
MBD	Methylated DNA binding domain
Mecp2	Methyl CpG binding protein 2 (mouse gene)
MECP2	Methyl CpG binding protein 2 (human gene)
MeCP2	Methyl CpG binding protein 2 (mouse and human protein)
miRNA	micro RNA
NCoR	Nuclear receptor co-repressor
NEB	New England Biolabs
NEO cassette	Neomycin resistance cassette
NID	NCoR interaction domain
NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance
NP	Neuronal precursor
N-terminus	Amino-terminus
OD	Optical density
PAC	P1-derived artificial chromosome
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline

PBST	Phosphate-buffered saline with Tween-20
PCR	Polymerase chain reaction
PMD	Partially methylated domain
qPCR	Quantitative PCR
R133C-GFP	MeCP2^{R133CEGFP}
R133C-GFP	*Mecp2^{R133CEGFP}*
R306C-GFP	MeCP2^{R306CEGFP}
R306C-GFP	*Mecp2^{R306CEGFP}*
RNA	Ribonucleic acid
RNAse	Ribonuclease
RNA-seq	High throughput sequencing of cDNA derived from RNA
RPM	Revolutions per minute
RTT	Rett-like phenotype in the mouse
SILAC	Stable isotope labelling with amino acids in culture
T	Thymine
T158M-GFP	MeCP2^{T158MEGFP}
T158M-GFP	*Mecp2^{T158MEGFP}*
TAB-seq	High throughput sequencing of DNA following Tet-assisted bisulfite treatment
TAE	Tris-acetate EDTA
TBE	Tris-borate EDTA
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween-20
TEMED	Tetramethylethylenediamine
Tet enzyme	Ten-eleven translocation dioxygenase
TRD	Transcriptional repression domain
Tris	Tris(hydroxymethyl)aminomethane
WT-GFP	MeCP2^{EGFP}
WT-GFP	*Mecp2^{EGFP}*
XCI	X chromosome inactivation

1. Introduction

1.1 Rett syndrome

Rett syndrome is a debilitating autistic spectrum disorder affecting 1 in 10,000 girls, making it one of the leading causes of genetic intellectual disability in females (Laurvick *et al*, 2006). Rett syndrome was first described in 1966 by Dr Andreas Rett, an Austrian paediatric neurologist. Although later to become his eponym, he referred to the characteristic constellation of symptoms and signs as "cerebral atrophy and hyperammonaemia in infancy" in his original publication in German. The syndrome only became internationally recognised in the 1980's when Dr Bengt Hagberg published a seminal paper in English describing the syndrome in 35 females and named it "Rett syndrome" (Hagberg *et al,* 1983). With the observation that Rett syndrome affected girls, a hunt for a genetic cause ensued. Exclusion mapping using rare familial cases identified Xq28 as the responsible region. Candidate genes were investigated and a landmark finding in Rett syndrome research came in 1999 when it was discovered that mutation in the *MECP2* gene, encoding Methyl-CpG binding protein 2, was responsible, thus making Rett syndrome one of an exclusive subset of complex neuropsychiatric disorders that is caused by mutation of one gene (Amir *et al,* 1999*)*. Therefore it is anticipated that modelling the disorder *in vitro* and *in vivo* will make a molecular understanding possible.

1.1.1 Clinical Features

Rett syndrome is characterised by a period of generally unremarkable development for 6 to 18 months. However, with retrospective assessment, there may be some hypotonia and subtle abnormality of hand-eye coordination during this time (Einspieler *et al,* 2005). Classically, the syndrome then unfolds in four clinical stages, although there is wide variability in the particular features presented by individual females (Kerr & Engerström, 2001; Smeets *et al,* 2011).

Stage I: The Early Onset Stagnation Period (between 6 and 18 months). There is an alteration in interaction that can be abrupt, for example infants may become irritable

or less demanding. Speech and motor development begin to show signs of delay however there can be babbling, single words and sitting.

Stage II: The Rapid Developmental Regression Period (between 18 months and 4 years). This phase may be heralded by days to weeks of general malaise and distress and is characterised by a loss of acquired skills. Autistic features become apparent, such as reduced communication and diminished interest in surroundings although eye contact is preserved. Speech and hand use regress and the child begins to self stimulate instead of interacting with her environment, for example by head-tapping. Vacant spells, blinking, panting, spitting, hypersalivation, hyperventilation, facial grimacing and twitching indicate brain stem immaturity. Head growth typically decelerates, resulting ultimately in microcephaly with head size around three standard deviations below the mean.

The Pseudo-Stationary Stage III. Characteristic hand stereotypies emerge with repetitive movements in the midline such as hand wringing, clapping or washing. Motor dysfunction is apparent in apraxia, ataxia, gait abnormality and muscle dystonia resulting in a progressive neurogenic scoliosis. Respiratory dysfunction occurs during wakefulness with hyperventilation, hypoventilation, breath holding, apnoeas, apneustic breathing, or Valsalva manoeuvre breathing which frequently leads to air swallowing. Seizures ensue and affect up to 80% of girls (Steffenburg *et al,* 2001). Somatic growth is reduced, ultimately falling to two standard deviations below the mean. Symptoms and signs of autonomic dysfunction manifest, such as cold, small feet, decreased intestinal motility and abnormalities of the electrocardiogram. Other frequent features are an uncoordinated swallow, gastro-oesophageal reflux, bruxism, impaired nociception and sleep disturbance. There may be some recovery of skills and character traits acquired prior to *Stage II*. The child appears happy and interested and will communicate needs with intense eye-pointing. Inattention, hyperactivity, periods of agitation and unexplained night laughing are common. Although severely intellectually disabled, girls are still able to learn and experience new things.

The Late Motor Deterioration Stage IV is prolonged and commences when mobility decreases with muscle wasting and rigidity. Girls ultimately become wheelchair dependent.

Each stage is associated with a characteristic pattern on electroencephalography (EEG): in *Stage I* the EEG is normal with focal spikes emerging in *Stage II*. *Stage III-IV* EEG's are more variable with pseudoperiodic delta pattern and generalised periodic spike activity (Hagne *et al,* 1989). Abnormalities have also been found in event related potentials with reduced amplitude and longer latency (Stauder *et al,* 2006).

1.1.2 Rett syndrome in males

Males only carry one copy of the *MECP2* gene as they only have one X chromosome. Rett syndrome was previously thought to be lethal prior to birth, however male infants have been reported with mutations in the *MECP2* gene that cause typical Rett syndrome in females (Kankirawatana *et al,* 2006). The clinical course is severe with a progressive neonatal encephalopathy, failure to thrive, respiratory insufficiency, microcephaly, intractable seizures, hypotonia leading to progressive rigidity and abnormal motor movements. Affected patients die in early infancy (Schüle *et al,* 2008). Alternatively, males can display signs of typical Rett syndrome in the context of an abnormal karyotype (Maiwald *et al,* 2002) or somatic mosaicism, where an *MECP2* mutation arises during development of the embryo and is not present in all cells (Down, 2000).

1.1.3 Diagnostic Criteria

Rett syndrome remains a clinical diagnosis, due to the association of *MECP2* mutations with other clinical presentations (described in 1.6.1). Criteria for diagnosing Rett syndrome were established in 1985 for females (Hagberg *et al*, 1985) and revised to include males in 1988 (The Rett Syndrome Diagnostic Work Group 1988), variant presentations in 1994 (Hagberg & Skjeldal, 1994) (see 1.1.4), and to broaden the diagnosis so that deceleration of head growth and a period of normal development were no longer mandatory in 2002 (Hagberg *et al,* 2002). In

2010, the RettSearch Consortium, an international network of clinically orientated Rett syndrome researchers, produced revised, simplified diagnostic criteria that emphasised the importance of the regressive phase with subsequent stabilisation and clarified the requirement for supportive criteria (see table 1) (Neul *et al,* 2010). There was good concordance with the 2002 criteria in a large population of patients (Percy *et al,* 2010). The goal was to streamline and unify the study of Rett syndrome, removing any ambiguity in making a diagnosis.

1.1.4 Phenotypically Similar Presentations

There are three common related disorders that present in a similar fashion to classical Rett syndrome. Historically they have been referred to as: the Preserved Speech Variant (Zapella, 1992), the Early Seizure Variant (Hanefeld, 1985) and the Congenital Variant (Rolando, 1985).

In the Preserved Speech Variant, the clinical picture is milder with a later regressive phase and prolonged pseudostationary phase. Patients tend to recover their language following the regressive phase and can use single words or phrases. Girls have an intelligence quotient (IQ) of up to 50 and some retention in hand use. Although autistic behaviours are common, epilepsy, autonomic dysfunction and reduced head circumference are rare. In addition, girls can have normal height and weight with milder spinal deformities (Neul *et al*, 2010).

The Early Seizure Variant does not present with as many features of typical Rett syndrome, but instead is characterised by onset of infantile spasms and refractory myoclonic epilepsy within the first five months of life and prior to the regressive phase (Neul *et al*, 2010). This variant is associated with mutation in the *CDKL5* gene on the X chromosome, which encodes cyclin-dependent kinase-like 5 (Tao *et al,* 2004; Bahi-Buisson *et al,* 2008). This protein shares homology with members of the mitogen-activated protein (MAP) kinase and cyclin-dependent kinase (CDK) families and has been proposed to phosphorylate MeCP2 *in vitro* (Mari *et al,* 2005). Interestingly, this activity is abolished with *CDKL5* mutations (Bertani *et al,* 2006).

Consider Rett syndrome diagnosis when postnatal deceleration of head growth is observed.
Required for typical or classic Rett syndrome A period of regression followed by recovery or stabilisation 1. All main and all exclusion criteria 2. Supportive criteria are not required, although often present in typical Rett syndrome
Required for atypical or variant Rett syndrome 1. A period of regression followed by recovery or stabilisation 2. At least 2 of the 4 main criteria 3. 5 of 11 supportive criteria
Main criteria 1. Partial or complete loss of acquired purposeful hand skills 2. Partial or complete loss of acquired spoken language 3. Gait abnormalities: impaired (dyspraxia) or absence of ability (apraxia) 4. Stereotypic hand movements such as hand wringing/squeezing, clapping/tapping, mouthing and washing/rubbing automatisms
Exclusion criteria for typical Rett syndrome 1. Brain injury secondary to trauma (peri- or postnatally), neurometabolic disease or severe infection that cause neurological problems 2. Grossly abnormal psychomotor development in the first 6 months of life
Supportive criteria for atypical Rett syndrome 1. Breathing disturbances when awake 2. Bruxism when awake 3. Impaired sleep pattern 4. Abnormal muscle tone 5. Peripheral vasomotor disturbances 6. Scoliosis/kyphosis 7. Growth retardation 8. Small cold hands and feet 9. Inappropriate laughing/screaming spells 10. Diminished sensitivity to pain 11. Intense eye communication – “eye-pointing”

Table 1. Rett syndrome diagnostic criteria (Neul *et al* 2010)

The Congenital Variant runs a severe course and is characterised by grossly abnormal early development with hypotonia, early postnatal microcephaly, regression within five months of life, frequent autonomic abnormalities as well as tongue stereotypies and specific jerky movements of the limbs. Girls rarely have the

intense Rett eye gaze (Neul *et al,* 2010). This variant has been associated with mutations in the autosomal *FOXG1* gene, encoding forkhead box protein G1, a brain-specific transcriptional repressor essential for telencephalon development (Ariani *et al,* 2008).

1.1.5 Clinical severity rating scales

There has been significant disparity in the means different Rett syndrome researchers have chosen to measure clinical severity, which has impacted on meaningful comparison of results between studies. Four commonly used scales will be briefly described as they are referenced repeatedly in later sections.

The Kerr scale was developed by an international group in 2001 in order to define a minimal data collection set for Rett syndrome researchers investigating genotype-phenotype correlation (Kerr *et al,* 2001). The scale contains twenty items, scored from 0–2 (maximum 40). A higher score indicates greater severity. Clinical features considered are: head circumference during the first year, developmental progress in the first year, present head circumference, weight, height, muscle tone, spine posture, joint contractures, gross motor function, hand stereotypy, other involuntary movements, voluntary hand use, oro-motor difficulty, intellectual disability, speech, epilepsy, awake breathing rhythm, peripheral circulation, mood disturbance and sleep disturbance.

The Percy scale contains fifteen items scored up to five, with a maximum score of 50 (Colvin *et al,* 2003; adapted from Amir *et al,* 2000). A higher score denotes a more severe phenotype. Clinical features considered are age at onset of regression, ambulation, respiratory dysfunction, head growth, somatic growth, sitting, crawling, hand use, non-verbal communication, language, seizures, feeding, onset of stereotypies, autonomic dysfunction and scoliosis.

The Pineda scale contains ten items scored up to four, with a maximum of 30 (Colvin *et al,* 2003; adapted from Monrós *et al,* 2001). A higher score denotes a more severe phenotype. Clinical features considered are age at loss of social interaction, head

growth, sitting alone, ambulation, language, epilepsy, respiratory function, hand use, air swallowing/bloating and onset of stereotypies.

The WeeFIM is a functional independence measure for children and measures independent functioning with a maximum composite score of 126. Subscales for self-care, mobility and cognition are scored out of 56, 35 and 35, respectively (Msall *et al,* 1994). A lower score denotes greater impairment and dependence.

1.1.6 Neuropathology

People with Rett syndrome have small brains (Casanova *et al*, 1991) with no sign of gross malformation, degeneration, demyelination or inflammation (Armstrong *et al,* 2002). The brain is the only organ that is disproportionately small in relation to the patient's height and this relative reduction in weight does not progress with age (Armstrong *et al,* 2005). There is a reduction of grey matter throughout the cortex, most so in prefrontal, posterior frontal and anterior temporal regions (Subramaniam *et al,* 1997). This is related to significantly reduced dendritic arborisation of frontal, motor and subicular cortical pyramidal neurons (Armstrong *et al,* 1998), dendritic spine loss in the frontal cortex (Belichenko & Dahlström, 1995), decreased neuronal size in the cortex, thalamus, basal ganglia, amygdala and hippocampus with increased packing density in the hippocampus (Bauman *et al,* 1995). The neurons of the pars compacta in the substantia nigra contain less melanin than age matched controls (Jellinger & Seitelberger, 1986) and a recent study documented significantly reduced levels of a dopamine metabolite, homovanillic acid and serotonin metabolite, 5-hydroxyindoleacetic acid (5-HIAA) in cerebrospinal fluid of 64 patients with Rett syndrome compared to 200 controls (Samaco *et al,* 2009). However, generally there have not been consistent differences found in neurotransmitter profiles in Rett syndrome patients (Armstrong, 2002). The neuropathological findings are consistent with a failure of normal brain development, rather than a degenerative process.

1.1.7 Prognosis

There is currently no cure for Rett syndrome and the management principally focuses on individualised rehabilitation programs led by a multidisciplinary team, as well as control of seizures. Mortality risk is increased, with sudden unexplained death mainly attributed to defective cardiorespiratory dysfunction. A database study of 332 Australian females with Rett syndrome revealed that in those that had died by the time of study, the average age of death was 16 years and 7 months (Freilinger *et al,* 2010). The majority of these early deaths were related to respiratory causes (infection, failure or aspiration) or seizures. The survival rate from this analysis was 64.9% at the age of 32. This corroborated well with a survival estimate of 70% at 35 years in North America (Kirby *et al,* 2010). Therefore Rett syndrome represents a severe, debilitating disease resulting in reduced quality of life and lifespan for those afflicted.

1.2 The methylation landscape of the brain

Since the discovery in 1999 that the gene responsible for Rett syndrome was *MECP2* (Amir *et al,* 1999), there have been significant advances in understanding the function of Methyl CpG binding Protein 2 (MeCP2) and the manner in which it interacts with the unique genomic environment of the brain, generating potential hypotheses for the molecular basis of Rett syndrome. Ultimately the goal of this research is to further understanding of neurodevelopmental disease and generate effective therapies to alleviate suffering in patients with Rett syndrome. Rett is a neuropsychiatric syndrome and MeCP2 is most highly expressed in neurons (Shabazian *et al*, 2002; Kishi *et al*, 2004), so in order to review the function of MeCP2, first the genomic epigenetic environment in which the protein operates will be considered.

Epigenetic mechanisms add an extra layer of complexity to the biological information stored in the deoxyribonucleic acid (DNA) sequence comprised of four bases – adenine (A), thymine (T), cytosine (C) and guanine (G). One such mechanism is methylation of cytosine (Figure 1.2), which has been referred to as the

"fifth base" (Lister *et al,* 2009). In mammals this predominantly occurs at cytosines when followed by guanines, forming self-complementary, symmetrical CpG pairs (Sinsheimer, 1955). The mark is established by the coordinated activity of three DNA (cytosine-5)-methyltransferase (Dnmt) enzymes. Dnmt3a and Dnmt3b place methylation *de novo* and Dnmt1 maintains methylation following DNA replication (Bestor, 1988; Okano *et al,* 1998). CpG methylation (mCpG) is established during embryogenesis in 70-80% of CpG's, widely distributed at low frequency and the genome is globally methylated. Unmethylated CpG's are generally clustered into discrete regions known as CpG islands, which are associated with gene promoters in approximately two thirds of cases (Antequera & Bird, 1999). When a CpG island gains methylation during development, the associated promoter is switched off. Thus, DNA methylation plays critical roles in genomic imprinting, X-chromosome inactivation (XCI), cellular differentiation and development (Bird, 2002).

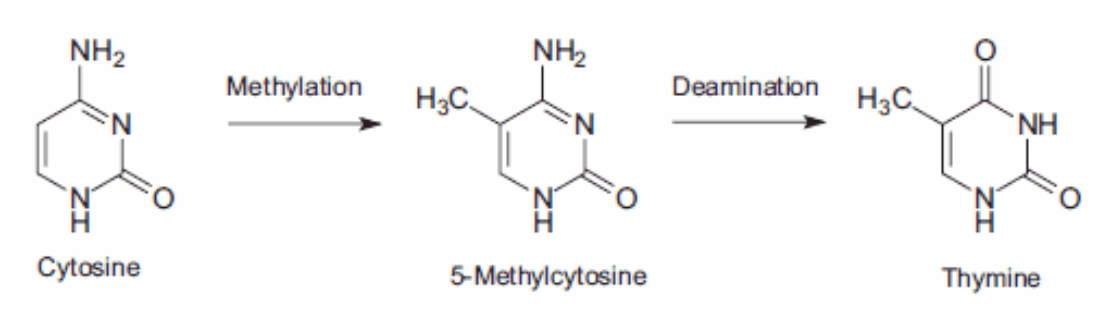


Figure 1.2 Cytosine methylation

Cytosine can be methylated at the carbon-5 position by Dnmt enzymes to form 5-methylcytosine (mC). Spontaneous deamination results in thymine. Adapted from sigmaaldrich.com.

The location of mCpG can be determined at single base pair resolution by bisulfite sequencing (Frommer *et al,* 1992). When DNA is treated with sodium bisulfite, unmethylated cytosines are converted to uracil, whereas mC's are not. Following amplification and sequencing, any remaining C's can be identified as methylated in the original material. Bisulfite treatment followed by high throughput sequencing (bisulfite-seq) has uncovered the mCpG landscape genome wide. The mCpG level is high over genes, intergenic regions and repetitive DNA and relatively low over

promoters and distal regulatory regions. Methylation frequency of individual CpG's is high (80-100%) with a separate low frequency population corresponding to CpG islands, so-called "low methylation regions" or LMRs and "partially methylated domains" or PMDs, which are dynamically methylated. This results in a bimodal distribution of mCpG (Lister *et al,* 2009; Stadler *et al,* 2011; Ziller *et al,* 2011). Interestingly there is little contribution of PMD regions in the neuronal methylome, suggesting alternative mechanisms of gene expression regulation (Lister *et al,* 2013).

In keeping with this, in recent years it has become apparent that the methylome of the brain is more complex than other terminally differentiated tissues. In addition to mCpG, the brain harbours significant proportions of cytosine that are alternatively modified, thus expanding the repertoire of methylated sites.

Non-CpG methylation, or CH methylation (mCH), is methylation occurring at cytosines followed by bases other than guanine, where H can be A, C or T. Previously thought to be an artefact, its existence was confirmed in mice (Ramsahoye *et al,* 2000) and humans (Lister *et al,* 2009) within the last twenty years. mCH is different from mCpG for a number of reasons. Firstly, the level of mCH is very low in most somatic tissues but high in neurons with appreciable amounts in embryonic stem (ES) cells, glia and oocytes (Ziller *et al,* 2013; Guo *et al,* 2014; He & Ecker, 2015). Secondly, for each CH site, only a small proportion of alleles is methylated, so the methylation level of each mCH site is approximately 20% (Lister *et al,* 2013). Thirdly, the signal is widely distributed with no obvious clustering of sites into island-like structures. mCH appears to correlate with the distribution of mCpG and is relatively depleted over highly expressed genes although it seems to be individually excluded from inaccessible regions of chromatin (Ziller *et al*, 2013; Lister *et al,* 2013). Fourthly, it accumulates with maturation of the nervous system, predominantly in neurons and most rapidly in the early postnatal brain corresponding to the primary phase of synaptogenesis in human and mouse (Lister *et al*, 2013). Finally, in brain the mark is established solely through *de novo* methylation with Dnmt3a being the likely writer. Dnmt3a expression in brain increases in the postnatal period in concert with the accumulation of mCH and knock down of Dnmt3a results

in the loss of mCH, but not mCpG (Ziller *et al,* 2013; Guo *et al,* 2014; Gabel *et al,* 2015). Interestingly disruption of Dnmt3a expression before this time results in a neurological phenotype (Nguyen *et al,* 2007; Tatton-Brown *et al,* 2014) whereas disruption following this developmental phase does not (Feng *et al,* 2010) indicating that this may be a critical period of mCH establishment.

Only 1.3-1.5% of CH is methylated in adult mouse and human brain (Lister *et al,* 2013). However, given the abundance of CH and the underrepresentation of CpG in the genome (Bird *et al,* 1980), this is still an appreciable proportion of the methylome. In a pure neuronal population mCH represents half of the human and over a third of the mouse methylome (Lister *et al,* 2013). In cells of the nervous system mCH occurs mainly in a CAC context whereas in ES cells it occurs mainly in a CAG context (Lister *et al,* 2009; Xie *et al,* 2012; Guo *et al,* 2014; Varley *et al,* 2013) implying a reconfiguration of the mark with differentiation of cells. This may be through passive dilution via cell division, followed by accumulation in differentiated non-dividing cells or through active and specific mechanisms in a non-random fashion. In favour of the latter, there is evidence that mCH placement can correlate between biological or sample replicates and is depleted in "mCH deserts" over olfactory receptor and immunoglobin genes. Furthermore, there appears to be a cell type-specific distribution in that glial-specific genes are CH hypermethylated in neurons and *vice versa* (Lister *et al,* 2013).

The function of mCH is yet to be determined, but it has been associated with repression of gene expression *in vivo* in mouse frontal cortex (Xie *et al,* 2012) and mouse dentate gyrus (Guo *et al,* 2014) and *in vitro* using methylated reporter constructs transfected into cultured neurons (Guo *et al,* 2014). Interestingly, when mCH occurs in a CAG context in pluripotent cells, it is not associated with repression of genes (Lister *et al,* 2009; Ziller *et al,* 2011). Therefore there may be two distinct functions of mCH depending on its sequence context and resulting binding partners. In either case, this potentially revises opinion on regulation of gene expression by methylation, which previously has centred on mCpG (Luo & Ecker, 2015).

In addition to mCH, hydroxymethylation (hm) of cytosines accumulates with maturation of the nervous system. HmC is formed when mC is oxidised by Ten-eleven translocation (Tet) dioxygenases (Tahiliani *et al,* 2009). This may be part of a demethylation pathway with subsequent oxidation steps or hmC may exist in the steady state as an epigenetic regulator in its own right (Kinde *et al,* 2015). The abundance of hmC has been overlooked, as traditional methods of identifying mC using sodium bisulfite treatment cannot distinguish the two marks. Its importance in mammals was brought to the forefront when it was identified in high abundance in cerebellar tissue using a combination of thin layer chromatography, high performance liquid chromatography and mass spectrometry (Kriaucionis *et al,* 2009). With the advent of Tet-assisted bisulfite (TAB) sequencing, where a glucosylation step prior to treatment with Tet enzymes protects hmC from further oxidation, the modification could be mapped at single base pair resolution across the genome (Yu *et al,* 2012). In neurons, hmC is enriched over distal regulatory elements and gene bodies, depleted at transcription start sites and correlates with gene activity (Song *et al,* 2011; Szulwach *et al,* 2011; Mellen *et al,* 2012).

Levels of hmC accumulate with postnatal development of the nervous system and are eventually 5-10 fold higher in the brain than in any other tissue, constituting approximately 40% of all modified CpG's in cerebellum and nearly 25% in mouse cortex: almost a fifth of the methylome (Kriaucionis *et al,* 2009; Lister *et al,* 2013). Nearly 100% of hmC is in a CpG context in mouse cortex (Lister *et al,* 2013) and human brain (Wen *et al,* 2014) suggesting that hmCH may not be a biologically relevant signal. These figures may be a product of biased Tet enzyme conversion in TAB-seq and it will be interesting to see if advances in technology revise these estimates.

The significant postnatal revision of the (hydroxy)methylome in neurons, with build-up of mCH and hmCpG establishes a unique cellular environment. The significance is unknown and linking its importance to the processes of neurodevelopment will be a future challenge. Deficiency of "writers" of this methylation lead to neurological

impairment (Nguyen *et al,* 2007; Zhang *et al,* 2013) Understanding "readers" of this (hydroxy)methylome will be key and in this regard, the study of *MECP2* and Rett syndrome may yield some insights.

1.3 Methyl CpG binding Protein 2

Methyl CpG binding protein 2 (MeCP2) was first characterised in 1992 in a search for proteins that might be candidates for interpretation of cytosine methylation (Lewis *et al,* 1992). It is one of a family of proteins able to bind methylated DNA, although mutation of the other members does not result in similar neurological phenotypes (Hendrich & Tweedie, 2003). MeCP2 is an abundant nuclear protein that is 5-10 times more highly expressed in post-mitotic neurons than other tissues with appreciable amounts also in glia, lung and spleen (Shabazian *et al,* 2002; Kishi *et al,* 2004). The number of molecules per neuronal nucleus in the adult mouse was estimated by quantitative western blotting to be approximately 16 million, which would be sufficient to nearly saturate the 40 million predicted mCpG sites in the neuronal genome (Skene *et al*, 2010). Expression of MeCP2 increases postnatally until adulthood. This closely parallels the remodelling of the neuronal methylome (see 1.2) and corresponds to a time of intense synaptogenesis, suggesting a linked function.

MeCP2 is present in all vertebrates and is well conserved in mammals with 95% identity between human and mouse (Guy *et al,* 2011). The protein has three main functional domains: the methylated DNA binding domain (MBD), nuclear localisation signal (NLS) and transcriptional repression domain (TRD). The MBD (amino acids 78-162) is the minimal domain required for binding to a mCpG dinucleotide *in vitro* (Nan *et al,* 1993). The NLS (amino acids 255-271) is the minimum polypeptide fragment with nuclear subcellular localisation of MeCP2 when transfected into murine fibroblasts (Nan *et al,* 1996). The TRD (amino acids 207-310) is the minimal polypeptide fragment able to repress transcription of reporter constructs using a GAL4 binding system in murine fibroblasts (Nan *et al,* 1997). This was later refined by exclusion mapping to pinpoint the minimum region required to repress transcription *in vitro* and interact with the nuclear receptor co-repressor

complex (NCoR) - the NCoR interaction domain (NID) (amino acids 285-309) (Lyst *et al,* 2013). MeCP2 with its main functional domains is shown in Figure 1.3.

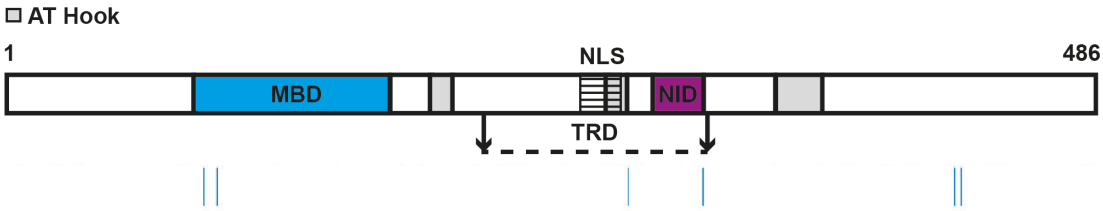


Figure 1.3 Methyl CpG binding Protein 2 (MeCP2)
Schematic representation of MeCP2 protein, isoform e2 (486 amino acids). The major functional domains are shown: the methylated DNA binding domain (MBD, blue), the nuclear localisation signal (NLS, shaded) and the NCoR Interaction Domain (NID, pink), along with three putative AT hooks (grey). The location of the originally defined transcriptional repression domain (TRD) is indicated with arrows and a dotted line. Regulatory phosphorylation sites are plotted in blue (adapted from Lyst & Bird, 2015).

The *MECP2* gene contains four exons and there are two possible protein isoforms, variant in their N-termini and inclusion of exon 2. MeCP2α/e1 contains exon 1, which is spliced directly to exon 3. MeCP2β/e2 contains exons 2-4. MeCP2α/e1 is the more ancestral form, is more efficiently translated and is approximately ten times more abundant in mouse brain than MeCP2β/e2 (Kraukionis *et al,* 2004). There are three possible *MECP2* transcripts from alternative use of polyadenylation sites in the 3' untranslated region (1.9, 7.5 and 10kb) (D'Esposito *et al,* 1996), which may allow regulation of MeCP2 by micro-RNA binding (Klein *et al*, 2007; Han *et al*, 2013). Additionally, MeCP2 contains multiple phosphorylation sites (Figure 1.3) and there is evidence that these may serve as a means of regulation in response to neuronal activity (Chen *et al,* 2003; Zhou *et al,* 2006; Tao *et al,* 2009; Ebert *et al,* 2013).

1.3.1 Binding Properties of MeCP2

MeCP2 expression correlates with evolution of the neuronal (hydroxy)methylome, but what is the evidence that it can interact with the various forms of cytosine modification?

1.3.2 Binding to mCpG

MeCP2 was identified by virtue of its ability to selectively bind a single symmetrical mCpG pair, independent of sequence context, in a southwestern assay (Lewis *et al,* 1992). This was confirmed with definition of the minimal fragment required for this function, the MBD (amino acids 78-162) using an electrophoretic mobility shift assay (EMSA) (Nan *et al,* 1993). In the mouse, major satellite DNA makes up 7% of the genome and is comprised of a 234 base pair repeat containing eight mCpG sites (Horz & Altenburger, 1981). Therefore the satellite DNA represents almost half of mouse genomic mCpG. MeCP2 bound with high affinity to fully methylated major satellite DNA probes by southwestern assay (Lewis *et al,* 1992). Following these early studies, mCpG binding specificity has been extensively replicated *in vitro* (Meehan *et al,* 1992; Hendrich & Bird, 1998; Ballestar *et al,* 2000; Yusufzai *et al,* 2000; Free *et al,* 2001; Valinuck *et al,* 2004; Galvao & Thomas, 2005).

Structural analysis of MeCP2 has reinforced this evidence and explained it in atomic detail. The solution structure of the MBD was predicted using nuclear magnetic resonance (NMR) spectroscopy (Wakefield et al, 1999). It appeared to form a novel wedge shape with a four-stranded antiparallel β-sheet comprising one face and a three-turn α-helix comprising the other. The wedge was asymmetrical, which was surprising given the symmetry of the mCpG dyad binding site. The authors predicted that aspartate at position 121 interacted with the flexible side chains of either arginine 111 or arginine 133 and buttressed them into an ordered conformation, in a manner similar to zinc finger motifs. Subsequently, and in agreement, the crystal X-ray structure of the MBD in complex with double-stranded DNA containing a single mCpG dyad was solved (Ho *et al,* 2008). MBD residues made multiple close contacts with the methylated cytosines. Two hydrophobic arginine fingers – arginine 111 and arginine 133 – projected in to the major groove of the DNA, making contact with the guanine residues. This structure was stabilised by several water molecules and constrained by salt bridges formed with the carboxylates of aspartate 121 and glutamate 137.

Therefore there was convincing evidence *in vitro* that MeCP2 could associate with fragments of methylated DNA. In eukaryotic cells DNA is further organised by wrapping 147bp stretches around histone protein octamers, forming nucleosomes. Nucleosomes are then spaced by linker DNA as chromatin with a "beads on a string" appearance (Luger et al, 1997). There is substantial evidence that MeCP2 associates with DNA in the context of chromatin and that in this respect also, there is a preference for mCpG.

MeCP2 initially appeared chromatin-associated *in vitro*, being more readily extracted from rat brain nuclei in the presence of increasing concentrations of micrococcal nuclease (Meehan *et al,* 1992). Murine major satellite DNA is concentrated in the pericentromeric heterochromatin and can be visualised easily using 4',6 diamidino-2-phenylindole (DAPI) or Hoechst 33258 staining by microscopy (Miller *et al,* 1974). MeCP2 was observed to be associated with this pericentromeric heterochromatin in murine fibroblasts and murine ES cells with a characteristic punctate binding pattern (Lewis *et al*, 1992; Nan *et al,* 1996). When methylation was reduced to 5% by homozygous mutation of a methyltransferase gene in ES cells, this binding pattern was abolished and MeCP2 appeared dispersed (Nan *et al,* 1996).

Association with methylated DNA has been confirmed by chromatin immunoprecipitation (ChIP) experiments at single genomic loci, such as the methylated maternal allele of imprinted genes (Gregory *et al,* 2001) or the brain derived neurotrophic factor (*BDNF)* locus (Chen *et al,* 2003). By utilising MeCP2 ChIP followed by high throughput sequencing (ChIP-seq), association with chromatin can be visualised genome-wide. One such study in mouse brain indicated that MeCP2 coated the genome and binding coincided with the occurrence of mCpG (Skene *et al,* 2010). This broad binding profile was corroborated in mouse brain (Baker *et al,* 2013) and cultured neurons (Cohen *et al,* 2011). Methylation-dependent binding was confirmed by ChIP-seq in mouse ES cells and derived neurons where MeCP2 preferentially bound mCpG rich sequences and linearly tracked local methylation density, regardless of the underlying genomic features (Baubec *et al,* 2013). Therefore relative enrichment was found over exons, promoters with

intermediate CpG content, and some classes of repetitive DNA, which contain the highest density of mCpG. The distribution of MeCP2 altered dramatically with depletion of methylation in ES cells to regions of accessible chromatin, marked by DNAse hypersensitivity and histone modifications representing actively transcribed chromatin (H3K4me1 and H3K27ac). Therefore genome wide assessment of binding in cultured neurons and brain indicates that MeCP2 tracks mCpG density.

1.3.3 MeCP2 binding to unmethylated DNA

In addition to binding with high specificity to methylated cytosines, there has been some evidence that MeCP2 can also bind non-specifically to unmethylated DNA. Early experimental evidence during the characterisation of MeCP2 detected lower affinity binding to unmethylated probes by southwestern assays and EMSA's (Lewis *et al,* 1992; Meehan *et al,* 1992). There was also evidence that the MBD alone could bind to unmethylated probes by EMSA with a dissociation constant that was 100-fold higher than for methylated probes (Valinuck *et al,* 2004). This binding has not just been reported for oligonucleotide probes, as in addition human MeCP2 could compact unmethylated reconstituted nucleosomal arrays *in vitro* into large ellipsoidal structures (Georgel *et al,* 2003). Interestingly, although an earlier study indicated a dependence on methylation for association with mouse pericentromeric heterochromatin (Nan *et al,* 1996), this finding was not replicated in ES cells that were triple knock-out for *Dnmt* genes. In fact, the characteristic punctate pattern of binding was unperturbed (Baubec *et al,* 2013). An association with unmethylated DNA has also been replicated by ChIP experiments at single genomic loci, for example the imprinted H19 locus, where MeCP2 bound preferentially to the unmethylated maternal allele (Kemohan *et al,* 2010) and genome wide: MeCP2 ChIP followed by a custom microarray analysis in a human neuronal cell line revealed that the majority of promoters bound by MeCP2 were of expressed genes and the vast majority (>90%) were not highly methylated (Yasui *et al,* 2007). However, an important caveat to this finding was that the MeCP2 antibody the authors used for ChIP was shown to discriminate poorly between protein extracts from wildtype mice and *Mecp2*-null mice, so these may not have been genuine MeCP2 binding sites (Guy *et al*, 2011).

The fact that binding to methylated DNA was preserved in the presence of an excess of unmethylated competitor DNA led to the speculation that binding to methylated DNA and unmethylated DNA was achieved with different regions of the protein. To this end, MeCP2 contains (R)GRP(K) motifs, or AT hooks, a feature of high mobility group (HMG) proteins that bind to AT rich DNA (Johnson *et al,* 1989) and it was suggested that this may be the basis of non-specific association with DNA. In keeping with this, MeCP2 required the presence of at least four AT bases adjacent to the mCpG dyad to bind to DNA probes with high affinity *in vitro* and *in vivo*, for example at the brain derived neurotrophic factor (*BDNF)* locus (Klose *et al,* 2005). However, the same study indicated that mutating an AT hook did not alter binding selectivity. The importance of the AT hook motifs has been investigated using mouse models and is discussed in section 1.7.1.

Structural studies have lent support to non-specific binding, indicating that the thermal stability of MeCP2 was increased on association with unmethylated DNA (Ghosh *et al,* 2008). Furthermore, by using hydrogen-deuterium exchange mass spectrometry it was shown that the MBD recognises methylated and unmethylated DNA in a similar manner (Hansen *et al,* 2011). Moreover there has been some evidence that additional regions of MeCP2 downstream of the MBD can independently bind DNA and compact chromatin (Adams *et al,* 2007; Nikitina *et al,* 2007; Kumar *et al,* 2008; Ghosh *et al,* 2010).

Therefore although there is a wealth of evidence to suggest that MeCP2 specifically binds to DNA methylated in a CpG context, this abundant protein can also associate with relatively unmethylated sites in the genome, perhaps utilising AT hook motifs and regions other than the MBD. This non-specific binding may be a mechanism for immediate association with DNA, in order to then migrate towards methylated binding sites (Hansen *et al,* 2010).

1.3.4 MeCP2 binding to hmC

Following the discovery of moderate levels of hydroxymethylation in the mature nervous system (Kriaucionis *et al,* 2009), one group searched for protein binders that may interpret this DNA moiety. A protein species was pulled down from rodent brain nuclear extracts using a hydroxymethylated probe, resolved by SDS-PAGE and identified by mass spectrometry as MeCP2. This species was not pulled down in *Mecp2*-null brain extracts. Interaction was also demonstrated by surface plasmon resonance and EMSA, which was sensitive to glycosylation of hmC (Mellen *et al,* 2012). No other hmC binders were identified in this study. MeCP2 was also identified as an hmC reader in mouse ES cell nuclear extracts (but not differentiated neuronal precursors or neurons) by quantitative mass spectrometry (Spruijt *et al,* 2013).

On the contrary, earlier work had suggested that MBD binding to a symmetrical hmCpG dyad was almost equivalent to binding unmodified cytosine by EMSA *i.e.* affinity was reduced 100-fold (Valinuck *et al,* 2004). Similar results were found using fluorescence polarisation and amino acids 77-205 of MeCP2 (Hashimoto *et al,* 2012). Furthermore, data from MeCP2 ChIP-seq combined with TAB-seq in mouse brain, suggests that MeCP2 does not bind to hmCpG *in vivo*. A relative depletion of MeCP2 was detected at gene bodies enriched for hmCpG (Gabel *et al,* 2015) and no correlation was found between MeCP2 binding and hmCpG content across genes (Chen *et al,* 2015).

These conflicting results can perhaps be explained when the probes used to bind MeCP2 are considered. The probe used by Spruijt *et al* contained four consecutively modified CpG dinucleotides in a 27 base pair oligonucleotide. This artificial concentration of the moiety may have captured low affinity binding. Additionally, their experimental protocol changed from ES cells (where they utilised stable isotope labelling with amino acids in culture (SILAC) medium to label cellular proteins) to neuronal precursors and brain extracts (as there was no SILAC-compatible neurobasal medium, so label-free quantification was employed instead). Binding was only detected in ES cells and so may have been an artefact of the protocol. The

binding probes used by Mellen *et al* contained hydroxymethylation in a CpG and CA context at multiple sites, with hmCpG being relatively underrepresented. This raises the possibility that MeCP2 was binding to hmCA in these experiments. Indeed a double-stranded oligonucleotide probe containing hmCA was found to be a superior competitor species to one with hmCpG by EMSA for both the MBD and full length MeCP2, indicating binding *in vitro* (Gabel *et al,* 2015). Moreover, MeCP2 binding was correlated with genes containing higher levels of hmCA *in vivo* in mouse brain (Gabel *et al,* 2015). Therefore overall, it seems unlikely that MeCP2 binds to the vast majority of hmC in the brain.

1.3.5 MeCP2 binding to mCH

Early work showed that MeCP2 could bind probes with multiple cytosines methylated in a CH context very weakly by southwestern assay (Meehan *et al,* 1992). In comparison to a single mCpG pair, the methyl group on thymine (see Figure 1.2) when placed opposite a methylated cytosine in a CA context did not present an adequate substitute. The probes used in this study contained CA methylation in a CAG and CAT context. A subsequent study showed that the MBD could bind to a single mCpG paired with a mismatched TG with comparable affinity to a symmetrically methylated CpG pair by EMSA, indicating that the methyl group on the thymine may act as a suitable substitute for cytosine methylation in the right sequence context (Valinuck *et al,* 2004).

A recent study investigated the distribution of mCH in adult mouse hippocampus (Guo *et al*, 2014). The authors noticed using published MeCP2 ChIP-seq data from sorted neurons that CH was hypermethylated near MeCP2 binding sites. MeCP2 ChIP was repeated and the DNA pulled down was bisulfite sequenced at individual loci. Notably the MeCP2-bound DNA was enriched for mCH over input DNA in the absence of any mCpG. This association was supported *in vitro* using the MBD in EMSA's. The MBD was shown to associate with probes methylated both in a CT and CAC context with lower affinity than to mCpG probes. Both modifications present on the same probe enhanced binding (Guo *et al,* 2014).

mCH binding has been replicated in two separate laboratories using MeCP2 ChIP-seq combined with high-resolution methylation maps in mouse brain tissues (Gabel *et al,* 2015; Chen *et al,* 2015). *In vitro* EMSA work using the MBD and full-length protein refined the "H" and showed that MeCP2 has high affinity for mCA probes, but not mCC or mCT (Gabel *et al,* 2015). Binding of full-length protein to mCA by EMSA has been replicated (Chen *et al,* 2015) and a preference for C as the third base has been determined (J Connelly unpublished data). This fits neatly with the most abundant form of mCH being mCAC in the brain (Guo *et al,* 2014) and explains why earlier studies using CAT or CAG probes did not detect this high affinity binding. Furthermore, it implies some requirement for symmetry in MeCP2 binding sites as mCAC also presents a methyl group on the thymine of the opposite complementary strand.

Thus, MeCP2 can bind to and interpret the unique pattern of cytosine methylation in the neuronal methylome. Exclusion at regions of hmCpG may also be functionally relevant. This leads to the questions: where do you need MeCP2 and what does it do there? In order to answer these questions, mouse modelling has been extensively employed.

1.4 Mouse models of *MECP2*

Mouse models have been widely utilised in biomedical science. The mouse is similar to the human in terms of anatomy, physiology and genetics and harbours a genome that is readily manipulated. In addition, mice are small, have a short reproductive cycle and lifespan, making them an efficient organism in which to study the effects of disease (research.jax.org).

Male mice that are *Mecp2*-null recapitulate the phenotype of Rett syndrome. *Mecp2*-null male mice appear normal at birth. However, their growth is restricted and they developed a stiff, uncoordinated gait from three weeks of age. There is a progressive phenotype of reduced spontaneous movement, tremor, hind-limb clasping, irregular breathing and bruxism. This is followed by rapid weight loss and death at around eight weeks. Affected females, heterozygous for *Mecp2*, develop symptoms from

three months onwards but have a normal lifespan (Guy *et al,* 2001). As in patients, subtle abnormalities of behaviour and vocalisation are detectable during the grossly normal, pre-symptomatic phase (De Filippis *et al,* 2010). *Mecp2*-null mice have reduced brain weight and small, tightly packed neurons in the hippocampus, cortex and cerebellum (Chen *et al,* 2001). The size of these regions is reduced by approximately 10% (Belichenko *et al,* 2008) and there is evidence of reduced dendritic spine density in the cortex and hippocampus (Belichenko *et al,* 2009). The period of unremarkable development, progressive neurological phenotype and neuropathology are all reminiscent of Rett syndrome in humans. Mice have therefore been widely accepted as an excellent tool in which to study Rett syndrome.

1.4.1 Global, life-long requirement for MeCP2

Mouse models have been used to try and disentangle the requirement for MeCP2 in specific tissues, cell types or stages of development in order to learn more about protein function and the pathogenesis of Rett syndrome.

The importance of MeCP2 function in the brain was emphasised by the observation that the mouse Rett-like phenotype (RTT) was grossly indistinguishable when a floxed *Mecp2* gene was knocked out in neurons and glia using Cre recombinase (Cre) expressed under the Nestin promoter (*Nestin*-Cre) (Guy *et al,* 2001; Chen *et al,* 2001). A landmark finding came in Rett syndrome research when it was proven that re-expression of *Mecp2* on a null background (using tamoxifen-inducible Cre to excise a floxed STOP cassette upstream of the *Mecp2* gene) could effectively reverse RTT (Guy *et al,* 2007), including neuronal soma size, dendritic complexity and length (Robinson *et al,* 2012). This exciting research raised the question of whether therapeutic strategies in Rett syndrome might be curative. It also highlighted that the nervous system appeared to develop sufficiently for full functioning to be restored at a later time point. It is remarkable that the emergence of RTT in female *Mecp2* heterozygous mice occurs not at the same developmental time as human patients, but the same chronological time, as if neuronal function can continue in the face of MeCP2 deficiency for a limited period. Rett syndrome appears to be a problem of neuronal maintenance, rather than gross abnormality of neurodevelopment with

studies failing to detect abnormalities of neuronal differentiation or proliferation in the absence of *Mecp2* (Kishi et al, 2004). In support of this statement, deletion of *Mecp2* in adult males results in emergence of a RTT phenotype and death after ten weeks (McGraw *et al,* 2011). This also occurs if *Mecp2* is deleted at a juvenile stage (Nguyen *et al,* 2012), although the onset of symptoms is slightly delayed with very early inactivation, indicating that there may be critical periods for MeCP2 function: postnatally when RTT symptoms would develop and in adulthood when animals would fail to survive, therefore MeCP2 is required throughout the lifespan (Cheval *et al,* 2012). To complicate matters, the level of MeCP2 expression appears to be critical. Mouse models have recapitulated the MeCP2 duplication syndrome in patients, where MeCP2 expression at higher than endogenous levels also causes a neurodevelopmental phenotype and reduced lifespan (Collins *et al,* 2004; Luikenhuis *et al,* 2004; Samaco *et al,* 2012). Furthermore a 50% reduction in MeCP2 expression caused by floxing the gene, resulted in abnormality of respiration, motor function, memory and social interaction (Samaco *et al,* 2008; Kerr *et al,* 2008). The level of MeCP2 expression must be just so.

Numerous studies have suggested that neuronal dysfunction in Rett syndrome is centred on the synapse, although this is unlikely to be the only manifestation of neuropathology. Reduced spontaneous activity related to smaller miniature excitatory post-synaptic currents (EPSC's), was detected using whole cell patch clamp recording in cortical pyramidal neurons (Dani *et al,* 2005). Reduced magnitude of evoked EPSC's and reduced frequency of spontaneous EPSC's were observed in hippocampal glutamatergic neurons (Chao *et al,* 2007). This was thought to be the result of a reduction in the presynaptic pool of readily-releasable neurotransmitter and related to a reduced number of glutamatergic synapses in the pre-symptomatic period. A further study demonstrated unimpaired hippocampal neurogenesis, but a reduction in synapse number in the dentate gyrus, related to reduced dendritic spine density and apparent immaturity of post-mitotic neurons (Smrt *et al,* 2007). Abnormalities in synaptic inhibitory/excitatory balance have been replicated in cortex (Wood *et al,* 2010), hippocampus (Nelson *et al,* 2006) and induced pluripotent stem cell-derived neurons from mouse (Farra *et al,* 2012) and human (Marchetto *et*

al, 2010). Moreover, synaptic dysfunction is evident from impaired long-term potentiation (LTP) in *Mecp2* heterozygous females (Guy *et al,* 2007) and impaired LTP and long-term depression (LTD) in mice expressing a truncated form of MeCP2, which was compatible with learning and memory deficits uncovered in these mice (Moretti *et al,* 2006). Thus Rett syndrome is at least in part related to synaptopathy.

Although studies are united in reporting abnormal synapse formation and function in Rett models, it is difficult to reconcile the paradoxical reduction in spontaneous excitation and excitatory synapse number observed in cortex and hippocampus with the high incidence of seizures and apparent hyperexcitability of the Rett brain. One possibility is that although reduced excitatory activity is observed at individual synapses, in the complexity of the developing brain, the overall network effect is different – for example reduced excitation of inhibitory interneurons. Indeed, Zhang and colleagues demonstrated diminished basal inhibitory rhythmic activity in *Mecp2*-null hippocampal slices, with the authors concluding that this rendered the tissue prone to hyperexcitability due to heightened synchronisation of cells (Zhang *et al,* 2008). Furthermore, re-examination of electrical activity in cultured hippocampal neurons from *Mecp2*-null animals revealed an increase in evoked excitatory transmission following short trains of action potentials and an increase in presynaptic release probability (inferred from reduced paired pulse ratios), consistent with hyperexcitability of the cells (Nelson *et al,* 2011). A third explanation is that the synaptic effects in various brain regions and cell types are diverse or that MeCP2 function is crucial in just a few cell types or regions.

Although the phenotype of Rett syndrome does not intuitively map to one region of the brain, as the brain functions in a network the impairment of one "node" could cause widespread dysfunction. Mouse models have been used to try and disentangle the issue of whether Rett syndrome is a global brain disorder or a disorder of specific brain regions or neuronal subtypes. This is possible by using a Cre expression under the control of region or cell-specific promoters to remove a floxed *Mecp2* gene. In this fashion several tissue-specific knock-outs have been studied allowing the

breakdown of the RTT phenotype into component parts. *Mecp2* knock-out in forebrain using calcium/calmodulin-dependent kinase II (*CamKII)*-Cre resulted in impaired motor coordination and social behaviour with increased anxiety, however mice had a normal lifespan (Chen *et al*, 2001). Knock-out in the hypothalamus using single-minded homolog 1 (*Sim1)*-Cre resulted in aggression, hyperphagia, obesity, and altered response to stress (Fyffe *et al,* 2008; Wang *et al,* 2014). *Mecp2* knock-out in the basolateral amygdala using stereotactic injection of adeno-associated virus packaged with Cre, caused heightened anxiety and deficits in cue-dependent fear learning (Adachi *et al,* 2009). *Mecp2* knock-out in caudal pons, medulla and spinal cord with homeobox B1 *(HoxB1)*-Cre reduced lifespan to around 13 weeks and caused abnormal heart rate and respiratory response to hypoxia, underscoring the importance of hindbrain in RTT (Ward *et al,* 2014). Although contributory to the phenotype, these studies reveal that *Mecp2* knock-out in solitary brain regions does not recapitulate RTT, but simply results in dysfunction of that region. Different cell types have been interrogated for *Mecp2* dependence in a similar fashion.

Mice with *Mecp2* knocked out in dopaminergic and noradrenergic neurons using tyrosine hydroxylase (*TH)*-Cre were hypoactive and had reduced levels of TH, dopamine and noradrenaline. Knock-out in serotonergic neurons using PC12 ets factor 1 *(Pet1)*-Cre resulted in aggression, hyperactivity, reduced tyrosine hydroxylase 2 (TH2) and serotonin levels (Samaco *et al,* 2009). In contrast to these mild phenotypes, mice lacking *Mecp2* in GABA-ergic neurons due to vesicular inhibitory amino acid transporter (*Viaat)*-Cre, displayed most of the features of RTT with some subtle variations such as slightly longer lifespan (26 weeks), excessive grooming, forepaw stereotyped movements and increased sociability, making the major inhibitory neurotransmitter in the brain a key player in Rett syndrome pathology (Chao *et al,* 2010). However, deletion of *Mecp2* in GABAergic neurons may simply be a more direct way of causing the excitatory/inhibitory imbalance that *Mecp2* deletion in whole brain achieves and so mirrors the phenotype, rather than recapitulates it.

Co-culture experiments indicated a non-cell autonomous role for glia in the pathogenesis of Rett syndrome. Culturing wildtype hippocampal neurons with *Mecp2*-null astrocytes resulted in abnormalities of the neurons, which grew at lower density and had fewer and shorter dendrites (Ballas *et al,* 2009). This effect was replicated with only the conditioned medium from astrocytic culture, indicating that an unknown secreted factor was responsible. This result was not replicated in a subsequent study, but instead the authors suggested a role for microglia in the pathogenesis of Rett syndrome when they observed detrimental effects with culture of wildtype neurons in conditioned medium from *Mecp2*-null microglia. They identified the toxic secreted factor as glutamate, which was five times the normal secreted level (Maezawa & Jin, 2010). Knock-out of *Mecp2* in astrocytes using glial fibrillary acidic protein *(GFAP)*-Cre resulted in small mice with hindlimb clasping and irregular breathing but normal lifespan, mobility and hippocampal dendritic arborisation (Lioy *et al,* 2011). Knock-out exclusively in oligodendrocytes using neural-glial antigen 2 *(NG2)*-Cre resulted in a very mild phenotype of hindlimb clasping and hyperactivity (Nguyen *et al,* 2013). A role for glial cells in pathogenesis was convincing, but could not account for the entire RTT phenotype.

Therefore deletion of *Mecp2* in specific brain regions and cell types indicates that MeCP2 is required throughout the whole brain, however dysfunction of GABAergic neurons and glial cells seem particularly influential for expression of the phenotype. An important question was whether these conclusions would be supported by complimentary rescue experiments when MeCP2 is re-expressed in particular regions or cell types on a null background.

The reversibility of the syndrome had been shown using re-expression from the endogenous locus in all cells (Guy *et al,* 2007), furthermore expression in all cells using a randomly integrated P1-derived artificial chromosome (PAC) (Collins *et al,* 2004) or expression in post-mitotic neurons using *Mecp2* knocked in to the endogenous *Tau* locus prevented onset of symptoms in null mice (Luikenhuis *et al,* 2004). On the contrary, incomplete improvement of symptoms and lifespan was found by reactivating a rescue *Mecp2* transgene with a floxed STOP cassette in brain

during embryogenesis (*Nestin-Cre*), post-mitotic neurons during embryogenesis (*Tau-Cre*) or forebrain in the postnatal period (*CamKII-Cre*) (Giacometti 2007). Other region-specific rescues have been unsuccessful such as striatum and cerebellum (Alvarez-Saavedra *et al,* 2007) and caudal pons, medulla and spinal cord (Ward *et al,* 2014). However rescue of respiratory response to hypoxia and abnormal heart rate – two key causes of lethality in Rett syndrome – were demonstrated in the latter, identifying the hindbrain as a potential therapeutic target.

Re-expression of MeCP2 in specific cell types has not resulted in complete rescue either. Astrocyte re-expression only extended lifespan by around four months, although weight and mobility improved and respiration, hippocampal neuronal soma size, dendritic complexity and expression of a glutamatergic transporter were all normalised (Lioy *et al,* 2011). Re-expression in oligodendrocytes lengthened lifespan by just 3 weeks, but increased weight, improved hindlimb clasping and motor function (Nguyen *et al,* 2013). One promising study reported a near reversal of RTT in *Mecp2*-null mice that received bone marrow transplantation following whole body irradiation, which allowed recolonisation of the brain's microglia population with wildtype cells (Derecki *et al,* 2012). This study also reported a dramatic improvement in symptoms with re-expression in myeloid cells using lysozyme 2 *(LysM)*-Cre. Mice had normal survival, increased weight, improved mobility and fewer apnoeas. The mechanism was proposed to be dependent on microglial phagocytosis. However, these findings were not replicated using three different mouse lines in independent laboratories and it was shown that the *LysM*-Cre rescue resulted in MeCP2 expression in only a quarter of microglia, but also in some neurons. Repeating the rescue with a more specific microglial Cre did not improve RTT (Wang *et al,* 2015).

These disappointing rescue efforts may indicate that to rescue a phenotype that has already manifested, MeCP2 function must be restored in all cells. Prevention of onset by expression of MeCP2 exclusively in neuronal cells as demonstrated by Luikenhuis and colleagues may be achievable prior to the development of the mature nervous system when compensatory strategies are still possible. An alternative

explanation is that there was also MeCP2 expression in glial cells under the *Tau* promoter in this study and so a pure neuronal rescue may not be possible. In summary, MeCP2 is required in all cells of the brain to prevent features of Rett syndrome.

1.4.2 Mouse modelling & therapeutics

There have been significant and rewarding advances in the Rett field from the use of mouse models. One important limitation of this line of investigation is whether the findings in the mouse can really be extrapolated to the human. For example, how can the effects of cognitive impairment and absence of speech be modelled in this organism? Is this disorder truly recapitulated when male mice survive into adulthood and human males die in infancy? Furthermore, much of the research has focused on male mice when *Mecp2* heterozygous females more realistically model the disorder. These aspects are tested when potential therapies identified in the mouse progress to clinical trials.

There have been a number of possible therapeutic avenues generated from mouse model experiments (Pozzo-Miller *et al,* 2015). Perhaps the most promising is insulin-derived growth factor 1 (IGF1) treatment, which was attempted following the observation that there are increased levels of IGF-binding protein 3 in patients and *Mecp2*-null mice and the RTT phenotype improved with IGF-1 treatment (Tropea *et al,* 2009). A safety trial was completed (Pini *et al,* 2012) and phase II trials are underway.

In addition, the expression of BDNF (an autocrine factor that supports differentiation and plasticity of neurons and promotes their survival) is reduced in *Mecp2*-null mice and replacing this with ectopic expression from a transgene ameliorated RTT (Chang *et al,* 2006). A subsequent study showed that a small molecule ligand (7,8-dihydroxyflavone) of the BDNF receptor, tropomyosin receptor kinase B (TrKB) improved symptoms in *Mecp2*-null mice (Johnson *et al,* 2011) and there are currently two clinical trials assessing the efficacy of BDNF-related compounds in patients (Pozzo-miller *et al,* 2015).

An alternative way of treating Rett syndrome would be to replace wildtype MeCP2 using gene therapy where the gene can be packaged into a non-pathogenic virus that can infect the central nervous system and be delivered intravenously, intrathecally or intracranially. Extended survival and symptom improvement was demonstrated in *Mecp2*-null male mice that had gene therapy in the postnatal period by intracranial injection (Gadalla *et al,* 2012). Another group delivered *Mecp2* to heterozygous female mice via tail vein injection and observed a reduction in phenotypic score, absence of seizures and improvement on behavioural tests (Garb *et al,* 2013). There are currently no gene therapy trials in patients with Rett syndrome.

Hopefully at some point the field will be in a position to test the reversibility of Rett syndrome in a patient group.

1.5 The functions of MeCP2

The key to understanding the molecular basis of Rett syndrome is to elucidate the function of MeCP2 and how its binding properties and global expression requirement can be translated into meaningful biological consequences in the brain. There have been a number of proposed functions, which are discussed in the following sections but after more than twenty years of research the true function of MeCP2 has not been indisputably demonstrated.

1.5.1 Transcriptional repression

MeCP2 has traditionally been thought of as a transcriptional repressor, given the association with cytosine methylation and the ability of the 207-310 amino acid (aa) fragment to repress transcription of reporter constructs *in vitro* (Nan *et al,* 1997). Early work showed that rat MeCP2 could repress transcription of the α-globin gene when it was both methylated and unmethylated in histone free rat liver nuclear extracts (Meehan *et al,* 1992). Later experiments using transient transfection of affinity purified rat MeCP2 revealed that transcription of a reporter construct was selectively repressed in HeLa cell nuclear extracts when from a methylated promoter

(Nan *et al,* 1997). This transcriptional repression was partially dependent on deacetylation of histones and subsequently, MeCP2 was found to interact with Sin3a, which recruits histone deacetylases (HDACs), *in vivo* (Jones *et al,* 1998; Nan *et al,* 1998). In addition to Sin3a, MeCP2 has also been shown to bind to other co-repressors, namely cSki, the NCoR complex, co-repressor for element-1-silencing transcription factor (CoREST) and Brahma (Guy *et al,* 2011). Recently, it was demonstrated that *in vitro* transcriptional repression of a reporter construct is dependent on interaction with NCoR and HDAC activity, thus refining the TRD to the NID (Lyst *et al,* 2013). Histone hyperacetylation has been observed in the brain of *Mecp2* deficient mouse models (Shabazian *et al,* 2002; Skene *et al,* 2010) and treatment with a histone decetylase inhibitor, Trichostatin A, recreates synaptic defects seen in *Mecp2* deficient neurons (Nelson *et al,* 2006). By binding methylated DNA in chromatin and linking this signal to chromatin modifiers, MeCP2 can therefore function as an epigenetic effector in the organisation and interpretation of information contained within the genetic code, ultimately resulting in repression of gene transcription. This hypothesis has been hampered by the lack of a consistently detectable subset of genes that are upregulated in *Mecp2*-null cells and tissues. This is discussed in detail in the following section.

1.5.2 Gene expression in *Mecp2* models

The search for a subset of deregulated genes in *Mecp2*-null brain led to glucocorticoid response genes (Nuber *et al,* 2005) and mitochondrial respiratory chain genes (Kraukionis *et al,* 2006), however alterations in expression were not consistent across studies or disease stages. Imprinted genes would be candidates for MeCP2 regulation, given that they are allele-specifically methylated. Their deregulation in the context of MeCP2 deficiency has been controversial and inconsistent (Lawson-Yuen *et al,* 2007). These potential subsets did not seem to encapsulate the entire Rett phenotype and were not universally altered across studies.

Several region-specific gene expression studies comparing *Mecp2*-null brain tissue to wildtype have been conducted in the search for patterns of gene misregulation. An initial report using cortical and hippocampal tissue from both pre- and symptomatic

mice found no significant alterations in global gene expression, but that genes were subtly up- and downregulated in comparison to wildtype brain (Tudor *et al,* 2002). Several hundred genes were deregulated in cerebellum, however gene expression in littermates was more similar than in mice of the same genotype and genes were not consistently altered across different time points (Jordan *et al,* 2007). There were no significant differences in gene expression found between cortex, midbrain and cerebellum and no genes that were significantly altered in all three regions in symptomatic males (Urdinguio *et al,* 2008). Subtle alteration in gene expression was found in just over 100 genes in striatum in symptomatic mice and by combining gene expression data from previous studies, the authors noted that the misregulated genes in different brain regions had a more distinctive transcriptional signature than that of the different genotypes (Zhao *et al,* 2013).

A new hypothesis arose, when gene expression patterns were compared in 6 week *Mecp2*-null and *Mecp2* overexpresser mice (where MeCP2 was expressed at 2-3 times the endogenous level). It was observed that several hundred genes were altered commonly in cerebellum and hypothalamus with the majority of genes being downregulated in the *Mecp2*-null and up-regulated in the *Mecp2* overexpresser tissue, leading to speculation that MeCP2 was acting as a transcriptional activator (Charhour *et al,* 2008; Ben-Shachar *et al,* 2009). These findings were replicated in amygdala (Samaco *et al,* 2012). This is discussed further in 1.5.4.

Most regions of the brain are collages of different cell types, each potentially with their own gene expression signature. MeCP2 may have different target genes depending on the neuronal cell type and this may have resulted in the mixed up- and downregulation and lack of consistent targets seen with analysis of whole regions. An alternative approach to investigation has been using gene expression analyses in MeCP2 deficient cultured cells and clonal cells derived from patient material. Analysis of dentate gyrus granule cells in symptomatic mice revealed 13 differentially expressed genes (12 of which were upregulated) including genes involved in cytoskeleton formation and synaptogenesis (Smrt *et al,* 2007). It has been shown that cultured astrocytes from neonatal mice express higher levels of astroglial

markers, glial fibrillary acidic protein (GFAP) and S100β (Okabe *et al,* 2012). Li *et al* analysed an isogenic population of neuronal cells differentiated from MeCP2 deficient human ES cells *in vitro*. Their microarray data was normalised for the total number of input cells and so corrected for reduced total RNA in the MeCP2 deficient cells. With this correction, they found that ten times more genes were downregulated than upregulated in the absence of *MECP2*. They related the downregulated set to high levels of hmC content and high expression level and concluded that MeCP2 acted as a transcriptional activator in their cultured neurons. They did not see the same effects in neuronal precursors (Li *et al,* 2013). This reduction in total RNA in cultured cells has been reported previously in neurons differentiated from *Mecp2*-null ES cells (Yazdani *et al,* 2012), but not been replicated in animal models and raises the question of whether these were primary effects of MeCP2 deficiency or secondary to the culture of "sick" cells.

Although initial efforts to examine gene expression patterns in clonal cell lines derived from patient material yielded inconsistent results (Traynor *et al,* 2002; Delgado *et al,* 2006) or revealed only a handful of significant expression changes in cell types not known to be affected in Rett syndrome (Nectoux *et al,* 2010), studies using induced pluripotent stem cells (iPSC) derived from Rett syndrome patient material have also suggested that *Mecp2* deficiency results in the alteration of transcription of many genes. Using RNA-seq, Tanaka and colleagues were able to distinguish the global gene expression pattern in induced pluripotent stem cells (iPSC's) with mutant *MECP2* from iPSC's with wildtype *MECP2* derived from female patient fibroblasts and found thousands of both up- and downregulated genes in the mutant cells. The authors suggested that there were different patterns of gene expression in iPSC's derived from patients carrying different *MECP2* mutations (Tanaka *et al,* 2014). One important limitation of this study is that the gene expression was analysed in stem cells, which have a different methylome and MeCP2 expression level to mature neuronal tissue. However the use of matched mutant and wildtype material derived from one patient provides a unique opportunity to control for the variable genetic background in different people.

An alternative view is that the lack of clearly distinguishable and consistent patterns of gene deregulation in these studies is not due to cell type-specific effects, but because genes are only regulated by MeCP2 in response to neuronal activity. Neurons are dependent on efficient interpretation of transient stimuli in order to store network information in the long term and so deregulation at this level could lead to neuropsychiatric disability. There is evidence to suggest that MeCP2 constitutively binds to the *BDNF* promoter region in cultured neurons, but with calcium influx coincident with neuronal activity, MeCP2 is phosphorylated at serine 421 and is released from the promoter, relieving the repression on *BDNF* transcription with a subsequent increase in histone acetylation (Chen *et al,* 2003; Martinowich *et al,* 2003). This precise mechanism may underlie the exclusive importance of MeCP2 in the nervous system and explain the emergence of the Rett phenotype at the time of intense synaptogenesis. Abnormalities of dendritic complexity in cultured hippocampal neurons overexpressing *Mecp2* were found to be dependent on phosphorylation of serine 421 (Zhou *et al,* 2006). A serine 421 phosphomutant mouse showed abnormalities of dendritic arborisation and abnormal responses to novel experiences however there were no differences detectable in MeCP2 ChIP and gene expression profiles in the resting state or in response to depolarisation (Cohen *et al,* 2011). The mechanism may depend on phosphorylation of serine 424 in concert, as a double phosphomutant mouse had enhanced long-term potentiation, context-dependent fear conditioning and spatial memory with an increased number of excitatory synapses in the cortex and hippocampus. MeCP2 was more tightly bound to target gene promoters (Li *et al,* 2011). Additionally, a small number of genes were deregulated in a serine 80 phosphomutant mouse that had abnormal locomotor activity and reduced association of MeCP2 with DNA, which is in keeping with serine 80 becoming dephosphorylated on neuronal stimulation (Tao *et al,* 2009). Due to these mild phenotypes and effects on MeCP2 binding and gene expression, it seems unlikely that phosphorylation of MeCP2 is the principal determinant of MeCP2 transcriptional regulation.

A fresh perspective on gene expression was offered when it was observed that genes of greater than 100kb in length were disproportionately represented in neuronal

specific processes and mutated in autistic spectrum disorders (King *et al,* 2013). A recent study analysed gene expression profiles using microarrays in four distinct cell types (motor cortical pyramidal cells, motor cortical interneurons, noradrenergic cells from the locus coeruleus and cerebellar purkinje cells) derived from symptomatic *Mecp2*-null mice and wildtype littermates (Sugino *et al,* 2014). Across all cell types there were over 800 misregulated genes: a small majority (56.2%) were upregulated in the absence of MeCP2 and only two genes were misregulated reproducibly in all four cell types underlining the importance of analysing individual cell types when searching for subsets of regulated genes. The authors observed that genes related to cell adhesion processes (which are long as a class) were over-represented in the group of altered genes. Misregulated genes were plotted according to length and a specific upregulation of genes greater than 100kb in length was noted in the *Mecp2*-null cells. This length dependent relationship did not hold for genes that were downregulated in the absence of MeCP2, indicating that MeCP2 was a repressor of long gene transcription (Sugino *et al,* 2014).

This finding was reproduced by another group, which analysed multiple previous studies with new data derived from combinations of microarrays, qPCR, RNA-seq and nCounter analysis in *Mecp2*-null mice, ES cell-derived neurons, $Mecp2^{R306C/y}$ brain and cortical tissue from a patient with Rett syndrome (Gabel *et al,* 2015). The authors found that the magnitude of long gene upregulation correlated with disease progression and severity. Using MeCP2 ChIP-seq data combined with published high resolution methylation maps, the authors proposed the mechanism of MeCP2 long gene regulation involved binding to mCA, as they observed that only long genes with high mCA content were upregulated in the absence of MeCP2. There was no effect of mCA content on the expression of short genes. Furthermore, reducing mCA by excising a floxed *Dnmt3a* gene in mouse brain using *Nestin-Cre* recapitulated the upregulation of long genes seen with a lack of functional MeCP2.

Support for a mCH mechanistic theory came from a third recent study that combined high coverage GFP ChIP-seq in hypothalamus of an enhanced green fluorescent protein (EGFP) tagged *Mecp2* transgenic mouse line with published methylation

maps and RNA-seq analysis of gene expression in *Mecp2*-null and overexpresser mouse hypothalamus (Chen *et al,* 2015). The authors found that over one thousand reciprocally misregulated genes in both the *Mecp2*-null and overexpresser mice had increased levels of MeCP2 binding and higher mCH, but not higher mCpG content in the gene body in comparison to genes that were unchanged. Genes that were uniquely misregulated in the overexpresser model had intermediate mCH content, perhaps indicating that extra MeCP2 had spilled over to lower affinity sites. Interestingly, MeCP2 appeared to bind to high mCH genes that were downregulated in the *Mecp2*-null hypothalamus, indicating that there may be separate gene-specific mechanisms for regulation by mCH (Chen *et al,* 2015).

To sum up, the only consistent subset of genes upregulated in the absence of MeCP2 are long. The gene length effect in Rett models is small, but global. This exciting new theory of MeCP2 function needs further investigation and leads to the question of whether manipulation of long gene expression in *Mecp2*-null mice would be a viable therapeutic strategy.

1.5.3 Regulation of transcriptional noise

Given (1) the abundance of MeCP2 in neurons, (2) the abundance of methylated cytosines, (3) the global binding pattern of MeCP2, (4) the increase in repetitive element RNA in *Mecp2*-null brain nuclei (Skene *et al,* 2010) and (5) the increase in L1 long interspersed nuclear element (LINE1) retrotransposition in *Mecp2*-null mouse brain and induced pluripotent stem cells derived from a Rett syndrome patient's fibroblasts (Muotri *et al,* 2010), one possibility is that MeCP2 acts to reduce transcriptional noise, preventing aberrant transcription on a genome-wide scale. Therefore in the absence of MeCP2, transcription of unessential repetitive material is prioritised with variable impact on individual genes. One line of evidence against this hypothesis is that no increase in variance of gene expression (which would be predicted with increased stochastic gene expression) was found in striatum, cerebellum or hypothalamus of *Mecp2*-null animals in comparison to wildtype (Zhao *et al,* 2013). The authors of this study suggested that MeCP2 was performing a

coordinated role in regulation of transcription, which is compatible with the long gene hypothesis described in the previous section.

1.5.4 Activation of transcription

It has been proposed that MeCP2 acts as a transcriptional activator, given that the majority of genes are downregulated in its absence (see 1.5.2) and the finding of reduced RNA synthesis in *Mecp2* deficient neurons in culture (Yazdani *et al,* 2012; Li *et al,* 2013). A plausible mechanism for this regulation would be recruitment of transcription factor, cyclic AMP-responsive element-binding protein 1 (CREB1), which has been reported (Charhour *et al,* 2008) and/or association with hmCpG, which has been linked to actively transcribed genes (Mellen *et al,* 2012). However, it is now known that MeCP2 does not bind to hmCpG *in vivo* (see 1.3.4) and by current estimates the amount of hmCA is too low to be biologically significant. Reduced transcription has not been replicated in mouse models and may be a secondary effect of unhealthy cells in culture. Another possibility is that in the absence of MeCP2, aberrant transcription of material that is not accounted for by microarray analysis and is filtered out in RNA-seq experiments, such as repetitive elements, results in a homeostatic downregulation of transcribed genes. Therefore the various transcriptional hypotheses are not mutually exclusive. It may be that MeCP2 exerts different effects depending on mode of interaction with DNA, cell type or developmental stage.

1.5.5 Regulation of chromatin

Another proposed function of MeCP2 that is related to transcriptional repression is the regulation of chromatin structure. It has been shown that the amount of histone 1 (H1) is doubled in the *Mecp2*-null brain (Skene *et al,* 2010) and given that injection of H1 accelerated the fluorescence recovery after photobleaching (FRAP) kinetics of MeCP2 in mouse fibroblasts, the two proteins may compete for binding sites *in vivo* (Ghosh *et al,* 2010). The ability of MeCP2 to assemble reconstituted nucleosomal arrays has been demonstrated *in vitro* (Georgel *et al,* 2003) and at the cellular level MeCP2 induced the formation of large chromocentre clusters when transfected into myoblasts (Brero *et al,* 2005). Chromatin looping was lost at an imprinted locus in

Mecp2-null brain indicating that the conformation of the chromatin had altered and previously juxtaposed sites had become separate (Horike *et al,* 2005). Together these data present convincing evidence that MeCP2 may be able to induce the formation of higher order chromatin structures, perhaps by displacing H1. This could potentially be the mechanism by which MeCP2 regulates gene expression.

1.5.6 Post-transcriptional regulation

Post-transcriptional roles have been proposed for MeCP2. An association with a Y-box transcription factor, YB1 that regulates alternative splicing has been demonstrated. In the same study, ectopic expression of MeCP2 in HeLa cells altered the splicing pattern of a co-transfected reporter minigene (Young *et al,* 2005). More recently, MeCP2 has been implicated in micro-RNA (miRNA) processing. When phosphorylated at serine 80, MeCP2 associated with DiGeorge syndrome critical region 8 (DGCR8), a component of the miRNA processing machinery. On neuronal firing, MeCP2 was dephosphorylated, dissociated from the complex and miRNA processing proceeded (Cheng *et al,* 2014). This suggests that there would be an upregulation of miRNA in the absence of MeCP2, however microarray analysis of whole brain from post-symptomatic *Mecp2*-null animals revealed that 70% of annotated miRNA's were downregulated (Urdinguio *et al,* 2010). A second study using high throughput sequencing of small RNAs contradicted this finding, reporting that the majority of miRNAs were upregulated in *Mecp2*-null cerebellum (Wu *et al,* 2010). The differing results may be the result of using a specific brain region, a different method of expression analysis or slightly younger mice in the second study.

1.5.7 A multifunctional protein

MeCP2 has a plethora of proposed interacting partners and consequently could be related to a number of cellular processes and functions (Lyst & Bird, 2015). Circular dichroism (CD) studies have indicated that MeCP2 is 60-65% unstructured and therefore intrinsically disordered (Ghosh *et al,* 2008). This conformational plasticity has been related to multifunctionality and the ability to bind multiple different protein partners (Hansen *et al,* 2010). It is possible that the function of MeCP2 is highly dependent on interaction with other protein complexes and the abundance of

these will vary between cell types. This leads to a complex hypothesis where the function of MeCP2 would be entirely cell type and state dependent, which would be difficult to disentangle.

1.6 Mutations in *MECP2* that cause Rett syndrome

Study of Rett-causing mutations is a way of highlighting key regions of MeCP2 that are necessary for function and in turn understanding the molecular basis of the syndrome. Mutations in *MECP2* have been identified in over 95% of cases of classical Rett syndrome (Moretti & Zoghbi, 2006) and in around 75% of atypical cases (Cuddapah *et al,* 2014). They almost always arise *de novo*, which means that Rett syndrome is very unlikely to run in families, especially as affected females usually do not reproduce. In fact recurrence risk with a negative family history is less than 0.1% (Smeets *et al,* 2011). The majority of mutations are of paternal origin, which explains the rarity in males (Trappe *et al,* 2001). Most are likely the result of the high mutability of mC, which can undergo spontaneous deamination to thymine (see Figure 1.2). The X chromosome is highly methylated in the male germ cells, but not in oocytes, hence the preponderance of paternal transmission (Wan *et al,* 1999).

More than 800 different mutations are described on the RettSyndrome.org Variation Database – a valuable resource of collated mutation information contributed by researchers and clinicians. Despite this vast number, eight common point mutations resulting from C to T transition account for approximately half. In descending order of frequency, these are p.Thr158Met (T158M, 8.8%), p.Arg168X (R168X, 7.3%), p.Arg255X (R225X, 6.4%), p.Arg270X (R270X, 5.8%), p.Arg294X (R294X, 4.9%), p.Arg306Cys (R306C, 4.8%), p.Arg133Cys (R133C, 4.24%) and p.Arg106Trp (R106W, 2.8%) (Christodoulou *et al,* 2003). Another common mutation class is carboxy-terminal deletions, accounting for approximately 10% of cases.

Around half of cases caused by common mutations occur in the MBD of *MECP2* and this region is completely conserved between human and mouse. Furthermore, if all causative missense mutations (where one amino acid is substituted for another) are plotted along the length of the protein, it can be seen that the mutations cluster over

the MBD and the NID (Figure 1.6). Plotting all polymorphisms recorded in unaffected controls (derived from the Exome Aggregation Consortium) excludes these domains. This indicates that the integrity of these regions is crucial and led to the hypothesis that a functional MBD and NID were all that was required to prevent Rett syndrome (Lyst & Bird, 2015). There are, however three known missense mutations that lie outwith these domains (A2V, P225R and P322L) and the sizeable group of C-terminal deletions, that require further investigation and are not accounted for by this new hypothesis.

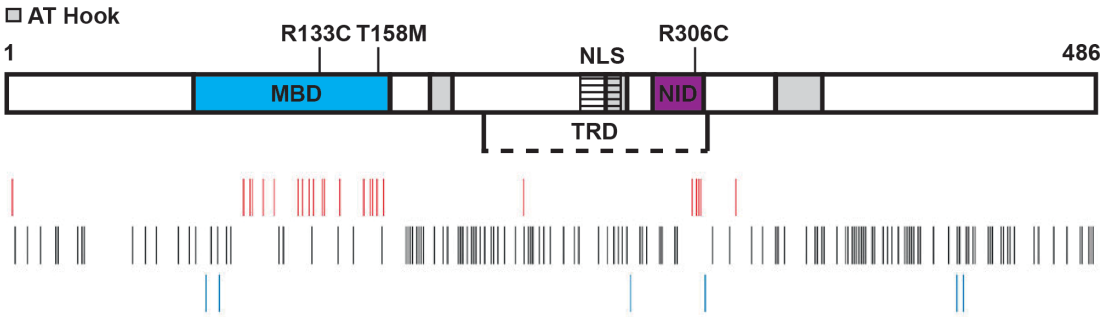


Figure 1.6 Rett syndrome mutation analysis of MeCP2

Schematic representation of MeCP2 protein, isoform e2 (486 amino acids) with major functional domains and putative AT hooks indicated. The locations of the Rett missense mutations are plotted in red and underline the importance of the MBD and NID. The locations of the R133C, T158M and R306C mutations pertinent to this study have been highlighted. Neutral amino acid variants from the Exome Aggregation Consortium are plotted in black and tend to lie outwith the MBD and the NID. Phosphorylation sites are shown in blue. Adapted from Lyst & Bird, 2015.

1.6.1 Mutations in *MECP2* with other consequences

As well as resulting in Rett syndrome, mutations in *MECP2* have been associated with X-linked mental retardation (Meloni *et al,* 2000; Orrico *et al,* 2000; Couvert *et al,* 2001; Yntema *et al,* 2002), non-progressive encephalopathy in males (Imessaoudene *et al,* 2001), Angelman syndrome-like phenotype (Watson *et al,* 2001), psychosis, pyramidal signs and macro-orchidism (PPM-X) (Klauck *et al,* 2002), early onset schizophrenia (Cohen *et al,* 2002) and infantile autism (Carney *et al,* 2003). In addition, there is a neurodevelopmental syndrome caused by duplication

of the *MECP2* gene (Peters *et al,* 2013). Therefore the relevance of research into MeCP2 is potentially widespread and will have significant impact further afield.

1.7 Study of Rett-causing mutations

The complete phenotype of Rett syndrome cannot be attributed to single regions or cell types and the function of MeCP2 is still unclear, although promising research implicates the regulation of long genes. It is now important to study patient groups and mouse models with commonly occurring mutations in *MECP2* in order to learn more about protein function and disease pathogenesis, to yield prognostic information and better inform recruitment to clinical trials There has already been some progress in this field, which is discussed in the following sections.

1.7.1 Mouse models of Rett-causing mutations

A mouse model of the rarely occurring T158A mutation revealed that instability of the protein was a major contributory factor to the RTT phenotype with mutation of this residue (Goffin *et al,* 2011). Using a mouse model, it was confirmed that MeCP2 with the R306C mutation failed to associate with the NCoR complex *in vivo*, which resulted in definition of the NID of the protein (Lyst *et al,* 2013). Modelling a mutation affecting a translation start site in exon 1 of *Mecp2* implicated this more abundant isoform in the RTT phenotype (Yasui *et al,* 2014), however this is likely the result of reduced protein abundance rather than isoform-specific effects, as rescue of an *Mecp2*-null phenotype using a transgene expressing the e2 isoform has been demonstrated (Collins *et al,* 2004).

It seems unlikely that disrupting regulatory phosphorylation of MeCP2 causes Rett syndrome as mouse models of MeCP2 phospho-mutants display very mild phenotypes (Tao *et al,* 2009; Cohen *et al,* 2011; Ebert *et al,* 2013) and there are no known missense mutations of phosphorylation sites. However, their disruption will be additive. It is possible that phosphorylation sites interact and their functioning is impaired by other Rett mutations, for example phosphorylation of threonine 308 is abolished in mice with the R306C mutation and was shown to be important for dissociation from NCoR in response to neuronal activity (Ebert *et al,* 2013).

Furthermore, phosphorylation sites will be disrupted in early truncation and some C-terminal deletion mutations. Another possibility is that the sort of phenotype caused by phosphomutants is not possible to measure in the mouse but has significant impact on the functioning of a human.

An observation made in patients indicated that a frameshift occurring at G273 results in a milder neuropsychiatric phenotype than frameshift at R270. Modelling these two mutations in transgenic mice recapitulated this severity with average survival of approximately 29 weeks and 12 weeks, respectively. Investigation revealed the importance of a putative AT hook domain in MeCP2 that was disrupted in R270X protein and resulted in subtle deficits in DNA binding, chromatin association, alpha thalassemia/mental retardation syndrome X-linked protein (ATRX) co-localisation and nucleosomal array assembly (Baker *et al,* 2013). It may also be possible that the earlier truncation disrupts the NLS (amino acids 255-271) however the protein appeared nuclear. One further caveat to this study is that there are no known Rett-causing missense mutations that disrupt the AT hook.

The same group later proposed that an intact MBD and TRD are required to produce toxic effects of MeCP2 overexpression, as transgenes with R111G or R306C did not produce a phenotype on a wild type background (Heckman *et al*, 2014). However, there was no wild type overexpresser presented in parallel for comparison. Interestingly, the authors noted that the phenotype of the R306C transgene on the *Mecp2*-null background was more severe than their previous G273X model where the C-terminal portion of the protein is lost, with the R306C mice only surviving for 18 weeks. They proposed a dominant negative effect of this mutation and attributed it to a deficit in DNA binding with the R306C mutation (shown by ChIP and EMSA) due to disruption of a “basic patch” of amino acids, presumably used to associate non-specifically with DNA. These results require further investigation, as some of the phenotypic differences may be attributable to variable protein expression from the transgene.

Thus, mouse models of Rett-causing mutations have helped to underline the importance of different domains of MeCP2 in Rett pathogenesis. The advantage to mouse modelling is that male mice with *Mecp2* mutations survive into adulthood and the effects of mutant protein can be studied without the complicating factor of skewed XCI. This is discussed in the following section.

1.7.2 X Chromosome Inactivation (XCI)

There is significant clinical heterogeneity in Rett syndrome, despite the classical course of development. Refining this heterogeneity by analysing subgroups of patients can aid in understanding the pathogenesis and this task has been addressed by dividing patients into categories based on their particular mutation in *MECP2*. Before moving on to review the study of Rett-causing *MECP2* mutations in humans, it is important to note another potential source of clinical variability.

As *MECP2* is located at Xq28, it is subject to XCI in females. This is the process by which one X is inactivated at random in every cell to compensate for the presence of two copies (Lyon, 1961). Usually the XCI pattern is considered skewed if more than 75-80% of cells express one allele and is typically calculated by analysing the methylation pattern of the androgen receptor in leukocytes. The pattern of XCI has been reported as random and non-random in Rett patients (Amir *et al,* 2000). It is potentially influential in determining phenotypic severity because it regulates the load of mutant protein in the brain. Convincing evidence of this phenomenon comes from rare examples of unaffected mothers with highly skewed XCI transmitting the mutant copy of the *MECP2* gene to their offspring, who then manifest the disease (Schanen *et al,* 1997; Masuyama *et al,* 2005) or cases of monozygotic twins with markedly different phenotypic severity and corresponding variance in XCI (Ishii *et al,* 2001). Additionally, studies focusing on particular mutations have statistically related skewing to phenotypic severity (Archer *et al,* 2007). Two caveats to the interpretation of this measure are (1) that skewing in XCI has been shown to increase with age, so may not correlate with retrospectively reported clinical features (Sharp *et al,* 2000) and (2) the pattern in leukocytes may not reflect that in the brain, furthermore inactivation in the brain may be region-specific, so it may be impossible

to know the true pattern of XCI in an individual with Rett syndrome in life (Gibson *et al,* 2005).

1.7.3 Clinical studies of genotype-phenotype correlation

There has been a wealth of work linking particular types of *MECP2* mutation to clinical severity in Rett syndrome. This is reviewed in the following section.

An early study found that patients with truncation mutations (n=30) were more likely to have lower homovanillic acid and patients with missense mutations (n=18) were more likely to have scoliosis (Amir *et al,* 2000). There was no difference in clinical severity between the two groups when 13 variables were considered. Patients with skewed XCI were excluded. Cheadle *et al* performed this analysis and detected significantly reduced clinical severity in the missense group (n=20) compared to the truncation group (n=24) when hand use, speech and walking were considered. Additionally, subdividing the truncation group revealed lower clinical severity in late truncations (n=16) compared to truncations upstream of the TRD (n=10). XCI was not considered (Cheadle *et al,* 2000).

There was a flurry of interest in genotype-phenotype correlation over the next four years, although investigating groups could not come to a consensus regarding which set of mutations predicted increased severity. Monros *et al* reported that truncation mutations were associated with greater clinical severity than missense mutations using an early version of the Pineda scale (Monrós *et al,* 2001). In contrast, Hoffbur *et al* reported that cases with MBD missense mutations or early truncations were clinically more severe than cases with mutations occurring within or downstream of the TRD when assessing individuals with random XCI using five clinical features. The authors suggested that the pattern of XCI *could* affect clinical outcome (Hoffbur *et al,* 2001). Huppke *et al* contradicted these findings and reported lower severity in patients with missense mutations and C-terminal deletions than truncations using six clinical features (Huppke *et al,* 2002). Smeets *et al* found no correlation between skewing of XCI and milder phenotype. They corroborated that early truncation mutations resulted in a more severe clinical phenotype than missense mutation cases

using the Kerr scale and a simplified six-point version (Smeets *et al,* 2003). Schanen *et al* stated that patients with missense mutations had lower overall severity scores and better language performance than those with nonsense mutations, however particular MBD and TRD mutations could not be separated by clinical severity in their study (Schanen *et al,* 2004). Three further studies did not find any influence of mutation type or XCI on clinical severity (Auranen *et al,* 2001; Nielsen *et al,* 2001; Weaving *et al,* 2003). These conflicting findings from many different countries were likely related to the small sample sizes (n≤123), meaning that diverse mutations were being grouped together on the basis of similar location within *MECP2*. Later, using alternative approaches to genotype-phenotype correlation and with the growth of patient databases and international collaboration, the frequency of common mutations was appreciated and it was possible to further refine mutation groups. A reproducible spectrum of phenotypic severity emerged.

1.7.4 The R133C mutation *in vivo*

Yamashita *et al* took a different approach to the question of phenotype-genotype correlation by focusing on cases with a specific clinical feature: preserved speech (Yamashita *et al,* 2001). At the time, there was diagnostic uncertainty over these children who met criteria for autistic spectrum disorder and Rett syndrome. The authors analysed *MECP2* mutations in 30 Japanese patients, five of whom had preserved speech. XCI was not assessed. They discovered that all of the five preserved speech cases had mutation in the *MECP2* gene and two of the five had the R133C mutation. None of the patients without preserved speech had the R133C mutation. The two patients with R133C appeared to have the mildest clinical presentations. For example, the first of these patients developed normally for three years prior to regression, she had normal growth, was able to walk and in-line skate, but was clumsy. She could still eat with a spoon and regained social interest at five years of age. The second patient had preserved ambulation. The authors speculated that the R133C mutation might result in a milder form of Rett syndrome.

With a similar approach, Leonard and colleagues asked whether there was evidence of a milder Rett syndrome phenotype in patients with the R133C mutation (Leonard

et al, 2003). This group analysed clinical data for 24 international patients with the R133C mutation and compared them to 98 Rett syndrome patients with other mutations from an Australian Rett syndrome database. The comparison mutations were distributed throughout the length of *MECP2* and different categories of mutation were equally represented. XCI patterns were assessed in 85 cases, however patients with skewing were not excluded. Clinical severity was assessed using the Kerr, Percy, Pineda and WeeFIM scales. They found that patients with the R133C mutation were significantly less likely to be in the severe category for voluntary hand use, speech, gross motor function and ambulation, age at losing social interaction, sleep disturbance, respiratory dysfunction, awake breathing rhythm, onset of and frequency of hand stereotypies, or scoliosis. This reduction in severity over a wide spectrum of clinical parameters translated into significantly lower composite severity scores in the R133C group than the comparison group according to all four clinical scales. The authors concluded that there was evidence of a milder clinical syndrome with the R133C mutation.

In a slightly larger follow up study with 147 cases, the same group distinguished two truncation mutations within a similar region of *MECP2* that would have been grouped together in previous studies. The authors found that R270X was significantly more severe than R294X on all of their four severity scales. They also noted that the T158M cases were conspicuously mid-range in severity, could not be distinguished from other mutations, and that the R306C cases were at the milder end of the spectrum with significantly later onset of regression (Colvin *et al,* 2004).

A UK study of 135 cases confirmed that patients with early truncating mutations had a more severe phenotype than missense or late truncating mutations using a dimensional approach to assessment that took into account typicality, severity (using a five-point scale) and clinical course. Analysis of individual mutations revealed that patients with the R133C mutation (n=9) had a significantly lower severity score than P255X and T158M patients (Charman *et al,* 2005).

The reduced severity of patients with the R133C mutation was corroborated by an American study, which considered 245 Rett syndrome patients for phenotype-genotype correlation (Neul *et al,* 2008). The authors excluded patients with atypical Rett syndrome from this study, which may have led to disproportionate exclusion of those with the R133C mutation and preserved speech. There were 12 patients with the R133C mutation and the other mutations were classified as one of the other most common point mutations, C-terminal truncations, large deletions, other or no *MECP2* mutation. Clinical severity was assessed using 13 traits. There was a significant difference in the overall clinical severity scores between the groups, with the R133C mutation group achieving the lowest severity score of all the groups studied (significantly lower than R168X and large deletions by pairwise analysis). This observation stood up when the R133C group was compared to the R168X group and participants with skewed XCI had been removed. Under these circumstances, the R168X group was also significantly more severely affected than R306C patients. Ambulation, hand use and language distinguished the mutation groups. R133C patients were significantly more likely to have preserved hand use and language and there was a trend towards lower severity for ambulation when compared to the R168X group (50% of the R133C group used words, 92% used hands and 75% could walk alone). This study also highlighted reduced severity in the C-terminal deletion category and reported the interesting observation that only a small proportion of patients with the R306C mutation could use words.

Bebbington *et al* studied 272 Rett syndrome patients, using InterRett, the international Rett syndrome database (Bebbington *et al,* 2008). Patients were assessed using the Kerr, Percy and Pineda scales and grouped by mutation: the eight most common point mutations, along with C-terminal deletions. The study did not take the pattern of XCI into account but excluded Australian cases from the Leonard *et al* study, to allow valid comparison. In addition, the Kerr scores were age-corrected to prevent any skewing towards increased severity for the older patients. There were 27 patients with the R133C mutation in the study. The R133C patients had the lowest severity score on all three scales with statistically significant comparisons to R168X, R255X and T158M on every scale. They were more likely to

have a delayed onset of regression and hand stereotypies and around two thirds had preserved hand function, compared to less than one third over the whole cohort. They were less likely to have severe feeding difficulties. All R133C cases learnt to walk and 21% of this group were able to use phrases in speech, compared to 2% of the rest of the group. Bebbington *et al* did not find any significant differences between the mutation groups in any other features indicating that R133C Rett syndrome is still a severe and debilitating condition. They were able to segregate R133C, R294X, R306C and C-terminal deletions as a generally milder group, R106W, R168X and T158M as an intermediate group, with R270X and R255X conferring the most severe phenotype. A spectrum of severity was beginning to form. In addition, the authors were able to demonstrate a good correlation of clinical severity scores for individual mutations with the earlier Australian Study (Leonard *et al,* 2003) and recommended routine genetic testing for individuals where a diagnosis of Rett syndrome was being considered.

The same group reported 832 Rett syndrome cases from the Australian Rett syndrome database and the InterRett database with a focus on C-terminal deletions (Bebbington *et al,* 2010). There were 56 patients with the R133C mutation included in their study. By direct comparison with the C-terminal deletion group (79 patients) the authors showed that the R133C cases had significantly lower severity scores on the Pineda and Percy scales, were more likely to learn and be able to walk, and were more likely to currently have speech. Although not the focus of this study, this data lends further weight to the conclusion that the R133C mutation causes the mildest form of Rett syndrome.

A Dutch group analysed 137 Rett syndrome patients from their Maastricht-Leuven Rett syndrome database using the Kerr scale (Halbach *et al,* 2012). The participants included girls with typical and atypical Rett syndrome. There was data on XCI pattern for 22 patients, but they were not excluded. Mutations were classified according to position in the *MECP2* gene: MBD, TRD, C-terminal and other. If a mutation occurred in more than five patients, it was analysed separately. There were 10 patients with the R133C mutation and this group achieved the lowest overall score

for clinical severity when compared to all other recurrent mutations (including C-terminal deletions). There was a statistically significant difference in the proportion of R133C patients who had normal development in the first year of life (80%) compared to patients with the R168X mutation (13%). The proportion of R133C patients with oro-motor difficulties (30%) was significantly lower than patients with the P152R mutation (60%). The authors noted that patients with the R133C mutation had consistently low scores in all domains, however they advised caution with emphasis on prognostic power, given the clinical heterogeneity observed in patients with the same mutation.

Urbanowicz *et al* looked specifically at language abilities in 766 Rett syndrome patients that were 15 years or younger at the time of enrolment to the Australian Rett syndrome database and the InterRett database (Urbanowicz *et al,* 2014). There were 55 patients with the R133C mutation in this study. Using enrolment questionnaire speech information, the authors categorised speech ability into no speech, use of babble or use of words. Mutation categories were the eight most common point mutations, early truncation, large deletion, C-terminal deletions or other. The R133C group were more likely to be able to say words before their regression than patients with large deletions and the most likely to say words following their regression of all the mutation groups. This study called into question the utility of distinguishing a preserved speech variant of Rett syndrome from classical cases, given that 185 of the 766 patients in the study did not experience a regression in their speech and language, which is a requirement for the diagnosis of classical Rett syndrome.

Finally, Cuddapah *et al* analysed a database of nearly 1000 American Rett syndrome patients for genotype-phenotype correlation, including 51 patients with the R133C mutation (Cuddapah *et al,* 2014). They divided patients into typical (n=815) and atypical (n=148) Rett syndrome (including 40 and 11 R133C cases, respectively) and used data gathered by experienced clinicians over several clinic visits. In this way, they were able to observe progression of disease over time. The study used 16 mutation groups: in addition to the eight most common point mutations, they had 3'truncations, deletions, exon 1 mutations, insertions, large deletions, other point

mutations, splice sites and no mutation. XCI was not recorded. They assessed clinical severity using thirteen separate traits. In the typical Rett category, the R133C patients had the lowest average severity score, except for the exon 1 mutation group. The score was significantly lower than twelve of the other mutation categories. This relationship persisted when individual clinical abilities were examined and the mutation was ranked similarly for the atypical Rett cases. The authors noted that the difference in phenotypic severity between the least severe mutations (exon 1 mutations, R133C, R306C, R294X, 3' truncations) and the most severe mutations (R168X, insertions, large deletions, R270X) widened in the atypical Rett syndrome analysis, which perhaps was an indication that these should not be considered separately and may have produced more robust differences in phenotypic severity if grouped together. Overall the authors concluded that (1) severity worsens with age, (2) hand use and age of onset of stereotypies predict disease severity for both groups, (3) ambulation and age of independent sitting predict severity in the typical cases and (4) language predicts severity in the atypical group (though this is not surprising, given that this includes the preserved speech variant of Rett). The interpretation of the R133C patient data was certainly influenced by over a quarter of patients being allocated to the atypical category. In this large study, R255X, R106W and T158M mutations were associated with Rett syndrome of intermediate severity.

The general consensus of these genotype-phenotype studies that the R133C mutation results in a milder form of Rett syndrome is supported by the existence of a male patient with the mutation. A case report of a Japanese boy with classical Rett syndrome, who was found to have the R133C mutation, was published (Masuyama *et al,* 2005). The patient had a history of developmental delay, autistic features, stereotypic hand movements, seizures, abnormal respiration in wakefulness, loss of purposeful hand use, growth retardation and was bed bound by the age of eight, due to motor impairment. The patient's sister also had Rett syndrome caused by the R133C mutation: their mother was a carrier with skewed XCI and mild intellectual disability. The authors were able to rule out somatic mosaicism in the boy. Additionally, a case of a male patient with classical Rett syndrome and the R133H mutation has also been reported (Monrós *et al,* 2001). Normally, males with *MECP2*

mutations that cause typical Rett in females and a normal karyotype die in the neonatal period with severe encephalopathy (Kankirawatana *et al,* 2006). Therefore the survival of a male into childhood underlines the reduced severity conferred by the R133C mutation.

There is wide phenotypic variability between individuals with the same mutation and although the R133C mutation is associated with preserved speech, not all patients with R133C have preserved speech. Equally, there don't seem to be abilities that are always lost in individuals with the R133C mutation, which would point towards a specific deficit in MeCP2 function conferred by this mutation. A recent study looked at 114 Italian patients, including 7 patients with the R133C mutation, using the Rett Assessment Rating Scale (RARS) (Fabio *et al,* 2014). This novel instrument assessed thirty clinical symptoms that were rated from 0-7 with the aim of achieving finer grain resolution of clinical profiles. The authors observed that patients with the R133C mutation deviated from a "mild" scoring profile most in the cognitive domain, which included visual and auditory perception, attention, spatial orientation, temporal orientation, memory, eye contact, gestures and, rather confusingly, verbal skills. Although this was a small study, it will be interesting to see if it's possible to elucidate any specific deficits conferred by the mutation using more detailed clinical analysis in the future. This would yield further clues towards understanding the molecular basis of MeCP2 dysfunction in these patients.

There are a number of important limitations to these clinical studies. Firstly, XCI status is rarely taken into account and it has been shown that skewing (at least from blood profiles) seems to affect phenotypic expression in some cases. Secondly, some of the data analysed is derived from voluntary and retrospective parent reporting, which is subject to ascertainment bias and recall bias. Thirdly, there was disparity in which clinical severity scale was used to rate patients between the studies and so different aspects of Rett syndrome were captured. Finally, separating the R133C participants into typical and atypical Rett syndrome in the largest study may have hampered the ability to detect subtle variation in phenotype caused by this mutation by reducing statistical power. Nevertheless, clinical studies from different countries

involving patients of different ages, using different means of assessment come to the same convincing conclusion that the R133C mutation is associated with a less severe form of Rett syndrome in patients.

1.8 The R133C mutation *in vitro*

The reduced clinical severity observed in patients with the R133C mutation is clear. This is surprising, as R133 is one of two arginine fingers proposed to confer binding specificity to MeCP2 (see 1.3.2). In the following sections, *in vitro* studies of $MeCP2^{R133C}$ are reviewed.

1.8.1 Binding to mCpG

DNA binding of *Xenopus laevis* MeCP2 was assessed by EMSA, using 42 base pair duplex oligomers with either one or twelve mCpG pairs. Binding of $MeCP2^{R133C}$ to the multiply methylated probe was virtually abolished, to the extent that saturation of the probe could not be achieved with increasing amounts of protein. By Southwestern assay with the same probes, $MeCP2^{R133C}$ binding to the singly and multiply methylated duplexes was 20% and 10% of wildtype protein binding, respectively (Ballestar *et al,* 2000). The same group revisited the Southwestern experiment with human MeCP2. They found that $MeCP2^{R133C}$ binding to the multiply methylated duplex was abolished (Yusufzai *et al,* 2000). Binding of an MBD fragment of human MeCP2 to a 28 base pair DNA duplex with a single mCpG pair was measured by EMSA. The wildtype MBD shifted 99% of the probe at the maximum peptide concentration assayed. MBD^{R133C} shifted 18% of the probe at an equivalent peptide concentration, which was in agreement with Ballestar *et al* (Free *et al,* 2001). However in an identical experiment (except the duplex was 17 base pairs) rat MBD^{R133C} binding to a single mCpG dyad was abolished by EMSA (Galavao & Thomas 2005). The focus of this study was to investigate MBD binding to unmethylated Holliday junction DNA, which may mimic MeCP2 binding to nucleosomes in order to compact chromatin. The authors observed that MBD^{R133C} had just a 2-fold reduction in binding the four-way junction DNA.

EMSA's were revisited in another study using human MeCP2 and a 208 base pair probe that was methylated at twelve CpG sites. At equivalent protein concentrations, there was 25-30% more unshifted probe with $MeCP2^{R133C}$ compared to wildtype, indicating that binding was not abolished. This experiment was repeated with twelve ligated 208bp probes, presenting 144 CpG sites. They found no impairment in the ability of $MeCP2^{R133C}$ to shift the unmethylated tandem, however there was only a minor enhancement of the shift when this large, complex probe was methylated. Wildtype protein showed a substantial enhancement of binding when the probe was methylated. The authors' interpretation was that $MeCP2^{R133C}$ was binding by methylation-independent mechanisms (Nikitina *et al,* 2007). This same experiment using MBD^{R133C} produced a very weak methylation-dependent enhancement of the shift, suggesting that the smaller segment also bound with low affinity in the presence of the R133C mutation (Ghosh *et al*, 2008).

Although MeCP2 is well conserved, these results collectively highlight the importance of species, protein fragment, and target probe in measuring DNA binding *in vitro*. To sum up, the R133C mutation significantly impairs specific binding to DNA probes methylated in a CpG context, but largely retained non-specific binding to unmethylated DNA.

1.8.2 Structural impact

Interestingly, despite the significant reduction in ability to bind methylated DNA, a minor impact of the R133C mutation on the secondary structure of the MBD was observed using circular dichroism (CD) (Ballestar *et al,* 2000) or by NMR spectroscopy with no evidence of unfolding or instability (Free *et al,* 2001). This issue was revisited using a combination of fluorescence spectroscopy and CD (Ghosh *et al,* 2008). By fluorescence spectroscopy, $MeCP2^{R133C}$ was partially unfolded, resulting in increased exposure of Trp104 to solvent and a reduction of emitted fluorescence. The Tm (temperature at which 50% of protein was unfolded) was lower for $MeCP2^{R133C}$ than wild type protein ($38^{o}C$ vs $44.5^{o}C$, respectively). In addition, when wildtype MeCP2 bound to a 45 base pair segment of DNA from the *BDNF* promoter IV region, which was methylated at one CpG pair, its Tm increased

by 18.6^{o}C, indicating an increase in stability induced by DNA binding. In contrast, the Tm of MeCP2^{R133C} only increased by 10.7^{o}C when bound to the same oligomer, suggesting lower affinity binding. Interestingly, the fluorescence emission spectra for wild type MBD and MBD^{R133C} were identical, but with a small reduction in thermal stability (the Tm of MBD^{R133C} was 3.4^{o}C lower). This was explained using CD, as the authors were able to distinguish the secondary structure of wildtype MBD and full-length protein. The CD spectra for R133C mutants were indistinguishable from wildtype equivalents, indicating minimal change in secondary structure and confirming earlier reports (Ballestar *et al,* 2000). When exposed to methylated DNA, however, there was no alteration in the CD spectrum for MBD^{R133C}, as there was for the wild type MBD with a small increase in alpha helicity, indicating that the conformation of the mutant protein was not altered or stabilised by interaction with methylated DNA. These results suggest that (1) the secondary structure of the full-length wildtype protein is different to that of the MBD fragment, (2) thermal instability of the R133C mutant is detectable using full-length protein, and (3) structural differences attributable to the R133C mutant are subtle, becoming apparent on binding to methylated DNA.

Recently, Kucukkal *et al* modelled the DNA binding deficit of MBD^{R133C} in silico using the previously determined X-ray structure (Ho et al *et al,* 2008) as a starting point (Kucukkal *et al,* 2015). They found that MBD^{R133C} still bound to methylated DNA, but its interaction was significantly reduced. This was attributable to a reduction in hydrogen bonding with methylated DNA (both direct bonding via arginine 133, but also an indirect reduction of hydrogen bonding by arginine 111) and the loss of two salt bridges in the vicinity of the arginine fingers. A predicted model of mutant binding is shown in Figure 1.8.2. The consequence of the mutation on secondary structure was proposed to be a reduction in alpha helicity of the bound protein (as found by Ghosh and colleagues) indicating a lack of specificity for methylated DNA.

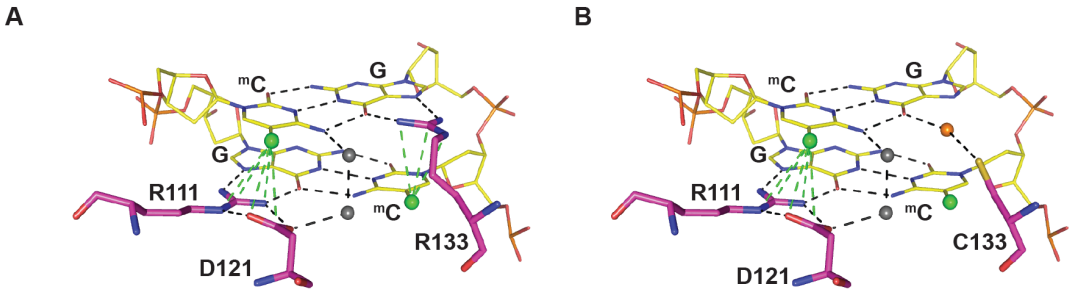


Figure 1.8.2 The structural impact of the R133C mutation

(A) Predicted model of MBD binding to a methylated CpG pair. Arginine 111 and 133 make contact with the guanines and form hydrogen bonds with the methylated cytosines of the CpG pair. Arginine 111 forms a hydrogen bond with aspartic acid 121.
(B) Predicted model of MBD^{R133C} binding to a methylated CpG pair. The cysteine 133 interaction with guanine is now dependent on water molecules and there is no bond with the methylated cytosine.
Hydrogen bonds are indicated by dotted lines, methyl groups by green balls & water molecules by grey & orange balls. Adapted from Brown *et al*, unpublished.

1.8.3 Binding to chromatin

In vitro assessment indicated that the R133C mutation had little impact on the ability of MeCP2 to compact unmethylated nucleosomal arrays (Georgel *et al,* 2003) or methylated nucleosomal arrays: $MeCP2^{R133C}$ protected nucleosomes from micrococcal nuclease digestion, indicating a preserved ability to bind and compact chromatin (Nikitina *et al,* 2007).

Chromatin association of R133C mutants has also been studied in cells. Kudo and colleagues transfected mouse fibroblast cells with EGFP-tagged MeCP2 and observed protein localisation to heterochromatic foci 48 hours later in fixed cells. Surprisingly, despite the previous literature indicating that the R133C mutation abolished specificity for methylated DNA, they observed a focal staining pattern for $MeCP2^{R133CEGFP}$ that was indistinguishable from wildtype (Kudo *et al,* 2003).

Baubec and colleagues studied the binding of $MeCP2^{R133C}$ in ES cells using ChIP-seq and published methylation maps (Baubec *et al,* 2013). The biotin-tagged MeCP2 constructs were expressed from a known site in the mouse ES cell genome (in addition to endogenous MeCP2 expression) and could be immunoprecipitated using

streptavidin. The authors revealed that wildtype MeCP2 tracked mCpG density but this specificity was lost with MeCP2^{R133C}. Mutant protein enrichment over wildtype protein enrichment was plotted against methylation density, so although it appears that MeCP2^{R133C} had no specificity for methylated DNA, any small increase in enrichment may have been masked by a larger fold change in enrichment for the wildtype protein. In addition, the MeCP2^{R133C} expressing ES cells had a normal pattern of localisation to heterochromatic foci by microscopy. An interpretation that consolidates the *in vitro* and structural R133C data would be that the mutant's apparently normal pattern of localisation to heterochromatin may rely on preserved non-specific binding.

The relative mobility of mutant protein on heterochromatic foci has been studied in cells by FRAP using murine fibroblasts transfected with EGFP-tagged MeCP2^{R133C}. The residence half time over heterochromatin was reduced for MeCP2^{R133C} at 12 seconds compared to wild type protein (21 seconds), indicating reduced affinity for chromatin (Kumar *et al,* 2008). These figures were in agreement with a similar FRAP study using EGFP-tagged MeCP2 transfected into murine *Mecp2*-null fibroblasts (Schmiedeberg *et al*, 2009). In the second study the residence half time was approximately 22 seconds for wild type protein and approximately 15 seconds for MeCP2^{R133C}. The advantage of the second study was that there was no competing endogenous MeCP2. Therefore by FRAP, there is a modest reduction in the affinity of EGFP-tagged MeCP2^{R133C} for heterochromatin that would be compatible with the apparently normal pattern of localisation to foci seen in fixed cells.

Agarwal and colleagues measured the ability to cluster chromatin as well as the association with heterochromatic foci using murine myoblasts transfected with EGFP-tagged MeCP2 constructs (Agarwal *et al,* 2011). They found a subtle but significant reduction in the relative amount of MeCP2^{R133CEGFP} accumulated in the heterochromatic dots versus the nuclear background (3.6 vs 5 for wildtype). Then taking the mean number of foci as a proxy measure of heterochromatin clustering, they found that the clustering ability of MeCP2^{R133CEGFP} was modestly impaired, with

a median number of 13 chromocentres per cell, compared to 11 with wild type MeCP2.

Taken together, these results indicate that there are subtle reductions in MeCP2^{R133C} localisation to and condensation of chromatin in some cellular assays that are not detected *in vitro*.

1.8.4 Binding to hmC

Mellen *et al* presented compelling data that human MeCP2(1-205)^{R133C} binding to a multiply hydroxymethylated 120bp oligonucleotide probe from the *BDNF* promoter IV region was impaired by EMSA (Mellen *et al,* 2012). The authors made the case that this was a specific DNA binding impairment because an equivalent assay using a multiply methylated *BDNF* probe showed only a modest DNA binding impairment. They supported this data with surface plasmon resonance assays where biotinylated probes that were either unmethylated, methylated or hydroxymethylated were immobilised on parallel flow cells of a streptavidin-coated sensor chip. Exposure to full-length recombinant MeCP2 altered the steady state response of the flow cells. Binding curves were generated from the read-out following exposure to serial dilutions of recombinant protein. MeCP2^{R133C} appeared to bind the hydroxymethylated probes non-specifically and had a maximum binding capacity of just 25% that of wild type protein. In contrast the maximum binding capacity for the methylated probes was 76% of the wild type value, which was not a statistically significant reduction. This differential result was at odds with the rest of the literature, which suggested a strong deficit in binding to methylated DNA. It was an intriguing finding: if MeCP2^{R133C} lost only one of the binding capacities of MeCP2, then that could be the explanation for the resulting milder clinical picture of R133C Rett syndrome.

1.8.5 Transcriptional repression

The ability of MeCP2 to repress transcription has been measured using luciferase reporter constructs in *Drosophila* cells. The authors used two methylated promoters – one with a high level of transcriptional activity and one with a low level. They found

that MeCP2^{R133C} could repress transcription with wildtype activity from both promoters (Kudo *et al*, 2003).

1.9 Aims of this thesis

The R133C mutation occurs in a critical region of MeCP2 that confers binding specificity yet results in a milder form of Rett syndrome in patients. The aim of this thesis is to create suitable models in which to study MeCP2^{R133C} function to understand this paradoxical observation. The proposed models are (1) cultured neuronal cells and (2) a mouse. In this way, MeCP2 ^{R133C} function can be studied in cells and tissue relevant to Rett pathogenesis. Initially, it will be important to establish whether the R133C mutation results in a milder Rett-like phenotype in the male mouse in the absence of genetic heterogeneity and skewed XCI. The molecular defects of MeCP2^{R133C} can then be investigated with particular attention to potential DNA binding deficits in order to resolve some of the inconsistencies reported in the literature. If DNA binding deficits are detected, the aim is to refine these as far as possible in terms of methylation status and sequence context using *in vitro* and *in vivo* techniques with the intention of understanding the relative importance of the different binding capacities of MeCP2. The overall aim is to determine the molecular basis of R133C Rett syndrome, thus advancing understanding of Rett pathogenesis and MeCP2 function.

2. Materials and Methods

2.1 Cell culture

2.1.1 Standard ES cell culture

ES cells were grown at 37^{o}C in 5% CO_2 in sterile, tissue culture gelatinised plastic ware (Greiner). Standard ES cell medium comprised Glasgow minimum essential medium (Invitrogen) supplemented with 15% (v/v) fetal bovine serum (Hyclone), 1mM sodium pyruvate, 50 μM β-mercaptoethanol, 2mM L-glutamine, non-essential amino acids (all Gibco) and recombinant human leukaemia inhibitory factor (a gift from J.Guy). Cells were passaged when around 80% confluent using Trypsin-EDTA (Gibco) in PBS. Sterile Dulbecco's phosphate-buffered saline (PBS, Gibco) was used for cell washes. Cell number was estimated using a Scepter 2.0 Cell Counter (Millipore).

2.1.2 ES cell targeting

Wildtype ES cells were targeted with *Mecp2* carrying the R133C mutation in order to establish a novel mouse line and differentiate mutant neuronal cells. A suitable targeting vector was kindly supplied by J.Guy (Lyst *et al*, 2013). It contained *Mecp2* exons 3 and 4 with the coding sequence of enhanced green fluorescent protein (EGFP) fused in frame to the end of *Mecp2* exon 4 with a downstream neomycin resistance (NEO) cassette. Point mutation, R133C was introduced by C.Merusi using the QuikChangeII XL Site-Directed Mutagenesis Kit (Agilent Technologies). The vector was linearised at the 5' end with SnabI (NEB). Gene targeting was performed on E14 TG2a ES cells, derived from the mouse 129/Ola substrain. ES cells (2 x 10^7) were transfected with 30μg linearised targeting vector in 600μl of HEPES buffered saline by electroporation (240V, 500μF, BioRad Gene Pulser), left for ten minutes on ice, then plated at 5×10^6 or $2.5x10^6$ cells per 10cm dish on day zero. Cells were selected by culturing in standard ES medium supplemented with 350μg/ml G-418 (Roche) from day one then 150 μg/ml G-418 from day three. Colonies were picked on day nine into a 96 well plate. On reaching confluence, cells were split into two 96 well plates. One plate was frozen down in standard culture medium adjusted to 20%

(v/v) FBS, 5% (v/v) dimethyl sulfoxide (Sigma) and the other was used for screening.

2.1.3 NEO cassette deletion

A pCAGGS vector that expressed Cre recombinase (Cre) under the control of the CAG promoter was used for transfection. The expression vector (20μg) was resuspended in 800μl of HEPES buffered saline and $1x10^7$ ES cells were transfected by electroporation (220V, 950μF, Biorad Gene Pulser). Cells were left for ten minutes at room temperature then resuspended in 18ml of standard ES medium and split equally between two 10cm dishes. Cells were fed the following day then split at low density to 1000 cells per 10cm dish in order to establish colonies derived from single ES subclones. After seven days in standard ES medium colonies were suitable for picking into a 96 well plate. They were immediately disaggregated and split into "master" and "selection" plates. The master plate was cultured in standard ES medium. The selection plate was grown on the following day in standard ES medium supplemented with 350μg/ml G-418 in order to reveal ES cell subclones that were now sensitive to the antibiotic and therefore had successfully had their floxed NEO cassette deleted by Cre. Such clones were expanded from the corresponding well of the "master" plate and used for neuronal differentiation.

2.1.4 Neuronal differentiation

ES cells were trypsinised and 4 x 10^6 cells were plated in 10cm bacterial dishes (Greiner) in embryoid body (EB) medium (standard ES cell medium with 10% (v/v) fetal bovine serum and no LIF). Medium was changed every two days. Four days after plating, new EB medium was supplemented with 5μM retinoic acid to encourage neuronal differentiation. Eight days after plating the EB's contained neuronal precursor cells. They were washed in PBS and disaggregated with fresh trypsin solution (0.05% (w/v) trypsin (Sigma), 1mM EDTA in PBS). The reaction was quenched with 10ml of EB medium and cells were pelleted for five minutes at 100 x g. The neuronal precursor cells were resuspended in N2 medium (advanced Dulbecco's modified Eagle medium, nutrient mixture F12, N2 supplement, 100units/ml penicillin and 100μg/ml streptomycin, (all Gibco)) and passed through a

40μm cell strainer (BD Falcon). Cells were counted, pelleted for five minutes at 100 x g then resuspended in an appropriate volume of N2 medium to be plated at 1.5 x 10^5 cells/cm^2. This was designated day zero. Culture dishes were washed with sterile water (Corning) and pre-treated with 0.1mg/ml poly-DL-ornithine (Sigma) in 30mM boric acid pH 8.3 48 hours before plating neuronal precursors. The day before plating, dishes were washed three times in sterile water and treated with 2μg/ml laminin (Roche) in PBS. This was washed briefly with PBS prior to plating. On day one, medium was replaced with 1:1 N2:Neurobasal medium (Neurobasal medium, B27 supplement, 100units/ml penicillin and 100μg/ml streptomycin, Gibco). On day three, half the medium was removed and replaced with Neurobasal medium. Neurons were harvested on day six.

2.2 The *R133C-GFP* mouse

2.2.1 Generation of the *R133C-GFP* mouse line

Correctly targeted ES cell clones with the NEO cassette still present were injected by J.Selfridge into blastocysts from C57BL/6J females 3.5 days after natural mating. Injections were performed in M2 medium (Sigma) with 10–15 ES cells being injected into each blastocyst before 6-12 were transferred to pseudopregnant recipient C57BL/6J females. Chimaeric pups were identified by agouti coat colour. On maturity, chimaeras were mated with C57BL/6J mice expressing Cre under the cytomegalovirus promoter to establish germline transmission and simultaneously delete the NEO cassette. Mice were then back-crossed for five generations onto a C57BL/6J genetic background (with the aid of C57BL/6J single nucleotide polymorphism analysis after four back-crosses). Unless otherwise stated, all mice used for biochemical analysis were 6 weeks of age.

2.2.2 Mouse phenotypic scoring & growth

Mice were weighed and scored for RTT on a weekly basis at the same time and in the same room. Scoring of symptoms was performed on the bench worktop and was done blind to the genotype of the mice. Six parameters were examined (activity, gait, hind-limb clasping, tremor, breathing and general condition) and given a score

between 0 and 2 (Guy *et al*, 2007; See Appendix 1). Given the slower and subtler development of the phenotype compared with *Mecp2*-null male mice, two intermediate scores (0.5 and 1.5) were utilised (Cheval *et al*, 2012). Animals scoring 2 for tremor, breathing or general condition, or those having reached the severity limit of the experiment according to the Home Office License, were automatically culled.

2.2.3 Behavioural analysis

R133C-GFP male mice were subjected to a battery of four behavioural tests at 9-10 weeks and 20-21 weeks. Female behavioural analysis was done at 1 year. Mice were assigned an animal unit number at birth so that they could be identified. Behavioural testing followed the chronological order of these numbers and so was random in relation to genotype. Firstly, mice underwent the elevated plus maze in order to assess anxiety level. Animals were placed individually in a cross-shaped maze 65cm above the floor with two open arms (20 x 8cm), two closed arms (20 x 8 x 25cm) and a central area (8 x 8cm). The experimenter left the room during trials, which were 15 minutes long and conducted in uniform dim lighting. Mice were tracked using ANY-maze software (Stoelting). The total time spent in open and closed arms along with the number of arm entries was recorded. If mice jumped from the maze during the trial, the trial was stopped and restarted a maximum of two times. If a mouse jumped from the apparatus three times, the data was discounted.

Secondly, in order to assess motor strength and coordination, mice underwent the hanging wire test. Animals were suspended from a 1.5mm diameter wire, 35cm above the bench worktop. The latency to bring one hindpaw up to grasp the wire was timed. An average time was taken from three trials, with an inter-trial interval of 15 minutes. Trials lasted for a maximum of 30 seconds.

Thirdly, mice underwent the open field test to assess anxiety and locomotor activity. Animals were placed individually in the centre of a 50x50cm square arena in uniform dim lighting. The experimenter left the room for the trial of 20 minutes. Mice were

tracked using ANY-maze software. The total time spent in the centre, total distance travelled, mean speed and time spent immobile were recorded.

Finally, to assess motor coordination and learning, mice underwent the accelerating rotatrod. Mice were trained to run on a 3cm diameter rod rotating at 4 revolutions per minute (RPM) on day one. If they fell, they were replaced on the rod over a 30 second period. Following this, there were three experimental days. Each day, mice completed four 5-minute trials, separated by an hour. During the trials, the rod accelerated between 4RPM and 40RPM and latency to fall was recorded.

2.3 DNA manipulation

2.3.1 Genomic DNA extraction

For screening of 96 well plates, ES cells were incubated in lysis buffer (100mM Tris HCl pH 8.5, 5mM EDTA, 0.2% (w/v) SDS, 200mM NaCl, 0.1mg/ml proteinase K) for 24 hours at 37^{o}C. Following incubation, DNA was precipitated with 1 volume of isopropanol and retrieved by centrifugation. Pellets were washed in 70% (v/v) ethanol before being dissolved in 50µl TE (10mM Tris HCl, 1mM EDTA; pH 8).

Otherwise, ES cells were pelleted by centrifugation, washed with PBS and incubated in lysis buffer (10mM Tris HCl pH 8, 400mM NaCl, 3mM EDTA, 1% (w/v) SDS, 0.4mg/ml proteinase K) overnight at 37^{o}C. The following day, 100µg/ml RNAseA was added for one hour at 37^{o}C. Samples were briefly centrifuged and the supernatant was treated with phenol/chloroform/isoamyl alcohol (1:1). Following centrifugation, DNA was precipitated from the aqueous phase using 1 volume of isopropanol, pelleted by centrifugation and washed with 70% (v/v) ethanol, before being air-dried and dissolved in TE.

Frozen mouse half brains were homogenised in phosphate buffered saline (140mM NaCl, 3mM KCl, 2mM KH_2PO_4, 10mM Na_2HPO_4, (PBS)) then genomic DNA was prepared using the Gentra Puregene Tissue Kit (Qiagen) according to manufacturer's instructions. DNA was cleaned up by phenol/chloroform/isoamyl alcohol (1:1)

extraction. Following centrifugation, DNA was precipitated from the aqueous phase by adding 1/10th volume of 3M NaOAc pH 5.2 and 2.5 volumes of ice-cold ethanol, pelleted by centrifugation and washed in 70% (v/v) ethanol before being air-dried and dissolved in TE.

2.3.2 Measurement of DNA concentration

DNA concentration was measured at OD_{260nm} and OD_{280nm} on a Nanodrop-1000 spectrophotometer (Thermo Scientific) and an automated concentration reading was calculated using Beer's law: concentration (ng/μl) = (absorbance OD_{260nm} x extinction coefficient dsDNA 50ng/μl/cm)/pathlength cm. Purity was determined by the OD_{260nm}:OD_{280nm} ratio. Values ≥1.8 indicated a lack of protein or phenol contaminants.

2.3.3 Restriction digestion

DNA was digested by restriction endonucleases according to manufacturer's instructions (NEB). Typically, DNA was digested using 6 units of enzyme per μg DNA with corresponding dilution buffer and 100μg/ml bovine serum albumin (if appropriate) in 20μl reactions for 1-2 hours at 37^{o}C.

2.3.4 DNA electrophoresis

DNA was loaded onto agarose gels using Orange G buffer (0.033% (w/v) orange G, 2% (w/v) Ficoll, 20mM EDTA pH 8, 0.7% (w/v) SDS) and resolved by electrophoresis at 120 volts (Sub-cell system, Bio-Rad). Agarose gel concentration 0.8-2% (w/v) and appropriate ladders in supplied loading buffer were selected (NEB/Bioline) depending on the size of DNA fragments to be resolved. Gels were prepared with Tris-acetate EDTA (40mM Tris-acetate, 1mM EDTA (TAE)) and 0.5μg/ml ethidium bromide as an intercalating agent, so that DNA could later be visualised using UV light.

2.3.5 Standard polymerase chain reaction (PCR)

PCR was primarily used to screen ES cell colonies firstly for homologous recombination of the *R133C-GFP* targeting vector and secondly to check for absence

of the NEO cassette following NEO deletion; to amplify genomic DNA prior to Sanger sequencing; in bisulfite sequencing firstly to amplify selected genomic regions of interest from mouse brain bisulfite-treated genomic DNA then secondly to screen bacterial colonies for presence of the bisulfite-treated PCR product following transformation; and also to test qPCR primers.

For homologous recombination screening, 1µl of genomic DNA was added to 0.625 units of Go Taq Polymerase with 1x Green Go Taq buffer (Promega), 200µM deoxynucleotide triphosphates (Thermo Scientific) and 0.5µM primers in 25µl on ice. Cycling conditions comprised an initial 5-minute denaturation at 95^{o}C, then 38 cycles of 1-minute denaturation at 95^{o}C, 1-minute annealing at 61^{o}C and 2 minutes 10 seconds of extension at 72^{o}C. There was a final 5-minute extension at 72^{o}C.

For NEO cassette screening, approximately 100ng of genomic DNA was added to 3mM $MgCl_2$, 0.5µm primers, 200µM deoxynucleotide triphosphates, 1 x reaction buffer IV and 2.5 units of Red Hot Taq (Thermo Scientific) in 50µl reactions. Cycling conditions were a 5-minute denaturation at 94^{o}C followed by 35 cycles of a 30-second denaturation at 94^{o}C, 45-second annealing at 64^{o}C and 1 minute 30-second extension at 72^{o}C. There was a final elongation step of 5 minutes at 72^{o}C.

Bacterial colony PCR, qPCR primer screening and amplification of bisulfite-treated genomic DNA all employed 1x DreamTaq Green PCR mastermix (Thermo Scientific) and 0.25µM primers. PCR conditions were individualised for specific primer pair melting temperatures and size of desired PCR product. A temperature gradient was employed to determine optimal annealing temperatures for new primer pairs (G-storm thermal cycler). Cycle number was 40 for general purposes except bacterial colony PCR, which was 25 cycles. Denaturation conditions were 1 minute at 95^{o}C initially and during cycles. There was also a final extension step at 72^{o}C for 5 minutes.

PCR amplification of bisulfite-treated chromatin immunoprecipitated DNA utilised JumpStart REDTaq polymerase (Sigma Aldrich), which has a high tolerance for

uracil in the sequence to be amplified. The 50μl reactions contained less than 50ng of DNA, 1x JumpStart REDTaq buffer, 2.5 units of JumpStartREDTaq, 0.25μM primers and 200μM deoxynucleotide triphosphates. Cycling conditions were 1 minute of denaturation at 94°C, then 40 cycles of 30-second denaturation at 94°C, 30-second annealing at 60°C and 1-minute extension at 72°C, with a final 5-minute 72°C elongation step.

PCR products were usually analysed by agarose gel electrophoresis.

2.3.6 Real-time quantitative PCR (qPCR)

Real-time qPCR of cDNA was used to determine relative levels of mRNA transcripts in neurons differentiated in culture and male mouse whole brain or cerebellar tissue. SYBR green technology was used where the amount of amplified DNA is determined from the integration of the green fluorescent dye during the reaction. 25ng (neurons) or 50ng (mouse brain or cerebellum) of template cDNA was combined with 0.25μM primers and 1x SensiMix SYBR & Fluorescein Master Mix (Bioline) in a 12.5μl reaction volume. Cycling reactions on a Roche Lightcycler were as follows: 5 minutes denaturation at 95°C then 45 cycles of 15-second denaturation at 95°C, 15-second annealing at 60°C and 20-second extension at 72°C (with a single acquisition at 522nm taken at the end of the extension phase). Melting curve analysis followed where the temperature was increased from 60°C to 95°C at a rate of 0.11°C/second with continuous acquisitions. A single melting curve indicated the presence of one PCR product and absence of non-specific products that would interfere with quantification. Cycle threshold (Ct) values were determined for each sample by Roche lightcycler software using the second derivative maximum method. A standard curve was calculated for all primer pairs in every qPCR run using dilution of input material. An average Ct value for each dilution was plotted against a log transformation of an arbitrary number corresponding to the dilution factor. The slope of the standard curve was used to calculate the efficiency of the reaction (a slope between -3.6 and -3.3 indicated 90-100% efficiency). The standard curve equation: $y = ax + b$ was used to transform the Ct values for samples by calculating a "real amount" = (Ct value – b)/a. This log of this real amount was then calculated, before

taking an average value from replicate wells for that sample (duplicate or triplicate). The Ct value was normalised to a housekeeping or control gene. The standard deviation was calculated from: √([standard deviation/average for gene of interest]^2 + [standard deviation/average for housekeeper]^2).

Real time qPCR was also used to determine the relative abundance of genomic regions of interest in input DNA and chromatin immunoprecipitated DNA in neurons differentiated in culture and male mouse whole brain tissue. Under these circumstances, 2.5µl of template and 0.5µM primers were combined with 1x SensiMix SYBR & Fluorescein Master Mix (Bioline) in a 12.5µl reaction volume. Cycling reactions were as follows: 10-minute denaturation at 95^{o}C, then 45 cycles of 1-minute denaturation at 95^{o}C, 15-second annealing at 60 or 62^{o}C (depending on the primers) and a 20-second extension at 72^{o}C. Melting curves and Ct values were calculated as above.

2.3.7 Southern blot

Southern blot was employed as a second screen for homologous recombination of the *R133C-GFP* targeting vector and also to check that the NEO cassette had been deleted from the targeted locus following exposure to Cre. For screening of 96 well plates, 25µl of genomic DNA was digested with 30 units of KpnI. Otherwise, approximately 15µg of ES cell DNA was digested with BamHI (for the 5´ end of the locus) or KpnI (for the 3´ end of the locus) in 40µl reactions. Digested DNA was resolved by electrophoresis at 1.5V/cm on a 0.8% (w/v) TAE agarose gel. The DNA was depurinated with 0.25M HCl for 15 minutes, rinsed with water then denatured in 0.4M NaOH for 45 minutes. It was then transferred overnight by dry blot to a Zeta-Probe GT Nylon Blotting Membrane (Bio-Rad), which had been pre-wet and incubated in NaOH for ten minutes. Blots were washed twice in 2xSSC (300mM NaCl, 30mM $Na_3C_6H_5O_7$, pH 7) before being blocked by modified Church & Gilbert hybridisation buffer (0.5M NaPi, 7% (w/v) SDS, 1mM EDTA, 10mg/ml BSA, 50µg/ml herring sperm ssDNA (Sigma)) at 65^{o}C for 30 minutes. For screening at the 5' end and 3' end, DNA was probed with a 1.15-kb HindIII fragment or 1.1kb NcoI-BamHI fragment, respectively. Approximately 30ng of probe was labelled with P^{32},

incorporated into cytosine nucleotides using the Megaprime DNA Labelling System (GE Healthcare) according to manufacturer's instructions. Radiolabelled probes were added to 40ml hybridisation buffer and hybridised overnight at 65°C. Blots were washed in 3xSSC (450mM NaCl, 45mM $Na_3C_6H_5O_7$, pH 7) and 1xSSC (150mM NaCl, 15mM $Na_3C_6H_5O_7$, pH7) then remaining activity checked with a Geiger counter, before being visualised by phosphorimagery (Typhoon FLA 9500).

2.3.8 Sequencing

Approximately 50ng DNA corresponding to regions of interest was sequenced using 2μl BigDye Terminator v3.1 ready reaction mix (Applied Biosystems) and 3.2 picomoles of primer in 10μl reactions. Cycling conditions were: 1 minute of denaturation at 96°C, then 25 cycles of 10 seconds of denaturation at 96°C, 5 seconds for annealing at 50°C and 4 minutes of extension at 60°C. Sequencing was completed by the University of Edinburgh Genepool facility.

2.3.9 Bisulfite sequencing

Bisulfite sequencing was used to determine the methylation status of genomic regions used in chromatin immunoprecipitation (ChIP) and DNA that was immunoprecipitated with MeCP2 during ChIP in mouse brain.

For genomic DNA, 2μg was bisulfite converted using the EpiTect Bisulfite Kit (Qiagen) according to manufacturer's instructions. In brief, DNA was incubated at 60°C and low pH (approximately 5) in the presence of sodium bisulfite to convert unmethylated cytosines to uracil, before being washed, desulfonated and purified. Approximately 50ng of converted DNA was amplified by PCR. The PCR product was gel analysed then bands of the correct size were cut out and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research), according to manufacturer's instructions. Recovered product was blunted and ligated (in 3 molar excess) into the pJET1.2/blunt vector using the CloneJET PCR Cloning Kit (Thermo Scientific) according to manufacturer's instructions. Competent DH5α E.coli (50μl) were transformed with 2.5μl of ligation reaction. For transformation, DNA was added to bacteria for 20 minutes on ice, heat-shocked at 42°C for 45 seconds, incubated on ice

for 2 minutes then grown in 0.9ml sterile Super Optimal broth with Catabolite repression (SOC) medium (20g/l bacto-tryptone, 5g/l bacto-yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM glucose, pH 7) at 37^{o}C for 1 hour. Cells were plated on Luria Bertani broth agar (10g/l bacto-tryptone, 5g/l bacto-yeast extract, 20mM NaCl, 20g/l bacto-agar, pH 7) with 50μg/ml ampicillin overnight at 37^{o}C. Colonies containing inserts were identified by growth on the plates, as recircularised vector expressed a lethal restriction enzyme. Inserts were PCR amplified from single colonies using pJET1.2 specific primers. 5μl of PCR reaction was resolved by agarose gel electrophoresis and reactions containing cloned fragments corresponding to the correct size were treated with 5 units of Exonuclease I (NEB) and 5 units of Antarctic Phosphatase (NEB) for 15 minutes at 37^{o}C. Enzymes were then heat inactivated at 80^{o}C for 15 minutes, before 3μl of reactions were sequenced. Sequencing data was aligned using Lasergene Software version 12.0.0 and data analysis performed with BiQ-Analyser software (Bock *et al*, 2005).

The protocol above was followed for methylation analysis of input and ChIP DNA from the 1-minute formaldehyde cross-link conditions in adult male mouse brain with the following modifications: (1) DNA was concentrated using the Agencourt AMPure XP Purification system (Beckman Coulter) prior to bisulfite conversion, (2) approximately 120ng of DNA was bisulfite converted, (3) approximately 12ng of converted DNA was PCR amplified using site-specific primers and (4) 5μl of PCR product was gel analysed, but the rest was purified using MSB Spin PCRapace (STRATEC Molecular), prior to blunting and ligation into pJET1.2/blunt.

2.4 RNA manipulation

2.4.1 RNA extraction

RNA was extracted from cells and tissues for the purposes of estimating relative mRNA transcript abundance by qPCR and microarrays.

ES cells, corresponding neuronal precursors and neurons differentiated in culture were mixed thoroughly with Tri Reagent (Sigma) by pipetting at a ratio of 1 x 10^7

cells: 1ml Tri Reagent. Frozen half brains were homogenised in 3ml Tri Reagent with the Ultra Turrax T25 for 15 seconds. After sitting at room temperature for 5 minutes, 0.1ml of 1-bromo-3-chloropropane was added per ml of Tri Reagent, then samples were vortexed for 15 seconds before standing at room temperature for 15 minutes. Samples were then centrifuged at 12,000g for 15 minutes at 4^{o}C and the aqueous phase was transferred to new tubes. Due to excess fatty tissue in the aqueous phase, whole brain RNA was cleaned up by re-extraction with 1 volume of chloroform, followed by another centrifugation step. RNA was precipitated with 0.5ml of isopropanol per ml of Tri Reagent: samples were mixed vigorously, left for 10 minutes at room temperature, then centrifuged at 12,000g for 10 minutes at 4^{o}C to pellet RNA. Pellets were washed with 75% (v/v) ethanol before being dissolved in nuclease free water (Ambion).

RNA extracted from 6-week male mouse cerebella was homogenised in 1ml of Tri Reagent with the Ultra Turrax T25 and cleaned up following isolation with the RNeasy Mini Kit (Qiagen) according to manufacturer's instructions. In brief, RNA was diluted in a buffer containing ethanol then bound to a silica-based membrane while contaminants and remaining genomic DNA were washed away. Molecules greater than 200 nucleotides in length were purified therefore there was enrichment for mRNA.

2.4.2 Measurement of RNA concentration & integrity

RNA concentration was measured at OD_{260nm} and OD_{280nm} on a Nanodrop-1000 spectrophotometer (Thermo Scientific) and an automated concentration reading was calculated using Beer's law with extinction coefficient for ssRNA 40ng/µl/cm. Purity was determined by the OD_{260nm}:OD_{280nm} ratio. Values ≥2.0 indicated a lack of protein or phenol contaminants.

Quality of cerebellar RNA was verified on the Agilent 2100 Bioanalyser by on-chip gel electrophoresis. This system uses integration of fluorescent dye and generation of electropherograms to detect the size and quantity of the migrating species with reference to a standard RNA 6000 ladder (Ambion) and marker. The RNA integrity

number (RIN) (which took the entire electropherogram trace into account) indicated RNA quality. A RIN of ≥8 was deemed good quality.

2.4.3 cDNA synthesis

RNA intended for qPCR was DNAse treated using the DNA-*free* kit (Ambion) then 500ng (cells differentiated in culture) or 1μg (mouse brain tissues) was reverse transcribed using the iScript cDNA Synthesis Kit (Biorad) in a reaction volume of 20μl. Cycling conditions were 5 minutes at 25^{o}C, 30 minutes at 42^{o}C then 5 minutes at 85^{o}C. Reactions containing everything except the iScript reverse transcriptase were included to control for potential contamination of DNA. Resulting cDNA was diluted in 100μl of nuclease free water (Ambion) and 5μl was used per reaction in qPCR.

2.5 Protein manipulation

2.5.1 Site directed mutagenesis

A recombinant human MeCP2 (amino acids 1-205) expression vector was constructed with a C-terminal histidine tag using the bacterial expression plasmid, pET30b (Novagen) by Dr R Klose (Klose *et al*, 2004). MeCP2(1-205) expression was therefore under control of the T7 promoter. The R133C mutation was placed using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) as per manufacturer's instructions. The PCR cycles included 6 minutes 10 seconds of elongation at 68^{o}C. The mutated plasmid was isolated using the Qiagen Qiaprep Spin Miniprep kit. The presence of the R133C mutation was confirmed by sequencing.

2.5.2 Expression of MeCP2(1-205) in *E.coli*

BL21(DE3)pLysS *E.coli* were transformed with MeCP2(1-205) plasmids because the bacteria express T7 RNA polymerase from the *lacUV5* promoter. Under basal conditions, this is minimised by the production of T7 lysozyme and lac repressor, but with the addition of isopropyl-β-D-thiogalactopyranoside (IPTG), expression can be strongly induced and reliably high transcription of MeCP2(1-205) was stimulated under control of the T7 promoter. 100μl of competent cells were transformed with

approximately 500ng of pet30bhMeCP2(1-205) and pet30bhMeCP2(1-205)^{R133C} under the same conditions described in 2.3.9, then plated on LB with 15μg/ml kanamycin overnight at 37^{o}C. Two scrapes of colonies were inoculated in 25ml LB medium (10g/l bacto-tryptone, 5g/l bacto-yeast extract, 20mM NaCl, pH 7) with 50μg/ml kanamycin & 17μg/ml chloramphenicol and shaken overnight at 37^{o}C. Cultures were inoculated in 500ml warm LB medium with 50μg/ml kanamycin & 17μg/ml chloramphenicol and shaken at 37^{o}C until the OD was 0.6-0.8 at 600nm. The addition of 1mM IPTG induced MeCP2 expression and cultures were transferred to 30^{o}C for 3 hours. *E.coli* was pelleted by centrifugation at 6,700 x g, 4^{o}C for 15 minutes and stored at -80^{o}C until use.

2.5.3 Purification of MeCP2(1-205)

Bacterial pellets were thawed on ice before being mashed in ice-cold lysis buffer (50mM NaH_2PO_4; 100mM NaCl; 10% glycerol; 30mM imidazole; 0.1% NP40; 2x protease inhibitor (complete EDTA free cocktail, Roche); pH 7.5) and passed through a 21G needle twice to break open cells. Following this, 750 units of benzonase were added before lysates were sonicated for 10 cycles of 30 seconds on/off, 30% amplitude on ice (Branson Digital Sonifier 450) to break down associated bacterial DNA. Sonicates were adjusted to 0.3M NaCl then spun down at 31,000g, 4^{o}C for 30 minutes.

Chelating Sepharose Fast Flow beads (GE Healthcare) were washed in double distilled water then coated in nickel by rotating in 0.2M $NiSO_4$ for 10 minutes. Coated beads were resuspended in 0.3M NaCl lysis buffer prior to use. Soluble protein was then precipitated from the sonication supernatant using its histidine tag by incubation with 0.5ml bead volume of nickel-coated beads for 1 hour at 4^{o}C. Protein-bound beads were washed three times in ice-cold 0.3M NaCl lysis buffer to remove non-specific binding. Lysis buffer was prepared with 250mM imidazole, which would displace protein from the nickel-coated beads. Beads were resuspended in 0.5ml of this buffer for 1 minute on ice, spun down briefly and eluted protein was collected in five fractions. Eluted fractions were pooled, made up to 1mM EDTA, transferred to a Slide-A-Lyzer Dialysis Cassette 7K MWCO (Thermo Scientific) and

dialysed against HEPES buffer (20mM HEPES pH 7; 100mM NaCl; 1mM EDTA; 5mM β-mercaptoethanol) overnight at 4^{o}C in two steps.

The protein underwent a further purification step by cation exchange chromatography using HiTRap SP Sepharose HP 1ml Columns (GE Healthcare). Protein bound to the negatively charged sulphopropyl coated agarose matrix was eluted with 0.7M NaCl in HEPES buffer. Protein concentration was measured and HEPES buffer was adjusted to 10% glycerol before snap freezing and storing suitably sized aliquots at -80^{o}C. Protein purity was assessed using SDS PAGE after appropriate dilution and addition of loading buffer.

2.5.4 Measurement of protein concentration

Protein concentration was measured using a Bradford Assay. By this method, differential colour change of acidic Coomassie Brilliant Blue G-250 dye occurs in response to variable concentrations of protein, because protein binding to dye causes a shift in absorbance maximum from 465nm to 595nm. Bio-Rad protein assay dye reagent concentrate was diluted 1 in 5 in distilled, deionised water. Known quantities of bovine serum albumin (BSA) and known volumes of MeCP2 polypeptides were added to 1ml aliquots, vortexed briefly and left for 15 minutes at room temperature. OD was measured at 595nm and a standard curve was derived from BSA readings. Protein concentration was estimated from comparison to the resulting BSA standard curve.

2.5.5 SDS-Polyacrylamide gel electrophoresis (PAGE)

ES cell pellets were resuspended in ice-cold NE1 buffer (20mM HEPES pH 7.9, 10mM KCl, 1mM $MgCl_2$, 0.1% (v/v) Triton X-100, 20% (v/v) glycerol, 0.5mM DTT, complete protease inhibitor cocktail (Roche)) at a ratio of 8 x 10^4 cells per μl of NE1. Then 250 units of benzonase were added per 2 x 10^7 cells and samples were left for 15 minutes at room temperature before adding 1 volume of 2x SDS loading buffer (125mM Tris HCl pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS, 0.25% (w/v) bromophenol blue, 20mM DTT, 0.3M β-mercaptoethanol) to leave a final

concentration of 4 x 10^4 cells/μl. Samples were boiled for 3 minutes, spun briefly then 12.5μl was loaded onto a 4%-12% Run Blue SDS precast gel (Expedeon).

For brain tissue, 6-week mice (adult) or postnatal day 0-2 (neonatal) were culled by cervical dislocation or decapitation, respectively. Brains were dissected, halved (adults) and snap frozen in isopentane on dry ice. They were stored at -80^{o}C until use. Half adult or whole neonatal brains were dounce homogenised with 10 strokes in 750μl of ice-cold NE1 buffer. Then 750 units of benzonase were added for 15 minutes at room temperature before adding 1 volume of 2x SDS loading buffer. After boiling for 3 minutes and a brief spin, 8μl was loaded onto 4%-12% Run Blue SDS precast gels (Expedeon). In the case of co-immunoprecipitation experiments, 10μl of sample were loaded onto a 4-15% Mini-PROTEAN® TGX™ Gel (Biorad).

Protein was resolved by electrophoresis at 120V for approximately 1 hour with PageRuler Prestained Protein Ladder (Thermo Scientific) in Rapid SDS Run Buffer (Expedeon). Co-immunoprecipitation gels were resolved in Tris-glycine SDS buffer (25mM Tris, 250mM glycine, 0.1% (w/v) SDS) at 270V for 30 minutes.

2.5.6 Wet transfer of proteins to a membrane

A Bio-rad mini trans-blot cell was used for wet transfer according to manufacturer's instructions. The transfer cassette was assembled as follows: 0.2μm pore size nitrocellulose membrane (Biorad) was cut to size and soaked in water for 5 minutes then transfer buffer (25mM Tris, 192mM glycine) along with 3mm Whatman paper; two pieces of wet Whatman were placed in the cassette with air bubbles carefully removed by rolling; the membrane was added, followed by the protein gel after a rinse in transfer buffer and another two wet pieces of Whatman paper. The cassette was immersed in ice-cold transfer buffer and transferred at a constant current of 400mA for 1 hour. Membranes were rinsed in Tris buffered saline (50mM Tris-HCl pH 8.0, 150mM NaCl (TBS)) briefly before western blotting.

2.5.7 Western blotting

Membranes were blocked and agitated gently by rocking in 5% milk (w/v), 0.1% Tween-20 (v/v) in TBS (milk TBST) overnight at 4°C, or for one hour at room temperature. Primary antibodies were diluted in milk TBST as follows: anti-MeCP2 1:1000 (mouse monoclonal, Sigma M6818), anti-Gamma tubulin 1:3000 (mouse monoclonal, Sigma T5326). They were incubated with the membrane for 4 hours at room temperature, or overnight at 4°C, with rocking agitation. The membranes were washed three times in 0.1% Tween-20 (v/v) in PBS (PBST) for 10 minutes. Secondary antibody was infra red dye-conjugated IgG (donkey anti-mouse 800, Licor 32212). It was diluted 1:10,000 in milk TBST and incubated with the membrane for 2 hours at room temperature. Following this, the blot was washed three times for 10 minutes in PBST and rinsed in PBS before imaging.

For co-immunoprecipitation experiments, blocking solution contained 0.05% (v/v) Tween-20. Primary antibodies, anti-MeCP2 (mouse monoclonal, Sigma M6818), anti-HDAC3 (rabbit polyclonal, Santa Cruz Biotechnology SC-11417) and anti-TBLR1 (rabbit polyclonal, Bethyl Laboratories A300 408A) were diluted in blocking solution at 1:1000. Secondary antibodies were horseradish peroxidase-conjugated. Anti-mouse or Anti-rabbit IgG (GE Healthcare NA931V or NA934V, respectively) was diluted 1:2000 in blocking solution.

Standard western blots using infra red-conjugated secondary antibody were imaged using the Licor Odyssey Infrared Imager. Quantification of bands was performed using Image Studio Lite Software (Licor, version 2.1.10). For co-immunoprecipitation experiments with HRP-conjugated secondary antibody, blots were treated with developing solution (100mM Tris HCl pH 8.8; 0.01% H_2O_2; 1.5mM luminol; 0.25mM p-coumaric acid). The blot was dried, wrapped in saran and exposed to Hyperfilm ECL (Amersham Biosciences).

2.5.8 Co-immunoprecipitation

Frozen half brains from 6-week male mice were dounce homogenised with 10 strokes in 1ml ice-cold NE1 buffer and pelleted at 800 x g for 5 minutes at 4°C.

Pellets were washed in NE1 then resuspended in 500µl NE1 with 375 units of benzonase for 10 minutes at room temperature. Samples were adjusted to 150mM NaCl and incubated at 4°C for 20 minutes. After centrifugation at 14,500 x g, 4°C for 20 minutes, 30µl of supernatant was set aside from each sample for input. Equal volumes of supernatant were incubated with 5µl bead volume of anti-GFP antibody coupled to agarose beads (GFP-Trap_A, Chromotek) for 1 hour at 4°C in order to immunoprecipitate $MeCP2^{EGFP}$ and associated binding partners. Beads were then washed four times in 150mM NaCl NE1 to minimise non-specific immunoprecipitation, removing as much as possible from the beads following the final wash. All samples were then resuspended in 20µl of 2x SDS loading buffer (inputs in 30µl), boiled for 3 minutes, spun briefly and equal volumes were resolved by SDS PAGE. Following this, protein was transferred to nitrocellulose membranes and visualised by western blotting.

2.6 Immunofluorescence

Immunofluorescence was conducted to confirm expression of neuronal proteins in cells differentiated in culture and assess the pattern of MeCP2 association with heterochromatin in neurons differentiated in culture and mouse brain.

2.6.1 Neurons differentiated in culture

Neurons grown on coverslips were washed twice in PBS on day six before being fixed for 15 minutes with 3.7% (w/v) paraformaldehyde in PBS, pH 7.4. They were then washed three times with PBS, before being permeabilised for 15 minutes with 0.1% (v/v) Triton X-100 in PBS then washed another three times in PBS. Coverslips were incubated for 20 minutes at room temperature in blocking solution (1.5% (v/v) goat serum (Sigma) in PBS). Primary antibodies were diluted in blocking solution as follows: anti-NeuN 1:300 (rabbit polyclonal, Millipore ABN78) and anti-neurofilament 1:500 (mouse monoclonal, Covance SMI 311R) and applied for 1 hour. Coverslips were washed three times in PBS then secondary antibodies were applied for 1 hour at a dilution of 1:400 in blocking solution (Alexa Fluor 594 goat anti-mouse IgG (Life A11032) and Alexa Fluor 633 goat anti-rabbit IgG (Life A21071)). Coverslips were washed with PBS then stained for 15 minutes with

1μg/ml DAPI, followed by another three PBS washes. Finally coverslips were mounted on slides using Prolong Gold Antifade Reagent (Molecular Probes). Slides were visualised on a Leica TCS SP5 Confocal Laser Scanning Microscope using 21-slice z-stacks taken at 100x magnification, 8 bit resolution, with pixel size 512 x 512. Stacks were compressed using Image J software version 1.46.

2.6.2 Cell dot quantification

Intensity of DAPI fluorescence in neuronal nuclei was analysed with Image-Pro Plus Software version 7.0 using a macroinstruction programme written by D.Kelly. Firstly, nuclei were highlighted then could be selected manually in order to discount dead cells. The maximum intensity was defined across a field of cells by the brightest DAPI pixel. To be considered a "DAPI dot", areas had to be 0.67 of this brightness and more than one pixel in area. The intensity of these "dot areas" was then measured on the corresponding GFP panel and compared to the average GFP intensity throughout the rest of the nucleus. These measurements were taken for 50 cells per genotype. Dot prominence was calculated as the average GFP intensity of the dots/(the average GFP intensity of the dots + the average GFP intensity of the nuclear background) to give an estimate of how much $MeCP2^{EGFP}$ was located within the dot areas.

2.6.3 Mouse brain

Adult (6-12 week) mice were culled by cervical dislocation and rapidly perfused with 3.7% (w/v) paraformaldehyde in PBS, pH 7.4 injected into the left ventricle by J.Selfridge. The quality of the anti-MeCP2 signal appeared to be dependent on a rapid fix. Brains were then harvested and slowly fixed overnight in paraformaldehyde at 4^{o}C. This solution was replaced with 30% (w/v) sucrose in PBS overnight to dehydrate the tissue. Following this step, brains were washed briefly in PBS then blotted dry, halved, snap frozen in isopentane cooled on dry ice and stored at -80^{o}C until used. Sections were taken of 10μm thickness on the Leica CM 1900 Cryostat and dried on slides at room temperature. Slides were rinsed in PBS then washed twice for 10 minutes in Coplin jars. Slices were permeabilised in 0.1% (v/v) Triton X-100 in PBS for 15 minutes, washed three times in PBS then blocked in

1.5% (v/v) goat serum in PBS for 1 hour at room temperature. Primary antibodies were diluted in this blocking solution as follows: anti-MeCP2 1:200 (rabbit monoclonal, 07_013 Millipore), anti-NeuN-Cy3 1:100 (mouse monoclonal, MAB377C3 Millipore). Sections were stained with primary antibodies overnight at 4^{o}C in a humidified chamber. Slides were washed in PBS three times then sections stained for 1 hour with secondary antibody at a dilution of 1:400 in blocking solution (Alexa Fluor 633, goat anti-rabbit IgG, Life A21071) if required. Finally, slides were washed for 10 minutes with PBS then sections stained for 10 minutes with 1μg/ml DAPI before two final 10-minute PBS washes. Coverslips were mounted on slides using Prolong Gold Antifade Reagent (Molecular Probes). Slides were visualised on a Leica TCS SP5 Confocal Laser Scanning Microscope using 21-slice z-stacks taken at 63x or 100x magnification, 12-bit resolution, with pixel size 512 x 512. Stacks were compressed using Image J software version 1.46.

2.7 Chromatin extraction and immunoprecipitation

Chromatin immunoprecipitation (ChIP) was performed in order to estimate binding of R133C-GFP relative to WT-GFP at selected regions in the genome with known methylation patterns. MeCP2^{EGFP} was precipitated using an anti-GFP antibody mounted on agarose beads, GFP-Trap_A (Chromotek).

2.7.1 Neurons differentiated in culture

Day six neuronal cells were fixed with 360mM formaldehyde in PBS for 10 minutes at 37^{o}C before quenching the cross-link with 125mM glycine in PBS for 5 minutes at room temperature, with gentle agitation. Dishes were washed twice with 5ml of ice-cold PBS before cells were harvested by scraping. Neurons were pelleted by centrifugation at 350 x g, 4^{o}C for 5 minutes then resuspended in WASH 1 (0.25% (v/v) Triton X-100, 10mM EDTA, 10mM EGTA, 10mM HEPES, protease inhibitor (complete EDTA free cocktail, Roche), 1μM PMSF). Samples were incubated for 10 minutes on ice to allow swelling of cell membranes before centrifugation at 4^{o}C. This was repeated with second wash, WASH 2 (200mM NaCl, 1mM EDTA, 10mM EGTA, 10mM HEPES, protease inhibitors, 1μM PMSF) in order to lyse the cell membrane, leaving nuclei in tact. The nuclear pellet was resuspended in lysis buffer

according to size (1% (w/v) SDS, 10mM EDTA, 50mM Tris HCl pH 8.1, protease inhibitors, 1μM PMSF) in order to lyse the nuclear membrane, resulting in a solution of cellular chromatin.

Chromatin was sonicated for 15 cycles of 30 seconds on-off on high power using the Diagenode Biorupter Twin Sonicator in order to generate 250-400bp fragments. Debris was pelleted by centrifugation for 10 minutes at 4^{o}C then the concentration of chromatin was measured. For ChIP, 50μg of chromatin was diluted 1 in 10 (170mM NaCl, 17mM Tris HCl pH 8.1, 1mM EDTA, 1.1% (v/v) Triton X-100, 0.01% (w/v) SDS) and 5μg was set aside for inputs. ChIP chromatin was pre-cleared for 1-2hours at 4^{o}C with 30μl of blocked agarose beads (bab-20, Chromotek). Following removal of these beads by centrifugation at 4^{o}C, 40μl of GFP-Trap_A slurry (washed in TE: 10mM Tris HCl pH 8, 1mM EDTA) was incubated with the ChIP chromatin by rotation overnight at 4^{o}C.

Beads were washed for 4 minutes at 4^{o}C in RIPA (150mM NaCl, 50mM Tris HCl pH 8.0, 0.1% (w/v) SDS, 0.5% (w/v) sodium deoxycholate, 1% (v/v) NP40), high salt buffer (500mM NaCl, 50mM Tris HCl pH 8.0, 0.1% (w/v) SDS, 1% (v/v) NP40), LiCl buffer (250mM LiCl, 50mM Tris HCl pH 8.0, 0.5% (w/v) sodium deoxycholate, 1% (v/v) NP40) followed by two TE washes. Between washes the beads were pelleted by centrifugation for 1 minute, 2,600 x g, 4^{o}C. Immunocomplexes were eluted in 400μl of elution buffer (2% (w/v) SDS, 100mM Na_2CO_3, 10mM DTT), along with 200mM NaCl to reverse the cross-link. Elution buffer and NaCl was also added to inputs and samples were shaken overnight at 65^{o}C. DNA was extracted by adjusting to 10mM EDTA, 40mM Tris HCl pH 6.5, 0.1μg/μl proteinase K for 1 hour at 55^{o}C and removing proteins by phenol-chloroform extraction. DNA was precipitated using NaOAc and ice-cold ethanol at -20^{o}C with 20μg glycogen as a carrier. DNA pellets were washed with 70% ethanol and dissolved in 200μl nuclease-free water. DNA from input samples was diluted 1:20 (overall 1 in 200 dilution compared to ChIP samples) and 2.5μl was used for qPCR analysis.

2.7.2 Whole mouse brain

Frozen half brains (adult) or whole brains (neonate) were dounce homogenised with 10 strokes in 1ml cold PBS. Cross-linking was either for 8 minutes at room temperature (standard), 1 minute at room temperature (short cross-link) or 1 minute straight from ice (stringent) with 355mM formaldehyde in PBS. Cross-link quench was either with 125mM glycine in PBS (standard) or 250mM glycine in PBS (short cross-link and stringent). The protocol was otherwise the same as for neuronal cells differentiated in culture with the following modifications: the chromatin was not pre-cleared; ChIP samples were incubated with GFP-Trap_A for 1.5 hours at 4^{o}C; and neonatal DNA was dissolved in 150µl nuclease-free water at the end of the protocol.

2.7.3 ChIP-seq

DNA libraries were prepared from the R133C-GFP short cross-link ChIP and input samples using the Nugen Ultralow Library System by S.Lagger. In brief, this involved ligation of unique barcoded adaptors onto the fragmented DNA and amplification by PCR. Libraries were sequenced at the Wellcome Trust Sanger Institute, Cambridge using Illumina Solexa high throughput paired end sequencing. The barcoded DNA was bound to complementary immobilised adapters on a flow cell. DNA was amplified generating clusters of identical amplicons, which were then sequenced. The unique barcoding meant that more than one sample could be processed in parallel in one flow cell lane. The sequencing data was analysed by S.Webb. MeCP2 ChIP summits (peaks of MeCP2 found in the ChIP sample compared to the input sample) were derived using the macs2 callpeaks algorithm with a minimum fold change of 1, maximum fold change of 30 and a p value cut off of 0.2. The maximum fold change meant that large counts were removed, as these tended to be artefacts, arising from the sequencing process. The ChIP summits were extended 4kb up- and downstream and divided into 20 nucleotide bins. Mean methylation density was calculated for CG and CA dinucleotides using a published bisulfite-seq data set from adult mouse frontal cortex (Lister *et al*, 2013). For each bin, the percentage methylation of individual cytosines (CG or CA) was summed and divided by 20. The mean bin methylation density was plotted for all summit regions.

2.8 Electrophoretic mobility shift assays (EMSA)

EMSA's were used to assess MeCP2(1-205) binding to methylation and hydroxymethylation in particular sequence contexts. Unphosphorylated single stranded oligonucleotides corresponding to the *BDNF* promoter region (plus complementary reverse strands) were obtained from Biocore and annealed by J.Connelly. In brief, oligonucleotides were resuspended in nuclease-free water, and 600 picomoles of complementary pairs were boiled for 20 minutes in 150mM NaCl & 10mM Tris HCl pH 8. Tubes were allowed to cool slowly so that complimentary oligonucleotides annealed to form double-stranded probes. Probes were then radio-labelled by adding 500ng to 2μl ATP gamma ^{32}P, with 25 units of T4 polynucleotide kinase (PNK) and 1x polynucleotide kinase buffer (NEB). Reactions were incubated at 37^{o}C for 30 minutes so that PNK could catalyse the transfer of phosphate from ATP to the probes. The radio-labelled probes were then extracted using the MinElute PCR Purification Kit (Qiagen), resuspended to a final concentration of 10ng/μl and stored at -20^{o}C until used.

Recombinant protein concentration was adjusted using ice-cold reaction buffer (0.1mg/ml BSA, 5% (v/v) glycerol, 0.1mM EDTA, 10mM Tris HCl pH 7.5, 150mM KCl) then added to 1μg polydeoxyadenylic-thymidylic acid competitor DNA (Sigma-Aldrich) and 1ng double-stranded ^{32}P-labelled probe in reaction buffer with careful mixing. Reactions were incubated on ice for 20 minutes. Loading buffer was added (4% (v/v) glycerol, 2mM Tris HCl pH 7.5, 5mM NaCl, 0.0002% (w/v) Bromophenol Blue) and reactions were resolved by PAGE. EMSA gels were 10% (w/v) acrylamide:bis-acrylamide (29:1, Bio-rad), Tris Borate EDTA (45mM Tris-borate, 1mM EDTA (TBE)), polymerised by ammonium persulphate with TEMED as a catalyst. Gels were self-poured using the Bio-rad Mini Protean-3 apparatus. Reactions were resolved by PAGE at 120V for 70 minutes in chilled TBE and visualised using phosphorimagery (Typhoon FLA 9500). Percentage of probe shifted was calculated using Image J software version 1.46.

2.9 Gene expression by microarray

2.9.1 RNA library preparation and microarray

RNA was prepared from 6-week male mouse cerebella and analysed as described in 2.4.1 and 2.4.2. Libraries of copy RNA (cRNA) were prepared from 400ng using the Illumina Total Prep RNA Amplification Kit (Ambion). In brief, cDNA was prepared with reverse transcription using a T7 oligo(dT) primer, followed by second strand synthesis and degradation of remaining RNA using RNAse H. cDNA was column purified before undergoing *in vitro* transcription with T7 RNA Polymerase and biotin-nucleotide triphosphates to synthesize cRNA that was biotinylated. This biotinylated cRNA was column purified then sent to the Wellcome Trust Clinical Research Facility (WTCRF) in Edinburgh to be assessed for quality. The WTCRF staff then measured gene expression on MouseWG-6 v2 BeadChips (Illumina) using a direct hybridisation assay, where signal detection is dependent on streptavidin-Cy3.

2.9.2 Analysis of microarray data

Gene expression levels were analysed using the Bioconductor Limma package by S.Webb. Probes were removed from the analysis if they were detected in fewer than three samples and background correction, normalisation, transformation, differential expression analysis and annotation were performed on the remaining set following standard procedures (Ritchie *et al*, 2011). For genes with multiple probes, mean log fold changes were calculated. Genes were ordered by length based on the Ensembl GRCm39 release 79 annotations and divided into bins using a window of 400 genes and a sliding step of 80 genes. The mean log fold change was plotted against the mean length of genes in each bin.

2.10 Primers

Primers were resuspended in nuclease-free water (Ambion) at a concentration of 100μM and stored at -20°C.

	Forward Primer	Reverse Primer
Site Directed Mutagenesis	CCCAGGGAAAAGCTTTT**TGC**T CTAAAGTAGAATTG	CAATTCTACTTTAGA**GCA**AAAA GCTTTTCCCTGGG
Targeting PCR	GGATCTGACATGGTAAGTAAG CTA	GCTGTCTCATTTGCTACTTTGA TA
NEO PCR	GTCATCTCACCTTGCTCCTGC C	GAAGGCGATAGAAGGCGATGC G
pET30b sequencing	GGGGTTATGCTAGTTATTGCTC A	GGTGATGTCGGCGATATAGG

2.10.1 Real time qPCR of ChIP DNA

	Forward Primer	Reverse Primer
Major Satellite*	GGCGAGAAAACTGAAAATCACG	AGGTCCTTCAGTGTGCATTTC
L1 LINE**	TTTGGGACACAATGAAAGCA	CTGCCGTCTACTCCTCTTGG
IAP Elements**	GAGATTGGACTTTTGACTTGT	TGTGGCTTGCTCATAGATTAG
BDNF	TTCGATTCACGCAGTTGTTC	CTGAGCCAGTTACGTGACCA
Cyp3a25	CAGGTTTGGGGTGTTGTGAA	CTGCAGCTGTTGTGGGAG
Cyp3a57	GTGCTGCTCTTACATGGCTG	GTGGGGCTACAGTCTATGCT
Cyp3a59	CCTGACTGGCTGCTCACTAT	AGGCTGTGAACTATAGGAGCC
Mad1|1	TGGTGAGACAGCCTTCTTCT	GTGCCCCTCTGACCTCTG
Egfl7	GCAGCTGGACCGAATTGATT	TGCCCATCCCAGAAGAAGTT
Olfr325	TCTTGGTTGCCTAGGAGACA	GGTGGGTGTCAGAGAGTACC
Cpa6	GCCAAGTACACCCTCTCCTT	TTCTCTACAGGGTCCAGTGC
Gata3	ACCCCTTTATTCCTCCGTGT	CCAGGAGAGGGGTCGTTTAA

*Skene *et al*, 2010
**Brunmeir *et al*, 2010

2.10.2 Real time PCR of cDNA

	Forward Primer	Reverse Primer
Mecp2	ACCTTGCCTGAAGGTTGGAC	GCAATCAATTCTACTTTAGAGCGAAAA
Cyclophillin A	TCGAGCTCTGAGCACTGGAG	CATTATGGCGTGTAAAGTCACCA
Abhd1	TTGGCGTTGCTCTCTACTTG	TCCAGGAAGGCCAGAAACT
Aplf	GGGGCAGACTGTGATAGGC	AATGGCATGTCTTCTGGATACTC
Dag1	CTGCTGCTGCTCCCTTTC	CCAGGCAGTGTTGAAAACCT
Krt222	GCTCAGGCTGAACTCAAGGA	GCCTGCAGGGAGTTCTCA
Zfp428	GATGAAGACCTTTCCCCAGA	GGGTCATCAGTGGTCTCCTCT
*Camk1d**	CCTCTACCTGGTCATGCAACT	TGTGTAAAACCCCTTCTCCACT
*Prkca**	CTGGTGCTTGGGTTGAATG	TAACTCCTGGGGCTGCAC
*Robo1**	AGGGAAGCCTACGCAGATG	TGGACAGTGGGCGATTTTAT
*Camk2d**	TTACAGTGAAGCTGATGCCAGT	TTAGGTCCCGATGGACTACG
*Plekha5**	ACAGGCTGTTAGTCCCCAGA	GCCGTCTCAGCTTTATTAGCA
*Arhgap26**	TGCACTTGCTTTTTGACAGG	TTGGCAGGCGTACAAAGC
*Lrrc8d**	CCAGGATGGAGGAGTGAAGT	CGCAAGGGTAAACATTCCTG
*Ptprk**	CATCAATGCTGCTCTTATGGAT	GCAGTGGGTATTGTGTGACG
*Magi1**	GGGAGGCTTGCTTCTACCTT	TTCCGGAACACCTTGTGC
*Tiam1**	CCAGTCCCTGGCTGAAAAT	GTCCCCGAAGTCGTCTAGG
*Ubp1**	ACGAGACACTTACTTACTTGAACCAA	TGTCTCCCATTTTGCGATTA
*Rufy3**	GACGGAGAATGGCTCTGC	GCCATGAGATAATTGGGATCTT
*Nr3c1**	CAAAGATTGCAGGTATCCTATGAA	TGGCTCTTCAGACCTTCCTT

*Gabel *et al*, 2015

2.10.3 Bisulfite sequencing

	Forward Primer	Reverse Primer
Colony PCR	CGACTCACTATAGGGAGAGCGGC	AAGAACATCGATTTTCCATGGCAG
Major Satellite (Skene *et al*, 2010)	GGAAAATGAGAAATATATATTTTAGGA	TCAATTTTCTTACCATATTCCA
L1 LINE Elements	TTTTTAATGATATTTTGGTTAAGGA	CACTTAATTAATTTCACCCCTAAA
IAP Elements	TTTGATTTTAGAGTAGAGAGATTGG	CAACTTACTCTAACTCTCCAAAAAT
BDNF	TAGTTTTTAAGAGGAAAAGGGA	AAAAAAACCAAAACAACCTC
Bai3 (Guo *et al*, 2014)	AATGAAAGTAATAGGAGTTAGGGGG	ACCCAATTCATAATTAAAATAAAACATCTT
Fgf1b (Guo *et al*, 2014)	TTTTGTTGATGAGTAAGGGTTAAGG	CAAACTAAAAAACTCTCTTCACTCCA

3. Construction of *Mecp2*^{R133C} models

3.1 Introduction

Arginine 133 occupies a key position in the MBD of MeCP2, yet its mutation to cysteine results in a form of Rett syndrome that is at the milder end of the phenotypic spectrum (Leonard *et al,* 2003). Several laboratories have assessed the impact of the mutation on DNA binding. Findings have been variable, indicating that binding to methylated DNA is impaired (Ballestar *et al,* 2000) and in contrast, nearly indistinguishable from wildtype by alternative assays (Mellen *et al,* 2012). The degree of binding impairment has varied according to the protein species, fragment length, target probe and assay. No consistent conclusions can be drawn regarding the impact of this mutation on DNA binding from the previous literature.

The explanation for this may lie in the choice of model system in which to study protein function. The R133C mutation in MeCP2 has only previously been studied *in vitro* or in non-neuronal cell types. ES cells and fibroblasts have been used in experimental paradigms where mutant protein was transfected or expressed from an artificial promoter at an alternative locus, on top of endogenous protein expression. Rett syndrome is a neuropsychiatric disorder and deletion of MeCP2 in the brain only is sufficient to cause RTT (Guy *et al*, 2001). MeCP2 is an abundant nuclear protein that binds methylated cytosines and is most highly expressed in neurons (Skene *et al,* 2010). It is appreciated that the methylome is quite unique in the mature nervous system, with a greater proportion of methylation occurring in a CAC context and a higher abundance of hmCpG than other tissues (Lister *et al,* 2013). Using more relevant model systems might yield more consistent conclusions regarding the impact of the R133C mutation on DNA binding, as well as other putative functions of MeCP2.

ES cells can be differentiated into a homogeneous population of mature cells in culture (Bibel *et al,* 2004). If ES cells were targeted by homologous recombination with *Mecp2* carrying the R133C mutation, they could be differentiated into a population of neuronal cells under defined culture conditions. Similar strategies have

been used previously with *Mecp2*-null ES cells (Li *et al,* 2013) and induced pluripotent ES cells from patient material (Marchetto *et al,* 2010). The major advantage is the clonal nature of the cells, thus eliminating the issue of potential cell type-specific effects obscured with the study of whole tissues (Sugino *et al,* 2014). MeCP2 would be abundantly expressed from the endogenous promoter with no competing alternative transfected species, in the context of a neuronal methylome.

Mouse models of *Mecp2* deficiency have been used extensively to study the impact of protein function in the whole organism, yielding interesting insights into the pathogenesis of Rett syndrome. For example, the seminal finding that replacement of the protein on an *Mecp2*-null background alleviates the Rett-like phenotype (RTT) in the mouse (Guy *et al,* 2007). The advantage of using a whole organism is that the interaction of the mutation and the phenotype can be studied. With mouse breeding protocols, it is possible to obtain a relatively homogeneous genetic background by backcrossing. This virtually eliminates confounding genetic variation when interpreting the impact of an individual mutation. In the mouse, unlike in patients, males with Rett syndrome survive into adulthood (Guy *et al,* 2001). In males, any manipulation of *Mecp2* can be studied without the added complication of the individual's pattern of XCI and its influence on phenotypic expression because there is no need for dosage reduction (Gibson *et al,* 2005). There are no existing mouse models of the R133C mutation. In this context, it would be possible to confirm whether the R133C mutation does result in a milder form of RTT than other *Mecp2* mouse models (Guy *et al,* 2001; Lyst *et al,* 2013), as is seen in patients (Cuddapah *et al,* 2014). This would yield important prognostic information for patients and provide a useful resource for trial of therapeutic interventions.

The enhanced green fluorescent protein (EGFP) tag is widely used as a means of easily identifying a protein of interest due to its bright and reliable expression in comparison to wildtype GFP. There are obvious advantages to using the tag for microscopy in that fluorescence is stable, species-independent and there is no need to apply additional cofactors (Zhang *et al*, 1996). There are also reliable anti-EGFP antibodies commercially available, making immunoprecipitation of tagged proteins

in the context of chromatin, or in association with other protein complexes achievable. An *Mecp2* targeting vector containing a C-terminal EGFP tag was selected for these reasons.

The objective was to knock $Mecp2^{R133C}$ in to the endogenous locus in wildtype ES cells. The resultant protein would have a C-terminal EGFP tag and be expressed from the endogenous promoter. These targeted cells could be differentiated along the neuronal lineage or injected into blastocysts to create a homogeneous population of $Mecp2^{R133C}$ neuronal cells and an $Mecp2^{R133C}$ mouse model, respectively. Together, these would provide pertinent model systems in which to study the impact of the R133C mutation on MeCP2 function.

3.2 $Mecp2^{R133C}$ targeting strategy

The objective was to create ES cells that were targeted by homologous recombination with *Mecp2* carrying the R133C mutation and a C-terminal EGFP tag. In this way, the effects of the mutation on the protein in neurons and mouse brain could be studied. A schematic of the desired protein with functional domains and putative AT hook motifs is shown in Figure 3.2. A suitable targeting construct created by J.Guy was available: a 7.2-kb plasmid subclone of mouse 129/Ola genomic DNA, including *Mecp2* exon 3 and exon 4, had the coding sequence of EGFP fused in-frame to the end of the coding sequence of exon 4, followed by a LoxP-flanked neomycin resistance (NEO) cassette as a selectable marker (plus STOP codon), retaining the first 2 kb of the *Mecp2* 3′ untranslated region. The targeting construct is depicted in Figure 3.2, along with the wildtype locus. Point mutation R133C was introduced into the wildtype $Mecp2^{EGFP}$ targeting vector by C.Merusi. The C to T substitution engineered in codon 133, resulted in the R133C mutation, which was confirmed by Sanger sequencing (data not shown). The targeting vector was diagnostically digested prior to use, following amplification (data not shown).

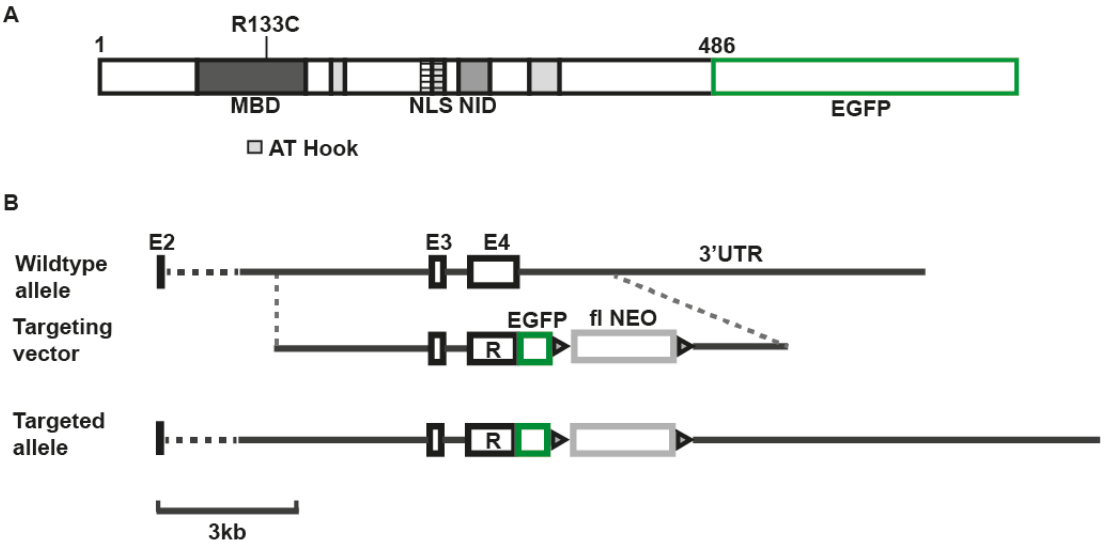


Figure 3.2 *Mecp2*^{R133C} targeting strategy

(A) Schematic representation of MeCP2 protein resulting from ES cell targeting. Isoform e2 is shown with a C-terminal EGFP tag. The major functional domains are shown: the methylated DNA binding domain (MBD) with position of the R133C missense mutation, the nuclear localisation signal (NLS) and the NCoR Interaction Domain (NID), along with three putative AT hooks.
(B) Diagram depicting the targeting strategy by homologous recombination. Exons 2-4 (black boxes) and the 3' untranslated region (UTR) of the endogenous *Mecp2* locus are represented. The targeting vector is shown with the R133C mutation (R) in exon 4, C-terminal EGFP coding sequence (green box) and floxed neomycin resistance cassette (fl NEO, grey box flanked by triangles) along with the targeted locus resulting from homologous recombination.

3.3 Targeting of wildtype ES cells with *Mecp2*^{R133CEGFP} (*R133C-GFP*)

After being linearised at the 5' end by SnabI, wild type 129/Ola ES cells were electroporated and selected by culture in G-418 supplemented ES cell medium. Colonies that were resistant to the antibiotic were picked and screened by polymerase chain reaction (PCR), Southern blot and finally sequencing. PCR primers were designed that flanked the NEO cassette and the 3' untranslated region, outwith the region that was included in the targeting vector (Figure 3.3A). The resulting PCR product of 2.2kb, should only be seen in ES cell clones correctly targeted by homologous recombination. A plate of colonies screened by PCR and resulting 2.2kb product is shown in Figure 3.3B.

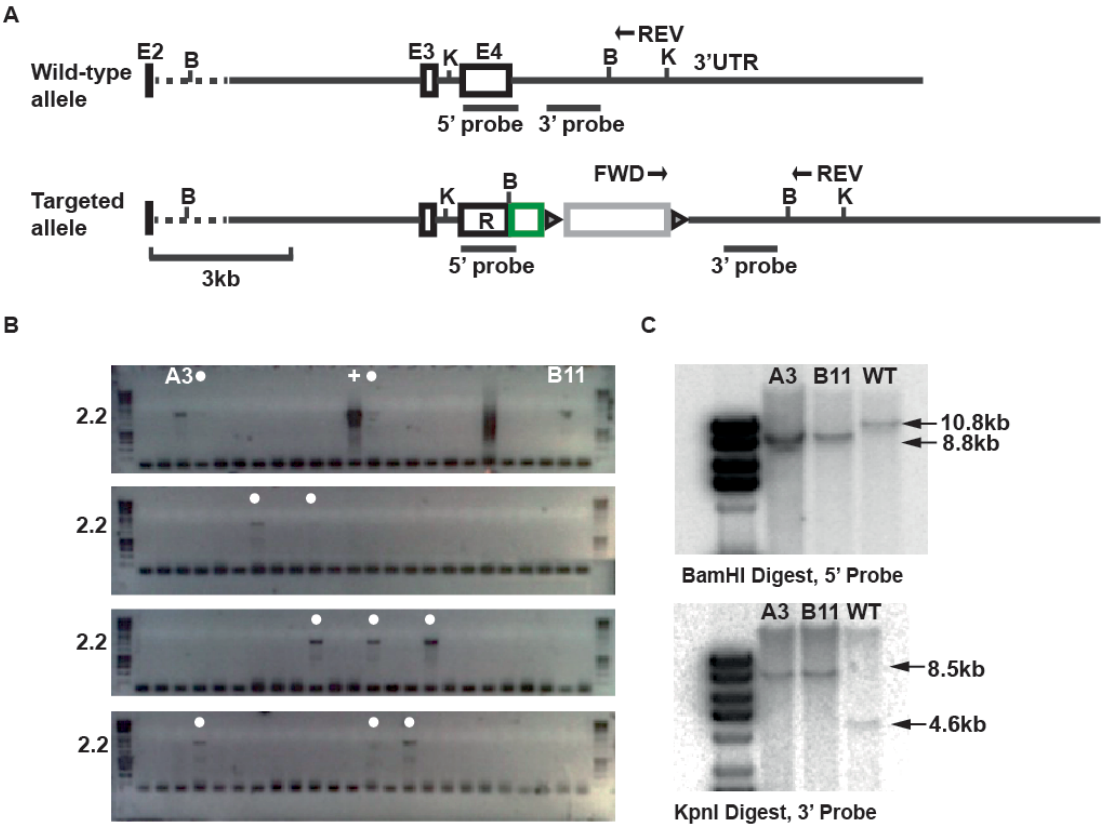


Figure 3.3 Successful targeting of wildtype ES cells with *Mecp2*^{R133CEGFP} (*R133C-GFP*)

(A) Maps of the wildtype and targeted loci are shown for reference including exons 2-4 (black boxes), EGFP tag (green box), floxed NEO cassette (grey box flanked by triangles) and R133C mutation (R). PCR primers (arrows), BamHI (B) and KpnI (K) restriction sites and Southern blot probes used to screen colonies are represented.
(B) Screening by PCR showed correctly targeted clones. An ethidium bromide gel with the desired 2.2kb PCR product in 11% of picked colonies indicates homologous recombination. Positive control DNA is indicated (+). Clones A3 and B11, which were taken forward to make neurons and the mouse, are labelled in white lettering.
(C) Clones A3 and B11 were screened for homologous recombination by Southern blot. Upper panel, screening at the 5' end of the locus following BamHI digestion resulted in an 8.8kb band at the targeted locus and a 10.8kb band at the wildtype (WT) locus. Lower panel, screening at the 3' end of the locus following KpnI digestion resulted in an 8.5kb band at the targeted locus and a 4.6kb band at the wild type (WT) locus.

Selected clones were then screened by Southern blot to check that the vector had been incorporated into the endogenous locus. Following digestion with BamHI, the 5' end of the locus was analysed. Digestion of the targeted locus would yield an 8.8kb band in comparison to the 10.8kb band seen following digestion of the wildtype locus, due to the introduction of an additional BamHI site by the targeting construct (Figure 3.3A&C). Clones were also screened at the 3' end following digestion with KpnI. The targeted locus yielded an 8.5kb band, whereas the wildtype

locus yielded a 4.6kb band. This was due to the introduction of the EGFP tag and floxed NEO cassette between KpnI restriction sites in the targeted locus (Figure 3.3A&C). Clones A3 and B11 were deemed correctly targeted following these two screening methods and so were also checked by Sanger sequencing (data not shown). The creation of two *Mecp2*^{R133CEGFP} ES cell lines was confirmed. From here on in, the *Mecp2*^{R133CEGFP} genotype will be referred to as *R133C-GFP*.

3.4 Creation of an *R133C-GFP* mouse model.

It was desirable to create a novel mouse line carrying the R133C mutation in order to study genotype-phenotype correlation and have access to mature brain tissue. Clone B11 was selected and injected into C57BL/6J mouse blastocysts by J.Selfridge. This was done prior to deletion of the NEO cassette so that cells had a lower passage number and spent less time in culture, thus reducing the opportunity to acquire additional random mutations of DNA. On average, 12 injected blastocysts were re-implanted per female recipient, primed for pregnancy by being mated with sterile males.

By using 129/Ola ES cells for targeting, it was apparent which resulting pups comprised a combination of the parent blastocyst cells and the targeted stem cells by nature of their mixed coat colour. Overall, from 9 surgeries, 11 chimaeras were generated. These chimaeras were test mated with C57BL/6J mice to establish whether the targeted cells had contributed to the germ line. The test breeding mice also expressed Cre recombinase under the cytomegalovirus (CMV) promoter, thus simultaneously excising the floxed NEO cassette from the endogenous locus. Germ line transmission was again apparent by the mixed (agouti) coat colour of the resulting pups. Mice were then back-crossed for a further four generations onto a C57BL/6J genetic background in order to minimise other sources of confounding genetic variation, before phenotypic analysis and harvesting of brain tissue. The injection strategy and breeding plan is depicted in Figure 3.4. The result was creation of a novel *R133C-GFP* mouse line on a C57BL/6J genetic background. Mutant pups were born at expected Mendelian ratios.

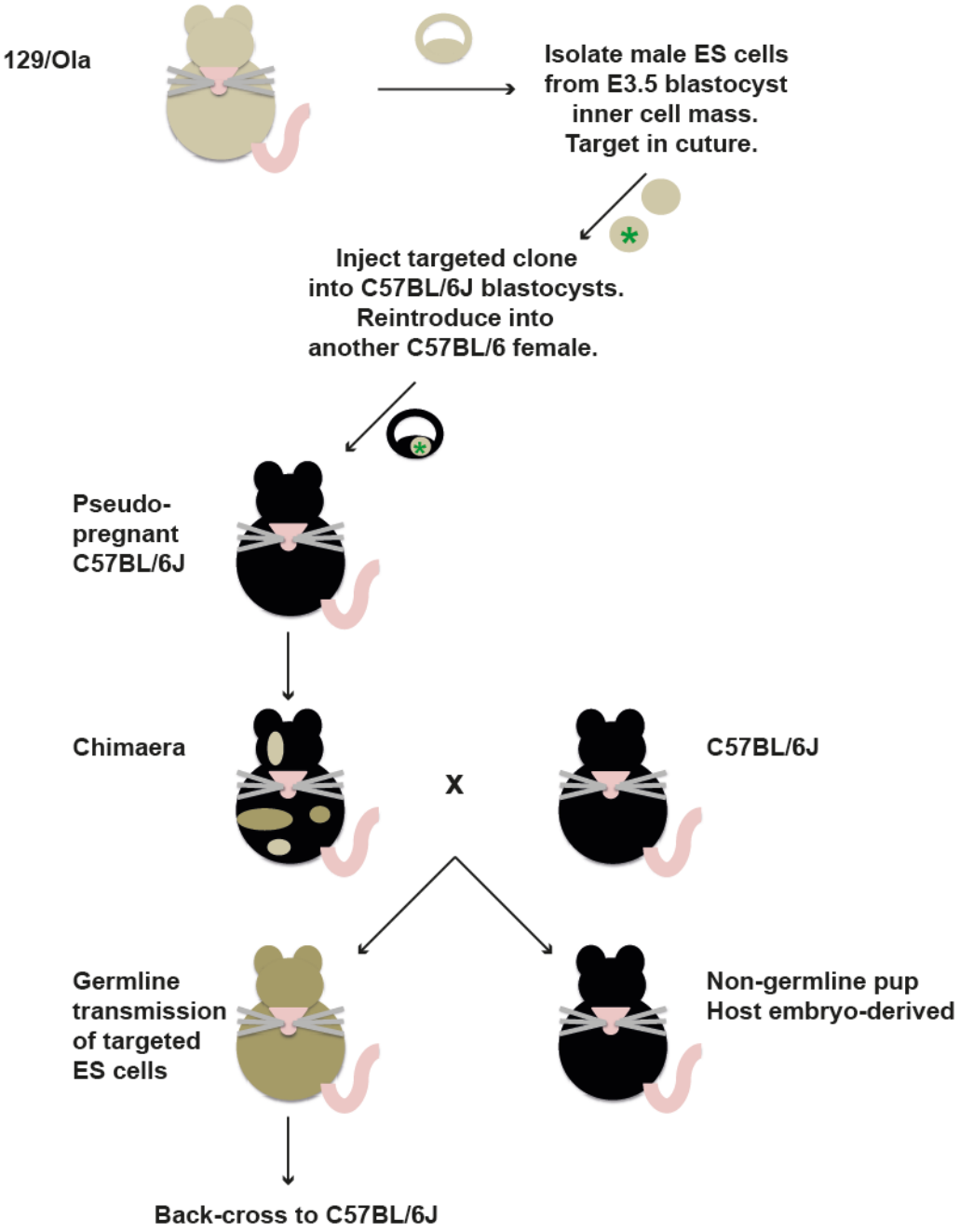


Figure 3.4 Creation of an *R133C-GFP* mouse model.

The injection and breeding strategy for creation of the *R133C-GFP* mouse line is shown. First correctly targeted (green asterisk) male ES cells derived from the 129-Ola mouse strain were injected into 3.5 day old C57BL/6J blastocysts. Blastocysts were immediately re-implanted into pseudo-pregnant females and successful integration into the developing embryo was determined by the birth of chimaeric pups. After one cross with C57BL/6J mice, it was apparent whether there was germ line transmission from chimaeras by the production of agouti pups. Mice were then back-crossed onto a C57/BL6J genetic background for further analysis.

3.5 Deletion of the NEO cassette from the *R133C-GFP* targeted locus

Following the establishment of two ES cell clones correctly targeted with *R133C-GFP*, it was necessary to delete the NEO resistance cassette from the targeted locus prior to neuronal differentiation, in order to avoid any undesirable effect it may have on protein expression or DNA binding. The appearance of the targeted locus following recombination is shown in Figure 3.5A, along with screening strategies.

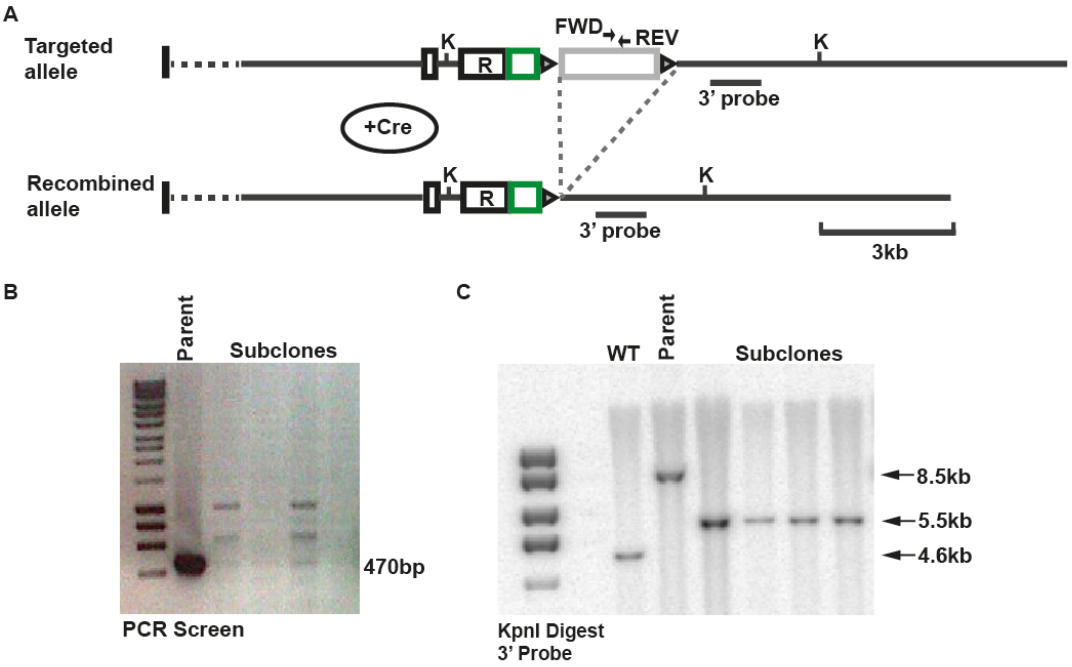


Figure 3.5 Deletion of the NEO Cassette from the *R133C-GFP* Targeted Locus

(A) A map of the targeted locus is shown including exons 2-4 (black boxes), EGFP tag (green box), floxed NEO cassette (grey box flanked by triangles) and R133C mutation (R). PCR primers (arrows), KpnI (K) restriction sites and Southern blot probe used to screen colonies following recombination are represented. The targeted locus following exposure to Cre recombinase (+Cre) is shown with the remaining single LoxP site (triangle).
(B) PCR screening for presence of the NEO cassette revealed a lack of 470bp product in the subclones of *R133C-GFP* ES cells transiently exposed to Cre.
(C) Southern blot screening following KpnI digest showed that the probe bound a 5.5kb fragment in the subclones of *R133C-GFP* ES cells transiently exposed to Cre. Wildtype genomic DNA was probed for reference (WT).

Targeted clones were transiently transfected with Cre plasmid by electroporation in order to excise the cassette. Post electroporation, cells were grown in normal ES cell medium at low density then colonies were picked and immediately split between two

plates, so that new sensitivity to G-418 could be assessed without killing the subclones. Sensitivity to G-418 indicated that the NEO cassette had been excised. Sensitive clones were screened by PCR: primers were within the NEO cassette and did not generate a 470bp product in correctly deleted subclones (Figure 3.5B). Secondly, clones were screened at the 3' end of the locus by Southern blot following KpnI digest as described above. Subclones with the cassette deleted would now have a 5.5kb band using this method (Figure 3.5C). In this fashion, daughter subclones were established that did not carry the NEO resistance cassette. They were denoted by .1, for example A3.1 from A3.

3.6 Differentiation of *R133C-GFP* ES cells into neuronal cells

R133C-GFP ES cells were differentiated into neuronal cells in order to create a suitable model system in which to study the effects of the mutation on protein function. It was possible to differentiate wildtype (*WT*) ES cells and ES cells targeted with the wildtype *Mecp2*^{EGFP} (which will be referred to as *WT-GFP*) in parallel for comparison. In this way, the effect of the EGFP tag and the R133C mutation could be determined. The differentiation strategy is shown in Figure 3.6A. *R133C-GFP* ES cells were cultured in embryoid body medium (which was essentially standard ES cell medium with no leukaemia inhibitory factor and lower fetal bovine serum content) in non-gelatinised dishes. This encouraged the pluripotent cells to form embryoid body aggregates (Figure 3.6A). The embryoid bodies were cultured in the presence of retinoic acid to encourage neuronal differentiation, before being disaggregated and plated in medium optimised for culture of post mitotic neurons in dishes coated with poly-DL-ornithine and laminin. This was gradually changed to neurobasal medium for culture of hippocampal and cortical neurons. This process took two weeks in total and was sensitive to variation in cell density and temperature. The neuronal cells were harvested on day six for further analysis. Day six cells stained positively for neurofilament, a major component of the cytoskeleton in neurons and NeuN, a marker of terminal neuronal differentiation, indicating that the process had been successful (Figure 3.6B). It was therefore possible to differentiate the targeted ES cells in culture into a relevant cell type for the study of MeCP2 function.

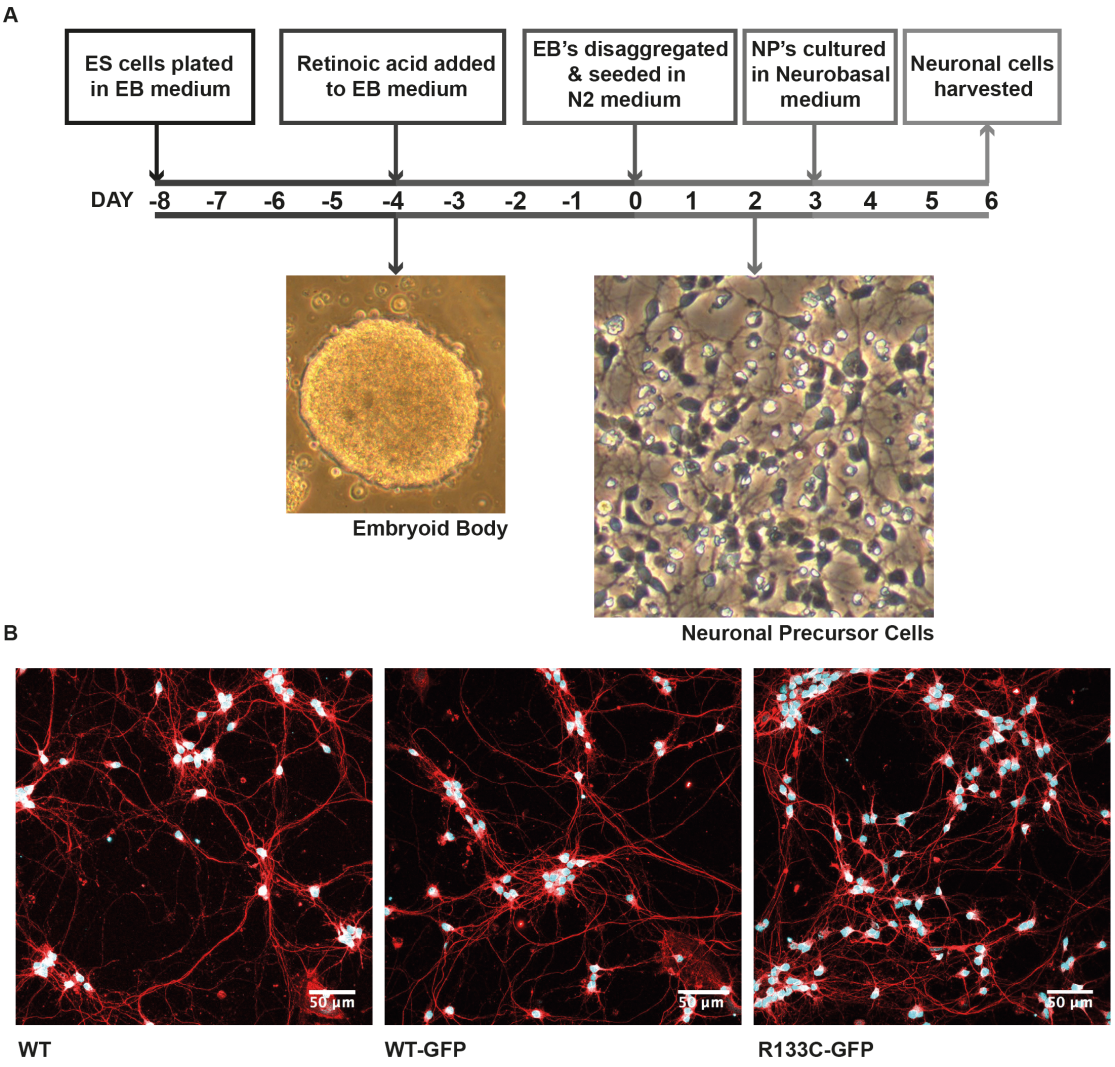


Figure 3.6 Differentiation of *R133C-GFP* ES Cells into Neuronal Cells

(A) Schematic of the differentiation strategy to produce a homogeneous population of neuronal cells from *R133C-GFP* embryonic stem cells. Representative light microscopy images of embryoid bodies and neuronal precursor cells are shown. Neuronal cells were harvested six days after plating. EB, embryoid body. NP, neuronal precursor.
(B) Confocal images of day six neuronal cells derived from *WT*, *WT-GFP* and *R133C-GFP* ES cells are shown. The neuronal cells were stained for neurofilament (red) and NeuN (pale green).

3.7 MeCP2 level increases during neuronal differentiation

MeCP2 expression increases with neuronal differentiation (Kishi *et al*, 2004). In keeping with this, the level of *Mecp2* mRNA increased during the process of differentiation in the ES cell-derived neurons for all genotypes studied (Figure 3.7A).

The relative transcript level in ES cells and neuronal precursors was approximately 15% and 50% of neuronal level, respectively when normalised to *Cyclophilin A* transcript levels. This was thought to be a suitable housekeeper for normalisation because the gene was highly expressed regardless of the stage of neuronal differentiation and the quantitative PCR cycle threshold values were similar and consistent across genotypes, suggesting that the expression of *Cyclophilin A* was not dependent on MeCP2 function or stage of differentiation. The increase in *Mecp2* mRNA transcript with neuronal differentiation was reflected in western blotting of MeCP2 protein from whole cell lysates (Figure 3.7B). For western blotting, Histone 3 was selected as a loading control because protein expression was equal throughout differentiation and between genotypes (see Figure 3.7B). These observations validated the use of the neuronal cells as a model system for the study of MeCP2 function.

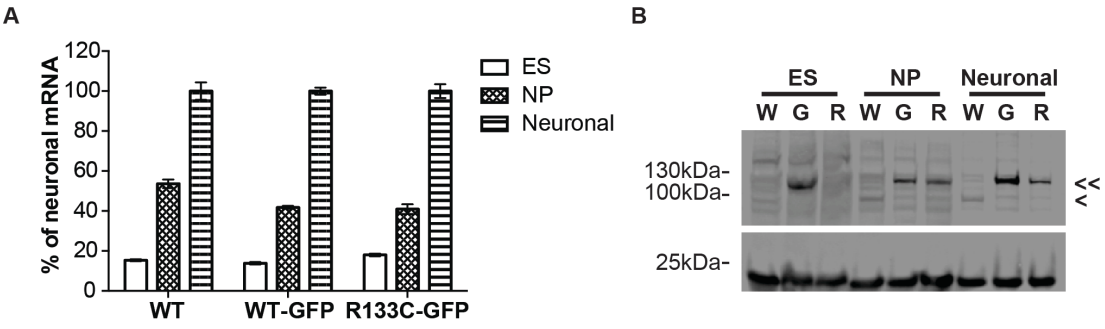


Figure 3.7 MeCP2 level increases during neuronal differentiation

(A) Graph depicting *Mecp2* mRNA levels relative to *Cyclophilin A* through the process of neuronal differentiation for *WT*, *WT-GFP* and *R133C-GFP* cells. For each genotype mRNA level increased in the neuronal cells compared to earlier time points. Representative graph from one differentiation is shown.
(B) Representative western blot of increasing MeCP2 level through the process of neuronal differentiation for wild type (W), *WT-GFP* (G) and *R133C-GFP* (R) cells. ES cells (ES), neuronal precursors (NP) and day six neuronal cells (neuronal) are shown. MeCP2 is indicated by < and $MeCP2^{EGFP}$ by <<. Histone 3 was the loading control.

3.8 Discussion

Previous studies on the impact of $MeCP2^{R133C}$ on DNA binding have not yielded consistent findings. The goal was to create two suitable model systems that were relevant to the pathogenesis of Rett syndrome, with the hypothesis that MeCP2

function would more accurately reflect function *in vivo*. The impact of the mutation could be analysed and this would lead to a greater understanding of its molecular consequences and why in patients, it results in expression of a milder phenotype (Cuddapah *et al*, 2014).

3.8.1 Optimisation of targeting protocol

The successful targeting reported was preceded by an unsuccessful attempt. There was no evidence of homologous recombination and on Southern blot screening it appeared that multiple copies of the targeting vector had been incorporated in concatemers. In order to rectify this, several steps were taken. Firstly, the targeting vector was linearised with SnabI at the 5' end instead of SalI at the 3' end, which had only created 2kb of homology. This resulted in an extra 1.1kb of homology upstream of exon 3, thus increasing the chance of homologous recombination. Secondly, ES cells were targeted when they were less confluent in the dish and therefore dividing rapidly. If the dish became confluent then it was reasoned that the chances of the cells entering a more stationary phase, with increased propensity to differentiate would be higher. Thirdly, the ratio of targeting vector:ES cells was reduced to decrease the likelihood of multiple integrants. Fourthly, the positive selection was modified such that the concentration of G-418 decreased over time. In this way, multiple vector integrants would not be selected for. Fifthly, the electroporator settings were modified to a slightly higher voltage, but lower charge as recommended by Millipore. Finally, the PCR screening was optimised: primers were redesigned and an alternative Taq polymerase was used.

With all these steps taken, the next targeting attempt worked with an overall frequency of 6.25% when all picked colonies were taken into account. Unfortunately when targeted clones identified by PCR were screened by Southern blot and sequencing, evidence of mixed clones was apparent, that is, having both a band corresponding to the wildtype locus and the targeted locus on Southern blot. This reduced the number of *R133C-GFP* clones, but there were two clean ones manufactured: A3 and B11. Both clones were taken forward to differentiate into neuronal cells: clone B11 made the mouse model.

3.8.2 Inconsistencies in neuronal culture

There are a number of advantages to having a homogeneous population of neuronal cells as a model system for studying MeCP2 function: higher levels of MeCP2 expression than in other cell types; endogenous levels of MeCP2 expression from the nascent promoter; the ability to study the pathogenesis of Rett syndrome in the principal tissue affected by the disorder; an ability to detect effects that may be cell type-specific; and the ability to compare different genotypes of neurons that are differentiated in parallel. This is dependent on reproducibility and reliability of the protocol, so that differences between the individual cultures are not inadvertently attributed to differences between the genotypes. In fact, the neuronal cells varied in appearance between different cultures using the same clone. There were multiple steps to the protocol that appeared sensitive to subtle variation. For example, the rate at which the ES cells were growing at the start of the process, the quality of the poly-DL-ornithine and laminin coating of the dishes the neurons were plated in, or variation in the temperature of the incubator (following completion of all neuronal experiments it was discovered that the incubator was nearly two degrees below the desired 37^{o}C). This may have impacted on the reproducibility of the results from experiments using this model and this will be discussed further in Chapter 5.

3.8.3 EGFP-tagging MeCP2

The protein was EGFP-tagged for ease of identification by microscopy and for straightforward immunoprecipitation. There were additional *Mecp2*^{EGFP} mouse lines available to which the *R133C-GFP* line could be compared. In this fashion, genotype-phenotype correlation could be studied. Western blotting of neuronal whole cell lysates gave an early indication that the MeCP2 expression may be higher with the EGFP-tagged protein in comparison to the untagged protein, which was an unanticipated observation. This will be discussed further in Chapter 5 with formal quantification.

3.8.4 Summary

In summary, *R133C-GFP* ES cell lines were manufactured. They could be differentiated into a population of neuronal cells and were used to create a novel

mouse line. These models were then used to study the impact of the mutation on MeCP2 function.

4. Phenotyping the *R133C-GFP* Mouse

4.1 Introduction

The R133C mutation consistently falls at the milder end of the spectrum when genotype is linked to phenotypic severity in patients, independent of patient age or geographical location of the study (Bebbington *et al*, 2008; Cuddapah *et al*, 2014). The significant limitation of these studies is that most groups have not measured the pattern of XCI in individuals participating in the study. In extreme circumstances, if patients with *MECP2*^{R133C} tended to have skewed XCI, where the mutant allele was expressed in less than 20% of cells (and this reflected the XCI pattern in the brain), this could artificially make the phenotype seem less severe. Given that R133 occupies a key position in the MBD of MeCP2 (Ho *et al*, 2008), the intuitive prediction would be that this mutation would be catastrophic to protein function and result in a severe form of Rett syndrome. It therefore becomes imperative to disentangle the effects of the potentially confounding pattern of XCI, in order to understand the molecular basis of the syndrome.

This problem can be circumvented using the *R133C-GFP* mouse model. Mouse models have been used with huge success in the Rett syndrome field (Guy *et al*, 2011). In contrast to humans, male *Mecp2*-null mice survive into adulthood. They display a constellation of signs that recapitulate the human Rett syndrome (Guy *et al*, 2001). First they develop a stiff, uncoordinated, wide-based gait. This is followed by a reduction in mobility and a clasping together of their hindlimbs. Later on, they develop tremors and breathing difficulties before early mortality. Their growth is also affected: null mice are smaller than their wildtype littermates. A well-established scoring system for measuring this phenotype exists (Guy *et al*, 2007; Appendix 1). Using the male mouse as a Rett syndrome model means that one can disregard the pattern of XCI because this does not take place when there is only one copy. MeCP2 function will be directly attributable to the genetically manipulated form. The male mouse therefore serves as an ideal platform to assess the severity of the R133C mutation in a whole organism. Using the scoring system, it is possible to

compare mouse lines and so the *R133C-GFP* line can be compared to other GFP-tagged lines.

Mice can also be phenotyped using behavioural tests. The elevated plus maze is a commonly used test to measure anxiety (Pellow *et al*, 1985). Mice are placed on a cross-shaped maze, elevated above the ground. Two of the maze arms are enclosed by high-sided walls and two are open. Wildtype mice find the unprotected open arms more anxiety provoking and will typically spend most of the trial in the safer closed arms. The open field test also measures anxiety as well as mobility (Prut & Belzung, 2003). Mice are placed in an enclosed field. Wildtype mice typically spend most of their time during the trial against the walls of the enclosure – a phenomenon known as thigmotaxis. This is preferable to venturing in to the centre of the open field, which is more anxiety provoking. The hanging wire test assesses motor strength and coordination (Gomez *et al*, 1997). Mice are suspended from a thin wire above the bench by their forepaws. The time taken for the mouse to bring one hind paw up to grasp the wire is recorded over a number of spaced trials. The accelerating rotarod also assesses motor strength and coordination (Jones & Roberts, 1968). Mice learn to run on a rotating rod during a training session. Over subsequent days, mice run on the rod in separate trials. The latency for the mouse to fall from the rod as the rotation accelerates is recorded. These four tests have previously been used to phenotype mice modelling Rett syndrome mutations (Goffin *et al*, 2011; Lyst *et al*, 2013; Heckman *et al*, 2014). They are suitable because Rett syndrome patients suffer from abnormal muscle tone and mobility, as well as intellectual disability and heightened anxiety (Smeets *et al*, 2011). Therefore it is necessary to establish whether these features are recapitulated in the *R133C-GFP* mouse.

In order to determine whether an entity is mild, it is necessary to have appropriate comparators, which would be judged as more severe. Three other mouse strains serve this purpose: $Mecp2^{EGFP}$ (*WT-GFP*), $Mecp2^{R306CEGFP}$ (*R306C-GFP)* and $Mecp2^{T158MEGFP}$ (*T158M-GFP)*. These lines were created by J.Guy, J.Selfridge & C.Merusi using the same targeting vector that was used to make *R133C-GFP*. The *WT-GFP* strain is simply wildtype *Mecp2* with a C-terminal EGFP tag. The *R306C-*

GFP strain additionally carries a missense mutation in the NID of *Mecp2*. Similar strains have previously been studied without the C-terminal EGFP tag (Lyst *et al*, 2013) and when expressed from a transgene on an *Mecp2*-null background (Heckman *et al*, 2014). The R306C mutation accounts for approximately 5% of Rett syndrome cases (Christodoulou et al, 2003). Patients with this mutation have mild to moderate clinical severity (Colvin *et al*, 2004; Neul *et al*, 2008; Bebbington *et al*, 2008; Cuddapah *et al*, 2014) and this was reflected in mouse phenotyping with an average survival of approximately 18 weeks. The *T158M-GFP* strain carries a missense mutation in the MBD of *Mecp2*. The T158M mutation is the most common Rett-causing mutation, accounting for nearly 9% of cases (Christodoulou *et al*, 2003). Patients with this mutation have moderate to severe clinical severity (Colvin *et al*, 2004; Charman *et al*, 2005; Bebbington *et al*, 2008; Cuddapah *et al*, 2014). This is a novel mouse line, although mice with the T158A mutation have previously been reported and displayed an average survival of 16 weeks (Goffin *et al*, 2011).

The hypothesis was that the *R133C-GFP* mouse would develop a Rett-like phenotype, but that this would be less severe than *R306C-GFP* and *T158M-GFP* mice, thus recapitulating and validating the patient genotype-phenotype correlation. The *R133C-GFP* mouse could then be used to study the molecular basis of Rett syndrome caused by the R133C mutation. The Rett-like phenotype displayed in the mouse will be referred to as "RTT" for brevity. When discussing phenotype in patients, "Rett syndrome" will be used.

4.2 Phenotypic analysis of male mice

It was not known whether the *R133C-GFP* mouse would develop RTT as the patient data suggests R133C would cause a milder syndrome and Rett syndrome appears to be milder in the mouse as males without *Mecp2* survive into adulthood (Guy *et al*, 2001). Mice were weighed and observed for emergence of RTT on a weekly basis at the same time and in the same room where possible. As is customary, mice were compared to their wildtype littermates, in order to minimise any impact that environmental factors might have on phenotypic observations. The observations

were done blind to genotype. The mice were on a 96% C57BL/6J genetic background having been backcrossed for five generations.

4.2.1 Growth

R133C-GFP mice weighed significantly less than their wildtype littermates over the 52-week period of observation, two-way repeated measures ANOVA $F(1,21)=61.49$, $p<0.001$ (Figure 4.2.1A). The absolute weight difference between the cohorts increased from 4.3g at 4 weeks to 11g by 52 weeks. There was an early flattening of the *R133C-GFP* growth curve, indicating an early stagnation in growth, which is also observed in human patients (Kerr & Engerström, 2001).

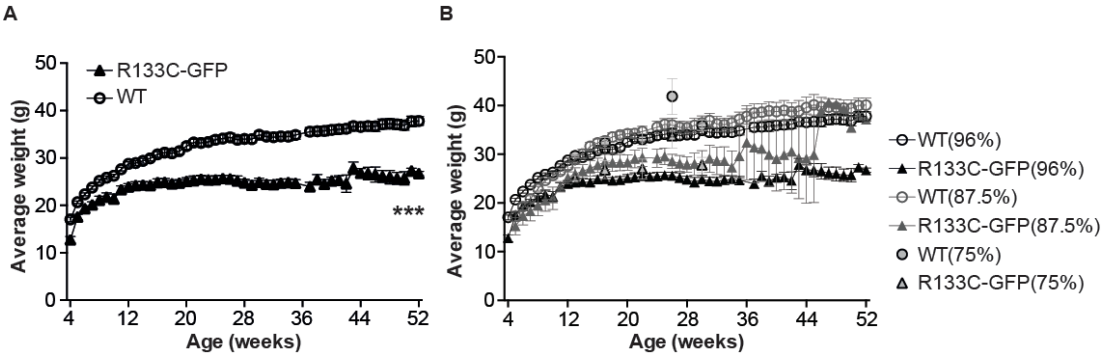


Figure 4.2.1 Reduced weight of *R133C-GFP* mice in comparison to wildtype littermates.

(A) Growth curves showing that *R133C-GFP* mice are significantly lighter than their wildtype (WT) littermates, n=10 & n=13, respectively. Mean±SEM plotted. ***p<0.001, two-way repeated measures ANOVA.
(B) Back-crossing onto a C57BL/6J background reduces the weight of the mice. C57BL/6J contribution is indicated in brackets. WT(87.5%), n=10; *R133C-GFP*(87.5%), n=10; WT(75%), n=5-17; *R133C-GFP*(75%), n=5-9. Means±SEM plotted.

Cohorts of *R133C-GFP* mice at earlier stages of back-crossing onto the C57BL/6J background were also observed: two and three back-crosses, equivalent to approximately 75% and 87.5% C57BL/6J contribution, respectively. The 96% C57BL/6J mice were lighter than 75% and 87.5% cohorts, indicating that the higher the contribution of C57BL/6J to the genetic makeup, the lighter the mouse. The effect was seen independently of *Mecp2* genotype (Figure 4.2.1B). As seen in Figure 4.2.1B, the variability in weight within the groups was greater for the 75% and 87.5% cohorts. Mice were humanely culled if they lost 20% of their body weight.

4.2.2 Phenotypic scoring

An established scoring system for mouse RTT exists (see Appendix 1). Mobility, gait, hindlimb clasping, tremor, breathing and general condition were observed and given a score from 0 (no symptom) to 2 (significant symptom) for each aspect with a maximum score of 12. Half scores were utilised as introduced previously (Cheval *et al*, 2012), given that the anticipated phenotype was subtle and the scoring system had been developed using mice null for *Mecp2*. Mice were humanely culled if they scored 2 for tremor, breathing or general condition.

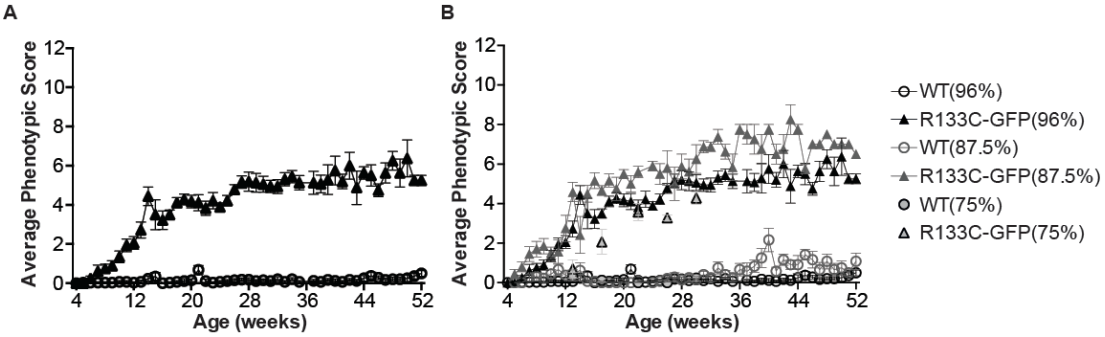


Figure 4.2.2 *R133C-GFP* mice develop RTT at 10-12 weeks of age.

(A) *R133C-GFP* mice (n=10) develop progressive RTT at 10-12 weeks of age. Wildtype (WT, n=13) almost invariably score 0. Mean score±SEM plotted.
(B) There is no consistent effect of back-crossing onto a C57BL/6J background on phenotypic score. C57BL/6J contribution is indicated in brackets. WT(87.5%), n=10; *R133C-GFP*(87.5%), n=10; WT(75%), n=5-17; *R133C-GFP*(75%), n=5-9.

R133C-GFP mice began to develop RTT between 10 and 12 weeks of age (Figure 4.2.2A). Symptoms worsened with time, reaching a maximum mean score of 6.4 at 50 weeks. Their wildtype littermates rarely displayed any symptoms that would be compatible with the disorder. *R133C-GFP* mice did not manifest any unusual or unanticipated symptoms that would indicate a unique phenotype associated with the mutation. Conversely, they did not have preservation of any particular measured faculty, although it was rare to observe breathing irregularity until a very late stage. The degree of backcross onto the C57BL/6J background did not have a consistent impact on the mean phenotypic score (Figure 4.2.2B). The *R133C-GFP* mice on the 96% C57BL/6J background scored intermediately between those on the 75% and those on the 87.5% backgrounds.

4.2.3 Survival

The *R133C-GFP* mice had a significantly reduced lifespan in comparison to their wildtype littermates (Figure 4.2.3A), with a median survival of 42 weeks. Mice were humanely culled when they reached the end point of the experiment. This corresponded to severe RTT that exceeded the severity limit of the Home Office License as detailed above. The reduced survival is analogous to the increased mortality seen in Rett patients (Freilinger *et al,* 2010). Survival of the wildtype mice was undetermined as only two mice died during the observation period of 52 weeks. These mice were humanely culled following injuries obtained during in-cage fighting. Both of these mice were from the same cage and were littermates. This in-cage fighting was prominent and problematic feature of the earlier generation of *R133C-GFP* mice that had 87.5% C57BL/6J contribution. Several mice were culled because of fighting injuries (four wildtype mice and four *R133C-GFP* mice), to the extent that it was considered whether the mutation produced an aggression phenotype (Figure 4.2.3B). The male mice had been pooled into larger cages following weaning and so a decision was taken to keep later generations in cages with only their littermates. This, along with the extra contribution of C57BL/6J, seemed to eradicate the phenomenon. The survival of the more outbred *R133C-GFP* mice was also significantly reduced, relative to wildtype, despite the confounding factor of the aggression (Figure 4.2.3B). There was no significant difference in the survival of the *R133C-GFP* cohorts. The median survival for the *R133C-GFP* mice with 87.5% C57BL/6J contribution was 24 weeks.

Given that the *R133C-GFP* mice displayed reduced body weight, an early stagnation in growth, reduced survival and RTT, it was considered that the novel mouse line was a suitable model for Rett syndrome, resulting from the R133C mutation.

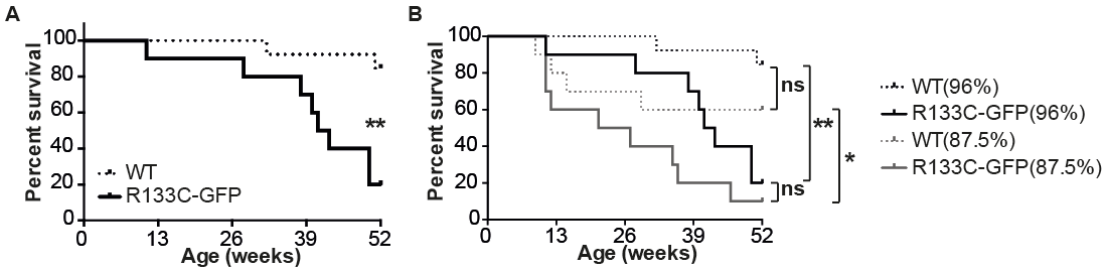


Figure 4.2.3 Reduced survival of *R133C-GFP* mice

(A) *R133C-GFP* mice (n=10) have significantly reduced survival compared to wildtype littermates (WT, n=13).
(B) Back-crossing onto a C57BL/6J background does not significantly alter median survival in wildtype or *R133C-GFP* mice. Contribution of C57BL/6J is indicated in brackets. WT(87.5%), n=10. *R133C-GFP*(87.5%), n=10. WT only comparison p=0.13, *R133C-GFP* only comparison p=0.08.
*p<0.05, **p<0.01, Log-rank Mantel Cox Test.

4.3 Behavioural analysis of 9-10 week male mice

Mice were further characterised using a battery of behavioural tests in order to more precisely describe their phenotype. Mice underwent the elevated plus maze, hanging wire test, open field test and accelerating rotarod. Behavioural testing was done at two time points: 9-10 weeks and 20-21 weeks (see 4.4) to represent pre-symptomatic and symptomatic stages. Separate cohorts of *R133C-GFP* mice and their wildtype littermates were used.

4.3.1 Elevated plus maze

First, the mice underwent the elevated plus maze to assess their level of anxiety. The trial was 15 minutes. As expected, the wildtype mice spent significantly more time in the closed arms of the maze than the open arms (643 seconds vs 91 seconds, Figure 4.3.1A). This was also true of the *R133C-GFP* mice (459 seconds vs 208 seconds) although the difference between the mean times was smaller (552 seconds for wildtype vs 250 seconds for *R133C-GFP* mice). Compatible with this observation, the *R133C-GFP* mice spent significantly more time in the open arms than their wildtype littermates (208 seconds vs 91 seconds, respectively) and significantly less time in the closed arms (459 seconds vs 643 seconds, respectively). This was not a result of the *R133C-GFP* mice moving to an open arm and remaining there for the

duration of the test, as there was no significant difference in the number of times the *R133C-GFP* mice entered the closed arms compared to the wildtype mice (45±4 vs 36±3, respectively, Figure 4.3.1B).

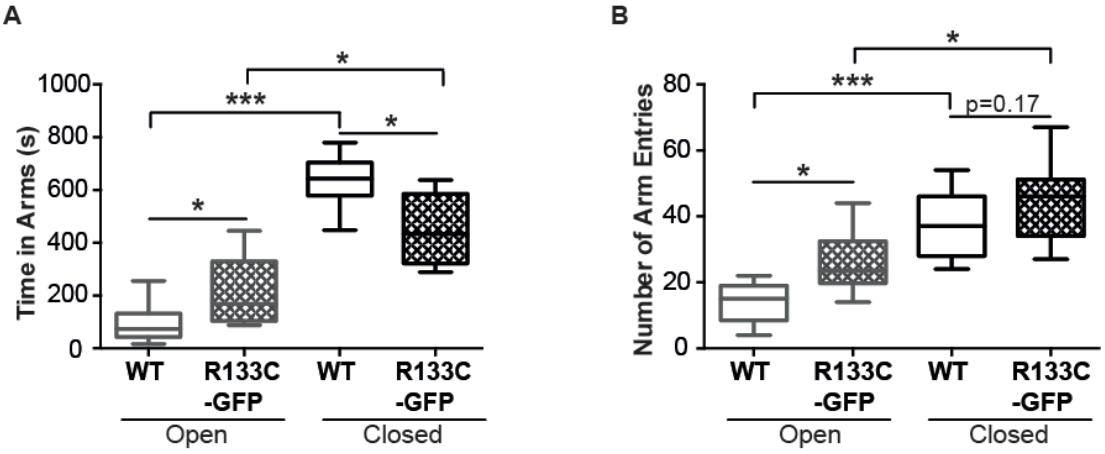


Figure 4.3.1 *R133C-GFP* mice display a mild reduction in anxiety on the elevated plus maze at 9 weeks.

(A) Box plots of individual animal times showing that *R133C-GFP* mice (n=10) spent significantly more time in the open arms than their wildtype littermates (WT, n=13), though still showed a preference for closed arms.
(B) Box plots of individual animal entries showing that the *R133C-GFP* mice entered the closed arms as many times as the wildtype mice.
Cohort median, interquartile range, maximum and minimum represented. *p<0.05, ***p<0.001, Kolmogorov-Smirnov Test.

In summary, the *R133C-GFP* mice spent significantly more time in the anxiety-provoking arms of the maze than their wildtype counterparts, although they still displayed a preference for the closed arms. This would indicate a marginally lower level of anxiety in the *R133C-GFP* mice.

4.3.2 Hanging wire test

Next, motor strength and coordination were assessed using the hanging wire test (Figure 4.3.2). The maximum time allowed per trial was 30 seconds. The *R133C-GFP* mice took significantly longer than their wildtype littermates – almost double the time - to bring one hind paw up to grasp the wire when suspended over three trials (19 seconds vs 10 seconds, respectively). This indicated an impairment of motor function.

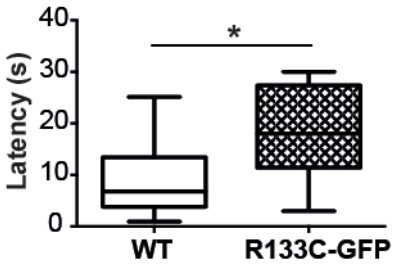


Figure 4.3.2 *R133C-GFP* mice show impaired motor strength and coordination in the hanging wire test at 9 weeks.

Box plots of mean individual latencies over three 30-second trials show that *R133C-GFP* mice (n=10) took significantly longer than wildtype (WT, n=13) mice to bring one hind paw to grasp the wire. Cohort median, interquartile range, maximum and minimum represented. *p<0.05, Kolmogorov-Smirnov Test.

4.3.3 Open field test

Thirdly, mice did the open field test to assess anxiety, as well as mobility in a single 20-minute trial (Figure 4.3.3). The *R133C-GFP* mice spent significantly more time than their wildtype littermates in the anxiety-provoking central area of the open field (42 seconds vs 23 seconds, respectively, Figure 4.3.3D). This was not the result of moving into the centre and staying there for the duration of the test, as there was no significant difference in the overall mean distance travelled (93m vs 93m), mean speed ($0.077ms^{-1}$ vs $0.077ms^{-1}$) or time spent immobile (11 seconds vs 12 seconds) during the test for the *R133C-GFP* mice compared to their wildtype littermates (Figure 4.3.3A-C). The *R133C-GFP* mice seemed less perturbed by or less aware of the anxiety-provoking central area.

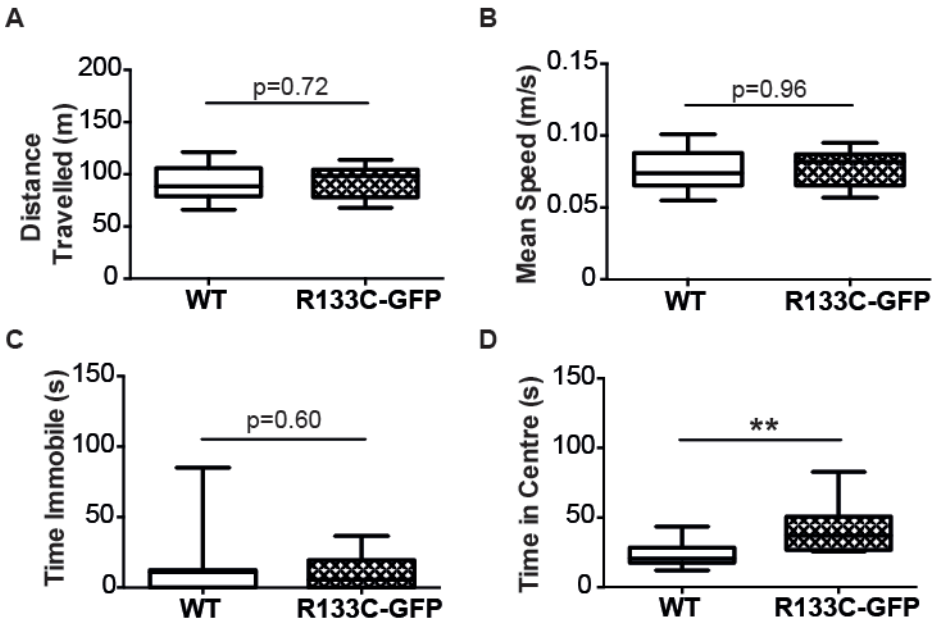


Figure 4.3.3 *R133C-GFP* mice display reduced anxiety in the open field test at 9 weeks.

(A) There was no significant difference in the distance travelled by the *R133C-GFP* (n=10) mice in comparison to their wildtype littermates (WT, n=13).
(B) There was no significant difference in the mean speed of the *R133C-GFP* mice in comparison to their wildtype littermates.
(C) The *R133C-GFP* mice did not spend significantly more time immobile in comparison to their wildtype littermates.
(D) The *R133C-GFP* mice spent significantly more time in the centre of the field in comparison to wildtype littermates.
Box plots of individual times. Cohort median, interquartile range, maximum and minimum represented. **p<0.01, Kolmogorov-Smirnov Test.

4.3.4 The accelerating rotarod

Finally, mice underwent the accelerating rotarod, another test of motor strength and coordination (Figure 4.3.4). Mice undertook four trials per day over three days: each trial was a maximum of five minutes. The *R133C-GFP* mice performed as well as their wildtype littermates on this test, across all three days (repeated measures two-way ANOVA $F(1,21)=0.21$, $p=0.65$). On day one it took wildtype mice 56 seconds to fall from the rod and *R133C-GFP* mice 41 seconds on average. On day two 70 seconds and 69 seconds, respectively and on day three, 77 seconds and 78 seconds, respectively. There was a significant difference in mean latency to fall over the different days of the test for both the genotypes as determined by repeated measures one-way ANOVA: wildtype animals ($F(2,12)=10.76$, $p<0.001$); *R133C-GFP* animals

(F(2,9)=18.95, p<0.001). A Tukey post-hoc test revealed a significantly higher mean latency to fall for wildtype animals on day two (+14.6 seconds 95%CI 2.8-26.3 seconds, p<0.05) and day three (+21.4 seconds 95%CI 9.6-33.1 seconds, p<0.001) versus day one. Similarly there was a significantly higher mean latency to fall for *R133C-GFP* animals on day two (+27.2 seconds 95%CI 11.5-42.8 seconds, p<0.001) and day three (+36.3 seconds 95%CI 20.6-51.9 seconds, p<0.001) versus day one. Both genotypes therefore showed evidence of motor learning with significant improvement across the three days of testing.

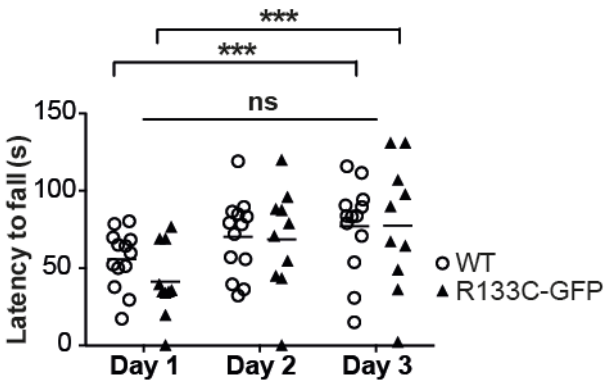


Figure 4.3.4 No significant motor impairment of 10-week *R133C-GFP* mice was detected on the accelerating rotarod.

Scatterplot of accelerating rotarod performance across days. Individual means are plotted with the bar representing cohort mean. The *R133C-GFP* mice (n=10) performed as well as their wildtype littermates (WT, n=13) on all days of the test (two-way repeated measures ANOVA, p=0.65). Both genotypes showed evidence of motor learning across days. ***p<0.001, one-way repeated measures ANOVA.

In summary, at 9-10 weeks the *R133C-GFP* mice displayed signs of reduced anxiety level on two behavioural tests and motor impairment on one behavioural test.

4.4 Behavioural analysis of 20-21 week male mice

The battery of tests was repeated at 20-21 weeks, when the *R133C-GFP* mice were obviously symptomatic with RTT (average score 4.1) to gain an appreciation for the evolution of the phenotype over time.

4.4.1 Elevated plus maze

Again, the wildtype mice showed a significant preference for the closed arms over the open arms of the elevated plus maze (652 seconds vs 89 seconds, Figure 4.4.1A). However, at this symptomatic time point, the *R133C-GFP* mice no longer showed a significant preference for the closed arms over the open arms, spending 414 seconds in the closed arms and 329 seconds in the open arms, on average. They spent significantly more time than the wildtype mice in the open arms (329 seconds vs 89 seconds respectively) and significantly less time than the wildtype mice in the closed arms (414 seconds vs 652 seconds, respectively). Again, the *R133C-GFP* mice entered the closed arms as many times as the wildtype mice (30±4 vs 31±5 times) and so spending more time in the open arms was informed after sampling the whole area of the apparatus (Figure 4.4.1B). Therefore at a symptomatic stage of RTT, the *R133C-GFP* mice show evidence of reduced levels of anxiety, or awareness of potential danger and no longer show a significant preference for the safer, closed arms of the maze.

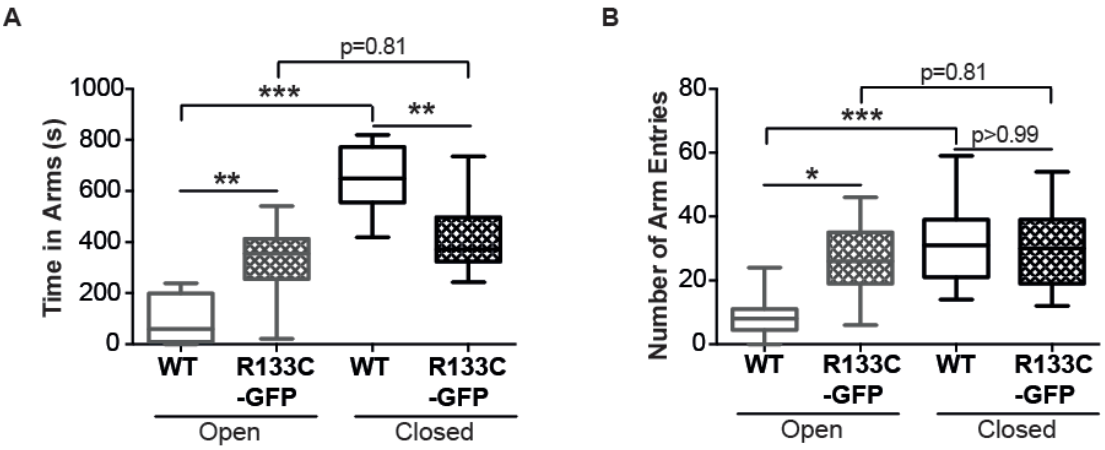


Figure 4.4.1 *R133C-GFP* 20-week mice display reduced anxiety in the elevated plus maze.

(A) *R133C-GFP* (n=11) mice spent significantly more time in the open arms than their wildtype littermates (WT, n=9) and showed no significant preference for the closed arms. (B) The *R133C-GFP* mice entered the closed arms as many times as the wildtype mice. Box plots representing individual animal times or entries with cohort median, interquartile range, maximum and minimum represented. *p<0.05, **p<0.01, ***p<0.001, Kolmogorov-Smirnov Test.

4.4.2 The hanging wire test

At 20-21 weeks, the *R133C-GFP* mice took significantly longer to bring their hind paw up to clasp the wire in comparison to their wildtype littermates on the hanging wire test (27 seconds vs 10 seconds, respectively, Figure 4.4.2). There was deterioration in motor strength and coordination as the difference between the mean latencies had increased to 17 seconds from 9.2 seconds at 9-10 weeks.

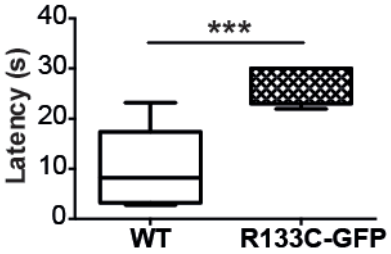


Figure 4.4.2 *R133C-GFP* mice show impaired motor coordination and strength on the hanging wire test at 20 weeks.

The *R133C-GFP* mice (n=11) took significantly longer to bring one hind paw up to the grasp the wire than their wildtype littermates (WT, n=9). Box plot of individual latencies with cohort median, interquartile range, maximum and minimum represented. ***p<0.001, Kolmogorov-Smirnov Test.

4.4.3 The open field test

In contrast to performance at 9-10 weeks, at 20-21 weeks the *R133C-GFP* mice did not spend significantly longer in the centre of the open field in comparison to wildtype littermates, although there was a trend towards this (74 seconds vs 44 seconds, respectively, Figure 4.4.3D). Again, there was no significant difference found between the *R133C-GFP* mice and their wildtype littermates in the mean distance travelled (83m vs 66m), the mean speed ($0.069ms^{-1}$ vs $0.055ms^{-1}$) or the time spent immobile (65 seconds vs 19 seconds) during the test (Figure 4.4.3A-C).

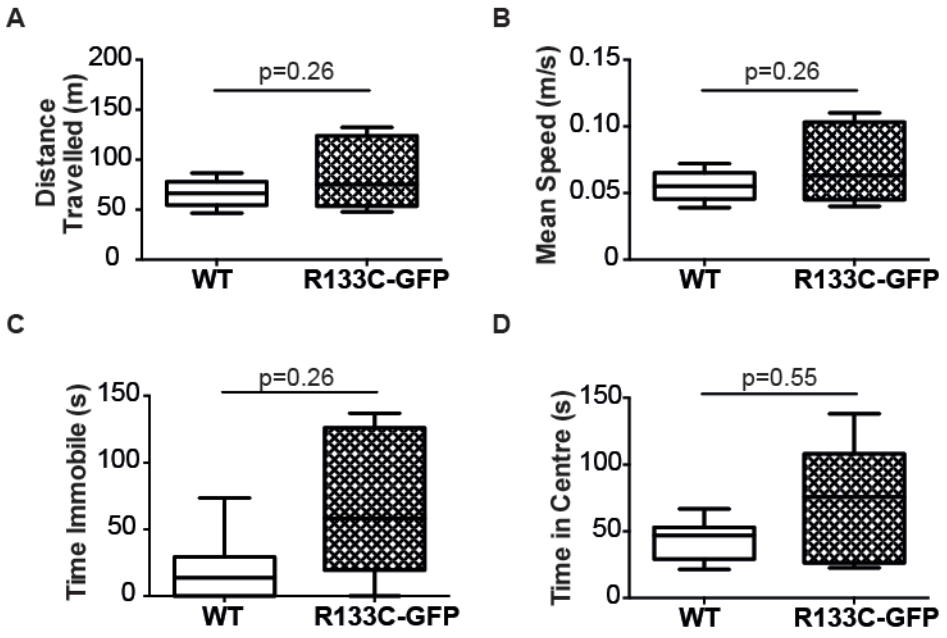


Figure 4.4.3 *R133C-GFP* mice show a trend towards reduced anxiety in the open field at 20 weeks.

(A) There was no significant difference in the distance travelled by the *R133C-GFP* (n=11) mice in comparison to their wildtype littermates (WT, n=9).
(B) There was no significant difference in the mean speed of the *R133C-GFP* mice in comparison to their wildtype littermates.
(C) The *R133C-GFP* mice did not spend significantly more time immobile in comparison to their wildtype littermates.
(D) There was a trend towards the *R133C-GFP* mice spending more time in the centre of the field in comparison to wildtype littermates.
Box plots of individual times. Cohort median, interquartile range, maximum and minimum represented. Kolmogorov-Smirnov Test.

There was a trend towards the *R133C-GFP* mice spending more time immobile, which did not meet significance. On further inspection of individual's performances the mice that spent the most time immobile were not the same individual animals that spent the longest times in the centre. Therefore the trend towards *R133C-GFP* mice spending more time in the centre cannot be explained by mice entering the centre and remaining immobile there for the duration. Although there was no significant difference in how the two genotypes performed in the open field, there was a trend towards the *R133C-GFP* mice spending more time in the centre, which would be compatible with a reduced level of anxiety or awareness of danger.

4.4.4 The accelerating rotarod.

Finally the 20-21 week mice underwent the accelerating rotarod test as a second motor paradigm (Figure 4.4.4). Again, there was no significant difference in the performance of the *R133C-GFP* mice compared to their wildtype littermates on this test, across all three days (repeated measures two-way ANOVA $F(1,18)=0.14$, $p=0.72$) although there was a trend towards the *R133C-GFP* mice falling off earlier. On day one it took wildtype mice 46 seconds to fall from the rod and *R133C-GFP* mice 55 seconds; on day two, 75 seconds and 68 seconds, respectively; and on day three, 93 seconds and 78 seconds, respectively. Again, there was a significant difference in mean latency to fall for the different days of the test for both the genotypes as determined by repeated measures one-way ANOVA: wildtype ($F(2,8) = 19.57$, $p<0.001$); *R133C-GFP* ($F(2,10) = 5.76$, $p<0.05$). A Tukey post-hoc test revealed a significantly higher mean latency to fall for wildtype animals on day two (+29.1 seconds 95%CI 9.5-48.8 seconds, $p<0.01$) and day three (+47.2 seconds 95%CI 27.5-66.8 seconds, $p<0.001$) versus day one. There was no significant increase in the mean latency to fall for *R133C-GFP* animals on day two (+13.0 seconds 95%CI -4.3-30.3 seconds) versus day one, however there was for day three (+23.1 seconds 95%CI 5.8-40.4 seconds, $p<0.01$) versus day one. Both genotypes therefore showed evidence of motor learning with significant improvement across the three days of testing.

Together, the behavioural analysis of the *R133C-GFP* males both at pre-symptomatic and symptomatic stages would suggest that the animals are less anxious (or cautious) than wildtype littermates and display a mild level of motor impairment that worsens with onset and progression of the disease.

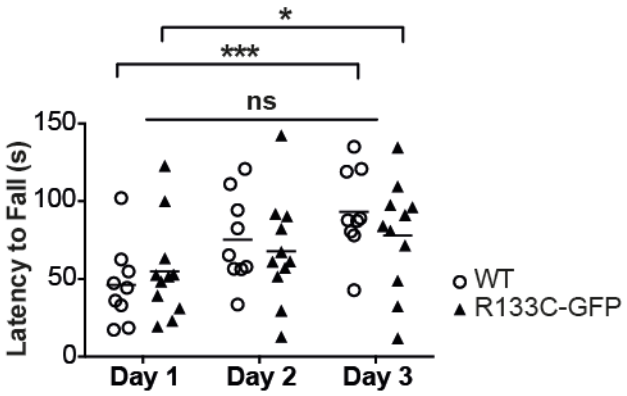


Figure 4.4.4 No significant motor impairment of *R133C-GFP* mice was detected on the accelerating rotarod at 21 weeks.

Scatterplot of accelerating rotarod performance across days. Individual mean latencies are plotted with the bar representing cohort mean. The *R133C-GFP* mice (n=11) performed as well as their wildtype littermates (WT, n=9) on all days of the test (repeated measures two-way ANOVA, p=0.72) and both genotypes showed evidence of motor learning across time *p<0.05, ***p<0.001, repeated measures one-way ANOVA.

4.5 Phenotypic analysis of female mice

Rett syndrome primarily affects girls, as boys carrying mutations in *MECP2* typically die in early infancy (Schüle *et al,* 2008). It follows that the true model of Rett syndrome would be the heterozygous female mouse. *Mecp2*$^{+/R133CEGFP}$ females (which will be referred to as *+/R133C-GFP*) were weighed and observed for the development of RTT on a monthly basis, under the same conditions as the male mice described above (Figure 4.5). The larger interval between observations was determined on the premise that the phenotypic development would be very gradual and there would be little change on a weekly basis. The observed females were 96% C57BL/6J.

There was no significant difference in the weight of the *+/R133C-GFP* females compared to wildtype littermates at any time point across the 50-week observation period (Figure 4.5A). The *+/R133C-GFP* females began to develop RTT at around 11 months, with a mean phenotypic score of 2.9 by 50 weeks, which was the last point of observation (Figure 4.5B). None of the *+/R133C-GFP* females died during the observation period (Figure 4.5C) however one wildtype mouse was humanely

culled for health reasons unrelated to RTT. Therefore the *+/R133C-GFP* females displayed very mild RTT, with normal weight and survival.

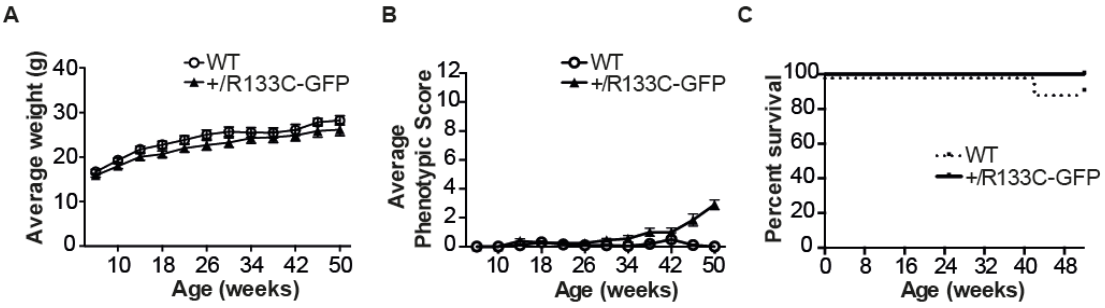


Figure 4.5 Phenotypic analysis of female *+/R133C-GFP* mice reveals mild RTT by 46 weeks.

(A) There was no significant difference in the mean weight of the *+/R133C-GFP* mice (n=10) compared to their wildtype littermates (WT, n=10), p=0.07, repeated measures two-way ANOVA. Mean±SEM plotted
(B) The *+/R133C-GFP* mice began to develop mild RTT at 11 months. Mean±SEM plotted.
(C) The survival of the *+/R133C-GFP* mice was indistinguishable from their wildtype littermates, p=0.32, Log-rank Mantel Cox test.

4.6 Behavioural analysis of female mice

A cohort of *+/R133C-GFP* females underwent the same battery of behavioural tests as male mice. It was decided to test females at one year of age, given the late-onset phenotype. Behavioural testing was done in females that had been back-crossed three times on to a C57BL/6J background: some of these females were still Cre positive following directed breeding to excise the NEO cassette that was introduced by the targeted ES cells.

4.6.1 The elevated plus maze.

Anxiety level was assessed on the elevated plus maze (Figure 4.6.1). Both the *+/R133C-GFP* females and their wildtype littermates spent significantly more time in the closed arms than the open arms of the maze: *+/R133C-GFP* (628 seconds vs 127 seconds); wildtype (629 seconds vs 128 seconds). There was no significant difference in the mean time spent in the open arms between the two genotypes (Figure 4.6.1A). There was a trend towards the *+/R133C-GFP* females entering the open and the closed arms more times than their wildtype littermates: open (15±6 vs 5±2); closed (20±5 vs 7±1), though this did not reach significance (Figure 4.6.1B).

Therefore the *+/R133C-GFP* performance was indistinguishable from wildtype on this test of anxiety.

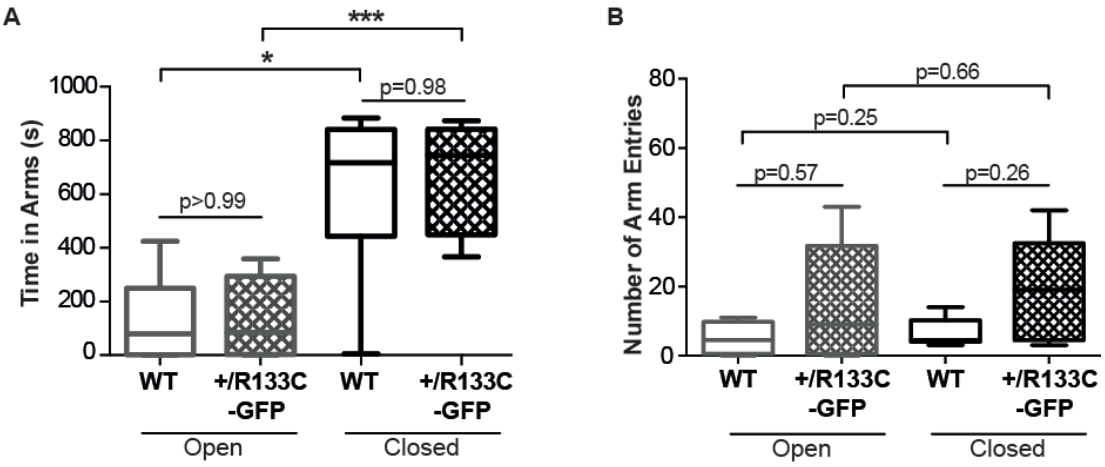


Figure 4.6.1 Equivalent performance of +/*R133C-GFP* mice and their wild type littermates on the elevated plus maze at 1 year.

(A) Both *+/R133C-GFP* mice (n=8) and their wildtype littermates (WT, n=8) spent significantly less time in the open arms than the closed arms of the maze.
(B) There was no significant difference in the mean number of times the *+/R133C-GFP* mice entered the open or the closed arms when compared to wildtype littermates.
Box plots of individual times or entries with cohort median, interquartile range, maximum and minimum represented. Kolmogorov-Smirnov Test.

4.6.2 The hanging wire test.

Next, the mice underwent the hanging wire test to look at motor coordination and strength. There was no significant difference in the average latency to bring one hind paw up to clasp the wire between the two genotypes: *+/R133C-GFP,* 19 seconds; wildtype, 21 seconds (Figure 4.6.2) and so there was no discernible difference in motor ability.

4.6.3 The open field test

The open field test did not reveal any significant differences in performance between the two groups (Figure 4.6.3), with both spending comparable time in the centre of the field (wildtype, 36 seconds; *+/R133C-GFP,* 45 seconds). This result echoed that of the elevated plus maze in finding no detectible difference in anxiety level between the *+/R133C-GFP* and wildtype mice.

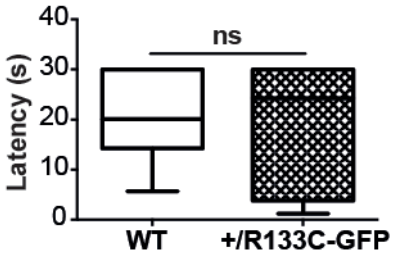


Figure 4.6.2 Equivalent performance on the hanging wire test for *+/R133C-GFP* mice and their wildtype littermates at 1 year.

There was no significant difference in the mean latency to bring one hind paw to the wire between the genotypes. *+/R133C-GFP* mice (n=8); wildtype (WT, n=8). Box plots of individual means with cohort median, interquartile range, maximum and minimum represented. Kolmogorov-Smirnov Test, p=0.95.

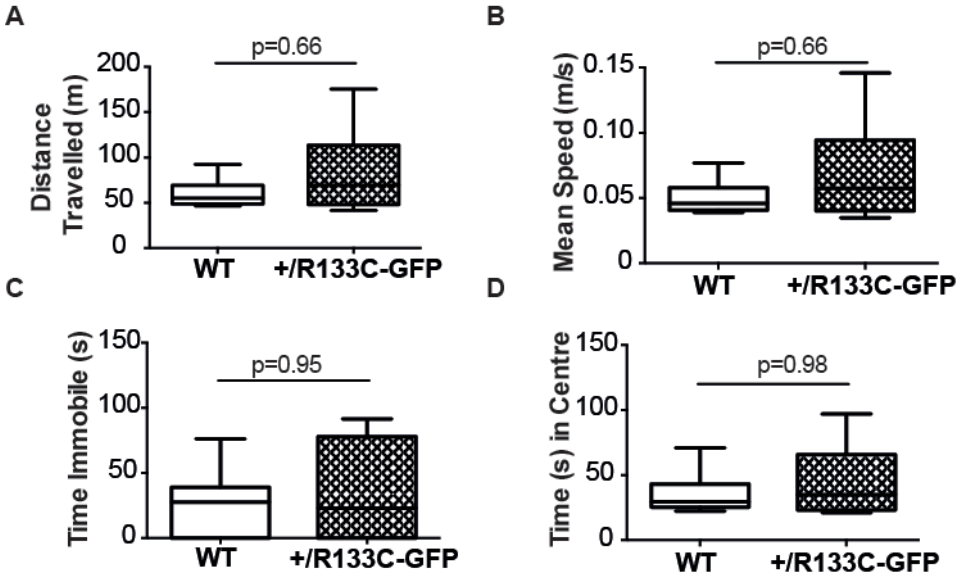


Figure 4.6.3 Equivalent performance in the open field test for *+/R133C-GFP* mice and their wildtype littermates at 1 year.

(A) There was no significant difference in the distance travelled by the *+/R133C-GFP* mice (n=8) in comparison to their wildtype littermates (WT, n=8).
(B) There was no significant difference in the mean speed of the *+/R133C-GFP* mice in comparison to their wildtype littermates.
(C) The *+/R133C-GFP* mice did not spend significantly more time immobile in comparison to their wildtype littermates.
(D) There was no significant difference in the time the *+/R133C-GFP* mice spent in the centre of the field in comparison to wildtype littermates.
Box plots of individual times. Cohort median, interquartile range, maximum and minimum represented. Kolmogorov-Smirnov Test.

4.6.4 The accelerating rotarod.

Finally, the females undertook the second motor test, the accelerating rotarod (Figure 4.6.4). The *+/R133C-GFP* mice performed as well as their wildtype littermates on this test, on all three days (repeated measures two-way ANOVA $F(1,14)=0.02$, $p=0.90$). On day one it took wildtype mice 26 seconds to fall from the rod and *R133C-GFP* mice 21 seconds; on day two, 43 seconds and 39 seconds, respectively; and on day three, 49 seconds and 55 seconds, respectively. Again, there was a significant difference in mean latency to fall across days of testing for both the genotypes as determined by repeated measures one-way ANOVA: wildtype ($F(2,7) = 10.64$, $p<0.01$); *+/R133C-GFP* ($F(2,7) = 18.46$, $p<0.001$). A Tukey post-hoc test revealed a significantly higher mean latency to fall for wildtype animals on day two (+17.0 seconds 95%CI 3.6-30.4 seconds, $p<0.05$) and day three (+22.7 seconds 95%CI 9.3-36.1 seconds, $p<0.01$) versus day one. There was also a significant increase in the mean latency to fall for *+/R133C-GFP* animals on day two (+18.0 seconds 95%CI 3.3-32.8 seconds, $p<0.05$) and day three (+34.2 seconds 95%CI 19.4-48.9 seconds, $p<0.001$) versus day one. Both genotypes therefore showed evidence of motor learning with significant improvement across the three days of testing.

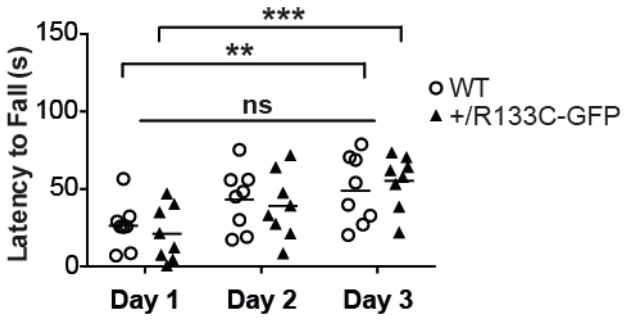


Figure 4.6.4 Equivalent performance on the accelerating rotarod for *+/R133C-GFP* mice and their wildtype littermates at 1 year.

Scatterplot of accelerating rotarod performance across days. Individual mean latencies to fall are plotted with the bar representing cohort mean. The *+/R133C-GFP* mice (n=8) performed as well as their wildtype littermates (WT, n=8) on all days of the test (repeated measures two-way ANOVA, $p=0.90$) and both genotypes showed evidence of motor learning across time **$p<0.01$, ***$p<0.001$, repeated measures one-way ANOVA.

To conclude, the *+/R133C-GFP* mice do not reveal signs of altered anxiety level or motor impairment on behavioural testing at one year of age.

4.7 Comparison with the EGFP-tagged allelic series

In order to answer the question of whether the R133C mutation results in a milder Rett phenotype the severity was placed in the context of *WT-GFP*, *R306C-GFP*, *T158M-GFP* and *Mecp2*-null phenotypic data that was gathered by other experimenters (J.Guy, J.Selfridge and D.De Souza). According to patient studies, the R306C and T158M mutations both result in Rett syndrome that is phenotypically more severe than that caused by the R133C mutation (Colvin *et al*, 2004; Bebbington *et al*, 2008; Cuddapah *et al*, 2014). The EGFP-tagged lines together comprised an "allelic series". All comparator lines were on a 94% C57BL/6J background, equivalent to four backcrosses. The hypothesis was that when data was collated, male mice from these lines would fall in order, from least to most severe: *WT-GFP*, *R133C-GFP*, *R306C-GFP*, *T158M-GFP* then *Mecp2*-null.

4.7.1 Phenotypic analysis

Figure 4.7.1 depicts the collated weekly weights (A) and phenotypic scores (B) of all the aforementioned mouse lines. Overall survival is shown in (C). The animals fell into the expected pattern of severity in terms of weight, score and survival. The *WT-GFP* mice weighed the most, then *R133C-GFP* and *R306C-GFP* fell together, followed by *T158M-GFP*, whose growth curve was similar to the *Mecp2*-null mice. The average phenotypic score at week ten was 0 for *WT-GFP*, 1.4 for *R133C-GFP*, 2.8 for *R306C-GFP*, 5.4 for *T158M-GFP* and 6.8 for the *Mecp2*-null mice. The median survival was undefined for *WT-GFP* (as no mice died during the observation period), 42 weeks for *R133C-GFP*, 30 weeks for *R306C-GFP*, 13 weeks for *T158M-GFP* and 9 weeks for the *Mecp2*-null mice. This data was in accordance with the patient studies.

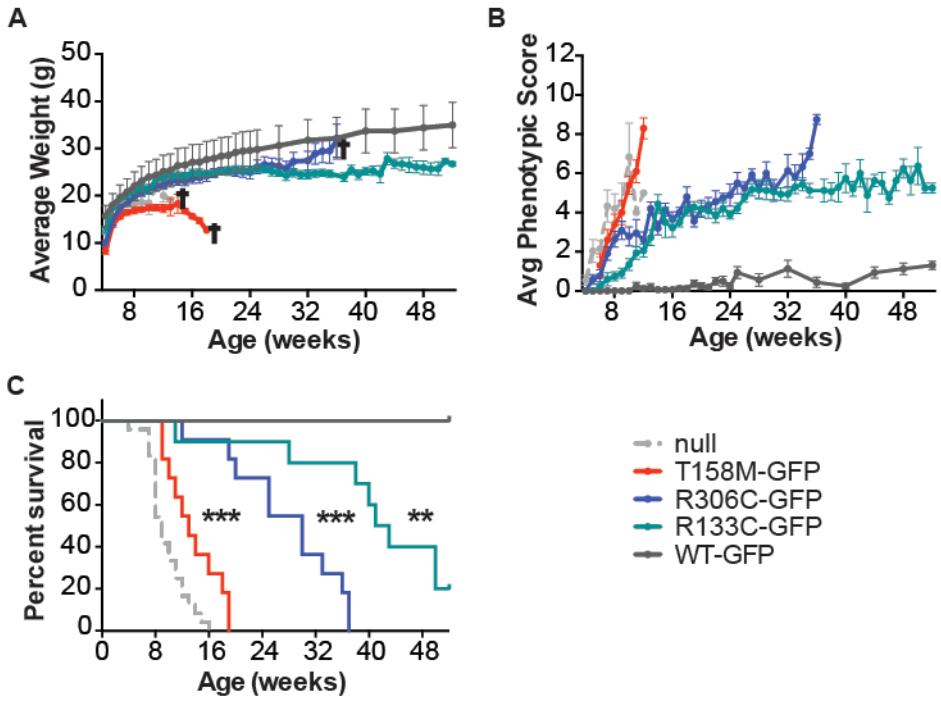


Figure 4.7.1 Phenotypic analysis reveals that the GFP-tagged allelic series recapitulates the spectrum of severity seen in patients.

(A) Graph showing that the GFP-tagged allelic series recapitulates the spectrum of severity observed in patients in terms of growth. The crucifix symbol indicates that there were no more surviving animals in the cohort. Mice were invariably culled due to severity of RTT symptoms. Mean±SEM plotted.
(B) Graph showing that the GFP-tagged allelic series recapitulates the spectrum of severity observed in patients in terms of RTT score. Mean±SEM plotted.
(C) The *R133C-GFP* mice have the longest mean survival of the allelic series. Survival was significantly reduced for all mutants in comparison to *WT-GFP*. **p<0.01, ***p<0.001, Log-rank Mantel Cox Test.
WT-GFP, n=8; *R133C-GFP*, n=10; *R306C-GFP*, n=11; *T158M-GFP*, n=11; *Mecp2*-null, n=12-24. Data from JS, JG & DD added to *R133C-GFP*.

4.7.2 Behavioural analysis

The *R133C-GFP* behavioural data was placed in the context of allelic series data gathered by J.Selfridge at 8-10 weeks (a time point when *T158M-GFP* mice were able to participate). Given that individual mutants were compared to corresponding wildtype littermates, the extra C57BL/6J back-cross in the *R133C-GFP* line was controlled for.

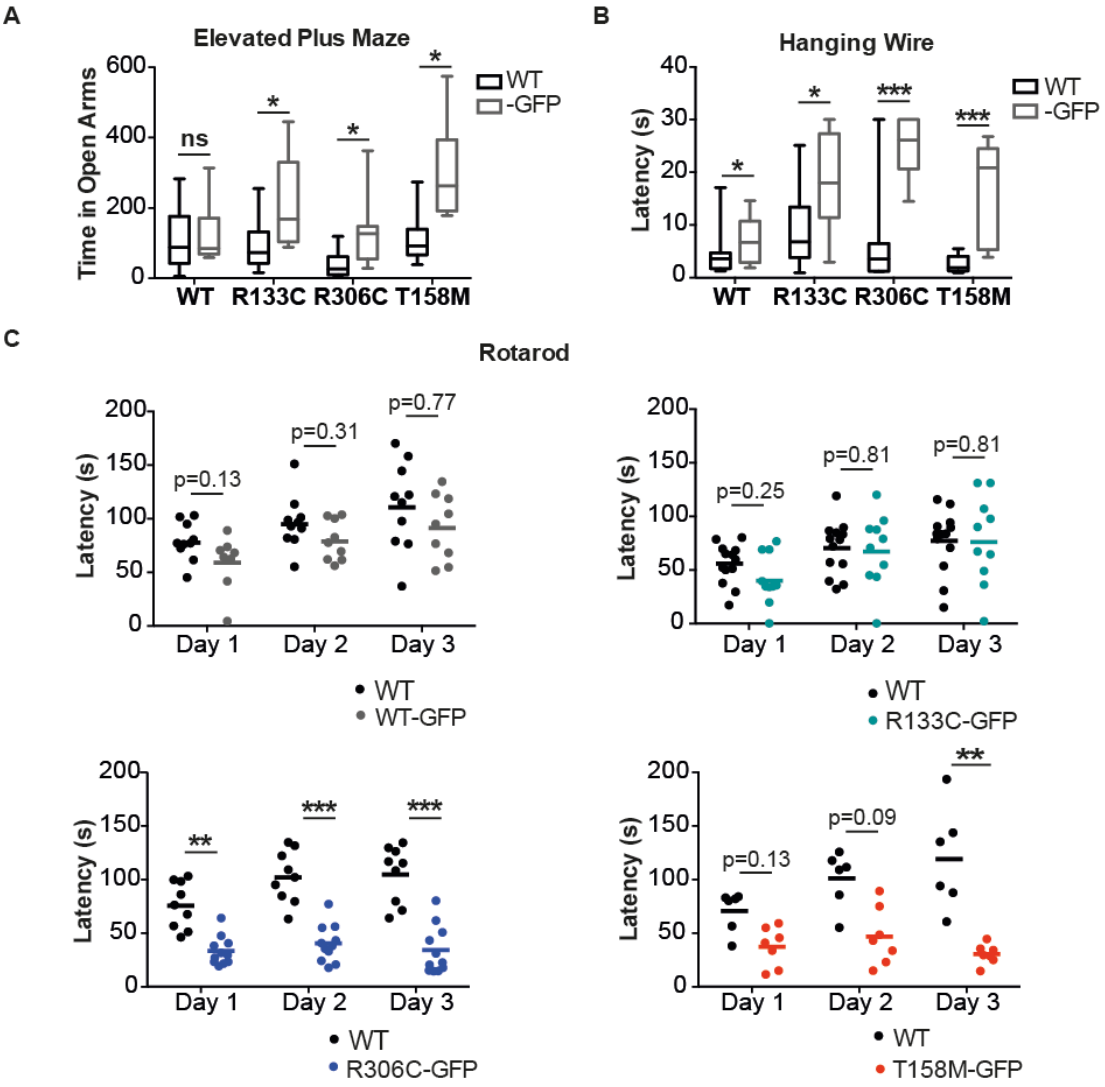


Figure 4.7.2 Behavioural analysis of the GFP-tagged allelic series recapitulates the spectrum of severity seen in patients.

(A) Box plots showing that all allelic series mutants spent significantly more time in the open arms of the elevated plus maze than their wildtype littermates. *WT-GFP* performed as wildtype (WT), p>0.99.
(B) Box plots showing that all cohorts of GFP-tagged mice took longer than their wildtype littermates on average to bring one hind paw to the wire in the hanging wire test.
(C) Scatterplots of individual mean latencies to fall from the accelerating rotarod (dots) and cohort means (lines). Only the *WT-GFP* and *R133C-GFP* performed as well as their wildtype littermates over all three test days.
WT-GFP n=9, WT n=10. *R133C-GFP*, n=10, WT n=13. *R306C-GFP* n=10-11, WT n=9. *T158M-GFP* n=7, WT n=6. Box plots of individual times. Cohort median, interquartile range, maximum and minimum represented. *p<0.05, **p<0.01, ***p<0.001 Kolmogorov-Smirnov Test.
Data from JS added to *R133C-GFP*.

All of the mutant mice spent significantly longer in the open arms than their wildtype littermates, indicating a reduction in anxiety or awareness of danger. Only the

T158M-GFP mice lost a preference for the closed arms over the open arms (384 seconds vs 302 seconds, data not shown). This would suggest that *T158M-GFP* was the most severely impaired in this trial, in keeping with the working hypothesis.

All of the mice, including the *WT-GFP* strain took significantly longer than their wildtype littermates to bring one hind paw up to grasp the wire in the hanging wire test (Figure 4.7.2B). However, the absolute difference in the mean times between the mutants and their corresponding wildtype littermates differentiated the genotypes: *WT-GFP* 5.1 seconds; *R133C-GFP* 9.2 seconds; *R306C-GFP* 18.9 seconds; and *T158M-GFP* 14.4 seconds. This result indicated that *R133C-GFP* was the least impaired of the allelic series lines on this test.

The accelerating rotarod test also supported the phenotypic severity hypothesis. Figure 4.7.2C shows that only *WT-GFP* and *R133C-GFP* performed as well as their wildtype littermates with no significant difference in the average latency to fall from the rod on any of the three test days. In contrast both the *R306C-GFP* and *T158M-GFP* mice fell significantly earlier than their wildtype littermates by day three of the motor paradigm, indicating impairment in motor coordination and ability.

In summary, from comparison with phenotypic and behavioural data from two other GFP-tagged missense mutant mouse lines, it can be concluded that the *R133C-GFP* mouse line has milder RTT. Therefore on a homogeneous genetic background with no possibility of skewing of XCI, the R133C mutation does result in a form of Rett syndrome that is less severe, in concordance with the patient studies.

4.8 Discussion

The objective of this chapter was to test the hypothesis that the R133C mutation in *MECP2* would result a milder form of Rett syndrome when the mutation was present on a homogeneous genetic background with no possibility of skewed XCI. In order to achieve this, the phenotype of the novel *R133C-GFP* mouse line was determined, before comparing this to other mouse models of Rett syndrome.

4.8.1 *R133C-GFP* mice as a model for human Rett syndrome

R133C-GFP mice were weighed and observed on a weekly basis for the development of RTT. The benefit of back-crossing onto an almost pure C57BL/6J genetic background was seen with the increased weight and variability in weight in the more outbred cohorts as well as significant aggression in the males, resulting in the early death of many mice in the experimental group. It is desirable that mice belonging to one experimental cohort differ only in the variable that is being measured and so introducing external variability from the degree of 129/Ola genetic background would act as a confounder of the experiment. However, from analysis of *R133C-GFP* mice at different degrees of back-cross, it was noted that there was little impact on the average phenotypic score. This is notable and interesting because it echoes what it seen in the patient population: with wide genetic diversity, patients carrying the R133C mutation still cluster together at one end of the spectrum of severity (Leonard *et al*, 2003).

4.8.2 Subjective phenotypic scoring does not reveal preserved capacities

The phenotypic scoring system is well established in the Rett research field (Guy *et al*, 2007; Appendix 1). Assessing six aspects of mouse appearance and behaviour, it encapsulates the key features of RTT. Although the mice were scored blind to genotype, as mice aged and became unwell, it became obvious which ones carried the mutation when observed next to their larger, healthier wildtype littermates. This compromised the blinding process. Another limitation to the scoring system is that there is an element of subjectivity in assigning a score to a particular symptom: one must judge the degree to which a mouse is impaired by a particular aspect of the phenotype. Inter-rater reliability has not been demonstrated where a second independent assessor's judgment of individual animals' phenotypic severity would consistently correlate well with the first's. This makes recording of weight, survival and behavioural analysis particularly important to include as more objective measures of phenotypic severity. There was no obvious preservation or loss of a specific capacity that would point to unique R133C Rett pathogenesis. Speech, hand use and ambulation are more often preserved in patients with R133C Rett

(Yamashita *et al*, 2001; Leonard *et al*, 2003). Two of these three features are difficult to assess in the mouse.

4.8.3 Progressive behavioural deficit in *R133C-GFP* mice

R133C-GFP mice underwent a battery of behavioural tests that measured motor and emotional function at a relatively pre-symptomatic age (9-10 weeks) and a symptomatic age (20-21 weeks). The pre-symptomatic age was chosen because RTT had not manifested in all of the *R133C-GFP* mice at this age and it would be possible to compare results to analyses done on other mouse lines at the same stage of development. The symptomatic age was chosen because all of the *R133C-GFP* mice were scoring for RTT by this time point. It was interesting that the motor impairment in the *R133C-GFP* mice at both time points was only detectable by the hanging wire test and not the accelerating rotarod. There are two possible explanations for this. Firstly, the tests measure different motor function: the hanging wire is a test of strength as well as coordination; the accelerating rotarod is a test of motor coordination as well as endurance, requiring adequate respiratory function. It could be that the *R133C-GFP* mice only have a deficit in motor ability that is specific to the hanging wire test, for example strength or hypertonicity. A second possible explanation is that the ability to complete the hanging wire test is affected by the presence of one of the core RTT symptoms in the mice – the hindlimb clasp. Once a mouse is suspended from the wire by its forepaws, if it clasps the hindlimbs together and cannot unclasp them during the 30-second trial, it will have no opportunity to bring its hind paws to the wire. Overall the behavioural testing showed evidence of a mild motor impairment on the hanging wire test and a deficit in the elevated plus maze and open field test in the *R133C-GFP* mice.

The elevated plus maze and open field tests have been extensively validated as assays of rodent anxiety, where animal performance is sensitive to anxiolytic medication (Pellow *et al*, 1985; Prut & Belzung, 2003). Therefore abnormalities detected in the *R133C-GFP* cohorts appeared to be characterized by a lack of wildtype anxiety about threatening environmental circumstances, spending significantly more time in the open arms of the elevated plus maze, or the centre of

the open field. This result seems incompatible with the heightened levels of anxiety seen in Rett syndrome patients (Smeets *et al*, 2011) although this has been reported previously in other *Mecp2* deficient mice, including female heterozygotes (Kerr *et al*, 2008; Samaco *et al*, 2008; Goffin *et al*, 2011; Samaco *et al*, 2012; Yasui *et al*, 2014). It is likely that the genetic background of the mice contributes to the particular anxiety phenotype (Crawley *et al*, 1997) as increased anxiety has also been reported in Rett models (Shahbazian *et al*, 2002; Heckman *et al*, 2014). Alternatively (or additively) it may relate to the particular *Mecp2* mutation. Two other factors may account for this apparent reduction in anxiety found on these two behavioural tests. Firstly, Rett patients are also moderately to severely intellectually disabled, however, and perhaps this mouse behaviour does not represent a calmer animal, but an animal that is less able to cognitively appraise its environment and act accordingly, a function that would be reliant on the frontal cortex as well as the limbic system. This would be a different quality to the muscle memory, training and motor learning displayed in the accelerating rotarod, which is perhaps more reliant on peripheral muscle, cerebellar and respiratory function. Further cognitive testing of the cohorts, for example using an object recognition test, may help to clarify whether this was a contributory factor. Secondly, the phenomenon of thigmotaxis, where a rodent remains close to the walls of tunnels in their natural habitat, or the walls of the testing apparatus, is dependent on vision and use of the whiskers and therefore adequate sensory capabilities. As sensory function was not tested in this study, it cannot be said conclusively that the *R133C-GFP* mice spent more time in the anxiety-provoking regions of the apparatus due to reduced anxiety and not impaired sensory function.

There was subtle progression in the behavioural phenotype between the pre-symptomatic and symptomatic age points. At 20-21 weeks *R133C-GFP* mice no longer spent significantly more time in the closed arms than the open arms of the elevated plus maze, indicating reduced anxiety or reduced appreciation of hazardous environmental cues. In addition, their performance deteriorated, relative to wildtype littermates, on the hanging wire test and there was a non-significant trend towards poorer performance on the accelerating rotarod. Although the symptomatic mice did

not spend significantly more time in the centre of the open field than their wildtype littermates, there was a strong trend towards this outcome. These observations help to validate the use of the behavioural paradigms as a correlate of progressive phenotypic severity. Moreover they are suggestive of increased sensitivity compared to simple observation as clear behavioural deficits were found at a relatively "pre-symptomatic" stage. The assumption was made that testing two separate cohorts of *R133C-GFP* mice and their wildtype littermates at the two time points was preferable to using the same cohort of mice in a longitudinal study design. This was because the mice would not have had the opportunity to learn, or acclimatise to the testing apparatus, which would appear less anxiety-provoking on the second presentation, thus potentially confounding the results of the elevated plus maze and the open field tests. Additionally, the mice may have achieved artificially loner times on the accelerating rotarod and reducedimprovement across days of testing due to retained motor learning from the experimental run at 9-10 weeks. On the contrary, the possibility that the second cohort of *R133C-GFP* mice and their wildtype littermates (tested at 20-21 weeks) were different in some unaccounted for way to the initial cohort (tested at 9-10 weeks) cannot be excluded. A longitudinal behavioural study design would therefore complement the behavioural, phenotypic scoring, growth and survival data from this study in concluding that there was progression in the behavioural phenotype of the *R133C-GFP* mice.

4.8.4 Possible confounders of +/R133C-GFP behavioural analysis

There was no discernible behavioural phenotype in the *+/R133C-GFP* mice – arguably the true model of Rett syndrome. The females began to phenotypically score at around eleven months and the behavioural testing was done at one year (due to time constraints), so perhaps this was too early to detect a significant difference from the wildtype littermates. Should the females have been tested at 1.5 or 2 years, for example, given the progressive nature of the RTT phenotype, a behavioural deficit may have been more obvious in the *+/R133C-GFP* mice. The behavioural cohort was more outbred than the scoring cohort (87.5% C57BL/6J) and the *+/R133C-GFP,* but not the wildtype, experimental group contained Cre positive mice (5 of 8), due to low availability of females of a suitable age. This was a remnant

of breeding with the C57BL/6J CMV-Cre deleter line, in order to remove the NEO cassette from the targeted *Mecp2* locus. It was not anticipated that Cre positivity would affect the behavioural performance. It may have acted as a confounder because it appeared that the Cre positive mice were more active than the Cre negative mice. Mice were monitored by tracking software during the elevated plus maze and open field trials. Maps of mouse activity showed that the Cre positive individuals travelled greater distances during the trials. In addition Cre positive mice were jumpy and hyperactive in the cage environment and when being handled. From analysis of individual animal's performances, this increased activity did not seem to impact on the decision to spend more time in the safer areas of the apparatus, although this is difficult to conclude with any certainty. It would be interesting to repeat the behavioural analysis in a group of females on a pure C57BL/6J background in the absence of Cre to see if the same results are achieved.

4.8.5 The R133C mutation produces a milder Rett-like phenotype than R306C or T158M

Once the *R133C-GFP* mouse phenotype was known, it was possible to compare the observational and behavioural data to that of other GFP-tagged mutant mouse lines to determine if the R133C mutation resulted in the least severe phenotype of the allelic series. Here the *WT-GFP* line was employed to control for the presence of the GFP tag in the mutants as a potential confounder of any difference from their wildtype littermates. This was a useful exercise, as although the *WT-GFP* mice did not weigh significantly less, have higher phenotypic scores, or reduced survival compared to their wildtype littermates (data not shown), they were impaired on the hanging wire test. Therefore determining the difference between wildtype and mutant mean latencies to bring one hind paw to the wire was useful as an alternative read-out of phenotypic severity. Despite this, the *WT-GFP* mice performed as well as their wildtype littermates on all other behavioural paradigms. The overwhelming conclusion therefore is that the *WT-GFP* mice do not have an overt phenotype and so can be used as a proxy wildtype comparator for the allelic series. There was some evidence that they had mild overexpression of MeCP2, which may be related to the detected motor deficit as mice also display neurological phenotypes with

overexpression of MeCP2 (Collins *et al*, 2004). This is discussed in more detail in Chapter 5.

4.8.6 Summary

Weight, average phenotypic score, survival and behavioural data all pointed convincingly towards the *R133C-GFP* mouse having the least severe phenotype of the allelic series. Therefore the *R133C-GFP* mouse nicely recapitulates human Rett syndrome caused by the R133C mutation and lends extra weight to conclusions from the patient studies (Leonard *et al*, 2003; Bebbington *et al*, 2008; Cuddapah *et al*, 2014). This supports important prognostic information for families of affected patients, at the point of diagnosis of Rett syndrome. Although Rett syndrome caused by the R133C mutation is still a severe and debilitating illness, the possibility of preservation of some faculties is hugely encouraging for affected families. Additionally, the mouse line will provide a useful resource for study of therapeutic strategies as well as a vehicle for determining the molecular basis of *MECP2*^{R133C} Rett syndrome, which is addressed in Chapter 5.

5. The Molecular Basis of Rett Syndrome Caused by the R133C Mutation

5.1 Introduction

Previous chapters have described creation of models in which to study MeCP2 function and verification that within a model organism, the R133C mutation in *Mecp2* produces a mild Rett-like phenotype, as is observed in Rett syndrome patients (Leonard *et al*, 2003). The goal of this chapter is to determine the specific deficits in MeCP2 function that are caused by the R133C mutation and their biological consequences. In other words, what is the molecular basis of $MECP2^{R133C}$ Rett syndrome?

The bridge model, which was conceived from observing the distribution of Rett-causing mutations, hypothesises that the critical function of MeCP2 is to connect the NCoR complex with methylated DNA in preventing Rett syndrome (Lyst & Bird, 2015). MeCP2 works as a transcriptional repressor according to this model. The complexity of this theory increases when it is considered that in the unique neuronal methylome (Lister *et al*, 2013), MeCP2 binds to mCpG across the whole genome (Skene *et al*, 2010) but also binds to methylation and hydroxymethylation in a non-CG context, about which very little is known (Guo *et al*, 2014; Gabel *et al*, 2015; Chen *et al*, 2015). It is possible that the capacity of MeCP2 to bind to different sequences and signalling moieties represents different biological functions of the protein. Support for this comes from the observation that in brain, although mCA content of long genes seems to relate to MeCP2 regulation (Gabel *et al*, 2015), genes that are downregulated in the absence of *Mecp2* are also rich in mCH (Chen *et al*, 2015). Furthermore, mCA in the neuronal genome has been associated with transcriptional repression (Xie *et al*, 2012; Guo *et al*, 2014), whereas mCA in the ES cell genome has not (Lister *et al*, 2009; Ziller *et al*, 2011). MeCP2 binding to methylation marks in different cellular contexts may result in alternative biological outcomes, perhaps dependent on the pool of binding partners in a particular cell type at a particular developmental stage.

Examining the specific deficits in MeCP2 function caused by Rett mutations will shed some light on these issues. The R133C mutation is well placed to disentangle the binding capacities of MeCP2 because it occurs in a key position of the MBD (Ho *et al*, 2008), yet results in a milder syndrome in humans (Cuddapah *et al*, 2014) and mice. One hypothesis is that the R133C mutation abolishes all of the protein's binding capacities because the mutation lies in a key position in the MBD. Therefore the mutation no longer connects the NCoR complex to DNA and the result is Rett syndrome. This theory does not explain the reduced severity of R133C Rett. An alternative hypothesis is that the R133C mutation loses one of the binding capacities in particular and, for this reason, results in a milder form of Rett syndrome. The biological consequences of the second hypothesis might illuminate any differential effects of binding to mCpG versus mCA. For example, a unique gene expression profile in tissue derived from the *R133C-GFP* mouse.

Previous studies have shown on one hand that $MeCP2^{R133C}$ binding to mCG is significantly impaired (Ballestar *et al*, 2000) and on the other hand that $MeCP2^{R133C}$ binding to mCG is near wildtype, but there is a specific impairment in binding to hmC in the context of DNA probes (Mellen *et al*, 2012). These conflicting findings have been replicated in cell culture experiments analysing localisation to heavily methylated heterochromatin, indicating both that binding is indistinguishable from wildtype (Kudo et al, 2003) and that it is modestly defective (Kumar *et al*, 2008; Schmeideberg *et al*, 2009; Agarwal *et al*, 2011). The only prior ChIP study of $MeCP2^{R133C}$ suggested a complete loss of binding selectivity for mCG (Baubec *et al*, 2013). $MeCP2^{R133C}$ binding to mCH has not previously been investigated.

In order to resolve these discrepancies and test the two hypotheses above, $MeCP2^{R133C}$ binding was analysed by confocal microscopy and ChIP in neuronal cells and mouse brain and by EMSA *in vitro*. ChIP regions included the *BDNF* promoter site, which has previously been shown to be an endogenous target of MeCP2 (Chen *et al*, 2003) and three classes of heavily methylated repetitive DNA: (1) highly abundant major satellite DNA, which constitutes almost half of murine

mCpG (Horz & Altenburger, 1981); (2) L1 long interspersed repetitive elements (L1 LINE), which are integrated throughout the murine genome with truncated versions of the full-length consensus sequence being represented hundreds of thousands of times (Boyle *et al*, 1990; Brunmeir *et al*, 2010); (3) Intracisternal A-particle (IAP) elements, which are related to retroviruses and transpose into new genomic sites through reverse transcription of an RNA intermediate. They are represented around 1000 times in the mouse genome (Barbot *et al*, 2002). MeCP2 binding has been shown to all three classes of repetitive DNA (Muotri *et al*, 2010; Skene *et al*, 2010; Chen *et al*, 2015). In addition, gene expression was investigated using microarrays in *R133C-GFP* cerebellum. These investigations were conducted with the purpose of determining the molecular basis of R133C Rett syndrome.

5.2 Investigation of hypotheses that do not relate to DNA Binding

5.2.1 R133C-GFP expression in neurons

Lack of sufficient protein would be an obvious molecular cause of disease and so it was crucial to rule this out in the first instance. *R133C-GFP* mRNA and protein levels were measured in the neuronal cells and adult male mouse brain tissue using reverse transcription qPCR and western blot, respectively. Wildtype and *WT-GFP* cells and brain tissue were included for comparison: for the brain tissue, the wildtype samples were littermates of the *R133C-GFP* mice.

The relative level of *Mecp2* mRNA in the *WT-GFP* cells was 152% that of the wildtype neuronal cells (Figure 5.2.1A). Interestingly, the presence of the R133C mutation reduced the level of *Mecp2* mRNA, so that it was not significantly different from that in the wildtype cells (96%). This made *R133C-GFP* mRNA approximately 63% of the measured level in *WT-GFP* cells. Although the level of MeCP2 protein appeared greatly increased by around four times in the *WT-GFP* neuronal cells, relative to wildtype, the difference did not reach significance due to variability between individual replicates (Figure 5.2.1B). The level of protein in the *R133C-GFP* cells also appeared increased by around 1.6 times relative to wildtype cells, but

this did not stand up to statistical testing, again due to variability in the results from replicate neuronal cultures. Comparing the two EGFP-tagged cell lines, the level of protein in *R133C-GFP* cells was 59% that in *WT-GFP* cells. Therefore from analysis of the neuronal cells, EGFP-tagging MeCP2 increased the level of the mRNA transcript by around 1.5 times, but having the R133C mutation on top abolished this effect. The *Mecp2* mRNA level in *R133C-GFP* cells was around 60% of *WT-GFP* mRNA. There was a trend towards the same phenomenon in MeCP2 protein levels, which did not reach statistical significance.

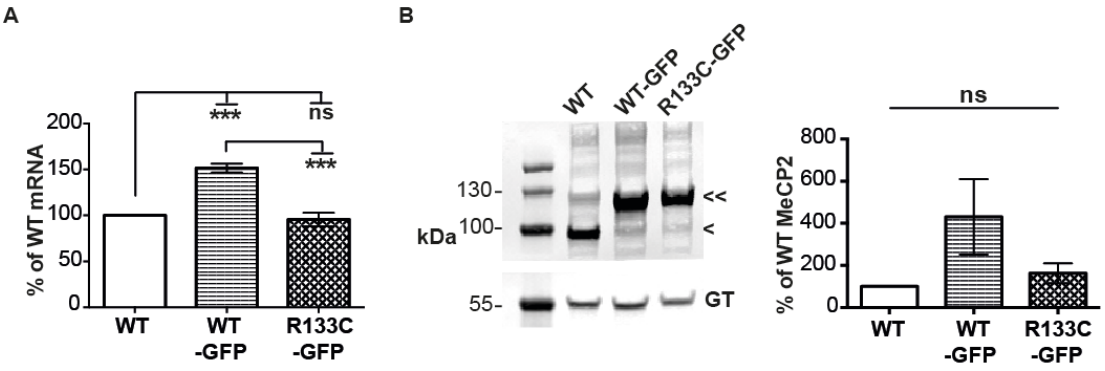


Figure 5.2.1 *R133C-GFP* neuronal cells have reduced *Mecp2* mRNA and a trend towards reduced MeCP2 expression in comparison to *WT-GFP* cells.

(A) Graph showing that *Mecp2* mRNA level is reduced in *R133C-GFP* neuronal cells relative to *WT-GFP*, but not wildtype (WT) cells. n=4, mean±SEM plotted, one-way ANOVA followed by Tukey's multiple comparison post-hoc testing, ***p<0.001.
(B) Left: representative western blot. Right: quantification of western blotting showing a trend towards reduced protein expression in *R133C-GFP*, relative to *WT-GFP* neuronal cells. n=3, mean±SEM plotted, one-way ANOVA, p=0.15. << $MeCP2^{EGFP}$, < MeCP2, GT gamma tubulin.

5.2.2 R133C-GFP expression in mouse brain

The level of *Mecp2* mRNA was also significantly increased in the *WT-GFP* male mouse brain, relative to wildtype (Figure 5.2.2A). *WT-GFP* brains had 152% the *Mecp2* mRNA level found in wildtype brains, which was exactly the same increase found in the neuronal cell model. Again, the presence of the R133C mutation in addition reduced this back to a similar level found in wildtype brain (89%). The level of *Mecp2* mRNA in *R133C-GFP* brain was therefore approximately 60% of that found in *WT-GFP* brain. These data were reflected in quantitative western blotting of whole brain lysates (Figure 5.2.2B). The level of MeCP2 in the *WT-GFP* brain was

163% the level in wildtype brain. Again, with the R133C mutation, the MeCP2 protein level dropped back to 93% of the level in wildtype brain. This means that the MeCP2 protein level in the *R133C-GFP* brain was 57% of that in *WT-GFP* brain. The relative mRNA and protein levels found in the neuronal cell model therefore support the relative levels found in using the whole brain tissue. To conclude, the R133C mutation reduces *Mecp2* mRNA and protein levels by approximately 40% *in vitro* and *in vivo*.

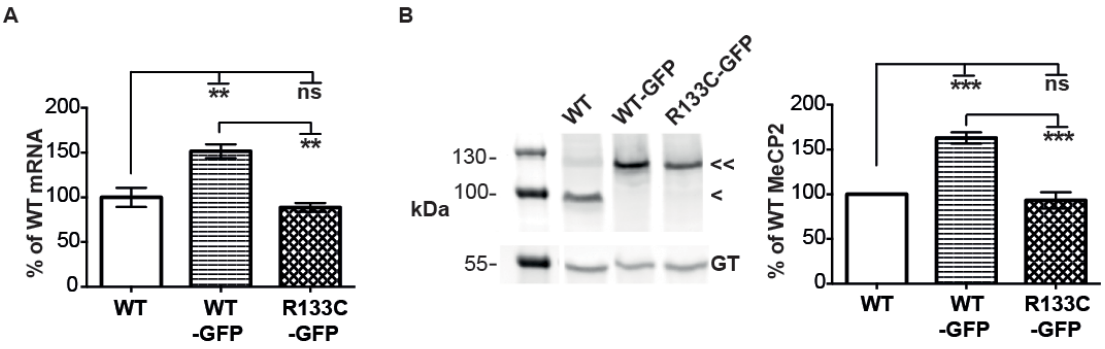


Figure 5.2.2 *Mecp2* mRNA and protein expression is reduced by 40% in adult male *R133C-GFP* mouse brain, relative to *WT-GFP*.

(A) Graph showing that *Mecp2* mRNA level is reduced in *R133C-GFP* mouse brain relative to *WT-GFP*, but not wildtype (WT) brain. n=4, mean±SEM plotted, one-way ANOVA followed by Tukey's multiple comparison post-hoc testing, **p<0.01..
(B) Left: representative western blot for mouse brain. Right: quantification of western blotting, showing a reduction in MeCP2 expression in *R133C-GFP* brain, relative to *WT-GFP*. n=3, mean±SEM plotted, one-way ANOVA followed by Tukey's multiple comparison post-hoc testing, ***p<0.001. << $MeCP2^{EGFP}$, < MeCP2, GT gamma tubulin.

5.2.3 WT-GFP expression in the GFP-tagged allelic series

These findings were placed in the context of data from the comparator mouse lines, comprising the EGFP-tagged "allelic series", which was quantified by J.Selfridge. With reference to Figure 5.2.3A, it can be seen that all the other EGFP-tagged mutants had significantly increased *Mecp2* mRNA, relative to their own wildtype littermates. However, figure 5.2.3B shows that the relative protein abundance does not reflect this pattern in brain. *R306C-GFP* protein was increased by 146% relative to wildtype, but *T158M-GFP* protein levels were only 50% of wildtype. Placing this in the context of protein abundance, R306C-GFP, R133C-GFP and T158M-GFP are 90%, 57% and 31%, respectively of WT-GFP. Therefore it is likely that reduced

protein abundance of MeCP2^{T158M} contributes to the severity of Rett syndrome and may be the first indication of the reason that *MECP2*^{R133C}, another MBD mutant, produces a milder syndrome in patients. It is unlikely that protein abundance contributes to the *MECP2*^{R306C} phenotype.

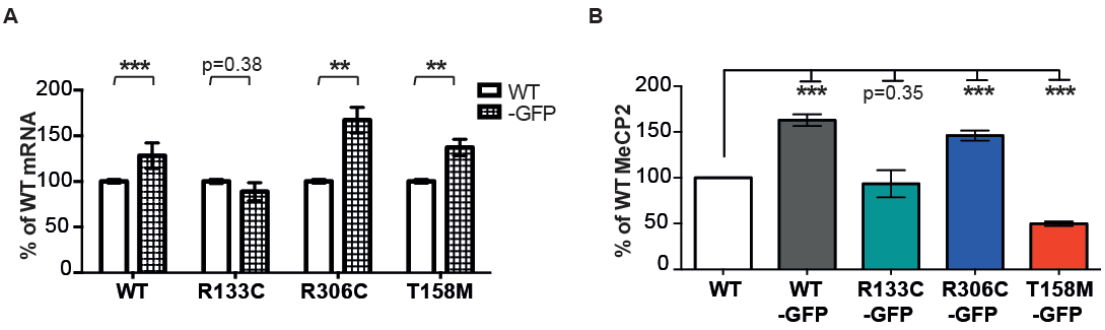


Figure 5.2.3 MeCP2 protein expression in adult male mouse brain partially explains the severity spectrum in the EGFP-tagged allelic series.

(A) Graph showing that *Mecp2* mRNA is significantly increased in GFP-tagged mouse lines, other than *R133C-GFP*, relative to their wildtype littermates. *WT-GFP* (n=9); *R133C-GFP* (n=4); *R306C-GFP* (n=3); *T158M-GFP* (n=3), mean ±SD plotted.
(B) Quantification of western blotting showing a significant reduction in T158M-GFP protein in mouse brain, relative to wildtype. *WT* (n=6); *WT-GFP* (n=6); *R133C-GFP* (n=3); *R306C-GFP* (n=4); *T158M-GFP* (n=6), mean±SEM plotted.
**p<0.01, ***p<0.001, unpaired t-test. Data from JS added for *WT-GFP*, *R306C-GFP* & *T158M-GFP*.

5.2.4 R133C-GFP association with the NCoR complex

Another important cause of Rett syndrome, according to the bridge model, is failure to associate with the NCoR transcriptional repressor complex (Figure 5.2.4A; Lyst & Bird, 2015). This is thought to be the molecular mechanism of Rett syndrome caused by *MECP2*^{R306C} (Lyst *et al*, 2013). It was important to rule this out as a contributory factor in the pathogenesis of Rett syndrome caused by *MECP2*^{R133C}. Co-immunoprecipitation experiments were done using whole brain lysates in order to check for association of MeCP2 with NCoR complex components, transducin β-like protein 1 related protein 1 (TBLR1) and histone deacetylase 3 (HDAC3) *in vivo*. MeCP2 was pulled down by virtue of its EGFP tag using an anti-GFP antibody and so in this experiment, the wildtype littermates acted as a negative control for the MeCP2 pull down, given that the protein was untagged. Association with NCoR components was detected using western blot of immunoprecipitates and the GFP-pull

down was checked with a C-terminal anti-MeCP2 antibody. *R306C-GFP* brains were included as a negative control for NCoR association.

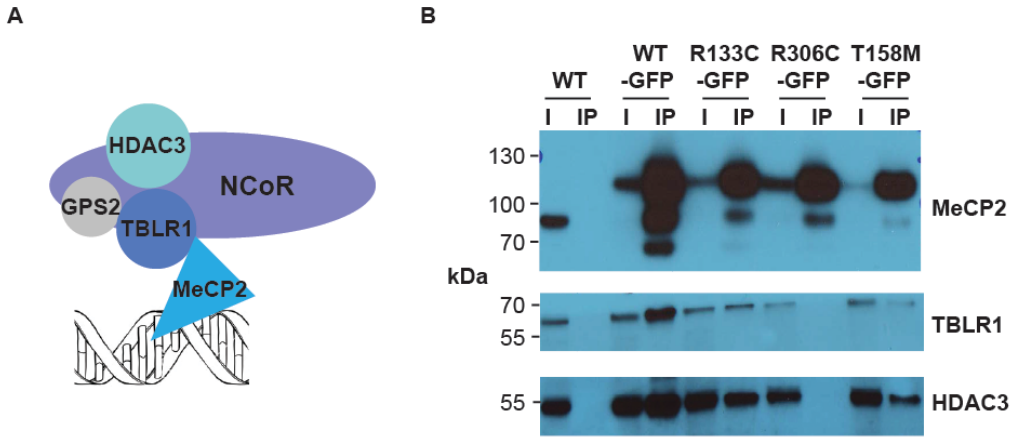


Figure 5.2.4 R133C-GFP associates with NCoR

(A) Simple cartoon depiction of the NCoR complex components in association with MeCP2, which, in turn, is making contact with DNA. GPS2, G protein pathway suppressor 2.
(B) Co-immunoprecipitation of MeCP2 and NCoR components in whole cell lysates from adult male mouse brain. MeCP2 was immunoprecipitated using an anti-GFP antibody and MeCP2-bound NCoR components were assayed using western blots. R133C-GFP associates with NCoR components, TBLR1 and HDAC3. Wildtype (WT), I Input, IP immunoprecipitate.

With reference to Figure 5.2.4B, it can be seen that there was a protein of between 70 and 100 kDa in the wildtype input lane of the MeCP2 blot, representing untagged MeCP2. The presence of TBLR1 (55-70 kDa) and HDAC3 (<55 kDa) was also detected. The wildtype immunoprecipitate (IP) lane was clear of all these proteins, indicating the specificity of the GFP pull-down. The *WT-GFP* input lane revealed a protein of between 100 and 130 kDa in the MeCP2 blot, which was the EGFP-tagged MeCP2. From scrutiny of all the EGFP-tagged protein input lanes, further evidence of the protein expression levels calculated in the preceding section could be seen: the band in the *WT-GFP* input lane was the darkest, followed by *R306C-GFP*, *R133C-GFP* then *T158M-GFP*. There was an enrichment of WT-GFP in the IP lane along with a band representing protein degradation (MeCP2 without the EGFP tag) and a non-specific band migrating at a smaller size. The presence of TBLR1 and HDAC3 could be seen in the *WT-GFP* IP lane, indicating that *WT-GFP* associated with these proteins in mouse brain. The same pattern is seen in the *R133C-GFP* co-IP,

indicating that MeCP2 with the R133C mutation still associates with NCoR components and therefore failure to interact with NCoR is not the cause of Rett syndrome in this model. The *R306C-GFP* IP lane was free of NCoR components, reproducing the data that this mutant protein does not associate with NCoR *in vivo* (Lyst *et al*, 2013). There was evidence that T158M-GFP also associated with NCoR from the co-IP. To summarise, the two *MECP2* missense mutants that are located in the MBD, R133C and T158M, still associated with NCoR *in vivo* and so it is unlikely that failure to associate with this complex plays a role in the pathogenesis of their resulting Rett syndromes.

5.3 R133C-GFP association with chromatin by microscopy

The abundance of R133C-GFP was not reduced to a degree that would account for RTT in the mice as absolute protein levels were equivalent to untagged wildtype protein and the mutant protein still associated with NCoR components *in vivo*. Therefore, according to the bridge model, the pathogenesis should involve a deficit in DNA binding, which would be predicted, given the location of the mutation in the MBD of *MECP2*. Association with heterochromatin was investigated in the neuronal cell model and adult mouse brain using confocal microscopy.

5.3.1 Verification of EGFP signal

Figure 5.3.1 demonstrates the specificity of the EGFP signal for MeCP2 in adult male mouse hippocampus. In fixed sections, the neurons of the CA1 region of the hippocampus were stained with NeuN and their nuclei with DAPI. The characteristic heterochromatic dots of the murine cell nucleus can be seen with the DAPI stain. MeCP2 typically associates with these heterochromatic dots, as they represent a substantial fraction of mCG in the mouse genome (Horz & Altenburger, 1981). *Mecp2-null*, wildtype and *WT-GFP* brains were compared to demonstrate (1) the lack of EGFP staining in the *Mecp2-null* and wildtype brains, (2) the presence of the MeCP2 signal in wildtype and *WT-GFP* brains using a C-terminal anti-MeCP2 antibody and (3) that this MeCP2 staining corresponds to the EGFP staining in the *WT-GFP* brains. Therefore EGFP signal is specific, representing MeCP2 with the

EGFP tag and the tagged protein displays the expected punctate pattern of localisation in the neuronal nucleus.

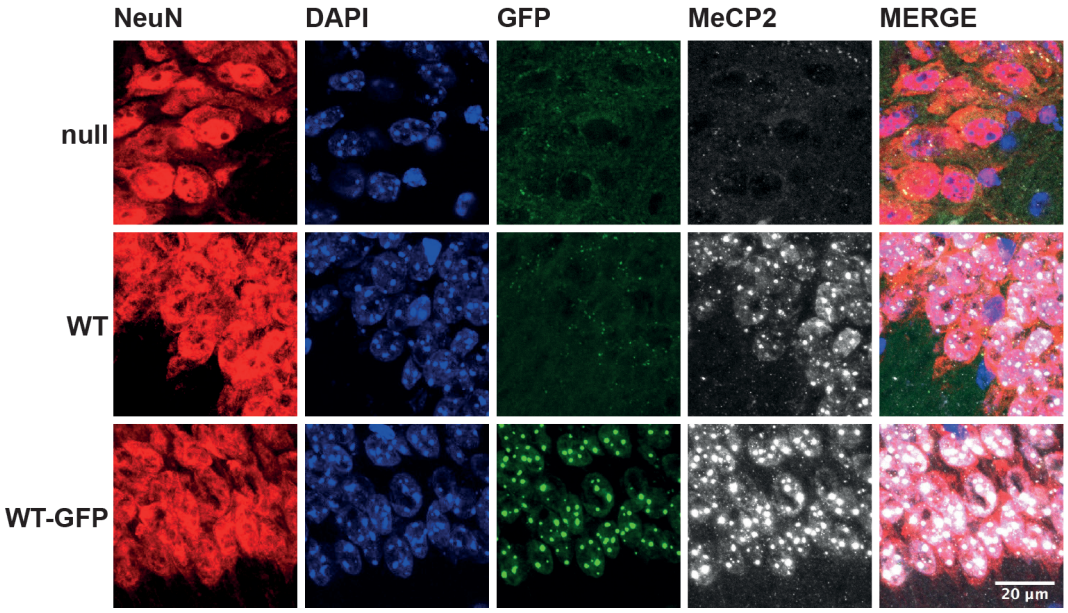


Figure 5.3.1 Verification that the GFP signal in tagged lines corresponds to MeCP2.

Sections of male mouse adult hippocampus were stained for NeuN (red), DAPI (blue) and MeCP2 (white) and visualised by confocal microscopy. *Mecp2-null* (null), wildtype (WT) and *WT-GFP* brains are shown to demonstrate the normal pattern of MeCP2 localisation to heterochromatic foci and that the GFP signal (green) in the *WT-GFP* brains is specific to MeCP2. Scale bar represents 20μm.

5.3.2 Reduced association of R133C-GFP with chromatin in neurons

R133C-GFP binding was analysed in the neuronal cells, differentiated from ES cells in culture, in comparison to the EGFP-tagged allelic series mutants (Figure 5.3.2A). Cells stained for NeuN, indicating that they were post-differentiation neuronal cells. R133C-GFP binding appeared to be abnormal when the protein was viewed by confocal microscopy using its EGFP tag (Figure 5.3.2A). In contrast to WT-GFP and R306C-GFP, the protein did not clearly correspond to the heterochromatic foci, visualised using the DAPI panel. This was also apparent in the DAPI/GFP merge panel, where DAPI dots with no corresponding GFP signal can be seen. The R133C-GFP protein appeared to be diffusely spread throughout the nucleus, with no clear

confinement to the dots. This was also apparent, to a greater degree, in the *T158M-GFP* cells, where it was difficult to appreciate discernible dots in the GFP panel.

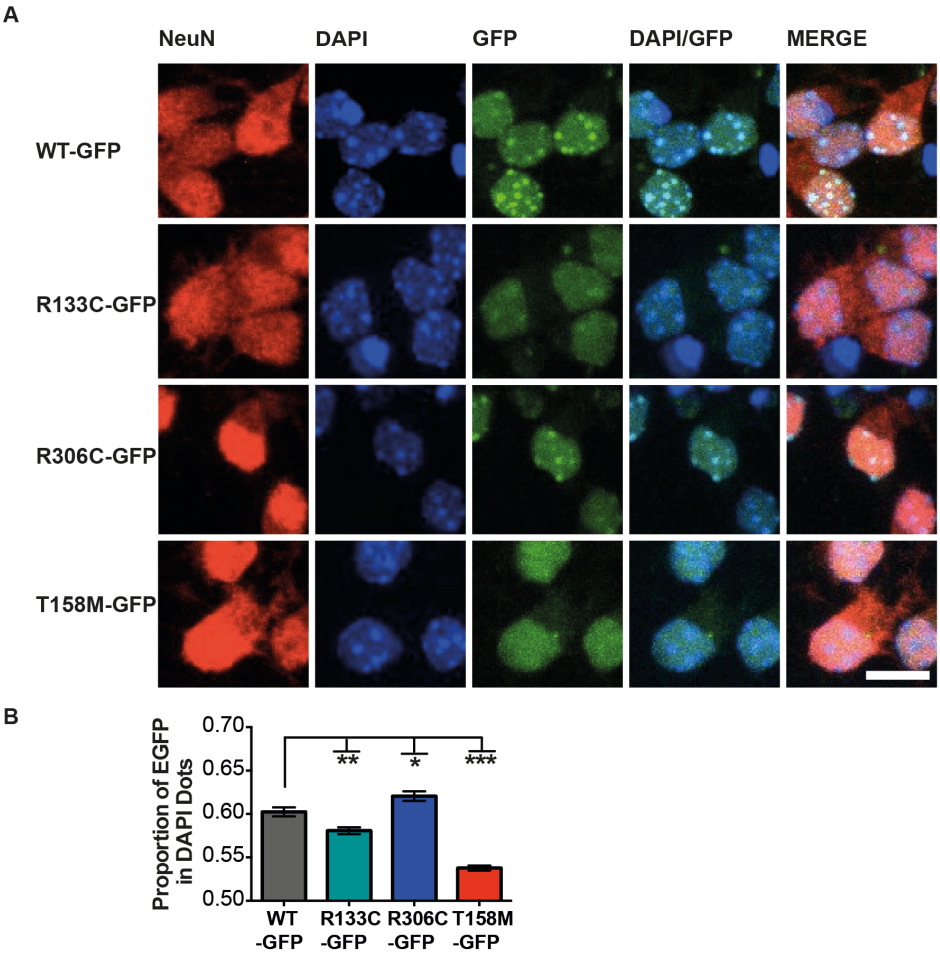


Figure 5.3.2 Abnormal localisation of R133C-GFP to heterochromatic dots in neuronal cells.

(A) Representative confocal microscopy images of *WT-GFP*, *R133C-GFP*, *R306C-GFP* and *T158M-GFP* neuronal cells. Localisation to heterochromatic foci was abnormal in the MBD mutants but not in *R306C-GFP*. NeuN (red), DAPI (blue), $MeCP2^{EGFP}$ (green). Scale bar 10μm.
(B) Graph showing quantification of $MeCP2^{EGFP}$ localisation to heterochromatic foci. A significantly lower proportion of $MeCP2^{EGFP}$ is bound to dots in the MBD mutants. Mean±SEM plotted. *p<0.05, **p<0.01, ***p<0.001 unpaired t-test.

The GFP signal was quantified for fifty cells of each genotype from compressed z-stack images. The proportion of GFP in DAPI dots was calculated as follows: individual cell nuclei were identified using Image Studio Software and verified manually; DAPI dot areas were identified by the software within each nucleus; the average GFP intensity within the DAPI dot defined areas was measured (a); the

average GFP intensity outwith the dot areas but within the nucleus was measured (b); the proportion of GFP in the dot areas was calculated as a/(a+b). This method controlled for overall MeCP2 expression variance between the genotypes. Using this technique, the proportion of MeCP2 in the nucleus that was located in the dots could be plotted and compared between the genotypes (Figure 5.3.2B). The *R306C-GFP* cells had a greater proportion of MeCP2 located in the dots than *WT-GFP* (0.62 vs 0.60). The two MBD mutant cell lines *R133C-GFP* and *T158M-GFP* had a lower proportion of MeCP2 located in the dots than *WT-GFP* (0.58 and 0.54 vs 0.60, respectively). This result was compatible with the imaging data and indicated that although the MBD mutant proteins still bound to heterochromatin, proportionately less of the total nuclear protein fraction was located there *i.e.* more protein was spread, presumably unbound, throughout the nucleus. The degree to which the protein was able to bind to the heterochromatic foci also reflected the severity of the Rett syndrome phenotype: R133C-GFP was localised to a greater degree than T158M-GFP and produced a milder phenotype. Therefore this suggests that heterochromatin localisation could be a proxy measure of phenotypic severity for Rett mutations in the MBD.

5.3.3 Reduced association of R133C-GFP with chromatin in male mouse brain

Confocal microscopy of fixed sagittal sections of adult male mouse brain was performed in order to ascertain what the localisation pattern of R133C-GFP was *in vivo* (Figure 5.3.3). All images were initially taken with the same confocal settings, however due to the lower levels of $MeCP2^{EGFP}$ expression in the mutant brains (Figure 5.2.3B), it was difficult to see the protein localisation. For clarity, a second set of images was taken where the GFP signal was optimised for each mutant individually. This was done by applying a look-up table (LUT) to an image of the CA1 region of the hippocampus. In LUT, blue pixels denote over-saturation of the image and green pixels denote zero. Photon multiplier tube gain was kept constant and laser power was adjusted until one or two pixels were blue and as few pixels as possible were green. These settings were saved and used for all subsequent images for that genotype.

R133C-GFP localisation was viewed in the CA1 region, CA3 region and dentate gyrus of the hippocampus and cortical neurons in the approximate region of the motor cortex (Figure 5.3.3A-D, respectively). Figure 5.3.3A shows the classic binding pattern of MeCP2 with tight localisation to heterochromatic dots in the *WT-GFP* hippocampus, CA1. This pattern was echoed in the *R306C-GFP* images. However, evidence of a more mixed punctate and diffuse binding pattern can be seen in *R133C-GFP* and *T158M-GFP* brains, indicating abnormal binding to DNA (represented in schematic GFP panel to the right of images). This is more clearly displayed in the lower panel of Figure 5.3.3A, taken with individual GFP settings for each genotype. Furthermore, in the DAPI/GFP merge panel on the GFP-optimised set of images, there is evidence of R133C-GFP and T158M-GFP binding in small foci outwith the DAPI dot regions (indicated more clearly in zoom images). These observations were made in all regions studied. Interestingly, in the cortical images (Figure 5.3.3D), there was evidence of reduced nuclear and soma size in the *T158M-GFP* brain, as reported in the literature for Rett patients (Bauman *et al,* 1995).

Figure 5.3.3A. CA1, hippocampus

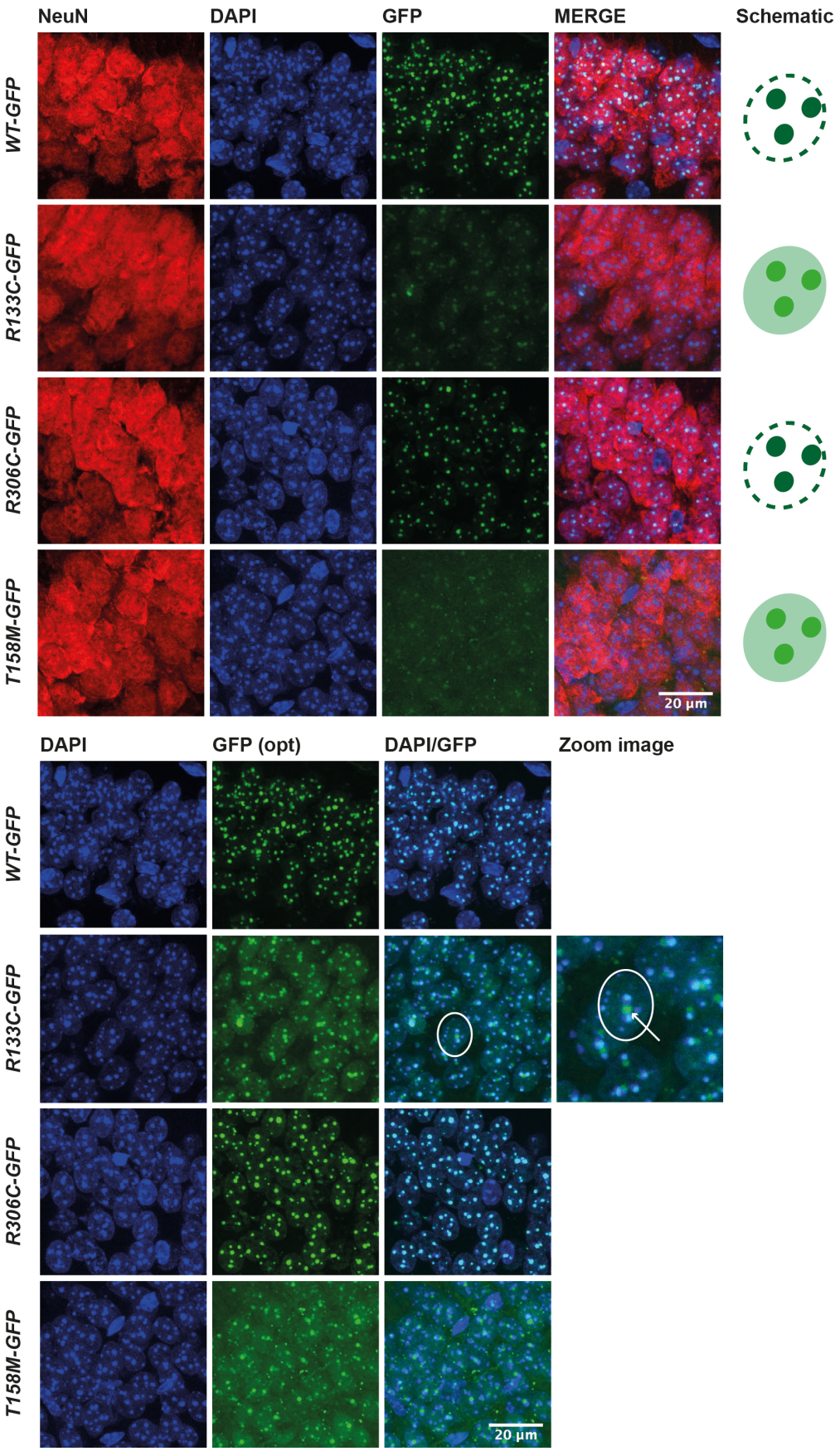


Figure 5.3.3B. CA3, hippocampus

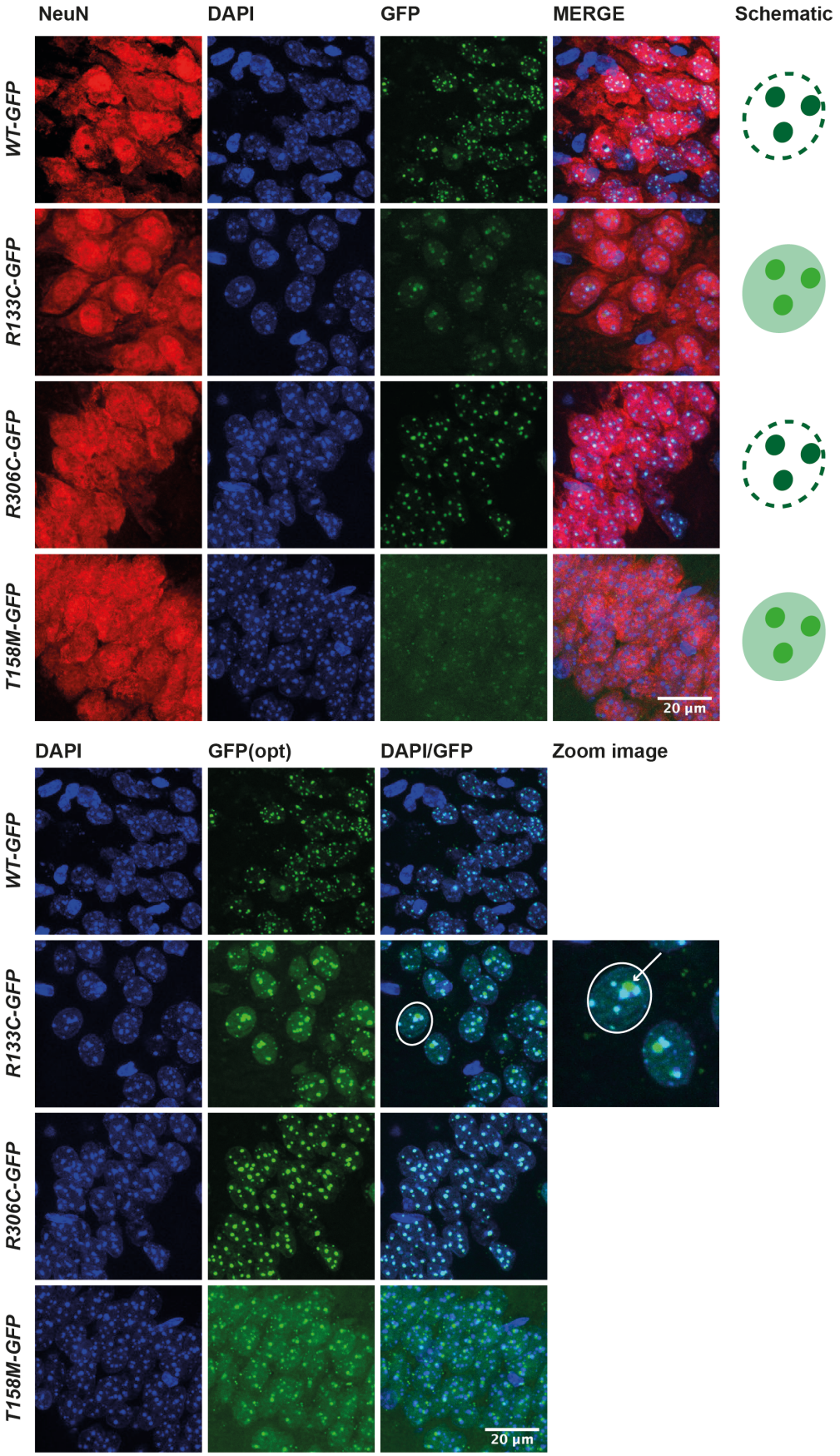
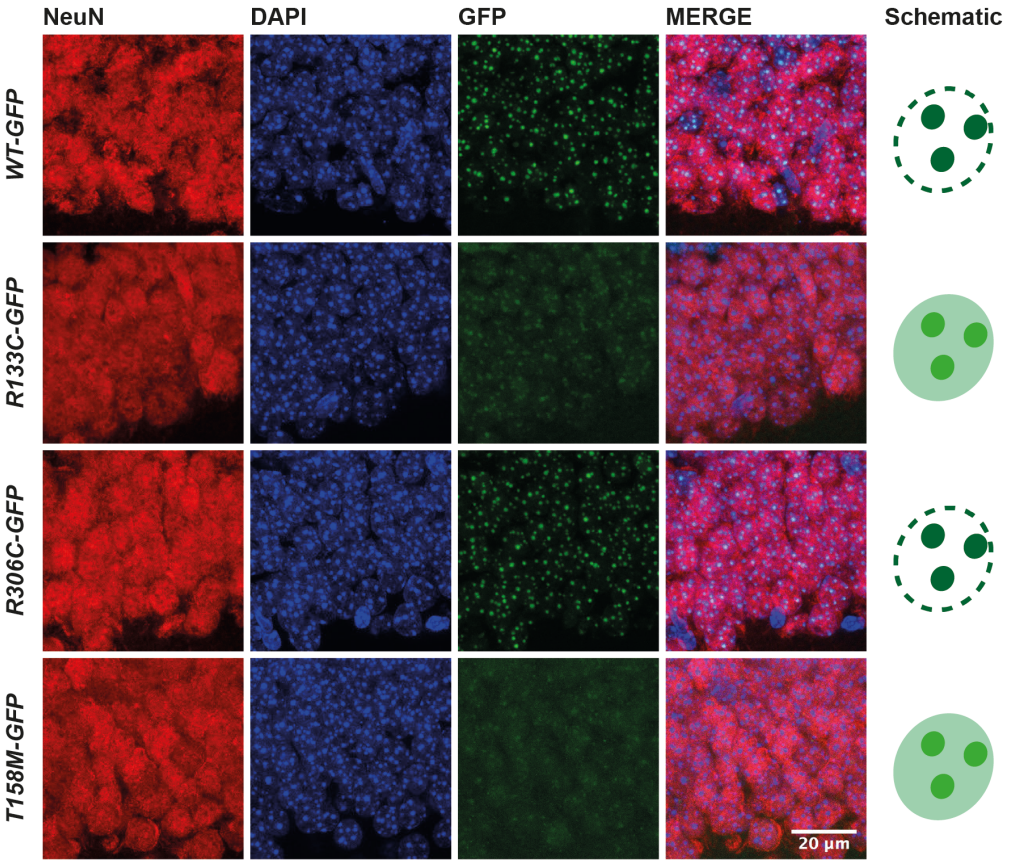


Figure 5.3.3C. Dentate gyrus, hippocampus



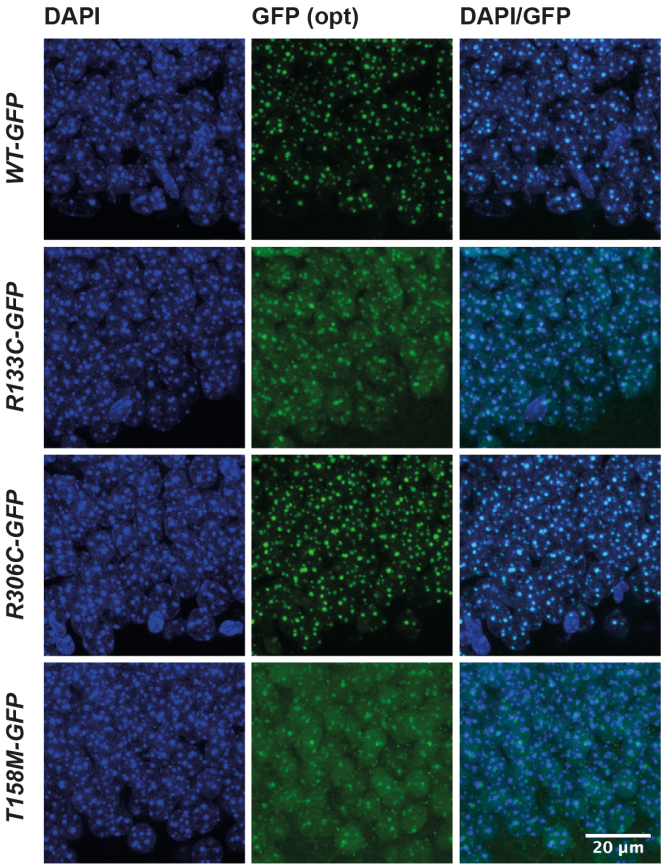


Figure 5.3.3D. Cortex

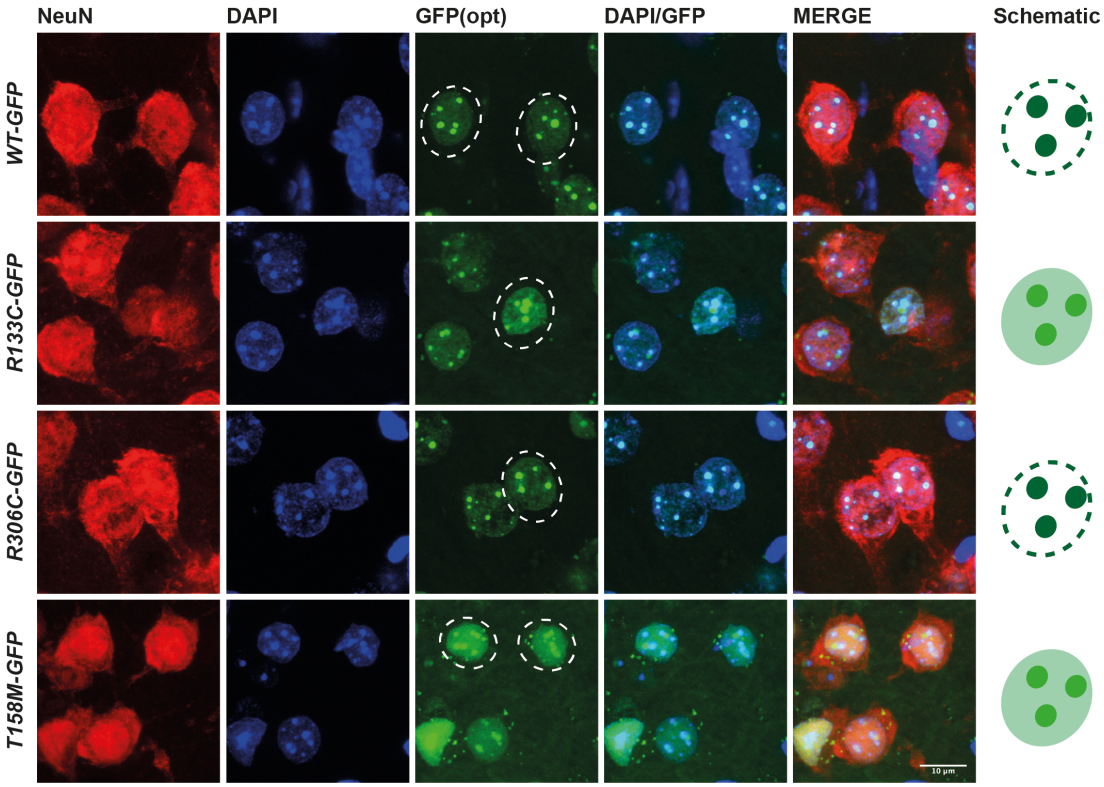


Figure 5.3.3 Abnormal localisation of R133C-GFP to heterochromatic dots in male mouse brain.

Confocal microscopy of male mouse brain revealed abnormalities of MeCP2^{EGFP} localisation to heterochromatic foci in the MBD mutants but not of R306C-GFP in the CA1 region (A), CA3 region (B) & dentate gyrus (C) of the hippocampus and in motor cortex (D). Observed localisation pattern is indicated by a schematic GFP panel to the right of the images. Additionally, R133C-GFP and T158M-GFP localise to foci that do not correspond to DAPI dots. This is highlighted with white arrows in zoom images of circled cells in (A) & (B). Fixed sagittal brain sections (10μm) were stained for NeuN (red) and DAPI (blue). MeCP2^{EGFP} (green). Upper panels were taken with the same confocal settings. In lower panels, GFP settings were optimised (opt) for each genotype to make binding pattern more visible. (A-C) scale bar represents 20μm; (D) scale bar represents 10μm, GFP optimised images only, white dotted lines indicate cells that were captured in their entirety by z-stacks.

5.3.4 Reduced association of R133C-GFP with chromatin in female mouse brain

The abnormal localisation of R133C-GFP is seen more easily in the images optimised for reduced protein abundance. To rule out the possibility that the diffuse binding pattern is an artefact of increasing the GFP laser intensity, a double mutant female mouse was created by crossing +/*R133C-GFP* mice with *WT-GFP* mice: *WT-GFP/R133C-GFP*. These mice expressed either wildtype protein with an EGFP tag

or R133C protein with an EGFP tag: the exact proportion of either was subject to the pattern of XCI. The binding of both proteins was therefore captured using the same confocal settings in the same image. Figure 5.3.4A shows control images taken from female mice heterozygous for *Mecp2*^{EGFP} (*+/WT-GFP* and *+/R133C-GFP*) prepared in parallel with the double mutant slices, for comparison.

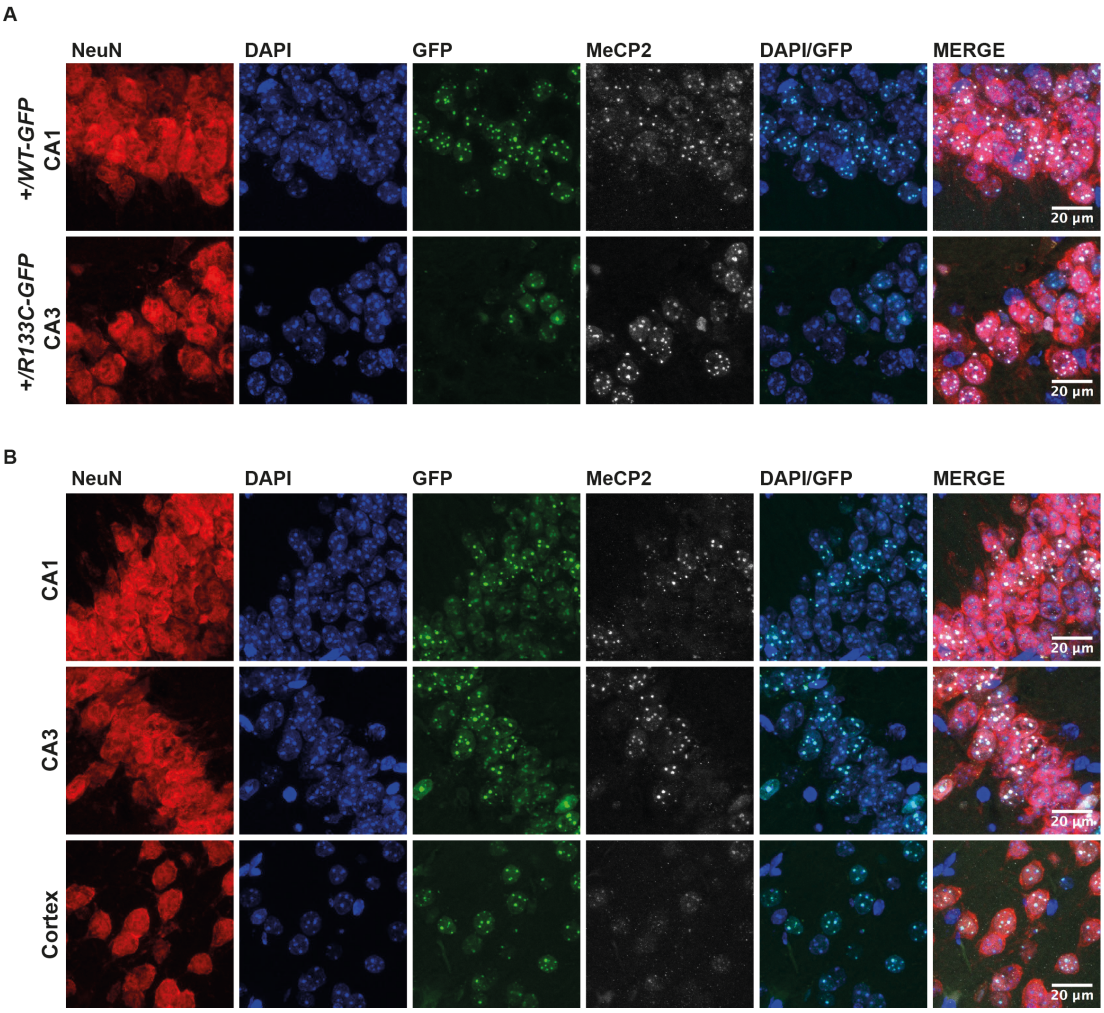


Figure 5.3.4 Abnormal localisation of MeCP2^{EGFP} in approximately 50% of cells in the *WT-GFP/R133C-GFP* mouse brain.

(A) Upper panel, CA1 region of hippocampus from *+/WT-GFP* brain shows EGFP expression in approximately 50% of the cells with punctate binding. Lower panel, CA3 region of hippocampus from *+/R133C-GFP* brain shows EGFP expression in approximately 50% of the cells with mixed punctate and diffuse binding.
(B) Images of CA1, CA3 and motor cortex from the *WT-GFP/R133C-GFP* brain. All cells now express EGFP: there is an abnormal pattern of localisation to heterochromatic foci in approximately 50% of neuronal nuclei.
Fixed sagittal sections were stained for NeuN (red), DAPI (blue) and MeCP2 (white). MeCP2^{EGFP} (green). Representative images shown. Scale bar 20µm.

The *+/WT-GFP* image (Figure 5.3.4A, upper panel) taken from the CA1 area of the hippocampus shows all of the neuronal nuclei in the NeuN, DAPI and anti-MeCP2 panels, but only half of the nuclei are visible in the GFP panel, indicating cells expressing WT-GFP from the active X chromosome. A typical punctate binding pattern can be seen in these nuclei. The pattern of XCI is also apparent in the DAPI/GFP merge panel, where the cells expressing wildtype MeCP2 only have blue DAPI dots. In contrast, the *+/R133C-GFP* image (Figure 5.3.4A, lower panel) from the CA3 area of the hippocampus again shows all the neuronal nuclei in the NeuN and DAPI panels and only some of the nuclei in the GFP panel, indicating cells expressing R133C-GFP from the active X chromosome, but this time the cells expressing the mutant protein, don't stain with the anti-MeCP2 antibody and the binding pattern is a mixture of punctate and diffuse. This is reaffirmed in the images taken from the double mutant *WT-GFP/R133C-GFP* across all brain areas (Figure 5.3.4B). All nuclei in the CA1 and CA3 regions of the hippocampus are stained with NeuN and DAPI, but only some (the WT-GFP expressing cells) stain with the anti-MeCP2 antibody. The GFP panel in these images shows a mixed population of cells: some with clear punctate binding where $MeCP2^{EGFP}$ is bound to heterochromatic foci and some with a mixed punctate and diffuse binding pattern. The background nuclear GFP intensity in these mixed pattern cells appears equivalent to the background GFP intensity in the punctate cells, but the foci are less intensely stained. The protein displays lower affinity for the heavily methylated dot regions. The lack of anti-MeCP2 staining in these cells also helps to identify them as the likely cells expressing R133C-GFP. The lack of anti-MeCP2 staining in the mutant cells was unexpected (see discussion).

In summary, investigating R133C-GFP association with chromatin by confocal microscopy revealed that the mutant protein had a subtly abnormal pattern of localisation to heterochromatic foci in neuronal cells differentiated in culture and in the adult male and female mouse brain. This is in contrast to previous literature suggesting an indistinguishable binding pattern for $MeCP2^{R133C}$ using transfection of recombinant protein in non-neuronal cells (Kudo *et al*, 2003). These findings indicate that $MeCP2^{R133C}$ association with heavily methylated chromatin is impaired.

5.4 R133C-GFP association with chromatin by ChIP

R133C-GFP binding to chromatin is impaired as measured by confocal microscopy. This DNA binding impairment can be further defined using ChIP where the amount of DNA, corresponding to specific genomic regions, that is pulled down by R133C-GFP can be measured by qPCR. From this one can infer the strength of the protein's association with the different regions.

5.4.1 Verification of methylation status of target regions

Four known MeCP2 binding sites were selected as ChIP regions in order to compare R133C-GFP binding to WT-GFP: major satellite DNA, L1 LINE and IAP elements (as methylated repetitive DNA) and the *BDNF* promoter region (as relatively unmethylated, CpG island-associated DNA). The methylation status of these regions in mouse brain was confirmed by bisulfite sequencing. Sodium bisulfite treatment of DNA converts all unmethylated cytosines to uracil, leaving methylated cytosines in tact. Regions of interest can be amplified by PCR using this bisulfite converted DNA, followed by sequencing. The methylation pattern of the original genomic DNA in this region can be inferred from any remaining cytosines in the sequenced PCR amplicon (Frommer *et al*, 1992).

Given that MeCP2 binds to methylated DNA, it is possible that the methylation pattern of the underlying DNA could change in the absence of MeCP2: is one of the functions of MeCP2 to maintain genomic methylation patterns? It follows that if there is reduced association with DNA in *R133C-GFP* brain, the underlying methylation status of the DNA might alter. For this reason, bisulfite sequencing was completed for wildtype, *WT-GFP* and *R133C-GFP* genomic DNA isolated from whole brain.

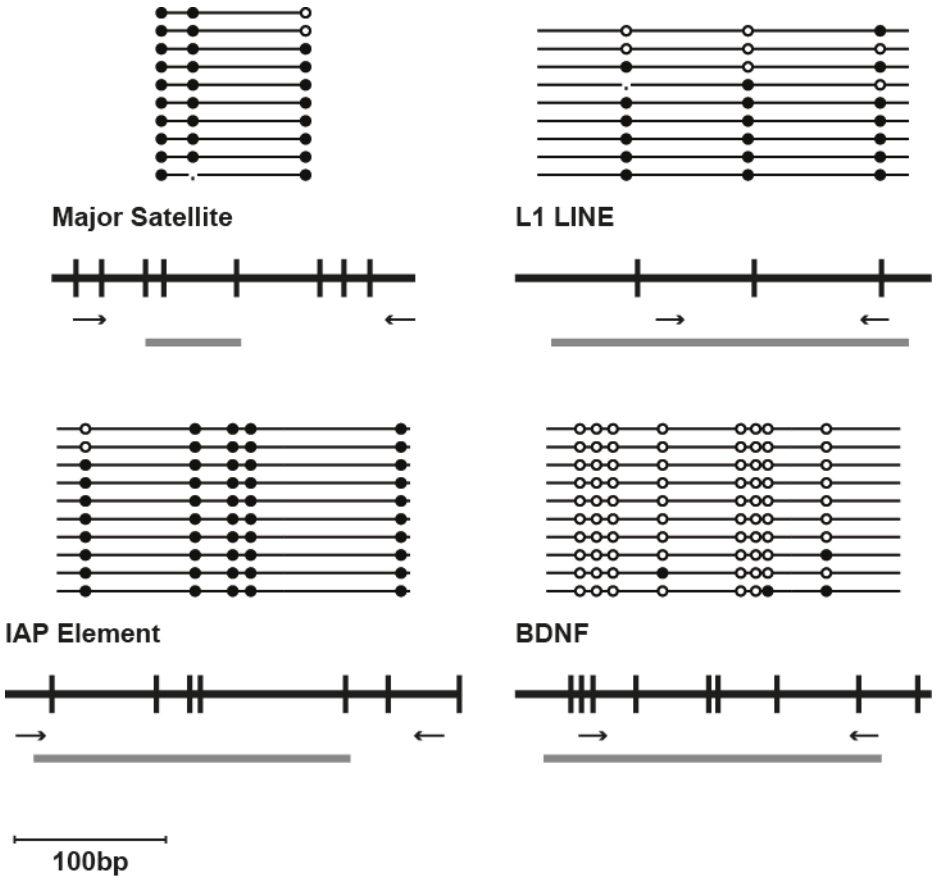


Figure 5.4.1 Bisulfite sequencing of ChIP regions confirms methylation status in mouse brain.

Bisulfite sequencing of repetitive elements and the *BDNF* promoter region in whole brain confirms that they are largely methylated and unmethylated, respectively. A map and bisulfite sequencing data are shown for each region. The black bar indicates genomic DNA with vertical black ticks representing CpG sites. Primers for qPCR are indicated as arrows and the location of the bisulfite amplicon by the grey bar. A scale bar is indicated. The methylation status of each CpG is represented by the circles: open are unmethylated; black filled are methylated; a dash indicates missing data. Each line represents a single clone. *WT-GFP* results are shown and were representative of wildtype and *R133C-GFP* sequencing.

Figure 5.4.1 shows the results of the bisulfite sequencing over major satellite, L1 LINE, IAP elements and the *BDNF* promoter for *WT-GFP* genomic DNA, along with the location of the bisulfite PCR amplicon. It can be seen that the bisulfite PCR amplicons correspond to the region amplified by qPCR following ChIP. Therefore the pattern of methylation was likely to be representative of the region bound by MeCP2 in the subsequent ChIP experiments. There were no major differences in the methylation pattern of the four regions between genotypes (wildtype and *R133C-GFP* data not shown). Major satellite DNA was heavily methylated in wildtype, *WT-*

GFP and *R133C-GFP* genomic DNA from whole brain (90%, 90% & 92% respectively); L1 LINEs were moderately methylated (67%, 63% & 72%); IAP elements were heavily methylated (98%, 96% & 80%); and the *BDNF* promoter region was largely unmethylated (3%, 5% & 8%). Therefore (1) it is unlikely that DNA methylation patterns change in the absence of MeCP2 binding and (2) these genomic regions could be considered representative of "methylated" and "unmethylated" DNA for all three genotypes for the purpose of ChIP analysis.

5.4.2 Reduced binding of R133C-GFP in cultured neurons

ChIP was performed to estimate R133C-GFP binding to major satellites, L1 LINEs and the *BDNF* promoter in neurons differentiated in culture. MeCP2 was immunoprecipitated using an anti-GFP antibody and so in this experiment *WT-GFP* cells and wildtype cells acted as positive and negative controls, respectively. The DNA immunoprecipitated by MeCP2 was expressed as a percentage of input DNA. Figure 5.4.2 shows that R133C-GFP pulled down 0.14% of the input DNA at major satellites. This was lower in comparison to WT-GFP (0.71%) but did not reach significance. For L1 LINEs, R133C-GFP pulled down a significantly lower percentage of the input (0.10%) in comparison to WT-GFP (0.32%). Finally, at the *BDNF* promoter site, R133C-GFP also pulled down a significantly lower percentage of the input (0.09%) in comparison to WT-GFP (0.24%). This would equate to approximately 20%, 31% and 38% of WT-GFP binding at major satellites, L1 LINEs and the *BDNF* promoter site, respectively. This data must be interpreted with the consideration that background protein abundance is only 60% of WT-GFP in the *R133C-GFP* cells. In summary, in the neuronal cells, there is evidence of a reduction in R133C-GFP association with both methylated and unmethylated genomic sites in comparison to EGFP-tagged wildtype protein.

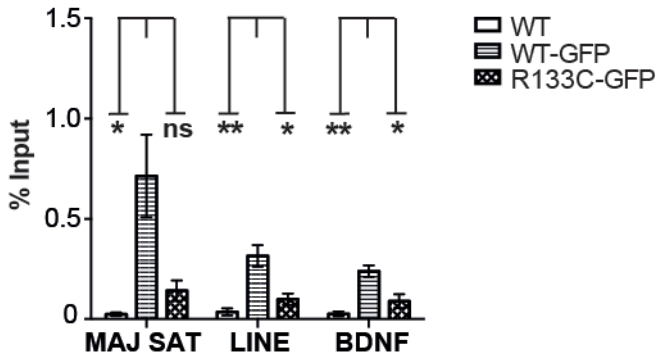


Figure 5.4.2 R133C-GFP binding to chromatin is significantly reduced, relative to WT-GFP by GFP ChIP.

Graph depicting significantly lower percentage of input DNA pulled down by R133C-GFP at L1 LINEs (LINE) and the *BDNF* promoter in comparison to WT-GFP in neuronal cells. n=3. Mean % input±SEM plotted. *p<0.05, **p<0.01, ns, p=0.05, unpaired t-test.

5.4.3 Reduced binding of R133C-GFP in mouse brain

The ChIP was repeated using the same regions plus IAP elements as another example of methylated repetitive DNA, in adult male mouse brain (Figure 5.4.3A). Relative to WT-GFP, there was a trend towards reduced R133C-GFP binding at major satellites (0.63% vs 0.19%), L1 LINEs (0.23% vs 0.11%), IAP elements (0.30% vs 0.13%) and the *BDNF* promoter (0.23% vs 0.11%). These relative reductions in percentage input did not stand up to statistical testing because the percentage of input DNA that was pulled down for both genotypes varied between replicate experiments (see discussion). In fact, the R133C-GFP percentage input expressed as a percentage of WT-GFP signal was quite consistent between replicates. Nevertheless, the results supported the findings in neuronal cells in that R133C-GFP binding was reduced relative to WT-GFP binding at methylated and unmethylated sites in the genome. R133C-GFP binding was approximately 30%, 48%, 43% and 48% of WT-GFP binding at major satellites, L1 LINEs, IAP elements and the *BDNF* promoter, respectively.

With the knowledge that the R133C mutation reduced MeCP2 protein abundance by around 40%, it was observed that the ChIP results could simply reflect protein abundance and not a deficit in DNA binding. FRAP data revealed that $MeCP2^{EGFP}$ dissociates from methylated DNA every 22 seconds and $MeCP2^{R133CEGFP}$ every 15

seconds (Schmiedeberg *et al*, 2009). Therefore in the eight minute ChIP cross-linking period, increasing amounts of MeCP2 would continue to bind to DNA. This might cloud the picture of relative binding of the two genotypes, as ambient nuclear protein could continue to associate until all binding sites are occupied and the readout reflected protein abundance instead of binding capacity (Poorey *et al*, 2013). The ChIP protocol was modified in order to address this issue, firstly by reducing the cross-linking time from eight minutes to one minute and secondly, by reducing the cross-linking temperature: samples were cross-linked straight from ice instead of being allowed to acclimatise to room temperature.

The results achieved following modification of the ChIP protocol are shown in Figure 5.4.3: shortening the cross-linking period (B) and lowering the cross-linking temperature (C). S.Lagger contributed replicate data to (B). Shortening the cross-linking time to one minute had little impact on the relative reduction in R133C-GFP binding. At major satellites mutant protein pulled down 0.62% of the input DNA, which was significantly lower than WT-GFP (1.31%) and represented 47% of binding. At L1 LINEs mutant protein pulled down 0.41% of the input DNA, which was significantly lower than WT-GFP (0.89%) and represented 46% of binding. At IAP elements mutant protein pulled down 0.56% of the input DNA, which was significantly lower than WT-GFP (1.20%) and represented 47% of binding. Finally, at the *BDNF* promoter, mutant protein pulled down 0.39% of the input DNA, which was significantly lower than WT-GFP (0.66%) and represented 59% of binding. These results again seemed to reflect relative protein abundance.

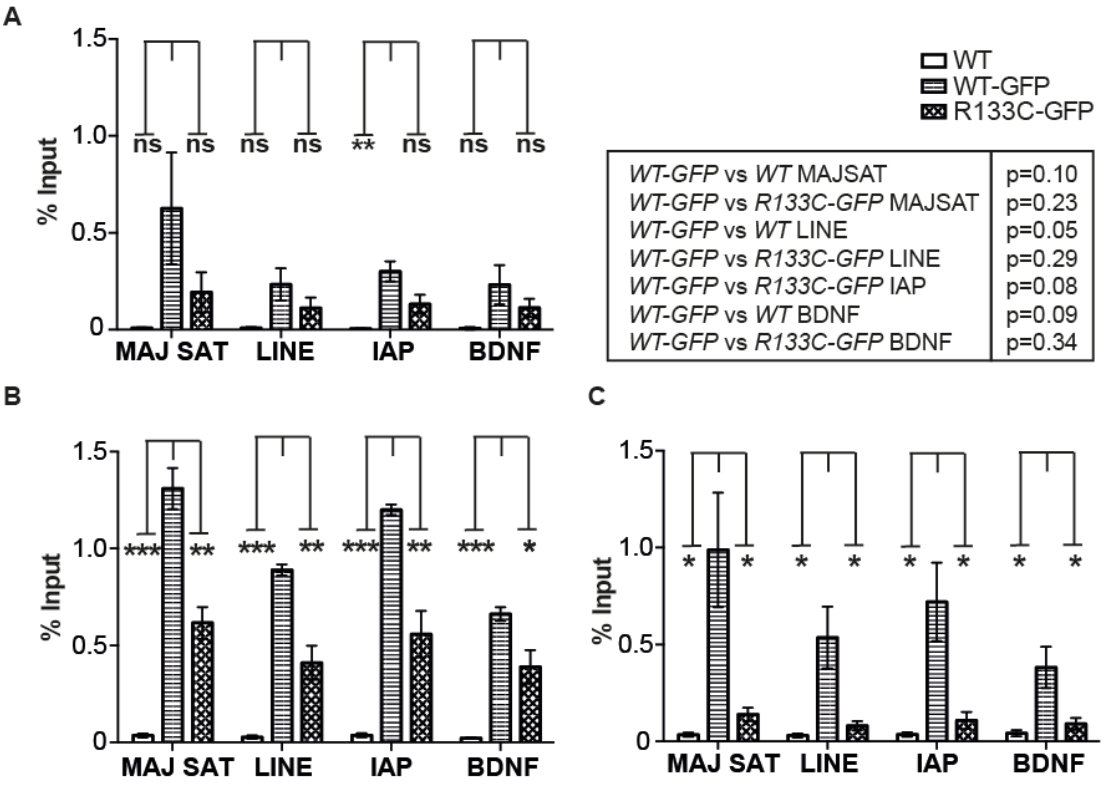


Figure 5.4.3 Increasing the stringency of GFP ChIP exposed a binding deficit in R133C-GFP at repetitive DNA and the *BDNF* promoter.

(A) Graph showing trend for reduced binding of R133C-GFP to major satellites, L1 LINEs, IAP elements and the *BDNF* promoter, relative to WT-GFP in mouse brain by GFP ChIP, n=3, p values for non-significant comparisons are detaied in adjacent table.
(B) Graph showing significant reduction in binding for R133C-GFP to methylated and unmethylated genomic regions relative to WT-GFP with 1 minute cross-link, n=3-4.
(C) Reducing the cross-linking temperature revealed a marked reduction in R133C-GFP binding at all ChIP regions, relative to WT-GFP, n=4.
Wildtype (WT). Mean % input±SEM plotted. *p<0.05, ** p<0.01, ***p<0.001, unpaired t-test.

Lowering the cross-linking temperature in addition to shortening the cross-linking time (Figure 5.4.3C) yielded different results. At major satellites mutant protein pulled down 0.14% of the input DNA, which was significantly lower than WT-GFP (0.99%) and represented 14% of binding. At L1 LINEs mutant protein pulled down 0.08% of the input DNA, which was significantly lower than WT-GFP (0.54%) and represented 15% of binding. At IAP elements mutant protein pulled down 0.11% of the input DNA, which was significantly lower than WT-GFP (0.72%) and represented 15% of binding. Finally, at the *BDNF* promoter, mutant protein pulled down 0.09% of the input DNA, which was significantly lower than WT-GFP (0.38%) and represented 24% of binding. These more stringent conditions appeared

to expose a deficit in binding to DNA in the mutant protein (relative to WT-GFP) that did not track protein abundance and had not been appreciable previously. It approximated to a 6-7 fold reduction in binding at methylated sites and a 4 fold reduction at the unmethylated site. This reduction was a combination of reduced protein abundance and chromatin binding, as would be the case in patients with the R133C mutation.

5.4.4 ChIP supports the phenotypic severity spectrum in the GFP-tagged allelic series

These results were placed in the context of the EGFP-tagged allelic series for comparison (Figure 5.4.4). The R306C-GFP and T158M-GFP ChIP experiments were done in parallel under the stringent conditions with lower cross-linking temperature. S.Lagger contributed one replicate. The hypothesis was that the amount of R133C-GFP bound to the four regions (as inferred from the percentage of input DNA measured by real time qPCR) would be greater than T158M-GFP. The reason was firstly, because R133C-GFP is more abundant than T158M-GFP and secondly, the DNA binding deficit was predicted to be less severe for R133C-GFP because, by confocal microscopy, the localisation to heterochromatic foci appeared to be more deranged in the *T158M-GFP* brain. The converse was predicted for R306C-GFP, which was more abundant, had a normal pattern of localisation by microscopy and carried a mutation in the domain of *Mecp2* that binds to the NCoR complex (Lyst *et al*, 2013).

Compared to WT-GFP pulling down 0.99% of input at major satellite DNA, the percentage input pulled down by all the mutants was lower: R133C-GFP (0.14%); R306C-GFP (0.38%); and T158M-GFP (0.07%). This equated with 14%, 38% and 7% of WT-GFP binding, respectively. At L1 LINEs, WT-GFP pulled down 0.54%, which was more than R133C-GFP (0.08%), R306C-GFP (0.19%) and T158M-GFP (0.05%). This was the equivalent of 15%, 35% and 9% of WT-GFP binding, respectively. Relative to the percentage of input DNA pulled down by WT-GFP at IAP elements (0.72%), the percentage bound by the mutants was lower: R133C-GFP (0.11%), R306C-GFP (0.24%) and T158M-GFP (0.05%). This corresponded to 15%,

33% and 7% of WT-GFP binding, respectively. The final region assessed was the *BDNF* promoter: compared to 0.38% for WT-GFP, again, all the mutants bound a lower percentage of their input DNA: R133C-GFP (0.09%), R306C-GFP (0.16%) and T158M-GFP (0.05%). This equated with 24%, 42% and 13% of WT-GFP binding, respectively. Therefore it was observed that the amount of R133C-GFP bound to these specific genomic regions under stringent conditions was intermediate between R306C-GFP and T158M-GFP. This was compatible with the hypothesis stated above. The non-significant trend towards reduced R306C-GFP binding at all ChIP regions, relative to WT-GFP was in line with a previous report that showed significantly reduced DNA binding relative to WT-GFP (Heckman *et al*, 2014).

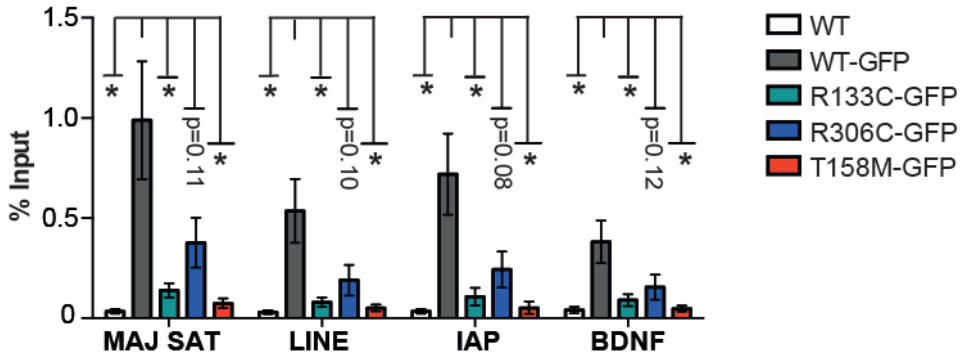


Figure 5.4.4 Significantly reduced binding of GFP-tagged MBD mutants, relative to WT-GFP at all ChIP regions.

Graph displaying GFP ChIP under stringent conditions for the EGFP-tagged allelic series at major satellite DNA, L1 LINEs, IAP elements and the *BDNF* promoter region in male mouse brain. Significantly less of the MBD mutant protein (R133C-GFP and T158M-GFP) is bound to all four sites, relative to WT-GFP. Wildtype (WT). n=3-4. Mean % input±SEM plotted. *p<0.05, unpaired t-test.

In summary, GFP ChIP revealed that R133C-GFP binding to heavily methylated repetitive DNA and an unmethylated site is reduced relative to WT-GFP in both neuronal cells and the male mouse brain. Under standard ChIP conditions in mouse brain, the reduction in binding was 2-3 fold, whereas under more stringent conditions, the reduction was 6-7 fold at methylated sites and 4 fold at the unmethylated *BDNF* promoter region. Placing this in the context of the EGFP-tagged allelic series, significantly less of the two MBD mutants, R133C-GFP and T158M-GFP, was bound at all assayed sites. This result was consistent with the confocal

microscopy findings that showed an abnormal pattern of localisation in these two mutants. In both cases, the deficit in R133C-GFP binding seemed to be less severe, in keeping with the decreased severity of the Rett syndrome phenotype in patients carrying this mutation. The objective was to further refine the DNA binding deficit to place it in the context of known MeCP2 binding capacities.

5.5 Mecp2^{R133C} binding *in vitro* by EMSA.

Thus far, it has been shown that R133C-GFP does not associate perfectly with heterochromatic foci and significantly less R133C-GFP was bound to methylated repetitive DNA and an unmethylated site, relative to WT-GFP. These *in vivo* experiments have the limitation that they are influenced by the reduced protein abundance of MeCP2^{R133C}. It has previously been suggested that MeCP2^{R133C} loses binding specifically to hmC (Mellen *et al*, 2012), however this does not seem compatible with the results presented so far. The goal was to answer the question of whether MeCP2^{R133C} has a sequence-specific DNA binding deficit, where only a subset of MeCP2 binding sites are affected.

EMSAs were employed as an experimental tool to answer this question. In this assay, protein binding to specific radio-labelled oligonucleotide probes can be quantitatively determined. It is possible to modify these oligonucleotides to be methylated or hydroxymethylated and so MeCP2^{R133C} binding to hydroxymethylated DNA could be investigated. Although it has been reported in the literature that MeCP2 binds to hmCpG *in vitro* (Mellen *et al,* 2012; Spruijt *et al,* 2013), this has since been disproved *in vivo* (Gabel *et al*, 2015; Chen *et al*, 2015) in agreement with earlier *in vitro* data (Valinuck *et al,* 2004; Hashimoto *et al,* 2012) and unpublished work from Professor Bird's lab: in fact MeCP2 only binds with high affinity to hmCA (Gabel *et al*, 2015). Additionally, J.Connelly found that MeCP2 binds with greatest affinity to methylation or hydroxymethylation in a CAC context (unpublished data). mCAC is overwhelmingly the most abundant form of mCH in the adult brain (Lister *et al*, 2013) and it has recently been shown that MeCP2 binds to mCA *in vivo* (Gabel *et al*, 2015). For these reasons, it was decided to investigate MeCP2^{R133C} binding to methylation in a CG and CAC context and

hydroxymethylation in a CAC context, as these are the principal binding sites of MeCP2 *in vivo*. J.Connelly's investigation of MeCP2 binding employed a 58-mer corresponding to a region of intron 3 of the *BDNF* locus. This probe contained a single central CpG site that could be altered in sequence context or methylation. The same probe was used so that the experiments were directly comparable (Figure 5.5.1A).

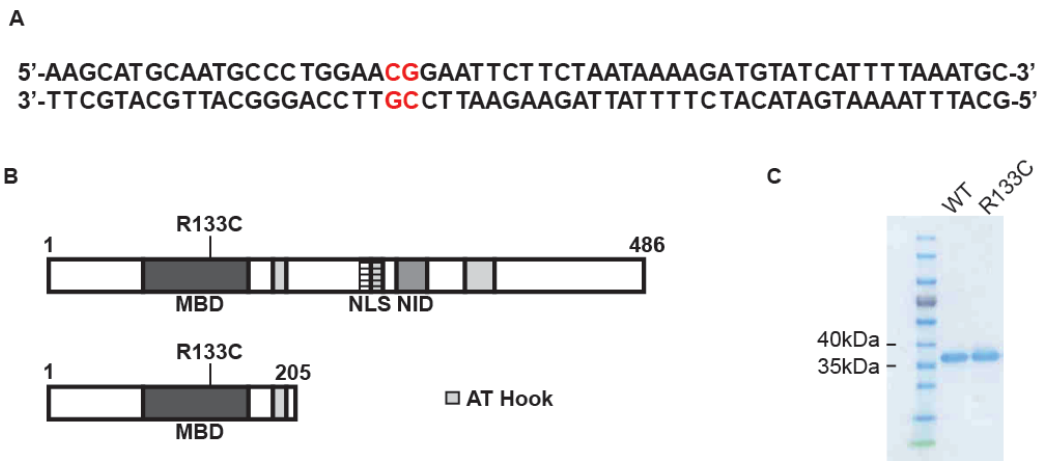


Figure 5.5.1 *In vitro* investigation of DNA binding by EMSA.

(A) The 58mer double stranded oligonucleotide from the *BDNF* promoter IV region used in EMSA's with a single central CpG site indicated in red. This site could be methylated or hydroxymethylated in a CpG or CAC context.
(B) Schematic representation of MeCP2 protein isoform e2. The major functional domains are shown: the methylated DNA binding domain (MBD) with position of the R133C missense mutation, the nuclear localisation signal (NLS) and the NCoR Interaction Domain (NID), along with three putative AT hooks. For comparison, the (1-205) polypeptide used in EMSA experiments is represented.
(C) SDS-PAGE showing MeCP2(1-205) (WT) and MeCP2(1-205)^{R133C} (R133C). The polypeptide was expressed in *E.coli* then purified by two steps. There is only one major migrating species between 35 and 40 kilodaltons with no evidence of polypeptide degradation.

5.5.1 Expression & purification of MeCP2(1-205)

The 1-205 fragment of human MeCP2 was employed for ease of purification and use in the EMSA. This fragment contains the N-terminal portion of the protein as well as the MBD and the first AT hook of MeCP2 (Figure 5.5.1B). A suitable recombinant polypeptide expression vector containing the 1-205 fragment with a C-terminal histidine tag was available in the lab. The R133C mutation was engineered by site-

directed mutagenesis. Recombinant polypeptide was expressed in *E.coli* then purified in two steps. Firstly, it was extracted from bacterial cell lysates by its histidine tag using nickel-coated sepharose beads. Secondly, the polypeptide was eluted from a sulphopropyl sepharose column by cation exchange. This two-step process resulted in minimal degradation. An example of the recombinant polypeptide used in the EMSAs is shown in Figure 5.5.1C. There was only one major migrating species corresponding to MeCP2(1-205).

5.5.2 Wildtype MeCP2(1-205)^{R133C} binding to mCpG

MeCP2(1-205)^{R133C} binding to the *BDNF* probe that was singly methylated at the central CpG, relative to MeCP2(1-205), is shown in Figure 5.5.2. Polypeptide concentration was decreased while the concentration of competitor DNA and radio-labelled probe were kept constant. The percentage of probe shifted was quantified using Image J software and plotted on a binding curve. Interestingly, MeCP2(1-205)^{R133C} binding to a single mCpG dyad was indistinguishable from MeCP2(1-205) binding. For example, at the highest polypeptide concentration MeCP2(1-205)^{R133C} shifted 66% of the probe and MeCP2(1-205) shifted 64%.

5.5.3 Reduced MeCP2(1-205)^{R133C} binding to mCAC

In contrast, when binding to mCAC was measured, MeCP2(1-205)^{R133C} was easily distinguishable from MeCP2(1-205) (Figure 5.5.3). At the maximum polypeptide concentration measured, MeCP2(1-205)^{R133C} shifted significantly less of the oligonucleotide probe compared to MeCP2(1-205) (27% vs 43%). The same was true for the lower concentrations of polypeptide.

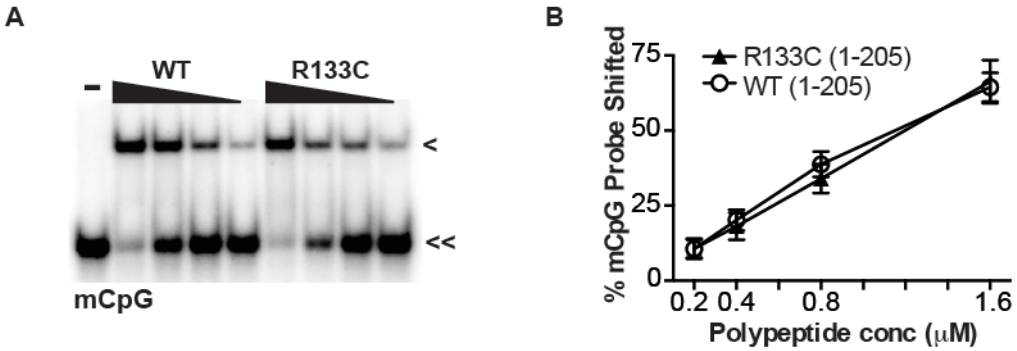


Figure 5.5.2 MeCP2(1-205)^{R133C} binding to mCpG is indistinguishable from wildtype.

(A) Representative gel showing mCpG probe bound and shifted by polypeptide (<) in comparison to free probe (<<) for decreasing concentrations (triangle) of MeCP2(1-205) (WT) and MeCP2(1-205)^{R133C} (R133C). - no protein.
(B) Quantification graph of EMSAs (n=3) showing that the percentage of mCpG probe shifted was highly similar for the two genotypes. Mean percentage of probe shifted is plotted±SEM.

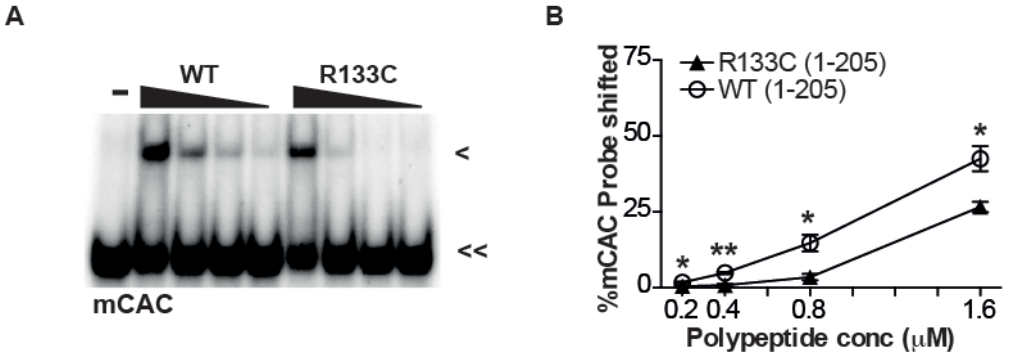


Figure 5.5.3 MeCP2(1-205)^{R133C} binding to mCAC is reduced relative to wildtype.

(A) Representative gel showing mCAC probe bound and shifted by polypeptide (<) in comparison to free probe (<<) for decreasing concentrations (triangle) of MeCP2(1-205) (WT) and MeCP2(1-205)^{R133C} (R133C). - no protein.
(B) Quantification graph of EMSAs (n=3) showing that the percentage of mCAC probe shifted was significantly reduced for MeCP2(1-205)^{R133C}. Mean percentage of probe shifted is plotted±SEM. *p<0.05, ** p<0.01, unpaired t-test.

5.5.4 Reduced MeCP2(1-205)^{R133C} binding to hmCAC

This differential binding was also apparent with the hmCAC probe (Figure 5.5.4). MeCP2(1-205)^{R133C} shifted significantly less of the oligonucleotide probe at all concentrations except the maximum (28% vs 40%), although there was a trend towards this. There was negligible binding of MeCP2(1-205) and MeCP2(1-205)^{R133C} to hmCpG probes (data not shown).

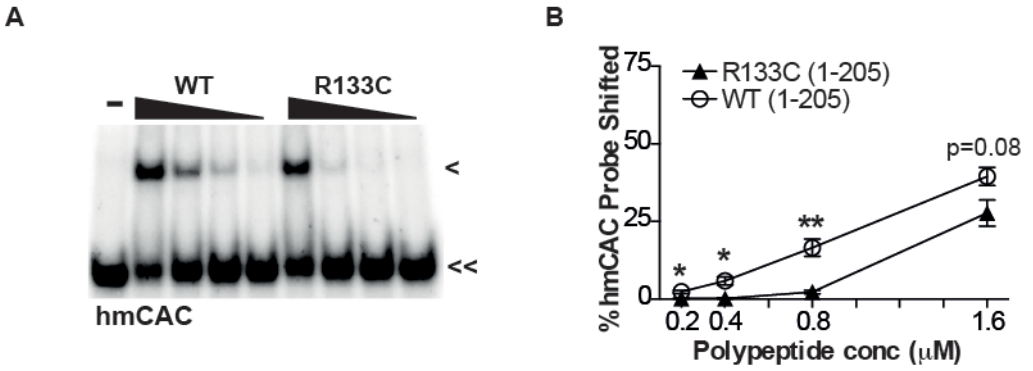


Figure 5.5.4 MeCP2(1-205)^{R133C} binding to hmCAC is reduced relative to wildtype.

(A) Representative gel showing hmCAC probe bound and shifted by polypeptide (<) in comparison to free probe (<<) for decreasing concentrations (triangle) of MeCP2(1-205) (WT) and MeCP2(1-205)^{R133C}(R133C). - no protein.
(B) Quantification graph of EMSAs (n=3) showing that the percentage of hmCAC probe shifted was significantly reduced for MeCP2(1-205)^{R133C} at all but the highest measured protein concentrations. Mean percentage of probe shifted is plotted±SEM. *p<0.05, ** p<0.01, unpaired t-test.

Therefore measuring MeCP2(1-205)^{R133C} DNA binding using EMSAs revealed a specific deficit where the mutant polypeptide had reduced binding to methylation and hydroxymethylation in a CAC sequence context, relative to wildtype. In contrast the binding to methylation in a CpG context was indistinguishable from wildtype.

5.5.5 Reduced MeCP2(77-167) binding to mCpG

Although this result was intriguing, an identical experiment by J.Connelly using a smaller fragment of MeCP2(77-167), corresponding to the MBD, yielded different results (Figure 5.5.5). He found that the MeCP2(77-167)^{R133C} polypeptide had virtually no binding to the mCpG probe (2% shift vs 68% wildtype shift). This raised the question of whether the specific binding deficit to CAC (hydroxy)methylation was an artefact of using a particular fragment of MeCP2 in this *in vitro* technique. The objective was now to investigate the possibility of a specific loss of binding to (hydroxy)methylation in a CAC context *in vivo*.

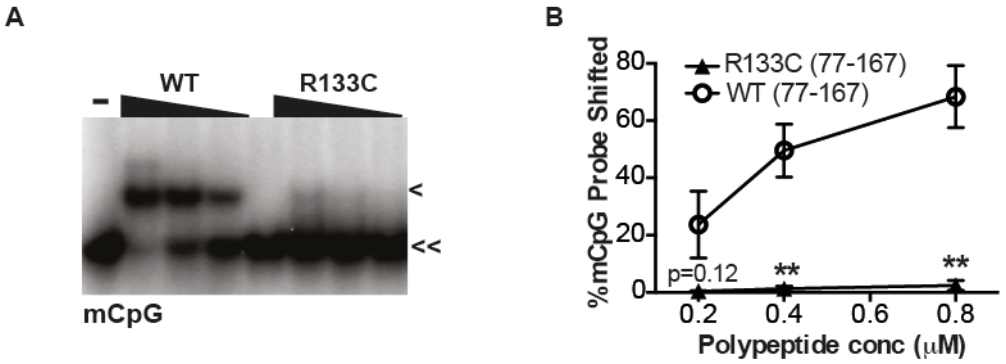


Figure 5.5.5 MeCP2(77-167)^{R133C} binding to mCpG is virtually abolished.

(A) Representative gel showing mCpG probe bound and shifted by polypeptide (<) in comparison to free probe (<<) for decreasing concentrations (triangle) of MeCP2(77-167) (WT) and MeCP2(77-167)^{R133C} (R133C). - no protein.
(B) Quantification graph of EMSA's (n=3) showing that the percentage of mCpG probe shifted was significantly reduced for MeCP2(77-167)^{R133C} at higher polypeptide concentrations tested. Mean percentage of probe shifted is plotted±SEM. ** p<0.01, unpaired t-test.
Data reproduced with permission from J.Connelly.

5.6 R133C-GFP binding to mCAC *in vivo*

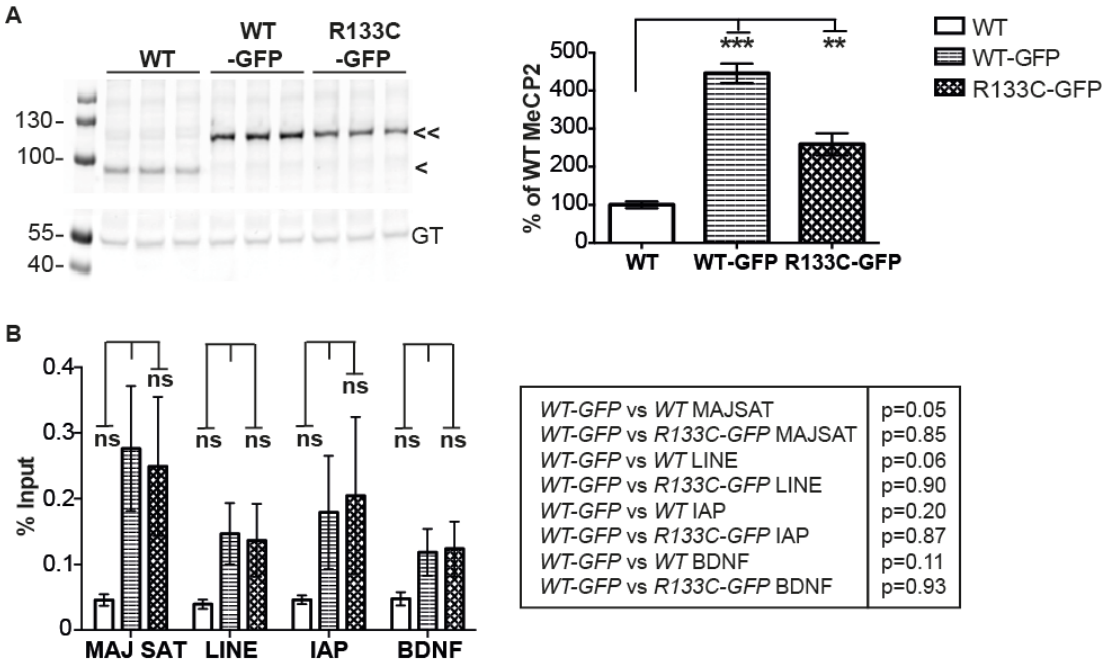
The idea that the reduced severity in Rett syndrome caused by the R133C mutation could be attributable to a loss of one of the specific binding capacities of MeCP2 was fascinating. The goal was to search for evidence in the mouse brain that would corroborate the *in vitro* findings. There were three main approaches to this: (1) repetition of the GFP ChIP experiment described above in tissue with low levels of mCH; (2) GFP ChIP in new areas of the genome enriched for mCpG over mCAC (and *vice versa*); and (3) bisulfite sequencing of ChIP DNA. The focus was on detecting differences in binding to mCpG sequences and mCAC sequences and not on binding to hmCAC sequences. The reason was that the level of hmCAC is very low by current estimates (Lister *et al*, 2013) and so associated MeCP2 binding is unlikely to be biologically significant.

5.6.1 ChIP in low mCH tissue

Repeating the GFP ChIP on repetitive sequences and the *BDNF* promoter region in a tissue with low levels of mCH was the first strategy employed. The hypothesis was that the percentage input pulled down by R133C-GFP would increase relative to the WT-GFP signal when there was only mCpG to bind to. Most adult peripheral tissues

have low levels of mCH (Guo *et al*, 2014). The liver was selected for GFP ChIP initially. Unfortunately the low level of MeCP2 expression in the liver meant that the EGFP-tagged wildtype protein was barely detectable over the background negative control signal (data not shown). Another strategy was to use neonatal brain tissue. Levels of mCH don't increase until mice are a couple of weeks old (Lister *et al*, 2013), so brains were harvested at postnatal day 0-2 and used for GFP ChIP.

Importantly, at this age the reduction in R133C-GFP expression, relative to WT-GFP, was also 40%, as found in both the neuronal cell model and the adult mouse brain, and so the ChIP would be comparable (Figure 5.6.1A). R133C-GFP expression was 259% untagged wildtype levels, whereas WT-GFP was 446% of untagged wildtype levels by quantitative western blot (Figure 5.6.1A). The GFP ChIP revealed that R133C-GFP pulled down 0.29% of the input at major satellites, compared to 0.28% for WT-GFP (Figure 5.6.1B). At L1 LINEs, it was 0.14% and 0.15% for R133C-GFP and WT-GFP, respectively. At IAP elements, the percentage of input DNA pulled down was 0.20% and 0.18% for R133C-GFP and WT-GFP, respectively. Finally, at the *BDNF* promoter, R133C-GFP pulled down 0.12%, compared to 0.12% for WT-GFP. There were no significant differences in the percentage of input DNA pulled down by R133C-GFP and WT-GFP at any of regions assayed. This was compatible with the hypothesis that the R133C-GFP ChIP signal would increase relative to WT-GFP in the neonatal brain in the absence of mCH, if R133C-GFP had a specific deficit in binding to mCAC. Although this appeared convincing, there was no significant difference in ChIP signal between the positive control (WT-GFP brain) and the negative control (wildtype brain). The variability of the results was high. Furthermore, the overall percentage input was low. These three points indicate that the abundance of WT-GFP was too low at this stage of development to detect meaningful differences in relative binding with confidence.



WT-GFP vs *WT* MAJSAT	p=0.05
WT-GFP vs *R133C-GFP* MAJSAT	p=0.85
WT-GFP vs *WT* LINE	p=0.06
WT-GFP vs *R133C-GFP* LINE	p=0.90
WT-GFP vs *WT* IAP	p=0.20
WT-GFP vs *R133C-GFP* IAP	p=0.87
WT-GFP vs *WT* BDNF	p=0.11
WT-GFP vs *R133C-GFP* BDNF	p=0.93

Figure 5.6.1 GFP ChIP in the male neonatal brain is compatible with a specific deficit in R133C-GFP binding to methylation in a CH context.

(A) Left, western blot of MeCP2 (<) and MeCP2^{EGFP} (<<) expression in wildtype (WT), *WT-GFP* and *R133C-GFP* neonatal brain. GT, Gamma tubulin. Right, graph showing a significant decrease in R133C-GFP abundance relative to WT-GFP, n=3. Mean % WT level±SEM plotted. ** p<0.01, ***p<0.001, one-way ANOVA with Tukey's multiple comparisons test.
(B) There was no significant difference in the % input pulled down by R133C-GFP and WT-GFP by GFP ChIP in neonatal brain, n=3-4. Mean % input±SEM plotted. Unpaired t-test, p values for non-significant comparisons are detailed in adjacent table.

5.6.2 No change in R133C-GFP ChIP signal at mCAC rich sites

The second strategy to investigate a specific deficit in R133C-GFP binding was to perform ChIP at selected sites in the genome that were enriched for mCpG and depleted of mCAC, "mCpG rich" sites or conversely, enriched for mCAC and depleted of mCpG "mCAC rich" sites. The Cytochrome p450 gene locus is a good example of a mCpG rich region in brain. Cytochrome p450 proteins are enzymes that utilise a haem cofactor in the metabolism of endogenous and exogenous chemicals (Danielson, 2002). They are highly expressed in the liver where they are essential for drug metabolism. In brain, the locus is coated in a blanket of CpG methylation, but contains regions where CH methylation and hydroxymethylation are virtually absent

(Figure 5.6.2A). Three of the cytochrome p450 genes – *Cyp3a25*, *Cyp3a57* and *Cyp3a59* were selected for ChIP. The mCAC rich sites were found by G.Schweikert. Regions of the genome were ranked by level of mCpG, mCAC and hmCpG (taking into account percentage methylation and number of sites) and outlier regions were chosen that were particularly enriched for mCAC and not mCpG (or hydroxymethylation). These regions were the Epidermal Growth Factor like domain 7 (*Egfl7)* gene, Olfactory Receptor 325 (*Olfr325*) gene and Mitotic arrest deficient 1|1 (*Mad1|1*) gene. Two regions that were devoid of methylation were included as controls: the Carboxypeptidase A6 (*Cpa6*) gene and Trans-acting T cell Specific Transcription Factor (*Gata3*) gene.

The selected regions are represented in Figure 5.6.2B, where they are plotted for mCAC and mCpG level. These levels represent the region spanned by the real time qPCR primers (approximately 190bp), as well as the 5' and 3' flanking 150bp. This was to exclude the possibility that the methylation context was significantly different over neighbouring regions of the genome, which could also be immunoprecipitated by chromatin fragments of $MeCP2^{EGFP}$, averaging 300bp in size.

The hypothesis was that there would be less of a reduction in the R133C-GFP ChIP signal relative to the WT-GFP signal at the mCpG rich sites, but not the mCAC rich sites if there was a specific deficit in R133C-GFP binding to mCAC. Figure 5.6.2C shows the results. The real time qPCR primers were run on the same ChIP DNA that was derived from the GFP ChIP under the stringent conditions in Figure 5.4.3C. Starting with the mCpG rich sites, the percentage of input DNA for WT-GFP at *Cyp3a25* was 0.70%. This was significantly higher than the percentage of input DNA for R133C-GFP (0.11%). At the second methyl-CG rich site, *Cyp3a57*, WT-GFP pulled down a significantly higher percentage of the input DNA than R133C-GFP (0.56% vs 0.11%). For the third methyl-CG rich site, *Cyp3a59* the picture was the same with WT-GFP pulling down a significantly higher percentage of the input DNA than R133C-GFP again (0.64% vs 0.11%). So for the mCpG rich sites, *Cyp3a25*, *Cyp3a57* and *Cyp3a59*, the percentage of WT-GFP ChIP signal for R133C-GFP was 16%, 20% and 17%, respectively. These figures are similar to the percentage of WT-

GFP signal for the mutant protein at repetitive elements and the *BDNF* promoter under the same ChIP conditions.

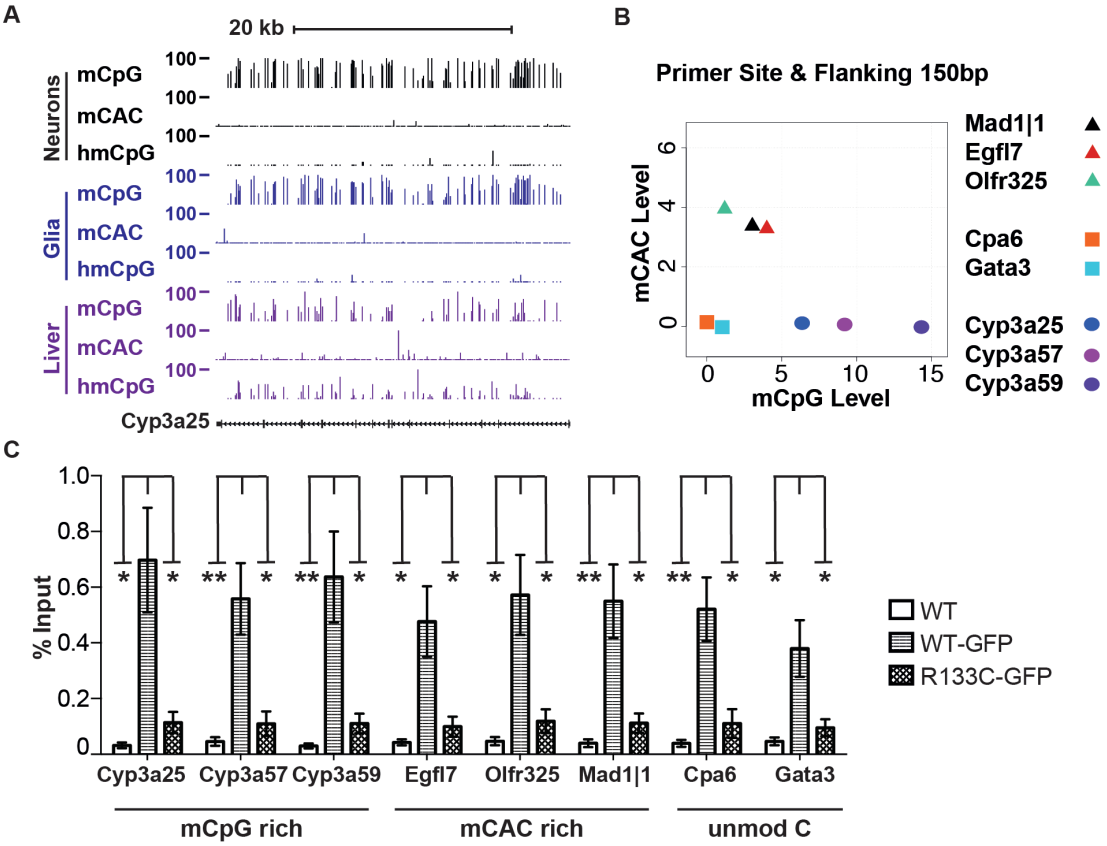


Figure 5.6.2 GFP-ChIP at mCAC regions in male mouse brain does not support the hypothesis that R133C-GFP has a specific DNA binding deficit.

(A) Schematic of (hydroxy)methylation context of representative Cytochrome p450 gene locus, *Cyp3a25*. Produced with permission from S.Lagger's unpublished TAB-seq data in sorted neurons, glia and liver for comparison. The site is devoid of mCAC and hmCpG in brain.
(B) Plot of methylation level for selected mCpG rich (circles), mCAC rich (triangles) and control regions (squares). Level represents 190bp plus flanking 150bp up and downstream.
(C) Graph of GFP ChIP showing that there is no sequence or methylation dependence on the relative percentage input for R133C-GFP compared to WT-GFP, n=4. Mean % input±SEM plotted. *p<0.05, ** p<0.01, unpaired t-test.

Moving on to the mCAC rich sites, the percentage of input DNA for WT-GFP immunoprecipitated at *Egfl7* was 0.48% (Figure 5.6.2C). This was significantly higher than the percentage of input DNA for R133C-GFP (0.10%). At the second mCAC rich site, *Olfr325*, WT-GFP pulled down a significantly higher percentage of the input DNA than R133C-GFP (0.57% vs 0.12%). For the third mCAC rich site,

Mad1|1 the picture was the same with WT-GFP pulling down a significantly higher percentage of the input DNA than R133C-GFP again (0.55% vs 0.11%). So for the mCAC rich sites, *Egfl7*, *Olfr325* and *Mad1|1*, the percentage of WT-GFP ChIP signal for R133C-GFP was 21%, 21% and 20%, respectively. These results were almost identical to those for the mCpG rich sites described above. There was no relative reduction in R133C-GFP binding over the mCAC rich sites, compared to the mCpG rich sites. Further to this, the R133C-GFP signal at the control regions, *Gata3* and *Cpa6*, was a very similar percentage of the WT-GFP ChIP signal. At the first of the two sites with no cytosine methylation, *Gata3*, R133C-GFP pulled down 0.09% of the input, which was significantly less than WT-GFP (0.38%) and represented 24% of the ChIP signal. At the second site, *Cpa6* R133C-GFP also pulled down a significantly lower percentage of the input compared to WT-GFP (0.11% vs 0.52%). This was approximately 21% of the EGFP-tagged wildtype signal.

To summarise, the R133C-GFP percentage of input DNA relative to that pulled down by WT-GFP was constant across all the sites, including the control regions with no methylation. Therefore, by this strategy of performing ChIP at specific sites in the genome, enriched for methylation in one sequence context over another, there was no evidence to support the hypothesis that R133C-GFP binding to mCAC was specifically impaired.

5.6.3 R133C-GFP ChIP DNA is enriched for mCpG but not mCAC

The third strategy to test the specific binding deficit hypothesis was to do bisulfite sequencing on DNA immunoprecipitated by R133C-GFP. Using this method, the reduced relative abundance of the mutant protein would be less contributory to the experiment. By comparing immunoprecipitated DNA to input DNA, the bisulfite sequencing would reveal whether (1) the DNA bound by WT-GFP was enriched for mCpG (which would be the prediction for a methyl CpG binding protein), (2) if the DNA bound by WT-GFP was enriched for mCAC (a recently reported phenomenon) and (3) if the DNA bound by R133C-GFP was enriched for mCpG but not for mCAC (as the specific binding deficit hypothesis would predict).

Two genomic sites were chosen from a recent publication, which showed that MeCP2 binds to mCH (Guo *et al*, 2014). One of these sites contained a CpG pair every 28bp with a low level of baseline methylation: the Fibroblast Growth Factor B (*Fgf1b*) gene. The second site contained a CAC sequence every 63bp with a low level of baseline methylation: the Brain Specific Angiogenesis Inhibitor 3 (*Bai3*) gene. WT-GFP and R133C-GFP GFP ChIP and input DNA were concentrated, treated with sodium bisulfite, then PCR amplified at these two regions. The PCR product was transformed into *E.coli* and the resulting colonies were screened for presence of the PCR insert then sequenced to reveal the methylation pattern. The location of the bisulfite amplicons is shown in Figure 5.6.3.

Figure 5.6.3A shows the results of the bisulfite sequencing of the MeCP2^{EGFP}-immunoprecipitated DNA from the *Fgf1b* locus. The proportion of mCpG in the input DNA was similar for *WT-GFP* and *R133C-GFP* brain (22% and 15%, respectively). The DNA immunoprecipitated by WT-GFP was significantly enriched for mCpG in comparison to the input DNA (44% vs 22%, respectively) as would be expected for a methyl CpG binding protein. Similarly, the DNA immunoprecipitated by R133C-GFP was significantly enriched for mCpG in comparison to the input (40% vs 15%, respectively). Therefore, independent of protein abundance, R133C-GFP showed evidence of binding to mCpG *in vivo*, with a comparable enrichment to WT-GFP over input DNA.

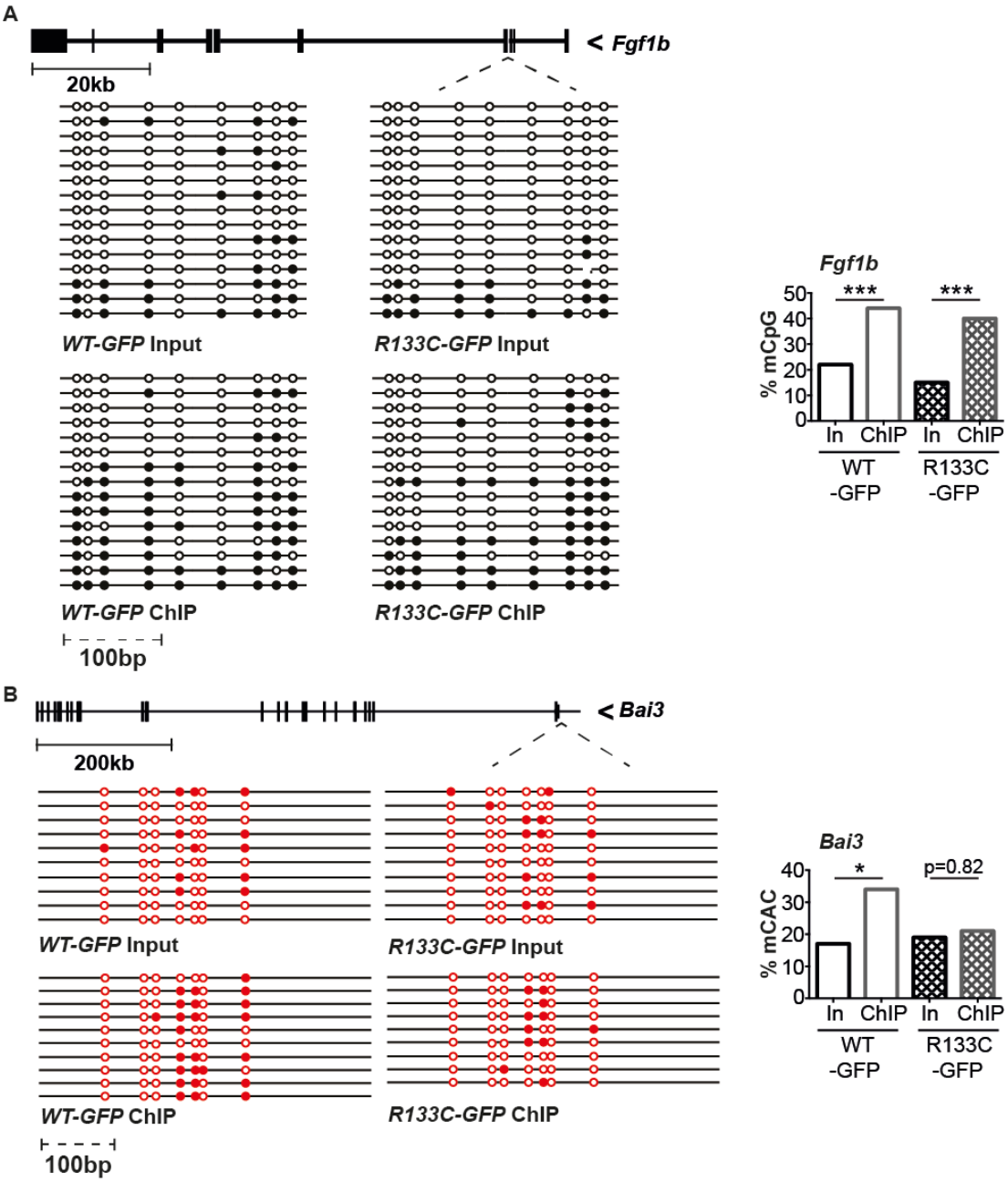


Figure 5.6.3 Bisulfite sequencing of ChIP DNA supports a specific deficit in binding to mCAC for R133C-GFP.

(A) Bisulfite sequencing of DNA immunoprecipitated by WT-GFP and R133C-GFP at the *Fgf1b* locus is significantly enriched for mCpG over input.
(B) Bisulfite sequencing of DNA immunoprecipitated by WT-GFP but not R133C-GFP at the *Bai3* locus is significantly enriched for mCAC over input.
Maps of the regions are shown with scale bars (solid scale) and location of the bisulfite amplicons. Bisulfite sequencing is shown with scale bars (dashed scale): the methylation status of each site is represented by circles: black open are unmethylated CpG; black filled are mCpG; red open are unmethylated CAC; red filled are mCAC. Each line represents a single clone. Graphs represent proportion of methylated residues at both loci. In, input.
* $p<0.05$, *** $p<0.001$, Fisher's exact test.

Figure 5.6.3B shows the results of the bisulfite sequencing of the MeCP2^{EGFP}-immunoprecipitated DNA from the *Bai3* locus. The proportion of mCAC in the input DNA was similar for *WT-GFP* and *R133C-GFP* brain (17% and 19%, respectively). The DNA immunoprecipitated by WT-GFP was significantly enriched for mCAC in comparison to the input DNA, which was compatible with the recent finding that MeCP2 binds mCA (Gabel *et al*, 2015). The proportion of mCAC in the *Bai3* amplicon was 34% in the DNA pulled down by WT-GFP. In contrast, the DNA immunoprecipitated by R133C-GFP was not significantly enriched for mCAC in comparison to the input. The proportion of mCAC in the DNA pulled down by R133C-GFP was 21%. Therefore, independent of protein abundance, R133C-GFP showed evidence of a deficit in binding to mCAC *in vivo*, in comparison to WT-GFP, with no enrichment for this modification in immunoprecipitated DNA over input DNA.

Taken together, these results provide *in vivo* evidence that there is a specific deficit in R133C-GFP binding to methylation in a CAC context. This supports the hypothesis generated by the EMSA DNA-binding assay *in vitro*.

5.6.4 R133C-GFP ChIP-seq is inconclusive

The results of bisulfite sequencing ChIP DNA at individual loci suggested that it was possible to detect a specific reduction in R133C-GFP affinity for mCAC. The next logical step was to look for this pattern genome wide. R133C-GFP ChIP and input samples from whole brain were taken for library preparation by S.Lagger in parallel with ChIPs and inputs for sorted neurons and glia. High throughput sequencing was performed at the Wellcome Trust Sanger Institute, Cambridge and data was analysed and combined with published bisulfite-seq data for frontal cortical neurons (Lister *et al*, 2013) by S.Webb. Peaks of R133C-GFP binding over input were identified and the methylation density with sequence context was plotted over 8 kilobases centred around this peak (averaged over 20 nucleotide bins). mCA density was considered to capture all potential mCAH binding, however mCAC will be overrepresented as it is

the most abundant form of mCA in the brain (S.Lagger, unpublished). With reference to Figure 5.6.4A, it can be seen that in neurons and glia, WT-GFP bound at regions of increased mCpG and mCA. Contrary to expectation, R133C-GFP also bound at regions of mCA, but not mCpG. However, if just the top 30,000 peaks are considered (when peaks are ranked by confidence, or p value) the opposite pattern is seen, with a comparable mCpG density at peaks of R133C-GFP and WT-GFP in neurons but reduced mCA density at R133C-GFP peaks compared to WT-GFP peaks in neurons and glia (Figure 5.6.4B). These results are inconsistent. Furthermore, there does not appear to be increased mCpG density at regions of WT-GFP binding in the top 30,000 peaks in neurons (Figure 5.6.4B), which was unexpected for a methyl CpG binding protein.

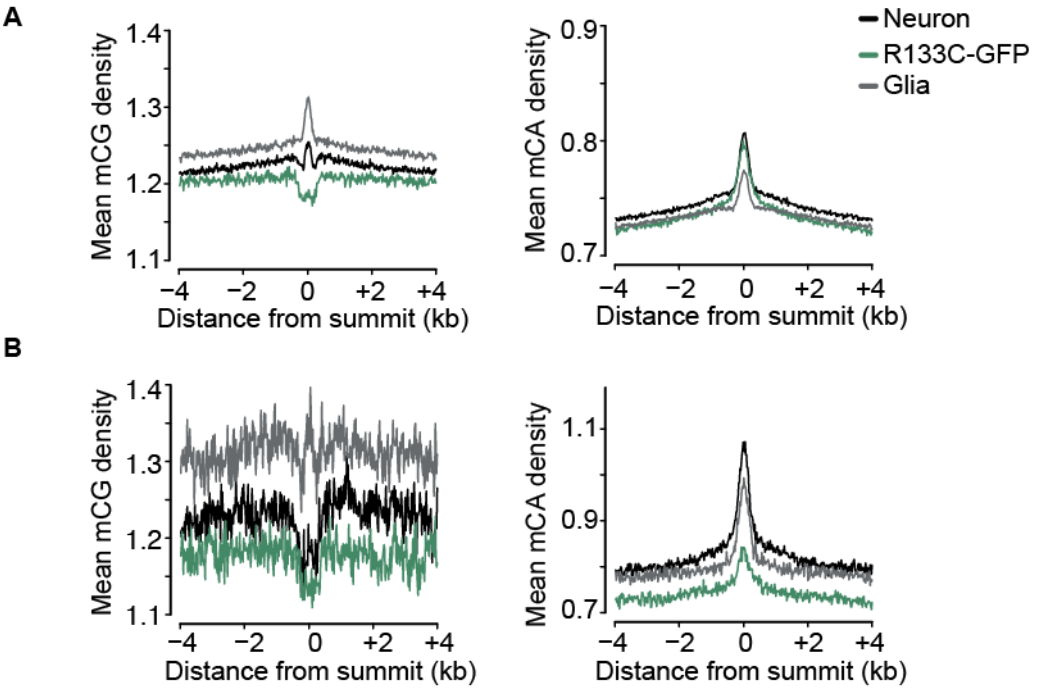


Figure 5.6.4 Inconclusive results from R133C-GFP ChIP-seq in whole brain.

(A) Plots of mean methylation density at all identified peaks of MeCP2^{EGFP} binding in R133C-GFP whole brain and WT-GFP sorted neurons and glia. R133C-GFP appears to bind at regions of similar mCA density but reduced mCpG density, in comparison to WT-GFP in neurons and glia.
(B) In contrast, the same plots of just the top 30,000 summits show the opposite pattern with R133C-GFP binding at regions of similar mCpG density but reduced mCA density, in comparison to WT-GFP in neurons and glia. There was no increase in mCpG density at peaks of WT-GFP binding in neurons.

5.7 Gene expression in *R133C-GFP* cerebellum

The biological consequence of losing one of the binding capacities of MeCP2 is not clear, as the function of MeCP2 is poorly understood. Many groups have investigated gene expression in the context of MeCP2 deficiency by gene expression microarrays (Tudor *et al*, 2002; Jordan *et al*, 2007; Urdinguio *et al*, 2008; Charhour *et al*, 2008; Ben Shachar *et al*, 2009; Samaco *et al*, 2012). The result appeared to be many small changes in gene expression - some genes were upregulated, some were downregulated. This was unexpected from the loss of a proposed transcriptional repressor. Until recently, there was no unifying feature of the genes that were upregulated in the absence of *Mecp2*. Within the last year, it was reported that in MeCP2 deficiency, there is an upregulation of long (>100kb) genes and that this was dependent on binding to mCA (Sugino *et al*, 2015; Gabel *et al*, 2015, Chen *et al*, 2015). So, if a long gene has a high level of mCA, it is more likely to be upregulated in the absence of MeCP2. If $MeCP2^{R133C}$ only loses the capacity to bind to mCA(C), then the hypothesis would be that the upregulation of long genes would be exaggerated by this mutation.

This was investigated with gene expression microarrays using cerebellar tissue from *WT-GFP* and *R133C-GFP* adult male mice. Wildtype and *Mecp2-null* mice were included for comparison. The cerebellum was selected as firstly, it is an accessible region; secondly, it is made predominantly of one cell type – granule cells and so this would largely control for cell type-specific effects; and thirdly, cerebellar dysfunction is relevant to the motor aspects of the Rett syndrome phenotype. Adult mice were culled at six weeks in order to be able to use *Mecp2-null* controls, which were still alive at this time point, and in effort to capture primary effects of Rett syndrome, prior to symptom onset in *R133C-GFP* mice. Total RNA was extracted from three cerebella for each genotype. Reverse transcription using an oligo(dT) primer generated cDNA, the second cDNA strand was synthesised and then RNA libraries were transcribed *in vitro*. This *in vitro* transcription involved biotinylation of the RNA, so it could then directly hybridise with streptavidin-Cy3 and be detected following association with Illumina microarray probes.

5.7.1 Handful of significantly altered genes

Figure 5.7.1 shows the results of the gene expression microarray analysis. This data was analysed by S.Webb using the Bioconductor Limma package. Figure 5.7.1A shows principal component analysis of the gene expression. This form of analysis defines the two greatest sources of variation within the samples, assigns them a numerical value and plots each individual data set according to these sources of variation. In this way, it can be determined whether the animals cluster by genotype, microarray chip or relatedness (some of the mice included in this study were littermates). The mice did not cluster by genotype or by relatedness. The *WT-GFP* mice seemed to be outliers according to the principal component analysis. This was an indication that the pattern of gene expression in these animals contributed the greatest source of variation in the experiment.

A second way of analysing variation in the data set was to create a heatmap of Euclidean distances between animals' expression values across all probes, followed by hierarchical clustering to look for patterns of similarity. The distance between the values is indicated by colour with paler colour indicating larger distance and dissimilarity. An example is shown in Figure 5.7.1B where the genotype of the samples has been considered. This analysis indicates that the animals do not cluster according to genotype, although the *R133C-GFP* animals seem to cluster with the *Mecp2-null* animals and the six wildtype animals (untagged and EGFP-tagged) cluster together. Similar plots were generated indicating that the animals did not cluster by litter or by the particular microarray chip used to analyse their RNA (data not shown).

These two plots combined would have revealed whether the *R133C-GFP* mouse had a unique pattern of gene expression, causing individual animals with this genotype to cluster together, appearing distinct from the other genotypes. There was no evidence that this was the case.

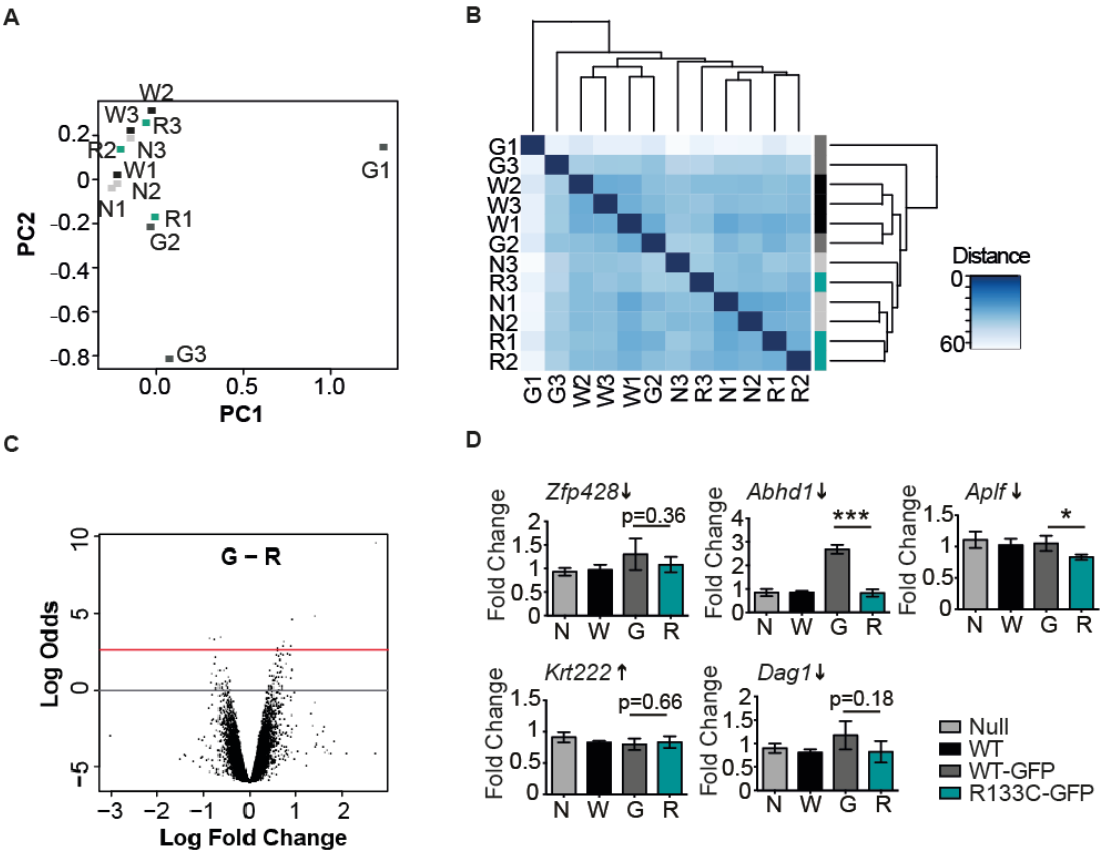


Figure 5.7.1 Microarray analysis of cerebellar gene expression in *R133C-GFP* compared to *WT-GFP* mice revealed only two genes that were significantly altered.

(A) PCA plot of variance in gene expression in wildtype (W1-3), *Mecp2-null* (N1-3), *WT-GFP* (G1-3) & *R133C-GFP* (R1-3) cerebella shows no clustering by genotype and that *WT-GFP* animals were outliers when the two leading source of variance were considered.
(B) Heatmap of Euclidean distance between samples shows that wildtype (tagged and untagged) mice clustered together and *Mecp2-null* and *R133C-GFP* mice clustered together. Paler shades indicate greater distance between samples.
(C) Volcano plot of *WT-GFP* & *R133C-GFP* comparison shows hundreds of subtly up- and downregulated genes, with very few significant changes. Red line indicates p<0.1; grey line indicates the cut-off for changes arising by chance; black dots are individual genes.
(D) Graphs showing relative transcript level in five of six significantly altered genes from (C) by qPCR of cerebellar cDNA. Only downregulation of *Abhd1* & *Aplf* was validated. Mean fold change relative to *Cyclophilin A*±SD plotted. Arrows indicate direction of expected change in *R133C-GFP* indicated by microarray. * p<0.05, *** p<0.001, unpaired t-test.

Volcano plots were used to obtain an impression of the overall pattern of gene expression change between the genotypes. This type of plot shows the relative change in expression of individual genes with a direct comparison of two genotypes (the fold change) and the confidence held in those changes (the log odds), where zero on the y-axis represents a 50:50 chance of expression change. A representative plot is

shown in Figure 5.7.1C comparing *WT-GFP* with *R133C-GFP* data. This plot shows that the expression level of hundreds of genes is different between the *WT-GFP* and *R133C-GFP* mice. Genes are both up and down-regulated in the *R133C-GFP* compared to the *WT-GFP* mice. The fold changes in gene expression were small and only a handful of these changes met statistical significance.

There were only six genes that showed significantly ($p<0.05$) altered expression between the *WT-GFP* and *R133C-GFP* animals: *Abhd1*, *Krt222*, *Mybpc3*, *Dag1*, *Aplf* and *Zfp428*. Only two of five tested were validated by qPCR (one primer pair did not work; qPCR kindly run by S.Lagger). These were Aprataxin and Polynucleotide Kinase-like factor (*Aplf*), encoding a protein involved in DNA double strand break repair (Iles *et al*, 2007) and Lung alpha/beta hydrolase domain gene 1 (*Abhd1*), encoding a protein containing an alpha/beta hydrolase catalytic domain, widely involved in enzymatic reactions (Edgar & Polak, 2002). Both genes were downregulated in *R133C-GFP* animals (Figure 5.7.1D).

5.7.2 Small upregulation of long genes

The data was analysed for changes in expression of long genes to find out if there was an exaggerated effect in the *R133C-GFP* cerebellum, in accordance with the hypothesis that there was a specific deficit in $MeCP2^{R133C}$ binding to CA(C) methylation and that this binding capacity was the basis of MeCP2 regulation of long genes (Gabel *et al*, 2015). Figure 5.7.2A plots fold change in gene expression against gene length for the *R133C-GFP*/*WT-GFP* comparison. There was a small upregulation of long genes (>100kb) in the *R133C-GFP* cerebellum relative to *WT-GFP*. The magnitude of this change, however was less than 1.1-fold. There also appeared to be a downregulation of short genes (<10kb) and so there was a clear length-dependent relationship in the perturbation of gene expression in the *R133C-GFP* mouse.

Figure 5.7.2B shows the *Mecp2-null*/wildtype comparison and the *R133C-GFP*/wildtype comparison on the same graph. Here *R133C-GFP* data is being compared to untagged wildtype MeCP2, so any potential effects of EGFP-tagging the protein are not being controlled for, however, the overall protein abundance

should be comparable in the two genotypes so this reveals the sole effect of the R133C mutation. Surprisingly, the *Mecp2-null* cerebellum showed minor upregulation of long gene expression in this experiment. Furthermore, there was little evidence of a length-dependent change in gene expression in the *R133C-GFP* cerebellum with this comparison. There was still a small degree of downregulation of short (<10kb) genes, but no upregulation of long genes. The explanation for this was apparent in the *WT-GFP*/wildtype comparison. Ideally, the gene expression pattern of these two controls would be indistinguishable. In fact, there is a clear length-dependent relationship, showing a small downregulation of long genes (Figure 5.7.2C). This may be due to the relative overexpression of tagged MeCP2 (Figure 5.2.2). Therefore the relative upregulation seen in *R133C-GFP* cerebellum is attributable to the derangement in the *WT-GFP* gene expression, relative to wildtype. Cytosine methylation was plotted against gene length using S.Lagger's unpublished mouse cerebellar bisulfite-SEQ data. In order that this analysis remained consistent with previously published data (Gabel *et al*, 2015), mCA level and not mCAC level was plotted. However, mCAC will be overrepresented in this plot, given that mCAC is the major form of mCA in the brain (Guo *et al*, 2014), representing approximately 50% of mCAH (S.Lagger, unpublished data). Additionally, the function of the lower affinity binding of MeCP2 to mCAT (J.Connelly, unpublished data) has not been explored and will be captured by this analysis. It can be seen in Figure 5.7.2D that mean mCA level and mCA density increase with gene length, but only mean mCpG level increases with gene length: mCpG density actually peaks in genes of average length. Therefore the deregulated genes correspond to those with both high and low mCA density, but only low mCpG density.

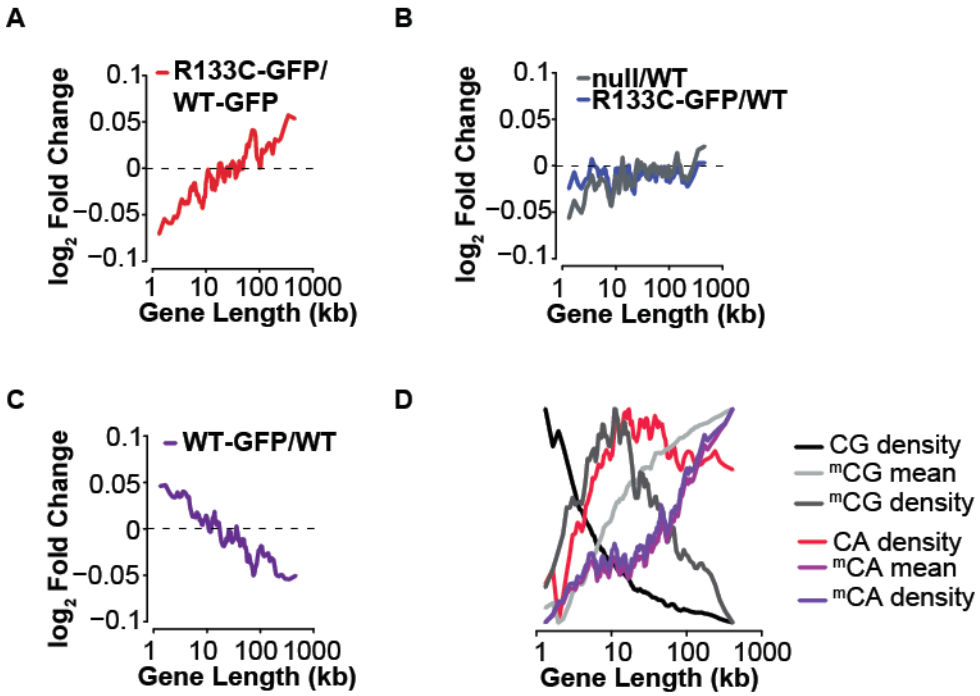


Figure 5.7.2 Subtle length-dependent change in gene expression in *R133C-GFP* cerebellum, relative to *WT-GFP*.

(A) Graph plotting mean log_2 fold change against gene length reveals a small downregulation of short genes and a small upregulation of long genes in the *R133C-GFP* cerebellum, compared to *WT-GFP*, n=3.
(B) Graph plotting mean log_2 fold change against gene length for *Mecp2-null* (null, n=3) and *R133C-GFP* compared to wildtype (WT, n=3). There was subtle evidence of length-dependent gene expression changes in *Mecp2-null* cerebellum. There appeared to be a small downregulation of short genes in *R133C-GFP* cerebellum.
(C) Graph plotting mean log_2 fold change against gene length for *WT-GFP* compared to wildtype. There was evidence of length-dependent gene expression changes in *WT-GFP* cerebellum with downregulation of long genes and upregulation of short genes.
(D) Graph plotting methylation sequence context in relation to gene length shows that only mCA and not mCpG density increases with length. Curve height is not to scale. Reproduced by S.Webb with permission from S.Lagger (unpublished data).

5.7.3 qPCR does not validate upregulation of selected long genes

A subset of long genes (>100kb) that are consistently upregulated in *Mecp2* knock-out models and downregulated in *Mecp2* overexpression models was recently identified (Gabel *et al*, 2015). Real-time qPCR was performed to estimate the expression of ten of these genes in the cerebellum of the same *Mecp2-null*, wildtype, *WT-GFP* and *R133C-GFP* mice in order to validate the length-dependent relationships seen in the microarray analysis. The results are shown in Figure 5.7.3.

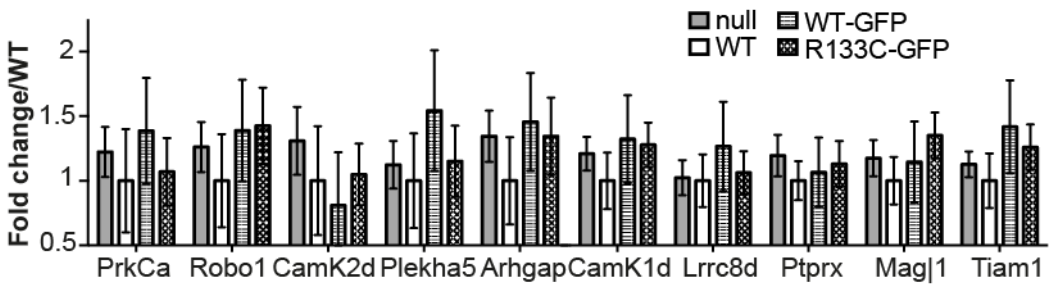


Figure 5.7.3 qPCR does not validate the upregulation of long genes seen in *R133C-GFP* cerebellum, relative to *WT-GFP*.

Graph of relative transcript levels in a subset of long genes that are consistently deregulated in Rett syndrome models. There was a trend towards increased relative transcript levels in *Mecp2-null* (null) and *R133C-GFP* cerebellum relative to wildtype (WT), but *WT-GFP* transcripts appeared to be more abundant than *R133C-GFP* for the majority of the long genes. Relative transcript levels were calculated by normalising qPCR signal to three control genes, *Ubp1*, *Rufy3* and *Nr3c1*. Relative transcript levels are plotted as mean fold change over wildtype±SD.

Relative transcript levels and fold-changes were calculated by normalising the real time qPCR signal for each sample to three genes from a set of control genes that do not show evidence of altered expression in *Mecp2* knock-out or over-expression models, but are also long (Gabel *et al*, 2015). The real-time qPCR did not show any significant differences in the relative transcript levels of the subset of long genes between the four genotypes. There was notable variation in levels between replicate animals within each genotype. This was in keeping with the microarray data. There was a trend towards higher relative transcript levels in the *Mecp2-null* and *R133C-GFP* cerebellum, in comparison to the wildtype cerebellum. This was also consistent with the microarray data. However, in six of the genes (*PrkCa, Plekha5, Arhgap, CamK1d, Lrrc8d, Tiam1*) there was a trend towards higher relative transcript levels in *WT-GFP* cerebellum relative to *R133C-GFP*. This was the opposite pattern to that seen in the gene expression microarray.

In summary, there was variability between *R133C-GFP* and *WT-GFP* replicates in the microarray analysis so only six gene expression changes were statistically significant overall. Just two of these were validated by qPCR. There was some evidence of a length-dependent change in gene expression in the *R133C-GFP* cerebellum. However, the changes were very small (less than 1.1 fold

downregulation of short genes and upregulation of long genes) and may have been attributable to mild overexpression of MeCP2 in *WT-GFP* cerebellum driving gene expression changes in the opposite direction. In addition, qPCR did not completely validate the microarray findings – although there was a trend towards higher relative transcript levels for long genes in *Mecp2-null* and *R133C-GFP* cerebellum relative to wildtype, there was a trend towards the relative transcript levels in *WT-GFP* cerebellum being higher than *R133C-GFP* in six of ten tested genes.

5.8 Discussion

The goal of this chapter was to investigate the molecular basis of Rett syndrome caused by the R133C mutation in order to explain why this results in a milder variant. As the mutation lies within the MBD of *MECP2*, the prediction was that $MeCP2^{R133C}$ would have a DNA binding deficit. It is now known that MeCP2 can bind to methylated cytosines, both in a CG, as well as a CA, context (Guo *et al*, 2014; Gabel *et al*, 2015). Furthermore, MeCP2 can bind to hmCA (Gabel *et al*, 2015). One hypothesis is that R133C Rett syndrome might be less severe because only some of the binding capacities of MeCP2 are impaired i.e. there is a specific deficit in DNA binding, rather than a general loss of binding to all sequences that MeCP2 recognises. This hypothesis was investigated using a combination of protein binding assays *in vitro*, a neuronal cell model and the *R133C-GFP* mouse.

5.8.1 Reduced protein abundance contributes to R133C Rett syndrome

The initial objective was to rule out alternative causes of Rett syndrome that would not result from an impairment of DNA binding. The first alternative considered was a lack of sufficient protein, affecting all proposed functions of MeCP2.

During investigation of relative mRNA level and protein abundance, it was discovered that EGFP-tagging MeCP2 was of consequence. The protein abundance was increased by around 1.5 times in the neuronal cell model and the mouse brain. This can only be an estimate due to the different rate of transfer of proteins of different size from a protein gel onto a nitrocellulose membrane for western blotting,

so it is difficult to accurately compare abundance. However, mRNA levels were increased by the same factor; gradient protein gels were employed to facilitate equal transfer of the differently sized proteins; and the increased protein abundance of WT-GFP was also supported by confocal microscopy, so there was reasonable confidence in the result. The molecular explanation is unclear - it is unlikely that the EGFP tag resulted in more frequent initiation of transcription or an increased rate of elongation of the transcript. It is possible that disruption of the 3' untranslated region in MeCP2 increased the stability of the transcript, the result being that more protein is subsequently translated. It is known that the different lengths of *MECP2* transcript have different relative abundance throughout development (Coy *et al*, 1999). In support of this, it was observed that WT-GFP was markedly more abundant than wildtype protein in neonatal brain tissue (4.5-fold) and cultured neuronal cells (4-fold). The observation of increased abundance of WT-GFP flagged the importance of including it as a control in all subsequent molecular experiments. Notably, although producing a subtle increase in protein abundance, EGFP-tagging the protein did not result in RTT in the mouse (see chapter 4).

Western blots of cultured neuronal cell lysates revealed different relative protein abundances with each differentiation. The quantifications derived from the blots were variable and no significant differences could be found as a result. This was for two reasons. Firstly, there was variation in the neuronal differentiation between genotypes, between clones and between dishes. It was later discovered that this was likely due to fluctuation in the temperature of the incubator, but inadvertent variation at each point in the differentiation protocol would have a similar effect. For example an inaccurate automated cell count at the time of plating the neuronal precursors would result in cells that were less dense, grew less profusely, showed early signs that connecting axons were disintegrating and presumably expressed less MeCP2. Secondly, this variation of input material could not be corrected for. It has previously been shown that MeCP2 deficient neuronal cells in culture have lower levels of RNA (Li *et al*, 2013). When estimating MeCP2 mRNA levels in this study, total RNA was quantified then a set amount was reverse transcribed to use in real time qPCR. The number of input cells was not taken into account. The protein lysate preparation had

to rely on a cell count prior to making a pellet of cells, which would often include dead cells in the final read-out. Although there was variability in the quantification of MeCP2 abundance derived from western blots of neuronal cell lysates, calculating the average from independent differentiations with different mutant ES cell clones yielded very similar results to those from mouse brain and so the results were considered supportive of conclusions drawn.

Interestingly, the abundance of MeCP2 mRNA and protein was reduced by approximately 40% in both the *R133C-GFP* neuronal cells and the *R133C-GFP* mouse, relative to the WT-GFP. This relative reduction in MeCP2 level should not be enough to cause RTT on its own, as a mouse with a floxed *Mecp2* gene and approximately 60% MeCP2 protein abundance, did not display full RTT, but had mild motor impairment and a normal lifespan (Samaco *et al*, 2008). Furthermore, there was no absolute reduction in MeCP2 abundance when comparing R133C-GFP to wildtype littermates, given that the additive result of the EGFP tag and the mutation was an endogenous MeCP2 level. It follows that the protein abundance would not fully account for the mouse phenotype. However, the relative reduction in ambient MeCP2 in the nucleus would amplify the effects of any other deficit in protein function caused by the R133C mutation when compared to the WT-GFP control.

Comparing the *R133C-GFP* mRNA level and protein abundance in adult mouse brain to *WT-GFP* and the other EGFP-tagged missense mutants, *R306C-GFP* and *T158M-GFP* was useful. The instability of MeCP2 when the threonine 158 residue is mutated has previously been reported (Goffin *et al*, 2011) and was recapitulated in this study where, despite equivalent *T158M-GFP* transcript levels, protein abundance was approximately 30% of WT-GFP. It is not clear why R133C did not cause protein instability: the abundance of R133C-GFP correlated with the relative mRNA transcript level. One possibility is that protein that is not associated with DNA is more likely to be degraded indicating that MeCP2^{T158M} has a weaker association with DNA. Another linked possibility, supported by structural data is that the degree of misfolding of MeCP2 is greater in the presence of the T158M mutation than the

R133C mutation, which leads to early degradation (Ghosh *et al*, 2008). The protein abundance data provided an initial explanation for the reduced phenotypic severity of R133C Rett syndrome – at least in relation to T158M Rett syndrome. This observation may be applicable to other mutations in the MBD that result in Rett syndrome. Protein abundance was not significantly reduced in the *R306C-GFP* brain and previous literature suggests that R306C Rett syndrome is caused by a lack of association with the NCoR complex (Lyst *et al*, 2013) so this trend did not seem applicable to Rett-causing mutations outwith the MBD.

In conclusion, MeCP2 abundance is reduced by approximately 40% in the adult brain in the presence of the R133C mutation and this will contribute to the Rett phenotype.

5.8.2 MeCP2^{R133C} associates with NCoR

The second alternative hypothesis for the molecular mechanism of R133C Rett syndrome was failure to perform functions of MeCP2 other than DNA binding. The bridge model of MeCP2 function was conceived from the observation that Rett-causing missense mutations cluster over the MBD and the NID of the *MECP2* gene, whereas neutral variants in the population lie outwith these regions. Therefore the hypothesis would be that failure to bind DNA or the NCoR complex is sufficient to cause Rett syndrome (Lyst & Bird, 2015). Using an anti-GFP antibody in adult mouse brain, R133C-GFP immunoprecipitated HDAC3 and TBLR1 (both components of the NCoR complex), confirming an association *in vivo* and ruling this out as contributory to R133C Rett syndrome.

5.8.3 The subnuclear localisation of MeCP2^{R133C} is abnormal

Having investigated these alternative causes of Rett syndrome, the objective was to establish whether there was evidence that R133C-GFP had a DNA binding deficit at all, as previous studies have indicated a wildtype pattern of subnuclear localisation (Kudo *et al*, 2003). To answer this question, R133C-GFP association with heterochromatin was visualised in mouse brain by confocal microscopy. This section discusses the microscopy findings along with technical challenges.

Notably, the pattern of WT-GFP localisation was indistinguishable from wildtype by confocal microscopy, as has previously been observed (Kudo *et al*, 2003; Schmiedeberg *et al*, 2009). There was an abnormal pattern of R133C-GFP subnuclear localisation in adult male mice across four separate brain regions: the CA1 area of the hippocampus, CA3 region of the hippocampus, the dentate gyrus and the motor cortex. There was a similar appearance in the cerebellum (data not shown). In contrast to EGFP-tagged wildtype protein, R133C-GFP did not appear tightly associated with heterochromatic foci: the protein was also diffusely spread throughout the nucleus. Despite there being 40% less protein than in the *WT-GFP* brain, the intensity of the background nuclear staining in *R133C-GFP* brain was grossly equivalent. In other words, the mutant protein appeared to be distributed throughout the nucleus to a greater degree than the EGFP-tagged wildtype protein. In addition to this, there were ectopic R133C-GFP dots that did not correspond to DAPI dots visible in several brain regions, raising the possibility that the protein bound non-specifically elsewhere in the genome, or displayed less functional ability to cluster chromatin, as has previously been suggested in cell culture experiments (Agarwal *et al*, 2011). These findings contradicted previous literature reporting a wildtype pattern of localisation with this mutation (Kudo *et al*, 2003) but were in accordance with studies that have suggested a mildly reduced association with heterochromatin (Kumar *et al*, 2008; Schmiedeberg *et al*, 2009; Agarwal *et al*, 2011).

Observing this pattern was challenging, given the reduced abundance of R133C-GFP. It was difficult to elucidate the binding pattern when confocal settings were optimised to *WT-GFP* brains because of the markedly reduced intensity of the *R133C-GFP* cells. When the confocal settings were optimised for the *R133C-GFP* brain, the abnormal localisation could be attributed to an artefact of "over-exposing" the image. This potential criticism was negotiated in two ways. Firstly, two sets of images were shown for each region – one set taken at the universal settings and one set taken at the individual settings for comparison. The same criteria were applied to optimise the settings for all genotypes. A look up table was employed to detect and minimise overexposure of the field. In this way, the sections were exposed to approximately the same degree. The second solution was to breed a double *WT-*

GFP/R133C-GFP female mutant, which would express either the wildtype EGFP-tagged protein or the mutant EGFP-tagged protein in any one cell (with the assumption that there were no non-cell-autonomous effects on localisation of protein). It was then possible to see the neighbouring cells in the same image, using the same, compromised settings. Indeed the double mutant appeared to have approximately 50% of neurons with punctate localisation and 50% with mixed punctate and diffuse localisation. It cannot be said with certainty that the "mixed" cells corresponded to the R133C-GFP expressing cells, however the lack of co-staining with the anti-MeCP2 antibody supported this conclusion.

It is not clear why the C-terminal anti-MeCP2 antibody did not stain R133C-GFP. It is conceivable that this staining was dependent on protein abundance. This phenomenon was also seen in the *T158M-GFP* brain (data not shown) and the staining less accurately reflected the EGFP signal in regions of the *WT-GFP* brain with lower levels of MeCP2 expression, for example the motor cortex (data not shown). As MeCP2 is a relatively unfolded protein, it is also possible that any conformational change in the MBD mutants (in combination with the EGFP tag) inhibited antibody binding. This may have been exacerbated by the long period of fixation used for the tissue preparation where the protein fixes internally as well as to DNA.

The abnormal localisation was corroborated by confocal microscopy of the differentiated neuronal cells. Furthermore, it was possible to quantify the altered distribution of protein, given that the cultured cells grew in a monolayer. The quantification supported the visual interpretation of the microscopy. There was a significantly lower proportion of R133C-GFP within the DAPI dot regions than WT-GFP. There were important limitations to the method used for quantification. Firstly heterochromatic dots were defined on the DAPI panel. The DAPI panel had to be utilised because of varying levels of WT-GFP expression in the different genotypes, so defining dots using EGFP intensity thresholds was problematic: one threshold would not fit all. To be a dot area, a region's DAPI intensity had to be at least 0.67 of the pixel with the greatest DAPI intensity in the field of cells (following removal of

dead cells), and had to be at least two pixels in size. There is a possibility that some dots were missed by this technique, by setting limits of DAPI brightness and size. In contrast, there may have been some areas defined as dots that were not, in areas of increased DAPI staining and overexposure for example, or in fields where the brightest DAPI dots were of relatively low intensity. Nevertheless, by using the DAPI panel to define the dot areas, it made the GFP intensity calculations relatively straightforward and independent of the overall protein abundance. The average GFP intensity in the dot region was divided by the sum of the average dot and average background intensity, to give a measure of protein distribution, or dot prominence. This method may have counted ectopic GFP dot areas as "background" and so if, as the brain microscopy suggested, the MBD mutant protein formed new loci outwith the heterochromatic foci, this may have artificially increased the GFP nuclear background intensity and therefore reduced the dot prominence. However, as this would still have represented an abnormal pattern of localisation, it is appropriate that it would result in a reduction of the "dot prominence" measure. One further limitation was that, because of the technical difficulties in running the macro on the GFP panel, the quantification was done on fifty cells per genotype from one differentiation instead of the originally planned, three differentiations. Therefore, it cannot be said definitively that the genotypes had a statistically significant difference in the prominence of their heterochromatic dots without further quantification.

The quantification agreed with a previous report using transfected EGFP-tagged MeCP2 in murine myoblasts (Agarwal 2011), which also found a statistically significant reduction in the accumulation of EGFP-tagged human $MeCP2^{R133C}$ at chromocentres. Agarwal and colleagues found a greater ratio of protein accumulated in dots than that found in this study (3.6 for $^{EGFP}MeCP2^{R133C}$, 5 for $^{EGFP}MeCP2$). However, to calculate their figure they took a simple ratio of mean GFP intensity from four chromocentre areas divided by the mean GFP intensity from four nucleoplasm areas in each cell. They did not take a proportion, by placing the total mean GFP intensity in the denominator. This was repeated for only ten cells from at least two transfections. Given the limitation introduced from using fewer cells and

fewer dots, in addition to taking a ratio, rather than a proportion, it is possible that the measure of accumulation was an overestimate in their study.

In order to determine the impact of the defect in association with heterochromatin on phenotypic severity, *T158M-GFP* and *R306C-GFP* brains were included for comparison. R306C-GFP appeared to have a normal punctate pattern of localisation despite having a slightly lower abundance than WT-GFP, which didn't alter in appearance under the optimised GFP settings, thus supporting the conclusion that the abnormal pattern of localisation in *R133C-GFP* brain was not an artefact of technique. Protein appeared more diffusely spread throughout the nucleus in the *T158M-GFP* brains, suggesting that in addition to reduced relative protein abundance, the T158M mutation impairs DNA binding to a greater degree. This would be a second reason for the milder phenotype in R133C Rett syndrome. Although not formally measured, it was noted that cortical neuronal nuclei in the *T158M-GFP* brain appeared smaller than those in the other genotypes. Soma size is reduced in Rett syndrome patients (Bauman *et al*, 1995). The *T158M-GFP* RTT would have been more advanced at the age mice were culled for the experiment, so perhaps this phenomenon would have been observed at an older age in the other mutant mice and reflects progression of disease.

To conclude, R133C-GFP did not localise normally to heterochromatic foci in cultured neurons or adult brain, indicating impairment in association with chromatin.

5.8.4 MeCP2^{R133C} binding to methylated and unmethylated genomic regions is defective.

There was an indication that binding to heterochromatin was impaired in the *R133C-GFP* cultured neuronal cells and brain and so this was followed up with ChIP experiments using an anti-GFP antibody to further refine this binding deficit. The following section discusses the necessary modification of experimental technique in order to delineate the binding defects.

Although heterochromatin association implies binding to mCpG, it is likely to represent a degree of non-specific binding also as the punctate binding pattern of MeCP2 was shown to be unperturbed in ES cells devoid of methylation (Baubec *et al*, 2013). The mCpG status of selected ChIP regions – major satellite DNA, L1LINEs, IAP elements and the *BDNF* promoter - was confirmed by bisulfite sequencing in adult mouse brain. It is accepted that there is likely to be variation in the level of methylation of selected repeat elements, occurring throughout the genome, which may not have been captured by the limited number of clones analysed. However, bisulfite sequencing of wildtype, *WT-GFP* and *R133C-GFP* genomic DNA yielded relatively consistent results, with varying degrees of MeCP2 expression, indicating that the general pattern of DNA methylation does not change with altered abundance of MeCP2. These sites were taken to represent high (repetitive DNA) and low (*BDNF*) levels of methylation. The pattern of mCpG was likely to have been the same in the relatively immature cultured neuronal cells, as mCpG is established early on in the developing embryo (Bird *et al*, 2002).

Both the cultured neuronal cell and brain ChIP revealed evidence of reduced binding to heavily methylated repetitive sequences and the relatively unmethylated *BDNF* promoter region with R133C-GFP. The interpretation of the ChIP was difficult because of the reduced abundance of R133C-GFP. It was not known how a 40% reduction in protein abundance would ultimately affect the ChIP signal and it appeared that the ChIP signal in the *R133C-GFP* brain relative to *WT-GFP* could have been attributable to the reduced background protein level without any superimposed DNA binding deficit (30% at major satellites; 48% at L1 LINE's; 43% at IAP elements; and 48% at the *BDNF* promoter). The second challenge was presented by variation in percentage input between experiments, which impacted on the ability to detect significant differences between genotypes, even though the R133C-GFP percentage of WT-GFP signal remained consistent. This variation may have been the result of using different anti-GFP antibody aliquots between experiments. Both of these challenges were addressed by refining the ChIP protocol.

Both the original neuronal and brain ChIP used long cross-linking periods (greater than seven minutes) and it was hypothesised that this time was far in excess of the DNA on-off time for WT-GFP, estimated as approximately 20 seconds from two FRAP experiments (Kumar *et al*, 2008; Schmiedeberg *et al*, 2009). Therefore it was possible that MeCP2^{EGFP} continued to accumulate on DNA during the crosslinking period to reach a point of saturation that correlated with overall protein abundance (Poorey *et al*, 2013). The ideal scenario when estimating the relative DNA binding of proteins would be to achieve an instantaneous snap shot of protein association with DNA, which could be approximated by reducing the cross-linking period. For handling samples, one minute was considered the shortest possible time that would be practically possible. Care was also taken to fully quench the formaldehyde by increasing the volume and concentration of the glycine, so that an excess molar ratio was achieved. Repeating the ChIP with a one-minute cross-link made little difference to the read-out. However, when the temperature of the one-minute cross-link was also reduced, the ChIP signal for R133C-GFP decreased relative to WT-GFP. Instead of the 2-3 fold reduction at tested sites, R133C-GFP ChIP signal was reduced by 6-7 fold at repetitive sequences and 4 fold at the unmethylated site, relative to WT-GFP.

With refinement of the protocol, the greater relative reduction in R133C-GFP binding at methylated regions was compatible with a loss of binding specificity, as suggested by structural studies (Ghosh *et al*, 2008; Kucukkal *et al*, 2015) and a previous ChIP-seq study in ES cells (Baubec *et al*, 2013). Additionally, with attention to using the same anti-GFP aliquot for experimental replicates, percentage input achieved was more consistent and statistical differences could be detected. The total percentage input did decrease slightly with the most stringent ChIP conditions, which would be expected if potential binding sites were not saturated during the cross-linking period, as discussed above. It is not possible to conclude that the stringent conditions reflect protein binding *in vivo* because the stringency may simply exaggerate the difference in protein abundance, rather than capturing the instantaneous DNA binding picture, as desired. However, it must be noted that in patients with the R133C mutation, cells expressing this protein will have both reduced abundance and a DNA binding deficit. So although the goal of the ChIP

experiments was to define the DNA binding deficit, with both factors operating, the result reflects the natural pathogenesis of Rett syndrome. Furthermore MeCP2^{R133C} DNA binding deficits have been found by ChIP previously in ES cells (Baubec *et al*, 2013) and in a cellular immunoprecipitation assay with transfected EGFP-tagged protein (S.Lagger, unpublished data), where there was no suggestion of reduced protein abundance. Therefore these results are in keeping with previous literature for MeCP2^{R133C}, suggesting impairment in binding to methylated DNA by ChIP. Additionally, impairment in binding to relatively unmethylated chromatin is suggested, which contradicts a previous *in vitro* study (Georgel *et al*, 2003).

Interestingly, by placing the R133C-GFP ChIP data in the context of the EGFP-tagged allelic series, there was an apparent binding deficit with R306C-GFP in all four regions tested, which did not reach statistical significance. This was not predicted, although one previous report suggested that MeCP2^{R306C} DNA binding may also be impaired by ChIP and EMSA, due to the disruption in a basic patch of amino acids (Heckman *et al*, 2014). This would place into question the assumption that Rett syndrome caused by this mutation is purely the result of a lack of association with the NCoR complex (Lyst *et al*, 2013). Less T158M-GFP than R133C-GFP appeared bound to all four regions, which was consistent with the reduced protein abundance and the appearance of the *T158M-GFP* brain by confocal microscopy. It was also compatible with the increased severity of the T158M Rett phenotype.

To conclude, R133C-GFP binding was reduced by approximately 6-7-fold at heavily methylated genomic regions and 4-fold at a relatively unmethylated site under stringent ChIP conditions in comparison to WT-GFP. This suggests a differential defect in specific and non-specific association with chromatin-associated DNA, with the former affected to a greater degree.

5.8.5 MeCP2^{R133C} does not specifically lose binding to hmC

It was desirable to refine this binding deficit further in order to explain the reduced severity of R133C Rett syndrome (Leonard *et al*, 2003) and shed light on the major

functions of MeCP2. For this reason, the sequence context and type of cytosine modification was investigated as a previous study had suggested that MeCP2^{R133C} lost binding to hmC, whereas binding to mC was relatively preserved (Mellen *et al*, 2012) and this could explain the distinguishable phenotype.

EMSAs were done using the (1-205) fragment of MeCP2 with or without the R133C mutation to address the question of any specific deficit in MeCP2^{R133C} binding. This *in vitro* technique had the advantage that binding to individual moieties could be studied in isolation with equal peptide concentrations. A probe taken from the *BDNF* locus was used, to recapitulate endogenous binding and was modified at a single cytosine in order to delineate the specific sequence and methylation context of the binding deficit. Interestingly, there appeared to be no binding deficit to a mCpG probe, but a specific deficit in binding to methyl- or hmCAC probes with the MeCP2(1-205)^{R133C} peptide. This explained the conclusion drawn by Mellen *et al*, as the probe used in their study contained methylation in a CpG and CA context, and so MeCP2^{R133C} binding to these probes when methylated would have appeared modestly reduced. However, given that MeCP2 doesn't bind to hmCpG (Gabel *et al*, 2015; Chen *et al*, 2015), in the hydroxymethylated probe, the reduced MeCP2^{R133C} binding to hmCA would have been readily apparent. The results presented here contradict the claim that MeCP2^{R133C} specifically loses binding to hmC.

The apparent wildtype binding of MeCP2(1-205)^{R133C} to the mCpG probe seemed at odds with the previous literature, which found mCpG binding to be virtually abolished by MeCP2^{R133C} in Southwestern assays and by EMSA (Ballestar *et al*, 2000; Yusufzai *et al*, 2000; Free *et al*, 2001; Galvão & Thomas, 2005). A conflicting result was also found in our laboratory using a shorter fragment of MeCP2 in exactly the same assay: MeCP2(77-167)^{R133C} binding to a singly methylated CpG dyad was abolished. How can these disparate findings be resolved? As a relatively unstructured protein, it is possible that MeCP2 folds in different ways depending on the length of the peptide fragment. Some support for this comes from one study that found an enhancement of MeCP2 binding to DNA probes with the addition of the N-terminus to the MBD, suggesting some sort of inter-domain stabilisation, or allosteric coupling

(Ghosh *et al*, 2010). Additionally, structural studies suggest quite distinctive CD spectra for the MBD of MeCP2, compared to the full-length protein (Ghosh *et al*, 2008). Notably, the only study to reach a similar conclusion that mC binding was near wildtype also made use of the (1-205) fragment of MeCP2 (Mellen *et al*, 2012), suggesting that this polypeptide may have unique binding characteristics.

To conclude, MeCP2^{R133C} does not specifically lose binding to hmC as previously suggested. The binding characteristics of MeCP2 polypeptides *in vitro* are dependent on the length of the fragment used.

5.8.6 A mCAC binding deficit contributes to the molecular basis of R133C Rett syndrome

The previous section ruled out specific hmC binding deficits using *in vitro* techniques, but this analysis uncovered a potential defect in MeCP2^{R133C} binding to mCAC and hmCAC, specifically. It was unclear whether the evidence of a particular deficit in (h)mCAC binding was relevant *in vivo*, as previous studies have not addressed binding to mCH with MeCP2^{R133C}. Here, it has been demonstrated that by ChIP, R133C-GFP showed reduced affinity for heavily methylated repetitive DNA, presenting a significant majority of mCpG sites in the mouse genome (Horz & Altenburger, 1981). A loss of specificity for mCpG has also been demonstrated by ChIP in ES cells (Baubec et al, 2013). Therefore it is unlikely that full-length protein retains wildtype binding for mCpG *in vivo*. However, the ChIP experiments were confounded by reduced R133C-GFP abundance and the ChIP-seq in ES cells showed a reduction in specificity by plotting MeCP2^{R133C}/MeCP2, so any smaller preference for mCpG may have been masked by a larger wildtype affinity, or higher abundance of protein expressed from the transgene. Loss, or impairment of one of the binding capacities of MeCP2 would be a compelling hypothesis to explain the milder nature of R133C Rett syndrome. The focus was then to determine whether there was any evidence of a specific, or disproportionate DNA binding deficit to mCAC *in vivo* and this is the topic of the following discussion. At this point, the investigation of hmCAC was concluded as these sites represent less than 2% of hmC by current estimates and so, are unlikely to be biologically significant (Lister *et al*, 2013).

The hypothesis was that MeCP2^{R133C} binding was more impaired to mCAC than mCpG. This was investigated initially using ChIP in mouse brain. According to this hypothesis, in neonatal brain tissue, with negligible mCAC, R133C-GFP ChIP signal would be less reduced relative to WT-GFP ChIP signal, in comparison to the same ChIP assay in adult tissue. Importantly, R133C-GFP abundance was also 40% less than WT-GFP at this stage. The hypothesis was correct, however neither the WT-GFP nor the R133C-GFP signal was significantly different from the negative control – wildtype brain with no EGFP tag. This was probably because the absolute protein abundance was low at this developmental stage. Furthermore, it's possible that there wasn't much mCAC in the repetitive sequences used in the ChIP assay in the adult brain. So a difference may have been difficult to detect in these particular regions.

Specific genomic loci enriched for mCpG or mCAC were then selected for GFP ChIP to see if the ChIP signal in *R133C-GFP* adult brain at the mCpG enriched sites would be higher than at the mCAC enriched sites, relative to *WT-GFP*. The results provided no evidence that this was the case, with an almost identical relative percentage input across all sites immunoprecipitated, including the control regions with no methylation. It was notable that the percentage of input DNA that was pulled down was slightly higher at the mCpG rich sites than at the mCAC rich sites in *WT-GFP* brain, indicating that these were higher affinity binding sites. This was in agreement with the EMSA findings where the percentage of mCpG probe shifted at the highest concentration of wildtype MeCP2(1-205) (64%) was higher than the percentage of mCAC shifted (43%). The lack of discrimination of R133C-GFP between the mCG-rich and mCAC-rich sites could have occurred for a number of reasons: firstly, the ChIP may not have been a sensitive enough assay to detect a distinctive binding pattern, especially with the reduced protein abundance; secondly, the sites may not have been equivalently enriched for the two signalling moieties – the mCAC sites had a lower "level" of enrichment (approximately 2) than the mCpG sites (approximately 8); thirdly, MeCP2 binding at the selected sites might have been at a low level, making it different to detect relative differences; fourthly, analysis of the methylation context in some of the mCpG-rich sites in a separate published

bisulfite-seq data set (Lister *et al*, 2013) following this experiment revealed that there were appreciable amounts of mCAC at this locus, at least in frontal cortex; finally, the hypothesis may have been incorrect.

To resolve the issue of lower protein abundance in *R133C-GFP* brain overshadowing any specific deficit in binding to mCAC, DNA that had already been immunoprecipitated by the protein was bisulfite-sequenced at two known MeCP2 binding sites: *Fgfb* (with several CpG's) and *Bai3* (with several CAC's) (Guo *et al*, 2014). Although both the WT-GFP and R133C-GFP immunoprecipitated DNA were enriched for mCpG over their respective input DNA at the *Fgfb* locus, only the WT-GFP immunoprecipitated DNA was enriched for mCAC over the corresponding input DNA at the *Bai3* locus. This finding was supportive of a specific deficit in mCAC binding.

An important limitation of this experiment was that it was not possible to calculate a non-conversion rate for the mCAC rich site (*Bai3*) bisulfite sequencing experiment due to the nature of conventional bisulfite sequencing protocols that use cytosines outwith a CpG context to calculate non-conversion rate following the bisulfite treatment (*i.e.* if a non-CpG cytosine remained a C, it was considered to have not been converted, rather than methylated). Given that the protocol was identical to that used for all of the bisulfite sequencing, which had achieved conversion rates of over 95%, it is unlikely that a significant majority of detected mCAC represented non-conversion. However, it would be important to repeat this experiment with a control probe, methylated in a known sequence context to have a definitive measure of non-conversion. With this limitation in mind, the experiment was supportive of the hypothesis that there was a greater deficit in R133C-GFP binding to mCAC, which presents a lower affinity endogenous binding site. Although inconclusive, the R133C-GFP ChIP-seq lent some support to the hypothesis – at least in the protein peaks identified with the highest confidence. It was unexpected that there appeared to be no increase in mCpG at WT-GFP binding sites in the sorted neurons. This may be because there was a notable AT-bias in the sequenced DNA following library preparation (S.Lagger, unpublished communication) and so the sites with the highest

mCpG content may not have been represented. Another possibility is that the published data set of frontal cortical neuron methylation that was used to plot the mean methylation density may not be representative of all regions of the brain, so the ChIP binding profiles may not match adequately. There is evidence to suggest that gene regulation by MeCP2 can be cell type-specific (Sugino *et al*, 2014). Therefore, this data must be interpreted with caution and it would be important to repeat this experiment with matching bisulfite sequencing data.

Taken together, the western blotting, confocal microscopy, ChIP, EMSA, and bisulfite sequencing data would suggest that the molecular basis of R133C Rett syndrome is a combination of a modest reduction in protein abundance and reduced affinity to both of the major binding sites of MeCP2: methylation in a CpG and CAC context, with the latter probably abolished. Non-specific binding and binding to a third moiety, hmCAC may also be reduced, although the biological significance of this is unknown.

5.8.7 Do deregulated long genes cause Rett syndrome?

There is recent evidence suggesting that repression of long genes by MeCP2 is dependent on binding to mCA, which in the brain occurs predominantly in a CAC context (Xie *et al*, 2012; Guo *et al*, 2014; Gabel *et al*, 2015; Lister *et al*, 2013). The prediction from this was that the long gene effect would be prominent in the *R133C-GFP* brain with little to no mCAC binding. To investigate this prediction and determine what the biological consequences of R133C Rett syndrome were, gene expression in adult mouse cerebellum was measured with microarrays in triplicate for six week-old *R133C-GFP*, *WT-GFP*, wildtype and *Mecp2-null* mice. This final section discusses the limited evidence from these experiments that deregulation of long genes causes Rett syndrome.

Microarray data did not display clustering with genotype, indicating (1) significant variability between the genotype replicates and (2) lack of a distinct pattern of gene expression with the R133C mutation. The expression of thousands of genes in the R133C mutant was altered to a small degree, both by up- and downregulation, in

comparison to the *WT-GFP* control, which was consistent with previous similar studies (Jordan *et al*, 2007; Urdinguio *et al*, 2008; Ben-Shachar *et al*, 2009), but there were only six significant changes in expression of individual genes (two of which were later validated by qPCR). Additionally, this handful of significantly altered genes was not the same as those found with the wildtype and *Mecp2-null* comparison. However, it was possible to observe an upregulation of long genes (>200kb) and down-regulation of short genes (<10kb) in *R133C-GFP* compared to *WT-GFP* cerebellum. This was not validated by real time qPCR and was significantly reduced in the *R133C-GFP* and wildtype comparison, indicating that most of the effect was attributable to a downregulation of long genes (and *vice versa*) in the *WT-GFP* cerebellum. This was compatible with the mild MeCP2 overexpression in this mouse and the observation that the *WT-GFP* samples appeared to be outliers according to principal component analysis and heat-mapping.

The relative fold changes detected in gene expression were small. Microarray analysis assumes that the overall level of transcription in the samples is the same, which appeared to be the case. Total RNA was isolated from half cerebella and normalised to total genomic DNA for all four genotypes and no significant differences were found (data not shown). Therefore the subtle upregulation of long genes in the mutant and *Mecp2*-null cerebella was not underestimated by reduced total RNA in these brains. Following the experiment, it emerged that the cerebellum had the smallest degree of long gene upregulation in *Mecp2*-null mice in comparison to other brain regions (Gabel *et al*, 2015), so it was not the optimal platform in which to detect effects specific to a loss in mCA(C) binding. Although the cerebellum is mostly comprised of granule cells, there are other cell types and cell type-specific gene expression changes may have been masked by analysing the whole tissue (Mellen *et al*, 2012; Sugino *et al*, 2014). It would be useful to repeat this experiment with a greater number of replicates, in several different brain regions and/or neuronal subtypes.

With these limitations in mind, importantly, these data may challenge the hypothesis that upregulation of long genes causes Rett syndrome. The magnitude of

downregulation in the *WT-GFP* and wildtype comparison was to the same degree as the magnitude of upregulation in the *R133C-GFP* and *WT-GFP* comparison and the *WT-GFP* mice did not show evidence of progressive neurological impairment previously reported in mouse models of *Mecp2* overexpression (Collins *et al*, 2004). In fact, the mice were overtly wildtype. This magnitude was greater than the magnitude of upregulation of long genes seen in the *Mecp2-null* and wildtype comparison. Additionally, the *R133C-GFP* cerebellum showed virtually no upregulation of long genes relative to untagged wildtype and the mice did have obvious RTT. It may be important that the cerebellum was harvested prior to manifestation of R133C RTT, but this would only suggest that upregulation of long genes represents a secondary effect, rather than primary effect of the disease. Perhaps lending weight to this argument was the observation of Gabel *et al*, that the upregulation of long genes worsened as Rett syndrome progressed (Gabel *et al*, 2015).

The obvious downregulation of genes less than 10kb in length in *R133C-GFP* cerebellum, although also apparent in the study by Gabel and colleagues in other Rett models, has not been widely discussed. One of the features that the short and long genes have in common is low mCpG density, suggesting that in the neuronal methylome, high density mCpG sites are not critically regulated by MeCP2 or on the contrary, that there are additional mechanisms safeguarding their regulation. There has been criticism of the claim that long genes are overrepresented in autism with the counter argument that this relationship is a product of the methods of mutation discovery (Shohat & Shifman, 2014). With further tissue and methylation-specific analysis, it will be interesting to see if shorter genes become relevant in the pathogenesis of Rett syndrome in the future. This study does not present convincing evidence that long gene deregulation plays a role in the molecular basis of R133C Rett syndrome.

6. Conclusions

Rett syndrome is a debilitating and progressive neuropsychiatric disorder caused by mutation in *MECP2*. Studies in mice and cells have suggested that Rett is a disorder of synaptogenesis and neuronal maintenance, in that it can be largely reversed following onset with restoration of lifelong MeCP2 expression in all brain cells. The onset of Rett syndrome coincides with maturation of the unique (hydroxy)methylation landscape of the brain (with expansion of mCH and hmCpG), increasing expression of MeCP2 and formation of synapses, suggesting the three are critically linked. MeCP2 is well placed to interpret this new repertoire of cytosine modification. The function of MeCP2 is not clearly elucidated, but undoubtedly involves linking methylation of DNA to remodelling of chromatin and at least in part, repression of transcription. The functional mechanisms are undetermined, but it has recently been proposed that MeCP2 binding to mCA in the brain results in the specific transcriptional repression of long genes. Regulation of MeCP2 by phosphorylation, coincident with neuronal firing, would implicate the protein in accurate interpretation of incoming sensory stimuli and translation into targeted gene expression. These actions would be essential for synaptogenesis and cognitive function and together, this forms an attractive hypothesis.

In order to further understand the specific functions of MeCP2 and their relevance to Rett syndrome, the field must progress to the study of specific Rett-causing mutations. In patients, classical Rett runs a stereotyped course but there is significant clinical heterogeneity. Despite this, it is apparent that there is a spectrum of phenotypic severity associated with different mutations. The reasons for this are uncertain. This work focused on the missense mutation, R133C, which is common, occurs in a key position of the MBD and results in a milder syndrome in patients. A novel mouse model of the R133C was created. With comparison to other mouse lines it was confirmed that the R133C mutation results in a milder Rett-like phenotype in the male in the absence of potentially confounding patterns of XCI. The molecular basis was the result of a modest reduction of protein abundance and reduced association with heterochromatin. This was a combination of decreased binding to

relatively unmethylated regions and repetitive DNA rich in mCpG, with the latter affected to a greater degree and only becoming apparent under stringent experimental conditions. There was evidence to suggest that binding to mCAC was more markedly impaired than binding to mCpG, both *in vitro* and *in vivo*. This work also contradicted a previous claim that the R133C mutation results in a specific deficit in binding to hmC (Mellen *et al*, 2012). The consequences of the R133C mutation in terms of gene expression require further investigation: results were variable and there was a subtle upregulation of long genes in the cerebellum. However this was largely the result of a downregulation of genes in the EGFP-tagged wildtype control cerebellum. This result questions the pathological relevance of deregulation of long genes in Rett models. Perhaps the pattern of long gene upregulation is a red herring and simply represents a final common pathway in brain tissue that is "sick" from a variety of causes, akin to the early stages of apoptosis or atrophy. There may be alternative cellular mechanisms to protect the expression of genes with dense CpG methylation, for example and the role of aberrant short gene expression has not been investigated.

This study and others suggest that phenotypic severity in Rett syndrome results from the degree to which key functions of MeCP2 are impaired, producing additive compound effects. R133C Rett may be less severe because protein expression and non-specific binding are only modestly reduced, mCAC binding is (potentially) abolished, but a degree of mCpG binding is retained. Transcriptional repression (Kudo et al, 2003), association with NCoR and (possibly) protein regulation by phosphorylation are preserved. In comparison, protein abundance and association with heterochromatin are reduced to a greater degree with $MeCP2^{T158M}$ (Kumar *et al*, 2008; Schmiedeberg *et al*, 2009). The R306C mutation abolishes association with NCoR (Lyst *et al*, 2013), but may also reduce DNA binding to a modest degree (Heckman *et al*, 2013) and disrupt regulation by phosphorylation (Ebert *et al*, 2013).

6.1 Future directions

Specifically, it would be useful to study gene expression in the context of $MeCP2^{R133C}$ in other brain regions or particular cell types on a larger scale. An RNA-

seq study may be useful in this respect, in order to delineate isoform-specific changes that are lost in microarray analysis. In combination with methylation maps corresponding to these brain regions and MeCP2^{R133C} ChIP-seq, this data could precisely define the binding profile of the protein and the impact on gene expression. The *R133C-GFP* mouse will be a useful tool for trialling therapeutic strategies, to investigate mutation-specific responses to treatments, allowing a personalised approach to Rett syndrome management. Functional impairment conferred by Rett mutations outwith the MBD and NID for example C-terminal deletions, A2V, P225R and P322L require further investigation and may highlight the importance of other protein domains, functions or forms of regulation in Rett pathogenesis, for example chromatin compaction, phosphorylation or non-specific association with DNA.

More generally, it will be fascinating to further delineate MeCP2 binding profiles *in vivo* and determine how this mechanistically leads to derangement of gene expression. The lack of deregulation of genes that are dense in mCpG is surprising, and further manipulation of methylation marks at different developmental stages, along with further classification of genes in terms of methylation content in specific cell types, may help to pinpoint MeCP2 transcriptional repression. The deregulation of short genes observed in several Rett models requires further investigation.

It is certainly an exciting time for epigenetics research in this complex field of neurodevelopment and the autistic spectrum. The discipline is edging closer to a complete understanding of Rett pathogenesis, with the ultimate goal of achieving phenotypic reversibility in affected girls.

Appendix 1: Mouse Phenotypic Scoring

	0	0.5	1	1.5	2
Mobility	as WT	reduced movement when compared to WT: brief freezing when first placed on the bench or short periods spent immobile	reduced movement when compared to WT: extended freezing period when first placed on bench and longer periods spent immobile	markedly reduced movement: long periods of freezing and majority of time spent immobile	no spontaneous movement when placed on the bench; mouse can move in response to a gentle prod or a food pellet placed nearby
Gait	as WT	hind legs are slightly wider than WT with feet pointing outwards	hind legs are spread wider than WT when walking or running with reduced pelvic elevation, resulting in a “waddling” gait	hind legs spread widely apart with pelvis flat to the ground	more severe abnormalities: tremor when feet are lifted, walks backwards or 'bunny hops' by lifting both rear feet at once
Hindlimb clasping	legs splayed outwards	legs briefly drawn towards each other or one leg drawn towards the body	hindlimbs are drawn towards each other (without touching) or one leg is drawn in to the body	both legs briefly pulled in tightly but not maintained	both legs are pulled in tightly, either touching each other or touching the body
Tremor	no tremor	suggestion of mild tremor	intermittent mild tremor	intermittent obvious tremor	continuous tremor or intermittent violent tremor
Breathing	normal breathing	suggestion of irregularity to breathing	regular breathing interspersed with short periods of more rapid breathing or with pauses in breathing	periods of regular breathing are less obvious	very irregular breathing - gasping or panting
General condition	clean shiny coat, clear eyes, normal stance	either eyes dull or coat dull & ungroomed, normal stance	eyes dull, coat dull & ungroomed, somewhat hunched	eyes dull, coat dull & ungroomed, obvious hunch	eyes crusted or narrowed, piloerection, hunched posture

(adapted from Guy *et al*, 2007)

Appendix 2: Publication

OXFORD

Human Molecular Genetics, 2016, Vol. 25, No. 3 558–570

doi: 10.1093/hmg/ddv496
Advance Access Publication Date: 8 December 2015
Original Article

ORIGINAL ARTICLE

The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome

Kyla Brown†, Jim Selfridge†, Sabine Lagger, John Connelly, Dina De Sousa, Alastair Kerr, Shaun Webb, Jacky Guy, Cara Merusi, Martha V. Koerner and Adrian Bird*

Wellcome Trust Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Max Born Crescent, Edinburgh EH9 3BF, UK

*To whom correspondence should be addressed. Tel: +44 1316505670; Fax: +44 1316505379; Email: a.bird@ed.ac.uk

Abstract

Rett syndrome is caused by mutations in the X-linked *MECP2* gene, which encodes a chromosomal protein that binds to methylated DNA. Mouse models mirror the human disorder and therefore allow investigation of phenotypes at a molecular level. We describe an *Mecp2* allelic series representing the three most common missense Rett syndrome (RTT) mutations, including first reports of *Mecp2*[R133C] and *Mecp2*[T158M] knock-in mice, in addition to *Mecp2*[R306C] mutant mice. Together these three alleles comprise ~25% of all RTT mutations in humans, but they vary significantly in average severity. This spectrum is mimicked in the mouse models; R133C being least severe, T158M most severe and R306C of intermediate severity. Both R133C and T158M mutations cause compound phenotypes at the molecular level, combining compromised DNA binding with reduced stability, the destabilizing effect of T158M being more severe. Our findings contradict the hypothesis that the R133C mutation exclusively abolishes binding to hydroxymethylated DNA, as interactions with DNA containing methyl-CG, methyl-CA and hydroxymethyl-CA are all reduced *in vivo*. We find that MeCP2[T158M] is significantly less stable than MeCP2[R133C], which may account for the divergent clinical impact of the mutations. Overall, this allelic series recapitulates human RTT severity, reveals compound molecular aetiologies and provides a valuable resource in the search for personalized therapeutic interventions.

Introduction

Mutations in the X-linked *MECP2* gene are implicated in several human disorders characterized by developmental delay and intellectual disability, including Rett syndrome (RTT) (1) and *MECP2* duplication syndrome (2). RTT is a condition with postnatal onset that predominantly affects girls, as males fail to survive beyond infancy. Animal models have proved useful for improving our understanding of MeCP2 function and for explaining in molecular terms the origin of the RTT phenotype. The first mouse models were simple loss-of-function alleles caused by gross deletion of most of the coding sequence (3,4), but knock-in mutations corresponding to specific RTT-causing mutations (5,6) offer the opportunity for deeper understanding. Of particular interest are missense RTT mutations leading to the substitution of a single amino acid, as these pinpoint critical regions of the protein that cannot be deduced from frameshift and nonsense mutations, or

†These authors contributed equally to this work.
Received: August 24, 2015. **Revised:** November 9, 2015. **Accepted:** November 30, 2015

mutations that affect splicing (7). Importantly, human and mouse MeCP2 are highly homologous proteins, being ~95% identical at the amino acid level across ~500 amino acids. This stringent functional conservation over evolutionary time makes it likely that mutations in the mouse gene provide appropriate models for determining the molecular aetiology of the human disorders.

In this study, we generated knock-in mice for the three most common RTT missense mutations: T158M, R306C and R133C, which together account for ~25% of all RTT cases. The most frequent of all RTT mutations is T158M (~12% of cases), followed by R306C (~9%) and R133C (~5%) (8). Two of the three mutations (T158M and R133C) localize to the methyl-CpG binding domain (9), and previous *in vitro* studies suggest that both can adversely affect DNA binding (10–12). The R306C mutation is located within the recently defined NCoR/SMRT Interaction Domain (NID) (13). This mutation, like others close by, prevents MeCP2 from interacting with the NCoR/SMRT corepressor complex and inhibits transcriptional repression in cell-based assays. Although all three mutations result in 'classical' RTT symptoms, there is a clear difference between them when clinical severity rating scales from many patients are analysed (8,14). T158M is more severe than R306C on average. Least severe of the most frequent RTT mutations is R133C, which is more often associated with preservation of walking and/or speech (8). We sought to explain this clinical spectrum in molecular terms.

The R133C mutation has not previously been modelled in mice, but it has been proposed that its milder phenotype is due to retention of binding to 5-methylcytosine accompanied by loss of binding to hydroxymethylcytosine (hmC) (15). Mice with the T158A RTT mutation were characterized (6), but the much more common T158M RTT allele has not yet been reported. The R306C mutation has been shown to cause Rett-like phenotypes in knock-in mice (13), and mice bearing an integrated mutant transgene have been comprehensively phenotyped (16). Here, we focus on the comparative phenotypes caused by these three mutants at molecular, cytogenetic and behavioural levels. This was achieved by comparing lines of male mice expressing MeCP2 variants from the endogenous *Mecp2* locus as fusions with the reporter protein enhanced green fluorescent protein (EGFP), including wild type (WT). We find that these RTT mutations recapitulate the severity spectrum of human RTT, and they offer a molecular explanation for this phenotypic range. In particular, the R133C mutant protein has a reduced affinity for all known MeCP2 target sequences, including methyl-CG (mCG) and is significantly reduced in abundance. We conclude that the RTT-like phenotype cannot be exclusively attributable to loss of hmC binding by this mutant protein as previously proposed (15). The greater severity of the T158M mutation is explained by more extreme destabilization of the protein, coupled with reduced affinity for modified DNA. This finding mirrors that reported previously for a less frequent allele, T158A, affecting the same amino acid (6). R306C, which has lost the ability to interact with the NCoR/SMRT co-repressor, persists at WT levels and largely retains WT chromatin binding characteristics. This *Mecp2* allelic series allows resolution of the compound phenotypes underlying these causes of RTT. In addition, the study supports available prognostic information regarding Rett syndrome and provides a resource in the search for individualized therapeutic approaches.

Results

WT MeCP2-GFP mice are essentially phenotypically WT

We generated an allelic series in which endogenous *Mecp2* or mutant *Mecp2* genes were fused in frame with EGFP at their C-termini (Fig. 1A; Supplementary Material, Fig. S1). Mice expressing knock-in WT *MeCP2-EGFP* fusion genes have been reported (13,17), but without extensive characterization. We initially looked for phenotypic defects due to the fusion of WT MeCP2 with EGFP by monitoring male mice at the molecular and whole organism levels. Analysis of hemizygous males has the advantage that phenotypic severity is not influenced by the pattern of X chromosome inactivation, and so severity of individual mutations can be assessed in an unbiased fashion. Quantitative polymerase chain reaction (PCR) and western blots indicated that both mRNA and protein products of the *Mecp2-GFP* gene (*WT-GFP*) were expressed in brain, though at somewhat higher levels than the endogenous *Mecp2* gene (*WT*) (Fig. 1B and C). Quantitative western blots suggested that the level of WT-GFP is ~1.6-fold higher than in untagged WT littermates. At the whole organism level, we analysed cohorts of *WT-GFP* mice back-crossed for four generations to give a genetic background that is ~94% C57BL/6J. *WT-GFP* knock-in mice were fertile and showed normal survival but tended to be smaller than WT littermates (Fig. 1D and E). Cohorts were monitored using a phenotypic scoring methodology that records breathing, tremor, gait, hindlimb clasping, mobility and general condition (18). This series of observational tests has the advantage that it is not affected by learning and can therefore be performed weekly over long periods, giving reproducible results. Using this method, *WT-GFP* mice showed no significant phenotypic deterioration compared with WT littermates (Fig. 1F), reinforcing the view that, despite the presence of the EGFP tag, they are essentially WT.

To search for neurological phenotypes in more detail, we subjected *WT-GFP* mice to a series of motor coordination and behavioural tests (Fig. 1G–I). Performance on the elevated plus maze was indistinguishable from *WT* and on the accelerating rotarod was also not significantly different from *WT* littermates. The hanging-wire test showed a weak but reproducible reduction in the ability to engage hindlimbs with the wire. We noted in addition that there was a trend towards a mild reduction in weight and a trend towards defective rotarod performance, but neither achieved significance. These very weak phenotypic effects may be attributable to the over-expression of MeCP2-GFP relative to untagged protein. Taking the findings together, however, we conclude that the addition of the C-terminal EGFP epitope and the moderate over-expression of the protein have minimal phenotypic consequences by these assays.

Allelic series of RTT missense mutations recapitulates severity in humans

Using the same knock-in technology, we generated the following mouse lines: *Mecp2*[T158M]^{EGFP}, *Mecp2*[R306C]^{EGFP} and *Mecp2* [R133C]^{EGFP}, referred to as *T158M-GFP*, *R306C-GFP* and *R133C-GFP*, respectively. Both MeCP2 isoforms, which differ only at their extreme N-termini, are affected by the knock-in. Each line was back-crossed to obtain a predominantly C57BL/6J genetic background equivalent to that of the *WT-GFP* mice (94%). Each of the mutants gave rise to males that exhibited overt phenotypic defects from ~6 weeks of age (see below), but their survival profiles were very different (Fig. 2A). *T158M-GFP* male survival was slightly prolonged compared with *Mecp2*-null mice from previous analyses (18) (median survival 13 weeks compared with 9 weeks), whereas *R306C-GFP* males survived significantly longer (median = 30 weeks). *R133C-GFP* males were most mildly affected, with a median survival of 42 weeks.

Phenotypic scoring of *T158M-GFP*, *R306C-GFP* and *R133C-GFP* mice matched survival curves and provided detail regarding

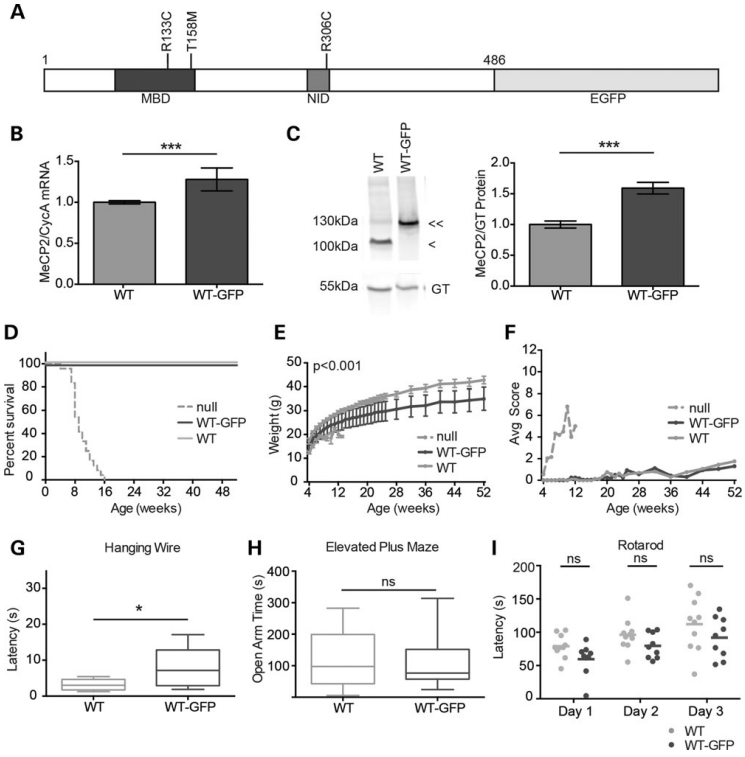


Figure 1. *WT-GFP* mice show minimal overt phenotype. (**A**) Schematic representation of MeCP2 with an EGFP tag. Missense mutations analysed in this study (T158M, R133C and R306C) are shown in relation to the MBD and NID. (**B**) Levels (mean ± SD) of *Mecp2* transcripts in *WT-GFP* mouse brain ($n = 9$) compared with *WT* littermates ($n = 9$), expressed relative to *Cyclophilin A* transcript (CycA). (**C**) Representative western blot and quantification comparing MeCP2 protein abundance in *WT-GFP* mouse brain (double arrow-head, $n = 6$) versus *WT* littermates (single arrow-head, $n = 6$). Gamma tubulin (GT) served as an internal control. Mean ± SEM plotted. (**D**) The Kaplan–Meyer plots showing survival of *WT-GFP* mice ($n = 8$) compared with their *WT* littermates ($n = 8$) and *Mecp2*-null mice [Ref. (18)] ($n = 24$). (**E**) Growth curve showing average weight of *WT-GFP* mice ($n = 8$) compared with their *WT* littermates ($n = 8$). Using repeated measures ANOVA, the difference was consistent and significant ($P < 0.001$) over time, but at any single time point, the difference was not significant. (**F**) Phenotypic scoring (see the text) of *WT-GFP* or *WT* mice. For comparison, Mecp2-null scoring is shown. (**G–I**) *WT-GFP* mice ($n = 9$) and their *WT* littermates ($n = 10$) were compared using three behavioural tests: (G) the hanging-wire test, (H) the elevated plus maze, and (I) the accelerating rotarod showing individual mean latencies (dots) and cohort mean latency (line) for each of three days of trials. Statistical analysis took all trials into account. Statistical tests were unpaired two-tailed *t*-test (B, C and E) and the Kolmogorov–Smirnov test (G–I). All behavioural paradigms were conducted on animals aged 8–10 weeks, and biochemical analyses were conducted using tissues from adults aged 6–12 weeks. Asterisks denote the following *P* values: $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$.

progression (Fig. 2B). *T158M-GFP* male mice acquired a range of phenotypic traits rapidly over a few weeks, whereas *R306C-GFP* mice, after an initially rapid onset, remained phenotypically relatively stable until 33 weeks, at which time phenotypes of survivors became more severe. Interestingly, this late inflection point coincides with an age identified previously as being particularly sensitive to MeCP2 deficiency (19). *R133C-GFP* males showed a more gradual build-up of phenotypic defects compared with either *R306C-GFP* or *T158M-GFP*. Weight profiles revealed that all the mutants were lighter than *WT-GFP* animals of equivalent age (Fig. 2C).

Cohorts of the allelic series underwent motor coordination and behavioural testing at 8–10 weeks of age at which time *T158M-GFP*, *R306C-GFP* and *R133C-GFP* mice scored 5.4, 2.8 and 1.4, respectively, using the observational scoring method (see Fig. 2B). None of the mutants performed as well as their corresponding *WT* littermates on the hanging wire and elevated plus maze (Fig. 3A and C). *T158M-GFP* and *R306C-GFP* were also significantly compromised as assayed by the 3-day accelerating rotarod-learning paradigm (Fig. 3E), but *R133C-GFP* mice showed no detectable defects compared with WT in this test.

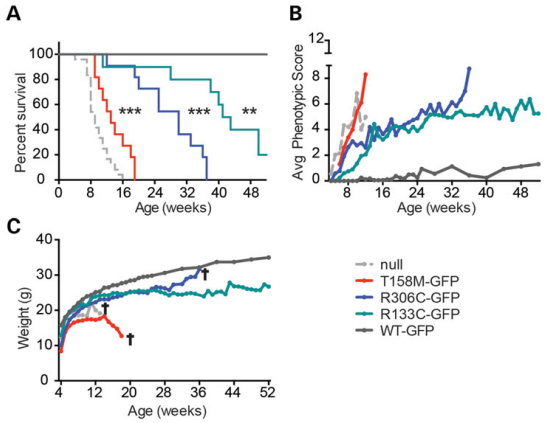


Figure 2. The *Mecp2-GFP* allelic series mimics the clinical severity of equivalent human mutations. (**A**) The Kaplan–Meyer plots showing reduced survival of *T158M-GFP* (red), *R306C-GFP* (blue) and *R133C-GFP* (green) mutants in comparison with *WT-GFP* (dark grey) and *Mecp2*-null (light grey) mice (18). Statistical significance is denoted as follows: *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$ (Mantel-Cox test) and calculated in comparison with the *WT-GFP* mice. (**B**) Graph showing average phenotypic score for each cohort over time. (**C**) Graph showing average weight over time. Cohorts comprised *WT-GFP* ($n = 8$), *R133C-GFP* ($n = 10$), *R306C-GFP* ($n = 11$), *T158M-GFP* ($n = 11$) and *Mecp2*-null mice ($n = 12$–24). A cross indicates that all mice in the cohort had been culled by this time point due to severity of the RTT-like phenotype.

In these analyses, WT littermates lacked the GFP tag on endogenous MeCP2, but as shown in Figure 1, the presence of this tag has minimal phenotypic consequences. We confirmed that the presence of C-terminal GFP was not contributing to phenotype by comparing the performance of mutant-GFP mice with age-matched WT-GFP mice in the same tests (Fig. 3B, D and F). Comparison between litters of different lines in this way is less well controlled, as illustrated by the variability of WT test scores in Figure 3A and C. Nevertheless, we observed that both R133C and R306C mice performed significantly less well than WT-GFP in the hanging-wire and rotarod tests, while performance of T158M mice in both elevated plus maze and rotarod was significantly inferior. In cases where differences between mutant and WT-GFP mice did not reach statistical significance (R133C and R306C in the elevated plus maze, and T158M on the hanging-wire test), the data showed trends towards defective performance that matched those detected in the comparison with littermates. The behavioural analyses therefore confirm phenotypic scoring and survival data in showing that the Rett missense mutations are responsible for the observed phenotypes.

Heterozygous females show RTT-like phenotypes

Rett syndrome affects females who are heterozygous for these *MECP2* mutations and consequently mosaic for expression of WT or mutant alleles, due to X-chromosome inactivation. In mice, heterozygotes for the null allele show a delayed-onset phenotype that is usually detectable between 4 and 12 months of age. We compared phenotypic scores of heterozygotes for each mutation over ~12 months. *WT-GFP* female heterozygotes were not scored in this fashion as they invariably displayed no signs of disease up to at least 1 year. All had RTT-like symptoms, with *R133C-GFP* and *R306C-GFP* being less severe than *T158M-GFP* (Supplementary Material, Fig. S2). In this respect, also the phenotypes of these frequent RTT mutations in mice reflect the spectrum of human severity.

MeCP2 R133C-GFP and T158M-GFP show reductions in both abundance and chromatin binding in neurons

The matching spectrum of severity between humans and mice with respect to *T158M-GFP*, *R306C-GFP* and *R133C-GFP* mutations provides validation that the mouse models are an appropriate system to investigate the mechanism by which these mutations cause RTT in humans. We used cell biological and biochemical techniques to understand the reasons for compromised MeCP2 function. At the level of mRNA, expression of the *Mecp2-GFP* knock-in alleles in brain was somewhat elevated compared with the native *Mecp2* gene, except in the case of *R133C-GFP* (Fig. 4A). In agreement with the mRNA analysis, WT-GFP and R306C-GFP proteins were 1.5-fold more abundant than native MeCP2 in the brain, with the expression of R306C-GFP modestly reduced relative to WT-GFP (Fig. 4B). In contrast, R133C-GFP protein was reduced to ~55%, and T158M-GFP was present at only ~30% of WT-GFP by western blotting (Fig. 4B and Supplementary Material, Fig. S3A) and fluorescence-activated cell sorting (Supplementary Material, Fig. S3B). Abundance was therefore equivalent to approximately 95 and 50% of untagged MeCP2 for R133C-GFP and T158M-GFP, respectively. Reduced stability of T158A, a distinct allele of T158, was reported previously (6), making it likely that this is an intrinsic property of MeCP2 lacking the hydrogen-bonding capability of threonine at this position (20). To determine whether the more subtle reduction in R133C-GFP seen in the brains is specific to this particular mouse line or is a reproducible property of the mutant protein, we analysed embryonic stem cell-derived neurons in culture using an independent knock-in clone. Again R133C-GFP was present at approximately half the level of WT-GFP, suggesting that the mutation itself is responsible for the deficiency, probably due to compromised RNA or protein stability (Supplementary Material, Fig. S3C). This relationship was also seen with an alternative *WT-GFP* clone (data not shown).

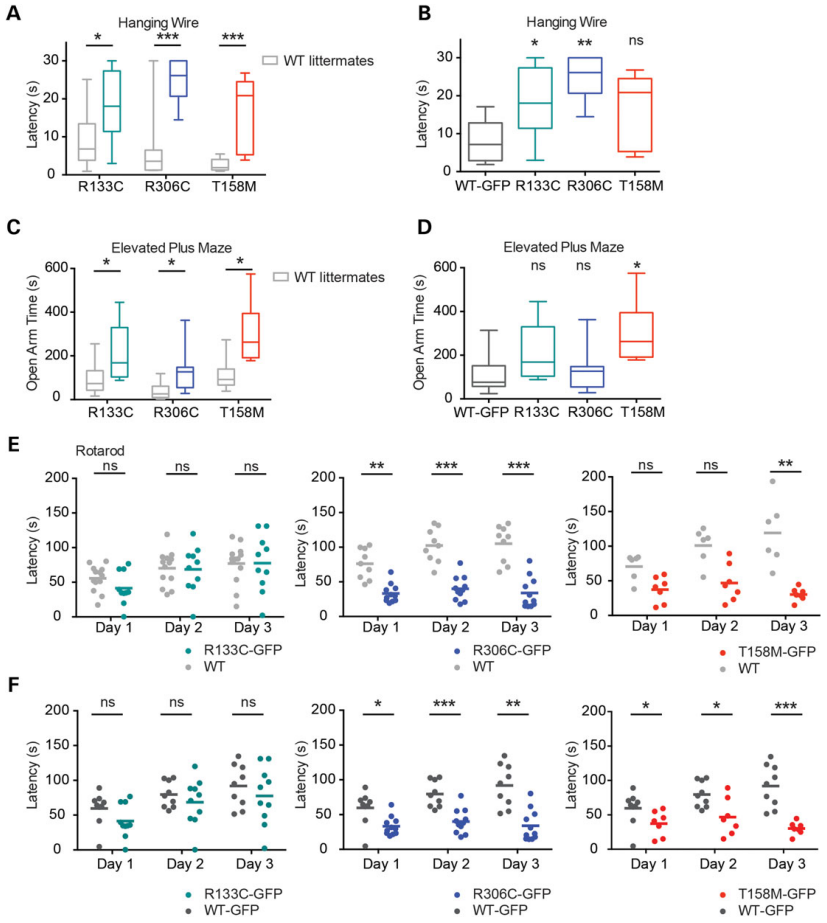


Figure 3. Behavioural analysis of the *Mecp2-GFP* allelic series indicates that *R133C-GFP* mice are less severely affected than *R306C-GFP* and *T158M-GFP*. Mutant males and WT male littermates underwent behavioural analysis as in Figure 1. (**A**) Hanging-wire test, (**C**) elevated plus maze and (**E**) accelerating rotarod. *R133C-GFP* ($n = 10$, green) plus WT littermates ($n = 13$, grey); *R306C-GFP* ($n = 10$–11, blue) plus WT littermates ($n = 9$, grey); *T158M-GFP* ($n = 7$, red) plus WT littermates ($n = 6$, grey). **B**, **D** and **F** show comparisons between mutant males and *WT-GFP* males in the same series of tests. (**E** and **F**) Graphs showing mean time to descent for individuals (dots) and cohorts (lines) for each trial day. Statistical analysis took all trials into account, and significance is denoted as follows: *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$ (Kolmogorov–Smirnov test). All behavioural paradigms were conducted on animals aged 8–10 weeks.

Taking advantage of GFP immunofluorescence, we examined localization of the different MeCP2 variants in hippocampal sections of hemizygous male mouse brains (Fig. 4C). The intensities of nuclear EGFP immunofluorescence in brain sections confirmed qualitatively that in comparison with WT-GFP protein, R306C-GFP expression was similar, R133C-GFP expression was moderately reduced and T158M-GFP was severely reduced in the brain. We focussed on sub-nuclear localization of the MeCP2 variants, as about half of all 5-methylcytosine in the mouse genome is concentrated in pericentromeric foci due to the abundance and CG richness of mouse satellite DNA sequences. Native MeCP2 localized to pericentromeric foci in a DNA methylation-dependent manner (21) as did WT-GFP and R306C-GFP fusion proteins. R133C-GFP and T158M-GFP, however,

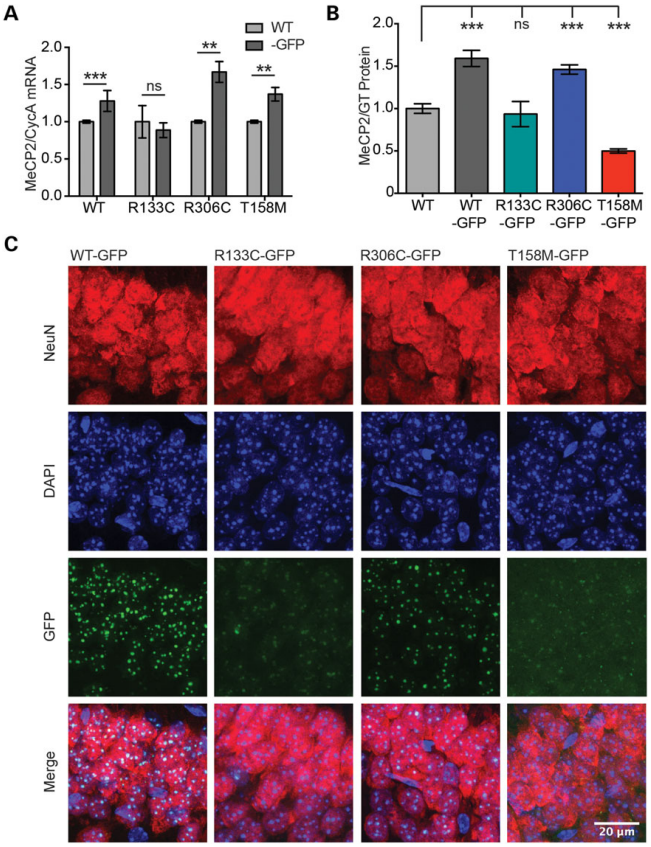


Figure 4. R133C-GFP and T158M-GFP are less abundant than WT-GFP and have an abnormal pattern of sub-nuclear localization. (**A**) Graph showing level of *Mecp2* transcript normalized to *Cyclophilin A* in mutant male mouse brains relative to WT littermates (mean ± SD). *WT-GFP*, $n = 9$; *R133C-GFP*, $n = 4$; *R306C-GFP*, $n = 3$; *T158M-GFP*, $n = 3$. (**B**) Quantification of western blots showing levels of WT-GFP, R306C-GFP, R133C-GFP and T158M-GFP protein in male mouse brain, relative to WT (mean ± SEM). GT served as an internal control; *WT*, $n = 6$; *WT-GFP*, $n = 6$; *R133C-GFP*, $n = 3$; *R306C-GFP*, $n = 4$; *T158M-GFP*, $n = 6$. (**C**) Representative images of the CA1 region of the hippocampus in adult male mice from the allelic series. Slices were imaged using the same confocal settings. Immunofluorescence was performed with DAPI (blue) and anti-NeuN (red). MeCP2 was imaged by virtue of its EGFP tag. R133C-GFP and T158M-GFP show mixed punctate and diffuse sub-nuclear localization. Scale bar = 20 µm. Statistical significance is denoted as follows: *$P < 0.05$, **$P < 0.01$ and ***$P < 0.001$ (two-tailed unpaired *t*-test). All biochemical analyses were conducted using tissues from adults aged 6–12 weeks.

showed a more diffuse distribution throughout the nucleoplasm coupled with reduced heterochromatic localization. Diffuse nuclear staining in comparison with WT-GFP and R306C-GFP mutant proteins was more apparent at higher confocal laser intensity, which showed background nuclear EGFP fluorescence in addition to heterochromatic localization (Supplementary Material, Fig. S3D). This effect was more severe for T158M-GFP than for R133C-GFP. An equivalent pattern was also seen in embryonic stem cell-derived neurons in culture (Supplementary Material, Fig. S3E). The results suggest compromised binding to mCG sites by both these mutants, as explored below.

Compromised binding of MeCP2[R133C] and MeCP2 [T158M] to modified DNA *in vivo* and *in vitro*

Published evidence regarding the DNA binding affinity of R133C protein is inconsistent. South-western (10,22) and EMSA (10,11,22,23) analyses revealed reduced binding of R133C to

methylated DNA. Expression of R133C protein from an integrated transgene in embryonic stem cells also revealed reduced specificity for methylated DNA (24). On the other hand, surface plasmon resonance indicated that the affinity of MeCP2 for CG-methylated DNA is unaffected by the R133C mutation, whereas binding to hydroxymethylated DNA is greatly reduced (15). In a transfection assay, over-expressed R133C-GFP localized normally to heterochromatin (12,24,25), although the mobility in a FRAP assay was somewhat increased, compatible with a modest reduction in affinity for DNA (25,26). In view of the somewhat variable observations, we repeated EMSAs using either the methylated DNA binding domain (MBD) alone (amino acids 77–167) or a larger N-terminal fragment of MeCP2 (amino acids 1–205). Interestingly, these two polypeptides showed very different DNA binding affinities when carrying the R133C mutation. The 77–167 fragment of R133C, which corresponds to the MBD alone, lost the ability to form a stable complex with methylated DNA, whereas 1–205 [R133C] was essentially WT by an equivalent assay (Fig. 5A and

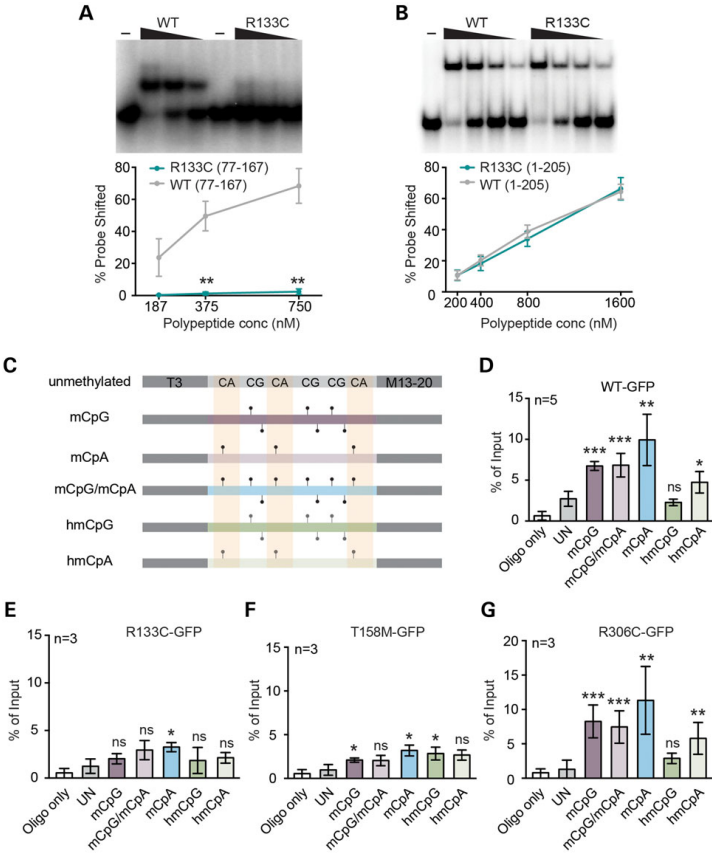


Figure 5. Defective binding of MeCP2[R133C] and T158M to modified DNA *in vitro* and *in vivo*. Representative gels and quantification of EMSAs measuring MeCP2 peptide binding to mCG probes using (**A**) aa77-167 or (**B**) aa1-205 fragments of MeCP2 with (green) or without (grey) the R133C mutation. Mean percentage of probe shifted ± SEM is plotted from three replicated experiments. Triangle represents decreasing peptide concentration; –, no peptide. Statistical significance is denoted as: $^{*}P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$ (unpaired two-tailed *t*-test). (**C**) ChIP oligonucleotide duplex probes containing symmetrical mCG or hmCG, or asymmetrical mCA or hmCA sites or both mCA and mCG sites together. Control DNA lacks all modifications. (**D–G**) Results of five independent transient transfection experiments followed by ChIP of MeCP2-associated probe oligonucleotide fragments. Cells transiently expressed WT-GFP, R133C-GFP, T158M-GFP and R306C-GFP, respectively. Mean % input ± SD plotted ($n \geq 3$).

B). These opposing findings with a common probe and under the same conditions recapitulate the spectrum of *in vivo* and *in vitro* findings in the literature for this mutant, and they leave open the question of whether this defect has any impact on chromatin association in living cells.

To address this issue, we used a cellular assay in which the various forms of MeCP2 are over-expressed transiently at equivalent levels in the presence of double-stranded oligonucleotides bearing specific modifications of cytosine (S. Lagger *et al.*, manuscript in preparation). The method has the advantage that it detects binding of full-length protein in living cells while allowing the modification status of the target DNA to be precisely defined. In this way, we could test binding not only to mCG, but also to methyl-CA (mCA) and hydroxymethyl-CA (hmCA), which are all implicated as target sequences for MeCP2 (27). Double-stranded oligonucleotides containing either three symmetrically methylated mCGs, three hmCGs or both mCA and mCG modifications (Fig. 5C) were transiently transfected into human embryonic kidney (HEK293) cells that transiently expressed WT MeCP2-GFP (WT-GFP) or mutant MeCP2-GFP from plasmid constructs. Unmodified oligonucleotides of the same sequence served as controls. Recovery of the target DNA by immunoprecipitation with anti-MeCP2 antibodies was monitored by quantitative PCR (qPCR). The results showed preferential binding of WT-GFP protein to DNA containing mCA, mCG and, less efficiently, hmCA compared with non-methylated DNA, confirming the specificity of the method (Fig. 5D). R306C-GFP, whose missense mutation is outside the DNA binding domain bound to methylated DNA indistinguishably from WT (Fig. 5G). Mutant R133C-GFP and T158M-GFP proteins, in contrast, showed greatly impaired binding to all modified oligonucleotides (Fig. 5E and F). Importantly, protein expression levels were closely similar between experiments (Supplementary Material, Fig. S4), and therefore the effects of the mutation on protein stability as seen in mouse brain do not affect these results. Of particular note, the R133C mutation significantly compromises the interaction between MeCP2 and all modified oligonucleotides *in vivo*, including mCG.

To examine the effects of mutations on MeCP2 binding in the native genome *in vivo*, we quantitatively analysed R133C-GFP and T158M-GFP binding in mouse brain by ChIP-qPCR using an anti-EGFP antibody. For comparison, we also analysed WT-GFP and R306C-GFP mutant proteins. The results showed that recovery of mouse satellite DNA, LINE1 elements, intracisternal A particles and the gene for brain-derived neurotrophic factor (*Bdnf*) from cross-linked brain chromatin by R133C-GFP was reduced to ~45–60% of the WT-GFP level, whereas T158M-GFP was 5–10% of WT-GFP (Fig. 6A). Binding of R306C-GFP protein was not significantly different from WT-GFP. We were concerned that reduced binding of the two MBD mutants might reflect their low abundance rather than an effect on binding affinities *per se*. We therefore increased the stringency of cross-linking by reducing the fixation temperature on ice. Under these conditions, ChIP of the R133C-GFP protein was reduced to ~15% of the WT-GFP level, which is significantly greater than the reduction in protein abundance of ~50%. It was noticeable that under these stringent conditions binding by R306C-GFP was also compromised, as reported by others (16), although this was not seen using our standard cross-linking protocol (Fig. 6B). The ChIP results, together with the *in vitro* data, transfection experiments and immunofluorescence, suggest that both T158M-GFP and R133C-GFP mutant proteins have a significantly reduced affinity for chromatin.

As MeCP2 is proposed to bind to both mCA- and hmC-containing DNA (16,27), we considered the possibility that the reduced ChIP signal from *R133C-GFP* brain may be due to compromised binding to these sequences rather than to mCG. To test this, we

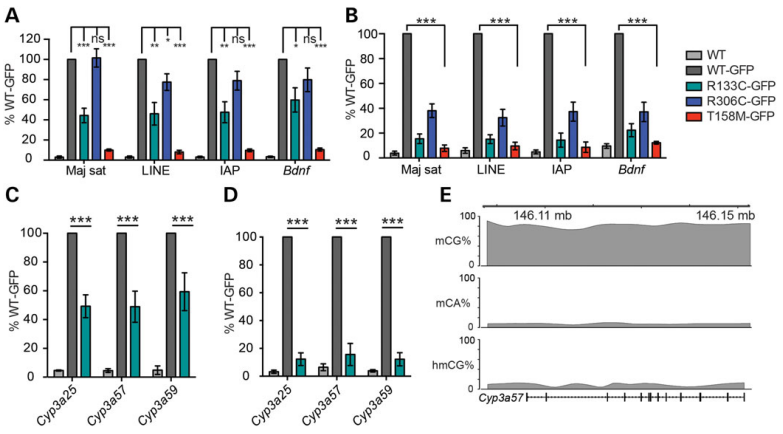


Figure 6. Chromatin immunoprecipitation reveals abnormal binding of R133C-GFP to mCG-rich repetitive sequences and genes in mouse brain. (**A**) ChIP-qPCR of WT-MeCP2 (lacking a GFP tag) and knocked-in WT-GFP, R133C-GFP, R306C-GFP and T158M-GFP in the adult male mouse brain. Primers amplify, respectively, major satellite, LINE and IAP transposable elements and the *Bdnf* gene locus. The antibody was against GFP and therefore does not precipitate untagged WT MeCP2 ($n \geq 3$). (**B**) Similar assay to (A) but with a lower and therefore more stringent cross-linking temperature ($n = 4$). (**C**) Column plot showing binding of R133C-GFP at the mCG-rich cytochrome p450 locus by GFP ChIP ($n = 3$). (**D**) The same assay as (C) but using a more stringent cross-linking temperature. Plots are expressed as % WT-GFP value. Error bars represent ± SEM, and statistical significance is denoted as: $^*P < 0.05$, $^{**}P < 0.01$ and $^{***}P < 0.001$ (unpaired, two-tailed *t*-test). Each replicate experiment used tissue from a separate animal. (**E**) Schematic of the *Cyp3a57* locus depicting cytosine methylation context. All biochemical analyses were conducted using tissues from adults aged 6–12 weeks.

probed immunoprecipitates for non-repetitive DNA sequences that contain mCG, but very low levels of mCA and hmCG. The ~900 kb cytochrome P450 gene locus, which is transcriptionally silent in the brain, meets this requirement as whole-genome bisulphite and TAB sequencing show a high level of CG methylation (28), while the other proposed MeCP2 binding sites, hmC and mCH, are rare (Fig. 6E). Three of the genes from this locus, *Cyp3a25*, *Cyp3a57* and *Cyp3a59* represented this pattern of methylation and were selected for ChIP analysis. Again, recovery of these sequences by ChIP was reduced in *R133C-GFP* brain extracts, reinforcing the conclusion that mCG binding by R133C-GFP is significantly compromised (Fig. 6C and D).

Modestly increased expression of long genes in the cerebellum of R133C-GFP mice

Previous studies have detected subtle changes in gene expression in the brains of mice that are *Mecp2*-null or carry RTT mutations (29). We compared gene expression in cerebellums of *R133C-GFP* and *WT-GFP* mice using expression microarrays, along with *WT* and *Mecp2*-null mice for reference. The results initially showed statistically significant differences at six individual loci in *R133C-GFP* cerebellum compared with *WT-GFP*, namely *Abhd1*, *Aplf*, *Dag1*, *Zfp428* (all down-regulated), *Mybpc3* and *Krt222* (*up-regulated*), but only two of these were reproduced by qPCR. In one case, the effect was very small. In the other case (*Abhd1*), the *R133C-GFP* expression was unchanged relative to *WT* or null comparators. In fact, the variance in expression was attributable to differences between WT-GFP cerebellum and the other tested tissues (Supplementary Material, Fig. S5A). We explored the data further by binning genes according to length to replicate recent studies showing that long genes are disproportionately up-regulated in *Mecp2* mutant brains (27,30). The results showed a small but sustained length-dependent increase in expression of longer genes in *R133C-GFP* versus *WT-GFP* cerebellum (Supplementary Material, Fig. S5B). In this study, the up-regulation of long genes was modest in the *Mecp2*-null versus *WT* cerebellum, which was in agreement with previous similar analyses of this brain region (27).

Discussion

MeCP2 is identical between human and mouse throughout the great majority of its sequence, suggesting that almost all residues of the protein are essential for full function. Despite long-range conservation of the primary sequence across evolution, the distribution of missense mutations causing RTT is highly non-random (13). Most missense mutations are concentrated in the MBD and NID domains, disrupting the interaction with DNA and NCoR/SMRT, respectively (7). All of the most common RTT mutations, including nonsense truncations, potentially involve C to T transitions at CG sites and are therefore most likely caused by 5-methylcytosine deamination coupled with failure of repair (31). Their high frequency of occurrence therefore reflects elevated mutability. The most common missense mutation is T158M, which disrupts hydrogen-bonding integral to maintaining the configuration of an ASX turn (Asp156–Phe157–Thr158) and following ST motif (Thr158–Val159–Thr160–Gly161) in the C-terminal region of the MBD. This has been shown to destabilize MBD structure and reduce affinity for methylated DNA (20). Next most frequent is R306C, which is located in the NID and prevents the interaction of MeCP2 with the NCoR/SMRT co-repressor complexes (13). Least severe on average of the common RTT mutations is R133C, which mutates an arginine residue that in the X-ray structure makes hydrogen bond contact with one of the guanines in the symmetrical CG dyad (20). Individuals with this mutation more frequently retain some speech and the ability to walk (8,14). We found that phenotypic severity in mice matches the respective clinical severity of these mutations in humans, suggesting that the molecular basis of the resulting disorders is the same in the two species.

The effects on DNA binding of MeCP2[T158M] to DNA are somewhat inconsistently reported in the literature. Previous *in vitro* studies of full-length T158M protein showed either mild (10) or severe (22) impairment of mCG binding by South-western assay. EMSA analysis, using the MBD fragment, also revealed ambiguous findings. Reduced binding to methylated DNA (10,20,23) was seen in some instances but not others (11). Some of these discrepancies may be attributed to the positions of N- and C-terminal affinity purification tags (23). *In vivo* analysis of murine fibroblasts over-expressing full-length MeCP2[T158M] indicated defective binding to heterochromatic foci, and FRAP analysis showed greater mobility of MeCP2[T158M] compared with WT protein, indicating a less stable association with chromatin (25,26). Most published evidence therefore supports the view that the T158M mutation compromises binding between MeCP2 and methylated DNA. Our ChIP data based on transient transfection of cultured cells and brain confirm this defect in chromatin binding. These findings agree with an analysis of a different *Mecp2* allelic variant involving the same amino acid, T158A (6).

Molecular and cell biological assays allowed us to critically assess a recent hypothetical explanation for the milder phenotype in *R133C-GFP* mice. It was proposed that MeCP2[R133C] retains WT levels of binding to mCG *in vitro*, but has specifically lost the ability to interact with hydroxymethylated DNA (15). Our findings place this hypothesis in doubt as several lines of evidence show that mCG binding by MeCP2 is significantly compromised by this mutation: (i) immunofluorescence analysis in differentiated cultured neurons and in hippocampal sections showed diffuse nuclear staining with reduced localization to densely methylated heterochromatic foci; (ii) MBD DNA binding *in vitro* was defective by EMSA, in agreement with most previous reports (though a longer fragment of MeCP2 displayed normal binding); (iii) a cellular assay showed that full-length R133C-GFP binding to mCG was greatly reduced; and (iv) ChIP analysis of genomic regions where mCG was the only detectable modification showed greatly reduced binding by the mutant protein in the brain. The brain ChIP data are particularly compelling, as mouse satellite sequences and a gene cluster, both of which lack hmC and mCH, showed impaired binding *in vivo*. In addition to this loss of mCG binding, transfection experiments showed a significant reduction in MeCP2[R133C] binding to the alternative canonical binding sites mCA and hmCA.

Despite occupying a key position in the MBD of MeCP2, the R133C mutation results in a milder RTT phenotype and displays residual mCG binding. These observations can be reconciled by modelling the MeCP2–DNA interaction at a molecular level (Fig. 7). The X-ray structure shows that two arginines, R133 and R111, form equivalent hydrogen bonds with the two guanine residues located on opposite strands at the self-complementary dinucleotide sequence 5′CG (20) and also interact with 5-methylcytosine (Fig. 7A). Replacement of R111 by glycine causes RTT, as expected if DNA binding is severely disrupted. Substitution of R133 by cysteine might be expected to have an equally severe effect, but in fact residual DNA binding persists. A notable difference between R111 and R133 is that the side chain of the former adopts an all-trans extended configuration and its side-chain amino group is held in this position by aspartic acid 121

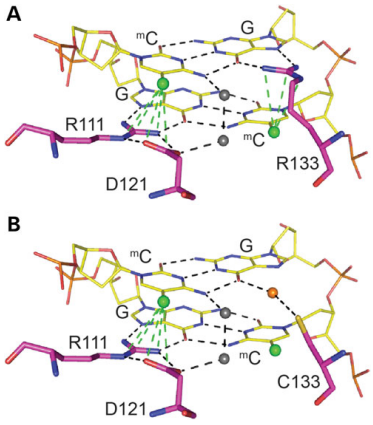


Figure 7. Predictive modelling of MeCP2 MBD binding to methylated DNA reveals a loss of specificity with the R133C mutation. (**A**) Predicted model of MBD binding to a methylated CpG pair. Arginine 111 and 133 make contact with the guanines and form hydrogen bonds with the methylcytosines of the CpG pair. Arginine 111 forms a hydrogen bond 'clip' with aspartic acid 121. (**B**) Predicted model of MBD [R133C] binding to a methylated CpG pair. The cysteine 133 interaction with guanine is now dependent on water molecules, and there is no significant interaction with the methylcytosine. (A and B) Hydrogen bonds are indicated by dotted lines, methyl groups by green balls, and water molecules by grey and orange balls.

via a hydrogen bond 'clip'. Loss of this interaction probably has structural consequences for the MBD as a whole. In contrast, the R133 side chain is relatively untethered, forming a single salt bridge with the carboxylate group of glutamate at position 137 (20). Replacement of R133 by the compact cysteine residue is unlikely to impact the integrity of the MBD, but due to loss of interactions with mC and G will reduce specificity. By placing a hypothetical water molecule in the major groove, it is possible to generate hydrogen bonding with G (Fig. 7B), which may account for the retention of weak DNA binding. We and others have previously reported that WT MeCP2 fails to interact with hmCG *in vitro* (27,32–34). As hmCG is by far the most abundant setting for hmC in the brain (28), it is unclear which hmC moieties might be targeted *in vivo*. Further work may be needed to define the full range of DNA binding specificities for both WT and mutant forms of MeCP2.

The R306C mutation occurs outside the MBD of MeCP2 and has no obvious effect on its stability, but may weakly affect chromatin binding (see also (16)). DNA binding impairment of this protein domain has not been characterized biochemically and was not detected in our cellular assay with full-length protein. It is possible that it is indirect, due for example to differences in extraction or altered association with partners. Further work is needed to address this issue. The two MBD mutations T158M and R133C, however, display both compromised stability and weaker DNA binding. Therefore, it is unlikely that a reduced affinity for chromatin alone underlies their clinical effects. In both cases, MeCP2 abundance was found to be lower than WT-GFP in knock-in mouse lines and when independent knock-in ESCs were differentiated into neurons in culture. R133C-GFP exhibited a less pronounced reduction in abundance (~55%) in mice compared with T158M-GFP (~30%). We conclude that the reduced abundance of T158M-GFP and R133C-GFP proteins is an intrinsic property conferred on the protein by the respective mutations. We speculate that the differing average severity associated with these two mutations is primarily due to their contrasting abundance. If both mutant proteins were present at the WT level, through pharmacological stabilization or over-expression, we speculate that they would exhibit similar pathologies at the milder end of the RTT spectrum.

The introduction of specific RTT mutations into mice confirms the value of this model system in studying the molecular basis of this condition. Generalizations are starting to emerge—for example, three missense mutations (T158A, T158M and R133C) have now been found to destabilize the protein and almost certainly make a significant contribution to the phenotype (this work and reference (6)). It is likely that the high abundance of MeCP2 in neurons is dependent to some extent on stability and the highly conserved amino acid sequence may serve to insure this. A measure of protein stability derived from gene ablation studies in adult mice suggested that MeCP2 has a half-life of ~2 weeks in the brain (19), which is longer than average. A second generalization from studies of missense mutations is that disruption of the interactions between MeCP2 and DNA (R133C, T158M, T158A) or between MeCP2 and NCoR (R306C) is associated with RTT (13). It will be of interest to discover if modelling of additional RTT mutations in mice adds to this list of molecular causes. This information is likely to be essential in the search for rational therapeutic approaches to mutation-specific aetiologies underlying this condition.

Materials and Methods

Knock-in of *Mecp2* alleles

Targeting vectors to create EGFP-tagged alleles for WT *Mecp2*, *Mecp2R133C*, *Mecp2T158M* and *Mecp2R306C* were constructed using a 7.2-kb plasmid subclone of mouse 129/Ola genomic DNA (3), including *Mecp2* exons 3 and 4 (Supplementary Material, Fig. S1). The coding sequence of *EGFP* was fused in-frame to the end of the coding sequence of *Mecp2* in exon 4 followed by a *loxP*-flanked NeoStop cassette as a selectable marker, retaining the first 1.8 kb of the *Mecp2* 3′ UTR as the 3′ homology arm. Point mutations R133C, T158M and R306C were introduced into the WT *Mecp2-EGFP* targeting vector using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). Linearized constructs were electroporated into 129/Ola E14 TG2a mouse ES cells and correctly targeted clones identified by PCR screening and Southern blotting (Supplementary Material, Fig. S1). Mice were generated from ES cells by standard procedures (3). The *loxP*-flanked selection cassette was either removed *in vitro* for ES cell differentiation into neurons by electroporation of targeted clones with pCAGGS-Cre plasmid, or *in vivo* by test mating chimeras with the CMV-Cre deleter strain. Mice have been submitted to Jackson Laboratories *T158M-GFP*: JAX Stock No. 26762 B6.Cg-*Mecp2*$^{tm4.1Bird}$/J, *R133C-GFP*: JAX Stock No. 26848 B6.Cg-*Mecp2*$^{tm6.1Bird}$/J and *R306C-GFP*:JAX Stock No. 26847 B6.Cg-*Mecp2*$^{tm5.1Bird}$/J.

RNA extraction and qPCR

Frozen half brains were homogenized in Tri Reagent (Sigma) using the Ultra-Turrax T25. RNA was extracted according to the manufacturer's instructions. RNA was DNase I treated (DNA-*free* kit, Ambion) and then reverse transcribed (iScript cDNA Synthesis Kit, Bio-Rad) according to the manufacturer's instructions. Fifty

nanograms of cDNA was amplified in qPCR using SensiMix SYBR & Fluorescein Master Mix (Bioline) with the following primers: Me2Ex3/4 FWD ACCTTGCCTGAAGGTTGGAC, REV GCAATCAA TTCTACTTTAGAGCGAAAA; CypA FWD TCGAGCTCTGAGCACTG GAG and REV CATTATGGCGTGTAAAGTCACCA. *Mecp2* mRNA was normalized to *Cyclophilin A* housekeeping gene expression. EGFP mutants were compared with WT littermates.

Protein extraction and western blot

One frozen half brain was homogenized in ice-cold NE1 buffer [20 mM Hepes pH 7.9, 10 mM KCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 20% glycerol, 0.5 mM DTT, protease inhibitor (complete EDTA free cocktail, Roche)] before adding 750 units of benzonase for 15 min at room temperature then an equal volume of 2× SDS loading buffer (0.125 M Tris pH 6.8, 20% glycerol, 4% SDS, 0.25% bromophenol blue, 20 mM DTT, 0.3 M β-mercaptoethanol). Samples were boiled for 3 min before loading equal volumes on a 4–12% Run Blue SDS precast gel (Expedeon). Gels were transferred to a 0.2-μm nitrocellulose membrane (Bio-Rad) and then blocked for 1 h (5% milk, 1× TBS, 0.1% Tween) before applying anti-Mecp2 1:1000 (Sigma M6818) or anti-γ-tubulin 1:3000 (Sigma T5326) overnight at 4°C, followed by IRDye 800CW donkey anti-mouse IgG 1:10 000 (LiCor) for 2 h at RT after washing. Membranes were imaged using the Odyssey Infrared Imager (LiCor) and quantified using Image Studio Lite Software (LiCor).

Phenotypic analysis

Mice tested were of fourth or fifth generation back-crossed to C57Bl/6J (N4 or N5). Scoring of symptoms and behavioural analyses were all performed blind to the genotype. Six parameters were examined at the same time each week (activity, gait, hindlimb clasping, tremor, breathing and general condition) and given a score between 0 and 2, as previously described (19). Animals that scored 2 for tremor, breathing or general condition, or which had lost 20% of their body weight had reached the severity limit of the experiment according to the Home Office License and were humanely culled. Mice were also weighed weekly. Behavioural analysis was done between 8 and 10 weeks in the order (1) elevated plus maze, (2) hanging wire and then (3) accelerating rotarod. For details of behavioural assays, see (19). In brief, for elevated plus maze, mice were placed in a cross-shaped maze 65 cm above the floor with 2 open arms (20 × 8 cm), 2 closed arms (20 × 8 × 25 cm) and a central area (8 × 8 cm) in uniform dim lighting. Time spent in open arms of the maze was visualized using ANY-maze software (Stoelting). For the hanging wire, time taken to bring one hind paw to a 1.5 mm diameter wire, 35 cm above the bench, after mice were suspended by forepaws was recorded for up to 30 s over 3 trials, inter-trial interval 15 min. For the accelerating rotarod, mice had one habituation training day then time taken to fall from a 3 cm diameter rotating rod, accelerating between 4 and 40 RPM over 5 min was recorded. Mice had 4 trials per day (inter-trial interval 1 h) over 3 days.

Immunofluorescence

Mice were perfused with 3.7% paraformaldehyde in PBS, pH 7.4, then immersion fixed overnight. The brains were then dehydrated in 30% sucrose for 24 h, washed briefly in PBS, blotted dry, halved and snap frozen in isopentane cooled on dry ice. The brains were sectioned at 10 μm (Leica CM 1900 Cryostat) and mounted. Slides were washed in PBS, permeabilized in 0.1% Triton X-100 for 15 min, washed, blocked for 1 h at RT in 1.5% normal goat serum, stained with anti-NeuN-Cy3 1:100 (MAB377C3 Millipore) overnight at 4°C then finally, washed, stained with 1 μg/ml DAPI and washed again. Slides were mounted in Prolong Gold Antifade Reagent (Molecular Probes). The brains were visualized on a LeicaSP5 Confocal Laser Scanning Microscope. For clarity, a set of images were taken where the EGFP signal was optimized for each genotype individually. A look-up table (LUT) was applied to hippocampal CA1 images, where blue pixels denoted over-saturation and green pixels denoted zero. Photon multiplier tube (PMT) gain was kept constant, and laser power was adjusted until one or two pixels were blue and as few pixels as possible were green. These settings were saved and used for all subsequent images for that genotype.

Electrophoretic mobility shift assays

Recombinant MeCP2 protein expression vectors were constructed with a C-terminal histidine tag in the bacterial expression plasmid pET30b (Novagen). The R133C point mutation was introduced into the WT plasmid using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) as per the manufacturer's instructions. BL21(DE3)pLysS competent cells were transformed, and a scrape of colonies was inoculated in a starter culture overnight, then into 500 ml warm LB (50 μg/ml kanamycin; 17 μg/ml chloramphenicol) and shaken at 37°C. Cultures were induced with 1 mM IPTG when the OD was 0.6–0.8 at 600 nm and then grown at 30°C for 3 h. Pellets were mashed in ice-cold lysis buffer [50 mM $NaH2PO_4$, 100 mM NaCl, 10% glycerol, 30 mM imidazole, 0.1% NP40, protease inhibitor (complete EDTA free cocktail, Roche), pH 7.5] and passed through a 21G needle prior to adding 750 units of benzonase then being sonicated 30 s on/off for 10 cycles (Branson Digital Sonifier). Samples were adjusted to 0.3 M NaCl, and then after centrifugation, 0.5 ml bead volume $NiSO_4$-coated Chelating Sepharose Fast Flow beads (GE Healthcare) in lysis buffer was added to lysates for 1 h at 4°C. Beads were washed 3 times in lysis buffer before MeCP2 polypeptides were eluted with 250 mM imidazole lysis buffer in 5 fractions. Eluted fractions were pooled, adjusted to 1 mM EDTA and then dialysed [Slide-A-Lyzer Dialysis Cassette 7 K MWCO (Thermo Scientific)] overnight at 4°C in 0.1 M HEPES buffer (20 mM HEPES pH7; 100 mM NaCl; 1 mM EDTA; 5 mM β-mercaptoethanol). The polypeptides underwent a second purification step and were eluted from a HiTrap SP HP 1 ml Column (GE Healthcare) using a 0.7 M NaCl HEPES buffer. An oligonucleotide probe from the mouse *Bdnf* promoter region 5′ AAGCATG CAATGCCCTGGAA**C**G**G**AATTCTTCTAATAAAAGATGTATCATTTTA AATGC 3′ (Biocore), plus complementary reverse strand, was methylated or unmethylated at the central CG indicated in bold italics. Five hundred nanograms of probe was radio-labelled using T4 polynucleotide kinase (NEB) and purified (MinElute PCR Purification Kit, Qiagen) according to the manufacturer's instructions. One nanogram probe and 1 μg poly deoxyadenylic-thymidylic acid competitor DNA (Sigma-Aldrich) were added to polypeptide in reaction buffer (5% glycerol, 0.1 mM EDTA, 10 mM Tris HCl pH7.5, 150 mM KCl, 0.1 mg/ml BSA) on ice for 20 min. Samples were run at 120 V for 70 min on a 10% acrylamide tris borate EDTA gel (0.075% APS, 0.00125% TEMED) in chilled TBE. The gels were exposed overnight and imaged using the Typhoon FLA 9500 scanner (GE Healthcare). Percentage shift was calculated using Image J software.

Cellular immunoprecipitation

Using Lipofectamine 2000 (Life technologies), 1.5×10^6 HEK 293 FT cells were transfected over night with 0.5 μg of MeCP2 WT or

R133C, T158M and R306C mutants tagged with EGFP, according to the manufacturer's instructions. After assessment of transfection efficiency, the medium was changed and cells were transfected with annealed unmodified or methylated oligonucleotides (100 nM) final concentration using TransIT Oligofect reagent (Mirus) for 4 h. Cells were washed with PBS and harvested by trypsinisation. In the next step, cells were cross-linked with 1% formaldehyde for 5 min at room temperature and quenched by the addition of glycine to a final concentration of 0.125 M for 5 min. At this stage, cell pellets could be flash frozen in liquid nitrogen and stored at –80°C or directly used for DNA–protein complex isolation and consecutive IP. Oligonucleotide sequence: 5′ ATGCTAATTAACCCTCACTAAAGGGAACTCGAGACATCGGAGAATTCACATCACCGGTGAATCAGTGCTACCCGCAAGTGCACTGGATCCACTGGCCGTCGTTTTACAA 3′. The transfected fragment consisted of an artificial sequence (49% GC content) flanked by T3 and M13-20 primer binding sites for unique amplification. Primer sequences are underlined. Differentially methylated CGs are highlighted in bold. Soluble chromatin preparation and chromatin immunoprecipitation assays were carried out as described previously with modifications. In short, chromatin was sonicated using a Twin Bioruptor (Diagenode) 30 s on/off for 15 cycles. Equal amounts of chromatin were used for IP with 5 μg MeCP2 M6818 antibody (Sigma) and incubated overnight. Protein–antibody complexes were bound to magnetic protein G beads (Life technologies) for 4–5 h and washed with standard IP wash buffers for 10 min at 4°C. The cross-link was reversed by addition of 0.05 volume of 4 M NaCl over night at 65°C. After proteinase K digestion, DNA was recovered by phenol–chloroform–isoamylalcohol extraction and dissolved in 200 μl H_2O. Real-time PCR of ChIP DNA and corresponding input DNA was performed using T3 and M13-20 primers.

Brain chromatin immunoprecipitation

Frozen half brains were Dounce homogenized in 1 ml PBS and then fixed in 1% formaldehyde for 1 min at room temperature or 60 s after being on ice (stringent ChIP) before quenching with 250 mM glycine. Chromatin immunoprecipitation assays were carried out as described previously with modifications. In short, chromatin was sonicated using a Twin Bioruptor (Diagenode) 30 s on/off for 15 cycles. Equal amounts of chromatin were used for IP with 40 μl GFP-Trap_A beads (Chromotek) and rotated for 1.5 h at 4°C. Protein–antibody complexes were washed with standard IP wash buffers for 4 min at 4°C. The cross-link was reversed by addition of 0.05 volume of 4 M NaCl overnight at 65°C. After proteinase K digestion, DNA was recovered by phenol–chloroform–isoamylalcohol extraction and dissolved in 200 μl H_2O. Real-time PCR of ChIP DNA and corresponding input DNA was performed using specific primers. Major Satellite FWD GGC GAGAAAACTGAAAATCACG, REV AGGTCCTTCAGTGTGCATTTC; LINE1 Elements FWD TTTGGGACACAATGAAAGCA, REV CTGCC GTCTACTCCTCTTGG; IAP FWD GAGATTGGACTTTTGACTTGT, REV TGTGGCTTGCTCATAGATTAG; *Bdnf* FWD TTCGATTCACG CAGTTGTTC, REV CTGAGCCAGTTACGTGACCA; Cyp3a25 FWD CAGGTTTGGGGTGTTGTGAA, REV CTGCAGCTGTTGTGGGAG; Cyp3a57 FWD GTGCTGCTCTTACATGGCTG, REV GTGGGGCTA CAGTCTATGCT; Cyp3a59 FWD CCTGACTGGCTGCTCACTAT, REV AGGCTGTGAACTATAGGAGCC.

Statistics

Behavioural data were analysed using the Kolmogorov–Smirnov test with the Benjamini and Hochberg method of correction for false discovery with multiple testing. Survival data were analysed using the Mantel-Cox test. Biochemical data were analysed using two-tailed, unpaired *t*-tests.

Study approval

Mice were maintained under standard conditions and in accordance with UK Home Office regulations and licenses.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

We are grateful to Alan McClure for animal husbandry, Dr David Kelly for imaging support and Professor M Walkinshaw for structural modelling. We also thank Dr Michael Greenberg and Dr Gail Mandel of the RSRT Consortium and their laboratory members for constructive comments during this work.

Conflict of Interest statement. None declared.

Funding

This work was supported by a consortium grant from the Rett Syndrome Research Trust, by a Wellcome Trust Edinburgh Clinical Academic Training studentship to Kyla Brown (100670), by a Wellcome Trust Programme Grant (091580) and by a Wellcome Trust Centre Core Grant (092076). S.L. was funded by a postdoctoral EMBO long-term fellowship (ALTF 1467-2011). Funding to pay the Open Access publication charges for this article was provided by The University of Edinburgh with support from The Wellcome Trust.

Authors' Contributions

K.B. and J.S. carried out mouse behaviour, biochemical analyses and immunohistochemistry. S.L. did cellular immunoprecipitation and qPCR validation of microarrays. S.L. and K.B. did ChIP. J.C. and K.B. did EMSAs. D.S., K.B. and M.K. did mouse phenotypic scoring. J.G. designed and engineered the targeting vector. J.S. and C.M. mutated targeting vectors. J.S., J.G., K.B. and C.M. did ES cell targeting. J.S. established all mouse lines. K.B. did neuronal differentiation and microarrays. J.S. and S.L. did FACS analysis. A.K. and S.W. analysed mouse behavioural data and microarray data, respectively. A.B. and K.B. produced the manuscript.

References

1. 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. *Nature Genet.*, **23**, 185–188.
2. Meins, M., Lehmann, J., Gerresheim, F., Herchenbach, J., Hagedorn, M., Hameister, K. and Epplen, J.T. (2005) Submicroscopic duplication in Xq28 causes increased expression of the MECP2 gene in a boy with severe mental retardation and features of Rett syndrome. *J Med Genet.*, **42**, e12.
3. 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.
4. 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.

5. Brendel, C., Belakhov, V., Werner, H., Wegener, E., Gartner, J., Nudelman, I., Baasov, T. and Huppke, P. (2011) Readthrough of nonsense mutations in Rett syndrome: evaluation of novel aminoglycosides and generation of a new mouse model. *J Mol Med (Berl)*, **89**, 389–398.
6. Goffin, D., Allen, M., Zhang, L., Amorim, M., Wang, I.T., Reyes, A.R., Mercado-Berton, A., Ong, C., Cohen, S., Hu, L. *et al.* (2012) Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. *Nat Neurosci.*, **15**, 274–e83.
7. Lyst, M.J. and Bird, A. (2015) Rett syndrome: a complex disorder with simple roots. *Nat Rev Genet*, **16**, 261–275.
8. Neul, J.L., Fang, P., Barrish, J., Lane, J., Caeg, E.B., Smith, E.O., Zoghbi, H., Percy, A. and Glaze, D.G. (2008) Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology*, **70**, 1313–1321.
9. 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.
10. Ballestar, E., Yusufzai, T.M. and Wolffe, A.P. (2000) Effects of Rett syndrome mutations of the methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry*, **39**, 7100–7106.
11. Free, A., Wakefield, R.I.D., Smith, B.O., Dryden, D.T.F., Barlow, P.N. and Bird, A. (2001) DNA recognition by the methyl-CpG binding domain of MeCP2. *J Biol Chem.*, **276**, 3353–3360.
12. Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Schanen, C. and Tamura, M. (2003) Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J Med Genet*, **40**, 487–493.
13. Lyst, M.J., Ekiert, R., Ebert, D.H., Merusi, C., Nowak, J., Selfridge, J., Guy, J., Kastan, N.R., Robinson, N.D., de Lima Alves, F. *et al.* (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci*, **16**, 898–902.
14. Cuddapah, V.A., Pillai, R.B., Shekar, K.V., Lane, J.B., Motil, K.J., Skinner, S.A., Tarquinio, D.C., Glaze, D.G., McGwin, G., Kaufmann, W.E. *et al.* (2014) Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. *J Med Genet*, **51**, 152–158.
15. Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S. and Heintz, N. (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell*, **151**, 1417–1430.
16. Heckman, L.D., Chahrour, M.H. and Zoghbi, H.Y. (2014) Rett-causing mutations reveal two domains critical for MeCP2 function and for toxicity in MECP2 duplication syndrome mice. *Elife*, **3**, e02676.
17. Schmid, R.S., Tsujimoto, N., Qu, Q., Lei, H., Li, E., Chen, T. and Blaustein, C.S. (2008) A methyl-CpG-binding protein 2-enhanced green fluorescent protein reporter mouse model provides a new tool for studying the neuronal basis of Rett syndrome. *Neuroreport*, **19**, 393–398.
18. 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.
19. Cheval, H., Guy, J., Merusi, C., De Sousa, D., Selfridge, J. and Bird, A. (2012) Postnatal inactivation reveals enhanced requirement for MeCP2 at distinct age windows. *Hum Mol Genet*, **21**, 3806–3814.
20. 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.
21. Nan, X., Tate, P., Li, E. and Bird, A. (1996) DNA methylation specifies chromosomal localization of MeCP2. *Mol Cell Biol*, **16**, 414–421.
22. 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.
23. Ghosh, R.P., Horowitz-Scherer, R.A., Nikitina, T., Gierasch, L.M. and Woodcock, C.L. (2008) Rett syndrome-causing mutations in human MeCP2 result in diverse structural changes that impact folding and DNA interactions. *J Biol Chem*, **283**, 20523–20534.
24. Baubec, T., Ivanek, R., Lienert, F. and Schubeler, D. (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell*, **153**, 480–492.
25. Schmiedeberg, L., Skene, P., Deaton, A. and Bird, A. (2009) A temporal threshold for formaldehyde crosslinking and fixation. *PLoS One*, **4**, e4636.
26. Kumar, A., Kamboj, S., Malone, B.M., Kudo, S., Twiss, J.L., Czymmek, K.J., LaSalle, J.M. and Schanen, N.C. (2008) Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J Cell Sci.*, **121**(Pt 7), 1128–1137.
27. Gabel, H.W., Kinde, B., Stroud, H., Gilbert, C.S., Harmin, D.A., Kastan, N.R., Hemberg, M., Ebert, D.H. and Greenberg, M.E. (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature*, **522**, 89–93.
28. Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D. *et al.* (2013) Global epigenomic reconfiguration during mammalian brain development. *Science*, **341**, 1237905.
29. Tudor, M., Akbarian, R., Chen, R. and Jaenisch, R. (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional chmages in the brain. *PNAS*, **99**, 15536–15541.
30. Sugino, K., Hempel, C.M., Okaty, B.W., Arnson, H.A., Kato, S., Dani, V.S. and Nelson, S.B. (2014) Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. *J Neurosci*, **34**, 12877–12883.
31. Cooper, D.N. and Youssoufian, H. (1988) The CpG dinucleotide and human genetic disease. *Hum Genet*, **78**, 151–155.
32. 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.
33. Khrapunov, S., Warren, C., Cheng, H., Berko, E.R., Greally, J.M. and Brenowitz, M. (2014) Unusual characteristics of the DNA binding domain of epigenetic regulatory protein MeCP2 determine its binding specificity. *Biochemistry*, **53**, 3379–3391.
34. Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X. and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res*, **40**, 4841–4849.

References

Adachi M, Autry AE, Covington HE & Monteggia LM (2009) MeCP2-mediated transcription repression in the basolateral amygdala may underlie heightened anxiety in a mouse model of Rett syndrome. *J. Neurosci.* **29:** 4218–4227

Adams VH, McBryant SJ, Wade P a., Woodcock CL & Hansen JC (2007) Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J. Biol. Chem.* **282:** 15057–15064

Agarwal N, Becker A, Jost KL, Haase S, Thakur BK, Brero A, Hardt T, Kudo S, Leonhardt H & Cardoso MC (2011) MeCP2 Rett mutations affect large scale chromatin organization. *Hum. Mol. Genet.* **20:** 4187–95

Alvarez-Saavedra M, Sáez M a, Kang D, Zoghbi HY & Young JI (2007) Cell-specific expression of wild-type MeCP2 in mouse models of Rett syndrome yields insight about pathogenesis. *Hum. Mol. Genet.* **16:** 2315–25

Amir RE, Van den Veyver IB, Schultz R, Malicki DM, Tran CQ, Dahle EJ, Philippi a, Timar L, Percy a K, Motil KJ, Lichtarge O, Smith EO, Glaze DG & Zoghbi HY (2000) Influence of mutation type and X chromosome inactivation on Rett syndrome phenotypes. *Ann. Neurol.* **47:** 670–9

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U & Zoghbi HY (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23:** 185–8

Antequera F & Bird A (1999) CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr. Biol.* **9:** 661–667

Archer H, Evans J, Leonard H, Colvin L, Ravine D, Christodoulou J, Williamson S, Charman T, Bailey MES, Sampson J, de Klerk N & Clarke A (2007) Correlation between clinical severity in patients with Rett syndrome with a p.R168X or p.T158M MECP2 mutation, and the direction and degree of skewing of X-chromosome inactivation. *J. Med. Genet.* **44:** 148–152

Ariani F, Hayek G, Rondinella D, Artuso R, Mencarelli MA, Spanhol-Rosseto A, Pollazzon M, Buoni S, Spiga O, Ricciardi S, Meloni I, Longo I, Mari F, Broccoli V, Zappella M & Renieri A (2008) FOXG1 Is Responsible for the Congenital Variant of Rett Syndrome. *Am. J. Hum. Genet.* **83:** 89–93

Armstrong D (2005) Neuropathology of Rett Syndrome. *J. Child Neurol.* **20:** 747–753

Armstrong DD (2002) Neuropathology of Rett syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.* **8:** 72–6

Armstrong DD, Dunn K & Antalffy B (1998) Decreased dendritic branching in motor cortex in Rett and DS. 1013–1017

Auranen M, Vanhala R, Vosman M, Levander M, Varilo T, Hietala M, Riikonen R, Peltonen L & Järvelä I (2001) MECP2 gene analysis in classical Rett syndrome and in patients with Rett-like features. *Neurology* **56:** 611–617

Bahi-Buisson N, Nectoux J, Rosas-Vargas H, Milh M, Boddaert N, Girard B, Cances C, Ville D, Afenjar A, Rio M, Héron D, N'Guyen Morel MA, Arzimanoglou A, Philippe C, Jonveaux P, Chelly J & Bienvenu T (2008) Key clinical features to identify girls with CDKL5 mutations. *Brain* **131:** 2647–2661

Baker SA, Chen L, Wilkins AD, Yu P, Lichtarge O & Zoghbi HY (2013) An AT-Hook Domain in MeCP2 Determines the Clinical Course of Rett Syndrome and Related Disorders. *Cell* **152:** 984–996

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

Ballestar E, Yusufzai TM & Wolffe a P (2000) Effects of Rett syndrome mutations of the methyl-CpG binding domain of the transcriptional repressor MeCP2 on selectivity for association with methylated DNA. *Biochemistry* **39:** 7100–6

Barbot W, Dupressoir A, Lazar V & Heidmann T (2002) Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic Acids Res.* **30:** 2365–2373

Baubec T, Ivánek R, Lienert F & Schübeler D (2013) Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153:** 480–92

Bauman ML, Kemper TL & Arin DM (1995) Microscopic observations of the brain in Rett syndrome. *Neuropediatrics* **26(2):** 105-8

Bebbington a, Bebbington A, Anderson A, Anderson A, Ravine D, Ravine D, Fyfe S, Fyfe S, Pineda M, Pineda M, Klerk N De, Klerk N De, Yatawara N, Yatawara N, Percy A, Percy A, Kaufmann WE, Kaufmann WE, Leonard H & Leonard H (2008) Investigating genotype–phenotype relationships in Rett syndrome using an international data set. *Neurology* **43100**

Bebbington a, Percy a, Christodoulou J, Ravine D, Ho G, Jacoby P, Anderson a, Pineda M, Ben Zeev B, Bahi-Buisson N, Smeets E & Leonard H (2010) Updating the profile of C-terminal MECP2 deletions in Rett syndrome. *J. Med. Genet.* **47:** 242–8

Belichenko NP, Belichenko P V, Li HH, Mobley WC & Francke U (2008) Comparative study of brain morphology in Mecp2 mutant mouse models of Rett syndrome. *J. Comp. Neurol.* **508:** 184–95

Belichenko P V & Dahlström a (1995) Studies on the 3-dimensional architecture of dendritic spines and varicosities in human cortex by confocal laser scanning microscopy and Lucifer yellow microinjections. *J. Neurosci. Methods* **57:** 55–61

Belichenko P V, Wright EE, Belichenko NP, Masliah E, Li HH, Mobley WC & Francke U (2009) Widespread changes in dendritic and axonal morphology in Mecp2-mutant mouse models of Rett syndrome: evidence for disruption of neuronal networks. *J. Comp. Neurol.* **514:** 240–58

Ben-Shachar S, Chahrour M, Thaller C, Shaw C a & Zoghbi HY (2009) Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. *Hum. Mol. Genet.* **18:** 2431–42

Bertani I, Rusconi L, Bolognese F, Forlani G, Conca B, De Monte L, Badaracco G, Landsberger N & Kilstrup-Nielsen C (2006) Functional consequences of mutations in CDKL5, an X-linked gene involved in infantile spasms and mental retardation. *J. Biol. Chem.* **281:** 32048–32056

Bestor TH (1988) Cloning of a mammalian DNA methyltransferase. *Gene* **74:** 9-12

Bibel M, Richter J, Schrenk K, Tucker KL, Staiger V, Korte M, Goetz M & Barde Y-A (2004) Differentiation of mouse embryonic stem cells into a defined neuronal lineage. *Nat. Neurosci.* **7:** 1003–9

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

Bird AP, Buildings K, Road WM & Court F (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 1499–1504

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

Boyle a L, Ballard SG & Ward DC (1990) Differential distribution of long and short interspersed element sequences in the mouse genome: chromosome karyotyping by fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **87:** 7757–7761

Brero A, Easwaran HP, Nowak D, Grunewald I, Cremer T, Leonhardt H & Cardoso MC (2005) Methyl CpG-binding proteins induce large-scale chromatin reorganization during terminal differentiation. *J. Cell Biol.* **169:** 733–743

Brunmeir R, Lagger S, Simboeck E, Sawicka A, Egger G, Hagelkruys A, Zhang Y, Matthias P, Miller WJ & Seiser C (2010) Epigenetic regulation of a murine retrotransposon by a dual histone modification mark. *PLoS Genet.* **6**

Carney RM, Wolpert CM, Ravan S a., Shahbazian M, Ashley-Koch A, Cuccaro ML, Vance JM & Pericak-Vance M a. (2003) Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatr. Neurol.* **28:** 205–211

Casanova MF, Naidu S, Goldberg TE, Moser HW, Khoromi S, Kumar A, Kleinman JE & Weinberger DR (1991) Quantitative magnetic resonance imaging in Rett syndrome. *J. Neuropsychiatry Clin. Neurosci.* **3(1):** 66-72

Chahrour M, Jung SY, Shaw C, Zhou X, Wong STC, Qin J & Zoghbi HY (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* **320:** 1224–9

Chang Q, Khare G, Dani V, Nelson S & Jaenisch R (2006) The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. *Neuron* **49:** 341–8

Chao H-T, Chen H, Samaco RC, Xue M, Chahrour M, Yoo J, Neul JL, Gong S, Lu H-C, Heintz N, Ekker M, Rubenstein JLR, Noebels JL, Rosenmund C & Zoghbi HY (2010) Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. *Nature* **468:** 263–9

Chao HT, Zoghbi HY & Rosenmund C (2007) MeCP2 Controls Excitatory Synaptic Strength by Regulating Glutamatergic Synapse Number. *Neuron* **56:** 58–65

Charman T, Neilson TCS, Mash V, Archer H, Gardiner MT, Knudsen GPS, McDonnell A, Perry J, Whatley SD, Bunyan DJ, Ravn K, Mount RH, Hastings RP, Hulten M, Orstavik KH, Reilly S, Cass H, Clarke A, Kerr AM & Bailey MES (2005) Dimensional phenotypic analysis and functional categorisation of mutations reveal novel genotype-phenotype associations in Rett syndrome. *Eur. J. Hum. Genet.* **13:** 1121–1130

Cheadle JP, Gill H, Fleming N, Maynard J, Kerr a, Leonard H, Krawczak M, Cooper DN, Lynch S, Thomas N, Hughes H, Hulten M, Ravine D, Sampson JR & Clarke a (2000) Long-read sequence analysis of the MECP2 gene in Rett syndrome patients: correlation of disease severity with mutation type and location. *Hum. Mol. Genet.* **9:** 1119–1129

Chen L, Chen K, Lavery LA, Baker SA, Shaw CA, Li W & Zoghbi HY (2015) MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome: Fig. 2. *Proc. Natl. Acad. Sci.* **112:** 201507794

Chen RZ, Akbarian S, Tudor M & Jaenisch R (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.* **27:** 327–31

Chen WG, Chang Q, Lin Y, Meissner A, West AE, Griffith EC, Jaenisch R & Greenberg ME (2003) Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science* **302:** 885–889

Cheng T-L, Wang Z, Liao Q, Zhu Y, Zhou W-H, Xu W & Qiu Z (2014) MeCP2 Suppresses Nuclear MicroRNA Processing and Dendritic Growth by Regulating the DGCR8/Drosha Complex. *Dev. Cell* **28:** 547–60

Cheval H, Guy J, Merusi C, De Sousa D, Selfridge J & Bird A (2012) Postnatal Inactivation Reveals Enhanced Requirement for Mecp2 At Distinct Age Windows. *Hum. Mol. Genet.* 1–9

Christodoulou J, Grimm A, Maher T & Bennetts B (2003) RettBASE: The IRSA *MECP2* Variation Database - A New Mutation Database in Evolution. *Hum Mutat* 21:466-472

Cohen D, Lazar G, Couvert P, Desportes V, Lippe D, Mazet P & Heron D (2002) MECP2 Mutation in a Boy With Language Disorder and Schizophrenia. *Am J Psychiatry* **159(1):** 148–9

Cohen S, Gabel HW, Hemberg M, Hutchinson AN, Sadacca LA, Ebert DH, Harmin D a, Greenberg RS, Verdine VK, Zhou Z, Wetsel WC, West AE & Greenberg ME (2011) Genome-Wide Activity-Dependent MeCP2 Phosphorylation Regulates Nervous System Development and Function. *Neuron* **72:** 72–85

Collins AL, Levenson JM, Vilaythong AP, Richman R, Armstrong DL, Noebels JL, David Sweatt J & Zoghbi HY (2004) Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.* **13:** 2679–89

Colvin L, Fyfe S, Leonard S, Schiavello T, Ellaway C, De Klerk N, Christodoulou J, Msall M & Leonard H (2003) Describing the phenotype in Rett syndrome using a population database. *Arch. Dis. Child.* **88:** 38–43

Colvin L, Leonard H, de Klerk N, Davis M, Weaving L, Williamson S & Christodoulou J (2004) Refining the phenotype of common mutations in Rett syndrome. *J. Med. Genet.* **41:** 25–30

Couvert P, Bienvenu T, Aquaviva C, Poirier K, Moraine C, Gendrot C, Verloes a, Andrès C, Le Fevre a C, Souville I, Steffann J, des Portes V, Ropers HH, Yntema HG, Fryns JP, Briault S, Chelly J & Cherif B (2001) MECP2 is highly mutated in X-linked mental retardation. *Hum. Mol. Genet.* **10:** 941–946

Coy JF, Sedlacek Z, Bächner D, Delius H & 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

Crawley JN, Belknap JK, Collins a, Crabbe JC, Frankel W, Henderson N, Hitzemann RJ, Maxson SC, Miner LL, Silva a J, Wehner JM, Wynshaw-Boris a & Paylor R (1997) Behavioral phenotypes of inbred mouse strains: implications and recommendations for molecular studies. *Psychopharmacology (Berl).* **132:** 107–24

Cuddapah VA, Pillai RB, Shekar K V, Lane JB, Motil KJ, Skinner S a, Tarquinio DC, Glaze DG, McGwin G, Kaufmann WE, Percy AK, Neul JL & Olsen ML (2014) Methyl-CpG-binding protein 2 (MECP2) mutation type is associated with disease severity in Rett syndrome. *J. Med. Genet.* **51(3)**:152-8

D'Esposito M, Quaderi N a, Ciccodicola a, Bruni P, Esposito T, D'Urso M & Brown SD (1996) Isolation, physical mapping, and northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2. *Mamm. Genome* **7:** 533–535

Dani VS, Chang Q, Maffei A, Turrigiano GG, Jaenisch R & Nelson SB (2005) Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **102:** 12560–5

Danielson PB (2002) The cytochrome P450 superfamily: biochemistry, evolution and drug metabolism in humans. Curr. Drug Metab. **3**: 561-97

Delgado IJ, Kim DS, Thatcher KN, LaSalle JM & Van den Veyver IB (2006) Expression profiling of clonal lymphocyte cell cultures from Rett syndrome patients. *BMC Med. Genet.* **7:** 61

Derecki NC, Cronk JC, Lu Z, Xu E, Abbott SBG, Guyenet PG & Kipnis J (2012) Wild-type microglia arrest pathology in a mouse model of Rett syndrome. *Nature* 1–7

Down C (2000) Somatic mutation in MECP2 as a non-fatal neurodevelopmental disorder in males For personal use only . Not to be reproduced without permission of The Lancet . *Lancet* **356:** 830–832

Ebert DH, Gabel HW, Robinson ND, Kastan NR, Hu LS, Cohen S, Navarro AJ, Lyst MJ, Ekiert R, Bird AP & Greenberg ME (2013) Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature* **499:** 341–5

Edgar AJ & Polak JM (2002) Cloning and tissue distribution of three murine alpha/beta hydrolase fold protein cDNAs. *Biochem. Biophys. Res. Commun.* **292:** 617–625

Einspieler C, Kerr AM & Prechtl HFR (2005) Is the early development of girls with Rett disorder really normal? *Pediatr. Res.* **57:** 696–700

Fabio RA, Colombo B, Russo S, Cogliati F, Masciadri M, Foglia S, Antonietti A & Tavian D (2014) Recent insights into genotype-phenotype relationships in patients with Rett syndrome using a fine grain scale. *Res. Dev. Disabil.* **35:** 2976–2986

Farra N, Zhang W-B, Pasceri P, Eubanks JH, Salter MW & Ellis J (2012) Rett syndrome induced pluripotent stem cell-derived neurons reveal novel neurophysiological alterations. *Mol. Psychiatry* 1–11

Feng J, Zhou Y, Campbell SL, Le T, Li E, Sweatt JD, Silva AJ & Fan G (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13:** 423–430

De Filippis B, Ricceri L & Laviola G (2010) Early postnatal behavioral changes in the Mecp2-308 truncation mouse model of Rett syndrome. *Genes. Brain. Behav.* **9:** 213–23

Free a, Wakefield RI, Smith BO, Dryden DT, Barlow PN & Bird a P (2001) DNA recognition by the methyl-CpG binding domain of MeCP2. *J. Biol. Chem.* **276:** 3353–60

Freilinger M, Bebbington A, Lanator I, De Klerk N, Dunkler D, Seidl R, Leonard H & Ronen GM (2010) Survival with Rett syndrome: comparing Rett's original sample with data from the Australian Rett Syndrome Database. *Dev. Med. Child Neurol.* **52:** 962–5

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL & Paul CL (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* **89:** 1827–1831

Fyffe SL, Neul JL, Samaco RC, Chao H-T, Ben-Shachar S, Moretti P, McGill BE, Goulding EH, Sullivan E, Tecott LH & Zoghbi HY (2008) Deletion of Mecp2 in Sim1-expressing neurons reveals a critical role for MeCP2 in feeding behavior, aggression, and the response to stress. *Neuron* **59:** 947–58

Gabel HW, Kinde B, Stroud H, Gilbert CS, Harmin D a., Kastan NR, Hemberg M, Ebert DH & Greenberg ME (2015) Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* **522(7554)**:89-93

Gadalla KK, Bailey ME, Spike RC, Ross PD, Woodard KT, Kalburgi SN, Bachaboina L, Deng J V, West AE, Samulski RJ, Gray SJ & Cobb SR (2012) Improved Survival and Reduced Phenotypic Severity Following AAV9/MECP2 Gene Transfer to Neonatal and Juvenile Male Mecp2 Knockout Mice. *Mol. Ther.* 1–13

Galvão TC & Thomas JO (2005) Structure-specific binding of MeCP2 to four-way junction DNA through its methyl CpG-binding domain. *Nucleic Acids Res.* **33:** 6603–9

Garg SK, Lioy DT, Cheval H, McGann JC, Bissonnette JM, Murtha MJ, Foust KD, Kaspar BK, Bird A & Mandel G (2013) Systemic delivery of MeCP2 rescues

behavioral and cellular deficits in female mouse models of Rett syndrome. *J. Neurosci.* **33:** 13612–20

Georgel PT, Horowitz-Scherer R a., Adkins N, Woodcock CL, Wade P a. & Hansen JC (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J. Biol. Chem.* **278:** 32181–32188

Ghosh RP, Horowitz-Scherer R a, Nikitina T, Gierasch LM & Woodcock CL (2008) Rett syndrome-causing mutations in human MeCP2 result in diverse structural changes that impact folding and DNA interactions. *J. Biol. Chem.* **283:** 20523–34

Ghosh RP, Horowitz-Scherer R a, Nikitina T, Shlyakhtenko LS & Woodcock CL (2010a) MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. *Mol. Cell. Biol.* **30:** 4656–4670

Ghosh RP, Nikitina T, Horowitz-Scherer R a., Gierasch LM, Uversky VN, Hite K, Hansen JC & Woodcock CL (2010b) Unique physical properties and interactions of the domains of methylated DNA binding protein 2. *Biochemistry* **49:** 4395–4410

Giacometti E, Luikenhuis S, Beard C & Jaenisch R (2007) Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. *Proc. Natl. Acad. Sci. U. S. A.* **104:** 1931–6

Gibson JH, Williamson SL, Arbuckle S & Christodoulou J (2005) X chromosome inactivation patterns in brain in Rett syndrome: Implications for the disease phenotype. *Brain Dev.* **27:** 266–270

Goffin D, Allen M, Zhang L, Amorim M, Wang I-TJ, Reyes A-RS, Mercado-Berton A, Ong C, Cohen S, Hu L, Blendy J a, Carlson GC, Siegel SJ, Greenberg ME & Zhou Z (2011) Rett syndrome mutation MeCP2 T158A disrupts DNA binding, protein stability and ERP responses. *Nat. Neurosci.* 1–12

Gomez CM, Maselli R, Gundeck JE, Chao M, Day JW, Tamamizu S, Lasalde J a, McNamee M & Wollmann RL (1997) Slow-channel transgenic mice: a model of postsynaptic organellar degeneration at the neuromuscular junction. *J. Neurosci.* **17:** 4170–4179

Gregory RI, Randall TE, Johnson C a, Khosla S, Hatada I, Neill LPO, Bryan M, Feil R & Turner BM (2001) DNA Methylation Is Linked to Deacetylation of Histone H3 , but Not H4 , on the Imprinted Genes Snrpn and U2af1-rs1 DNA Methylation Is Linked to Deacetylation of Histone H3 , but Not H4 , on the Imprinted Genes Snrpn and U2af1-rs1. *Mol. Cell. Biol.* **21:** 5426–5436

Guo JU, Su Y, Shin JH, Shin J, Li H, Xie B, Zhong C, Hu S, Le T, Fan G, Zhu H, Chang Q, Gao Y, Ming G-L & Song H (2014) Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat. Neurosci.* **17:** 215–22

Guy J, Cheval H, Selfridge J & Bird A (2011) The Role of MeCP2 in the Brain. *Annu. Rev. Cell Dev. Biol.* **27:** 631–52

Guy J, Gan J, Selfridge J, Cobb S & Bird A (2007) Reversal of neurological defects in a mouse model of Rett syndrome. *Science* **315:** 1143–7

Guy J, Hendrich B, Holmes M, Martin JE & Bird a (2001) A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27:** 322–6

Hagberg B, Aicardi J, Dias K & Ramos O (1983) A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann. Neurol.* **14:** 471–479

Hagberg B, Goutieres F, Hanefeld F, Rett A & Wilson J (1985) Rett syndrome: criteria for inclusion and exclusion. *Brain Dev.* **7(3):** 372–3

Hagberg B & Skjeldal OH (1994) Rett variants: a suggested model for inclusion criteria. *Pediatr Neurol.* **11:** 5-11

Hagberg B, Hanefeld F, Percy A, Skjeldal O (2002) An update on clinically applicable diagnostic criteria in Rett syndrome. Comments to Rett Syndrome Clinical Criteria Consensus Panel Satellite to European Paediatric Neurology Society Meeting, Baden Baden, Germany. *Eur J Paediatr Neurol*. **6**:293–297

Hagne I, Witt-Engerström I & Hagberg B (1989) EEG development in Rett syndrome. A study of 30 cases. *Electroencephalogr. Clin. Neurophysiol.* **72:** 1–6

Halbach NSJ, Smeets EEJ, van den Braak N, van Roozendaal KEP, Blok RMJ, Schrander-Stumpel CTRM, Frijns JP, Maaskant M a. & Curfs LMG (2012) Genotype-phenotype relationships as prognosticators in Rett syndrome should be handled with care in clinical practice. *Am. J. Med. Genet. Part A* **158 A:** 340–350

Han K, Gennarino VA, Lee Y, Pang K, Hashimoto-Torii K, Choufani S, Raju CS, Oldham MC, Weksberg R, Rakic P, Liu Z & Zoghbi HY (2013) Human-specific regulation of MeCP2 levels in fetal brains by microRNA miR-483-5p. *Genes Dev.* **27:** 485–490

Hanefeld F (1985) The clinical pattern of the Rett syndrome. *Brain Dev*. **7**:320–325

Hansen JC, Ghosh RP & Woodcock CL (2010) Binding of the Rett syndrome protein, MeCP2, to methylated and unmethylated DNA and chromatin. *IUBMB Life* **62:** 732–8

Hansen JC, Wexler BB, Rogers DJ, Hite KC, Panchenko T, Ajith S & Black BE (2011) DNA binding restricts the intrinsic conformational flexibility of methyl CpG binding protein 2 (MeCP2). *J. Biol. Chem.* **286:** 18938–48

Hashimoto H, Liu Y, Upadhyay a. K, Chang Y, Howerton SB, Vertino PM, Zhang X & Cheng X (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40:** 4841–4849

He Y & Ecker JR (2015) Non-CG Methylation in the Human Genome. *Annu. Rev. Genomics Hum. Genet.* **16:** 150615185749007

Heckman LD, Chahrour MH & Zoghbi HY (2014) Rett-causing mutations reveal two domains critical for MeCP2 function and for toxicity in MECP2 duplication syndrome mice. *Elife* e02676

Hendrich B & Bird A (1998) Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins Identification and Characterization of a Family of Mammalian Methyl-CpG Binding Proteins. *Mol. Cell. Biol.* **18:** 6538–6547

Hendrich B & Tweedie S (2003) The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet.* **19:** 269–277
Ho KL, McNae IW, Schmiedeberg L, Klose RJ, Bird AP & Walkinshaw MD (2008) MeCP2 binding to DNA depends upon hydration at methyl-CpG. *Mol. Cell* **29:** 525–31

Hoffbuhr K, Devaney JM, LaFleur B, Sirianni N, Scacheri C, Giron J, Schuette J, Innis J, Marino M, Philippart M, Narayanan V, Umansky R, Kronn D, Hoffman EP & Naidu S (2001) MeCP2 mutations in children with and without the phenotype of Rett syndrome. *Neurology* **56:** 1486–1495

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

Horz W & Altenburger W (1981) Nudeotide sequence of mouse satellite DNA. *Nucleic Acids Res.* **9:** 683–696

Huppke P, Held M, Hanefeld F, Engel W & Laccone F (2002) Influence of mutation type and location on phenotype in 123 patients with Rett syndrome. *Neuropediatrics* **33:** 63-8

Iles N, Rulten S, El-Khamisy SF & Caldecott KW (2007) APLF (C2orf13) is a novel human protein involved in the cellular response to chromosomal DNA strand breaks. *Mol. Cell. Biol.* **27:** 3793–3803

Imessaoudene B, Bonnefont J, Royer G, Cormier-daire V, Lyonnet S, Lyon G, Munnich A & Amiel J (2001) Encephalopathy in a Male. *J. Med. Genet.* 171–174

Ishii T, Makita Y, Ogawa A, Amamiya S, Yamamoto M, Miyamoto A & Oki J (2001) The role of different X-inactivation pattern on the variable clinical phenotype with Rett syndrome. *Brain Dev.* **23:** 161–164

Jellinger K & Seitelberger F (1986) Neuropathology of Rett syndrome. *Am. J. Genet. Suppl.* **1:** 259-88

Johnson KR, Lehn D a & Reeves R (1989) Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins HMG-I and HMG-Y. *Mol. Cell. Biol.* **9:** 2114–2123

Johnson R a, Lam M, Punzo AM, Li H, Lin BR, Ye K, Mitchell GS & Chang Q (2011) 7,8-dihydroxyflavone (7,8-DHF) exhibits therapeutic efficacy in a mouse model of Rett syndrome. *J. Appl. Physiol.* **112**: 704-10

Jones BJ & Roberts DJ (1968) The quantitative measurement of motor incoordination in naïve mice using an accelerating rotarod. *J. Pharm. Pharmacol*. **20**: 302-4.

Jones PL, Veenstra GJ, Wade P a, Vermaak D, Kass SU, Landsberger N, Strouboulis J & Wolffe a P (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19:** 187–91

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

Kankirawatana P, Leonard H, Ellaway C, Scurlock J, Mansour a., Makris CM, Dure IV LS, Friez M, Lane J, Kiraly-Borri C, Fabian V, Davis M, Jackson J, Christodoulou J, Kaufmann WE, Ravine D & Percy a. K (2006) Early progressive encephalopathy in boys and MECP2 mutations. *Neurology* **67:** 164–166

Kernohan KD, Jiang Y, Tremblay DC, Bonvissuto AC, Eubanks JH, Mann MRW & Bérubé NG (2010) ATRX Partners with Cohesin and MeCP2 and Contributes to Developmental Silencing of Imprinted Genes in the Brain. *Dev. Cell* **18:** 191–202

Kerr AM, Nomura Y, Armstrong D, Anvret M, Belichenko P V., Budden S, Cass H, Christodoulou J, Clarke A, Ellaway C, D'Esposito M, Francke U, Hulten M, Julu P, Leonard H, Naidu S, Schanen C, Webb T, Engerstrom IW, Yamashita Y, et al (2001) Guidelines for reporting clinical features in cases with MECP2 mutations. *Brain Dev.* **23:** 208–211

Kerr AM & Engerström IW (2001) The clinical background to the Rett disorder. Rett Disorder and the Developing Brain, pp 1-26 (Oxford University Press).

Kerr B, Alvarez-Saavedra M, Sáez M a, Saona A & Young JI (2008) Defective body-weight regulation, motor control and abnormal social interactions in Mecp2 hypomorphic mice. *Hum. Mol. Genet.* **17:** 1707–17

Kinde B, Gabel HW, Gilbert CS, Griffith EC & Greenberg ME (2015) Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc. Natl. Acad. Sci.* 201411269

King IF, Yandava CN, Mabb AM, Hsiao JS, Huang H-S, Pearson BL, Calabrese JM, Starmer J, Parker JS, Magnuson T, Chamberlain SJ, Philpot BD & Zylka MJ (2013) Topoisomerases facilitate transcription of long genes linked to autism. *Nature* **501:** 58–62

Kirby RS, Lane JB, Childers J, Skinner S a, Annese F, Barrish JO, Glaze DG, Macleod P & Percy AK (2010) Longevity in Rett syndrome: analysis of the North American Database. *J. Pediatr.* **156:** 135–138.e1

Kishi N & Macklis JD (2004) MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol. Cell. Neurosci.* **27:** 306–21

Klauck SM, Lindsay S, Beyer KS, Splitt M, Burn J & Poustka A (2002) A mutation hot spot for nonspecific X-linked mental retardation in the MECP2 gene causes the PPM-X syndrome. *Am. J. Hum. Genet.* **70:** 1034–1037

Klein ME, Lioy DT, Ma L, Impey S, Mandel G & Goodman RH (2007) Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nat. Neurosci.* **10:** 1513–1514

Klose RJ, Sarraf S a, Schmiedeberg L, McDermott SM, Stancheva I & Bird AP (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell* **19:** 667–78

Kriaucionis S & Bird A (2004) The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucleic Acids Res.* **32:** 1818–23

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

Kriaucionis S, Paterson A, Curtis J, Guy J, Macleod N & Bird A (2006) Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. *Mol. Cell. Biol.* **26:** 5033–42

Kucukkal TG & Alexov E (2015) Structural , Dynamical , and Energetical Consequences of Rett Syndrome Mutation R133C in MeCP2. *Comput. Math. Methods Med.* **2015:** ID 746157

Kudo S, Nomura Y, Segawa M, Fujita N, Nakao M, Schanen C & Tamura M (2003) Heterogeneity in residual function of MeCP2 carrying missense mutations in the methyl CpG binding domain. *J. Med. Genet.* **40:** 487–93

Kumar A, Kamboj S, Malone BM, Kudo S, Twiss JL, Czymmek KJ, LaSalle JM & Schanen NC (2008) Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J. Cell Sci.* **121:** 1128–37

Laurvick CL, de Klerk N, Bower C, Christodoulou J, Ravine D, Ellaway C, Williamson S & Leonard H (2006) Rett syndrome in Australia: a review of the epidemiology. *J. Pediatr.* **148:** 347–352

Lawson-Yuen A, Liu D, Han L, Jiang ZI, Tsai GE, Basu AC, Picker J, Feng J & Coyle JT (2007) Ube3a mRNA and protein expression are not decreased in Mecp2R168X mutant mice. *Brain Res.* **1180:** 1–6

Leonard H, Colvin L, Christodoulou J, Schiavello T, Williamson S, Davis M, Ravine D, Fyfe S, de Klerk N, Matsuishi T, Kondo I, Clarke a, Hackwell S & Yamashita Y (2003) Patients with the R133C mutation: is their phenotype different from patients with Rett syndrome with other mutations? *J. Med. Genet.* **40:** e52

Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F & Bird a (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69:** 905–14

Li H, Zhong X, Chau KF, Williams EC & Chang Q (2011) Loss of activity-induced phosphorylation of MeCP2 enhances synaptogenesis, LTP and spatial memory. *Nat. Neurosci.* **14:** 1001–1008

Li Y, Wang H, Muffat J, Cheng AW, Orlando D a, Lovén J, Kwok S-M, Feldman D a, Bateup HS, Gao Q, Hockemeyer D, Mitalipova M, Lewis C a, Vander Heiden MG, Sur M, Young R a & Jaenisch R (2013) Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons. *Cell Stem Cell* **13:** 446–58

Lioy DT, Garg SK, Monaghan CE, Raber J, Foust KD, Kaspar BK, Hirrlinger PG, Kirchhoff F, Bissonnette JM, Ballas N & Mandel G (2011) A role for glia in the progression of Rett's syndrome. *Nature* **475:** 497–500

Lister R & Ecker JR (2009) Finding the fifth base: Genome-wide sequencing of cytosine methylation. *Genome Res.* **19:** 959–966

Lister R, Mukamel E a, Nery JR, Urich M, Puddifoot C a, Johnson ND, Lucero J, Huang Y, Dwork AJ, Schultz MD, Yu M, Tonti-Filippini J, Heyn H, Hu S, Wu JC, Rao A, Esteller M, He C, Haghighi FG, Sejnowski TJ, et al (2013) Global epigenomic reconfiguration during mammalian brain development. *Science* **341:** 1237905

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V,

Millar a H, Thomson J a, Ren B & Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462:** 315–322

Luger K, Mäder a W, Richmond RK, Sargent DF & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389:** 251–260
Luikenhuis S, Giacometti E, Beard CF & Jaenisch R (2004) Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc. Natl. Acad. Sci. U. S. A.* **101:** 6033–6038

Luo BC & Ecker JR (2015) Exceptional epigenetics in the brain. *Science* **348:** 1094–1095

Lyon M (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* **190**:372-3

Lyst MJ & Bird A (2015) Rett syndrome: a complex disorder with simple roots. *Nat. Rev. Genet.* 1–13

Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, Rappsilber J, Greenberg ME & Bird A (2013) Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16:** 898–902

Maezawa I & Jin L-W (2010) Rett syndrome microglia damage dendrites and synapses by the elevated release of glutamate. *J. Neurosci.* **30:** 5346–5356
Maiwald R, Bönte A, Jung H, Bitter P, Storm Z, Laccone F & Herkenrath P (2002) De novo MECP2 mutation in a 46,XX male patient with Rett syndrome. *Neurogenetics* **4:** 107–108

Marchetto MCN, Carromeu C, Acab A, Yu D, Yeo G, Mu Y, Chen G, Gage FH & Muotri AR (2010) A model for neural development and treatment of Rett Syndrome using human induced pluripotent stem cells. *Cell* **143:** 527–539

Mari F, Kilstrup-Nielsen C, Cambi F, Speciale C, Mencarelli MA & Renieri A (2005) Genetics and mechanisms of disease in Rett syndrome. *Drug Discov. Today Dis. Mech.* **2:** 419–425

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

Masuyama T, Matsuo M, Jing JJ, Tabara Y, Kitsuki K, Yamagata H, Kan Y, Miki T, Ishii K & Kondo I (2005) Classic Rett syndrome in a boy with R133C mutation of MECP2. *Brain Dev.* **27:** 439–42

McGraw CM, Samaco RC & Zoghbi HY (2011) Adult neural function requires MeCP2. *Science* **333:** 186

Meehan RR, Lewis JD & Bird a P (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.* **20:** 5085–92

Mellén M, Ayata P, Dewell S, Kriaucionis S & Heintz N (2012) MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell* **151:** 1417–1430

Meloni I, Bruttini M, Longo I, Mari F, Rizzolio F, D'Adamo P, Denvriendt K, Fryns JP, Toniolo D & Renieri a (2000) A mutation in the rett syndrome gene, MECP2, causes X-linked mental retardation and progressive spasticity in males. *Am. J. Hum. Genet.* **67:** 982–985

Miller OJ, Schnedl W, Allen, J & Erlanger BF (1974) 5-Methylcytosine localised in mammalian constitutive heterochromatin. *Nature* **251(5476):** 636-7

Monrós E, Armstrong J, Aibar E, Poo P, Canós I & Pineda M (2001) Rett syndrome in Spain: mutation analysis and clinical correlations. *Brain Dev.* **23 Suppl 1:** S251–S253

Moretti P, Levenson JM, Battaglia F, Atkinson R, Teague R, Antalffy B, Armstrong D, Arancio O, Sweatt JD & Zoghbi HY (2006) Learning and memory and synaptic plasticity are impaired in a mouse model of Rett syndrome. *J. Neurosci.* **26:** 319–27

Moretti P & Zoghbi HY (2006) MeCP2 dysfunction in Rett syndrome and related disorders. *Curr. Opin. Genet. Dev.* **16:** 276–281

Msall ME, DiGaudio K, Rogers BT, LaForest S, Catanzaro NL, Campbell J, Wilczenski F & Duffy LC (1994) The Functional Independence Measure for Children (WeeFIM). Conceptual basis and pilot use in children with developmental disabilities. *Clin. Pediatr. (Phila.)* **33:** 421-30

Muotri AR, Marchetto MCN, Coufal NG, Oefner R, Yeo G, Nakashima K & Gage FH (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468:** 443–6

Nan X, Campoy FJ & Bird A (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88:** 471–81

Nan X, Meehan RR & Bird A (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* **21:** 4886–92

Nan X, Ng HH, Johnson C A, Laherty CD, Turner BM, Eisenman RN & Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393:** 386–9

Nan X, Tate P, Li E & Bird A (1996) DNA methylation specifies chromosomal localization of MeCP2. *Mol. Cell. Biol.* **16(1):** 414-21

Nectoux J, Fichou Y, Rosas-Vargas H, Cagnard N, Bahi-Buisson N, Nusbaum P, Letourneur F, Chelly J & Bienvenu T (2010) Cell cloning-based transcriptome analysis in Rett patients: Relevance to the pathogenesis of Rett syndrome of new human MeCP2 target genes. *J. Cell. Mol. Med.* **14:** 1962–1974

Nelson ED, Bal M, Kavalali ET & Monteggia LM (2011) Selective impact of MeCP2 and associated histone deacetylases on the dynamics of evoked excitatory neurotransmission. *J. Neurophysiol.* **106:** 193–201

Nelson ED, Kavalali ET & Monteggia LM (2006) MeCP2-Dependent Transcriptional Repression Regulates Excitatory Neurotransmission. *Curr. Biol.* **16:** 710–716

Neul JL, Fang P, Barrish RN, Lane J, Caeg E, Smith EO, Zoghbi H, Percy A & Glaze DG (2008) Specific Mutations in Methyl-CpG-Binding Protein 2 Confer Different Severity in Rett Syndrome. *Neurology* **70:** 1313–1321

Neul JL, Kaufmann WE, Glaze DG, Clarke AJ, Leonard H, Bailey MES, Carolyn N, Zappella M & Renieri A (2010) Rett Syndrome: Revised Diagnostic Criteria and Nomenclature. *Ann Neurol* **68:** 944–950

Nguyen MVC, Du F, Felice C a, Shan X, Nigam A, Mandel G, Robinson JK & Ballas N (2012) MeCP2 Is Critical for Maintaining Mature Neuronal Networks and Global Brain Anatomy during Late Stages of Postnatal Brain Development and in the Mature Adult Brain. *J. Neurosci.* **32:** 10021–34

Nguyen MVC, Felice C a, Du F, Covey M V, Robinson JK, Mandel G & Ballas N (2013) Oligodendrocyte lineage cells contribute unique features to Rett syndrome neuropathology. *J. Neurosci.* **33:** 18764–74

Nguyen S, Meletis K, Fu D, Jhaveri S & Jaenisch R (2007) Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Dev. Dyn.* **236:** 1663–1676

Nielsen JB, Henriksen KF, Hansen C, Silahtaroglu a, Schwartz M & Tommerup N (2001) MECP2 mutations in Danish patients with Rett syndrome: high frequency of mutations but no consistent correlations with clinical severity or with the X chromosome inactivation pattern. *Eur. J. Hum. Genet.* **9:** 178–184

Nikitina T, Ghosh RP, Horowitz-Scherer R a, Hansen JC, Grigoryev S a & Woodcock CL (2007a) MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. *J. Biol. Chem.* **282:** 28237–45

Nikitina T, Shi X, Ghosh RP, Horowitz-Scherer R a, Hansen JC & Woodcock CL (2007b) Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Mol. Cell. Biol.* **27:** 864–77

Nuber U a, Kriaucionis S, Roloff TC, Guy J, Selfridge J, Steinhoff C, Schulz R, Lipkowitz B, Ropers HH, Holmes MC & Bird A (2005) Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. *Hum. Mol. Genet.* **14:** 2247–56

Okabe Y, Takahashi T, Mitsumasu C, Kosai K, Tanaka E & Matsuishi T (2012) Alterations of gene expression and glutamate clearance in astrocytes derived from an MeCP2-null mouse model of Rett syndrome. *PLoS One* **7:** e35354

Okano M, Xie S & Li E (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet*. **19:** 219-20

Orrico a, Lam C, Galli L, Dotti MT, Hayek G, Tong SF, Poon PM, Zappella M, Federico a & Sorrentino V (2000) MECP2 mutation in male patients with non-specific X-linked mental retardation. *FEBS Lett.* **481:** 285–288

Pellow S, Chopin P, File SE & Briley M (1985) Validation of open:closed arm entries in an elevated plus-maze as a measure of anxiety in the rat. *J. Neurosci. Methods* **14:** 149–167

Percy AK (2010) Rett syndrome diagnostic criteria: Lessons from the Natural History Study. *Ann Neurol* **68:** 951–955

Peters SU, Hundley RJ, Wilson AK, Warren Z, Vehorn A, Carvalho CMB, Lupski JR & Ramocki MB (2013) The behavioral phenotype in MECP2 duplication syndrome: a comparison with idiopathic autism. *Autism Res.* **6:** 42–50

Pini G, Scusa MF, Congiu L, Benincasa A, Morescalchi P, Bottiglioni I, Di Marco P, Borelli P, Bonuccelli U, Della-Chiesa A, Prina-Mello A & Tropea D (2012) IGF1 as a Potential Treatment for Rett Syndrome: Safety Assessment in Six Rett Patients. *Autism Res. Treat.* **2012:** 679801
Poorey K, Viswanathan R, Carver MN, Karpova TS, Cirimotich SM, McNally JG, Bekiranov S & Auble DT (2013) Measuring chromatin interaction dynamics on the second time scale at single-copy genes. *Science* **342:** 369–72

Pozzo-Miller L, Pati S & Percy AK (2015) Rett Syndrome: Reaching for Clinical Trials. *Neurotherapeutics* **12(3)**: 631-40

Prut L & Belzung C (2003) The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: A review. *Eur. J. Pharmacol.* **463:** 3–33

Ramsahoye BH, Biniszkiewicz D, Lyko F, Clark V, Bird a P & Jaenisch R (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U. S. A.* **97:** 5237–5242

Ritchie ME, Dunning MJ, Smith ML, Shi W & Lynch AG (2011) Beadarray expression analysis using bioconductor. *PLoS Comput. Biol.* **7:** e1002276

Robinson L, Guy J, McKay L, Brockett E, Spike RC, Selfridge J, De Sousa D, Merusi C, Riedel G, Bird A & Cobb SR (2012) Morphological and functional reversal of phenotypes in a mouse model of Rett syndrome. *Brain.* **135**:2699-710

Rolando S (1985) Rett syndrome: report of eight cases. *Brain Dev*. **7**:290–296

Samaco RC, Fryer JD, Ren J, Fyffe S, Chao H-T, Sun Y, Greer JJ, Zoghbi HY & Neul JL (2008) A partial loss of function allele of methyl-CpG-binding protein 2 predicts a human neurodevelopmental syndrome. *Hum. Mol. Genet.* **17:** 1718–27

Samaco RC, Mandel-Brehm C, Chao H-T, Ward CS, Fyffe-Maricich SL, Ren J, Hyland K, Thaller C, Maricich SM, Humphreys P, Greer JJ, Percy A, Glaze DG, Zoghbi HY & Neul JL (2009) Loss of MeCP2 in aminergic neurons causes cell-autonomous defects in neurotransmitter synthesis and specific behavioral abnormalities. *Proc. Natl. Acad. Sci. U. S. A.* **106:** 21966–21971

Samaco RC, McGraw CM, Ward CS, Sun Y, Neul JL & Zoghbi HY (2012) Female Mecp2+/- mice display robust behavioral deficits on two different genetic backgrounds providing a framework for pre-clinical studies. *Hum. Mol. Genet.* 1–14

Samaco RC, Mandel-Brehm C, McGraw CM, Shaw C a, McGill BE & Zoghbi HY (2012) Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. *Nat. Genet.* **44:** 206–211

Schanen C, Houwink EJF, Dorrani N, Lane J, Everett R, Feng A, Cantor RM & Percy A (2004) Phenotypic manifestations of MECP2 mutations in classical and atypical Rett syndrome. *Am. J. Med. Genet. A* **126A:** 129–40

Schanen NC, Dahle EJ, Capozzoli F, Holm V a, Zoghbi HY & Francke U (1997) A new Rett syndrome family consistent with X-linked inheritance expands the X chromosome exclusion map. *Am. J. Hum. Genet.* **61:** 634–641

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

Schüle B, Armstrong DD, Vogel H, Oviedo a & Francke U (2008) Severe congenital encephalopathy caused by MECP2 null mutations in males: central hypoxia and reduced neuronal dendritic structure. *Clin. Genet.* **74:** 116–26

Shahbazian M, Young J, Yuva-Paylor L, Spencer C, Antalffy B, Noebels J, Armstrong D, Paylor R & Zoghbi H (2002a) Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron* **35:** 243–54

Shahbazian MD, Antalffy B, Armstrong DL & Zoghbi HY (2002b) Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Hum. Mol. Genet.* **11:** 115–24

Sharp A, Robinson D & Jacobs P (2000) Age- and tissue-specific variation of X chromosome inactivation ratios in normal women. *Hum. Genet.* **107:** 343-9

Sinsheimer RL. (1955) The action of pancreatic deoxyribonuclease. II. Isomeric dinucleotides. *J. Biol. Chem*. **215**: 579-83

Skene PJ, Illingworth RS, Webb S, Kerr ARW, James KD, Turner DJ, Andrews R & Bird AP (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell* **37:** 457–68

Smeets E, Schollen E, Moog U, Matthijs G, Herbergs J, Smeets H, Curfs L, Schrander-Stumpel C & Fryns JP (2003) Rett syndrome in adolescent and adult females: clinical and molecular genetic findings. *Am. J. Med. Genet. A* **122A:** 227–233

Smeets EEJ, Pelc K & Dan B (2011) Rett Syndrome. *Mol. Syndromol.* **2:** 113–127
Smrt RD, Eaves-Egenes J, Barkho BZ, Santistevan NJ, Zhao C, Aimone JB, Gage FH & Zhao X (2007) Mecp2 deficiency leads to delayed maturation and altered gene expression in hippocampal neurons. *Neurobiol. Dis.* **27:** 77–89

Song C-X, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen C-H, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley L a, Hicks LM, Lahn BT, Jin P & He C (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29:** 68–72

Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PWTC, Bauer C, Münzel M, Wagner M, Müller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Müller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T, et al (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152:** 1146–59

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK & Schübeler D (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480:** 490–5

Stauder JE a, Smeets EEJ, van Mil SGM & Curfs LGM (2006) The development of visual- and auditory processing in Rett syndrome: An ERP study. *Brain Dev.* **28:** 487–494

Steffenburg U, Hagberg G & Hagberg B (2001) Epilepsy in a representative series of Rett syndrome. *Acta Paediatr.* **90:** 34–39

Subramaniam B, Naidu S & Reiss a L (1997) Neuroanatomy in Rett syndrome: cerebral cortex and posterior fossa. *Neurology* **48:** 399–407

Sugino K, Hempel CM, Okaty BW, Arnson H a., Kato S, Dani VS & Nelson SB (2014) Cell-Type-Specific Repression by Methyl-CpG-Binding Protein 2 Is Biased toward Long Genes. *J. Neurosci.* **34:** 12877–12883

Szulwach KE, Li X, Li Y, Song C-X, Wu H, Dai Q, Irier H, Upadhyay AK, Gearing M, Levey AI, Vasanthakumar A, Godley L a, Chang Q, Cheng X, He C & Jin P (2011) 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14:** 1607–16

Tahiliani M, Koh KP, Shen Y, Pastor W a, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L & Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324:** 930–5

Tanaka Y, Kim K-Y, Zhong M, Pan X, Weissman SM & Park I-H (2014) Transcriptional regulation in pluripotent stem cells by methyl CpG-binding protein 2 (MeCP2). *Hum. Mol. Genet.* **23:** 1045–55

Tao J, Van Esch H, Hagedorn-Greiwe M, Hoffmann K, Moser B, Raynaud M, Sperner J, Fryns J-P, Schwinger E, Gécz J, Ropers H-H & Kalscheuer VM (2004) Mutations in the X-linked cyclin-dependent kinase-like 5 (CDKL5/STK9) gene are associated with severe neurodevelopmental retardation. *Am. J. Hum. Genet.* **75:** 1149–1154

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

Tatton-Brown K, Seal S, Ruark E, Harmer J, Ramsay E, Del Vecchio Duarte S, Zachariou A, Hanks S, O'Brien E, Aksglaede L, Baralle D, Dabir T, Gener B, Goudie D, Homfray T, Kumar A, Pilz DT, Selicorni A, Temple IK, Van Maldergem L, et al (2014) Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* **46:** 385–8

The Rett Syndrome Diagnostic Criteria Work Group (1988) Diagnostic criteria for Rett syndrome. *Ann Neurol.* **23(4):** 425-8

Trappe R, Laccone F, Cobilanschi J, Meins M, Huppke P, Hanefeld F & Engel W (2001) MECP2 mutations in sporadic cases of Rett syndrome are almost exclusively of paternal origin. *Am. J. Hum. Genet.* **68:** 1093–101

Traynor J, Agarwal P, Lazzeroni L & 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

Tropea D, Giacometti E, Wilson NR, Beard C, McCurry C, Fu DD, Flannery R, Jaenisch R & Sur M (2009) Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. *Proc. Natl. Acad. Sci. U. S. A.* **106:** 2029–34

Tudor M, Akbarian S, Chen RZ & 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:** 15536–41

Urbanowicz A, Downs J, Girdler S, Ciccone N & Leonard H (2014) Aspects of speech-language abilities are influenced by MECP2 mutation type in girls with Rett syndrome. *Am. J. Med. Genet. A*

Urdinguio RG, Fernandez AF, Lopez-Nieva P, Rossi S, Huertas D, Kulis M, Liu CG, Croce C, Calin G a. & Esteller M (2010) Disrupted microrna expression caused by Mecp2 loss in a mouse model of Rett syndrome. *Epigenetics* **5:** 656–663

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

Valinluck V, Tsai H-H, Rogstad DK, Burdzy A, Bird A & Sowers LC (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:** 4100–8

Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams B a, Stamatoyannopoulos J a, Crawford GE, Absher DM, Wold BJ & Myers RM (2013) Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res.* **23:** 555–67

Wakefield RI, Smith BO, Nan X, Free a, Soteriou a, Uhrin D, Bird a P & Barlow PN (1999) The solution structure of the domain from MeCP2 that binds to methylated DNA. *J. Mol. Biol.* **291:** 1055–65

Wan M, Lee SS, Zhang X, Houwink-Manville I, Song HR, Amir RE, Budden S, Naidu S, Pereira JL, Lo IF, Zoghbi HY, Schanen NC & Francke U (1999) Rett syndrome and beyond: recurrent spontaneous and familial MECP2 mutations at CpG hotspots. *Am. J. Hum. Genet.* **65:** 1520–1529

Wang J, Wegener JE, Huang T-W, Sripathy S, DeJesus-Cortes H, Xu P, Tran S, Knobbe W, Leko V, Britt J, Starwalt R, McDaniel L, Ward C, Parra D, Newcomb B, Lao U, Nourigat C, Flowers DA, Cullen S, Jorstad NL, et al (2015) Wild-type microglia arrest pathology in a mouse model of Rett syndrome. *Nature* **521**

Wang X, Lacza Z, Sun YE & Han W (2014) Leptin resistance and obesity in mice with deletion of methyl-CpG-binding protein 2 (MeCP2) in hypothalamic pro-opiomelanocortin (POMC) neurons. *Diabetologia* **57:** 236–245

Ward CS, Arvide EM, Huang T-W, Yoo J, Noebels JL & Neul JL (2011) MeCP2 is critical within HoxB1-derived tissues of mice for normal lifespan. *J. Neurosci.* **31:** 10359–10370

Watson P, Black G, Ramsden S, Barrow M, Super M, Kerr B & Clayton-Smith J (2001) Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein. *J. Med. Genet.* **38:** 224–228

Weaving LS, Williamson SL, Bennetts B, Davis M, Ellaway CJ, Leonard H, Thong M-K, Delatycki M, Thompson EM, Laing N & Christodoulou J (2003) Effects of MECP2 mutation type, location and X-inactivation in modulating Rett syndrome phenotype. *Am. J. Med. Genet. A* **118A:** 103–114

Wen L, Li X, Yan L, Tan Y, Li R, Zhao Y, Wang Y, Xie J, Zhang Y, Song C, Yu M, Liu X, Zhu P, Li X, Hou Y, Guo H, Wu X, He C, Li R, Tang F, et al (2014) Whole-genome analysis of 5-hydroxymethylcytosine and 5-methylcytosine at base resolution in the human brain. *Genome Biol.* **15:** R49

Wood L & Shepherd GMG (2010) Synaptic circuit abnormalities of motor-frontal layer 2/3 pyramidal neurons in a mutant mouse model of Rett syndrome. *Neurobiol. Dis.* **38:** 281–287

Wu H, Tao J, Chen PJ, Shahab A, Ge W, Hart RP, Ruan X, Ruan Y & Sun YE (2010) Genome-wide analysis reveals methyl-CpG-binding protein 2-dependent regulation of microRNAs in a mouse model of Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **107:** 18161–18166

Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL & Ren B (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* **148:** 816–31

Yamashita Y, Kondo I, Fukuda T, Morishima R, Kusaga a, Iwanaga R & Matsuishi T (2001) Mutation analysis of the methyl-CpG-binding protein 2 gene (MECP2) in Rett patients with preserved speech. *Brain Dev.* **23 Suppl 1:** S157–60

Yasui DH, Gonzales ML, Aflatooni JO, Crary FK, Hu DJ, Gavino BJ, Golub MS, Vincent JB, Carolyn Schanen N, Olson CO, Rastegar M & Lasalle JM (2014) Mice with an isoform-ablating Mecp2 exon 1 mutation recapitulate the neurologic deficits of Rett syndrome. *Hum. Mol. Genet.* **23:** 6695

Yasui DH, Peddada S, Bieda MC, Vallero RO, Hogart A, Nagarajan RP, Thatcher KN, Farnham PJ & Lasalle JM (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

Yazdani M, Deogracias R, Guy J, Poot R a, Bird A & Barde Y-A (2012) Disease Modeling Using Embryonic Stem Cells: MeCP2 Regulates Nuclear Size and RNA Synthesis in Neurons. *Stem Cells*

Yntema HG, Kleefstra T, Oudakker AR, Romein T, de Vries BB a, Nillesen W, Sistermans E a, Brunner HG, Hamel BCJ & van Bokhoven H (2002) Low frequency of MECP2 mutations in mentally retarded males. *Eur. J. Hum. Genet.* **10:** 487–490

Young JI, Hong EP, Castle JC, Crespo-barreto J, Bowman AB, Rose MF, Kang D, Richman R, Johnson JM, Berget S & Zoghbi HY (2005) Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *PNAS* **102:** 17551–17558

Yu M, Hon GC, Szulwach KE, Song C-X, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min J-H, Jin P, Ren B & He C (2012) Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149:** 1368–80

Yusufzai TM & Wolffe a P (2000) Functional consequences of Rett syndrome mutations on human MeCP2. *Nucleic Acids Res.* **28:** 4172–9

Zapella M (1992) The Rett girls with preserved speech. *Brain Dev*. **14**:98–101

Zhang G, Gurtu V & Kain SR (1996) An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem. Biophys. Res. Commun.* **227:** 707–711

Zhang L, He J, Jugloff DGM & Eubanks JH (2008) The MeCP2-null mouse hippocampus displays altered basal inhibitory rhythms and is prone to hyperexcitability. *Hippocampus* **18:** 294–309

Zhang RR, Cui QY, Murai K, Lim YC, Smith ZD, Jin S, Ye P, Rosa L, Lee YK, Wu HP, Liu W, Xu ZM, Yang L, Ding YQ, Tang F, Meissner A, Ding C, Shi Y & Xu GL (2013) Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* **13:** 237–245

Zhao YT, Goffin D, Johnson BS & Zhou Z (2013) Loss of MeCP2 function is associated with distinct gene expression changes in the striatum. *Neurobiol. Dis.* **59:** 257–266

Zhou Z, Hong EJ, Cohen S, Zhao W-N, Ho H-YH, Schmidt L, Chen WG, Lin Y, Savner E, Griffith EC, Hu L, Steen J a J, Weitz CJ & Greenberg ME (2006) Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron* **52:** 255–69

Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT-Y, Kohlbacher O, De Jager PL, Rosen ED, Bennett D a, Bernstein BE, Gnirke A & Meissner A (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500:** 477–81

Ziller MJ, Müller F, Liao J, Zhang Y, Gu H, Bock C, Boyle P, Epstein CB, Bernstein BE, Lengauer T, Gnirke A & Meissner A (2011) Genomic distribution and inter-

sample variation of non-CpG methylation across human cell types. *PLoS Genet.* **7:** e1002389