THE UNIVERSITY
of EDINBURGH

Identifying the key functions of MeCP2 via genetic manipulation in mice

Rebekah Tillotson

Thesis presented for the degree of Doctor of Philosophy

The University of Edinburgh

2017

Declaration

I declare that this thesis was composed by me and the research presented is my own unless otherwise stated. This work has not been submitted for any other degree or personal qualification.

Rebekah Tillotson

Abstract

MeCP2 was identified by its ability to bind DNA in a methylation-specific manner. Yet, how it interprets the DNA methylome remains unclear. Several mechanisms have been proposed, including a role in transcriptional repression. MeCP2 is highly abundant in the brain, and loss-of-function mutations result in a neurological disorder called Rett syndrome (RTT). Strikingly, RTT-causing missense mutations are almost all located in either the methyl-CpG-binding domain (MBD) or a region that has been shown to bind the NCoR/SMRT co-repressor complex (NID). This suggests that the MBD and the NID are the key functional domains in MeCP2, and that the role of MeCP2 is to form a 'bridge' between chromatin and the co-repressor complex to regulate gene expression. To test this 'bridge' hypothesis, I have made an allelic series of knock-in mice with truncated forms of MeCP2 to determine whether the other regions are dispensable for protein function. The three other regions of MeCP2 (the N-terminus before the MBD, the Intervening region between the MBD and the NID, and the C-terminus after the NID) were deleted in a step-wise manner to produce progressively smaller truncated proteins. Knock-in mice which lack just the N-terminus or both the N- and C-termini are phenotypically normal. Therefore, these regions, which together make up 46% of the protein sequence, are dispensable for MeCP2 function *in vivo*. Additional deletion of the Intervening region, retaining only 34% of the original sequence, results in mild RTT-like symptoms in the knock in mice. This is likely to be caused by this protein's decreased stability and reduced ability to bind the NCoR/SMRT complex in the brain. The most severely truncated protein is nevertheless able to reverse the *Mecp2*-null phenotype when reactivated after the onset of symptoms. Together, these findings strongly support the 'bridge' hypothesis.

Lay Summary

Every cell in our bodies contains the same set of instructions, our DNA. Yet, in different cell types, different genes must be expressed at different levels. Chemical marks on the DNA and its packaging proteins help regulate whether genes are turned up or down. These marks are recognised by proteins that interpret their signal. One of the proteins that recognises marked DNA is called MeCP2 and is the subject of this study. It is present in high levels in the brain, and loss of its function causes a neurological disorder called Rett syndrome that affects 1 in 10-15,000 girls. Analysis of which parts of MeCP2 protein are damaged in patients with Rett syndrome highlights two essential regions: the segment that binds to marked DNA and a second segment that binds to a complex of proteins, which turn down gene expression. I therefore proposed that MeCP2 functions as a bridge between marked DNA and this protein complex. To test this theory, I removed all other parts of the protein leaving just these two segments. I have made genetically modified mice where MeCP2 has been replaced with this minimal version. These mice only have very mild symptoms compared to mice that model Rett syndrome, suggesting that this bridge is indeed the most important role of MeCP2. This study furthers our understanding of gene regulation in the brain and how it is affected in Rett syndrome.

Acknowledgements

The biggest thank you goes of course to Adrian, first for advising that I joined a lab that was carrying out research that I thought was exciting and worthwhile, and second for the excellent supervision he gave me once I had. I would also like to thank my thesis committee members, Atlanta Cook and David Finnegan, for their helpful suggestions and advice. Especially Atlanta, who helped me decide how to chop up MeCP2.

It would not be an understatement to say that this project would have not been possible without the help and support I received from Jacky, Jim, Martha and Dina. Jacky for teaching everything I know (and a small proportion of what she knows) about genetic editing in ES cells, and Jim and Martha for turning these knock-in ES cells into mice. Further thanks go to Jim for the huge amount of work he put into managing all of these mouse lines, including setting up the breeding pairs for the reversal experiment without my knowledge so that I would get started with it and face my fear of performing IP injections. Thank you to Dina for doing the huge volume of genotyping that this project created and for guiding me through the mouse scoring system. A big thank you to all other members of the Bird lab for helpful, stimulating and fun discussions and sharing protocols.

I would also like to thank all the animal unit staff for their help, particularly Alan McClure who trained me to perform IP injections in baby-steps. Thank you to David Kelly for help with microscopy, and to Martin Waterfall for help with flow cytometry.

Finally, thank you to my family for all their support and to everyone who has made my last eight years of student life in Edinburgh so enjoyable! I feel privileged to have had the chance to do two degrees here, enabling me to meet so many wonderful people in this remarkable city.

My studentship was funded by the BBSRC EASTBio DTP scheme.

Table of contents

List of figures

List of tables

Abbreviations

A	Adenine
ASO	Antisense oligonucleotide
ATRX	α-thalassemia/mental retardation, X-linked protein
BDNF	Brain-derived neurotropic factor
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
CamKII kinase	Calcium/calmodulin-dependent protein kinase II
Cas9	CRISPR-associated protein 9
CDKL5	Cyclin-dependent kinase-like 5
CGI	CpG island
ChIP	Chromatin immunoprecipitation
ChIP-PCR	Chromatin immunoprecipitation followed by polymerase chain reaction
ChIP-seq	Chromatin immunoprecipitation followed by high throughput sequencing
CMV	cytomegalovirus
CNS	Central nervous system
CpG	Cytosine followed by guanine of the same strand
CREB1	Cyclic-AMP responsive element binding protein 1
CRISPR	Clustered regularly interspaced short palindromic repeats
DAPI	4’,6-diamidino-2-phenylindole
DGCR8	DiGeorge syndrome critical region 8
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DTT	dithiothreitol
DHX9	DExH-box helicase 9
EDTA	Ethylenediaminetetraacetic acid
EEG	Electroencephalography
EGFP	Enhance green fluorescent protein

EMSA	Electrophoretic mobility shift assay
ES cells	Embryonic stem cells
FBP11	Formin-binding protein 11
FBS	Foetal bovine serum
FOXG1	Forkhead box protein G1
FRAP	Fluorescence recovery after photobleaching
G	Guanine
H1	Histone 1
HDAC	Histone deacetylase
HIPK2	Homeodomain-interacting protein kinase 2
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HLCS	Holocarboxylase synthetase
hmC	Hydroxymethylated cytosine
HMGB1	high mobility group protein 1
HP1γ	Heterochromatin protein 1 γ
HYPC	Huntingtin-interacting protein C
IKKα	IκB kinase α
IP	Immunoprecipitation
kDa	Kilodaltons
KPNA3	Karyopherin α3/ importin subunit α3
KPNA4	Karyopherin α4/ importin subunit α4
LEDGF	lens epithelium-derived growth factor
LIF	Leukemia inhibitory factor
MeCP2	Methyl-CpG binding protein 2
Mecp2	Mouse gene encoding methyl-CpG binding protein 2
MECP2	Human gene encoding methyl-CpG binding protein 2
MBD	Methyl-CpG binding domain
MBD2	Methyl-CpG binding domain protein 2
mC	Methylated cytosine
miRNA	Micro RNA
NA	Nucleosomal array
NCoR	Nuclear receptor co-repressor

NID	NCoR/SMRT interaction domain
NLS	Nuclear localisation signal
p300	histone acetyltransferase p300
PBS	Phosphate buffered saline
PCH	Pericentromeric heterochromatin
PCR	Polymerase chain reaction
PIAS1	Protein inhibitor of activated STAT1
PRMT6	Protein arginine N-methyltransferase 6
PRPF3	Pre-mRNA-splicing factor 3
PTM	Post-translational modification
qPCR	Quantitative PCR
RNA	Ribonucleic acid
RNA-seq	High throughput sequencing of RNA
rpm	Revolutions per minute
RTT	Rett Syndrome
scAAV9	Self-complementary adeno-associated virus serotype 9
SDS	Sodium dodecyl sulphate
SIRT1	Sirtuin 1
SMC1/3	Structural maintenance of chromosomes protein 1 and 3
SMRT	Silencing mediator for retinoid or thyroid-hormone receptors
SOX2	SRY-box 2
SUMO	Small ubiquitin-like modifier
T	Thymine
TBL1X	Transducin β-like protein 1X
TBLR1	Transducin β-like 1 related protein
TBS	Tris buffered saline
TET	Ten-eleven translocation dioxygenase
TFIIB	Transcription factor IIB
TRD	Transcriptional repression domain
Tris	Tris(hydroxymethyl)aminomethane
TSA	Trichostatin A
UTR	Untranslated region

XCI	X chromosome inactiviation
YB-1	Y-box binding protein 1
YY1	Ying yang 1

Chapter 1 – Introduction

1.1 Epigenetic modifications: what, where, when and how?

The human genome consists of three billion nucleotides. Each cell in our bodies has two copies - one paternal and one maternal – divided into 23 pairs of chromosomes. This DNA is neatly packaged around proteins called histones, together making up 'chromatin'. The first level of this packaging is the wrapping of 147 base pair (bp) segments of DNA around histone octamers to form nucleosomes. The histone octamers consist of two molecules of each of the core histones: H2A, H2B, H3 and H4. This structure can then be folded further to form higher order structures of varying accessibility. These chromatin states are crucial for the proper regulation of gene transcription, DNA replication and DNA damage repair. In this section, I will give a brief overview of the small molecular marks (known as 'epigenetic modifications') that are added to the DNA itself and its packaging histones, affecting chromatin structure. This will cover (i) what the modifications are and their relevant 'writers', 'readers' and 'erasers'; (ii) where in the genome they are found; (iii) when during development they are added; and (iv) how they are added to the correct genomic sites.

1.1.1 DNA methylation and histone marks

In vertebrates, epigenetic modification of DNA occurs at the fifth carbon in cytosine rings. The most abundant and well-studied cytosine modification is methyl-cytosine (mC). DNA is methylated predominantly at symmetrical CpG dinucleotides by the *de novo* DNA methyltransferases, DNMT3A and DNMT3B. This mark is heritable through DNA replication, via the maintenance DNA methyltransferase, DNMT1, which targets hemi-methylated DNA (reviewed by Reik, 2007; and Seisenberger et al., 2013). DNA methylation can be lost passively via lack of maintenance over repeated cell divisions, or actively removed via its oxidation by the Ten-eleven Translocation (Tet) enzymes (reviewed by Dahl et al., 2011). This pathway creates several forms of modified cytosine: hydroxyl-methyl cytosine (hmC), fC, and CaC.

Demethylation is completed by conversion of fC or CaC to unmodified cytosine by the Base Excision Repair (BER) pathway (reviewed by Klungland and Robertson, 2016). These other forms were originally thought to simply be intermediates in this demethylation process, but the most abundant and stable of these marks, hmC, is now considered to be a separate epigenetic signal. Unlike mC, hmC has no mechanism for inheritance upon cell division and can only be created by oxidation of mC. Recent studies have mapped mC in the asymmetric CpH context (where 'H' is A, C, or T, but not G). Establishment of CpH methylation is catalysed by the *de novo* DNA methyltransferases, but it cannot be maintained through DNA replication by DNMT1. The majority of CpH methylation is found at CpA dinucleotides, due to sequence preferences of the *de novo* DNA methyltransferases (Ramsahoye et al., 2000; Ziller et al., 2011; Shirane et al., 2013; Varley et al., 2013; Guo et al., 2014; Gabel et al., 2015).

DNA methylation is associated with transcriptional silencing (reviewed by Bird, 2002). This can occur directly by preventing binding of transcriptional activators at their sequence-specific binding sites, or indirectly via the recruitment of proteins that bind specifically to methylated sites (reviewed by Klose and Bird, 2006). Methyl-CpG-binding proteins induce a repressive state by recruiting additional chromatin modifying enzymes such as histone deacetylases (discussed in section 1.2.1). Lastly, methylation of CpG dinucleotides prevents binding by CxxC domain-containing proteins, which recognise unmethylated CpG sites and instead promote an active chromatin environment by recruiting enzymes such as H3K4 methyl-transferases (reviewed by Voigt et al., 2013). DNA methylation at CpH sites is also thought to promote silencing, as shown using a reporter gene by Guo et al. (2014). To date, no proteins that specifically bind mCpH have been identified and only of the mCpG-binding proteins (MeCP2) can also bind mCpH sites (Guo et al., 2014). Conversely, hmC is thought to be activating, with a positive correlation between hmC enrichment and gene transcription in the nervous system (Mellén et al., 2012). The mechanism for this is unclear due to the lack of readers (except for proteins involved in demethylation) that recognise hmC in its most abundant hmCpG dinucleotide context (Spruijt et al., 2013; Gabel et al., 2015).

All four of the core histones can acquire post-translational modifications (PTMs) that affect chromatin structure, either directly by altering their charge or indirectly by recruiting modifying or remodelling enzymes that recognise these marks (reviewed by Bannister and Kouzarides, 2011). The X-ray crystal structure of the nucleosome (Luger et al., 1997) showed that the core histones are arranged with their highly basic N-terminal tails protruding from the main structure, enabling them to make extensive interactions with the negatively charged DNA and other proteins. The majority of the PTMs are added to these accessible tails. These can be acetylated, methylated, phosphorylated, ubiquitinated, SUMOylated and poly(ADP-ribosyl)ated.

Acetylation occurs at multiple lysine residues in the N-terminal tails of Histones H3 and H4, including: H3K9, H3K14, H3K18, H4K5, H4K8 and H4K12. Histone acetylation is added by Histone acetyltransferases (HATs, which require acetyl-CoA) and removed by Histone deactylases (HDACs). There are two classes of HATs: Type A and Type B. Type A HATs function in the nucleus, modifying chromatin. These include GNAT, MYST and CBP/p300. In contrast, Type B HATs are found in the cytoplasm, catalysing the acetylation of free histones before their incorporation into chromatin. There are relatively few HDACs, which alone have limited specificity. *In vivo*, they are incorporated into large multi-unit complexes that regulate their activity, e.g. NuRD, NCoR/SMRT, Sin3a and Co-REST (reviewed by Bannister and Kouzarides, 2011; Oberoi et al., 2011).

Acetylation neutralises the positive charge of the lysine residue, weakening the interaction between histones and DNA. Histone acetylation is an active mark that is highly enriched in more open chromatin structures. This mark is ‘read’ by bromodomain-containing proteins, e.g. Swi/Snf2 (part of the SWI/SNF remodelling complex) and PHD finger-containing proteins, e.g. DPF3b (part of the BAF remodelling complex). Both of these complexes open up chromatin structure (reviewed by Bannister and Kouzarides, 2011). Lastly, acetylation of some residues (e.g. H3K9) prevents their methylation, which is associated with a repressive state (discussed below).

Methylation can occur at both lysine and methionine residues: lysines can be mono-, di- and tri-methylated; and methionines can be mono-, symmetrically di- and asymmetrically di-methylated. Histone methyltransferases contain a catalytic SET domain and require S-adensylmethionine (SAM) to modify lysines and methionines. Histone methylation and demethylation involves specific enzymes for each modification. Histone methylation does not alter the charge of these residues but introduces bulky chemical groups, which affect the binding of 'readers'. Histone methylation can promote either an active or repressive state, depending on the residue that is modified (reviewed by Bannister and Kouzarides, 2011).

Histone H3K4 methylation promotes an open chromatin structure. It is added by SET1A/B complexes and MLL1/2 histone methyltransferases (members of the Trithorax group). This mark is read by a variety of proteins including the ING protein family (via their PHD finger domain), which recruit HATs and HDACs; CHD1 (via its chromodomain), which remodels chromatin structure; and by JMJD2A (via its Tudor domain), which removes methylation from H3K9 and H3K36). Mono- and di- methylation of H3K4 is removed by LSD1 when it is incorporated into the Co-REST repressor complex (reviewed by Bannister and Kouzarides, 2011, and Voigt et al., 2013).

Methylation of H3K9 and H3K27 both promote a repressive chromatin state. H3K9 methylation is added by SUV39 and G9a/GLP enzymes. This mark is read by HP1, which dimerises and recruits both the linker histone H1 and SUV39. Together, these proteins established a condensed 'heterochromatic' state and provide a mechanism of inheritance through cell division (reviewed by Bannister and Kouzarides, 2011). H3K27 methylation is catalysed by EZH1 or EZH2, mutually exclusive components of the Polycomb Repressive Complex 2 (PRC2). This complex can methylate H3K27 to a tri-methyl state. H3K27me3 is bound by PRC2 and Polycomb Repressive Complex 1 (PRC1). Together, these complexes establish repressive polycomb domains that and provide a mechanism for inheritance upon cell division (reviewed by Blackledge et al., 2015).

Ubiquitination is the addition of ubiquitin, a 76 amino acid peptide. Single or multiple copies of ubiquitin can be added to proteins by ubiquitin ligases. Histone H2A is mono-ubiquitinated at K119 by RING1A or RING1B, mutually exclusive enzymes in PRC1. This mark is read by PRC2 (containing AEBP2 and JARID2), and is therefore important in the establishment and maintenance of polycomb domains (reviewed by Blackledge et al., 2015). All of the four core histones can also be SUMOylated. This mark antagonises ubiquitination/acetylation of the same residues (reviewed by Bannister and Kouzarides, 2011).

Phosphorylation of histone tails at serine, threonine and tyrosine residues introduces a negative charge. Yet, unlike histone acetylation, this does not confer a more open chromatin structure. Instead, it has specific roles, such as in the DNA damage response (phosphorylation of Ser139 of variant Histone H2AX) and establishment of mitotic chromatin structure (phosphorylation of H3S10) (Kuo and Yang, 2008; reviewed by Bannister and Kouzarides, 2011).

1.1.2 The genomic location of epigenetic modifications

The distribution of both unmethylated and methylated CpG dinucleotides in vertebrate genomes is highly non-random. CpG dinucleotides are grossly depleted in the bulk genome but are enriched in so-called ‘CpG Islands’ (CGIs). These are often founds at gene promoters and tend to be unmethylated. In contrast, CpG dinucleotides in the bulk genome are highly methylated (Bird et al, 1985; reviewed by Bird 2002; and Deaton and Bird, 2011). The non-random distribution of CpG dinucleotides can be explained by the higher efficiency of C→T mutation at methylated sites, leading to their depletion in the bulk genome (Bird, 1980). The frequency of CpG dinucleotides in CGIs is closer to the expected frequency (observed/expected ratio >0.65) based on base composition (%G+C). This is consistent with the absence of methylation in these regions over an evolutionary timescale. A very small proportion (3-4%) of promoter-proximal CGIs are methylated in differentiated cells, causing silencing of the associated genes

(reviewed by Illingworth and Bird, 2009). The majority of facultatively inactive genes are instead repressed by polycomb marks at their promoters (reviewed by Reik, 2007; and Barth and Imhof, 2010). The location of DNA methylation within gene bodies is also important for reducing transcription, as shown by a reverse correlation between methylation and expression levels (Mellén et al., 2012).

The location of CpH (which is predominantly CpA) methylation is highly correlated with CpG methylation (Ziller et al., 2011; Guo et al., 2014). This is unsurprising given that DNMT3A and DMNT3B are responsible for both CpG and CpH methylation. Strikingly, both CpG and CpH methylation are enriched over gene bodies, consistent with a role of both in regulating transcriptional elongation (Shirane et al., 2013; Guo et al., 2014; Kinde et al., 2016). A major difference between CpG and CpH methylation is that high levels of CpH methylation are only found in a few cell types: ES cells, oocytes, and mature neurons (Ramsahoye et al., 2000; Shirane et al., 2013; Varley et al., 2013; Lister et al., 2013). Another difference is the reduced percentage of methylation at each CpH sites within the cell population or tissue. However, the much higher abundance of CpH dinucleotides compared to CpG dinucleotides in the genome (due to absence of depletion) means that the total number of mCpH sites in the brain is similar to mCpG sites (Lister et al., 2013; Lagger, Connelly, Schweikert, et al., 2017).

Hydroxy-methyl cytosine (hmC) is also highly abundance in mature neurons (~40% of modified cytosines). It is found in gene bodies and repeat regions but depleted from promoters (Szulwach et al., 2011). Unlike CpH, its enrichment in gene bodies anti-correlates with CpG methylation (Mellén et al., 2012). This may be because its creation by oxidation of mCpG leads to the depletion of this mark. This observation questions the role of hmC as an activating mark, as the higher level of expression of hmC-enriched genes may instead be due to lower levels of mCpG in their gene bodies.

The location of various histone modifications has been best characterised in active genes and enhancers. Promoters are enriched in H3K4me2/3 and H3K9ac and gene

bodies are enriched in H3K36me1/2/3, H3K79me1/2/3, H2BK5me1, H3K9me1, H3K27me1 and H4K20me1 (Barth and Imhof, 2010). These marks are important for transcriptional initiation and elongation for RNA polymerase II. Active enhancers are enriched with H3K4me1 and H3K27ac - the mechanisms by which these marks and others regulate transcription is the subject of current research (reviewed by Pradeepa, 2017).

The discovery of CGI promoters marked with both H3K4me3 and H3K27me3 was a surprise given their opposing roles in transcriptional activation and repression, respectively (reviewed by Voigt et al., 2013). Further analysis found that a single Histone H3 tail cannot be modified with both these marks but sister H3 molecules within a single nucleosome can be asymmetrically modified (Voigt et al., 2012). 'Bivalent' domains containing both H3K4me3 and H3K27me3 are found at the CGI promoters of genes encoding developmental transcription factors in cultured ES cells. Analysis of pluripotent epiblast cells from post-implantation mouse embryos found that these domains also occur *in vivo*, but are fewer in number due to lower levels of H3K27me3 (Rugg-Gunn et al., 2010; reviewed by Voigt et al., 2013). Genes with bivalently marked promoters are expressed at low levels: undergoing transcriptional activation but inefficient elongation (reviewed by Voigt et al., 2013).

Lastly, targeting of DNA methylation and histone modifications to specific genomic regions is important for their roles in associated processes. DNA methylation is required for sex-specific imprinting (at Differentially Methylated Regions, DMRs) and silencing of transposons (reviewed by Reik, 2007). The stable silencing of one X chromosome (X chromosome inactivation, XCI) in humans and mice is controlled by multiple epigenetic marks: DNA methylation, H3K27me3, H2AK119ub, H3K9me2 and H4K20me1. The inactive X chromosome in human cells also has domains of H3K9me3and its reader, HP1, and H3K20me3. Strangely, the DNA hypermethylation on the inactive X is restricted to the 5' end of genes, with lower overall DNA methylation than the active X - including in gene bodies (Escamilla-Del-Arenal et al., 2011).

1.1.3 The establishment of epigenetic modifications during mammalian development

An important feature of epigenetic modifications is that they are stably inherited though cell division, providing a mechanism for gene expression signatures to be remembered and transposons to remain silenced. There are times in development, however, when epigenetic marks must be erased and rewritten *de novo*: gametogenesis and embryogenesis. Primordial germ cells (PGCs) in developing embryos undergo extensive reprogramming at embryonic day E10.5-12.5. DNA methylation, including at imprinted Differentially Methylated Regions (DMRs), is erased at this point and then gamete-specific DNA methylation patterns are established by DNMT3A and its cofactor, DNMT3L. This includes sex-specific imprints. The second reprogramming stage occurs after fertilisation, when gamete-specific DNA methylation is erased: first actively from the paternal genome and then passively from both the paternal and maternal genomes. This time, DNA methylation is retained at imprinted regions and at transposons (reviewed by Reik, 2007). As pluripotent cells differentiate in embryogenesis, they acquire cell type-specific epigenetic marks controlling transcription. This process includes the resolution of the vast majority of bivalently marked genes but shifting the equilibrium to either the active (H3K4me3-marked) or repressed (H3K27me3-marked) state (reviewed by Voigt et al., 2013). The later establishment of high levels CpH methylation and hmC makes them stand out from other epigenetic modifications. Their lack of maintenance mechanism means that they accumulate in non-replicating cells, e.g. post-mitotic neurons, with levels increasing during neuronal maturation after birth (Lister et al., 2013; Szulwach et al., 2011).

After fertilisation of female zygotes, dosage compensation by X chromosome inactivation (XCI) must occur. This happens in early embryogenesis, and its then maintains through subsequent cell divisions. In humans, random X chromosome inactivation occurs in both embryonic and extra-embryonic tissues. The process in mice is slightly more complicated: first imprinted inactivation of the paternal X chromosome (Xp) occurs at the 4-8 cell stage, Xp reactivation then occurs in the inner cell mass of the blastocyst, followed by random X chromosome inactivation in

embryonic cells around implantation (embryonic day 5). In both humans and mice, active marks on the inactive X chromosome are lost and repressive marks are enriched, resulting in a highly compact chromatin structure (Escamilla-Del-Arenal et al., 2011).

1.1.4 How epigenetic states are established: developmental and environmental signals, underlying DNA sequence

The establishment of epigenetic states has been a subject of extensive study. Of particular interest is how different epigenetic signatures are established in different cell types in multicellular organisms. While epigenetic marks have been suggested to drive transcriptional profiles, it is more likely that they serve to maintain transcriptional states that have been dictated by transcription factors. To support this, blocking of transcription by small molecules led to the establishment of polycomb marks at previously active genes (Riising et al., 2014). Possible mechanisms for remembering transcriptional states involve nascent transcription or long non-coding RNAs to recruit epigenetic modifiers. Nascent transcription have been shown to lead to the acquisition of active marks at genes in yeast, with initiating RNA polymerase II (phosphorylated at Ser5 in its C-terminal domain) binding the methyltransferase that lays down H3K4 methylation, and elongating RNA polymerase II (phosphorylated at Ser2 in its C-terminal domain) binding the methyltransferase that lays down H3K36 methylation (reviewed by Bannister and Kouzarides, 2011). Recruitment of epigenetic modifiers to genomic loci by non-coding RNA has been a favoured hypothesis for some time. However, to date, only a few examples exist (e.g. *HOTAIR* and at the imprinted region, *kcnq1ot1*), suggesting this mechanism is not global (reviewed by Blackledge et al., 2015).

A lot of research (and media attention) has been devoted to the idea that the environment can alter your epigenome. This is most convincing illustrated by a study that determined the DNA methylation and histone acetylation patterns in pairs of identical twins of different ages. Strikingly, they found that older sets of twins had

more divergent epigenetic signatures and consequential gene expression profiles than younger sets of twins (Fraga et al., 2015).

Recently, Quante and Bird (2016) proposed that epigenetic modifications are influenced in part by the underlying DNA sequence – meaning they are hard-wired in the genome. The discovery of sequence-specific polycomb response elements (PREs) that recruit polycomb complexes to chromatin in *Drosophila* promoted the search for equivalent sequences in mammals (reviewed by Ringrose and Paro, 2007). Mendenhall et al. (2010) and Wachter et al. (2014) both integrated exogenous DNA sequences into mouse ES to determine which chromatin marks they acquired. Mendenhall et al. (2010) found that a CpG island (CGI) without its associated gene was sufficient to acquire bivalent chromatin marks in these cells. Wachter et al., (2014) used artificial sequences to alter the CpG density and GC-richness independently. They found that both features of CGIs (high CpG density and high GC-content) were necessary for these sequences to obtain bivalent marks. Reducing the GC-content led to DNA methylation in this sequence. This was found to be dominant, with its removal (by deletion of DNMT3A and DNMT3B) allowing the acquisition of bivalent marks. This result shed light on the hierarchy of crosstalk mechanisms. CGIs stand out from the rest of the genome by containing a high density of unmethylated CpG dinucleotides. These are 'read' by CxxC domain-containing proteins. The H3K4 methyltransferase complexes, SET1A/B and MLL1/2, all contain subunits with CxxC domains, consistent with the presence of this mark in the majority of unmethylated CGIs (reviewed by Voigt et al., 2013). The recruitment of polycomb to these domains was more of a mystery. Farcas et al. (2012) showed that another CxxC domain-containing protein, KDM2B, could be the missing link, as it recruits PRC1 to chromatin. Quante and Bird (2016) build on these observations to hypothesise that DNA sequence composition is a major determinant of how epigenetic marks are deposited, proposing a mechanism involves a number of proteins that interact with short (2-4 bp) sequences e.g. AT runs. Base composition can be used to divide mammalian genomes into large domains, with higher gene density associated with GC-richness. The observation by Quante and Bird that the frequency of short AT-runs is affected non-linearly by base composition provides a

mechanism by which AT-binding proteins can differentiate between these domains. The expression levels of these proteins vary between different cell types, so could contribute to their distinct epigenetic signatures.

1.2 Readers the DNA methylome: MeCP2 and other members of the MBD protein family

1.2.1 The MBD protein family

This study focuses of Methyl-CpG Binding Protein 2 (MeCP2), the founding member of the MBD protein family. It was discovered in Adrian Bird's lab by its ability to bind to DNA in a methyl-specific manner (Lewis et al., 1992). The domain responsible was mapped to residues 78-162, and termed the methyl-CpG binding domain (MBD) (Nan et al., 1993). The MBD sequence was use to screen expressed sequence tags (ESTs) to identify other methyl-CpG-binding proteins. To date, eleven proteins have been discovered that contain this domain: MeCP2, MBD1-6, BAZ2A/TIP5, BAZ2B, CLLD8 and SETDB1 (Cross et al., 1997; Hendrich and Bird, 1998; Roloff et al., 2003; reviewed by Du et al., 2015). However, only three of these (MBD1, MBD2 and MBD4) have been shown to bind DNA in a methyl-specific manner (Hendrich and Bird, 1998; reviewed by Du et al., 2015). The high sequence conservation of this domain and identical location of an intron within it (Fig. S1) indicate that these proteins arose from gene duplication during evolution. Strikingly, the sequences of these proteins outside the MBD are highly divergent, and contain different functional domains (Fig. 1.2.1). Despite this, all four proteins have been implicated in transcriptional repression, by recruiting chromatin modifiers such as HDACs and H3K9 methyltransferases to methylated DNA (Table 1.2.1; reviewed by Du et al., 2015). This role is consistent with DNA methylation being a repressive mark. Extensive study of these proteins has identified additional functions of these proteins. The functionality of MeCP2 is the most controversial, with proposed roles in transcriptional activation, chromatin compaction, alternative splicing and miRNA processing (reviewed by Lyst and Bird, 2015; see section 1.5). As a result, the general view is that MeCP2 is a 'multifunctional protein'. MBD1 has been reported

to recruit several different effector proteins to chromatin, but all of these promote a repressive state, either by histone deacetylation (by HDAC3) or by H3K9 methylation (by SUV39 and SETDB1). MBD1 has been shown to be involved in heterochromatin formation and its inheritance through DNA replication (Villa et al., 2006; reviewed by Du et al., 2015). MBD2 is well-characterised as a transcriptional repressor, recruiting the HDAC1/2-containing NuRD complex to chromatin, resulting in histone deacetylation and ATP-dependent chromatin remodelling (reviewed by Wood and Zhou, 2016). Additionally, it can recruit PRMT1 or PRMT5 arginine methyltransferases, which modify H4R3 with asymmetrical di-methylation (an active mark) and symmetrical di-methylation (a repressive mark), respectively (Le Guezennec et al., 2006; reviewed by Du et al., 2015). MBD4 is unique among the MBD protein family in having a glycosylase domain. This is required for its well-characterised role in base excision repair (BER) repair at CpG dinucleotides; with the MBD domain of MBD4 binding to 5mCpG/TpG mismatches as well as symmetrically modified mCpG/mCpG dinucleotides (Hendrich et al., 1999; reviewed by Bellacosa and Drohat, 2015).

Another key difference between these proteins is their spatial and temporal expression patterns, which are closely linked to the phenotypic effects of their mutation and overexpression. MeCP2 is expressed in all somatic tissues, with extremely high levels in post-mitotic neurons (Skene et al., 2010; Ross, Guy et al., 2016). It is believed to be required for the proper function of mature neurons, with loss of function mutations and overexpression both resulting in severe neurological symptoms in human patients (see section 1.3) and mouse models (see section 1.4). As the *MECP2* gene (*Mecp2* in mice) is X-linked and subjected to X chromosome inactivation (Quaderi et al., 1994; Alder et al., 1995; Vilain et al., 1996; D'Esposito et al., 1996), these phenotypes are sex-specific. Loss of function mutations cause a severe neurological disorder called Rett syndrome in heterozygous female patients (Amir et al., 1999) and the same mutations result in neonatal encephalopathy and early death in hemizygous males (Kankirawatana et al., 2006). Duplication of the *MECP2* gene also results in neuronal dysfunction, with more severe symptoms in male patients as the duplication usually occurs on the X chromosome (van Esch et

al., 2005). These "*MECP2*-related disorders" have all been modelled in mice, which recapitulate the neurological symptoms, dramatically furthering our understanding about the underlying biology of these diseases (Guy et al., 2001; Chen et al., 2001; Collins et al., 2004).

MBD1 is also most highly expressed in the brain, but decreases rather than increases with age (Wood et al., 2016). Like MeCP2, MBD1 is more highly expressed in neurons than glia (Zhao et al., 2003). This expression pattern fits with a role of MBD1 in neurogenesis, which is decreased in *MBD1-/-* mice (Zhao et al., 2003). Loss of MBD1 has also been reported to cause autism-like features such as reduced social interaction in these mice (Allan et al., 2008), and MBD1 mutations have been identified in individuals with autism (Li et al., 2005). Although mutations in MeCP2 and MBD1 both result neurological symptoms in humans and mice, it is important to note that *MECP2*-related disorders are much more severe. MBD2 and MBD4 both show a broader expression pattern in somatic tissues (Wood et al., 2016; Hendrich and Bird, 1998) and MBD4 is the only member of the family to be expressed in ES cells (Hendrich and Bird, 1998). Like *MBD1-/-* mice, both *MBD2-/-* and *MBD4-/-* mice have mild phenotypes (Table 1.2.1). The only adverse phenotypes reported in MBD2-/- mice are hypoactivity (Wood et al., 2016) and defects in maternal behaviour (Hendrich et al., 2001). Loss of MBD4 results in a three-fold increase of C→T transitions at CpG sites, consistent with its role in base excision repair. While this increased mutation rate is insufficient to cause tumorigenesis alone, this occurred more frequently in *MBD4-/-* mice when one copy of the *Apc* gene was mutated (Millar et al., 2002; Wong et al., 2002). Like numerous other DNA repair proteins, MBD4 is frequently mutated in cancer – mostly in tumours with microsatellite instability (reviewed by Du et al., 2015). The other three MBD protein members have also been found to be mutated and/or overexpressed in a variety of types of cancer (reviewed by Du et al., 2015). Surprisingly, deletion of MBD2 prevents the development of intestinal tumours in *Apc* mutant mouse models (Sansom et al., 2003). Epigenetic modifiers are another group of proteins frequently mutated in cancer, as altered epigenetic programming can cause gene expression changes that promote tumour formation.

Despite their diverse domain structure and biological roles, the shared binding sites of the MBD protein family and their repressive effect of gene transcription beckons the question of whether these proteins can compensate for each other. To date, there is little evidence that this is the case. Deletion of these proteins does not result in upregulation of the others (Martín Caballero et al., 2009), nor does double mutation result in compound phenotypes. In fact, double knock-out mice lacking both *Mecp2* and *MBD2* resembled *Mecp2*-nulls in terms of disease severity (Guy et al., 2001; Martín Caballero et al., 2009). Double knock-out mice do, however, have slightly but significantly reduced survival compared to mice only lacking *Mecp2* after backcrossing onto a C57BL/6 inbred background (Martín Caballero et al., 2009). The greater expression pattern overlap between MeCP2 and MBD1 proteins would predict that they could compensate for each other more than the others. While a reduced survival effect upon deletion of *MBD1* in an *Mecp2*-null background was hinted at by Allan et al. (2008), this has never been followed up.

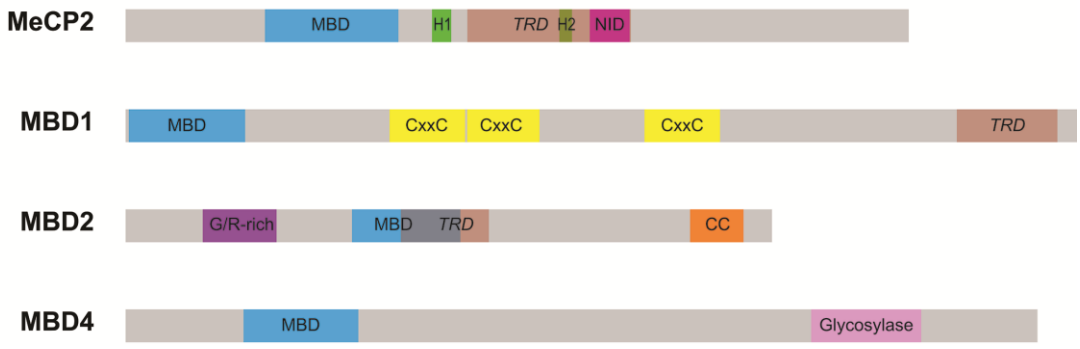


Figure 1.2.1: The MBD protein family
Schematic diagram of the members of the MBD protein family that can bind DNA in a methyl-specific manner: MeCP2 (Accession NP_001104262), MBD1 (Q9UIS9), MBD2 (Q9UBB5.1) and MBD4 (O95243). The major human isoform of each is shown. Domains: MBD (Methyl-CpG Binding Domain); TRD (Transcriptional Repression Domain); NID (NCoR/SMRT Interaction Domain); H1/2 (AT-hook); CxxC (Unmethylated-CpG-binding zinc finger); G/R-rich (Glycine/arginine-rich domain); CC (Coiled-coil domain); Glycosylase (DNA glycosylase). Note: the TRD overlaps other domains in MeCP2 and MBD2. *Modified from Du et al., 2015.*

	MeCP2	MBD1	MBD2	MBD4
Proposed function(s)	Multiple functions, including transcriptional reprepression[1] *(See section 1.5)*	Transcriptional repression and hetero-chromatin formation/ inheritance[12].	Transcriptional repression (and activation)[12].	Base excision repair[20] and transcriptional repression[12].
Spatial/ temporal expression	Expressed in all tissues, very highly in mature neurons[2,3] *(See section 1.2.3)*	High in brains of juvenile animals (decreased in adults) and low in all other tissues[13]. Expressed in neurons and immature cells, not glia[14].	All somatic tissues in adult animals (low in spleen and small intestine in juvenile animals)[13].	Lowly expressed in somatic tissues and ES cells[21].
Associated human diseases	Rett syndrome[4], *MECP2* Duplication syndrome[5], milder intellectual disability/ autism[6], neonatal encephalopathy[7] *(See section 1.3)*; overexpressed in gastric cancer[8] *(See section 1.5.6)*	Mutated in autism[15]; mutated or overexpressed in cancers[12].	Mutated or overexpressed in cancers[12].	Mutated in autism[22]; mutated or overexpressed in cancers[12].
Phenotype in mice	Deletion[9,10] and overexpression[11] both cause severe neurological symptoms. Hemizygous (*-/y*) males have reduced lifespan[9,10]. *(See section 1.4)*	*MBD1 -/-* mice are healthy, fertile, with normal lifespan but have defects in adult neurogenesis and impaired spatial learning[16] and autism-like behaviours[17].	*MBD2 -/-* mice are healthy and fertile but have mild defects in maternal behaviour[18] and are hypoactive[13]. *MBD2 -/- Apc*-mutant mice are resistant to intestinal cancer formation[19].	*MBD4 -/-* mice are healthy, fertile, with normal lifespan. Combination with *Apc* mutation causes increased tumour formation[23,24.]

Table 1.2.1: MBD protein family References: 1. reviewed by Lyst and Bird, 2015; 2. Skene et al., 2010; 3. Ross, Guy et al., 2016; 4. Amir et al., 1999; 5. Van Esch et al., 2005; 6. Neul, 2012; 7. Zeev et al., 2002; 8. Wada et al., 2010; 9. Guy et al., 2001; 10. Chen et al., 2001; 11. Collins et al., 2004; 12. reviewed by Du et al., 2015; 13. Wood et al., 2016; 14. Zhao et al., 2003; 15. Li et al., 2005; 16. Zhao et al., 2003; 17. Allan et al., 2008; 18. Hendrich et al., 2001; 19. Sansom et al., 2003; 20. reviewed by Bellacosa and Drohat, 2015; 21. Hendrich and Bird, 1998; 22. Cukier et al., 2010; 23. Millar et al., 2002; 24. Wong et al., 2002.

1.2.2 The MECP2 gene encodes two isoforms: e1 and e2

The *MECP2* gene consists of four exons, from which mRNA is transcribed and spliced to produce two isoforms (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004). The more abundant isoform, e1, is translated from mRNA comprising of

exons 1, 3 and 4, from a start codon in exon 1. A splice variant containing all four exons encodes the shorter and less abundant isoform, e2, which is translated from a downstream start codon in exon 2 (Fig. 1.2.2). Isoform e1 is considered to be the ancestral form, with higher conservation across evolution including vertebrate species that have only one isoform (Fig. S2). The additional isoform, e2, is only present in mammalian species. The majority of the open reading frame for both isoforms is encoded in exons 3 and 4, so they only differ at the extreme N-terminus. These different N-terminal sequences are both very short: 21 amino acids in the human e1 isoform (26 amino acids in mice); and 9 amino acids in the e2 isoform. By convention, residues are numbered according to the e2 isoform since it was discovered first. This has the advantage that the numbering is the same for the human and mouse proteins for the majority of the sequence: until a two amino acid insertion (residues 385-386) in human MeCP2. Human and mouse e2 isoforms are 486 and 484 amino acids in long, respectively. The longer N-terminal sequences in the e1 isoforms means that they are 498 and 501 amino acids long in humans and mice, respectively.

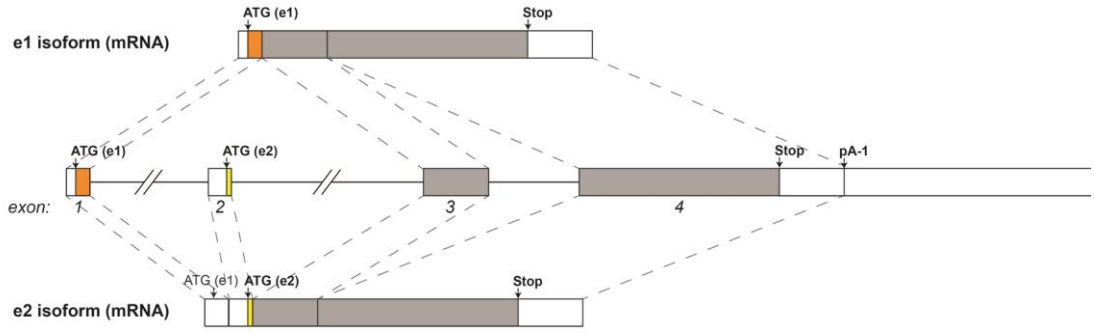


Figure 1.2.2: Alternative splicing of the mammalian *MECP2* gene to produce two isoforms: e1 and e2
Schematic diagram of the *MECP2* gene structure, which consists of four exons. Above: Isoform e1 mRNA formed by splicing together exons 1,3 and 4. The protein is translated from exon 1. Below: Isoform e2 mRNA formed by splicing together all four exons. The protein is translated from exon 2. (The upstream start codon causes translational interference.) Only the first of four alternative polyadenylation sites was drawn for simplicity. This produces transcripts 1.8 kb in length. The open reading frame is shown is grey except for the different N-termini, which are shown in orange (e1) and yellow (e2). The UTRs are shown in white.

1.2.3 Spatial and temporal MeCP2 expression in mammalian tissues

In adult mice, MeCP2 is most highly expressed in the brain, around one order of magnitude higher per cell than all in other tissues tested: heart, lung, liver, spleen, kidney and skeletal muscle (Meehan et al., 1992; Ross, Guy et al., 2016). This high expression is observed in the forebrain, mid/hindbrain and spinal cord, but is considerably lower in the cerebellum (Ross, Guy et al., 2016). Immunohistochemistry using mouse and human brain slices shows that MeCP2 is highly expressed in neurons but much lower in glia (Shahbazian et al., 2002a; Balmer et al., 2003; Kishi and Macklis, 2004). Quantitative western blotting was used to estimate the number of MeCP2 in neuronal nuclei (positive for the marker NeuN) and non-neuronal brain nuclei (NeuN negative – mostly glia). MeCP2 was found to approach histone levels in neuronal nuclei: around $16x10^6$ molecules per nucleus. The level in non-neuronal brain nuclei was 6-8 fold lower: ~$2x10^6$ molecules per nucleus. In the liver, MeCP2 levels were lower still: $0.5x10^6$ per nucleus (Skene et al., 2010).

Analysis of embryonic tissue found much lower levels of MeCP2 in the brains of both humans and mice. Proteins levels increase with neuronal maturation during postnatal development (Shahbazian et al., 2002a; Balmer et al., 2003). Levels reach a plateau at 5 weeks of age in mice (Skene et al., 2010) and around 10 years of age in humans (Shahbazian et al., 2002a). This increase was recapitulated using an *in vitro* system: neural precursors derived from embryonic mice were differentiated in culture into a mixture of neurons and glia. MeCP2 was shown to increase during differentiation reaching high levels in the resulting neurons but not in the glia. Despite its increased protein levels in maturing neurons, MeCP2 was found not to be required for determination of cell fate, as *in vitro* differentiation of neural precursors derived from *Mecp2*-nullmouse embryos resulted in the same ratio of neurons to glia (Kishi and Macklis, 2004). It is now believed that MeCP2 is required for the maintenance of neuronal function but not for neuronal development (Guy et al., 2007).

The regulation of MeCP2 protein levels is highly complex. Northern blot analysis detects *MECP2/Mecp2* transcripts of four different lengths: 1.8 kb, 5.4 kb, 7.5 kb and 10.2 kb (D'Esposito et al., 1996; Shahbazian et al., 2002a; Pelka et al., 2005). These are produced by the use of different polyadenylation sites in the long 3'UTR. Of these, the longest (10.2 kb) transcript is most abundant in human and mouse brains. It is also the most dynamic during embryonic and postnatal development: first decreasing in abundance and then increasing again in adulthood (Samaco et al., 2004; Pelka et al., 2005). Shahbazian et al. (2002a) compare the levels of the two most abundant *MECP2/Mecp2* transcripts (1.8 kb and 10.2 kb) to MeCP2 protein levels in a variety of human and mouse tissues and show that there is no correlation. The discovery of multiple miRNA target sites in the long 3'UTR partly explains this result (reviewed by McGowan and Pang, 2015). Several of these miRNAs have been shown to regulate MeCP2 translation, for example miR-132 (Klein et al., 2007).

The Northern blot analysis mentioned above is unable to distinguish between the two splice variants of *MECP2/Mecp2* due to the small size (124 bp) of the variant exon (exon 2). Kriaucionis and Bird (2004) showed that translation of the e2 isoform is less efficient due to interference from an upstream start in exon 1 (Fig. 1.2.2), adding further complexity to the regulation of MeCP2 protein levels. mRNA encoding the longer isoform, e1, is most abundant in the brains of human and mice. The increased abundance and translational efficiency together result in 10-fold higher levels of e1 protein compared to e2 protein in the brain. The higher ratio of e2 transcripts in other tissues (e.g. liver) may be partly responsible for the lower levels of MeCP2 protein detected in these tissues (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004).

1.2.4 The DNA binding specificities of MeCP2

The ability of MeCP2 to bind DNA probes that had been symmetrically methylated at CpG dinucleotides led to its initial discovery using Southwestern blotting (Lewis et al., 1992). This binding specificity has been proved repeatedly with *in vitro* techniques using a variety of DNA probes containing single or multiple mCpG sites

(Meehan et al., 1992; Valinluck et al., 2004; Klose et al., 2005; Mellén et al., 2012). MeCP2 binding to the methylated pericentromeric repeats *in vivo* was visualised in both mitotic (Lewis et al., 1992) and interphase (Nan et al., 1996) cultured cells and later in sections of brain tissue (Shahbazian et al., 2002a). In interphase mouse cells, the pericentromeric repeats can be visualised by Hoechst or DAPI staining as bright spots. The localisation of MeCP2 at these heterochromatic foci has been detected by immunohistochemistry or by fusion with visible tag (Nan et al., 1996; Shahbazian et al., 2002a; Kudo et al., 2003; Baubec et al., 2013). Chromatin immunoprecipitation followed by PCR (ChIP-PCR) or high throughput sequencing (ChIP-seq) has been used in a great number of studies to find where MeCP2 binds in the genome. MeCP2 was shown to preferentially bind to the methylated allele of the imprinted gene, *U2af1-rs1*, in mouse liver (Gregory et al. 2001). Unlike transcription factors, which show distinct peaks at their target sites, the MeCP2-binding signal in the murine brain is relatively flat, indicating binding over the majority of the genome. This is consistent with its high abundance, allowing it to occupy a large fraction of methylated cytosines, which occur at an average frequency of one every ~150 bp in the bulk genome. The MeCP2 binding signal was shown to track CpG methylation, decreasing dramatically at unmethylated CpG islands (Skene et al., 2010; Lagger, Connelly, Schweikert, et al., 2017).

The discovery that hmC and mCpH (where H is A, T or C) accumulate during neuronal maturation, correlating with the upregulation of MeCP2 (Lister et al., 2013; Guo et al., 2014; Szulwach et al., 2011) prompted investigations into whether MeCP2 is able to bind these modifications. MeCP2 was shown to bind hmC-containing probes *in vitro*, leading to the proposal that is also acted as a reader of this mark *in vivo* (Mellén et al., 2012). More recently, studies have shown that this binding is restricted to hmC in the CAC trinucleotide (and to a lesser extent the CAT trinucleotide) context (Lagger, Connelly, Schweikert, et al., 2017). This is consistent with earlier studies that show MeCP2 is unable to bind symmetrically modified hmCpG/hmCpG, and has weak affinity for asymmetrical hmCpG/mCpG sites (Valinluck et al., 2004; Hashimoto et al., 2012). The extremely low abundance of

hmCAC or hmCAT sites in neuronal DNA (Lister et al., 2013) suggests that its ability to recognise them has little biological relevance.

MeCP2 was also able to bind a probe containing mCpH sites *in vitro* (Guo et al., 2014). As with hmCpH, its recognition sequence was refined to mCAC and (more weakly) mCAT (Lagger, Connelly, Schweikert, et al., 2017). Despite the relatively low modification frequency (1.3-1.5% of CpH residues are methylated), the higher abundance of CpH compared to CpG nucleotides enables this mark to reach similar levels to mCpG sites (Lister et al., 2013). Furthermore, motif analysis of CpH methylation in adult dentate gyrus found a strong preference for the mCAC dinucleotide (Guo et al., 2014), making the vast majority of methylated CpH dinucleotide potential MeCP2 binding sites. This finding greatly increases the number of MeCP2 binding sites in the genome, including in CpG-depleted regions (Guo et al., 2014). Recent examination of MeCP2 ChIP-seq datasets using summit analysis confirms a preference for mCAC (but not mCAT) in mouse brains (Lagger, Connelly, Schweikert, et al., 2017).

Overall, these findings are consistent with the view that MeCP2 is a reader of the DNA methylome, recognising both mCpG and mCAC sites. How it interprets these signals has been the subject of extensive study, with multiple proposed roles that involve its ability to bind and recruit cofactors to genomic regions (see section 1.5). Analysis of these roles has furthered our understanding of how MeCP2 regulates proper neuronal function and how it is affected in *MECP2*-related disorders, such as Rett syndrome.

1.3 Rett syndrome and other *MECP2*-related disorders

1.3.1 Rett syndrome

Rett syndrome (RTT; Online Mendelian Inheritance in Man #312750) was first described in 1966 in a report published in German by Andreas Rett, an Austrian paediatrician. Subsequent reports describing patients in multiple countries attribute

Rett's name to the disorder (Hagberg et al., 1983; Kerr and Stephenson, 1985). RTT patients are almost all female, and female gender was initially a requirement for the diagnosis (Hagberg et al., 1985). Its prevalence is estimated to be 1 in 10-15,000 female live births (Hagberg, 1985). RTT is a severe neurological disorder and is often referred to as an autism spectrum disorder (Hagberg et al., 1983; Neul, 2012). Before the disease was fully described or well-known, patients were often misdiagnosed as autistic (Kerr and Stephenson, 1985). However, the differences in these conditions, particularly the motor defects displayed in RTT, suggest that this association is misleading (Kerr and Stephenson, 1985; Percy et al., 1988).

'Classical' RTT follows a well characterised clinical progression beginning at 6-18 months of age. Crucially, before the onset of symptoms at this time, patients present no neurological symptoms and at birth their weight and head circumference are within normal limits (Hagberg et al., 1985; Nomura et al., 1984). Most patients are born at term without complications after a usually uneventful pregnancy (Hagberg et al., 1983). This period of normal development for up to 6 months is essential for the diagnosis of RTT (The Rett syndrome Diagnostic Criteria Work Group, 1988).

The disease presents itself in four stages (Haas, 1988; Neul et al., 2010). Firstly, a period of stagnation (stage 1) occurs where patients fail to meet developmental milestones at the expected ages (Neul, 2012). Although this stage is often associated with onset of acquired microcephaly (Hagberg et al., 1985; Nomura et al., 1984), this is no longer deemed essential for diagnosis (Hagberg et al., 2001).

The most striking stage of Rett syndrome progression is the period of regression (stage two), beginning at 1 – 4 years of age (Hass, 1988; The Rett syndrome Diagnostic Criteria Work Group, 1988). Key features of this stage are the partial or complete loss of purposeful hand movements and speech, and gait ataxia (Hagberg et al., 1985; Hass, 1988; Neul et al., 2010). Hand movements begin to be replaced with stereotypies such as 'wringing', and some patients are no longer able to walk independently. Autistic behaviour is frequently reported at this stage, with patients displaying increased anxiety and loss of interest in their surroundings (Hagberg et al.,

1983; Neul 2012). Other common symptoms include progression of microcephaly (Kerr and Stephenson 1985), respiratory problems (reviewed by Lombardi et al., 2015), tremors (Neul, 2012) and the onset of seizures in some patients (Hagberg et al., 1983).

Stage three involves a period of relative stabilisation beginning at 4 – 10 years of age (The Rett syndrome Diagnostic Criteria Work Group, 1988; Neul, 2012), where patients may even show some improvement (Neul et al., 2010). Autistic behaviour often improves during this stage, with patients showing increased interest in social interaction (Neul, 2012). Despite a relatively stable mental status over this period, motor problems slowly progress. This includes further manifestation of hand stereotypies, bruxism (teeth grinding), worsening of gait impairment and the onset of scoliosis (sideways curvature of the spine). Seizure onset is also common at this stage (Neul, 2012; reviewed by Lombardi et al., 2015).

The fourth and final stage of disease progression is late motor deterioration starting in patients' teens or early twenties (Neul, 2012). Historically, this stage was associated with complete loss of ability to walk, but this varies between patients (Neul, 2012). Patients in later stages of RTT suffer from progressive scoliosis, muscle wasting and Parkinsonism (Syndrome Diagnostic Criteria Work Group, 1988; Neul, 2012). The rate of symptom progression at this stage is much slower than during stage two.

Consistent with reduced head circumference, RTT patients have a smaller brain size (Jellinger and Seitelberger, 1986). This is not due to neuronal death, but to reduced cell size. These smaller cells are more densely packed in RTT brains (Bauman et al., 1995). They also show decreased complexity, with a reduced density of dendritic spines (Armstrong et al., 1995).

Despite the appearance of severe symptoms in early childhood, RTT patients usually live long into adulthood. A recent longitudinal study based in the USA estimated survival at 50 years to be 70% (Tarquinio et al., 2015). The most common cause of

death is cardiorespiratory problems. Shorter lifespan correlates with factors indicating disease severity, e.g. reduced ambulation or low body weight. Survival may be improved by treatments that alleviate the symptoms of the disease: anti-seizure medication, physiotherapy, and nutritional plans or the use of a feeding tube in severe cases (Tarquinio et al., 2015; Neul, 2012). The decrease in number of deaths from frail condition supports the positive effect of these therapies (Tarquinio et al., 2015).

The fact that almost all RTT patients are female made a genetic cause seem likely. This was backed up by reports of complete concordance in monozygotic twins and the absence of female dizygotic twins who are both affected (Naidu et al., 1988; Zoghbi, 1988). It was predicted that RTT was a dominant X-linked disorder, with the absence of male cases explained by male embryonic lethality (Hagberg et al., 1983). The gene was difficult to map as cases were usually sporadic. However, a small number of RTT families (0.5-1% of cases) were reported containing multiple females with the diagnosis of RTT. In these families, obligate carrier females were protected from the disease by skewed X chromosome inactivation (XCI) or somatic mosaicism. Studies involving these families facilitated exclusion mapping using microsatellite markers on the X chromosome, pinpointing Xq28 as the responsible locus (Schanen et al., 1997; Sirianni et al., 1998). This finally led to the identification of *MECP2* as the causative gene in 1999 (Amir et al., 1999). This paper reported several causative mutations in *MECP2* - both missense and truncating - suggesting that the mutations disrupt protein function. RTT patients are heterozygous for a loss-of-function mutation in *MECP2*, with mosaic expression of the wild-type and mutant alleles due to XCI. Now, this locus is routinely sequenced in RTT patients and mutations are submitted to RettBASE, an online database (http://mecp2.chw.edu.au/). Sequencing data from patients' parents distinguish between inherited and sporadic mutations and help to determine causality.

Analysis of specific *MECP2* mutations in patients diagnosed with classical RTT has identified genotype-phenotype correlations. Unsurprisingly, large-scale genomic changes such as large deletions, insertions and splice-site mutations, result in more

severe disease symptoms. Within the protein coding sequence, early truncations (up to R270X) are more severe than more C-terminal truncations (Cheadle, 2000; Cuddapah et al., 2014). In general, point mutations located within the methyl-CpG binding domain (MBD) e.g. R106W, are more severe than those in the C-terminus, e.g. R306C. An exception to this is the milder mutation, R133C - an MBD mutant that retains some DNA binding (Cuddapah et al., 2014; Kudo et al., 2003; Brown, Selfridge et al., 2016). The presence of mutations in exon 1 suggests that loss of the more abundant isoform (e1) is sufficient to cause RTT. Truncating mutations in this exon result in disease at the more severe end of the RTT spectrum, consistent with other early truncating mutations (Cuddapah et al., 2014). To date, no RTT-causing mutations have been reported in exon 2, which would only affect isoform e2 (http://mecp2.chw.edu.au/). Although RTT-causing mutations in the 3'UTR have been identified in several patients (Fendri-Kriaa et al., 2010; http://mecp2.chw.edu.au/), it is yet to be established how they impact MeCP2 protein levels.

Despite the identification of *MECP2* mutations in 95-97% of RTT patients, it has not been added to the list of inclusion criteria for diagnosis, which remains purely clinical (Neul et al., 2010). The major cause of phenotypic variation is protective skewed X chromosome inactivation (reviewed by Chahrour and Zoghbi, 2007). A recent report of four patients with well-established common RTT-causing mutations and random XCI that exhibit much milder symptoms indicates that *MECP2* mutation and RTT are not always associated (Suter et al., 2014; see section 1.4.7).

1.3.2 Atypical Rett syndrome

Patients who meet some but not all the essential criteria for classical Rett syndrome are given the diagnosis of 'atypical RTT' (also referred to as 'variant RTT'). Atypical RTT can be subdivided into three clinically different groups: preserved speech variant, early-seizure onset variant, and congenital variant. Surprisingly, causative mutations in *MECP2* could only be identified in 50-70% of cases of atypical RTT (Neul et al., 2010). The latter two subgroups have been shown to

mostly be due to mutations in different genes: *CDKL5* and *FOXG1* in early-seizure onset variant and congenital variant, respectively. Recent reports have suggested that both these conditions often lack a clear period of regression – which is essential for a diagnosis of atypical RTT (Neul et al., 2010). Therefore, these disorders could instead be classed as distinct clinical entities (Neul, 2012). The third class of atypical RTT, the preserved speech variant, is mostly caused by mutations in *MECP2* (Neul et al., 2010). As its name suggests, patients are capable of some spoken language, sometimes forming sentences. Overall, it is a milder form of RTT, with increased mobility and a closer resemblance to autism than classical RTT (Neul, 2012). Interestingly, patients diagnosed with the preserved speech variants have the same causative mutations as classical RTT patients (http://mecp2.chw.edu.au/), further suggesting that *MECP2* mutations in patients cannot completely predict clinical manifestation.

1.3.3 RTT-causing mutations in males

The mosaic expression of the wild-type and mutant alleles of *MECP2* in RTT patients can only occur in heterozygous females. Exceptions where similar patterns of expression occur in males are a result of additional mechanisms: either Kleinfelter syndrome (XXY) or somatic mosaicism (resulting from a somatic mutation). The extension of the diagnostic criteria for Rett syndrome to include males (The Rett syndrome Diagnostic Criteria Work Group, 1988) permitted the identification of such cases (Salomão Schwartzman et al., 1999; Silberstein et al., 2001; Clayton-Smith et al., 2000).

For a long time it was believed that inheritance of RTT-causing mutations by hemizygous males resulted in lethality (Hagberg et al., 1983). This theory was put into question, however, by the absence of spontaneous abortions in RTT families (Schanen et al., 1997). In 1998, the first report of two cases of males with neonatal encephalopathy in RTT families was published, predicting the male phenotype of RTT-causing mutations (Schanen et al., 1998). This was confirmed by the discovery of *MECP2* as the causative gene the following year (Amir et al., 1999), which led to

sequencing of this allele in male patients with this condition: firstly in congenital cases where family members had been diagnosed with RTT and later in sporadic cases (Zeev et al., 2002; Kankirawatana et al., 2006). The remarkably low number of sporadic cases in hemizygous males may have been a contributing factor to the long-held belief that these mutations are lethal. This phenomenon is instead explained by the fact that most RTT patients inherit *de novo* mutations from their father due to the higher mutation rate in spermatogenesis compared to oogenesis (Trappe et al., 2001). Male babies inherit their sole *MECP2* allele on their mother's X chromosome, meaning their chance of inheriting *de novo* RTT-causing mutations is substantially lower.

It is now established that mutations that cause RTT in heterozygous females cause severe neonatal encephalopathy in karyotypically normal hemizygous males. These patients are diagnosed as new-borns with severe neurological problems, microcephaly, respiratory insufficiency, hypotonia and feeding difficulties (Kankirawatana et al., 2006). This condition results in premature death before the age of three due to respiratory problems (Kankirawatana et al., 2006; Schüle et al., 2008). A less severe condition has been reported in hemizygous males carrying certain RTT-causing mutations, including the common mild RTT mutation, R133C (Masuyama et al., 2005).

1.3.4 Other mutations in MECP2

Neul et al. (2010) proposed the term '*MECP2*-related disorders' to encompass all neurological conditions caused by mutations in *MECP2*. This includes classical and atypical RTT and male neonatal encephalopathy (described above) as well as milder forms of mental retardation.

Some *MECP2* mutations have been reported to cause 'male variant' RTT in karyotypically normal hemizygous boys. These patients present a disease similar to female RTT patients. The mutations responsible have not been found to cause RTT in females and would be predicted to retain more protein function. One example of

this is a 21 year old male with a *de novo* point mutation (P225L) in an uncharacterised region of the protein (Moog et al., 2003).

Milder mutations in *MECP2* have been identified as the cause of X-linked intellectual disability and autism in both male and female patients (Neul, 2012). Most of these mutations are single amino acid changes in the C-terminal region or short truncations of the end of the C-terminus (Zeev et al., 2002; Moog et al., 2003). Interestingly, the most common and well-studied of these mutations (A140V) is located in the methyl-CpG binding domain (MBD), but it does not disrupt DNA binding (Nan et al., 2007). This mutation causes mental retardation in males and very minor abnormalities in females, including their carrier mothers (Moog et al., 2003; Neul, 2012).

Despite the high level of evolutionary conservation and the range of disease-causing mutations in *MECP2*, the Exome Aggregation Consortium (ExAC) database reports a large number of mutations in *MECP2* in healthy individuals (http://exac.broadinstitute.org/). To date, this database contains genetic information from 60,706 people from diverse ethnic backgrounds (Lek et al., 2016). Within the coding sequence of the *MECP2* gene, the vast majority of mutations are single amino acid changes. Several short in-frame deletions and truncating mutations near to the C-terminus have also been found. Importantly, many of these mutations were found in hemizygous males, the most common of which were also found in homozygous females, providing strong evidence that these mutations do not disrupt protein function.

1.3.5 MECP2 duplication syndrome

Following the characterisation of MeCP2 overexpression in mice, which resulted in neurological symptoms (Collins et al., 2004; see section 1.4.3), *MECP2* duplication was identified a cause of X-linked intellectual disability in male patients (Meins et al., 2005; van Esch et al., 2005). The patient described in the first reported case study had a duplication of a 430 kb region in Xq28 containing the complete coding

sequence of 12 genes. *MECP2* duplication was predicted to be the cause of his condition as it is the only candidate gene with previous links to intellectual disability (Meins et al., 2005). Following the identification of more patients, molecular analysis of their DNA narrowed down the minimal duplication region to *MECP2* and its neighbouring gene, *IRAK1*, which encodes IL receptor-associated kinase 1(van Esch, 2011).

Patients with *MECP2* duplication syndrome have moderate to severe intellectual disability with impaired spoken language and autistic features. They have early infantile hypotonia and develop movement problems, such as gait abnormalities. They also suffer from tremors and seizures (van Esch et al., 2005; van Esch, 2011; Neul et al., 2012). This disorder results in premature death (around 25 years of age) from infections due to immunological dysfunction, which may be connected to the duplication of *IRAK1* (van Esch et al., 2005; van Esch, 2011). Unlike Rett syndrome, most cases of *MECP2* duplication syndrome are familial: with carrier mothers displaying no or much milder symptoms, such as anxiety and depression. Strikingly, this is due to extreme skewing (>85%) of X chromosome inactivation in most carriers (Esch et al., 2005; reviewed by Lombardi et al., 2015). The degree of disease severity is closely linked to XCI skewing: female family members with skewed XCI ratios around 70:30 display some mild symptoms (van Esch, 2011). Cases of sporadic intellectual disability in females that are due to *MECP2* duplication have been reported. The majority of these patients have a duplication of *MECP2* onto an autosome, resulting in symptoms resembling male patients (van Esch, 2011). To date, only two female patients with *de novo MECP2* duplications on their X chromosome and random XCI have been reported. They display moderate intellectual disability due to absence of protective XCI skewing (Grasshoff et al., 2011). The high incidence of phenotypically protective XCI skewing in females in familial cohorts suggests the presence of a mechanism for preferential silencing of an inherited X chromosome containing the duplicated region (van Esch, 2011). Overall, characterisation of the effects of *MECP2* gene duplication to all or a fraction of cells furthers our understanding of the importance of correct MeCP2 dosage.

1.4 Mouse models with *Mecp2* mutations

Mouse models have been used extensively to study the function of MeCP2 *in vivo*. There are several key advantages of using mouse models to study MeCP2 function and RTT. Inbred laboratory mouse strains enable the production of large cohorts of isogenic animals, resulting in strong genotype-phenotype correlations. It is also possible to carry out histological and biochemical analysis of relevant tissues (including the brain) at multiple ages or stages of disease progression. Finally, genetic editing techniques enable the design of mutant mouse models to answer specific scientific questions, e.g. tissue-specific knock-outs.

Both *Mecp2*-null and RTT-mutation knock-in mice recapitulate the symptoms of the disease and facilitate characterisation of molecular phenotypes. Cell type-specific knock-outs show that these phenotypes are due to neuronal dysfunction, and hypomorph and overexpression models have highlighted need for correct MeCP2 dosage. Excitingly, it has been shown that reintroduction of MeCP2 after the onset of symptoms in null mice using genetic manipulation techniques leads to disease reversal. This ground-breaking experiment proves that Rett syndrome is theoretically curable. MeCP2 mouse models are now being used to test therapeutic strategies with the hope that they will be used for RTT patients in the future.

1.4.1 Mecp2-null mice recapitulate the symptoms of patients with Rett syndrome

Initial attempts to produce *Mecp2*-null mice in Adrian Bird's lab resulted in lethality of the chimaeras, except when the contribution of null ES cells was negligible (Tate et al., 1996). In this study, exons 3 and 4 – which contain almost all of the protein coding sequence – were replaced with β-galactosidase fused to a Neomycin resistance gene (β-geo). Without the knowledge that mutations in *MECP2* cause RTT in humans, it was accepted that the presence of *Mecp2*-null cells in the chimaeras was fatal. To determine whether this is the true mouse phenotype or whether lethality was due an adverse effect of the β-geo transgene, an alternative approach was employed. The Bird and Jaenisch labs both used Cre-*loxP* technology to produce conditional *Mecp2* alleles (Guy et al., 2001; Chen et al., 2001). *LoxP* sites were

inserted into the *Mecp2* locus flanking vital exons: exons 3 and 4 (Guy et al., 2001) or exon 3 (Chen et al., 2001) in ES cells. These targeted ES cells were used to produce viable chimaeras, leading to the establishment of conditional mouse lines. Both labs successfully produced viable heterozygous female (-/+) and hemizygous male (*-/y*) mice after recombination of the conditional alleles with ubiquitous deleter (*CMV-Cre*) mice, proving that loss of *Mecp2* does not cause lethality. Furthermore, wild-type and null males were born at Mendelian ratios (Guy et al., 2001). This result is consistent with the effect of loss-of-function mutations in *MECP2* in humans (Amir et al., 1999; Zeev et al., 2002).

Since RTT occurs in female patients who are heterozygous for *MECP2* mutations, heterozygous female mice could be considered to be the 'true' models of RTT. Both studies detected disease symptoms in these animals, including inertia, gait ataxia and hind limb clasping (Guy et al., 2001; Chen et al., 2001). Astonishingly, these symptoms did not appear until 3-12 months of age. Comparison to human RTT patients suggests that disease progression due to the absence of *Mecp2* more closely follows absolute time than developmental time. While this discovery furthers our knowledge of how the disease develops, and enables heterozygous females to be used for breeding purposes before symptom onset, it makes the study of RTT phenotypes in female mouse models very slow.

Another surprising result was that hemizygous male mice are not only viable at birth, they remain symptom-free for 3-5 weeks (Guy et al., 2001; Chen et al., 2001). After this period, symptoms rapidly progress, resulting in premature death between 6 and 12 weeks of age. Symptoms include hypoactivity, gait abnormalities, hind limb clasping, trembling, irregular breathing, irregular wearing of the teeth and jaw misalignment. Both studies report an obesity phenotype in null animals on a mixed genetic background, but the biological relevance of this is unclear as Guy et al. (2001) show that backcrossing onto a C57BL/6 background causes *Mecp2*-null animals to instead be underweight compared to their wild-type littermates. Perhaps more relevant is the finding that the outbred *Mecp2*-nulls have smaller brains, despite their increased body weight. Further analysis of the brains of null mice shows that

they consist of smaller, more densely packed neurons with fewer dendritic branches (Chen et al., 2001; Kishi and Macklis, 2004). Behavioural analysis has confirmed the reduced activity phenotype, and shown that these mice have motor impairment on the Accelerating Rotarod. They also display decreased anxiety: spending more time in the open regions of the Elevated Zero Maze. Lastly, abnormal electrophysiology has been reported in these animals, including by EEG (electroencephalogram) recordings, although none of these studies report the presence of seizures (Asaka et al., 2006; Goffin et al., 2012). These results are summarised in Table S1A.

Overall, the symptoms present in the *Mecp2*-null mice recapitulate many features of RTT, including motor defects, gait abnormalities, breathing problems, tremors, abnormal electrophysiology, decreased brain and nuclear size and reduced dendritic branching (Guy et al., 2001; Chen et al., 2001; Kishi and Macklis, 2004; Asaka et al., 2006; Goffin et al., 2012). While it is tempting to make a connection between the characteristic hand stereotypies of RTT patients and the hind limb clasping displayed by *Mecp2*-null mice, there is no evidence to support this. In fact, hind limb clasping is a common feature of mice with neurological dysfunction (reviewed by Lalonde and Strazielle, 2011). The decreased anxiety phenotype of the *Mecp2*-null mice is surprising as RTT patients tend to display autistic behaviours with heightened anxiety. This may be explained by the fact that mice are naturally very anxious with high risk perception: behaviours that could become impaired with suboptimal neurological function.

1.4.2 Mecp2 knock-in mice with RTT-causing mutations and milder mutations reflect the severity scale in patients and help determine the importance of domains in MeCP2

Several knock-in mouse models with *Mecp2* mutations have been produced to date (Fig. S3). In the majority of cases, it is hemizygous males rather than heterozygous females that have been studied. This is due to the delayed onset and mildness of the symptoms in heterozygous females: for example they are only just detectable at ~10 months in R133C knock-in mice. Mice with RTT-causing mutations include three of

the most common missense mutations (R133C, T158M and R306C; summarised in Table S1B-C) and three of the most common nonsense mutations (R168X, R255X and R270X; summarised in Table S1D). A mouse carrying another RTT-causing mutation, T158A, demonstrates the importance of the threonine residue at position 158 without introducing the bulky methionine residue. The R273X mouse indicates that truncating mutations located further towards the C-terminus cause disease symptoms that progress more slowly.

The mouse models of RTT resemble the phenotype described in the male null mice, developing symptoms after weaning, which result in premature death. However, the presence of mutant forms of MeCP2 protein tends to lead to slightly milder phenotypes. This is particularly evident in the median survival of the hemizygous males, for example the median survival of R273X mice is 29 weeks, compared to ~9 weeks in nulls (Baker et al., 2013; Guy et al., 2001; Chen et al., 2001). Guy et al., (2007) developed a scoring method for quantifying six symptoms associated with the RTT-like phenotype in mice: hypoactivity, gait abnormalities, hind-limb clasping, tremor, breathing problems, and degeneration of general condition. This method can be repeated multiple times with the same animals, tracking symptom progression. It is a vital tool for comparing the severity of different mouse lines. Brown, Selfridge et al. (2016) show that mice with the three most common missense mutations mirror the severity spectrum of RTT patients (T158M>R306C>R133C), in terms of symptom progression and age of death. A separate study characterising the phenotype of mice with the A140V mutation (which causes intellectual disability in boys) has shown that hemizygous male mice have a normal lifespan and no RTT-like symptoms were detected (Table S1E). They do, however, have abnormal electrophysiology (Ma et al., 2014) and histological analysis shows increased cell packing density in the brain and decreased dendritic branching (Jentarra et al., 2010).

Analysis of truncating mutations in both male and female patients has shown that early truncations (up to R270X) have a more severe condition than later truncations (G273fs or later) (reviewed by Villard, 2007; Cuddapah et al., 2014). The three knock-in male mice with early truncations (R168X, R255X and R270X) all closely

resemble *Mecp2*-nulls, with rapid symptom onset and a median survival of 9-12 weeks (Lawson-Yuen et al., 2007; Baker et al., 2013; Pitcher et al., 2015). This phenotype could be explained in the R168X and R255X mice by the great reduction or absence of the mutant protein. The presence of a normal level of protein in the R270X mice proves that these symptoms must be due to reduced functionality of the truncated protein (Baker et al., 2013). Crucially, G273X mice display a milder phenotype with slower symptom progression and increased survival time, despite only having three additional three amino acids. The most C-terminally truncated mouse produced so far, '308/y', expresses a protein consisting of residues 1-308. This mutation has not been identified in patients, but was predicted to enable prolonged survival in male mice (Shahbazian et al., 2002b). This was indeed the case, with 90% surviving more than one year (the others died of unknown causes after 10 months). Although their phenotype is much milder, extensive testing has identified hypoactivity, increased anxiety, impaired motor function, abnormal social behaviour, abnormal electrophysiology and seizures in these mice (Table S1E; Shahbazian et al., 2002b; Moretti et al., 2005; McGill et al., 2006).

It is clear from the results of these studies that mouse models provide a reliable and relatively quantifiable assay of MeCP2 protein activity; and can be used to determine the importance of specific residues or domains for protein function *in vivo*.

1.4.3 Mecp2 overexpression and hypomorph mice highlight the important of correct dosage

Prior to the discovery of *MECP2* duplication syndrome, Collins et al., (2004) produced a transgenic mouse line that overexpressed MeCP2 at ~2 fold. This was achieved by the random insertion of a human transgene containing the entire *MECP2* genomic locus. When bred onto a *Mecp2*-null background, this transgene could rescue the loss-of-function phenotype: indicating that the human and mouse genes are functionally equivalent. This study extensively characterised the mouse overexpression phenotype: the animals showed no symptoms for the first 10-12 weeks, followed by the development of hypoactivity, forepaw clasping, seizures,

abnormal electrophysiology and eventually kyphosis. Overexpression also affects survival, with 30% of animals dying between 20 weeks and one year (Table S1F). Many of these symptoms resemble those in loss-of-function mouse models, although contrastingly these mice displayed increased anxiety and enhanced learning in both motor function and contextual fear conditioning (Collins et al., 2004; Sztainberg et al., 2015).

Mild RTT-like phenotypes have also been reported in a floxed allele of *Mecp2*, which has been shown to have reduced expression. The insertion of a human β-globin intron and polyadenylation signal followed by a *TK-Neo* selection cassette and *loxP* site at the beginning of the long 3'UTR before the alternative polyadenylation sites prevents their usage (Guy et al., 2001; Kerr et al., 2008). This lowers protein levels to between 30-90% of wild-type controls in different brain regions, resulting in it being referred to as a 'hypomorph'. This reduction in MeCP2 levels has no effect on survival, but causes weight gain, breathing problems and hind limb clasping. Extensive behavioural testing has detected decreased anxiety, motor defects, learning defects and abnormal social interaction in these mice (Table S1F; Samaco et al., 2008; Kerr et al., 2008).

Together, these studies indicate that the levels of MeCP2 protein are of vital importance for proper neurological function.

1.4.4 Tissue-specific deletion or expression of Mecp2 shows that RTT-like phenotypes are due to abnormal neuronal function

Although MeCP2 is most highly expressed in the brain, it is present in all tissues. To determine whether the RTT-like symptoms displayed by the null mice were due to abnormal CNS function or to the absence of MeCP2 in peripheral tissues, both original studies also produced brain-specific knock-outs (Guy et al., 2001; Chen et al., 2001). To do this, they crossed *floxed Mecp2* mice with mice containing a transgene expressing Cre recombinase under a *Nestin* promoter (*Nestin-Cre*). This results in the deletion of *Mecp2* from over 90% of neuronal and glial cells in the

brain (Chen et al., 2001). The phenotype of these mice is indistinguishable from the nulls in both studies, showing that CNS dysfunction is the major contributor to RTT pathogenicity.

Several other cell-type specific *Mecp2* knock-out mice have now been produced to determine the importance of different neuronal subtypes to the RTT phenotype (reviewed by Guy et al., 2011). *Mecp2* deletion from GABAergic neurons (using *Viaat-Cre*) has the most severe phenotype, and is the only one of these lines that has a reduced survival. The importance of MeCP2 function in GABAergic neurons is consistent with its high protein levels: ~50% higher than other neuronal subtypes (Chao et al., 2010). Despite this strong correlation, it is important to consider the number of cells from which MeCP2 has been deleted in each cell-type specific knock-out as well as their type. Removal of MeCP2 from only the GABAergic neurons in the forebrain (using *Dlx5/6-Cre*) results in a much milder phenotype (Chao et al., 2010). Another complicating factor may be the difference in the timing of these deletions (which occur when the *Cre* transgenes are activated).

A recent study produced a novel mouse model with brain-only *Mecp2* expression to determine the phenotypic consequences of the loss of MeCP2 in peripheral tissues (Ross, Guy et al., 2016). The *Mecp2* gene was inactivated by the insertion of a floxed transcriptional stop cassette in intron 2, which terminates transcription (Guy et al., 2007). This silencing was retained in the periphery but alleviated in 90% of neurons and glia by excision of the stop cassette using a *Nestin-Cre* transgene. These mice did not present any of the overt symptoms associated with RTT models, but rigorous testing at 14-16 weeks of age identified milder phenotypes including reduced spontaneous activity, increased exercise fatigue, abnormal kidney histology and bone defects. Additional abnormalities in peripheral tissues observed in null controls were not present in the peripheral knock-out mice. This mouse model reveals important roles of MeCP2 in peripheral tissues but confirms that other peripheral abnormalities are due to neuronal dysfunction (Ross, Guy et al., 2016).

Luikenhuis et al. (2004) showed that neuronal-only expression of MeCP2 from the *Tau* locus rescues the *Mecp2*-null phenotype. The absence of symptoms in these mice is surprising as the level of the Tau-MeCP2 protein was 2-4 times that of wild-type MeCP2. Heterozygous Tau-MeCP2 expression on a wild-type background (bringing the total protein levels to 3-5 fold) still did not result in neurological dysfunction. Symptoms were only apparent when these levels were elevated even further in homozygous *Tau-Mecp2* mice on either a null (total levels 4-8 fold) or wild-type (total levels 5-9 fold) background (Luikenhuis et al., 2004). This study was carried out before the characterisation on the transgenic 2-fold overexpression mice (Collins et al., 2004) or the identification of *MECP2* duplication in patients with intellectual disability (van Esch et al., 2005). With this knowledge, these results suggests that neurons are more tolerant to high levels of MeCP2 than other cell types, and that *MECP2* duplication syndrome may partly be a consequence of elevated protein levels in other cell types. Although a stronger phenotypic effect by additional expression of the human protein in the transgenic mice (Collins et al., 2004) should also be considered when comparing these two mouse models.

1.4.5 Inactivation of Mecp2 after development causes symptom onset and premature death

The low levels of MeCP2 protein in the early stages of postnatal development (Shahbazian et al., 2002a) and the delay of symptom onset in null mice (Guy et al., 2001; Chen et al., 2001) questioned the requirement for MeCP2 function in embryonic development. Furthermore, postnatal deletion of *Mecp2* in forebrain excitatory neurons using CamK-Cre results in the development of RTT-like symptoms, albeit with delayed onset (Chen et al., 2001). It therefore seemed likely that *Mecp2*-null phenotypes are not due to its absence during development: rather MeCP2 is required for proper function of mature neurons. To test this, an inducible *Cre* allele (*CreER*^{T}, fused to a modified oestrogen receptor), facilitated the deletion of the conditional *Mecp2* allele at a range of stages of postnatal development upon Tamoxifen injection (McGraw et al., 2011; Cheval, Guy et al., 2012; Nguyen et al., 2012; Du et al., 2016). At all ages, removal of *Mecp2* resulted in the development of

RTT-like symptoms (quantified by the scoring system described above), resulting in premature death. Although comparison of the dynamics of disease progression post-injection in different studies is complicated by differences in injection regimes, two separate studies show that symptoms progress more rapidly when *Mecp2* is inactivated at later stages in development (Cheval, Guy et al., 2012; Du et al., 2016). This effect is particularly dramatic in the most recent study where *Mecp2* inactivation in adult mice ≥15 weeks of age results in rapid symptom progression and lethality with 50% survival after 7 days (Du et al., 2016).

Microcephaly and reduced brain size in RTT patients and hemizygous male mice is associated with deceleration of head growth as symptoms progress during post-natal development. Strikingly, removal of *Mecp2* from fully grown adult mice (10 or 15 weeks of age) also results in reduced brain weight and neuronal nuclear size, suggesting shrinkage has occurred. The dynamics of brain shrinkage mirror the rate of symptom progression in these mice (Nguyen et al., 2012; Du et al., 2016).

1.4.6 Proof-of-principle experiments using genetically engineered mice show that Rett syndrome and MECP2 duplication syndrome are both theoretically curable

A major breakthrough in the RTT field was the discovery that induced activation of the *Mecp2* gene in null mice before and after the onset of symptoms prevented or reversed the disease, respectively. The initial study by Guy et al. (2007) and a follow-up study by Robinson et al. (2012) used genetic manipulation in mice to engineer a proof-of-principle experiment showing that RTT is theoretically curable. A transcriptional stop cassette (flanked with *loxP* sites) was inserted into intron 2 of the endogenous *Mecp2* allele, leading to transcriptional silencing. The resulting male mice had ~2.5% of wild-type levels of MeCP2 protein in the brain, which led to the onset of severe symptoms and premature death, closely resembling null mice. Heterozygous *Stop/+* females were crossed to males that carried a transgene encoding Tamoxifen-inducible Cre (*CreER^T*) to produce *Stop/y CreER^T* males. In the absence of Tamoxifen, this has no effect on the phenotype, proving the robustness of the system. Following injection with Tamoxifen, the CreER^T protein was

translocated into the nucleus where it excised the transcriptional stop cassette, leading to gene activation. Unexpectedly, rapid activation of the *Mecp2* gene before symptom onset (3-4 weeks of age) led to toxicity. This was overcome by a more gradual injection regime, preventing disease onset in these mice. This same regime was used in symptomatic animals (11-17 weeks of age), leading to amelioration of RTT-like symptoms and improved survival. Disease reversal was also possible in symptomatic heterozygous female mice (with injections beginning between 20 and 42 weeks of age).

A recent study has demonstrated rescue of the *Mecp2* overexpression phenotype in mouse models using either genetic manipulation or antisense oligonucleotides (Sztainberg et al., 2015). *Mecp2* overexpressing mice with a human transgene (Collins et al., 2004; described in section 1.4.3) were crossed with 'floxed' mice – with *loxP* sites either side of exon 3 (Chen et al., 2001). Unlike the hypomorphic allele (described in section 1.4.3), this alternative floxed allele has no effect on protein level. The resulting $Mecp2^{lox}$ $MECP2_{tg}$ male mice resemble the overexpression mice, displaying enhanced motor performance on the Accelerating Rotarod, impaired social behaviour and abnormal electrophysiology. The presence of Tamoxifen-inducible Cre recombinase ($CreER^{T}$) facilitated genetic rescue by excision of the floxed allele. Treatment with a 4-week long regime of Tamoxifen injections (starting at 8-9 weeks of age) led to an almost 2-fold reduction in the amount of MeCP2 protein. A second cohort of Tamoxifen-treated mice underwent extensive behavioural testing, displaying no phenotype in any of the tests performed. Control $Mecp2^{lox}$ $MECP2_{tg}$ mice underwent vehicle-only injections and displayed symptoms such as hypoactivity, increased anxiety and increased motor learning in these tests. This study also demonstrated a more therapeutic approach to reduce protein level in the transgenic mice: using antisense oligonucleotides (ASOs). ASOs were designed to target the human transgenic allele, reducing total protein levels by almost 2-fold. As observed with the mice in the genetic rescue cohort, the ASO-treated mice did not display any of the phenotypes associated with *Mecp2* overexpression. This study shows that, like Rett syndrome, *MECP2* duplication syndrome is theoretically curable.

1.4.7 Several therapeutic therapies are currently being developed, with relative success in preliminary studies using mouse models

After the discovery that delayed activation of the endogenous *Mecp2* gene in null mice was able to reverse neurological symptoms (Guy et al., 2007), the search began for a cure that could be administered to RTT patients. Mouse models have been used to test multiple strategies, each with their own benefits and complications, which I shall discuss briefly here.

The first strategy is introduction of exogenous MeCP2, either by Gene Therapy or by Protein Replacement Therapy (reviewed by Katz et al., 2016). AAV viruses are the preferred choice for Gene Therapy studies as they can cross the blood brain barrier (BBB); infect both neuronal and non-neuronal cells with a high transduction efficacy; and the viral DNA persists as a stable episome – providing long-lasting treatment and avoiding problems associated with insertional mutagenesis (reviewed by Gadalla et al., 2011). To date, several studies have used AAV vectors to introduce human or mouse *Mecp2*/*MECP2* cDNA into *Mecp2*-null male mice, resulting in phenotypic rescue (Table 1.4.7; Gadalla et al., 2013; Garg et al., 2013; Matagne et al., 2017; Gadalla et al., 2017). These studies have varied a number of parameters to maximise phenotypic rescue (prolonged survival and amelioration of neurological symptoms) and to increase transduction efficiency in the brain whilst reducing toxic effects from overexpression (which most frequently occurred in the liver). The varied parameters include: vector design (regulatory sequences), AAV serotype, injection route and dosage (summarised in Table 1.4.7). So far, vectors have only been introduced into pre-symptomatic mice, the oldest of which were 4-6 weeks of age (Gadalla et al., 2013; Garg et al., 2013; Matagne et al., 2017). Further work is needed to determine whether virally-delivered *Mecp2*/*MECP2* can result in symptom reversal. At present, there have been no experimental trials of Protein Replacement Therapy in mice. A priority for developing this strategy is designing modifications to MeCP2 protein that enable it to cross the brain barrier and enter cells. Possible tags include the TAT sequence from HIV (Lin et al., 2004). A major concern for both

Gene Therapy and Protein Replacement Therapy is MeCP2 dosage given the deleterious effects of its overexpression. The mosaic pattern of expression in female RTT patients further complicates this. It is important to note, however, that female patients with X-linked *MECP2* duplication and random XCI have much milder symptoms than males with *MECP2* duplication syndrome (Grasshoff et al., 2011), suggesting that introduction of MeCP2 to all cells will still greatly improve RTT patients' quality of life. Phenotypic improvement was been demonstrated in heterozygous (-/+) female mouse models after systemic injection of scAAV9/*Mecp2* at 10-12 months of age. Vector-treated mice displayed symptom stabilisation and performed better in behavioural paradigms than controls that had been injected with scAAV9 lacking a viral insert (Table 1.4.7; Garg et al., 2013). High transduction efficiency and near-physiological protein level is also a requirement for Protein Transduction Therapy. Though, as the exogenous protein would not persist, dosage can be altered over time.

An alternative method for introducing wild-type MeCP2 protein into only the mutant cells in female RTT patients is activation of the silenced allele on the inactive X chromosome. Current work is developing small molecules that either activate several X-linked genes or just MeCP2 (reviewed by Katz et al., 2016).

If successful, all of the strategies described above would result in the presence of both wild-type and mutant proteins in the same cells. This could be deleterious if the mutant protein has a dominant negative effect. The use of *Mecp2*-null males and -/+ females by the Gene Therapy studies does not address this concern. Only the most recent paper (Gadalla et al., 2017) used *Mecp2*-mutant mice, successfully demonstrating phenotypic rescue in T158M-EGFP males. Additional experimental evidence using overexpressing mouse models suggests that the presence of mutant MeCP2 on top of normal levels of wild-type protein does not have phenotypic consequences (Heckman et al., 2014; Koerner, unpublished).

Reference	Vector design	Injection method	Results
Gadalla et al., 2013	Human *MECP2*-e1 cDNA with C-terminal Myc tag; *chicken β-actin (CBA)* promoter. ssAAV2 vector backbone with AAV9 capsid	Intracranial injection into *Mecp2*-null males at P0-2. Dose: $4.8x10^{10}$ vg (viral genomes)/mouse	~7-42% cells in different brain regions transduced, around physiological levels. Prolonged lifespan (increased by ~7 weeks), reduced symptom severity, reduced phenotypes in behavioural tests (hypoactivity and motor ability). No rescue of reduced body weight phenotype or breathing apneas. Note: no adverse effects in vector-treated WT animals
Gadalla et al., 2013	Human *MECP2*-e1 cDNA with C-terminal Myc tag;229 bp fragment of mouse *Mecp2* promoter scAAV2 vector backbone with AAV9 capsid	Systemic injection (tail vein) into *Mecp2*-null males at 4-5 weeks of age. Dose: $5x10^{11}$ vg/mouse	~2-4% cells in different brain regions transduced (mostly in neurons). Highest transduction in liver and spleen. Prolonged lifespan (increased by ~5 weeks). No rescue of reduced body weight phenotype. Liver toxicity reported.
Garg et al., 2013	Mouse *Mecp2*-e1 cDNA (untagged); *Mecp2* promoter (730 bp) scAAV2 vector backbone with AAV9 capsid)	Systemic injection (tail vein) into *Mecp2*-null males at 4-6 weeks of age; -/+ females at 10-12 months of age. Dose: $3x10^{12}$ vg/mouse	Males: ~8-25% cells in different brain regions transduced (proteins expressed at around physiological levels). Prolonged lifespan, reduced symptom severity. Females: scores stabilised ~1 (vehicle-only controls scores ~6), reduced phenotypes in behavioural tests (e.g. motor ability). No seizures (observed in controls). Survival not compromised.
Matagne et al., 2017	Mouse *Mecp2*-e1 cDNA (untagged) – codon optimised (most frequently used codons); shortened (223 bp) *Mecp2* promoter scAAV9	Systemic injection (tail vein) into *Mecp2*-null males at P30 (4 weeks of age). Dose: $2x10^{11}$ vg/mouse	~10-20% cells in different brain regions transduced. Prolonged lifespan (increased by ~6 weeks), reduced phenotypes in behavioural tests (hypoactivity and motor ability), partial rescue of decreased weight phenotype, decreased breathing defects (apneas).
Gadalla et al.,2017 **Best conditions**	Human *MECP2*-e1 cDNA with C-terminal Myc tag; 426 bp of mouse *Mecp2* promoter and selected regulatory elements from mouse *Mecp2* 3'UTR. scAAV2 particles in AAV9 capsid.	Intracranial injection at P0-3 into *Mecp2*-null males. Dose: $2x10^{11}$ vg/mouse	~20-40% cells transduced, most at physiological levels (some overexpressing) Greatly prolonged lifespan (increased by ~26 weeks), reduced symptom severity. No rescue of decreased body weight phenotype.

Table 1.4.7: Gene Therapy studies using *Mecp2*-null mice

Several other therapeutic strategies are being developed that focus on repairing the mutant allele. These could be more promising given the importance of obtaining the correct levels of MeCP2 protein expression. These strategies include gene editing, RNA editing and the read-through of nonsense mutations (reviewed by Katz et al., 2016). Gene editing would provide a permanent cure with the correct level of expression, making it the most appealing method. Unfortunately, despite vast development in the relevant technologies, this is not yet possible (discussed in more detail below). RNA editing enzymes, Adenosine deaminases, perform transitions from adenine to iosine (read as guanine). Adenosine deaminase has been engineered to require an RNA guide (ADAR: Adenosine Deaminase that Acts on RNA), enabling transcript specificity (Schneider et al., 2014). Development of this technology is required to optimise delivery and efficiency of this enzyme (reviewed by Katz et al., 2016). While off-target editing may occur, this is less of a problem than in gene editing as edited RNAs do not persist in cells. Nonsense mutations can be targeted with small molecules that promote read-through. This was shown to be successful in Gentamicin-treated fibroblasts derived from R255X knock-in mouse embryos (Pitcher et al., 2015). Although both of these strategies are limited to specific kinds of RTT-causing mutations, a high proportion of patients have relevant mutations (particularly non-sense mutations).

Advances in gene editing using CRISPR/Cas9 technology have brought great hope to the field. The CRISPR (clustered regularly interspaced short palindromic repeats)-associated RNA-guided endonuclease, Cas9, is a bacterial DNA endonuclease from *Streptococcus pyogenes* that is involved in its adaptive immune system. Cas9 makes a double-stranded cut in DNA at a site determined by the sequence of a guide RNA (gRNA). This break is repaired in one of two ways: either by non-homologous end joining, which is usually imprecise; or by homology-directed repair using a homologous template (reviewed by Hsu et al., 2014). Over the last few years, substantial progress has been made in our understanding of this technology, particularly in the design of guide RNAs to minimise off-target hits and maximise homology-dependent repair. Several alternative endonucleases from different

bacterial species have now been identified and characterised, including ones that are small enough to be packaged into viral vectors (Ran et al., 2015). Despite these advancements, rates of homologous repair using an introduced donor molecule are very low in somatic cells (Shah et al., 2016). Gene editing protocols using cell lines involve steps to select and grow up correctly targeted clones, but this is impossible in existing tissues. Sadly, trials using drugs to promote homology-directed repair by inhibiting the alternative mechanism (non-homologous end joining, NHEJ) have so far been unsuccessful. Prospects are perhaps more hopeful for patients with *MECP2* duplication syndrome: using CRISPR-mediated cleavage followed by improper repair by NHEJ to destroy the additional *MECP2* allele. Strategies would have to be carefully designed, potentially using patient-specific SNPs, to only target one of the copies of *MECP2*.

The study by Sztainberg et al. (2015), described in section 1.4.6, proposes antisense oligonucleotides (ASOs) as a potential therapeutic strategy for patients with MECP2 duplication syndrome. The initial experiments with mouse models presented in this study look promising: protein levels were successfully reduced and the mice did not display neurological symptoms. Administration to patients will be more challenging as the design of this study takes advantage of the differences in the sequences of the endogenous mouse allele and the human transgene, enabling ASOs to be made that specifically target the human allele. To develop this strategy further to administer ASO treatment to patients, great care will have to be taken to determine dosage. This paper also highlights the need for continuous treatment, with the levels of MeCP2 protein increasing after its cessation.

An alternative genetic strategy for the development of potential therapies for RTT and other *MECP2*-related disorders is the identification of modifying mutations. The case report describing four patients with common RTT-causing mutations but much milder symptoms proposes that they carry additional genetic modifying mutations (Suter et al., 2014). However, until a larger number of patients like this are discovered, identification of the relevant mutations from genome sequencing would be very difficult. Buchovecky et al. (2013) have carried out a mutagenesis screen

using the chemical mutagen N-ethyl-N-nitrosourea (ENU) to discover mutations that suppress the phenotype of *Mecp2*-null male mice. They describe a nonsense mutation in *Sqle*, the gene encoding squalene epoxidase, that improves longevity, activity and motor abilities. Since SQLE is involved in cholesterol biosynthesis, it was proposed that statins might have a similar effect on the null phenotype. This proved to be the case, with better results than the *Sqle* mutation. Several other candidate genes were identified in the screen, including several known interaction partners of MeCP2 (Monica Justice, personal communication). Further work is required to understand the modifying mechanisms of these proteins. Additionally, it is important to determine whether genetic modifiers identified in studies using null mice have the same effect in mice expressing mutant forms of the protein that retain some functionality.

1.5 MeCP2: a multifunctional protein?

While the studies described above detail the spatial and temporal requirement of MeCP2 activity *in vivo*, they tell us very little about its molecular function. These *Mecp2*-mutant mouse models as well as cell culture-based systems and *in vitro* assays have been used to determine the functions of MeCP2 protein and how these are disturbed in *MECP2*-related disorders. MeCP2 has been implicated in several cellular processes leading to it commonly being referred to as a multifunctional protein. These processes involve its interaction with over forty binding partners (Fig. 1.5). Here, I will present the evidence for each of these in turn. At the end of this section, I will discuss how these activities may be regulated by neuronal activity-dependent post-translational modifications.

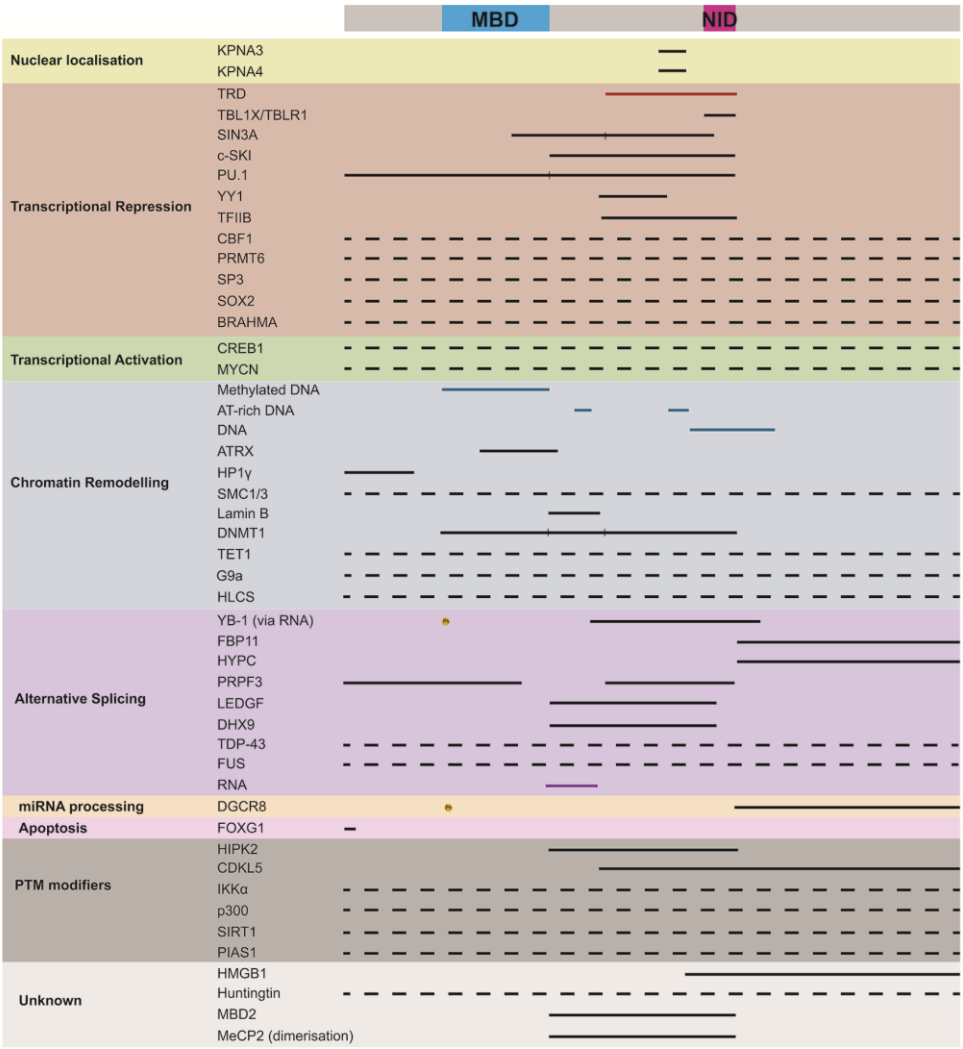


Figure 1.5: Binding sites of MeCP2 interaction partners
Schematic showing the binding sites of MeCP2 interaction partners characterised by their proposed function when complexed with MeCP2. Protein interaction sites are shown in black, DNA in blue and RNA in purple. The minimal domain required for transcriptional repression of a reporter gene (the TRD) is shown in red. Unmapped interactions are shown as dotted lines. Interactions with YB-1 and DGCR8 require phosphorylation of Ser80 (shown in orange).
References: KPNA3/4 (Baker et al., 2015); TRD (Nan et al., 1998); NCoR/SMRT complex components - directly with TBLX1/TBLR1 (Lyst et al., 2013); SIN3A (Nan et al., 1998; Lyst et al., 2013); c-SKI (Kokura et al., 2001); PU.1 (Suzuki et al., 2003); YY1 (Forlani et al., 2010); TFIIB (Kaludov et al., 2000); CBF1 (Mann et al., 2007); PRMT6 (Chawan et al., 2011); SP3 (Hwang et al., 2010); SOX2 (Szulwach et al., 2010); BRAHMA (Harikrishnan et al., 2005); CREB1 (Chahrour et al., 2008); MYCN (Murphy et al., 2011); methylated DNA [MBD] (Nan et al., 1993); AT-rich DNA [AT-hook 1/2] (Baker et al., 2013; Lyst, Connelly et al., 2016); DNA ['basic patch'] (Heckman et al., 2014); ATRX (Nan et al., 2007); HP1γ (Agarwal et al., 2007); SMC1/3 (Kernohan et al., 2010); Lamin B (Guarda et al., 2009); DNMT1 (Kimura et al., 2003); TET1 (Cartron et al., 2013); G9a (Subbanna et al., 2014); HLCS (Xue et al., 2013); YB-1 (Young et al., 2005; Gonzales et al., 2012); FBP11 and HYPC (Buschdorf and Strätling, 2004); PRPF3 (Long et al., 2011); LEDGF, DHX9, TDP-43 and FUS (Li et al., 2016); RNA (Jefferey and Nakielny, 2004);miRNA processing complex components – directly with DGCR8 (Cheng et al., 2014a); FOXG1 (Dastidar et al., 2012); HIPK2 (Bracaglia et al., 2009); CDKL5 (Mari et al., 2005); IKKα (Knoshman et al., 2012); p300 and SIRT1 (Zocchi and Sassone-Corsi, 2012); PIAS1 (Tai et al., 2016); HMGB1 (Dintilhac et al., 2002); Huntingtin (McFarland et al., 2014); MBD2 and MeCP2 (Becker et al., 2013).

1.5.1 MeCP2 as a transcriptional repressor

Methylation of DNA sequences such as repeat elements, imprinted control regions, the inactive X chromosome, and CpG islands has long been associated with transcriptional repression (reviewed by Bird, 2002). More recently, both mCpG and mCpA methylation in gene bodies was shown to negatively correlate with transcription levels (Mellén et al., 2012; Guo et al., 2014). The discovery of MeCP2 as a methylated DNA-binding protein led to the prediction that it acted as a mediator for indirect silencing of methylated regions in the genome (reviewed by Tate and Bird, 1993).

Transcriptional repressive activity of MeCP2 was first demonstrated by Nan et al., (1997), who showed that MeCP2 could preferentially silence a methylated reporter gene over a non-methylated one. This repressive activity required the methyl-CpG binding domain (MBD) to be intact and was affected by mCpG density. To map the region required for recruiting co-repressors, fragments of MeCP2 were fused to the GAL4 DNA-binding domain. Residues 207-310 were sufficient to repress a reporter gene downstream of GAL4 binding sites: a region that became known as the Transcriptional Repression Domain (TRD; Fig. 1.5.1). The involvement of histone deacetylases (HDACs) was indicated by the partial disruption of this repressive activity by the HDAC inhibitor, Trichostatin A (Nan et al., 1998). The discovery that two HDAC-containing co-repressor complexes, SIN3A and NCoR/SMRT, directly interact with the TRD of MeCP2 provides a mechanism for its role in repression (Nan et al., 1998; Kokura et al., 2001; Lyst et al., 2013). Furthermore, mutations in MeCP2 that specifically disrupt these interactions lead to a loss of repressive activity (Yu et al., 2000; Lyst et al., 2013). Several other relevant binding partners have been reported including three co-repressors that interact with the SIN3A or NCoR/SMRT complexes: cSki (Kokura et al., 2001), PU.1 (Suzuki et al., 2003) and CBF1 (Mann et al., 2007). An HDAC-independent function in transcriptional repression has also been proposed (Yu et al., 2000). This may occur via the recruitment of transcription factors such as YY1 (Forlani et al., 2010), BRAHMA (Harikrishnan et al., 2005),

SOX2 (Szulwach et al., 2010) and SP3 (Hwang et al., 2010), all of which have been reported to have repressive activity when bound to MeCP2. Alternative mechanisms include MeCP2 recruitment of histone methyltransferase activity either at histone H3 Arg2 (H3R2) by PRMT6 (Dhawan et al., 2011) or at histone H3 Lys9 (H3K9) by G9a or HCLS (Fuks et al., 2003; Subbanna et al., 2014; Xue et al., 2013). Lastly, MeCP2 has been proposed to inhibit transcription directly by binding TFIIB, preventing its incorporation into the preinitiation complex (Kaludov et al., 2000). Fig. 1.5 shows the domains in MeCP2 required for these interactions, several of which have not been mapped.

The repressive activity of MeCP2 described in many of these studies was characterised using artificial reporter genes. Subsequent studies have used mouse models, patient tissues and cell lines to characterise its repressive function *in vivo*. Global transcriptional effects were predicted as the extremely high levels of MeCP2 in neurons mean that it is able to coat chromatin, tracking mCpG and mCAC sites genome-wide (Skene et al., 2010; Lagger, Connelly, Schweikert, et al., 2017). Additionally, the loss of MeCP2 activity in mouse neurons results in increased total histone acetylation (Shahbazian et al., 2002b; Skene et al., 2010). Two studies showed that MeCP2 is important for silencing transcriptional noise, resulting in increased expression of repetitive elements in the brains of *Mecp2*-null mice. Although these transcripts do not persist (Skene et al., 2010), their transcription leads to increased retrotransposition (Muotri et al., 2010). The mutagenic effect of retrotransposition is no doubt dangerous, but is unlikely to be the cause of RTT pathology as these events are irreversible, whereas phenotypic effects of MeCP2-deficiency can be reversed in mouse models (Guy et al., 2007).

High throughput methodologies, such as microarrays and RNA-sequencing, have been used to attempt to identify MeCP2 target genes. Several studies report large numbers of both up- and down-regulated genes with small changes in gene expression. Some of these genes show reciprocal dysregulation in *Mecp2*-null and overexpression systems (reviewed by Guy et al., 2011). Recently, Sugino et al. (2014) proposed that these weak effects may be due to a ‘dilution effect’ caused by

homogenising brain tissues that are made up of multiple cell types whose gene expression patterns are affected differently. They isolated cells of five subtypes and identified genes that are dysregulated by >1.5 fold in each. Notably, there is little overlap in the dysregulated genes, supporting the hypothesis of 'dilution'. Surprisingly, this study also found that genes that are upregulated in the absence of MeCP2 tend to be longer than average. This 'long gene effect' is robust enough to be detected in multiple datasets obtained from RTT patient samples and *Mecp2*-mutant mouse tissues without cell sorting (Gabel et al., 2015). Additionally, the reverse effect (downregulation of long genes) was observed in the brains of mice overexpressing MeCP2. The extent of long-gene upregulation correlates with symptom progression, suggesting a link to pathogenicity. This is consistent with the fact that a disproportionately high number of long genes have neuronal functions.

In order to distinguish between direct and indirect effects of MeCP2 activity on gene expression, chromatin immunoprecipitation (ChIP) datasets have been used to determine the level of MeCP2 binding at dysregulated genes. Recent analysis shows that MeCP2 is most highly enriched on genes that are upregulated in *Mecp2*-null mice and down-regulated in *Mecp2*-overexpression mice, indicating direct repressive activity of MeCP2 at these loci (Lagger, Connelly, Schweikert, et al., 2017). These dysregulated genes contained higher levels of CA(C) methylation in their gene bodies, providing a link between MeCP2 binding at these methylated sites and transcriptional repression (Gabel et al., 2015; Chen et al., 2015; Lagger, Connelly, Schweikert, et al., 2017). While MeCP2 has many interaction partners associated with gene silencing, strong evidence supports the importance of its role in the recruitment of the NCoR/SMRT co-repressor complex to chromatin. This interaction is abolished by the RTT-causing missense mutation, R306C, in MeCP2 (Lyst et al., 2013; Fig. 1.5.1). Strikingly, analysis of gene expression on the cerebellum of knock-in mice expressing this mutated allele also showed dysregulation of a large number of genes, including the upregulation of long genes. (Gabel et al., 2015).

A recent study by Lyu et al. (2016) may provide a link between transcriptional repression by MeCP2 and reduced cell size, a common phenotype of RTT patients

and models. This study shows that loss of MeCP2 in cultured primary mouse neurons results in indirect downregulation of PTEN (a negative regulator of the AKT signalling pathway that promotes cell growth). This is mediated by a miRNA, miR-137, which targets PTEN. These findings are consistent with an earlier report showing that MeCP2 directly represses miR-137 transcription by recruiting the transcription factor, SOX2, to its promoter (Szulwach et al., 2010). While the AKT/mTOR pathway has been shown to be impaired in the absence of MeCP2 (Ricciardi et al., 2011), further work is needed to determine whether MeCP2-mediated repression of miR-137 actually affects cell size.

Figure 1.5.1: The Transcriptional Repression Domain and NCoR/SMRT Interaction Domain of MeCP2
Schematic diagram of MeCP2 protein (e2) annotated with the Methyl-CpG binding domain (MBD, residues 78-162); the Transcriptional repression domain (TRD, residues 207-310); and the NCoR/SMRT Interaction Domain (NID, residues 285-309). The RTT-causing missense mutation, R306C, abolishes the interaction with the NCoR/SMRT co-repressor complex. *References: Nan et al., 1993; Nan et al., 1998; Lyst et al., 2013.*

1.5.2 MeCP2 as a transcriptional activator

As mentioned above, genome-wide studies have found large numbers of genes that are dysregulated in the absence of MeCP2 (reviewed by Guy et al., 2011). Contrary to a repressive function of MeCP2, these studies identified a greater number of downregulated than upregulated genes in tissues from RTT patients and *Mecp2*-mutant mice. Furthermore, quantification of the levels of RNA per cell in human ESC-derived neurons and specific mouse brain regions, shows that loss of MeCP2 activity results in a total reduction in both ribosomal and messenger RNA (Li et al., 2013; Lagger, Connelly, Schweikert, et al., 2017). Improved normalisation techniques used to analyse RNA-sequencing data that account for cell number instead of total amount of RNA mean that the percentage of dysregulated genes that

are upregulated in the absence of MeCP2 is likely to be even higher than originally predicted. The most recent estimate is ~400 upregulated genes and ~8000 downregulated genes in mouse hypothalamus (Lagger, Connelly, Schweikert, et al., 2017). Overall, these genome-wide transcription changes would predict that MeCP2 can also function as a global transcriptional activator.

MeCP2 has been reported to directly interact with two transcriptional activators (Fig S2): CREB1 (Chahrour et al., 2008) and MYCN (Murphy et al., 2011). Whereas MYCN only reduces the repressive activity of MeCP2 when they co-localise at genomic sites, CREB1 is thought to provide a mechanism for MeCP2-mediated gene activation. Sequential ChIP of MeCP2 and CREB1 showed that these proteins preferentially co-occupy the promoters of MeCP2 target genes: those which are down-regulated in the hypothalamus of *Mecp2*-null mice and upregulated in *Mecp2* overexpression mice.

In contrast to the high percentage of reciprocally dysregulated genes reported (Chahrour et al., 2008; Chen et al., 2015), the global effect on RNA level was not reciprocal. Overexpression of a transcriptional activator would predict an increase in total RNA, but levels were not significantly different from wild-type controls (Lagger, Connelly, Schweikert, et al., 2017). It is therefore possible that transcriptional changes could be a result of an upper limit on total transcription, complicating their interpretation.

It is likely that there is a strong connection between decreased RNA levels and the reduced cell size observed in RTT patients and models (Bauman et al., 1995; Chen et al., 2001; Yazdani et al., 2012), but it is unclear whether decreased transcription results in reduced cell size or vice versa. Li et al. (2013) reported that *Mecp2*-null cells have reduced transcription of both ribosomal RNA and transcripts encoding ribosomal proteins resulting in decreased nascent protein synthesis, which is required for cell growth. Alternatively, the transcriptional effect may be secondary: the smaller size of sub-optimal RTT neurons could reduce the transcriptional potential of these cells (Lagger, Connelly, Schweikert, et al., 2017).

Regardless of whether transcriptional changes are a direct or indirect result of the absence of MeCP2 activity, they are likely to contribute to RTT pathology. Deletion of *Mecp2* in adult mice led to shrinkage of the animals' brains and neuronal nuclei with symptom progression (Du et al., 2016; see section 1.4.5). Downregulation of individual genes can also have an impact on the RTT phenotype. Several relevant genes have been identified, including *A2BP1* (encoding ataxin 2 binding protein 1) and *GAMT* (encoding guanidinoacetate methyltransferase), both of which are reduced in patients with intellectual disability and seizures (Chahrour et al., 2008). The most well studied phenotypically-relevant downregulated gene is brain-derived neurotrophic factor (*BDNF*), a growth factor essential for neuronal differentiation and survival. Chang et al. (2006) show that neuronal overexpression of *BDNF* from a human transgene in *Mecp2*-null mice improves their lifespan and motor function. Drugs that increase *BDNF* levels are now being tested as therapies for RTT (reviewed by Katz et al., 2016). *BDNF* partially rescued the reduced brain size phenotype in *Mecp2*-null mice (Chang et al., 2006) and the reduced cell size phenotype in *MECP2*-null cultured human neurons (Li et al., 2013). In the cultured *MECP2*-null neurons, ectopic *BDNF* expression resulted in the upregulation of the AKT/mTOR pathway, providing a mechanism for cell growth (Li et al., 2013). Interestingly, the rescue of cell size was not combined with a rescue of total RNA levels (Li et al., 2013), arguing against the hypothesis that cell size imposes a limit on total transcription.

A2bp1, Gamt and *Bdnf* are all reciprocally dysregulated in *Mecp2*-null and overexpression models (Chahrour et al., 2008), but their contribution to the *MECP2* duplication syndrome pathology has not been investigated. Two further relevant genes, *Crh* (encoding corticotrophin-releasing hormone) and *Oprm1* (encoding G protein-coupled μ-opioid receptor) are down-regulated in *Mecp2*-null mice and upregulated in *Mecp2* overexpression mice. Reducing the expression of either gene by heterozygous deletion partially rescues the *Mecp2* overexpression phenotype in mice: reduction of *Crh* alleviated their anxiety phenotype, and reduction of *Oprm1* improved social behaviour (Samaco et al., 2012).

1.5.3 MeCP2 as a chromatin remodeller

Several studies have described a role of MeCP2 in remodelling chromatin structure. Its ability to compact unmethylated nucleosomal arrays (NAs) *in vitro* was demonstrated by Georgel et al. (2003). Later, this activity was confirmed using methylated NAs (Nikitina et al., 2007; Baker et al., 2013). All three papers show that the C-terminus of MeCP2 is important for this function, as mutants truncated after the methyl-CpG binding domain (MBD) have little or no activity in these assays. In addition to the MBD, MeCP2 has three other short basic regions that have been shown to interact with non-methylated DNA: two AT hooks and a 'basic patch' (Baker et al., 2013; Heckman et al., 2014; Lyst, Connelly et al., 2016; Fig. 1.5.3; Fig. 1.5). AT hook 1 (residues 184-195) was first noted in the initial study describing MeCP2 (Lewis et al., 1992), but its ability to preferentially bind AT-rich DNA was only demonstrated recently (Lyst, Connelly et al., 2016). Baker et al. (2013) identified a second AT hook (residues 265-272), which is also able to bind AT-rich DNA, albeit more weakly (Lyst, Connelly et al., 2016). AT hook domains are best characterised in the HMGA protein family, which each have two or three of these domains. Sequence alignment of MeCP2 with HMGA1 identified a third basic region with high homology to the third AT hook in HMGA1. However, it contains an insertion that destroys the consensus motif so is unlikely to have AT-hook activity (Baker et al., 2013). The highly conserved C-terminus of MeCP2 contains an additional DNA-binding domain, referred to as the 'basic patch' consisting of residues 274-340 (Heckman et al., 2014). Truncated proteins that are missing one (G273X) or two (R270X) of these DNA-binding domains (Fig. 1.5.3) have a reduced ability to condense methylated NAs *in vitro*. This impairment is worse in the shorter truncation (R270X), indicating a cumulative effect of these DNA-binding domains required for full activity (Baker et al., 2013).

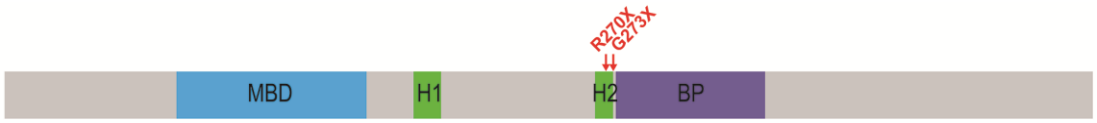


Figure 1.5.3: The DNA binding domains of MeCP2
Schematic diagram of MeCP2 protein (e2) annotated with the Methyl-CpG binding domain (MBD, residues 78-162); AT-hook 1 (H1, residues 184-195); AT-hook 2 (H2, residues 265-272); and the 'basic patch' (BP, residues 274-340). The first nonsense mutation, R270X, removes both AT-hook 2 and the 'basic patch'. The second nonsense mutation, G273X, removed the 'basic patch' only. *References: Nan et al., 1993; Baker et al., 2013; Heckman et al., 2014; Lyst, Connelly et al., 2016).*

Although MeCP2 alone is sufficient to condense nucleosomal arrays, *in vivo* this activity may also involve its interaction partners. MeCP2 has been shown to bind directly to several chromatin remodelling and chromatin modifying proteins (Fig. 1.5), including ATRX (Nan et al., 2007; Kernohan et al., 2010), HP1γ (Agarwal et al., 2007), SMC1/3 (Kernohan et al., 2010), DNMT1 (Kimura et al., 2003), TET1 (Cartron et al., 2013) and Lamin B (Guarda et al., 2009). Methylation of H3K9 by G9a and HLCS could affect global chromatin structure as well as repress transcription locally (Fuks et al., 2003; Subbanna et al., 2014; Xue et al., 2013). The HDAC-containing repressor complexes that interact with MeCP2 (described in section 1.5.1) can also alter chromatin structure by the removal of acetyl groups from histone tails.

The role of the SWI/SNF2 ATPase, ATRX (α-thalassemia/mental retardation, X-linked), in mouse models of RTT has been the most widely studied. In wild-type mice, ATRX co-localises with MeCP2 at heterochromatic foci in neurons; but this localisation is reduced in *Mecp2*-null mice. This delocalisation of ATRX does not occur in peripheral tissues (such as kidney and liver) or in the brains of new-born mice (Nan et al., 2007; Baker et al., 2013). This indicates that the neuronal-specific MeCP2-dependent mechanism for proper ATRX localisation becomes dominant upon neuronal maturation as MeCP2 levels increase, with alternative mechanisms of localisation present in peripheral tissues and developing brains (Nan et al., 2007;

Baker et al., 2013). Indeed, overexpression of ATRX fragments in cultured mouse fibroblasts shows that its N-terminus is sufficient for localisation to heterochromatic foci. However, overexpression of its C-terminal half, which interacts with residues 108-169 of MeCP2, requires MeCP2 to recruit it to heterochromatin (Nan et al., 2007). Surprisingly, mice expressing the two truncated alleles mentioned above, R270X and G273X, both show delocalisation of ATRX, despite retaining the ability to bind ATRX. Interestingly, the loss of ATRX localisation in the neurons of these mice follows symptom progression: occurring earlier in the more severe R270X truncation. The authors of this study propose a model where MeCP2 is recruited to chromatin via its MBD and manipulates the chromatin structure via multiple DNA interactions, promoting a chromatin environment that retains ATRX. The progressively weak remodelling activity with the loss of one or two DNA-binding domains affects the rate of chromatin de-condensation, leading to ATRX loss, and to disease progression (Baker et al., 2013). A further link between ATRX and the pathology of *MECP2*-related disorders is the discovery that the point mutation in the MBD, A140V, specifically reduces the affinity of MeCP2 for ATRX without affecting DNA binding (Nan et al. 2007). This mutation causes intellectual disability in male patients (Neul, 2012).

A role of MeCP2 in chromatin compacting activity *in vivo* would predict increased chromatin density upon its upregulation. This was demonstrated when MeCP2 was ectopically overexpressed in mouse myoblasts: chromocentres became larger and fewer in number due to clustering (Brero et al., 2005). This activity was weakened by RTT-causing missense mutations that disrupt MBD function, showing a requirement for this domain (Agarwal et al., 2011). The authors of these studies propose that upregulation of MeCP2 during myogenesis is responsible for the clustering of pericentromeric heterochromatin that occurs during differentiation (Brero et al., 2005), and that this process is furthered by the recruitment of HP1γ, resulting in increased levels of H3K9me3 (Agarwal et al., 2007). Recently, Linhoff et al. (2015) developed a 3D imaging technique called ChromATin to analyse these effects in mouse brains. Surprisingly, they show that *Mecp2*-null neurons in heterozygous females have more densely packed chromatin than wild-type neurons in the same

tissue. The magnitude of the change varied between neuronal subtypes: the smallest change was in cerebellar granule cells, which express the lowest levels of MeCP2. A possible explanation for the reverse effects observed in myoblasts and neurons is that these different cell types contain different levels of other chromatin architectural proteins, such as the linker histone H1. Normally, neurons have one molecule of histone H1 every two nucleosomes, half the amount than in other cell types (Pearson et al., 1984). Strikingly, in *Mecp2*-null neurons, histone H1 levels are doubled (Skene et al., 2010). MeCP2 has been shown to preferentially displace histone H1 (Nan et al., 1997; Ghosh et al., 2010), providing a potential mechanism for reduced histone H1 levels in wild-type neurons. Further work is needed to confirm whether mutants of MeCP2 that contain an intact MBD (e.g. R270X and G273X) retain the ability to displace histone H1, keeping its levels down. If so, it will be important to analyse how their presence on chromatin affects its structure given that they lack one or two of the additional DNA binding domains.

MeCP2 has also been implicated in chromatin-looping at imprinted gene loci, resulting in transcriptional repression of these genes. This was first described at the *Dlx5*/*Dlx6* locus by Horike et al. (2005). In wild-type mouse brains, Chromosome Conformation Capture (3C) was used to detect two different loop structures: one associated with the active maternal allele and one with the silent paternal allele. In *Mecp2*-null brains, *Dlx5* and *Dlx6* are both upregulated due to loss of repression of the paternal alleles. This is combined with increased histone acetylation and reduced H3K9 dimethylation at this locus. Furthermore, only the loop structure associated with active chromatin could be detected. The authors propose that MeCP2 recruits histone modifying enzymes such as HDAC-containing complexes and H3K9 methyltransferases to this locus to enable the formation of chromatin loops that maintain silencing. This is consistent with a loss of HDAC1 recruitment to *Dlx5*/*Dlx6* in *Mecp2*-null brains. Although MeCP2 has also been reported to interact with the maintenance DNA methyltransferase, DNMT1, no changes in DNA methylation were observed at this locus. Biallelic expression of *DLX5* was also detected in lymphoblastoid cell lines (LCLs) derived from RTT patients. Two of these patients have missense mutations in the MBD (T158M and G161V) indicating

the importance of this domain for the function of MeCP2 in regulating chromatin loops. Two more recent studies by Kernohan and colleagues (2010 and 2014) describe a more direct mechanism for MeCP2 regulation of chromatin looping at imprinted genes. They demonstrate that MeCP2 interacts with two subunits of the cohesin ring complex (SMC1 and SMC3) as well as ATRX in mouse brains (Kernohan et al., 2010) and that MeCP2 is required for ATRX and CTCF recruitment at the *H19* locus. Together, these proteins regulate proper long-range chromatin interactions at this locus, which are disrupted in the absence of MeCP2, resulting in upregulation of gene expression (Kernohan et al., 2014). Surprisingly, these proteins were found to preferentially bind the unmethylated maternal allele of H19, enabling its silencing in postnatal brains (Kernohan et al., 2010). Further work is needed to determine whether MeCP2 binding at this locus is methylation independent. This mechanism provides an additional link between MeCP2 and ARTX function, further connecting the two forms of intellectual disability caused by their mutation.

It is possible that loss of imprinting at the *Dlx5*/*Dlx6* locus could contribute to the RTT phenotype as *Dlx5* is important for the function of GABAergic neurons. Conversely, loss of imprinting was only detected in the LCLs derived from three out of four RTT patients tested: the fourth has the common C-terminal nonsense mutation, R294X (Horike et al., 2005).

1.5.4 MeCP2 as a regulator of alternative splicing

MeCP2 has been reported to directly interact with a number of splicing factors (Fig. 1.5), including FBP11, HYPC (Buschdorf and Strätling, 2004), YB-1 (Young et al., 2005), PRPF3 (Long et al., 2011), LEDGF, DHX9, TDP-43 and FUS (Li et al., 2016). Aberrantly spliced mRNAs have been detected in RNA-sequencing datasets obtained from three different regions of the brains of *Mecp2*-null mice. A range of different aberrant splicing events occurred in these transcripts, including skipped exons and retained introns. Interestingly, expression levels were unaffected in most of the genes that showed splicing changes (Li et al., 2016).

The use of *Mecp2*-null animals for this study gives no indication about which of the splicing factors that interact with MeCP2 might be responsible for these changes. The binding sites for three of these splicing factors have been mapped to the C-terminus of MeCP2: the related proteins, FBP11 and HYPC, require residues 311-486 (Buschdorf and Strätling, 2004) and YB-1 requires residues 195-329 (Young et al., 2005). Altered splicing events were also detected in knock-in mice expressing a C-terminally truncated allele of MeCP2 (lacking residues 309-CT), highlighting the involvement of these three proteins. The presence of YB-1 binding sites (ACE elements) near alternatively spliced exons in the top 10 aberrantly spliced genes indicates an importance for YB-1. The deletion of an ACE element from a *CD44* splicing reporter minigene abolished MeCP2-mediated inclusion of two alternative exons, providing further evidence for a role of YB-1 binding in MeCP2-mediated alterative splicing (Young et al., 2005). Another splicing factor, LEDGF, is thought to be for required MeCP2-mediated splicing of the *Gria2* gene, as knockdown of MeCP2 and LEDGF had the same effect on the inclusion ratio of two mutually exclusive exons in a *Gria2* reporter minigene (Li et al., 2016).

Maunakea et al. (2013) demonstrated a link between high DNA methylation levels and inclusion of alternatively spliced exons in two human cell lines: lung fibroblasts (IMR90) and colorectal cancer cells (HCT116). Data from these cell lines shows that DNA methylation and MeCP2 binding are more highly enriched at included alternatively spliced exons than excluded exons. While DNA methylation depletion and MeCP2 knockdown both result in both aberrant exclusion and inclusion of exons (as seen in *Mecp2*-null mice, Li et al., 2016), the aberrantly spliced exons that normally have the highest levels of both DNA methylation and MeCP2 binding tended to be excluded when these were reduced. Overall, this data indicates that MeCP2 binding to highly methylated exons promotes their inclusion in transcripts. This study does not test whether any of the splicing factors shown to interact with MeCP2 are involved in these DNA methylation-dependent splicing events. Instead, they show that this activity is HDAC-dependent, with strong overlap between aberrant splicing events in MeCP2-depleted and Trichostatin A-treated cells. Their data is consistent with a kinetic model of splicing regulation, whereby MeCP2

recruits HDAC-containing complexes to highly methylated exons, resulting in slower RNA pol II elongation and increased exon inclusion (Maunakea et al., 2013).

Li et al. (2016) show that MeCP2 is also enriched over exons and exon/intron boundaries in the mouse cortex. As yet, there is no evidence so far to determine whether this regulates MeCP2-mediated splicing in the brain: via its interactions either with splicing factors or with HDAC-containing complexes. Surprisingly, a RTT-causing mutation that destroys MBD function (R106W) has no effect on the ability of MeCP2 to mediate exon inclusion of the *CD44* minigene reporter in transfected HeLa cells (Young et al., 2005). Instead this activity requires the RNA-binding ability of MeCP2 as its interaction with YB-1 is RNA-dependent (Jeffery and Nakielny, 2004; Young et al., 2005). Conversely, the interactions between MeCP2 and several other splicing factors, PRPF3, LEDGF, DHX9, TDP-43 and FUS, have been shown to be RNA-independent (Long et al., 2011; Li et al., 2016), so DNA binding may play a role in recruiting these MeCP2-splicing factor complexes to their target sites.

Whatever the mechanism for MeCP2 regulation of alternative splicing, the mis-spliced transcripts are likely to have phenotypic consequences. Alternative splicing is thought to be particularly important in the central nervous system since almost all neurotransmitter receptors and ion channels have multiple alternatively spliced isoforms (reviewed by O'Donovan and Darnell, 2001). Potentially relevant affected genes include *Gria1/2/3/4*, components of the AMPA receptor (Li et al., 2016), and *Dlx5*, which is involved in the function of GABAergic neurons (Young et al., 2005). Unlike *Dlx5*, which has elevated expression in *Mecp2*-null mice due to altered chromatin looping (Horike et al., 2005; see section 1.5.3), the expression levels of *Gria1/2/3/4* are unaffected in *Mecp2*-null mice (Li et al., 2016). Therefore, any phenotypic consequences caused by these genes will be solely a result of aberrant splicing.

1.5.5 MeCP2 as a regulator of miRNA processing

A recent study has described a role of MeCP2 in promoting miRNA processing (Tsujimura et al., 2015). MeCP2 was shown to interact with core components of the nuclear miRNA processing complex: DGCR8, Drosha and DDX5. RNA-sequencing of miRNAs identified 60 and 75 miRNAs that were downregulated by >1.5 fold in *Mecp2 KO* neurons and neural stem cells (NSCs), respectively. This study goes on to focus on miR-199a, the only one of the nine miRNAs downregulated in both cell types that affects cell size. They propose a link between MeCP2-mediated processing of miR-199a and cell growth: miR-199a targets three inhibitors of the mTOR signalling pathway, SIRT1, HIF1a and PDE4D. The model is supported by the characterisation of a *miR-199a KO* mouse, which resembles many of the features of *Mecp2*-null animals, including the development of hypoactivity, gait abnormalities, trembling and irregular breathing after a symptom-free period of 3 weeks. These symptoms lead to premature death from six weeks of age. These mice also have reduced body weight and brain size, consistent with a defect in cell growth. Surprisingly, a separate study published the previous year described a suppressive role of MeCP2 in regulating miRNA processing (Cheng et al., 2014a). Using the same threshold of >1.5 fold, they identified 314 (out of 720) miRNAs that are upregulated and only 63 that are downregulated in the hippocampus of *Mecp2*-null mice. Of the 314 upregulated miRNAs, 106 were downregulated in cortical neurons overexpressing MeCP2. While this study reports an interaction between the miRNA processing component, DGCR8, and the C-terminus of MeCP2 (Fig. 1.5), it does not propose that MeCP2 functions as a component of this complex. Instead, they show that MeCP2 competes with Drosha for DGCR8 binding, negatively regulating miRNA processing complex formation. Tsujimura et al. (2015) state that they cannot replicate this result, but suggest that the differences observed in the two studies could be a result of different post-translational modifications altering MeCP2 activity in different neuronal subtypes.

Importantly, both studies show that MeCP2 affects the processing and not the transcription of its target miRNAs as the levels of the primary miRNA transcripts are

unaffected. Altered levels of both the precursor and mature miRNAs confirm that MeCP2 is involved in the primary (Drosha-mediated) nuclear processing step rather than the secondary (DICER-mediated) cytoplasmic processing step, consistent with its nuclear localisation (Cheng et al., 2014a; Tsujimura et al., 2015).

1.5.6 MeCP2 as a regulator of proliferation and apoptosis

Loss of MeCP2 has been shown to inhibit growth in a range of normal and cancerous cell lines in culture (Balmer et al., 2002; Nagia et al., 2005; Babbio et al., 2012; Bergo et al., 2015; Yang et al., 2016a). Bergo et al. (2015) suggest that MeCP2 is involved in organisation of the mitotic spindle since depletion of MeCP2 in fibroblasts (MRC5 cells) results in a large percentage of cells with abnormal spindle morphology, mostly monopolar. They propose that a subpopulation of MeCP2 (<1% of molecules) is responsible for this activity. This subpopulation is phosphorylated on Tyr120 and localises at the centromeres throughout the cell cycle. This study was prompted by the report of an atypical RTT patient with a Y120D mutation, which may mimic the phosphorylated state. Overexpression of the wild-type but not this mutated protein rescued the proliferation defect in *Mecp2*-null MEFs, disproving the hypothesis that it is constitutively active for this function. The importance of Tyr120 phosphorylation for cell proliferation was not tested with the phospho-abolishing mutation, Y120F. Yang et al. (2016a) propose an alternative explanation for the proliferation defect in MeCP2 deficient cells whereby MeCP2 represses H19, leading to elevated levels of Insulin-like growth factor (IGF-1) in hepatic stellate cells.

A proliferation defect due to MeCP2 depletion has also been reported in cultured non-neuronal cell types derived from embryonic mouse brains, suggesting that reduced proliferation of glial cells may contribute to microcephaly in RTT patients and mouse models (Nagai et al., 2005). Conversely, it is unlikely that this proliferation defect occurs *in vivo* as RTT patients have no weight or head circumference defects at birth (Hagberg et al., 1985). Furthermore, analysis of X chromosome inactivation (XCI) ratios in RTT patients' peripheral blood cells has

shown that the majority of patients do not have skewed ratios of wild-type to mutant cells (http://mecp2.chw.edu.au).

Ectopic overexpression of MeCP2 in cancer cells promotes proliferation (Bernard et al., 2006) and increased levels of MeCP2 have been reported in gastric tumours (Wada et al., 2010; Tong et al., 2016). Conversely, ectopic overexpression of MeCP2 in normal cell lines results in increased apoptosis (Bracaglia et al., 2009). In *Mecp2*-null MEFs, overexpression of wild-type but not MeCP2-R106W resulted in increased cell death, suggesting that this role of MeCP2 is dependent on its ability to bind DNA. This apoptotic activity was enhanced by co-overexpression with homeodomain-interacting protein kinase 2 (HIPK2), which can interact with MeCP2 and phosphorylate it on Ser80 (Fig. 1.5; Bracaglia et al., 2009) increasing its affinity for DNA (Tao et al., 2009). The role of Ser80 phosphorylation is supported by the fact that the phospho-defective mutant, S80A, is unable to promote cell death, and the phospho-mimetic mutant, S80E, functions independently of HIPK2 overexpression. This effect was also seen in P19 (mouse embryonic teratocarcinoma) cells differentiated into neurons (Bracaglia et al., 2009) and cerebellar granule cells (Dastidar et al., 2012). This latter paper shows that this effect is specific to the shorter and less abundant e2 isoform; and that e2-mediated apoptosis is normally prevented in healthy neurons by a direct interaction with FOXG1 (the protein that is mutated in congenital variant RTT; Fig. 1.5). Williams et al. (2016) recently used a Drosophila-based system to show that overexpression of human MeCP2e2 with C-terminal truncations (residues 431-CT or more), as well as the phospho-mimetic mutant S80E, have a stronger effect of neurotoxicity than the full-length protein. Together, these studies identify two inhibitory processes at the extreme N- and C-termini of e2 but shed no light on why e2 but not e1 promotes apoptosis. To my knowledge, this is the only isoform-specific function of MeCP2 that has been reported.

Neurodegeneration has not been described in *Mecp2*-overexpression mouse models but has been reported in patients with *MECP2* duplication syndrome (Reardon et al., 2010). Two groups have characterised mouse models with very high levels of

MeCP2e2: one with a *CMV*-driven human e2 cDNA transgene at 40-140% of wild-type levels replacing endogenous MeCP2 (Kerr et al., 2012) and one with 2-8 fold levels of mouse e2 expressed in neurons from the *Tau* locus (Luikenhuis et al., 2004). Although, in both studies, expression of MeCP2e2 rescues RTT-like symptoms in null males, the apoptotic activity of these proteins may be disrupted by N-terminal fusion with EGFP or the first 31 amino acids of Tau in Kerr et al. (2012) and Luikenhuis et al. (2004) respectively.

Williams et al. (2016) hypothesised that this apoptotic activity of MeCP2e2 might provide a mechanism for the high incidence of skewed XCI in women who have *MECP2* duplicated on one X chromosome (van Esch et al., 2005). Loss of overexpressing cells in these individuals would result in a higher proportion of wild-type cells, providing protection against more severe symptoms (see section.1.3.5). While Sztainberg et al. (2015) show that male mouse models of *Mecp2*-duplication whose MeCP2 levels have been halved by genetic deletion or ASO-mediated knockdown do not display neurological symptoms, it is unclear from the authors' description whether these mice were treated before or after the onset of symptoms. Further work is needed to determine the extent that neurotoxicity contributes to *MECP2* duplication syndrome pathophysiology and how time-dependent a possible cure could be.

1.5.7 Regulation of MeCP2 function by post-translational modification

MeCP2 protein has been reported to be post-translationally modified at multiple sites by phosphorylation, acetylation, ubiquitination, SUMOylation and poly(ADP-ribosyl)ation (Table S2). The discovery of neuronal-activity dependent phosphorylation and de-phosphorylation of specific residues provides a mechanism for the regulation of MeCP2 activities in response to neuronal signalling. Several serine and threonine residues have been found to be phosphorylated in seized mouse brains and/or depolarised cultured neurons (Fig. 1.5.7): Ser86, Thr148/Ser149, Ser164, Ser229, Ser274, Thr308, Ser421 [human residue number 423] and Ser424 [human residue number 426] (Ebert et al., 2013; Tao et al., 2009; Zhou et al., 2006).

The opposite effect was described for Ser80, which is phosphorylated under basal conditions and dephosphorylated upon membrane depolarisation (Tao et al., 2009).

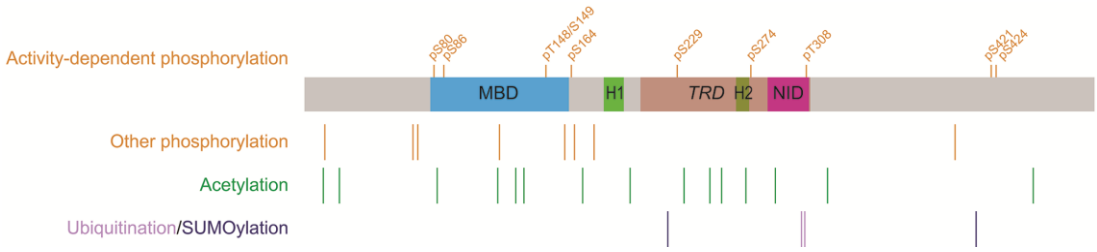


Figure 1.5.7: Post-translational modifications in MeCP2
Schematic diagram of mouse MeCP2 protein (e2) annotated with the activity-dependent phosphorylation sites above and other post-translational modifications below. All activity-dependent residues are phosphorylated upon neuronal activity except Ser80, which is phosphorylated under basal conditions and dephosphorylated upon neuronal activity. All residues numbers are the same between human and mouse except S421 [S423 in human] and S424 [S426 in human]. All modified are conserved between human and mouse except acetylated Lys22, which is Arg in mice. *See Table S2 for more information and references.*

The effect of phosphorylation and de-phosphorylation on the ability of MeCP2 to bind DNA has been widely studied. Two separate knock-in mouse lines with phospho-abolishing alanine mutations at both Ser421 and Ser424 recapitulate some of the features of *Mecp2*-overexpression mice, including improved motor function, heightened context-dependent learning and enhanced LTP (Table S1G; Tao et al., 2009; Li et al., 2011). This phenotype is consistent with the inability to phosphorylate these sites having a gain of function effect. Indeed, the mutated protein shows increased binding at several selected loci, resulting in similar expression changes of these genes to those observed in *Mecp2*-overexpression mice (Li et al., 2011). The first of these phospho-sites, Ser421, has been characterised further. It was shown to be phosphorylated by CamKII kinase in response to membrane depolarisation or seizure induction (Zhou et al., 2006; Tao et al., 2009). Knock-in mice with only the S421A mutation have increased dendritic branching similar to pre-symptomatic overexpression models (Cohen et al., 2011; Jiang et al., 2013), but their mild motor and social behaviour defects are more reminiscent of RTT models (Table S1G). This singularly mutated protein did not have increased

chromatin binding at any of the tested loci (Cohen et al., 2011), suggesting that S424A may have been responsible for this phenotype in the double-mutant mice.

Phosphorylation of Ser80 by HIPK2 kinase has been reported in both cycling cells and neurons (Bracaglia et al., 2009). Dephosphorylation of this site upon neuronal activity is thought to reduce chromatin binding, as demonstrated by the S80A mutant protein at selected gene promoters in cultured cortical neurons. Mice with the same mutation display mild RTT-like symptoms from 2-3 months of age, including impaired motor function (Table S1G; Tao et al., 2009). Although this is consistent with reduced MeCP2 function through weakened DNA binding, the phenotype of these mice is less severe than those carrying RTT-causing MBD mutations (R111G, R133C and T158M/A) that partially or completely abolish DNA binding (Goffin et al., 2012; Heckman et al., 2014; Brown, Selfridge et al., 2016).

A recent paper proposed that phosphorylation of MeCP2 at Ser164 also reduces its affinity for both methylated and unmethylated DNA. This effect is demonstrated using the phospho-mimetic mutant S164D. It remains unclear whether a phosphorylated serine would have the same consequence, but computational modelling using the crystal structure predicts this to be the case (Stefanelli et al., 2016).

Notably, phosphorylation or dephosphorylation upon neuronal activity at all of these sites in MeCP2 was initially proposed to reduce chromatin binding. To support this, depolarisation of cultured neurons with KCl led to MeCP2 release from the *BDNF* promoter (Chen et al., 2003). This model was later revised by a study that performed genome-wide ChIP-seq analysis on untreated and depolarised cultured neurons and failed to identify any regions of the genome with reduced MeCP2 binding (Cohen et al., 2011).

The location of Ser80 at the N-terminal end of the MBD suggested that it could decrease DNA binding directly by steric hindrance. Recently Cheng et al. (2014a) showed that phosphorylation of Ser80 disrupts intramolecular interaction between

the N- and C- termini of MeCP2. Therefore, removal of the phosphate from this residue upon membrane depolarisation promotes protein folding into a less active state. This model is consistent with the requirement of Ser80 phosphorylation for its interactions with YB-1 and DGCR8, when C-terminal fragments of MeCP2 are sufficient for their binding (Gonzales et al., 2012; Cheng et al., 2014a; Young et al., 2005). Similarly, it explains why mutating Ser80 to alanine in the context of an N-terminal fragment of MeCP2 (residues 1-205) displayed no DNA-binding deficit in electrophoretic mobility shift assays (Tao et al., 2009). This protein folding model was proposed as an explanation for the increased apoptotic activity in both S80E and C-terminally truncated mutants, whereby deletion of the extreme C-terminus (residues 431-CT) could destroy intramolecular binding, rendering the protein constitutively active (Williams et al., 2016).

The phosphorylation of Thr308 was shown to disrupt binding to the NCoR/SMRT co-repressor complex (Ebert et al., 2013). This finding fits with its location within the region of MeCP2 required for this interaction: residues 285-309 (Lyst et al., 2013). This finding provides an activity-dependent mechanism of alleviating MeCP2-mediated gene repression without disrupting DNA-binding, which could have additional consequences by altering global chromatin structure. The importance of Thr308 phosphorylation *in vivo* was demonstrated by the presence of mild RTT-like symptoms in T308A knock-in mice. These animals have reduced brain weight, impaired motor function and a lower pentylenetetrazol-induced seizure threshold (Table 1G; Ebert et al., 2013). This phenotype is surprising as this mutation could be predicted to have a gain-of-function effect due to the inability to release the NCoR/SMRT complex from chromatin-bound MeCP2.

Two studies describe altered cellular localisation due to the phosphorylation state of MeCP2. The absence of phosphorylation at Ser30 [residue 13 in e2] was observed in a small proportion of MeCP2e1 molecules that reportedly localised to the nucleoplasm in wild-type mouse brains (Yasui et al., 2014). This study does not go on to identify a specific role for these molecules, nor does it determine whether unphosphorylated e2 is localised similarly. As described in section 1.5.6, a small

subpopulation of MeCP2 is phosphorylated at Tyr120 in cycling cells and has been proposed to have a role in mitotic spindle assembly (Bergo et al., 2015). The relevance of this phospho-site for neuronal function was proposed due to its mutation to a phospho-mimetic state (Y120D) in a patient with variant RTT. In two separate studies, this mutation has been shown to disrupt DNA-binding (Kudo et al., 2003; Agarwal et al., 2011), providing a more likely explanation for this patient's condition. It is yet to be determined whether phosphorylation of Tyr120 also affects DNA binding. Examining the effect of the phospho-abolishing mutation Y120F on mitotic spindle assembly and regulation of DNA binding will help determine the importance of this phospho-site. A phenylalanine at this site has not been reported in patients with *MECP2*-related disorders or in the general population (http://mecp2.chw.edu.au/; http://exac.broadinstitute.org/), but is present in both the Xenopus and zebrafish homologues, indicating that this mutation – and therefore loss of Tyr120 phosphorylation - may not have an adverse effect of MeCP2 function *in vivo*.

Two recent papers report that MeCP2 can be SUMOylated, and that this affects its function. The first, by Cheng et al. (2014b), showed that MeCP2 can be immunoprecipitated from cultured mouse cortical neurons with a SUMO antibody, and that SUMOylation was increased when these cells were depolarised with KCl. They identify Lys223 as a SUMOylated residue, as mutating this residue to arginine (K223R) led to a much lower level of MeCP2 SUMOylation when it is co-overexpressed in HEK293 cells with SUMO1. They go on to characterise the activity of the K223R mutant protein: showing it has reduced HDAC1/2 binding (the catalytic components of the SIN3A co-repressor complex) and reduced ability to repress a GFP reporter gene. Lastly, they show that expression of wild-type MeCP2 but not K223R can rescue the reduced synaptic density phenotype in MeCP2-depleted hippocampal neurons. While these results indicate that Lys223 is important for MeCP2-mediated repression via the SIN3A complex and for neuronal growth, it is unclear whether this is due to the presence of SUMO at this residue. A separate study, by Tai et al., (2016), found that MeCP2 can be SUMOylated by PIAS1 at K412R [human residue 414] using an *in vitro* SUMOylation assay. They showed that

SUMOylation at this site is facilitated by phosphorylation of Thr308 and Ser421, and so is increased by neuronal stimulation. Finally, they showed that SUMOylation of this residue disrupts CREB binding, leading to a model whereby SUMOylation of MeCP2 releases CREB allowing it to activate its target genes, e.g. *Bdnf*. Despite the extensive characterisation of this modification using *in vitro* and cell-based assays, there is no evidence presented in this study that it can occur naturally. It is also worth noting that neither of these lysine residues (Lys223 or Lys412) is located in a consensus SUMO-substrate motif, ψ-K-X-E (where ψ is any hydrophobic amino acid).

A recent study reports poly(ADP-ribosyl)ation of MeCP2 in mouse brains (Becker et al. 2016). This modification was mapped to two regions, residues 163-206 and 244-275, by expressing MeCP2 fragments in cultured cells. Poly(ADP-ribose) polymerase (PARP) 1 was shown to be responsible for this modification, with lower levels of poly(ADP-ribosyl)ation on ectopically expressed MeCP2 in *PARP1-/-* fibroblasts. This study reports a role for poly(ADP-ribosyl)ation of MeCP2 in regulating its chromatin condensing activity. As discussed previously, ectopic expression of wild-type MeCP2 in myoblasts results in clustering of chromocentres (Brero et al., 2005; see section 1.5.3). This activity was increased upon treatment with the PARP1 inhibitor 3AB. This effect was also demonstrated in the *PARP1-/-* fibroblasts, where MeCP2 was shown to have increased localisation and longer residence time on heterochromatin. Overall, these results suggest a negative effect of poly(ADP-ribosyl)ation on the chromatin remodelling activity of MeCP2 (Becker et al., 2016). Further work is needed to determine the impact of poly(ADP-ribosyl)ation on MeCP2-mediated chromatin remodelling in neurons.

Two of the modifying enzymes mentioned above, HIPK2 kinase and PIAS1 SUMO E3 ligase, are reported to directly interact with MeCP2 (Bracaglia et al., 2009; Tai et al., 2006). In addition to these, four modifying enzymes have been proposed to directly interact with and modify MeCP2: CDKL5 kinase, IKKα kinase, p300 acetylase and SIRT1 deacetylase (Fig. 1.5; Mari et al., 2005; Knoshman et al., 2012; Zocchi and Sassone-Corsi, 2012). Inhibition of SIRT1 in a variety of cell lines

resulted in increased acetylation of Lys22, Lys135, Lys171, Lys200, Lys256, Lys271, Lys289 and Lys447 [human residue 449] (Pandely et al., 2015; Zocchi and Sassone-Corsi, 2012). Of these modifications, Lys171 acetylation could be of particular interest as it has been suggested to decrease binding affinity to ATRX and HDAC1 (Pandely et al., 2015). The discovery that CDKL5 kinase can interact with MeCP2 (residues 202-CT) and phosphorylate it links classical and early seizure onset RTT, as the latter is usually caused by mutations in *CDKL5* (Mari et al., 2005).

1.6 Analysis of RTT-causing mutations highlights the key function of MeCP2

1.6.1 The RTT mutation spectrum highlights the importance of the MBD and the NID

A common method for predicting the importance of protein regions or specific residues is their evolutionary conservation. MeCP2 is highly conserved along its entire length, with 95% identity between human and mouse - greater than the average of 86.4% (Makałowski and Boguski, 1998). This suggests that all its domains are vital for its function. Strikingly, analysis of the missense mutations that cause RTT gives a contrasting picture. These are almost entirely located in two clusters: a large cluster in the Methyl-CpG Binding Domain (MBD) and a smaller cluster in the more recently defined NCoR/SMRT Interaction Domain (NID) (Fig. 1.6.1; Nan et al., 1993; Lyst et al., 2013). Analysis of the polymorphisms present in the general population (recorded in the ExAC database) helps to determine regions of the protein that may be dispensable for its function. Evidence for their benign effect is particularly strong in males, who are hemizygous for *MECP2*. Mapping these mutations along the length of MeCP2 shows a reciprocal picture to the RTT-causing mutations (Fig. 1.6.1). Together, this suggests that the role of MeCP2 in recruiting the NCoR/SMRT co-repressor complex to methylated DNA could be its key function *in vivo*. This is referred to in this study as the ‘bridge hypothesis’, with MeCP2 forming a bridge between chromatin and the NCoR/SMRT complex. It was first proposed by Lyst et al. (2013) upon characterisation of the NID where they showed that missense mutations within it abolish binding to components of the NCoR/SMRT complex. As the majority of RTT-causing missense mutations in the MBD disrupt

DNA binding (Table S3), this hypothesis predicts that Rett syndrome occurs if this bridge is broken at either end. While the other proposed roles of MeCP2 may be important for full neuronal function, their importance is likely to be secondary, with their loss resulting in much less severe neurological symptoms.

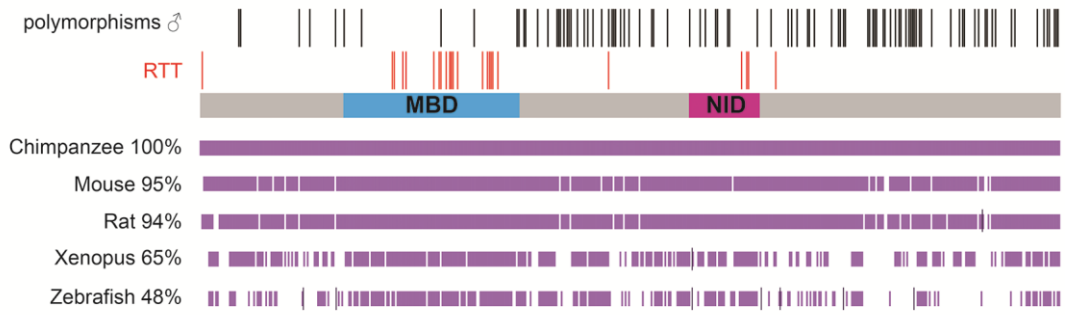


Figure 1.6.1: RTT-causing mutations highlight the importance of the MBD and the NID
Schematic diagram of human MeCP2 protein (e1) annotated with the Methyl-CpG binding domain (MBD, residues 78-162) and NCoR/SMRT interaction domain (NID, residues 285-309). Above: polymorphisms found in hemizygous males in the general population (shown in black) and RTT-causing missense mutations (shown in red). Below: alignments showing identical amino acids conserved between the human protein and homologs in other species (chimpanzee, mouse, rat, Xenopus, zebrafish). *References: Nan et al., 1993; Lyst et al., 2013; http://exac.broadinstitute.org/; http://mecp2.chw.edu.au/ (last updated September 2016).*

1.6.2 The 'unexplained' RTT-causing mutations

While the vast majority of RTT-causing missense mutations fit this bridge hypothesis, three lie outside these regions: A2V (e1 only), P225R, P322L/A (Fig. 1.6.1). Another category of mutations that does not fit the hypothesis is those where a portion of the C-terminus after the end of the NID is missing. Importantly, these C-terminal truncating mutations are the cause of ~10% of cases of RTT. How these mutations cause RTT is the subject of an ongoing study in the lab by Jacky Guy. To date, she has shown that both P225R and P322L are unstable, resulting in low protein levels in neurons derived from knock-in ES cells and knock-in mice. Both mutant proteins also show reduced ability to recruit a component of the NCoR/SMRT

complex (TBL1X) to heterochromatin foci in a cell-based assay. Hemizygous male mice expressing these mutant proteins have undergone phenotypic characterisation, and both resemble other RTT mouse models (Table S1H). Currently, there is also no experimental evidence for the pathogenicity of the e1 missense mutation, A2V, though its location at the beginning of the protein suggests that it could affect translation.

1.7 Aims of this project

The key functional importance of only the Methyl-CpG Binding Domain (MBD) and the NCoR/SMRT Interaction Domain (NID) as indicated by the locations of the RTT-causing missense mutations is in stark contrast to the common view of MeCP2 as a multifunctional protein. The aim of my PhD project was to test the much simpler model that MeCP2 functions as a 'bridge' between methylated DNA and the NCoR/SMRT co-repressor complex. This activity should only require these two domains: MBD and the NID. To determine whether the two domains are sufficient for MeCP2 function, I made a truncated form of the protein consisting solely of the MBD and the NID. This protein lacks three potential dispensable regions of MeCP2: the N-terminus before the MBD, the Intervening Region between the MBD and the NID, and the C-terminus after the NID. To test the functionality of this protein *in vivo*, I produced a mouse line expressing it from the endogenous *Mecp*2 locus. Given that this was a high-risk strategy and mouse production takes a long time, I simultaneously made two other knock-in mouse lines expressing less severe truncations of MeCP2. The first is missing only the N-terminus, and the second is missing both the N- and the C-termini. Together, these proteins form a stepwise series of MeCP2 deletions, enabling any RTT-like symptoms displayed by the knock-in mice expressing the most severe truncation to be attributed to one or more of the deleted regions. This project ran alongside the characterisation of the 'unexplained' RTT-causing mutants by Jacky Guy, which addresses the same hypothesis.

In 2007, Guy et al. showed that the RTT-like symptoms displayed by *Mecp2*-null mice are reversible upon reintroduction of wild-type MeCP2. This was done by alleviating transcriptional silencing of the endogenous *Mecp*2 allele using genetic manipulation techniques. This study showed that MeCP2 is not required during development and is instead essential for the maintenance of proper function in mature neurons. In this study, I set out to determine whether the 'bridge function' of MeCP2 was sufficient to restore neuronal function after the onset of symptoms in MeCP2-deficient mice by activating a gene expressing the most severely deleted protein (consisting of just the MBD and the NID).

Chapter 2 – Materials and Methods

2.1 Materials

2.1.1 Standard Buffers

Name	Recipe
L broth	1% (w/v) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 1% (w/v) NaCl (Fisher Scientific 5/3160/63) (adjusted to pH 7.0 with 5M NaOH)
Terrific broth	1.2% (w/v) Bacto tryptone, 2.4% (w/v) Bacto yeast extract, 2.4% (w/v) glycerol. Add 17 mM KH_2PO_4 and 72 mM K_2HPO_4 before use.
Ampicillin	Stock 50 mg/ml in water (Fisher Scientific T-22943)
Kanamycin	Stock 20 mg/ml in water (Gibco 11815-024)
TE	10 mM Tris HCl (from 1M pH 7.5 stock; Invitrogen 15504-020), 1mM EDTA (VMR 20302.236; from 0.5M pH 8.0 stock)
6x Orange G	12% (w/v) ficoll (Sigma G8790), 1% (w/v) Orange G (Sigma O3756-25G) in TE
1X TAE	40 mM Tris (Invitrogen 15504-020), 1.142% (v/v) glacial acetic acid (Fisher Scientific A/0400/PB17), 1 mM EDTA (VMR 20302.236; from 0.5M pH 8.0 stock)
1X SDS running buffer	24.8 mM Tris (Invitrogen 15504-020), 192 mM glycine (Sigma G8790), 1% (w/v) SDS (Sigma L4509)
1X Transfer buffer	25 mM Tris (Invitrogen 15504-020), 190 mM glycine (Sigma G8790)
TBS (Tris buffered saline)	Dissolve 0.5 M Tris (Invitrogen 15504-020) and pH to 8.0 with HCl (Fisher Scientific H/1200/PB17), add 1.5 M NaCl (Fisher Scientific 5/3160/63).
PBS	1 tablet (Fisher BR14a) per 100ml
NaPi (1 M pH7.2)	1 M Di-sodium hydrogen phosphate (Fisher Scientific S/4440/60), pH to 7.2 using orthophosphoric acid (Fluka 79617)
20X SSC	300 mM Trisodium citrate (Fisher Scientific S/3320/60), 3 M NaCl (Fisher Scientific 5/3160/63), pH to 7.0 with HCl
Church buffer	0.5 M NaPi (from 1 M stock) (pH 7.2), 1mM EDTA (VMR 20302.236), 7% (w/v) SDS (Sigma L4509), 1% (w/v) BSA (Sigma A9418)
PBS-gelatine	1 tablet (Fisher BR14a) per 100ml, 0.1% (w/v) gelatin
IPTG	Stock 100mM in water (Melford MB1008)
Ponceau	0.1% (w/v) ponceau (Sigma P 7767) in 5% acetic acid (Fisher Scientific A/0400/PB17) in water

Table 2.1.1: List of standard buffers

2.1.2 Antibodies

Name	Species (monoclonal/ polyclonal)	Source	Application (dilution)
GFP	Mouse (monoclonal)	CRUK	WB (1:2,000-4,000)
GFP	Rabbit (monoclonal)	NEB (#2956)	WB (1:1,000)
MeCP2 (C-term)	Mouse (monoclonal)	Sigma (M6818)	WB (1:1,000)
HDAC3	Mouse (monoclonal)	Sigma (3E11)	WB (1:1,000)
TBL1XR1	Rabbit (polyclonal)	Bethyl (A300-408A)	WB (1:1,000)
NCoR	Rabbit (polyclonal)	Bethyl (A301-146A)	WB (1:1,000)
H3	Rabbit (polyclonal)	Abcam (ab1791)	WB (1:10,000)
NeuN A60	Mouse (monoclonal)	Millipore (MAB377)	FC (after conjugation to Alexa Fluor 647)
Mouse (800)	Donkey	LI-COR (926-32212)	WB secondary (1:10,000)
Rabbit (800)	Donkey	LI-COR (926-32313)	WB secondary (1:10,000)
Rabbit (680)	Donkey	LI-COR (926-32223)	WB secondary (1:10,000)

Table 2.1.2: List of antibodies
WB = western blot, FC = flow cytometry

2.1.3 Cell lines

NIH3T3 and HeLa cells were used for cell culture assays. The wild type ES cells were 129/Ola E14 TG2a. Culture conditions are described below.

2.1.4 Mice

All mice used in this study were bred and maintained at the University of Edinburgh animal facilities under standard conditions and procedures were carried out by staff licensed by the UK Home Office and according to the Animal and Scientific Procedures Act 1986. Brains were harvested at 6-8 weeks of age (unless stated otherwise). Two existing mouse lines were used in addition to the novel lines produced in this study: WT-EGFP and R306C-EGFP mice (Brown, Selfridge et al. 2016). All mice were backcrossed onto a C57BL/6J background.

2.1.5 Primers

Name	Sequence	Use
WTe2XhoI For	TGTCCTCGAGATGGTAGCTGGGATGTTAGGG	Cloning
WTe2BamHI Rev	CGGTGGATCCTTACAGCTAACTCTCTCGGTCACG	Cloning
pEGFPN1seq For	AAATGTCGTAACAACTCCGC	Sequencing
MBD seq For	AGCTTAAACAAAGGAAGTCTGG	Sequencing
NID seq For	AAGGAGTCTTCCATACGGTC	Sequencing
EGFP seq Rev	GGCTGTTGTAGTTGTACTCC	Sequencing
pEGFPN1seq Rev	ACAAACCACAACTAGAATGCAG	Sequencing
NID seq Rev	GACCGTATGGAAGACTCCTT	Sequencing
pBS seq For1	CAGCCATACCACATTTGTAGAG	Sequencing
pBS seq For2	TCTGTATGTTAAGAAGCCAACC	Sequencing
intron3 seq For	TAGAAGAAAGCTTGTCAGCA	Sequencing
pBS seq Rev2	CCCTACCCATAAGGAGAAGAG	Sequencing
pBS seq Rev1	AGAGTCCCATAGTTTCTCCTG	Sequencing
exon4 For	CAGCATCAGAAGGTGTTCAG	Sequencing
MBD seq Rev	CCAGTTACCGTGAAGTCAAA	Sequencing
CTD seq Rev	ACCATAGGCTGAGTCTTAGC	Sequencing
5LoxP_seq_F	GAATTGGAAAGTCAGGAGGCTGGAAC	Sequencing
NeoSTOP seq F1	GACGTGACAAATGGAAGTAGC	Sequencing
NeoSTOP seq F2	TTTGTCAAGACCGACCTGTC	Sequencing
NeoSTOP seq F3	CTGCTTGCCGAATATCATGG	Sequencing
NeoSTOP seq F4	GGAGGATTGGGAAGACAATAGC	Sequencing
NeoSTOP seq F5	CTTGTTGTTCTTACGGAATACC	Sequencing
NeoSTOP seq F6	GTAATAGAACTCTTGCTTGC	Sequencing
R111G For	TGGACACGAAAGCTTAAACAAGGGAAGTCTGGCC	Mutagenesis
R111G Rev	GGCCAGACTTCCCTTGTTTAAGCTTTCGTGTCCA	Mutagenesis
R306C For	CTCCCGGGTCTTGCACTTCTTGATGGGGA	Mutagenesis
R306C Rev	TCCCCATCAAGAAGTGCAAGACCCGGGAG	Mutagenesis
5’ screen F4	TCACCATAACCAGCCTGCTCGC	5’ PCR screen
5’ screen R5	TGCTAAAGCGCATGCTCCAGACT	5’ PCR screen
5’ screen R3	ATTCGATGACCTCGAGGATCCG	5’ PCR screen
3’ screen F1	AGTTCTGGAATGGTGAGCAAGG	3’ PCR screen
3’ screen F2	GACGTAAACGGCCACAAGTTCA	3’ PCR screen
3’ screen R2	TCAAAAGCAGAGCAGGCAAAAG	3’ PCR screen
Dock5 For	AGTAAAGTAGATTTGTGGGTTCCG	Dock5 seq
Dock5 Rev	GTAGAATGTGGAGATTTCCTCTAGC	Dock5 seq
Gen P5	TGGTAAAGACCCATGTGACCCAAG	Genotyping
Gen P7	GGCTTGCCACATGACAAGAC	Genotyping
Gen Cre For	GACCGTACACCAAAATTTGCCTGC	Genotyping
Gen Cre Rev	TTACGTATATCCTGGCAGCGATC	Genotyping
qPCR MeCP2 For	ACCTTGCCTGAAGGTTGGAC	qPCR
qPCR MeCP2 Rev	GCAATCAATTCTACTTTAGAGCGAAAA	qPCR
qPCR CycA For	TCGAGCTCTGAGCACTGGAG	qPCR
qPCR CycA Rev	CATTATGGCGTGTAAAGTCACCA	qPCR

Table 2.1.5: List of primers

2.2 Methods

The methods are listed in order of their appearance in results (Chapters 3-6).

2.2.1 Cloning

For the cell culture-based assays, cDNA encoding the wild-type and truncated MeCP2 proteins used in this study (ΔN-EGFP, ΔNC-EGFP and ΔNIC-EGFP) was cloned into the pEGFPN1 expression vector (Kanamycin resistant). The shorter isoform (e2) was used for this and all sequences were mouse (producing exactly the same proteins as the knock-in mice). Full length MeCP2e2 was amplified by PCR using primers containing XhoI (5' end) and BamHI (3' end), enabling it to be cloned into the multiple cloning site (MCS) of pEGFPN1 using these enzymes. This results in the desired linker sequence (CKDPPVAT) between MeCP2 and the C-terminal EGFP tag - consistent with the pre-existing WT-EGFP mice (and later the novel ΔN-EGFP mice). cDNA encoding ΔNC-EGFP protein was synthesised (GeneArt, Life Technologies) with a 5' XhoI site and a 3' NotI site. This enabled it to be cloned into the pEGFPN1 vector using these enzymes, with the synthesised EGFP sequence replacing the one originally in the vector. This is required to obtain the desired linker sequence (GSSGSSG) between MeCP2 and the C-terminal EGFP tag. These two plasmids (pEGFPN1_WT-EGFP and pEGFPN1_ΔNC-EGFP) were used to make the other two expression plasmids (pEGFPN1_ΔN-EGFP and pEGFPN1_ΔNIC-EGFP). To make pEGFPN1_ΔN-EGFP, the N-terminal region on the pEGFPN1_WT-EGFP plasmid was replaced with that of pEGFPN1_ΔNC-EGFP using XhoI and BbvCI restriction enzymes. To make pEGFPN1_ΔNIC-EGFP, the Intervening region in the pEGFPN1_ΔNC-EGFP plasmid was replaced by a synthesised fragment (GeneArt, Life Technologies) using HindIII and SacI restriction enzymes. Point mutations (R111G and R306) were inserted into the pEGFPN1_WT-EGFP plasmid using the QuikChange II XL Site-Directed Mutagenesis Kit (ThermoFisher).

For ES cell targeting, genomic sequences encoding exons 3 and 4 of the truncated MeCP2 proteins (ΔN-EGFP, ΔNC-EGFP and ΔNIC-EGFP) were cloned into the

pBS_NeoSTOP_MeCP2 vector (Guy et al., 2007; Ampicillin resistant). This vector contained flanking sequences homologous to the *Mecp2* locus and a Neomycin resistance gene and transcriptional STOP cassette flanked by *loxP* sites in intron 2. Genomic sequence encoding ΔNC-EGFP up to the MfeI sites in the pBS_NeoSTOP_MeCP2 was synthesised (GeneArt, Life Technologies) and cloned into the vector using MfeI restriction digestion. This vector is referred to as pBS_NeoSTOP_ ΔNC-EGFP. This vector was used to make the pBS_NeoSTOP_ΔNIC-EGFP vector by replacing the Intervening region (all encoded in exon 4) with the same synthesised fragment used to make pEGFPN1_ ΔNIC-EGFP, using BstXI and BbvCI restriction enzymes. To make pBS_NeoSTOP_ΔN-EGFP, a fragment with the C-terminal sequence and almost all of EGFP (from the targeting vector used by Brown, Selfridge et al., 2016) was used to replace the deleted C-terminus and EGFP in the pBS_NeoSTOP_ΔNC-EGFP vector using BbvCI and AflII restriction enzymes (after sub-cloning the ΔNC-EGFP genomic sequence into the pEGFPN1 vector using MfeI so all restriction sites used were unique).

All enzymes used were from NEB (used with the recommended buffers): XhoI (R0146S); BamHI-HF (R3136S); NotI-HF (R3189S); BbvCI (R0601S); HindIII-HF (R3104S); SacI-HF (R3156S); MfeI-HF (R3589S); BstXI (R0113S); AflII (R0520S). Digested fragments were run on agarose gels and purified using the QIAGEN gel extraction kit. Fragments were ligated using T4 ligase (M0202S) at a ratio of 3:1 insert to vector for 15 min at room temperature or overnight at 16°C. All plasmid sequences were verified by Sanger sequencing using BigDye (ThermoFisher Scientific). PCR conditions: 96°C 1 min, followed by 25 cycles of: 96°C 1 min, 50°C 5 s, 60°C 4 min. Sequencing was performed by Edinburgh Genomics. Plasmids were purified using QIAGEN Miniprep (Cat #27106) and Maxiprep (Cat #12262) kits.

Targeting vectors were linearised with NotI, and purified by adding an equal volume of PCI (phenol:chloroform:isoamyl alcohol, Sigma 77617-500ML), centrifuged at 4,500 rpm 10 min, 4°C. The (top) aqueous layer was transferred to new tubes and an equal volume of chloroform (Fisher Scientific C/4960/17) was added. Samples were

centrifuged again at 4,500 rpm 10 min, 4°C. The (top) aqueous layer was transferred to new tubes and DNA was precipitated with 0.1 volume 3M NaOAc (Sigma S7899-100ML) and 1 volume isopropanol. Precipitated DNA was pelleted at 13,000 rpm 30 min, 4°C. Pellets were washed in 80% ethanol and resuspended in sterile TE.

2.2.2 Cell culture (HeLa and NIH-3T3)

Both HeLa and NIH-3T3 cells were grown in DMEM (Gibo 41966-029) supplemented with 10% foetal bovine serum (Gibco) and 1% Pen Strep (Gibco 15140-122). Cultures were maintained in T75 or T175 flasks and split regularly using TrypLE Express (Gibco 12604-013) and DPBS (Gibco 14190-094).

2.2.3 Immunoprecipitation and western blotting

HeLa cells were seeded into 15 cm dishes (split 1 in 8 from a T175 flask). The following day, cells were transfected with 30 μg pEGFPN1-MeCP2 plasmids using JetPEI (Polyplus Transfection). Briefly, DNA was vortexed in 0.5 ml 150 mM NaCl and 60 μl of JetPEI was vortexed in a separate 0.5 ml 150mM NaCl, before the two were combined. After incubating for 15-30 min at room temperature, this solution was added drop-wise to cells and left overnight before the media was replaced. 24-48 hours after transfection, cells were harvested by trypsinisation (TrypLE Express) into 15 ml falcon tubes. Cell pellets were washed with DPBS and stored at -80°C.

Cells were thawed and resuspended in 1.5 ml cold NE1 (20mM HEPES pH 7.5; 10 mM NaCl; 1mM $MgCl_2$; 0.1% Triton X-100; 10mM β-mercaptoethanol; protease inhibitor tablets (Roche)). Resuspended pellets were Dounce homogenised and nuclei were pelleted at 3,000 rpm (5 min, 4°C). The supernatant (cytoplasmic fraction) was discarded and nuclei were suspended in 200 μl NE1. Nucleic acids were digested with 1 μl (250 units) Benzonase (Sigma E1014-25KU) for 5 min at room temperature. An equal volume of NE300 (NE1 plus 300 mM NaCl) was added, bringing the NaCl concentration to 150 mM. Samples were incubated under rotation for 20 min at 4°C, and then centrifuged at 13,000 rpm (20 min, 4°C). The

supernatant was used as the nuclear extract, from which 20-30 µl was saved as 'inputs'. The rest of the nuclear extract was incubated with GFP-TRAP_A beads (Chromotek gta-20) that had been pre-washed in NE150 (NE1 plus 150 mM NaCl). Samples were rotated at 4°C for 30-60 min. Beads were centrifuged and washed four times with 1 ml NE150. Proteins were dissociated from the beads by the addition of 2x Laemmli Sample Buffer (Sigma S3401-1VL) and boiled for 5 min at 100°C. Sample buffer was also added to 'input' samples, which were also boiled.

Samples were run on 4-15% SDS gels (BioRad Mini-PROTEAN TGX) at 150-200V, and transferred onto nitrocellulose membranes (BioRad) at 100V for 1 hour. Good quality transfer was verified using Ponceau staining. Membranes were blocked with 4-5% (w/v) milk in TBST (TBS with 0.05% Tween-20), and incubated with primary antibodies added to blocking solution at 4°C overnight. Membranes were then washed three times with PBS+0.05% Tween-20, blocked again for 30 min at room temperature, incubated with LI-COR secondary antibodies in blocking solution for 2-3 hours at room temperature, and washed a further three times with PBS+0.05% Tween-20. Primary antibodies used: GFP (NEB #2956) or GFP (CRUK), NCoR (Bethyl A301-146A), HDAC3 (Sigma 3E11), TBL1XR1 (Bethyl A300-408A). LI-COR secondary antibodies used: Mouse 800 (926-32212) and Rabbit 680 (926-32223). Conditions are listed in section 2.1.2. Membranes were scanned using an Odyssey Infrared Imager with Image Studio Lite Ver 4.0 software.

2.2.4 Microscopy and TBL1X recruitment assay (NIH-3T3)

NIH-3T3 were seeded in 6-well plates containing HCl-treated and sterilised coverslips (~25,000 cells per well). The following day they were transfected with 2 µg pEGFPN1-MeCP2 plasmids using 6 µl JetPEI per well. Transfections were performed as described in section 2.2.3 (using 2x 100 µl 150 mM NaCl). The transfection mixture was left on cells overnight before the media was changed. 48 hours after transfection, cells were washed twice with DPBS and fixed with 4% paraformaldehyde (w/v in PBS, pH 7.4) for 15 min at room temperature. Fixed cells were washed three times with PBS and then incubated with 1 µg/ml DAPI in PBS (in

the dark) for 5 min. Cells were washed a final two times and then the coverslips were mounted onto glass slides using ProLong Diamond (Life Technologies P36961). For the TBL1X recruitment assay, cells were transfected with 1 μg pEGFPN1-MeCP2 plasmids and 1 μg pmCherryN1-TBL1X. The cells were then visualised and photographed using confocal microscopy (Leica SP5). DAPI was visualised with the 405 nm laser (15% power) and detected between 420-470 nm; EGFP with the 488 nm laser (3% power) and detected between 510-550 nm; and mCherry with the 594 nm laser (12% power) and detected between 610-700 nm.

2.2.5 ES cell culture and gene targeting (CRISPR). PCR strategy for screening ES cell clones

ES cells (wild-type line 129/Ola E14 TG2a) were grown in Glasgow MEM (Gibco 21710-025) supplemented with foetal bovine serum (FBS; Gibco- batch tested), 1% Non-essential amino acids (Gibco 11140-035), 1% Sodium Pyruvate (Gibco 11360-039), 0.1% β-mercaptoethanol (Gibco 31350-010) and 1000 units/ml LIF (ESGRO #ESG1107). ES cells were grown in 10 cm dishes coated in PBS-gelatin (coated for at least 20 min before adding media and cells). Cells were routinely split 1/8 using Trypsin (Gibco 15090-046) every two days with media changed on intermediate days. Very low passage cells were used for gene editing. These cells were grown for one week before electroporation with 15 μg linearised targeting vector (and 15 μg pX300 CRISPR plasmid). $20x10^6$ cells were trypsinised, spun down and resuspended in 600ul HBS (1.4 M NaCl, 50 mM KCl, 10 mM Na2HPO4[dihydrate], 55 mM D-glucose, 20 mM HEPES) containing the plasmids and transferred to a 0.4 cm cuvette (BioRad #1652088) for electroporation using a GenePulser X cell machine (240V 500μF ∞ resistance). After 10 min, cells were diluted in growth medium and plated at clonal density: $0.5\text{-}2.5x10^6$ cells per 10 cm dish. The following day, the media was replaced with G418 selection media (normal growth media with 350 μg/ml G418 (Gibco 108321-42-2)). The G418 concentration was reduced to 150 μg/ml after a further 48 hours and kept at this concentration. Clones were grown up until visible by eye (~1 week), before being picked in DPBS using a P20 pipette into wells of a round-bottomed 96-well plate. These colonies were then trypsinised (20μl) to break

up clumps, before (selection) media was added and they were transferred to a gelatinised flat-bottomed 96-well plate. The colonies were grown up with regular media changes until the majority of colonies fill the wells, turning the media yellow overnight. At this point, colonies were given fresh media and then split 1/2 after 3-4 hours for cryopreservation. For this, trypsinised (30 µl trypsin) colonies were 'quenched' with 70 µl Quench medium (1:1 selection medium and FBS). Half of this volume was transferred to a new 96-well plate with an additional 150 µl selection media for continued growth (for genomic DNA purification). Cold 2x Freeze medium (25% DMSO (v/v) in selection medium) was added to the other half, and the plate was then sealed and stored at -80°C.

Colonies were thawed by placing the 96-well plate into a 37°C water bath. Positive clones were then transferred to 24-well plates. These were grown up until they filled two T25 flasks, from which one was cryopreserved (1xFreezing medium: growth medium supplemented with 12% FBS and 10% DMSO) and the other was used for genomic DNA purification.

2.2.6 Genomic DNA purification and Southern blotting (and sequencing Mecp2 and Dock5)

After splitting the colonies 1/2 (see section 2.2.5), the half that was grown on was used for genomic DNA purification. First, these cells were lysed in 100 µl lysis buffer (100 mM Tris pH8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl, 100 µg/ml Proteinase K, 50 µg/ml RNase A (Thermo Scientific #EN0531)) overnight at 37°C. Lysates were transferred to V-bottomed 96-well plates using wide-bore tips. DNA was precipitated with 100 µl isopropanol (Fisher Scientific P/7500/17) and plates were centrifuged at 3,000 rpm, 15 min. Pellets were washed with 75% ethanol and plates were centrifuged at 3,000 rpm, 15 min. Pellets were air-dried and then dissolved in 50-100 µl TE.

Genomic DNA was purified from ES cell pellets harvested from T25 flasks using Puregene Core Kit A (Qiagen 1042601) according to manufacturer's instructions for cultured cells. DNA pellets were dissolved in 400 µl DNA hydration solution.

Colonies were screened by PCR at the 5' end (primer pairs F4/R3 and F4/R5) and at the 3' end (primer pairs F1/R2 and F2/R2). This was done using Phusion polymerase (NEB #M0530L) with the HF buffer. PCR conditions: 98°C (hot start) for 30 s, followed by 35 cycles of: 98°C 10 s, annealing temperature 15 s, 72°C 1 min 45 s; and then 10 min elongation at 72°C. Annealing temperatures: 72°C for F4/R3 and F4/R5, 71°C for F1/F2 and 70°C for F2/F2. PCR reactions were run on 2% agarose gels. Positive clones give products of 3.16 kb for both F4/R3 and F4/R5, and 3.18 kb and 3.11 kb for F2/R2 and F1/R2, respectively.

For Southern blot analysis, 10 µg of genomic DNA was digested with 3 µl restriction enzyme (Bsu36I, BsrGI or BamHI-HF) in a total volume of 40 µl containing the recommended NEB buffer for >3 hours at 37°C. 8 µl 6x Orange G was added and the digested fragments were separated on 0.8% agarose gels run at 36V overnight. The next day, gels were incubated in 0.25M HCl (Fisher Scientific H/1200/PB17) for 15 min and then 0.4 M NaOH (Fisher Scientific S/4880/53) for 45 min. DNA fragments were transferred to ZetaProbe membranes (BioRad) that had been pre-wetted with 0.4 M NaOH. DNA probes homologous to either exon 4 or the end of the 3' homology arm were radioactively labelled using the Promega Prime-a-Gene Labeling System. For each blot, 25 ng probe DNA was boiled at 100°C for 2 min in a total volume of 30 µl water. The boiled probe was rapidly chilled on ice before adding 10 µl 5x reaction buffer, 2 µl dNTPs (dATP, dGTP and dTTP) and 2 µl BSA [all reagents are included in this kit]. Next, 5 µl [α^{32}]dCTP (Perkin Elmer) and 1 µl (5 units) DNA polymerase (Klenow) Fragment were added and the reaction mixture was incubated for 1 hour at room temperature. The reaction was stopped with 20 mM EDTA and the probe was purified using a G50 column (GE Healthcare). The probe was then boiled at 100°C for 5 min, cooled rapidly on ice and added to membranes that had been blocked in 25 ml Church buffer containing 125 µl Herring Sperm DNA (Sigma D7290-1ML) at 65°C for >30 min. Probes were left to hybridise at 65°C

overnight. The next day, membranes were washed three times with 3xSSC/0.2% SDS that had been preheated to 65°C. Radioactivity was monitored during this process. If further washing was needed, 1xSSC/0.2% SDS was used. The blots were left to expose for 1-5 days in Phosphorimager cassettes (GE Healthcare) and then scanned using a Typhoon FLA 7000. Bands were quantified using ImageQuant software.

The *Mecp2* locus (exons 3 and 4) was sequenced in the positive clones. This region was amplified with Phusion polymerase (primers pBS seq For2 and pBS seq Rev2) and the PCR product was then sequenced by Sanger Sequencing. The top-intragenic predicted off-target site (*Dock5*) was also PCR amplified (using DreamTaq, Thermo Scientific) and Sanger Sequenced.

2.2.7 Karyotyping

To determine the karyotype of the ES cell clones, they were grown normally and then treated for 3 hours with 0.1 μg/ml colemid (KARYOMAX, Gibco 1512-012) to cause metaphase arrest. Cells were then trypsinised and pelleted and resuspended in 300 μl of growth media and 5 ml of 0.4% (w/v) KCl and incubated at 37°C for 10 min. Next, 100 μl of freshly prepared fixative (3:1 methanol and acetic acid) was added and mixed gently. Cells were pelleted at 300xg, 5 min and the supernatant was discarded. The cell pellet was resuspended in 5 ml fixative and incubated at room temperature for 20 min. Cells were pelleted again at 300xg, 5 min, and then resuspended in 200-500 μl fixative. This suspension was dropped onto pre-chilled glass slides from a height of ~30 cm to spread out the cells. Slides were left to dry overnight before covering with coverslips using Vectashield containing DAPI (H-1200). Cells were photographed under the microscope and chromosomes were counted manually using ImageJ.

2.2.8 The production of knock-in mouse lines

One positive ES cell clone was used to make each knock-in mice line. These cells were injected into blastocysts (embryonic day 3.4) obtained from C57BL/6J females after natural mating. Typically, 15 ES cells were injected into each blastocyst and 12 injected blastocysts were transferred to pseudo-pregnant recipient females (either C57BL/6J or F1). Blastocyst injections were performed by Jim Selfridge and Martha Koerner. Chimaeric pups were recognised by varied coat colour as the ES cells were 129/Ola (chinchilla) and the injected blastocysts were C57BL/6J (black). The majority of chimaeras were males, consistent with the ES cells being male. Male chimaeras were bred with homozygous *CMV-Cre* (ubiquitous deleter) mice (C57BL/6J). The resulting germline female pups were recognisable by their agouti (brown) coat colour. They were heterozygous for the knocked-in truncated allele, which was expressed due to Cre-mediated recombination (excising the Neomycin resistance gene and STOP cassette). These heterozygous females were bred with wild-type C57BL/6J males to produce the first hemizygous males expressing the truncated proteins. Some pups in this generation had both the truncated *Mecp2* allele and the (X-linked) *CMV-Cre* transgene due to crossing over during meiosis. Cre-negative heterozygous females from this generation were used for further breeding onto the C57BL/6J background. Mice were backcrossed for four generations to a 94% C57BL/6J background for use in phenotypic characterisation.

2.2.9 Genotyping

Mice were genotyped using ear punches. 25 µl 'ear biopsy buffer' (50 mM KCl, 1.5 mM $MgCl_2$, 10mM Tris HCl pH8.5, 0.01% PBS-gelatine, 0.45% NP-40, 0.45% Tween 20, 0.45% Triton X-100, 100 µg/ml Proteinase K) was added to each. These were then heated using the following programme: 65°C 15 min, 96°C 2 min, 65°C 4 min, 96°C 1 min, 65°C 1 min, 96°C 2 30 s. Genotypes were determined by PCR using DreamTaq polymerase. *Mecp2* genotypes were determined using P5/P7 primers, which are located in intron 2 spanning the site from which the NeoSTOP cassette was excised. Excision of this cassette leaves a *loxP* site and polylinker

sequence so mutant animals have a larger PCR product. *Cre* genotypes were determined using primers located in this region so can only produce a PCR product when the transgene is present. Genotyping was performed by Dina De Sousa.

2.2.10 Genomic DNA purification from brains

Brains were harvested and snap-frozen in liquid nitrogen and stored at -80°C. Frozen half-brains were thawed and Dounce homogenised in 3 ml lysis buffer (50 mM Tris HCl pH7.5, 100 mM NaCl, 5mM EDTA). Proteinase K was added to a final concentration of 0.4 mg/ml and SDS to 1%. Samples were incubated overnight at 55°C. The next day, RNase A was added to a final concentration of 0.1 mg/ml and incubated at 37°C for 1-2 hours. An equal volume of PCI (phenol:chloroform: isoamyl alcohol, Sigma 77617-500ML) was added and samples were centrifuged at 3,000 rpm, 15 min. The supernatant (aqueous phase) was transferred to new tubes and DNA was precipitated with 0.1 volume 3M NaOAc and 2.5 volumes of ethanol. Precipitated DNA was extracted with a heat-sealed glass Pasteur pipette, washed in 80% ethanol, and dissolved in 400 µl TE. Genomic DNA was analysed by Southern blotting, as described in section 2.2.6.

2.2.11 Western blot analysis of protein levels in whole brain crude extracts

Frozen half-brains were Dounce homogenised in 750 µl cold NE1 (20 mM HEPES pH7.9, 10 mM KCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 20% glycerol, 0.5 mM DTT or 14.3 mM β-mercaptoethanol (Sigma M6250-250ML), and protease inhibitors (Roche)). Nucleic acids were digested with 3 µl (750 units) Benzonase (Sigma E1014-25KU) for 15 min at room temperature. Next, 750 µl of 2x Laemmli Sample Buffer (Sigma S3401-1VL) was added and samples were boiled for 3 min at 100°C, snap-frozen on dry ice, and boiled again for 5 min. Aliquots were stored at -80°C. Samples were diluted 1/6 with 1x sample buffer before protein levels were analysed by western blotting (method as described in section 2.2.3, except 4-20% gels were used, and transfer onto nitrocellulose membranes was done overnight at 30V). Antibodies used: GFP (NEB #2956), MeCP2 C-terminal (Sigma M6818), H3

(Abcam ab1791) and Rabbit 800 (LI-COR 926-32313). Western blots were quantified using Image Studio Lite Ver 4.0 and genotypes were compared using t-tests. Novel knock-in lines were compared to WT-EGFP controls (Brown, Selfridge et al., 2016).

2.2.12 Flow cytometry analysis of protein levels in whole brain nuclei

Fresh brains were used for Flow Cytometry analysis, harvested from animals aged 12 weeks. WT-EGFP mice (Brown, Selfridge et al., 2016) were used as controls. Brains were Dounce homogenised in 2 ml cold homogenisation buffer (320 mM sucrose, 5 mM $CaCl_2$, 3 mM $Mg(Ac)_2$, 10 mM Tris HCl pH.7.8, 0.1 mM EDTA, 0.1% NP40, 0.1 mM PMSF, 14.3mM β-mercaptoethanol, protease inhibitors (Roche)). A further 3 ml homogenisation buffer was added, followed by 5 ml of 50% OptiPrep gradient centrifugation medium (50% Optiprep (Sigma D1556-250ML), 5 mM $CaCl_2$, 3mM $Mg(Ac)_2$, 10 mM Tris HCl pH7.8, 0.1M PMSF, 14.3mM β-mercaptoethanol). First, 10 ml of 29% OptiPrep solution (diluted in water) was added to Ultra clear Beckman Coulter centrifuge tubes, and then the brain lysate was layered on top. Samples were centrifuged at 7,500 rpm (equivalent to 10,100xg) for 30 min, 4°C. Pelleted nuclei were resuspended in Resuspension buffer (20% glycerol in DPBS with protease inhibitors (Roche)). Aliquots (four per brain) were snap-frozen in liquid nitrogen and stored at -80°C.

For flow cytometry analysis, frozen nuclei were thawed on ice and pelleted at 600xg (5 min, 4°C). Nuclei were resuspended in 1 ml PBTB (5% (w/v) BSA, 0.1% Triton X-100 in DPBS with protease inhibitors (Roche)). Nuclei were pelleted again and resuspended in 250 µl PBTB. To stain for NeuN, 10 µl of NeuN-A60 antibody (Millipore MAB377) was conjugated to Alexa Fluor 647 using the APEX Antibody Labelling Kit (Invitrogen A10475). The final eluate from this reaction, which contained the conjugated antibody, had a volume of ~50 µl. 2ul of this eluate was added to the nuclei and incubated under rotation for 45 min at 4°C. These samples were analysed by flow cytometry using a BD LSRFortessa SORP: 50,000 nuclei (events) for each sample. Nuclei were gated by size (to remove debris and clumps)

and NeuN levels. The mean EGFP expression was obtained for the total nuclei and the high NeuN (neuronal) subpopulation (n>8,000). Fluorescence was measured in arbitrary units (au). Background EGFP fluorescence (94 arbitrary units) was determined by analysing nuclei obtained from an untagged wild-type mouse, and this value was subtracted from the other samples. Genotypes were compared using t-tests.

2.2.13 RNA purification from brains and qPCR

Frozen half brains were homogenised in 3 ml TRI Reagent (Sigma T9424) by Polytron and incubated at room temperature for 10 min. Next, 300 µl 1-Bromo-3-chloropropane (Sigma B9673-200ML) was added and samples were vortexed and then transferred to 15 ml PhaseLock tubes (5Prime). Sampled were then centrifuged at 3,500 rpm for 15 min and the supernatant was transferred into new tubes. RNA was precipitated by adding 1.5 ml isopropanol (incubated for 10 min at room temperature). RNA was then pelleted at 3,500 rpm for 15 min. Pellets were washed with 3 ml 70% ethanol and centrifuged for another 5 min. RNA pellets were then air-dried and dissolved in 300 µl nuclease-free water. RNA samples were stored at -80°C.

DNA was removed from 10 µg of these RNA preps using the DNA-free kit (Ambion). Next, 0.5 µg of this DNA-free RNA was reversed transcribed using the iScript cDNA Synthesis Kit (BioRad). cDNA samples were made up to 200 µl and stored at -20°C. *Mecp2* RNA levels were analysed by qPCR, normalised to *Cyclophilin A*. 2.5 µl of cDNA was added to the qPCR reaction mixture with 1X SYBR+Fluorescein (from 2X stock = Sensimix, Bioline) and 500 nM each of forward and reverse primers in a total volume of 12.5 µl in 384-well plates. Samples were run on a LightCycler 480WS1.5. PCR conditions: 95°C 10 min, followed by 45 cycles of: 95°C 15 s, 60°C 15 s, 72°C 20 s (during which DNA levels were recorded). Melting curves of the final PCR products were obtained with continuous measurement from 60°C to 95°C (ramp 0.11, 5 acquisitions per °C). Standard curves for each primer pair were calculated using serial dilutions of cDNA samples. The gradient of these graphs was used to determine primer efficiency

(http://www.genomics.agilent.com/biocalculators/calcSlopeEfficiency.jsp). Samples were pipetted in triplicate, and the mean Cp values for each sample were used. Primer efficiencies (E) were 1.995 and 1.879 for *Mecp2* and *Cyclophilin A*, respectively (average of all experiments). *Mecp2*/*Cyclophilin A* ratios were calculated for each sample using the formula: E_{cycA}^(Cp_{cycA}) / E_{Mecp2}^(Cp_{Mecp2}). These ratios were averaged between replicate experiments for each animal. Graphs show the mean (and standard deviation) values between three biological replicates, normalised to the wild-type littermates for each line. Genotypes were compared using t-tests.

2.2.14 Immunoprecipitation from brain nuclear extracts

Immunoprecipitations were performed using nuclear extracts from the brains of the knock-in mice. WT-EGFP and R306C-EGFP mice (Brown, Selfridge et al., 2016) were used as controls. Frozen brains were Dounce homogenised in 1 ml NE1 (20 mM HEPES pH7.9, 10 mM KCl, 1 mM $MgCl_2$, 0.1% Triton X-100, 20% glycerol, 14.3 mM β-mercaptoethanol, and protease inhibitors (Roche)). Nuclei were pelleted at 3,000 rpm, 5 min, 4°C and the supernatant (cytoplasmic fraction) was discarded. Nuclei were washed in 500 µl NE1 and pelleted again before being resuspended in 500 µl NE1. Nucleic acids were digested with 1.5 µl (375 units) Benzonase (Sigma E1014-25KU) for 10 min at room temperature. An equal volume of NE300 (NE1 plus 300 mM NaCl) was added bringing the NaCl concentration to 150 mM. Samples were incubated under rotation for 20 min at 4°C, and then centrifuged at 13,000 rpm (20 min, 4°C). The supernatant was used as the nuclear extract and its protein concentration was determined using BioRad Protein Assay Dye Reagent (#500-0006) according to manufacturer's instructions. Samples were adjusted to the same protein concentration with NE150 (NE1 plus 150 mM NaCl) and 30 µl was saved as 'inputs'. Equal volumes of the remaining nuclear extract (using the maximum volume possible) were incubated with GFP-TRAP_A beads (Chromotek gta-20) that had been pre-washed in NE150 (NE1 plus 150 mM NaCl). Samples were rotated at 4°C for 1 hour. Beads were centrifuged and washed four times with 1 ml NE150. Proteins were dissociated from the beads by the addition of 2x Laemmli Sample

Buffer (Sigma S3401-1VL) and boiled for 5 min at 100°C. Sample buffer was also added to 'input' samples, which were also boiled. Samples were analysed by western blotting (described in section 2.2.3). Primary antibodies used: GFP (CRUK), HDAC3 (Sigma 3E11). LI-COR secondary antibody used: Mouse 800 (926-32212). Conditions are listed in section 2.1.2. Membranes were scanned using an Odyssey Infrared Imager with Image Studio Lite Ver 4.0 software.

2.2.15 Scoring of mouse phenotypic severity

Backcrossed (94% C57BL/6J) scoring cohorts each consisted of 10 mutant animals and 10 wild-type littermates. Animals were housed with genotypes intermixed in two large cages per cohort (10 animals per cage). Preliminary outbred (75% C57BL/6J) cohorts were also scored for the ΔNC-EGFP and ΔNIC-EGFP lines. The ΔNC-EGFP outbred cohort consisted of 7 mutants and 9 wild-types, and the ΔNIC-EGFP outbred cohort consisted of 10 mutants and 1 wild-type. (Note: some animals in these outbred cohorts also carried the *CMV-Cre* transgene.) Animals were scored from 4 weeks up to a maximum of a year using a scoring system developed by Guy et al. (2007). Mice were scored in six categories: spontaneous activity, gait, hind-limb clasping, tremor, abnormal breathing and general appearance. Mice receive a score between 0 and 2 for each category, where 0 = as wild-type, 1 = present, and 2 = severe. Intermediate scores of 0.5 and 1.5 were also used in all categories except hind-limb clasping (Cheval, Guy et al., 2012). The scores in each category are added together to give the aggregate symptomatic score for each animal. The mean scores for all animals are plotted over time.

Animals were also weighed during scoring sessions. Animals were culled if they lost more than 20% of their maximum body weight (unless they had been obese). Growth curves were compared by Repeated Measures ANOVA (the animals that died within the experimental period – one wild-type in each ΔNC cohort and one ΔNIC in their cohort – were excluded from this analysis to enable a balanced design).

2.2.16 Behavioural tests

A second backcrossed (94% C57BL/6J) cohort for each mouse line underwent a series of behavioural tests. These cohorts also consist of 10 mutant animals and 10 wild-type littermates. Tests were performed over a two week period when the animals were 20-21 weeks of age.

On day 1, anxiety levels were tested using the Elevated Plus Maze. This is 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). It is set up in a dimly lit room with uniform lighting between each of the closed arms and each of the open arms. Animals were placed in the central area and left to explore for 15 min (during which the experimenter left the room). Mice were tracked using ANYmaze software (Stoelting). Mice were determined to be in a particular area of the maze if ≥70% of their body was inside this region. Several parameters for this test were analysed: time spent in areas, total distance travelled, distance ratios (each area: total distance), and number of arm entries. Genotypes were compared using t-tests if both datasets fitted a normal distribution (determined using D'Agostino-Pearson omnibus and Shapiro-Wilk normality tests). Otherwise, Kolmogorov-Smirnov tests were used.

On day 2, activity and anxiety levels were tested using the Open Field test. The apparatus used consists of a square arena measuring 50 by 50 cm, which is evenly lit (in a dimly lit room) and littered with fresh wood chippings. The animals were placed in the centre of the area and left to explore for 20 minutes each (during which the experimenter left the room). Mice were tracked using ANYmaze software (Stoelting), which divided the area into nine squares of equal size. Mice were determined to be in a particular area of the maze if ≥70% of their body was inside this region. Activity was assessed by total distance travelled and time immobile. An immobile episode is defined as lack of movement for a minimum of 10 seconds. Anxiety was assessed by the time spent in the central square time or Centre:Total ratio of distance travelled. Data was analysed for the full test period and also after division into two 10 min segments. Genotypes were compared using t-tests if both

datasets fitted a normal distribution (determined using D'Agostino-Pearson omnibus and Shapiro-Wilk normality tests). Otherwise, Kolmogorov-Smirnov tests were used.

On day 3, muscle strength was analysed using the Hanging Wire test. The apparatus consists of a 1.5mm diameter wire, 35 cm above the bench. Animals are placed on a horizontal wire with their forepaws by the experimenter, and the time taken to bring a hind paw to the wire is recorded. Animals are given a maximum of 30 seconds to complete this task. Animals that take longer or fall off the wire are given the maximum score of 30. The test was performed three times for each animal (with an inter-test interval of 30 min) and the mean of the three tests was calculated. Genotypes were compared using t-tests if both datasets fitted a normal distribution (determined using D'Agostino-Pearson omnibus and Shapiro-Wilk normality tests). Otherwise, Kolmogorov-Smirnov tests were used.

On days 6-9, motor coordination and learning was analysed using the Accelerating Rotarod. This consists of a 3 cm diameter rod that can rotate between 4 and 40 revolutions per minute (rpm). On the first day, the animals are accustomed to the apparatus in a short training session where they must stay on the track for 30 seconds at its lowest speed (4 rpm). On the subsequent three days, each animal undergoes four trials (with an inter-trial interval of 90 min) where the speed is slowly increased from 4 rpm to 40 rpm over 5 min. The time taken to fall (latency) is recorded and the average time for each day is calculated. Genotypes were compared using t-tests if both datasets fitted a normal distribution (determined using D'Agostino-Pearson omnibus and Shapiro-Wilk normality tests). Otherwise, Kolmogorov-Smirnov tests were used. Learning (and worsening) were analysed by comparing the datasets for each genotype over time. As several datasets did not fit a normal distribution, the non-parametric Friedman test was used for all experiments.

2.2.17 Genetic reversal experiment

The knock-in ES cells for all lines contained a Neomycin resistance gene and transcriptional STOP cassette in intron 2. This cassette was not deleted before any of

the ES cell lines were used in blastocyst injections. 'STOP' mice unable to express the ΔNIC-EGFP allele were produced simply by breeding ΔNIC-EGFP (NeoSTOP) chimaeras with wild-type C57BL/6J females. STOP mice were scored and weighed (see section 2.2.15). Survival was graphed using Kaplan Meier plots and genotypes were compared using the using Mantel-Cox test.

Mice with the potential to activate the silenced ΔNIC-EGFP gene (STOP CreER) were produced by crossing heterozygous STOP females with males that were heterozygous for the $CreER^T$ transgene (C57BL/6J background). Male STOP animals that did not carry the transgene and wild-types with and without the transgene were also produced in this cross.

Male mice of all four genotypes (STOP $CreER^T$; STOP; WT $CreER^T$; WT) were used in the reversal experiment. All animals were scored and weighed weekly from 4 weeks (see section 2.2.15). From 6 weeks of age, they were all given a series of Tamoxifen (Sigma T5648-1G) injections. The regime consisted of two weekly injections followed by five daily injections, each at a dose of 100 μg Tamoxifen/g weight. Tamoxifen was solubilised by sonication in corn oil by sonication. The experiment was continued until 2 weeks after all the STOP controls had died, when the animals were 28 weeks of ages. At this point their brains were harvested for DNA and protein analysis (methods described above).

Chapter 3 – Designing truncations of MeCP2 lacking the potentially dispensable domains and verifying that these proteins retain 'bridge' function

3.1 Introduction and aims

The Rett syndrome (RTT) missense mutation spectrum highlights the Methyl-CpG Binding Domain (MBD) and the NCoR/SMRT Interaction Domain (NID) as the key functional domains of MeCP2. This suggests that MeCP2 forms a bridge between chromatin and the NCoR/SMRT co-repressor complex. My PhD project was to test this 'bridge' hypothesis by determining whether these two domains are necessary and sufficient for MeCP2 function *in vivo*. The strategy used was to make a stepwise series of truncated MeCP2 alleles where the regions outside the MBD and NID have been deleted and then test their function *in vivo* by producing mouse lines expressing the truncated alleles from the endogenous *Mecp2* locus. Three truncated proteins of decreasing length were designed and tested in this study (Fig. 3.1): the first (ΔN) is missing the N-terminal region before the MBD; the second (ΔNC) is missing both this N-terminal region and the long C-terminal region after the NID; and the third (ΔNIC) is missing both the N- and C-termini, and the Intervening region between the MBD and the NID has been replaced by an exogenous linker sequence and a nuclear localisation signal (NLS). This chapter outlines the initial stages of this project: the design of the truncated alleles and the assays that I carried out to determine that these two domains had retained their ability to interact with both methylated DNA and the NCoR/SMRT complex. The truncated proteins were designed to remove as many as possible of the other annotated domains in MeCP2 whilst retaining bridge function. It is important to note that these considerations were based on the functions of the various regions of MeCP2 known at the time (October 2013). I will discuss how novel findings might have impacted these designs at the end of this chapter.

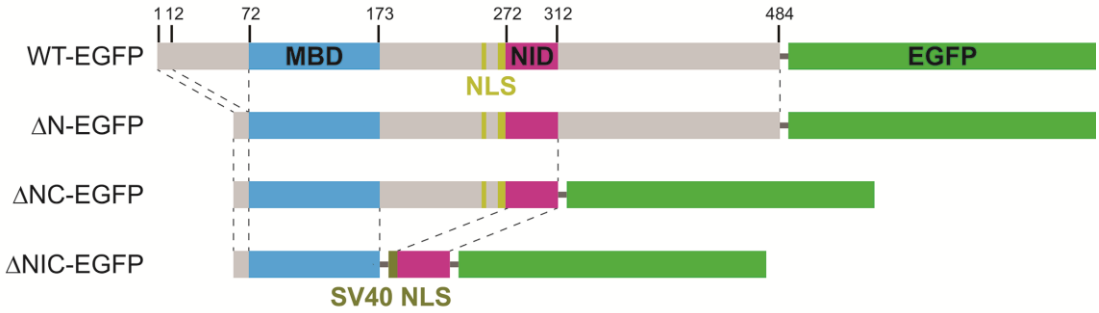


Figure 3.1: Stepwise truncations of MeCP2 protein
Schematic diagram of the truncated proteins designed in this study, compared to the full length protein (WT-EGFP). Several features shown here are described below: retention of the extreme N-terminus; the boundaries of the MBD and the NID; the use of the exogenous NLS from SV40; the presence of the EGFP tag; and the design of linker sequences. Drawn to scale (e2 isoforms). Linker sequences are shown as dark grey bars. Dotted lines denote corresponding regions between the proteins.

3.2 Considerations when designing truncated proteins

3.2.1 Complications at the N-terminus due to the Mecp2 gene structure

The *Mecp2* gene consists of four exons, and mRNA transcripts undergo alternative splicing resulting in the production of two protein isoforms: e1 and e2. The more abundant isoform, e1, is translated from mRNA consisting of exons 1,3 and 4, with its start codon in exon 1. The mRNA encoding the shorter and less abundant isoform, e2, contains all four exons and is translated from a downstream start codon in exon 2. These two isoforms are almost identical since the vast majority of the coding sequence for both is in exons 3 and 4. They differ only by the short extreme N-termini encoded in exons 1 and 2 for isoform e1 and e2, respectively (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004). The use of two translational start codons, one for each isoform (Fig. 3.2.1) makes deletion of the N-terminal sequence at the endogenous locus in the knock-in mice very complicated. I therefore considered the consequences of retaining the extreme N-terminal sequences in the truncated proteins. These are 26 amino acids and 9 amino acids long in e1 and e2, respectively (Fig. S2). No role has been attributed to the longer (e1) terminus, consistent with its poor evolutionary conservation (Fig. S2). The e2 isoform has, however, been

reported to promote apoptosis, a function that is negatively regulated by its interaction with FOXG1. Although the mechanism for how this isoform promotes apoptosis has not been identified, it is unlikely that this 9 amino acid sequence is sufficient for this activity. Furthermore, the e1 isoform can interact weakly with FOXG1, suggesting that the e2 N-terminal sequence is not sufficient for this interaction (Dastidar et al., 2012). I therefore decided to retain these extreme N-terminal sequences, leaving exon 1 and 2 of the endogenous gene unedited in the knock-in mice. To facilitate proper splicing into exon 3, the first 10 base pairs of this exon encoding Glu10-Glu11-Lys12 were also included in the truncated proteins (Fig. 3.2.1). Therefore, in the truncated proteins, the retained region contains a total of 29 and 12 amino acids in the e1 and e2 isoforms, respectively.

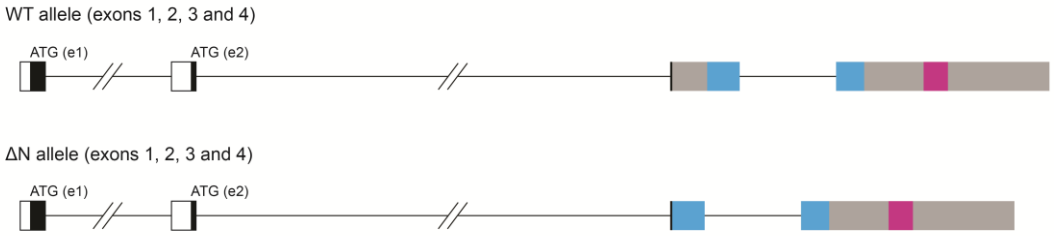


Figure 3.2.1: The short extreme N-termini of both isoforms were retained due to the MeCP2 gene structure
Schematic of the wild-type MeCP2 gene structure, encoded by four exons (upper); and how the N-terminus was deleted in the truncated proteins (lower). The 5'UTRs of e1 and e2 isoforms, located in exons 1 and 2, respectively, are shown in white. The retained extreme N-terminal sequences in exons 1, 2 and the very beginning of exon 3 are shown in black. The MBD is shown in blue and the NID is shown in pink. The deleted N-terminal, intermediate and C-terminal sequences are shown in grey.

3.2.2 Deciding the boundaries of the MBD

It is well established that the Methyl-CpG Binding Domain (MBD) of MeCP2 binds to DNA in a methylation-specific manner (Nan et al., 1993). However, the residues that define the MBD vary between different studies. Deletion analysis identified residues 78-162 to be the minimal region that could bind methylated DNA (Nan et al., 1993). A slightly longer sequence, residues 77-167, was used to determine the X-

ray structure, co-crystallised with a methylated DNA oligonucleotide (Ho et al., 2008).

The ‘bridge’ hypothesis predicts that loss of function mutations in MeCP2 must disrupt its ability to form the bridge. If this hypothesis is correct, all RTT-causing missense mutations located in the MBD must affect DNA binding – a consequence that has been reported for many of them (Table S3). Therefore, in this study it is important to include the entire cluster of missense mutations in the region defined as the MBD. This cluster extends from residue 100 to 161 (Fig 3.2.2), so is entirely within the minimal DNA-binding region identified in deletion analysis (Nan et al., 1993).

A large number of post-translational modifications (PTMs) have been identified along the length of MeCP2, many of which lie within the MBD. This includes four activity-dependent phosphorylation sites: pSer80, pSer86, Thr148/Ser149, and Ser164 (Fig. 3.2.2; Zhou et al., 2006; Tao et al., 2009; Ebert et al., 2013). At the time of designing these proteins, only phosphorylation of Ser80 had been implicated in the regulation of DNA binding (Tao et al., 2009). In neurons, Ser80 is phosphorylated under basal conditions and dephosphorylated upon membrane depolarisation. De-phosphorylation of this site has been reported to reduce DNA binding affinity, suggesting that this activity is regulated by neuronal signalling (Tao et al., 2009). At the time, the most likely mechanism for its effect on DNA binding affinity was steric hindrance due to the location of Ser80 at the N-terminal end of the MBD. It therefore seemed essential to include this residue in the defined MBD. The kinase responsible for its phosphorylation, HIPK2, recognises the motif S/T-P (Bracaglia et al., 2009), so further N-terminal extension was not required to ensure that this residue could be phosphorylated. Regardless of its functional importance, the location of Ser80 within the minimal region required for DNA binding (Nan et al., 1993) ensured its inclusion. Very little was known about the other modified residues around this region but all except pSer164 lie within this minimal binding region so could not be excluded.

When designing these truncated proteins, particularly the most severe (ΔNIC), it was tempting to remove as much of the potentially dispensable sequence as possible, leaving the minimal length required to form a bridge between chromatin and the NCoR/SMRT complex. Whilst the MBD has undergone precise deletion analysis (Nan et al., 1993), it is impossible to predict whether this minimal sequence would retain its function when fused to the sequences that end up adjacent to it as a result of deletions in these truncated proteins. The MBD in this study is preceded by the extreme N-terminal residues in all three truncated proteins, and followed by a sequence that links to an exogenous NLS in ΔNIC. To increase the chance that these neighbouring sequences would not interfere with DNA binding and that the protein would be properly folded, a larger region consisting of residues Pro72-Pro173 was used. This encompasses the peptide used to determine the crystal structure (Ho et al., 2008). As proline residues introduce flexibility to protein structure, they serve as good boundaries either side of the MBD. This region is very highly conserved throughout evolution, with reduced conservation either side (Fig. 3.2.2).

3.2.3 Deciding the boundaries of the NID

The region of MeCP2 required for binding to the NCoR/SMRT co-repressor complex has been identified using deletion analysis. It consists of residues 285-309, and is located at the C-terminus of the historical Transcriptional Repression Domain (Nan et al., 1998; Lyst et al., 2013). All RTT-causing missense mutations in this cluster are located within this region (Fig. 3.2.3) and several have been shown to abolish binding to members of the NCoR/SMRT co-repressor complex (Lyst et al., 2013). The binding region also contains the phospho-site Thr308 (Fig. 3.2.3), which is important for activity-dependent regulation of NCoR/SMRT complex binding (Ebert et al., 2013).

The structure of MeCP2 outside the MBD is unknown, and has been suggested to be largely disordered (Ghosh et al., 2011). I therefore had to rely on secondary structure prediction programmes to help with the design of this region of the truncated proteins. PHD software (https://npsa-prabi.ibcp.fr/NPSA/npsa_phd.html) predicted

the presence of two alpha helices around the NID: consisting of residues 274-287 and 300-307 (Fig. 3.2.3; Fig S4). The region defined as the NID in this study is extended at the N-terminus until Pro272 to include the first of these helices. Unfortunately, the inclusion of these extra amino acids results in the retention of an uncharacterised activity-dependent phosphorylated residue, Ser274 (Ebert et al., 2013). It may be, however, that this residue cannot be phosphorylated in ΔNIC due to the loss of the upstream residues that could be required for the kinase responsible for its phosphorylation to recognise this site.

The C-terminus of the NID was easier to define as the deletion analysis was more precise at this end, with the removal of a single amino acid (peptides ending with residue 308) resulting in reduced binding (Lyst et al., 2013). This is particularly interesting as a well characterised knock-in mouse expressing MeCP2 1-308 exhibits mild RTT-like symptoms (Shahbazian et al., 2002b). The NID used in this study was only extended by three amino acids beyond the identified binding domain to Val312, taking it to the end of an evolutionarily conserved region (Fig. 3.2.3). This is consistent with another knock-in mouse expressing MeCP2 residues 1-312, which was already being characterised in the lab when I began this project (Lagger and Selfridge, unpublished).

3.2.4 Exogenous NLS

The endogenous bipartite nuclear localisation signal (NLS) of MeCP2 is located in the Intervening domain, directly upstream of the boundary of the NID used in this study (Fig. 3.2.3). Its functionality and ability to bind nuclear import machinery have been confirmed (Nan et al., 1996; Baker et al., 2015; Lyst, unpublished), yet its requirement for MeCP2 protein function remains controversial. This is because the MBD is sufficient for MeCP2 localisation to heterochromatin (Nan et al., 1996). Despite the attraction of reducing MeCP2 to just the MBD and the NID, it seemed safer to include an NLS as it may be necessary for function *in vivo*. As the second part of the endogenous bipartite NLS has also been shown to be a weak AT-hook

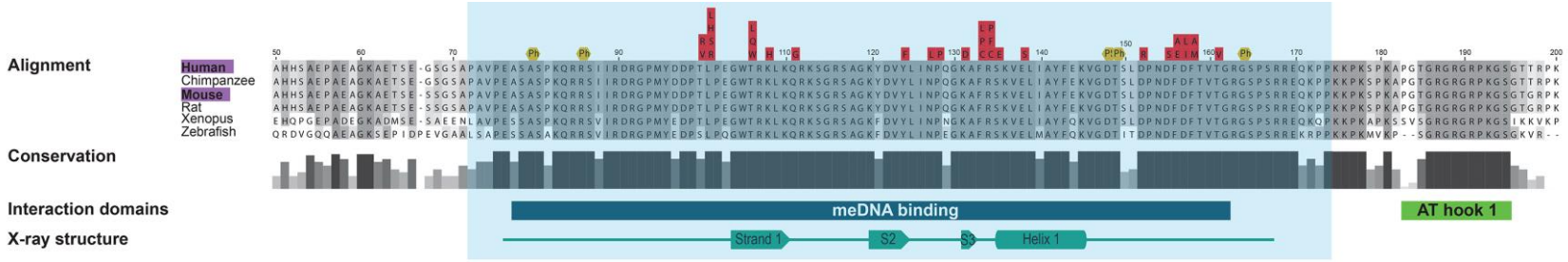


Figure 3.2.2: Annotated features around the MBD
Protein sequence alignment of the MBD region using ClustalWS, shaded according to BLOSUM62 score. Annotated with (from top): RTT-missense mutations (red) and activity-dependent phosphorylation sites (orange); sequence conservation; interaction domains (minimal MBD required for binding, Nan et al., 1993; AT hook 1 peptide 183-195, Lyst, Connelly et al., 2016); X-ray structure with methylated DNA (peptide and secondary structure, Ho et al., 2008). Residue numbers correspond to that of mammalian e2 isoforms. The region defined as the MBD in this study is highlighted by the blue rectangle.

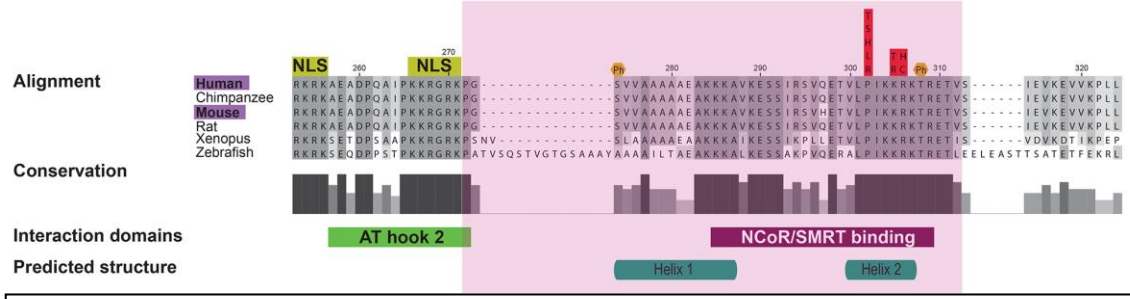


Figure 3.2.3: Annotated features around the NID
Protein sequence alignment of the NID region using ClustalWS, shaded according to BLOSUM62 score. Annotated with (from top): RTT-missense mutations (red), activity-dependent phosphorylation sites (orange), and the bipartite nuclear localisation signal (yellow); sequence conservation; interaction domains (AT hook 2 peptide 257-272, Baker et al., 2013; NCoR/SMRT interaction domain, Lyst et al., 2013); predicted secondary structure (PHD prediction software, October 2013). Residue numbers correspond to that of mammalian e2 isoforms. The region defined as the NID in this study is highlighted by the pink rectangle.

(Baker et al., 2013; Lyst, Connelly et al., 2016), I chose to delete the endogenous NLS and replace it with an exogenous one from SV40. This is a commonly used strong monopartite NLS (sequence PKKKRKV) that lacks the AT hook motif (Kalderon et al., 1984).

3.2.5 The inclusion of the EGFP tag

There are several good quality antibodies that recognise MeCP2, but they all detect epitopes that are located in either the N- or C-termini. Therefore, the two smaller truncated proteins, ΔNC and ΔNIC, would be undetectable. This was solved by the inclusion of an exogenous tag. The structure of the *Mecp2* gene (as described in section 3.2.1) means that the tag must be added to the C-terminus. In the case of the C-terminally truncated proteins, ΔNC and ΔNIC, it would be directly after the NID (joined by a short linker, described below). Whereas several short tags such as HA or an *in vivo* biotinylated peptide (de Boer et al., 2003) have been shown to work efficiently for western blotting, immunofluorescence and ChIP, they are unstructured and therefore might interfere with NCoR/SMRT complex binding. EGFP was chosen as it is globular (Royant and Noirclerc-Savoye, 2011) and therefore less likely to interfere with the structure of the truncated MeCP2 proteins. Additionally, knock-in mice expressing wild-type MeCP2 tagged with EGFP show no RTT-like symptoms (Brown, Selfridge et al., 2016). These mice and the R306C-EGFP mice provided vital controls in the biochemical characterisation of the mice presented in this study.

3.2.6 Linker sequences

To increase the chance that the MBD and NID retained their function in the truncated proteins, flexible linker sequences were included between these two domains. As the N-terminal extension of the NID to Pro272 reached the endogenous NLS (Fig. 3.2.3), I decided not to insert additional amino acids between the SV40 NLS and the NID. The first three amino acids of the NID, Pro272-Gly273-Ser274, provide flexibility between the SV40 NLS and the functional portion of the NID. As the existing WT-EGFP knock-in mice contain a short linker (CKDPPVAT) between the C-terminus of

MeCP2 and the EGFP tag (Brown, Selfridge et al., 2016), I chose to use the same sequence in the design of the ΔN-EGFP protein for consistency. However, a more flexible linker was preferred to connect the NID and the EGFP tag for the two C-terminally truncated alleles. Gly-Ser linkers are commonly used as they are both flexible and polar, so do not introduce hydrophobicity (e.g. Jørgensen et al., 2006). The sequence GSSGSSG was inserted between the NID and EGFP tag, and also between the MBD and the SV40 NLS in ΔNIC-EGFP. Codons were varied to avoid repetitive sequence, taking mouse codon usage into account. Sequence alignments of the truncated proteins and cDNA sequences are included in the Appendix (Fig. S5-6).

3.3 Verification of bridge function in the truncated proteins

Before making knock-in mice with the series of truncated *Mecp2* alleles, it was important to test whether the proteins I designed retained the ability to form the bridge between methylated DNA and the NCoR/SMRT co-repressor complex. This was done with established transient transfection assays using cultured mammalian cells. For this, cDNA sequences encoding e2 isoforms of the MeCP2 truncated proteins were cloned into the commercial overexpression vector, pEGFPN1.

3.3.1 All truncated proteins are successfully expressed in HeLa cells where they are able to bind endogenous NCoR/SMRT complex components

HeLa cells have successfully been used to determine the ability of FLAG-tagged fragments and mutants of MeCP2 to bind endogenous NCoR/SMRT complex components, mirroring the results from mouse brain (Lyst et al., 2013). Briefly, MeCP2 constructs are overexpressed transiently in HeLa cells and harvested after 24-48 hours. The MeCP2 variants are pulled down from nuclear extracts and the presence of NCoR/SMRT complex components in the immunoprecipitate is determined by western blotting. I performed this assay with the truncated proteins that I had designed, using GFP-TRAP beads. Western blotting using a GFP antibody shows that all proteins were successfully expressed and immunoprecipitated, with

bands at the expected sizes (Fig. 3.3.1; Table S4). Furthermore, all three truncated proteins were able to bind members of the NCoR/SMRT complex: NCoR, HDAC3 and TBL1XR1. As previously reported (Lyst et al., 2013), this interaction was abolished in the NID mutant, R306C, which was used as a negative control. This experiment shows that all three truncated proteins retain NID functionality.

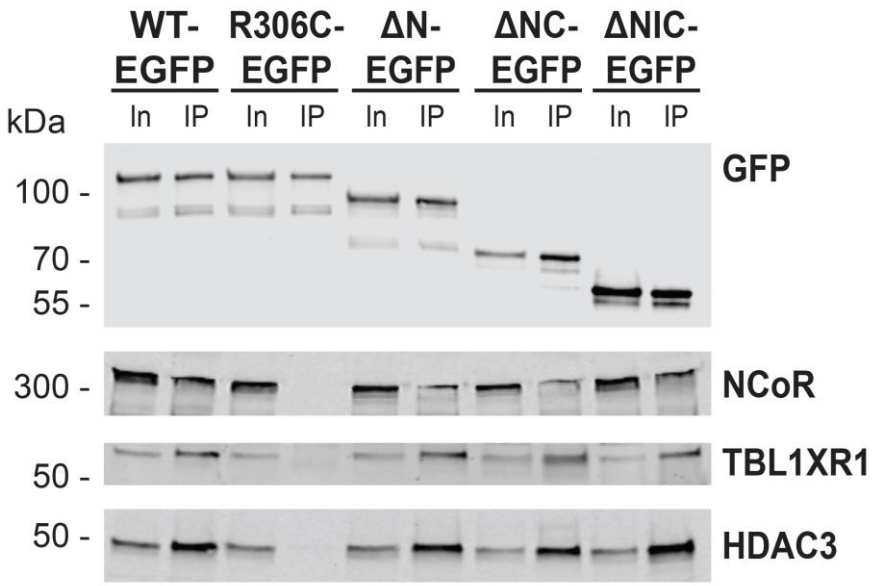


Figure 3.3.1: Truncated MeCP2 proteins can bind NCoR/SMRT complex components
Immunoprecipitation of EGFP-tagged MeCP2 proteins: full length (WT), with the R306C mutation (R306C), and the three truncations (ΔN, ΔNC, ΔNIC), from HeLa nuclear extracts using GFP-TRAP beads. Western blots are shown to indicate successful expression and immunoprecipitation of the MeCP2 proteins (probed using a GFP antibody), and to assess binding to NCoR/SMRT complex components (NCoR, HDAC3 and TBL1XR1 antibodies). 'In' = input, 'IP' = immunoprecipitate, m = size marker. Representative blot of eight experiments (three of which were probed with antibodies against all three NCoR/SMRT complex components).

3.3.2 Verification of binding to methylated DNA in mouse fibroblasts

In mouse cells, around 40% of mCpG marks are present in pericentromeric heterochromatin, which appears as DAPI bright spots (Lewis et al., 1992; Nan et al., 1996). Co-localisation of proteins such as MeCP2 with these foci is a well-established assay for determining whether they can bind to DNA in a methyl-specific

manner (Nan et al., 1996). All three truncations of MeCP2 co-localised with the DAPI bright spots when they were over-expressed in NIH-3T3 cells (mouse fibroblasts). The MBD mutant, R111G, which abolishes binding to methylated DNA (Kudo et al., 2003), was used as a negative control and shows diffuse nuclear localisation (Fig. 3.3.2).

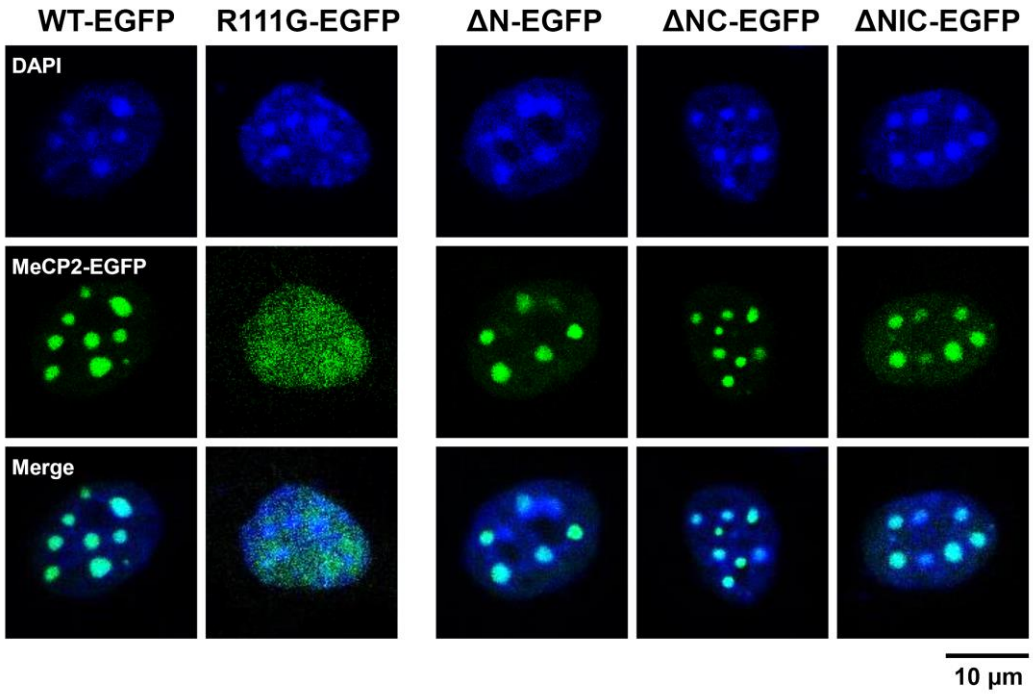


Figure 3.3.2: Truncated MeCP2 proteins can bind methylated DNA
EGFP-tagged MeCP2 proteins were overexpressed in NIH-3T3 cells and the EGFP tag was visualised using confocal microscopy. DAPI staining shows pericentromeric heterochromatin as bright spots. All three truncated proteins (ΔN, ΔNC, ΔNIC) co-localise with these foci in all cells, resembling the wild-type protein (WT). The DNA binding mutant, R111G, instead shows diffuse nuclear localisation.

3.3.3 Simultaneous binding to methylated DNA and a component of the NCoR/SMRT complex, TBL1X

The 'bridge hypothesis' predicts that MeCP2 recruits the NCoR/SMRT co-repressor complex to chromatin. It is therefore vital that the truncated proteins retain the ability to bind DNA and NCoR/SMRT components simultaneously. This can be determined using an assay developed in our lab (Lyst et al., 2013). The NCoR/SMRT complex

member TBL1X lacks an NLS and, when overexpressed alone in NIH-3T3 cells, it remains in the cytoplasm. Co-overexpression with MeCP2 results in its translocation to heterochromatic foci. This is dependent on NID binding activity as it is abolished by the R306C mutation (Lyst et al., 2013; Fig. 3.3.3). I performed this assay using the truncated MeCP2-EGFP proteins. All three were able to bring TBL1X-mCherry to the DAPI bright spots, confirming bridge function (Fig. 3.3.3).

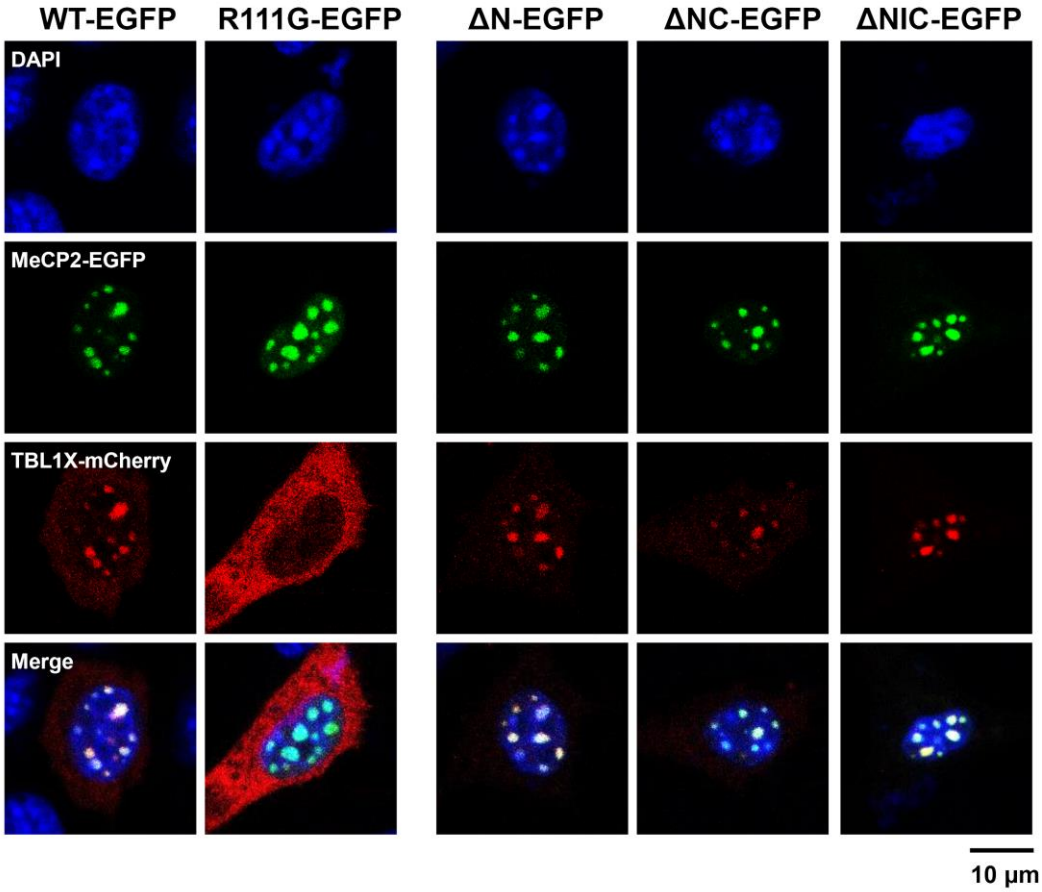


Figure 3.3.3: Truncated MeCP2 proteins can bind methylated DNA and the NCoR/SMRT complex member, TBL1X, simultaneously
MeCP2-EGFP proteins were co-overexpressed with TBL1X-mCherry in NIH-3T3 cells, and the fluorescent tags were visualised by confocal microscopy. DAPI staining shows pericentromeric heterochromatin as DAPI bright spots. All three truncated proteins (ΔN, ΔNC, ΔNIC) recruit TBL1X-mCherry to heterochromatic foci, resembling the wild-type protein (WT). The NID mutant, 306C, is unable to recruit TBL1X-mCherry, which remains in the cytoplasm.

3.4 Discussion

In order to test the 'bridge' hypothesis, I have designed a truncated MeCP2 protein that retains only the Methyl-CpG Binding Domain (MBD) and NCoR/SMRT interaction domain (NID) of the wild-type sequence. This protein also contains an exogenous NLS (from SV40) between the two domains to facilitate nuclear import and a C-terminal EGFP tag to enable its detection and purification. These domains are connected with short flexible linker sequences in case their direct fusion affected their activity. The protein is referred to as ΔNIC-EGFP, because it lacks all three other regions of full-length MeCP2: the N-terminus before the MBD, the Intervening region between the MBD and the NID, and the C-terminus after the NID. ΔNIC-EGFP retains only 34% of the full-length MeCP2 amino acid sequence (Table S4). In this chapter, I have explained how I designed this truncated protein and demonstrated that it retains the ability to bind methylated heterochromatin and the NCoR/SMRT co-repressor complex simultaneously.

If the 'bridge' hypothesis is correct, replacement of endogenous *Mecp2* with the genomic DNA sequence encoding ΔNIC-EGFP will not have any phenotypic consequences. If, however, these mice display RTT-like symptoms, I would want to determine which of the deleted regions is responsible for this protein's reduced function. To enable this distinction (if required), I also designed two less severely truncated MeCP2 proteins: ΔN-EGFP, which only lacks the N-terminus; and ΔNC-EGFP, which lacks both the N- and C-termini. Together, these three truncated proteins form a step-wise deletion series, retaining 88%, 54% and 34% of the full-length MeCP2 protein sequence. The pre-existing knock-in line expressing C-terminally truncated MeCP2 (residues 1-312), which retains 66% of the protein sequence, can also be considered to be a member of this series (Lagger and Selfridge, unpublished). Unlike the three novel truncations, the 1-312 protein does not have an exogenous tag. The ability to bind methylated heterochromatin and the NCoR/SMRT co-repressor complex simultaneously was also confirmed for all three of these less severe deletions (this study; Guy and Alexander-Howden, unpublished).

3.4.1 Impact of the Mecp2 gene structure of protein design

All three of the truncated alleles designed in this study lack the N-terminus before the MBD. Deletion of this region was designed in such a way that genomic DNA sequences encoding the truncated protein can be inserted into the endogenous *Mecp2* locus in mice, i.e. exons 1 and 2 (containing the start codons for isoforms e1 and e2, respectively) were retained as well as the first 10 base pairs of exon 3 to enable proper splicing. This results in the retention of 29 or 12 amino acids in the e1 or e2 isoforms, respectively.

3.4.2 Deletion of the N- and C-termini and Intervening region is likely to abolish almost all other reported functions, including those discovered since designing these proteins

The main aim when designing the most severely truncated protein (ΔNIC-EGFP) was to retain enough of the Methyl-CpG Binding Domain (MBD) and NCoR/SMRT interaction domain (NID) sequences to preserve the interactions with methylated DNA and the NCoR/SMRT complex, whilst abolishing the other proposed functional domains of MeCP2. The sequences defined as the MBD (resides 72-173) and the NID (272-312) are extended by several amino acids either side of each of the minimal binding domains to increase the chance that they retain their functionality within the context of the truncated proteins. Although the vast majority of the mapped interaction sites reported in the literature are wholly or partially deleted in ΔNIC-EGFP, more precise mapping is needed to predict whether some interactions are likely to be lost, for example with FOXG1 and SIN3A (Fig. S7; Dastidar et al., 2012; Nan et al., 1998; Lyst et al., 2013). Due to overlap with the MBD, it was impossible to delete the sequences required for either ATRX or DNMT1 binding (Nan et al., 2007; Kimura et al., 2003).

Several novel MeCP2 interaction partners have been reported since I finalised the design of these proteins. Three of these have been mapped, all to the deleted regions: LEDGF and DHX9 to the Intervening region (Li et al., 2016) and DGCR8 to the C-

terminus (Cheng et al., 2014a). Heckman et al. (2014) also reported a novel DNA-binding domain termed the 'basic patch' which overlaps with the NID. Although a longer peptide was used to demonstrate its DNA-binding activity (Fig. S7), the basic residues required for DNA binding are all located in the NID. Therefore, of these recent findings, only the function of 'Basic Patch' is likely to be retained in ΔNIC-EGFP, and this would have been unavoidable anyway due to its overlap with the NID. Another novel finding is that activity-dependent phosphorylation of Ser164 (located just after the minimal MBD) has been reported to disrupt DNA binding (Stefanelli et al., 2016). Since I wanted to retain activity-dependent regulation of MBD activity, it is fortunate that the slightly extended MBD used in this study (residues 72-173) includes this phospho-site and any nearby residues that might be needed for recognition by the kinase responsible for its phosphorylation.

3.4.3 Deletion of these regions may have structural consequences that affect its activity

It has long been discussed in the literature that MeCP2 is highly disordered outside the MBD (Ghosh et al., 2011), making it seem a perfect candidate for deletion analysis. Increasing evidence now indicates the presence of intramolecular interactions within its length (Cheng et al., 2014a; Connelly and Belsom, unpublished). Cheng et al. (2014a) demonstrate that the two halves of the protein can directly interact, and that this is disrupted by Ser80 phosphorylation. Ser80 is normally phosphorylated at high levels in neurons and is dephosphorylated upon membrane depolarisation (Bracaglia et al., 2009; Tao et al., 2009). Phosphorylation of this site has been shown to be promote DNA binding (Tao et al., 2009) and be essential for binding to YB-1 (Gonzales et al., 2012) and DGCR8 (Cheng et al., 2014a). Together, these findings fit a model whereby neuronal signalling promotes intramolecular interactions by dephosphorylation of Ser80, reducing its affinity for DNA, YB-1 and DGCR8. The N- and C- terminal fragments used by Cheng et al. (2014a) to show intramolecular binding in MeCP2 consisted of residues 1-305 and 306-CT. It is therefore highly likely that this interaction is lost by deletion of the C-terminus in this study (residues 313-CT). Consequentially, characterisation of

knock-in mice expressing the C-terminally truncated proteins (ΔNC-EGFP, ΔNIC-EGFP and 1-312) will help determine the functional importance of the intramolecular interaction *in vivo*.

3.4.5 Concluding remarks

I have successfully designed three truncations of MeCP2 where the regions outside the MBD and NID have been progressively deleted. The shortest of these proteins, ΔNIC-EGFP, retains only the MBD and NID of MeCP2. All other reported functional domains are wholly or partially deleted in this protein except for a small number that are unavoidable. Using cell culture-based assays, I have proved that all three of these proteins retain the ability to bind methylated heterochromatin and the NCoR/SMRT complex simultaneously. Therefore, these designs meet all the necessary criteria required to be used in the production of knock-in mice. Characterisation of the resulting mouse lines will determine the importance of each of the three potentially dispensable regions. If the ΔNIC-EGFP mice display no RTT-like symptoms, this will provide a strong indication that 'bridge' function is sufficient for proper MeCP2 activity *in vivo*.

Chapter 4 – Generation of knock-in mouse lines from targeted ES cells

4.1 Introduction and aims

In order to test the functionality of the three truncated MeCP2 proteins (see Chapter 3), I had to produce a knock-in mouse line for each. *Mecp2*-null and loss-of-function mouse models recapitulate many of the features of Rett syndrome, including hypoactivity, tremors, abnormal gait, and breathing problems (e.g. Guy et al., 2001; Chen et al., 2001; Baker et al., 2013; Brown, Selfridge et al., 2016). Although the most appropriate model of Rett syndrome is the heterozygous female, the symptoms exhibited by the mice are relatively mild and only develop at 3-12 months of age (Guy et al., 2001). Hemizygous male *Mecp2*-null mice develop symptoms around 4 weeks of age, resulting in premature death at 6-12 weeks (Guy et al., 2001; Chen et al., 2001). The use of hemizygous males also simplifies biochemical characterisation of *Mecp2* knock-in lines as heterozygous females are mosaic due to X chromosome inactivation (XCI). In this chapter, I will outline the production process of the knock-in mice lines carrying truncated alleles of MeCP2. I will also present the biochemical characterisation of these mice: determining the expression levels of the truncated MeCP2 proteins and mRNA transcripts and the ability of these proteins to interact with NCoR/SMRT complex components in the brain.

4.2 ES cell targeting strategies and screening processes

All mouse lines were produced from targeted male ES cells in which the truncated alleles had been knocked into the endogenous *Mecp2* locus. This was the predominant method at the time, and avoids the complications of sterility that may occur in hemizygous male or homozygous female *Mecp2* mutants produced by CRISPR/Cas9 injection protocols using zygotes. CRISPR/Cas9 technology was used instead to increase gene editing efficiency in the production of ΔN-EGFP and ΔNIC-EGFP knock-in ES cells.

4.2.1 Generation of one positive ΔNC-EGFP ES cell line using the traditional method

The first truncated allele of MeCP2 used for ES cell targeting was ΔNC-EGFP: the intermediate of the three, missing both the N- and C-termini (Fig. 3.1). Prior to ES cell targeting, a synthesised DNA fragment encoding the ΔNC-EGFP genomic sequence was cloned into the pBS_NeoSTOP_MeCP2 targeting vector using MfeI restriction sites (Fig. 4.2.1A). This fragment contained exons 3 and 4 and the intervening intron, the EGFP tag, and flanking regions up to the restriction sites. Exons 1 and 2 are located upstream of the 5' homology arm in this vector and remain unedited (see Chapter 3).

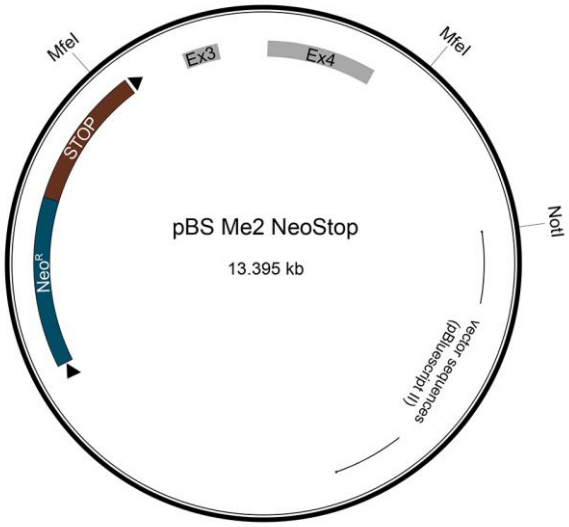


Figure 4.2.1A: pBS_NeoSTOP_MeCP2 targeting vector
Diagram of the targeting vector cloned by Jacky Guy (Guy et al., 2007). It has a pBluescript II vector backbone and contains the wild-type mouse MeCP2 genomic DNA sequence of exons 3 and 4 flanked by parts of intron 2 and the 3' UTR. A Neomycin resistance gene, 'Neo', (conferring resistance to G418) and a transcription stop cassette, 'STOP', flanked by *loxP* sites (black arrows) has been inserted into intron 2. MfeI sites were used in this study to clone in the fragments encoding the EGFP-tagged truncated MeCP2 alleles. NotI was used for linearisation of the plasmid.

The targeting vector was linearised using the NotI restriction site (Fig. 4.2.1A) and introduced into male ES cells via electroporation. G418 resistance was used to select

for cells that had integrated the vector. In total, 192 clones were picked and grown up in 96 well plates, from which cells were frozen and genomic DNA was extracted. Potential positive clones were identified by PCR screening (Fig. 4.2.1B). Only clones that have integrated the vector at the correct locus will produce a PCR product as primers span the end of the 5' homology arm. Eight positive clones were identified in an initial screen using F4/R3 PCR primers, all of which were confirmed with a second PCR reaction using F4/R5 primers.

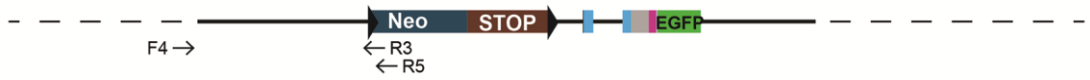


Figure 4.2.1B: 5' PCR screen
Diagram of the ΔNC-EGFP locus showing the location of PCR primers used for screening ES cell clones. The solid black line represents the sequence encoded in the targeted vector and the dotted lines indicate the flanking regions present in mouse genomic DNA. The ΔNC-EGFP locus is drawn to scale, but the primer arrows are enlarged for visibility. Clones were screened using F4/R3 and positives were confirmed using F4/R5. Both reactions give a 3.16 kb PCR product.

All eight clones were successfully grown up to fill T25 flasks, from which they were cryopreserved or used for genomic DNA extraction. Four clones that had a different sized band in the initial PCR screen were also grown up as negative controls. The higher quality genomic DNA prepared from the cells grown up in T25 flasks was further screened using Southern blotting (Fig 4.2.1C). Since the PCR screen had spanned the 5' end, I ensured that Southern analysis would be able to distinguish between partial and full integration of the ΔNC-EGFP genomic sequence. An internal probe located within the 3' homology arm was used to detect both the correctly targeted locus and random integrants. Analysis with Bsu36I digestion shows that only one clone, 2G7, had been correctly targeted, as the other seven clones all contain the 4.7 kb wild-type band. Two clones (1D7 and 1E12) contain additional bands of different sizes, which are likely to be due to random integration of the vector. The second blot, using BsrGI digestion, confirms that 2G7 has been correctly targeted. In this blot, the seven clones that appeared positive in the PCR have a band that is larger than the 9.7 kb wild-type band present in the negative

control clones. This is likely to be due to recombination within the length of the donor sequence, resulting in integration of the NeoSTOP cassette and N-terminal deletion but not of the C-terminal deletion and EGFP tag. This may have occurred due to the length of the DNA sequence to be integrated (5.5 kb); and the fact that the sequence of the ΔNC-EGFP donor molecule only differs from the wild-type genomic sequence at the ends (the NeoSTOP cassette and N-terminal deletion at the 5' end and the C-terminal deletions and EGFP tag at the 3' EGFP). In between, there is a long region (1.7 kb) of DNA encoding the truncated protein that is homologous to the wild-type sequence where recombination could occur. Sequencing of clone 1A2 confirmed that this is indeed was had happened, and these clones contain an untagged ΔN allele. The presence of both bands in clone 2C2 suggests that it is a mixed clone.

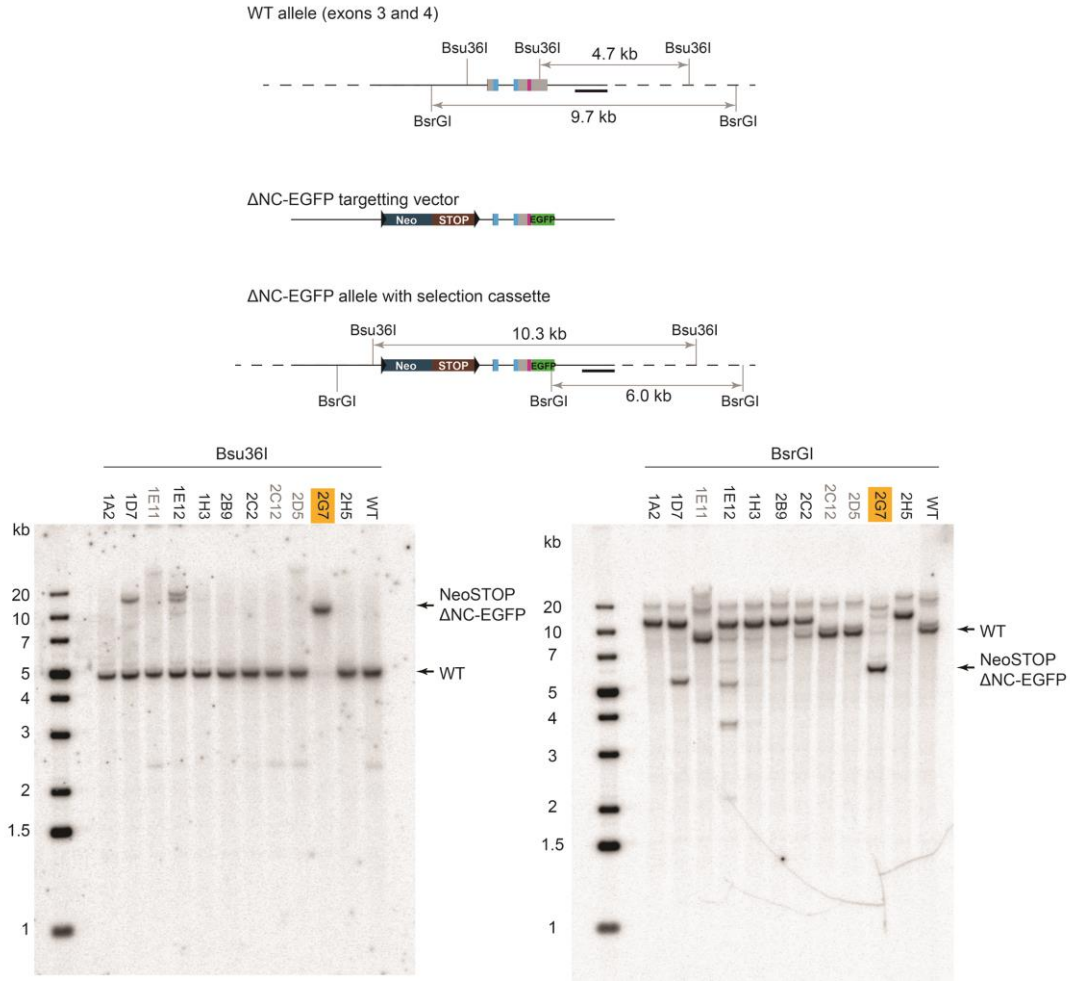


Figure 4.2.1C: legend on next page

Figure 4.2.1C: Southern blot analysis shows one successfully targeted ΔNC-EGFP clone
Screening of potential positive clones by Southern blotting. The upper portion shows the ΔNC-EGFP targeting strategy with maps of the restriction sites of the two enzymes used in Southern analysis, Bsu36I and BsrGI. An internal probe, recognising the 3' homology arm was used, drawn as a black line. The positive clone, 2G7, is highlighted in orange.

The single positive clone, 2G7, was verified further before it was used to make knock-in mice. Sequencing of the *Mecp2* locus in genomic DNA confirmed that the ΔNC-EGFP allele had been incorporated successfully without mutations. I also checked that these cells had the normal karyotype of 40 chromosomes (Fig. 4.2.1D). This clone passed all requirements so could be used in blastocyst injections to produce ΔNC-EGFP knock-in mice.

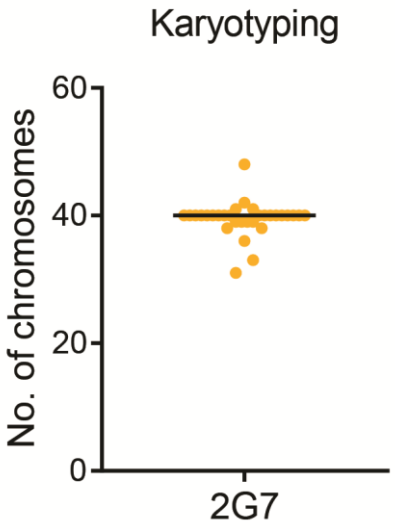


Figure 4.2.1D: Clone 2G7 has a normal karyotype
Cells were arrested in metaphase with colcemid and stained with DAPI. Chromosomes were counted manually in 31 cells. Median = 40, shown by black bar. 40 chromosomes were counted in 58% of cells.

4.2.2 Increased efficiency and extra considerations with the addition of CRISPR/Cas9 technology

The gene editing method described above was used successfully to produce a knock-in ΔNC-EGFP ES cell line. Nevertheless, the frequency of correct targeting was very low: the donor plasmid was integrated at the correct locus in 4.2% (8/192) G418 resistance clones and the entire desired sequence was integrated in only one of these.

To increase the gene editing frequency of the two remaining constructs to generate ΔN-EGFP and ΔNIC-EGFP knock-in ES cells, I took advantage of CRISPR/Cas9 technology. Cas9 is a bacterial DNA endonuclease from *Streptococcus pyogenes* that makes a double-stranded cut in DNA at a site determined by the sequence of a guide RNA (gRNA). The gRNA sequence must be followed by the Protospacer adjacent motif (PAM), which has the consensus 'NGG'. A gDNA targeting the wild-type sequence at the site in intron 2 where the NeoSTOP cassette had been inserted was designed and tested by Jacky Guy (Fig. 4.2.2A). The absence of this sequence in the donor molecule prevents its excision by Cas9. It also prevents subsequent cutting of the targeted allele after integration of the donor plasmid.

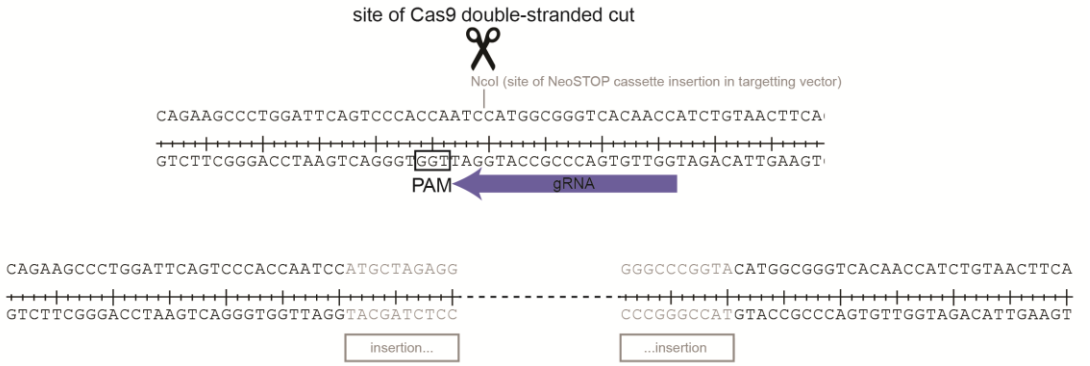


Figure 4.2.2A: CRISPR/Cas9 targeting strategy
(Upper) Sequence in intron 2 of the wild-type locus showing the location of the guide RNA (gRNA), the PAM, and the Cas9 double-stranded cut site. The guide RNA spans the NcoI site that was used to insert the NeoSTOP cassette into intron 2 in the targeting vector. (Lower) The same region of the intron 2 sequence after integration of the targeting vector. The inserted NeoSTOP cassette (labelled 'insertion', shown in grey) in this sequence disrupts the guide RNA sequence, preventing further cleavage by Cas9.
CRISPR strategy designed and tested by Jacky Guy

The pX330 plasmid expressing both the gRNA and Cas9 was introduced into ES cells with the linearised targeting vector via electroporation. A higher level of plasmid integration was evident by the density of G418 resistant clones: single clones were most easily picked from plates seeded at one fifth of the density used previously. Ninety-six clones were picked for each allele, and genomic DNA was screened by PCR using F4/R3 primers as before (Fig. 4.2.1B). Of these, 67% (58/86,

excluding clones that did not grow up) and 66% (61/92) of clones gave PCR products of the correct size for ΔN-EGFP and ΔNIC-EGFP, respectively. Therefore, the introduction of CRISPR/Cas9 has greatly enhanced the targeting frequency at the correct locus.

The low editing frequency in the production of a ΔNC-EGFP clone was due to two reasons: low insertion rate at the correct locus and only partial insertion of the donor molecule at this site. It is clear from the 5' PCR screen that CRISPR/Cas9 mediated cleavage of the endogenous *Mecp2* locus overcame the first of these problems. The problem of partial insertion of the donor molecule due to recombination within its length was likely to occur in the production of ΔN-EGFP and ΔNIC-EGFP knock-in ES cells since both of these vectors have long stretches of wild-type sequence within their length. To distinguish between full and partial insertion of the desired sequence in these clones, I developed a 3' PCR screen (Fig. 4.2.2B). Only a small proportion of clones that were positive in the 5' PCR screen (F4/R3 primers) were also positive in the 3' screen using F2/R2 primers: 22% (13/58) and 36% (22/61) of ΔN-EGFP and ΔNIC-EGFP clones, respectively. Positive clones were confirmed with both the 5' F4/R5 primers and the 3' R1/R2 primers (suitable for ΔNIC-EGFP clones only). These results show that the high level of partial recombination persists in the clones targeted using CRISPR/Cas9. This is notably lower in the ΔNIC-EGFP clones, which have a shorter region of internal homology due to the deletion of the Intervening region and C-terminus.

Figure 4.2.2B: 5' and 3' PCR screens
Diagram of the ΔNIC-EGFP locus showing the location of PCR primers used for screening ES cell clones. The solid black line represents the sequence encoded in the targeted vector and the dotted line indicates the flanking regions present in mouse genomic DNA. The ΔNIC-EGFP locus is drawn to scale, but the primer arrows are enlarged for visibility. Clones were screened using F2/R2 and ΔNIC-EGFP positives were confirmed using F1/R2. The reaction products are 3.18 kb and 3.11 kb for F2/R2 and F1/R2, respectively. F1 is not suitable for screening ΔN-EGFP clones as it binds to the sequence encoding the GSSGSSG linker.

Clones that were positive in the PCR screens were then screened by Southern blotting. Firstly, this was done using the genomic DNA purified from the 96-well plates that had been used for the PCR screening (Fig. S8A). While all the clones contained a 6.0 kb band, suggesting correct targeting, a few contained one or more (often multiple) additional bands representing random integration. For this reason, four ΔN-EGFP clones and six ΔNIC-EGFP clones were excluded at this stage (not highlighted in Fig. S8A). The remaining nine ΔN-EGFP clones and sixteen ΔNIC-EGFP clones were thawed and grown up in T25 flasks for cryopreservation and high quality genomic DNA purification.

Genomic DNA purified from the nine potential positive ΔN-EGFP clones was analysed by Southern screening with a second enzyme, BamHI. This experiment identified two correctly targeted clones, 4E1 and 4G9 (Fig S6B). Clone 4H12 was excluded due to the extra band on the preliminary BsrGI blot. Other clones contain additional random integrants (e.g. 3A11), or are mixed clones consisting of both ΔN-EGFP and untagged ΔN cells (e.g. 4F3). The sixteen potentially positive ΔNIC-EGFP clones were also screened with a second enzyme, Bsu36I. This identified five positive clones: 3E1, 3F1, 3F5, 3H10 and 4D7, which all only contained a band of the correct size in both blots (Fig S6C, highlighted in pink). Several clones (e.g. 3E3) were excluded due to the presence of a band 4.7 kb in size. This is likely to be due to the presence of untagged ΔN rather than wild-type cells (which give the same sized band) in these clones as the cells underwent G418 selection. Other clones (e.g. 4C7) were rejected due to the presence of additional random intergrants.

The ΔN-EGFP and ΔNIC-EGFP positive clones underwent the same checks as described above. Sequencing of genomic DNA showed that the *Mecp2* locus had been correctly edited. The CRISPR design tool (http://crispr.mit.edu/) that was used to design the gRNA predicts the top off-target sites. The highest scoring intragenic locus, Dock5, was also sequenced. None of the clones obtained off-target mutagenesis at this site. Although all clones appeared to have a normal karyotype

(Fig. 4.2.2C), I chose ΔN-EGFP clone 4G9 and ΔNIC-EGFP clone 3F1 for the production of knock-in mice as these had the clearest results.

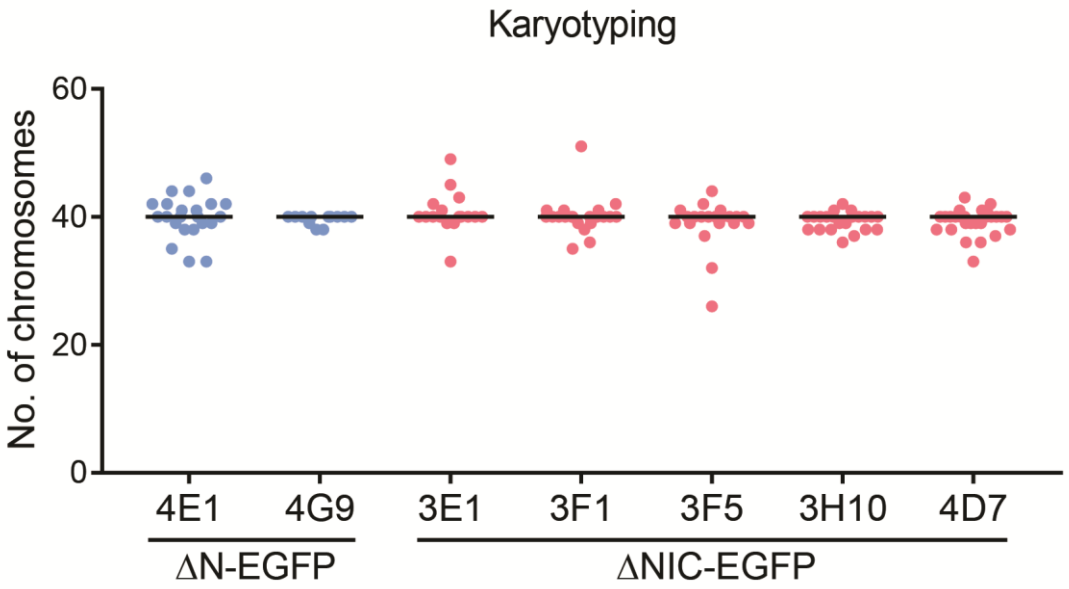


Figure 4.2.2C: All ΔN-EGFP and ΔNIC-EGFP clones have a normal karyotype Cells were arrested in metaphase with colcemid and stained with DAPI. Chromosomes were counted manually. Median = 40 in all, shown by black bar. Percentage of cells with 40 chromosomes: 4E1 27% (n=22), 4G9 75% (n=12), 3E1 53% (n=17), 3F1 52% (n=21), 3F5 43% (n=21), 3H10 45% (n=22), and 4D7 41% (n=27).

4.3 Generation of knock-in mice using positive ES cell clones

4.3.1 Successful generation of germline transmitting chimeras from blastocyst injections

The knock-in ES cells were used for blastocyst injections to produce chimaeras. This procedure was performed by Jim Selfridge and Martha Koerner. Each line underwent 3-5 rounds of injection resulting in 6, 7 and 16 chimeras for ΔN-EGFP, ΔNC-EGFP and ΔNIC-EGFP lines, respectively (Table 4.3.1). Chimaeras are recognisable by coat colour as they are made up of a mixture of C57BL/6J (black) cells from the blastocyst and injected 129/Ola (chinchilla) ES cells (Fig. 4.3.1). Breeding the chimaeras resulted in germline transmission for all three lines.

	ΔN-EGFP	ΔNC-EGFP	ΔNIC-EGFP
ES cell clone	4G9	2G7	3F1
Blastocysts	100	115	130
Transfers into recipient ♀s	9	15	10
Pregnancies	4	12	6
Pups born	14	51	28
Chimaeras	6	7	16
Germline transmission/ total tested	2 ♂/3 ♂	3 ♂/ 7 (6 ♂, 1 ♀)	2 ♂/2 ♂ STOP line (xC57BL/6J): 2 ♂/8 (3 ♂, 5 ♀) – *see Chapter 6*

Table 4.3.1: Production of knock-in mice. Notes from blastocyst injections (totals of all sessions for each line) and outcomes of chimaera test matings.

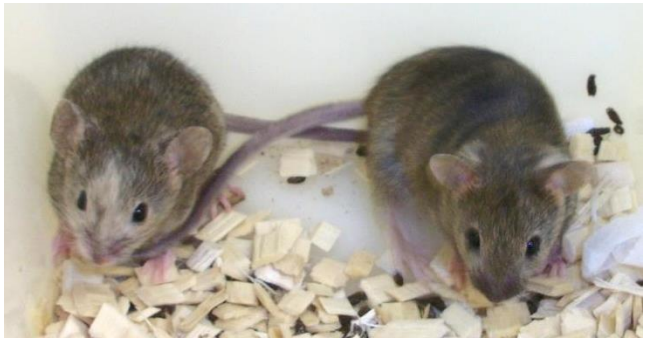


Figure 4.3.1 Chimaeras
Two of the ΔNC-EGFP chimaeras
Photograph by Jim Selfridge

4.3.2 Breeding program to generate hemizygous males for phenotypic and biochemical analysis

To reduce the number of ES cell passages, the floxed NeoSTOP cassette in intron 2 was retained in the cells used for blastocyst injections, and was instead deleted by crossing with mice that ubiquitously express Cre recombinase. This was done at the first stage: breeding chimaeric males to homozygous *CMV-Cre* females (Fig 4.3.2A; Table 4.3.1). Only female pups produced in this cross will carry the truncated *Mecp2* allele, located on their chimaeric father's X chromosome. These females will also inherit the *CMV-Cre* gene (which is also X-linked) from their mother. Ubiquitous expression of *CMV-Cre* in early development will result in the excision of the floxed NeoSTOP cassette. These heterozygous females will therefore express the truncated *Mecp2* alleles. Not all female pups produced in this cross will carry the desired allele: the germline transmitting chimaeras produce sperm from both their C57BL/6J and introduced 129/Ola cells (Fig. 4.3.2A). Germline transmitted female pups are

recognisable by their brown coat colour, as they inherit chinchilla coat colour genes from their father (129/Ola) and black coat colour genes from their mother (C57BL/6J).

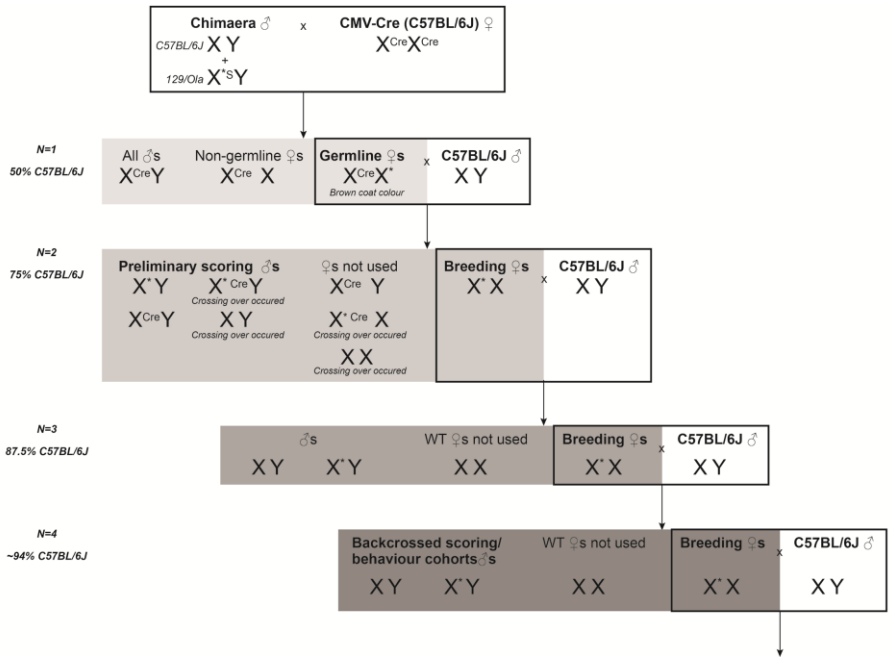


Figure 4.3.2A: Knock-in mouse breeding programme
Diagram of crosses and genotypes of offspring produced at each stage. The knocked-in truncated allele is denoted by *, the NeoSTOP cassette by 'S', and *CMV-Cre* by 'Cre'. Breeding pairs are indicated by black rectangles. Each generation is labelled by the number of times is has been backcrossed (N) and the resulting percentage of C57BL/6J background.

These heterozygous females were backcrossed with C57BL/6J males to produce the first hemizygous males carrying the truncated alleles. Some animals also carried the *CMV-Cre* transgene, due to crossing over in meiosis in their mothers (Fig. 4.3.2A). Both *CMV-Cre* negative and *CMV-Cre* positive knock-in and wild-type males were used in preliminary scoring cohorts for the two more severe mutations: ΔNC-EGFP and ΔNIC-EGFP. Heterozygous female pups that didn't carry the *CMV-Cre* transgene were used for further backcrossing with C57BL/6J males. Generation N=4 (94% C57BL/6J) was used for extensive phenotypic analysis: animals were divided into two cohorts, one for scoring and one for behavioural testing. This is consistent with previous studies in the lab (Brown, Selfridge et al., 2016).

Southern blot analysis of genomic DNA purified from the brains of these knock-in mice show that these lines carry the desired allele after successful Cre-mediated recombination (Fig. 4.3.2B-D).

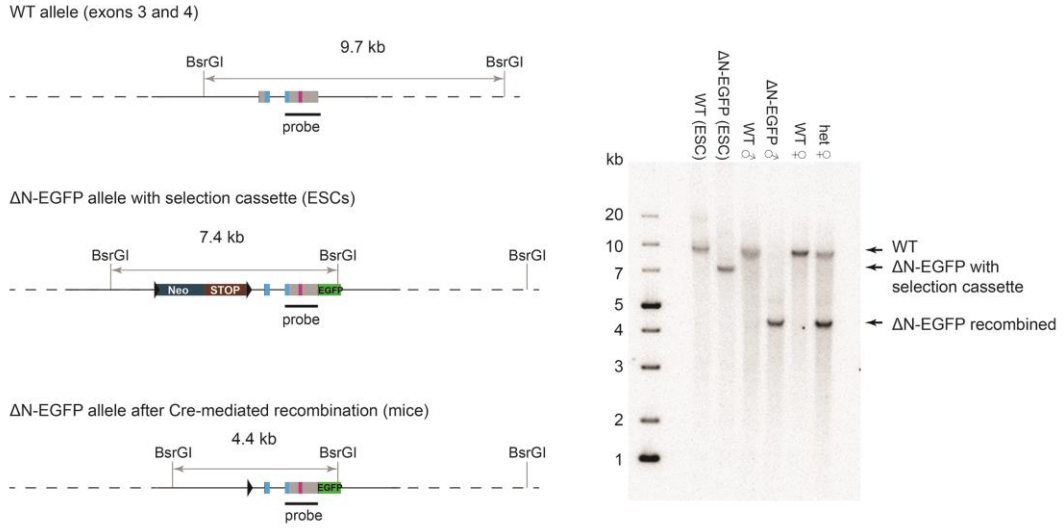


Figure 4.3.2B: ΔN-EGFP knock-in mice carry the Cre-recombined allele
(Left) Genomic DNA sequences of the WT locus, the targeted NeoSTOP ΔN-EGFP allele, and the Cre-recombined allele. Diagrams show maps of the restriction sites of the enzyme used in Southern analysis, BsrGI. An internal probe, recognising exon 4 was used, drawn as a black line. (Right) Southern blot of genomic DNA purified from ES cells and mouse brains.

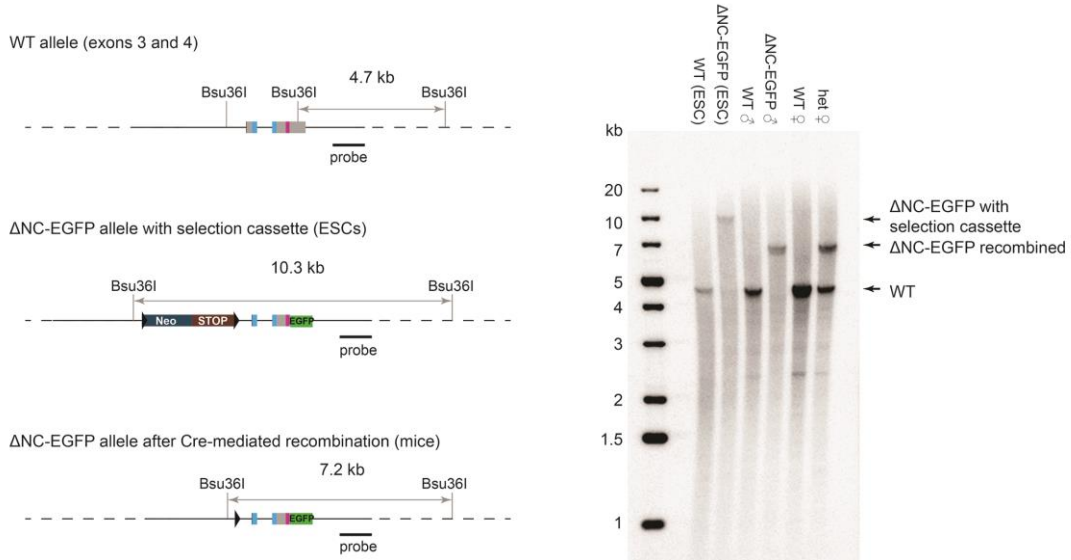


Figure 4.3.2C: legend on next page

Figure 4.3.2C: ΔNC-EGFP knock-in mice carry the Cre-recombined allele
(Left) Genomic DNA sequences of the WT locus, the targeted NeoSTOP ΔNC-EGFP allele, and the Cre-recombined allele. Diagrams show maps of the restriction sites of the enzyme used in Southern analysis, Bsu36I. An internal probe, recognising the 3' homology arm was used, drawn as a black line. (Right) Southern blot of genomic DNA purified from ES cells and mouse brains.

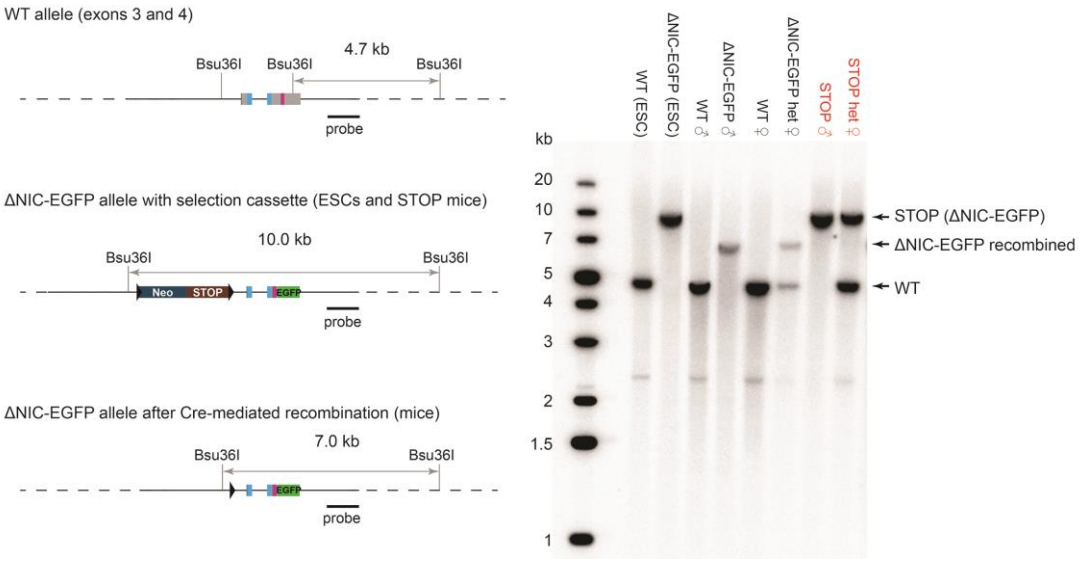


Figure 4.3.2D: ΔNIC-EGFP knock-in mice carry the Cre-recombined allele
(Left) Genomic DNA sequences of the WT locus, the targeted NeoSTOP ΔNIC-EGFP allele, and the Cre-recombined allele. Diagrams show maps of the restriction sites of the enzyme used in Southern analysis, Bsu36I. An internal probe, recognising the 3' homology arm was used, drawn as a black line. (Right) Southern blot of genomic DNA purified from ES cells and mouse brains. *STOP mice retain the NeoSTOP cassette – see Chapter 6*

4.4 Expression levels of truncated alleles in knock-in mice

The first step in the biochemical characterisation of these mouse lines was to look at the expression levels of the truncated MeCP2 proteins. The level of MeCP2 is very important, as lower levels (~50%) result in mild RTT-like symptoms (Kerr et al., 2008; Samaco et al., 2008). Approximately equivalent expression levels of the different truncated proteins is required for proper comparison of their function *in vivo*.

4.4.1 Analysis of protein levels by western blotting and flow cytometry

The protein level in the brains of the knock-in mice was determined using two independent assays: western blot analysis and flow cytometry. Due to the location of the epitopes detected by antibodies, only the ΔN-EGFP mice could be compared directly to their untagged littermates, using an antibody that recognises the C-terminus of MeCP2. Western blot analysis on whole brain crude extract shows that ΔN-EGFP protein levels are not significantly different from wild-type controls (Fig. 4.4.1A).

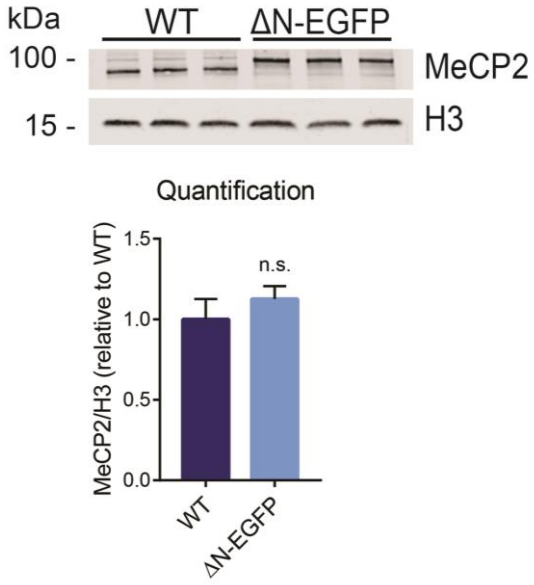


Figure 4.4.1A: Quantification of ΔN-EGFP protein levels using western blotting
(Upper) Western blot of crude whole brain extract derived from hemizygous male ΔN-EGFP and wild-type controls mice (three biological replicates of each). MeCP2 was detected with a C-terminal antibody (Sigma 6818). Histone H3 was used as a loading control. 'm' = size marker.
(Lower) Quantification of protein levels by western blotting. The samples were run on two blots and the values were combined (protein levels were calculated relative to WT controls). Graph shows mean and standard deviation. Protein levels were not significantly different: p=0.215, t-test.

All three proteins were quantified with a GFP antibody, using WT-EGFP mice as a comparator. The levels of ΔN-EGFP were found to be ~74% of WT-EGFP controls (Fig. 4.4.1B). Taking account of the slightly elevated levels of WT-EGFP protein in these mice (159% as much protein as untagged controls) (Brown, Selfridge et al. 2016), ΔN-EGFP is close to wild-type levels (~118%), which is consistent with the result obtained with the MeCP2 antibody. The levels of ΔNC-EGFP quantified from western analysis are slightly lower: ~45% of WT-EGFP, equivalent to ~72% of untagged wild-type. The most severe truncation, ΔNIC-EGFP, has dramatically low

protein levels when analysed in this way: ~18% of WT-EGFP, equivalent to ~29% of untagged wild-type.

Although MeCP2 mutations frequently affect protein levels in knock-in mice (Goffin et al., 2012; Brown, Selfridge et al., 2016), it is surprising that the levels correlate with protein size. To overcome inaccuracies of quantifying proteins of different sizes by western blotting, I exploited the endogenous fluorescent EGFP tag for flow cytometry analysis. Nuclei were purified from whole brains and EGFP fluorescence was used to measure protein levels. Using this method, all three truncated proteins have higher levels relative to WT-EGFP than in the western blot analysis, and protein levels do not correlate with protein size (Fig. 4.4.1C). The protein level in ΔN-EGFP mice is not significantly different from WT-EGFP controls and the ΔNC mice have slightly higher protein levels: 127% of WT-EGFP mice (** p=0.003, t-test). Although the levels of ΔNIC-EGFP are still lower than the other lines, the difference is less dramatic: 48% of WT-EGFP mice (*** p=0.0002, t-test) instead of ~18%.

It has previously been shown that MeCP2 is most highly expressed in neurons (Skene et al., 2010), and neuronal-only expression of *Mecp2* can rescue the null phenotype (Luikenhuis et al., 2004). Additionally, deletion of *Mecp2* in various neuronal subtypes (e.g. GABAergic neurons) results in RTT-like symptoms (Chao et al., 2010; reviewed by Guy et al., 2011). Altered protein levels in neurons are therefore most likely to have the greatest effect on the phenotype of the knock-in mice presented in this study. Flow cytometry analysis enables the separation of neuronal nuclei by staining with an Alexa Fluor 647-conjugated antibody that recognises the neuronal marker, NeuN. As expected, nuclei in the neuronal subpopulation have higher levels of MeCP2 than the whole brain average (Fig. 4.4.1C). As with the total nuclei, protein levels in ΔN-EGFP mice are not significantly different from WT-EGFP controls, and ΔNC-EGFP mice have very slightly elevated protein levels: 117% of WT-EGFP controls (* p=0.014). The neuronal levels of ΔNIC-EGFP were slightly more reduced: 40% (*** p=0.0001) of WT-EGFP controls.

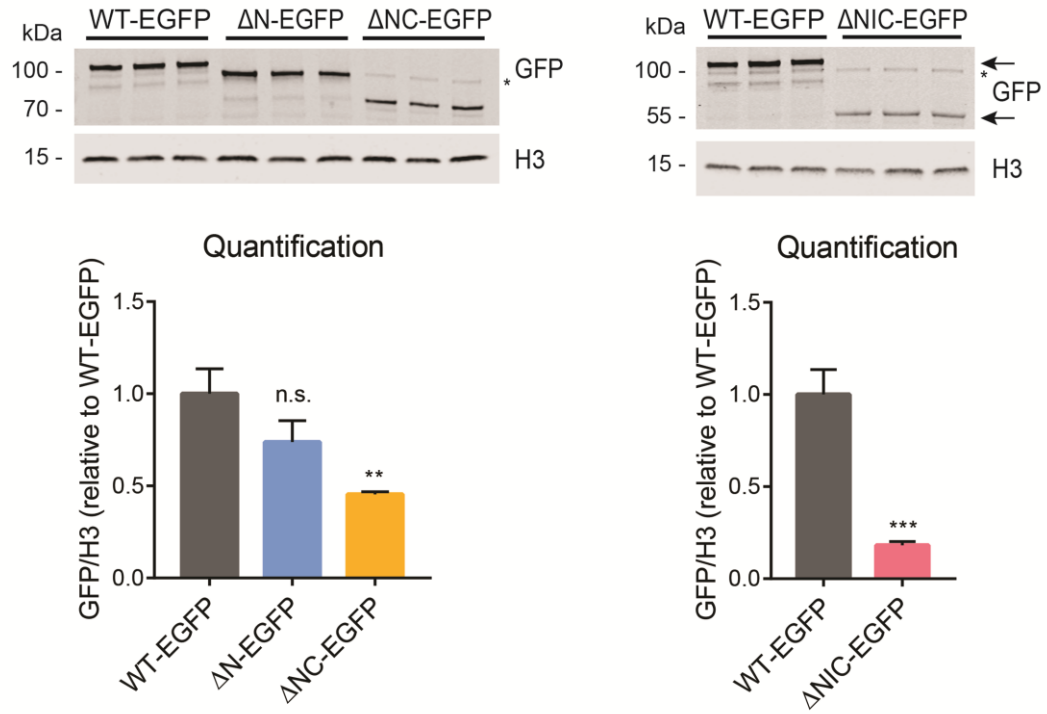


Figure 4.4.1B: Quantification of protein levels using western blotting
(Upper) Western blots of crude whole brain extract derived from hemizygous male WT-EGFP, ΔN-EGFP, ΔNC-EGFP and ΔNIC-EGFP mice (three biological replicates of each). MeCP2 was detected with a GFP antibody. Histone H3 was used as a loading control. 'm' =size marker, * = non-specific band.
(Lower) Quantification of protein levels by western blotting. The samples were run on two blots and the values were combined (protein levels were calculated relative to WT-EGFP controls). Graphs show mean and standard deviation. Statistical analysis was performed using t-tests: ΔN-EGFP 74% not significant p=0.063, ΔNC-EGFP 45% ** p=0.002, ΔNIC-EGFP 18% *** p=0.0005.

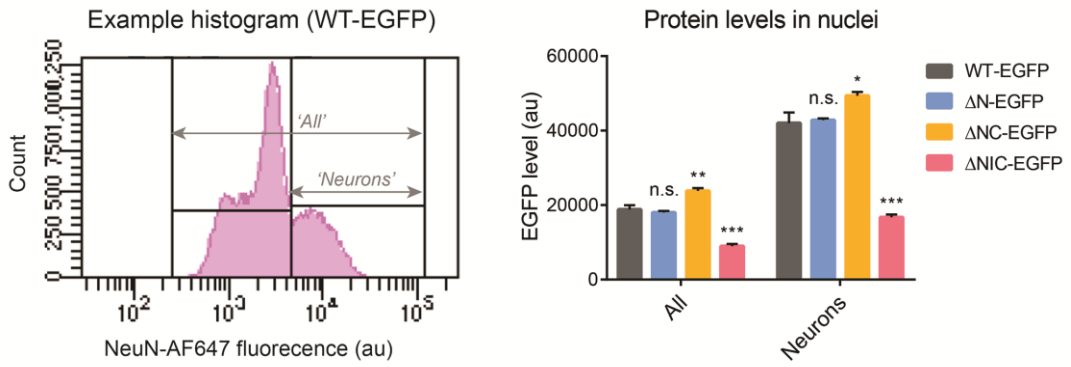


Figure 4.4.1C: Quantification of protein levels using flow cytometry
(Left) Example histogram of NeuN-Alexa Fluor 647 fluorescence, used to obtain the 'neuronal' population (high NeuN). (Right) Quantification of protein levels by flow cytometry. Nuclei were purified from the brains of knock-in mice (three biological replicates of each genotype) and stained with anti-NeuN-Alexa Fluor 647. Protein level was determined by EGFP fluorescence in all nuclei (n=50,000), and the neuronal (high NeuN) subpopulation (n>8,000). Graph shows mean and standard deviation (au = arbitrary units). Statistical analysis was performed using t-tests to compare each line to WT-EGFP controls. All nuclei: ΔN-EGFP not significant p=0.338, ΔNC-EGFP 127% ** p=0.003, ΔNIC-EGFP 48% *** p=0.0002. Neuronal nuclei: ΔN-EGFP not significant p=0.672, ΔNC-EGFP 117% ** p=0.014, ΔNIC-EGFP 40% *** p=0.0001.

4.4.2 Analysis of mRNA levels by quantitative PCR

To determine whether the low level of protein in the ΔNIC-EGFP mice was due to reduced gene expression or protein instability, I performed quantitative PCR (qPCR) on cDNA prepared from RNA purified from the brains of these mice. Here, I was able to compare the knock-in mice directly to their wild-type littermates, using primers located within the MBD, spanning exon 3. qPCR analysis shows that RNA levels are actually elevated by almost 2-fold in these mice (Fig. 4.4.2). The low protein level is therefore likely to be due to protein instability. The high level of RNA observed may indicate the presence of a negative feedback loop regulating MeCP2 expression.

ΔN-EGFP mice also have elevated levels of RNA, compared to their wild-type littermates: ~1.5 fold (Fig. 4.4.2). This is consistent with other EGFP-tagged *Mecp2* knock-in mice (Brown, Selfridge et al., 2016). However, the levels of RNA in ΔNC-

EGFP mice are not significantly different from their wild-type littermates (Fig. 4.4.2). This is surprising given the higher levels of ΔNC-EGFP protein measured using flow cytometry.

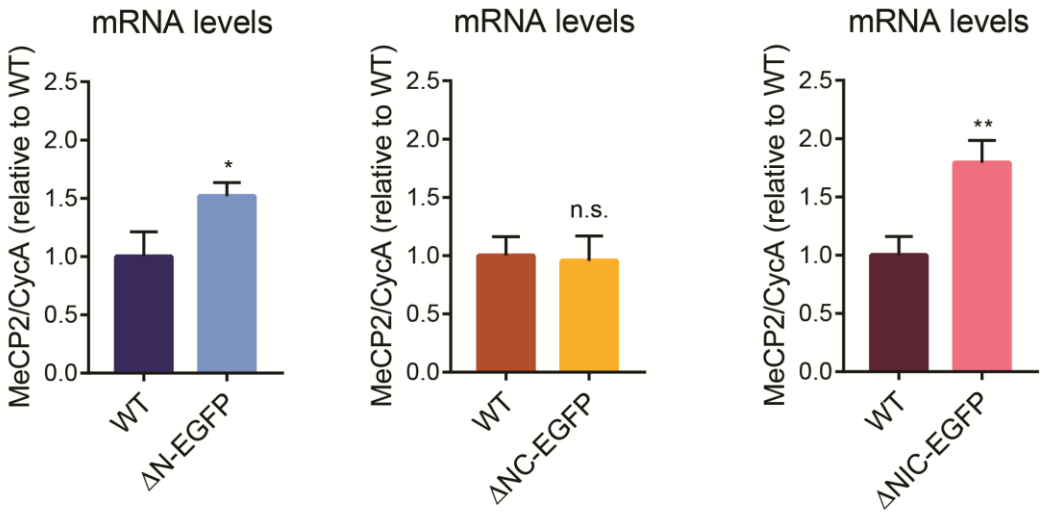


Figure 4.4.2: Quantification of RNA levels by qPCR
RNA was purified from whole brains and converted to cDNA. This was analysed by qPCR using primers located within the MBD that spanned exon 3. This allowed direct comparison of each line to their wild-type littermates (three biological replicates of each). Graphs show the mean and standard deviation, (combined quantifications of ≥3 experiments, each pipetted in triplicate). Statistical analysis was performed using t-tests. ΔN-EGFP: 152% of wild-type * p=0.021 (four experiments), ΔNC-EGFP: n.s. p=0.791 (three experiments), ΔNIC-EGFP: 179% * p=0.005 (five experiments).

In summary, at the RNA level, ΔN-EGFP and ΔNC-EGFP alleles are both expressed in the brain at the same level or slightly higher than the wild-type controls. This results in slightly higher levels of protein for both ΔN-EGFP and ΔNC-EGFP lines - at a level comparable to the WT-EGFP mice. The use of flow cytometry analysis in this study overcame the problem of comparing the levels of proteins of different sizes by western blotting. This technique also facilitated easy quantification of the levels of each truncated protein in the neuronal nuclei. The levels of ΔN-EGFP and ΔNC-EGFP were comparable to WT-EGFP in the neuronal subpopulation. The most severely truncated allele, ΔNIC-EGFP, is present at a lower level than the others. It is likely that its level is underestimated by western blot analysis and that is it closer to

the quantification obtained by flow cytometry: 48% of WT-EGFP in whole brain. Taking the elevated expression of WT-EGFP into account (159% of wild-type controls using western blotting; Brown, Selfridge, et al., 2016), its levels could be as high as 76% of wild-type. Its relative expression level in the neuronal nuclei is only slightly lower: 40% of WT-EGFP. Finally, quantification of ΔNIC-EGFP mRNA levels by qPCR shows that it is overexpressed to around 2-fold compared to wild-type controls. This rules out transcriptional silencing as the cause of low ΔNIC-EGFP protein levels and suggests that this protein may have reduced stability.

4.4.3 All three truncated proteins can interact with HDAC3, the catalytic component of the NCoR/SMRT complex, in the brain

I have previously shown that all three truncated proteins can interact with endogenous NCoR/SMRT co-repressor complex components when over-expressed in HeLa cells (section 3.3.1). To determine whether these proteins retained this activity when expressed in the brains of knock-in mice, I performed immunoprecipitation experiments on brain nuclear extracts obtained from the novel knock-in mouse lines. The two pre-existing knock-in mouse lines, WT-EGFP and R306C-EGFP (Brown, Selfridge et al., 2016), were used as positive and negative controls, respectively. All MeCP2 proteins were successfully immunoprecipitated with GFP-TRAP beads from the nuclear extracts, shown by western blot probed with a GFP antibody (Fig. 4.4.3). The catalytic component of the NCoR/SMRT complex, HDAC3, was co-immunoprecipitated with all proteins except the negative control, R306C-EGFP. The level of binding, however, varied between proteins: ΔN-EGFP appears to bind HDAC3 at WT levels, ΔNC-EGFP pulls down slightly less HDAC3, and ΔNIC-EGFP pulls down very little (Fig. 4.4.3). This result is in contrast to observations in HeLa cells (Fig. 3.3.1).

To investigate this result further, I followed the levels of the MeCP2 proteins throughout the steps of this experiment to determine whether they had been equally extracted and immunoprecipitated. It seemed surprising that the amount of immunoprecipitated protein was very even between samples, contrasting with the

level of these proteins quantified by western blotting using crude brain extract (Fig. 4.4.1B). Although I had calculated a volume of GFP-TRAP beads that should have been in excess, it was a possibility that the beads were saturated. To determine if this was the case, I analysed the unbound fraction by western blotting after incubation with the beads. Only very little MeCP2 remained in the nuclear extracts, indicating complete immunoprecipitation of these proteins (Fig. S9). An alternative explanation was a difference in extraction efficiency between these proteins. Nuclei were treated with benzonase and then incubated in 150 mM NaCl before centrifugation to obtain soluble nuclear extracts that were used as inputs. Insufficient extraction of MeCP2 would result in it being discarded in the pellet at this stage. Western blot analysis of the pellets shows higher levels of WT-EGFP, Δ-EGFP and ΔNC-EGFP than of ΔNIC-EGFP, indicating that this may be the cause of the differences (Fig. S9). Unfortunately, equal protein extraction of these proteins may be impossible to obtain as increasing the salt concentration disrupts the interaction between MeCP2 and the NCoR/SMRT complex. Further work is needed to determine whether ΔNIC-EGFP has reduced 'bridge' function *in vivo*.

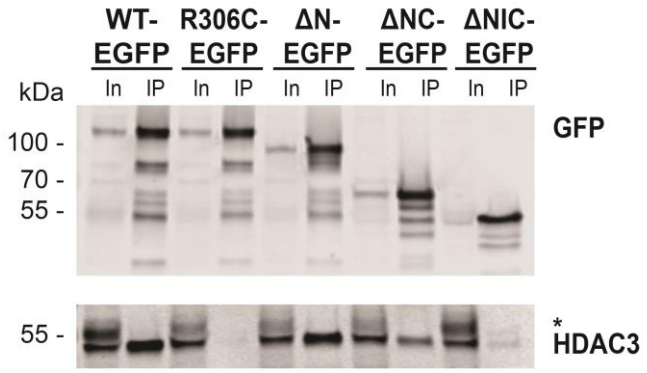


Figure 4.4.3: Truncated MeCP2 proteins can bind HDAC3, the catalytic component of the NCoR/SMRT complex, in the brain
Immunoprecipitation of EGFP-tagged MeCP2 proteins: full length (WT), with the R306C mutation (R306C), and the three truncations (ΔN, ΔNC, ΔNIC), from brain nuclear extracts using GFP-TRAP beads. Western blots show successful immunoprecipitation of the MeCP2 proteins (probed using a GFP antibody), and co-immunoprecipitation with HDAC3. 'In' = input, 'IP' = immunoprecipitate, * denotes the presence of IgG from small amounts of blood in these samples, detected by the anti-mouse secondary antibody. (Representative blots from one of two replicate experiments.)

4.5 Discussion

I have successfully produced knock-in mouse lines expressing each of the three truncated alleles: ΔN-EGFP, ΔNC-EGFP and ΔNIC-EGFP. First, I cloned genomic DNA sequences encoding these proteins into a targeting vector that contained a Neomycin resistance cassette and *Mecp2* homology arms. I then used these vectors to target the endogenous *Mecp2* locus on the single X chromosome in male mouse ES cells. Positive ES cell clones were identified by PCR and Southern blot screening. Of the 192 ΔNC-EGFP clones screened, only one was correctly edited. For the other two alleles, I took advantage of CRISPR/Cas9 technology, increasing the editing efficiency to 2/86 (2.3%) for ΔN-EGFP and 5/92 (5.4%) for ΔNIC-EGFP. One ES cell clone for each line was injected into blastocysts to produce chimaeras. Chimaeras from all three lines produced heterozygous female pups carrying the truncated allele. These were backcrossed onto a C57BL/6J background to produce hemizygous males for biochemical and phenotypic analysis. The two less severe truncations, ΔN-EGFP and ΔNC-EGFP, contain slightly elevated levels of MeCP2 in the brain, comparable with the levels previously quantified in WT-EGFP mice (Brown, Selfridge et al., 2016). The ΔNIC-EGFP protein, which lacks the Intervening region as well as both the N- and C-termini, is present at lower levels in the brains of the knock-in mice: 40-48% of WT-EGFP controls. High levels of ΔNIC-EGFP mRNA suggest that deletion of this region results in reduced protein stability. Lastly, I performed GFP-pulldowns using whole brain nuclear extracts and showed that all three truncated proteins are able to bind the NCoR/SMRT co-repressor complex in the brains of these mice but with varying affinities.

4.5.1 CRISPR/Cas9 technology increased the gene editing efficiency at the Mecp2 locus in ES cells

The conventional targeting method used to produce the ΔNC-EGFP knock-in ES cells resulted in a low frequency of insertion at the *Mecp2* locus (8/192 clones). Furthermore, the entire ΔNC-EGFP sequence was correctly incorporated in only one of these eight clones. I made use of CRISPR/Cas9 technology to boost the targeting

efficiency during subsequent production of ΔN-EGFP and ΔNIC-EGFP knock in ES cell lines. CRISPR/Cas9-mediated cleavage of the *Mecp2* locus dramatically increased the frequency of donor molecule insertion. It did not, however, promote full over partial insertion of the desired sequence. One possible reason for this is that the Cas9 cut site and the Neomycin resistance gene are both located at the 5' end of the donor molecule. Therefore, the selection cassette will be readily inserted upon DNA cleavage, but the resulting G418-resistant clones may not contain the entire desired sequence. Relocation of either the CRISPR target site or the selection cassette to the 3' end of the gene could result in a higher percentage of correctly targeted G418-resistant clones.

An alternative method for selecting positively targeted clones with fluorescent tags was recently developed in the lab (Shah et al., 2016). This study used Fluorescence-Activated Cell Sorting (FACS) to isolate human neuronal progenitor cells that have incorporated an mCherry tag at the C-terminus of *MECP2*. Fluorescent cells were sorted into 96 well plates, and grown up as clonal populations. As neuronal progenitor cells express higher levels of MeCP2 than ES cells, it must first be established whether the EGFP-tagged MeCP2 proteins expressed in ES cells are detectable by FACS before using this method for future ES cell targeting. If possible, this method would bypass the need for the insertion of the floxed selectable marker, which had to be subsequently deleted in the knock-in mice.

Unfortunately, the use of CRISPR/Cas9 for gene editing increases the frequency of random integration or multi-copy insertion of the donor molecules (this study and Schick et al., 2016). Recent studies have used circular rather than linearised donor plasmids to reduce this. Another possible consequence is integration of the CRISPR pX330 plasmid. I performed Southern blot analysis using a probe that is homologous to the Bluescript vector backbone to check whether this had occurred in any of the chosen clones. All three clones used to make the knock-in mice (2G7, 4G9 and 3F1) were negative in this screen (data not shown).

4.5.2 Flow cytometry analysis provides accurate quantification of the levels of the truncated proteins in the brains of knock-in mice

A crucial part of the characterisation of knock-in mouse lines is quantifying the expression level of the mutant proteins. This is usually done using western blot analysis. Brown, Selfridge et al. (2016) used flow cytometry to measure the levels of EGFP-tagged MeCP2 alleles with missense mutations in four knock-in mouse lines, confirming the results they had obtained using western blotting. Unlike the truncated MeCP2 mutants produced in this study, those characterised by Brown, Selfridge et al. (2016) are all the same size. Surprisingly, in this study, estimations of the MeCP2 protein levels in the truncated knock-in mouse lines varied greatly between these two analysis methods, and the values obtained by western blot analysis correlated with protein size. The most likely explanation for this effect is that proteins of different sizes behave differently in the western blot procedure, i.e. they are transferred with difference efficiencies from SDS gels to nitrocellulose membranes. The use of gradient gels (4-20%) and slow low-voltage transfer conditions did not alter this result. I therefore conclude that the levels estimated by flow cytometry are more reliable. To confirm this, western blot analysis could be performed using the nuclei purified for flow cytometry analysis to ensure that the difference between the techniques due to a bias in the western blotting procedure and is not a result of the nuclei-purification method used for flow cytometry.

4.5.3 The Intervening region of MeCP2 between the MBD and the NID may be required to maintain protein stability

The low protein levels but high mRNA levels in mice expressing ΔNIC-EGFP, the smallest truncation missing all three potentially dispensable domains, suggests that this protein is unstable. As protein levels are not reduced in mice expressing ΔNC-EGFP, missing only the N- and C-terminal regions, it is likely that the Intervening region of MeCP2 is required for protein stability. This region contains one of the three ‘unexplained’ RTT-causing missense mutations that lie outside the MBD or NID, P225R. Knock-in mice expressing MeCP2-P225R also have lower levels of

protein: 22% compared to wild-type controls (and wild-type levels of mRNA). Further analysis using a cell culture-based stability assay has confirmed that P225R is unstable (Guy and Alexander-Howden, unpublished). Further work is needed to determine whether ΔNIC-EGFP is unstable or inefficiently translated. The elevated levels of ΔNIC-EGFP mRNA detected in these mice indicated the presence of a possible negative feedback loop increasing its expression. This, however, is not observed in the P225R knock-in mice (Guy, unpublished) or R168X knock-in mice, which also have very low protein levels (Lawson-Yuen et al., 2007). An alternative explanation for the higher ΔNIC-EGFP mRNA levels is increased stability of one or more of the alternatively spliced/polyadenylated transcripts.

4.5.4 ΔNIC-EGFP may have reduced 'bridge' activity

Immunoprecipitation of the truncated proteins from brain nuclear extracts was performed to determine HDAC3 (the catalytic component of the NCoR/SMRT complex) binding. The results suggest the ΔNC-EGFP has a slightly reduced affinity for HDAC3, and additional deletion of the Intervening region in ΔNIC-EGFP results in very weak binding. This result contrasts with those obtained using HeLa cells (see Chapter 3). Further investigation of this is required to enable proper quantification of binding affinity. Additionally, the binding of other members of the NCoR/SMRT complex needs to be assessed - particularly TBL1XR1/TBL1X, which binds directly to MeCP2 (Lyst et al., 2013). Experimental approaches to determine 'bridge' function of these proteins are discussed in Chapter 7.

4.5.5 Concluding remarks

In conclusion, I have successfully produced knock-in mouse lines from targeted ES cells for all three of the truncated *Mecp2* alleles. The two less severe alleles: ΔN-EGFP (missing only the N-terminus) and ΔNC-EGFP (missing both the N- and C-termini) are expressed at slightly elevated levels in the brain, comparable to the expression of WT-EGFP protein. As the WT-EGFP mice are essentially phenotypically normal, the slightly elevated levels of these proteins should not affect

the phenotype of these mice. My results suggest that additional deletion of the Intervening region in the ΔNIC-EGFP protein results in instability and reduced binding to HDAC3. Unfortunately, the lower ΔNIC-EGFP protein levels and reduced 'bridge' function could make it harder to address the bridge hypothesis as it will be impossible to determine whether any RTT-like symptoms detected in these mice are the result of the loss of additional functions that require the Intervening region or due to these factors.

Chapter 5 – Phenotypic analysis of mice expressing truncations of MeCP2

5.1 Introduction and aims

In this chapter, I will describe the phenotypes of all three truncated *Mecp2* knock-in mouse lines. Hemizygous males were used for all phenotypic analysis as they provide a more sensitive assay for protein function. Hemizygous male mouse models lacking *Mecp2* or expressing mutant versions that cause RTT in patients have been widely studied (e.g. Guy et al., 2001; Goffin et al., 2012; Baker et al., 2013; Brown, Selfridge et al., 2016). They display a severe phenotype, with rapid symptom progression from 4 weeks of age, resulting in premature death. The mouse models expressing RTT-causing MeCP2 mutant proteins mirror the severity spectrum in patients (Baker et al., 2013; Brown, Selfridge et al., 2016). Hemizygous males have also been studied to detect more subtle phenotypes caused by milder mutations in MeCP2, including A140V (which causes intellectual disability in patients) and phospho-abolishing mutants (e.g. Jentarra et al., 2010; Tao et al., 2009; Cohen et al., 2011). This means that characterisation of knock-in mouse models expressing MeCP2 mutants is a semi-quantitative assay of these proteins' function. Here, I present the phenotypic analysis of the truncated MeCP2 mouse lines using an established scoring system and commonly used behavioural tests. Overall, this analysis gives a robust assessment of the phenotypes of these lines. Mice were backcrossed onto a 94% C57BL/6J background for phenotypic analysis to allow comparison between the novel lines and to other *Mecp2*-mutant mice in previous studies (Brown, Selfridge et al., 2016). Preliminary outbred cohorts of the two more severe truncations, ΔNC-EGFP and ΔNIC-EGFP, were scored in addition. Both backcrossed and preliminary cohorts consist of mutant mice and wild-type littermates, which serve as controls. These wild-type controls express the endogenous full-length protein without an EGFP tag.

5.2 Phenotypic scoring using a well-established system

The phenotype of the mice was assessed over a period of one year using a well-established scoring system developed in the lab to record the presence and progression of RTT-like symptoms in mouse models. This system was based on SHIRPA screening of *Mecp2*-null*l* mice and was first published by Guy et al. (2007). Mice were scored in six categories: spontaneous activity, gait, hind-limb clasping, tremor, abnormal breathing and general appearance. Mice received a score between 0 and 2 for each category, where 0 = as wild-type, 1 = present, and 2 = severe. Intermediate scores of 0.5 and 1.5 were also used in all categories except hind-limb clasping (Cheval, Guy et al., 2012). The scores in each category were added together to give the aggregate symptomatic score for each animal. Loss of MeCP2 function results in reduced spontaneous activity; abnormal gait where hind legs are splayed; hind limb clasping of one (score = 1) or both (score = 2, or 1 if partial) legs towards the body; development of a tremor; development of breathing problems; and degradation of general condition including dull coat, hunched posture, and dull or crusty eyes. This system is quick (2-3 minutes per animal), non-invasive, and can be repeated multiple times with the same animals. Aggregate scores for each animal were averaged to show symptom progression over time for each mouse line. Animals were culled if they reached a score of 2 in any of the last three categories: tremor, breathing problems or general condition. The mice in the scoring cohorts were also weighed as several *Mecp2*-mutant lines have altered body weight (Guy et al., 2001; Brown, Selfridge et al., 2016). Additionally, symptom progression is often combined with weight loss (Guy et al., 2001). Mice were culled if they lost >20% of their maximum body weight (unless they had been obese). Animals were caged together with their wild-type littermates and scored blind. Scoring began soon after weaning, at 4 weeks of age, when symptoms begin to develop in *Mecp2*-null animals.

5.2.1 ΔN-EGFP mice are indistinguishable from their wild-type littermates

The ΔN-EGFP mice, expressing the least severe of all the three truncations - missing only the N-terminal region (Fig. 3.1) - are viable, healthy and fertile. A backcrossed

(94% C57BL/6J) cohort of ten ΔN-EGFP animals and ten wild-type littermates was scored every four weeks for 52 weeks. Over this period, all ten ΔN-EGFP animals survived and were indistinguishable from their wild-type littermates. Overall, all animals of both genotypes received very low total scores: up to a maximum of 2.5 (except at 52 weeks when three ΔN animals scored more) (Fig. 5.2.1i). This was mainly due to mild gait defects or reduced activity from 4-12 weeks. Very rarely, both wild-type and ΔN-EGFP animals were given a low score (0.5) for general appearance due to piloerection or irregular breathing, but these symptoms did not persist when these individuals were scored over subsequent months. No tremor or hind-limb clasping was detected in any individuals until the final time point (but these were detected in both mutant and wild-type animals). ΔN-EGFP mice displayed no weight difference from their wild-type littermates (Fig. 5.2.1ii, repeated measures ANOVA, p=0.362).

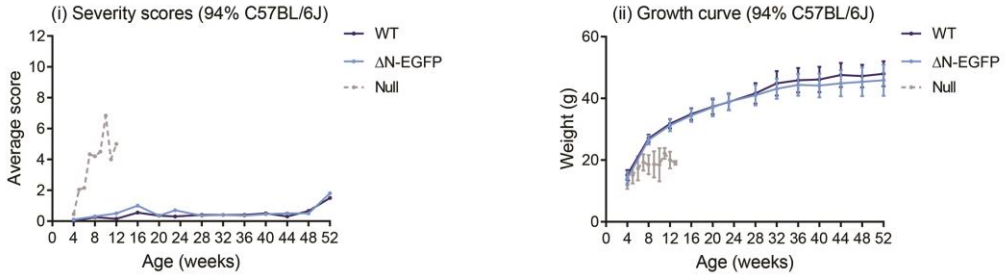


Figure 5.2.1: ΔN-EGFP mice are phenotypically normal
Scoring (i) and growth (ii) curves for hemizygous male ΔN-EGFP mice and their wild-type (WT) littermates. Graphs show mean values (scores) or mean and standard deviation (weights). Mice were scored every 4 weeks from 4-52 weeks. WT n=10, ΔN-EGFP n=10. (Backcrossed) null scores and weight data was obtained by Jacky Guy (Brown, Selfridge et al., 2016). Null animals: scoring n=12, weights n=20. WT and ΔN growth curves were compared using repeated measures ANOVA: not significant p=0.362.

5.2.2 ΔNC-EGFP mice are asymptomatic but have a mild weight phenotype

Mice expressing the second truncated protein, ΔNC-EGFP, missing both the N- and C-termini (Fig. 3.1) are also viable, healthy and fertile. Outbred and backcrossed cohorts of mice were scored over one year and were indistinguishable from their

wild-type littermates throughout this period (Fig. 5.2.2Ai). As described above for the ΔN-EGFP line, some mice received low scores, rarely exceeding a score of 2. The maximum score obtained by ΔNC-EGFP individuals in the backcrossed cohort was 4.5. However, this was also the case for the wild-type littermates, and two wild-types were given scores of 6 and 7.5 - the latter of which was culled due to intestinal blockage at 10 weeks of age. As with the ΔN-EGFP line, these scores were mostly due to mild gait defects or reduced activity from 4-5 weeks. Mild breathing problems (score of 0.5 or 1) were recorded for both wild-type and ΔN-EGFP individuals, but they did not persist over subsequent weeks. Only the animal that developed an intestinal blockage had more severe breathing problems (on the day that it was culled). Mild tremors (score of 0.5 or 1) were more rare: recorded occasionally in animals of both genotypes, and only persisting in two wild-type mice. Hind-limb clasping was absent for the majority of the scoring period, but partial clasping (score of 1) was recorded for a few animals of both genotypes from 40 weeks. One ΔNC-EGFP animal was an exception to this: it began partial clasping at 24 weeks and persisted until the end of the experiment. Some wild-type and ΔNC-EGFP animals received low scores (0.5) for general condition due to slight piloerection, which became more common with increased age (>32 weeks). Again, only the culled wild-type received a higher score (of 2) in this category.

The results for the outbred cohort were very similar, with animals of both genotypes mostly receiving maximum scores of 2 (Fig. 5.2.2Aii). These scores were almost entirely recorded for activity and gait, plus general appearance as the animals got older. Breathing problems and tremors were even less frequent than in the backcrossed cohort, and no hind-limb clasping was observed. All seven ΔNC-EGFP mice survived for the duration of the experiment (one year), but one wild-type littermate was culled due to weight loss at 44 weeks of age. Although some individuals of both genotypes in this cohort carried the *CMV-Cre* transgene, this did not affect their scores (Fig. S10i).

The backcrossed ΔNC-EGFP mice displayed a mild weight phenotype: they were heavier than their wild-type littermates (Fig. 5.2.2Bi). This is not significant at any

individual time point (analysed using t-tests) but is significant overall (repeated measures ANOVA, **** $p<0.0001$). Importantly, the weakness of this phenotype meant that the animals were not distinguishable by size, so the scoring was truly blind. The biological significance of weight phenotypes in *Mecp2*-mutant mice is unclear as it has previously shown to be linked to genetic background (Guy et al., 2001). To support this, no weight phenotype was observed between ΔNC-EGFP mice and their wild type littermates in the outbred (75% C57BL/6J) cohort (Fig. 5.2.2Bii). This was not affected by the presence of the *CMV-Cre* transgene in some individuals in this cohort (Fig. S10ii).

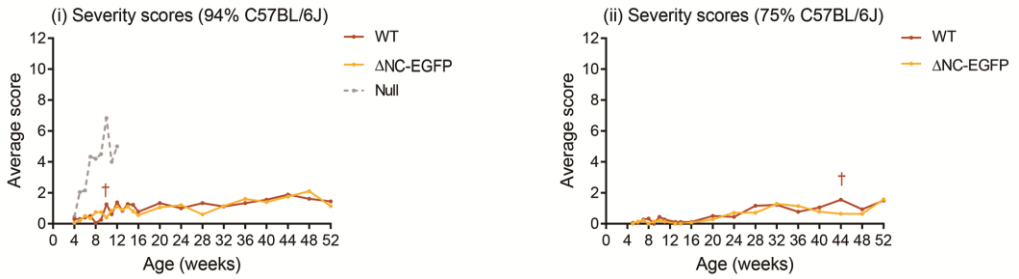


Figure 5.2.2A: ΔNC-EGFP mice are asymptomatic over one year
Scoring of backcrossed (i) and outbred (ii) hemizygous male ΔNC-EGFP mice and their wild-type (WT) littermates. Graphs show mean values. Backcrossed (94% C57BL/6J) mice were scored weekly from 4-16 weeks, and every 4 weeks thereafter up to one year. WT n=10, ΔNC-EGFP n=10. Outbred (75% C57BL/6J) mice were scored weekly from 5-16 weeks (excluding weeks 12 and 15) and every 4 weeks thereafter up to one year. WT n=9, ΔNC-EGFP n=7. † denotes the time points where one wild-type was culled due to weight loss: aged 10 weeks in the backcrossed cohort and aged 44 weeks in the outbred cohort. Backcrossed null scoring data was obtained by Jacky Guy (Brown, Selfridge et al., 2016). Null animals: n=12. (Deaths of null animals are not shown.)

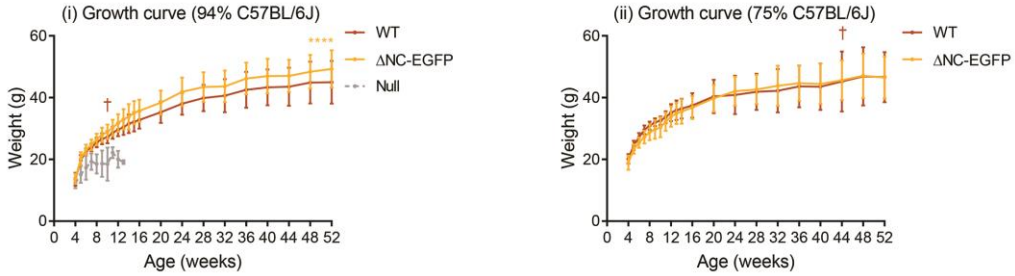


Figure 5.2.2B: ΔNC-EGFP mice have a slight weight phenotype after backcrossing
Growth curves of backcrossed (i) and outbred (ii) hemizygous male ΔNC-EGFP mice and their wild-type (WT) littermates. Graphs show mean and standard deviation. Backcrossed (94% C57BL/6J) mice were weighed weekly from 4-16 weeks, and every 4 weeks thereafter up to one year. WT n=10, ΔNC-EGFP n=10. Outbred (75% C57BL/6J) mice were weighed weekly from 4-16 weeks (excluding week 15) and every 4 weeks thereafter up to one year. WT n=9, ΔNC-EGFP n=7. † denotes the time points where one wild-type was culled due to weight loss: aged 10 weeks in the backcrossed cohort and aged 44 weeks in the outbred cohort. WT and ΔNC growth curves were compared using repeated measures ANOVA: backcrossed cohort **** p<0.0001, outbred cohort not significant p=0.739. Backcrossed null weight data was obtained by Jacky Guy (Brown, Selfridge et al., 2016). Null animals: n=20. (Deaths of null animals are not shown.)

5.2.3 ΔNIC-EGFP mice exhibit mild RTT-like symptoms

Mice expressing ΔNIC-EGFP, the third and most severely truncated protein retaining only the MBD and the NID, are also viable and fertile, but exhibit mild RTT-like symptoms from around 7 weeks (Fig. 5.2.3Ai). Unlike *Mecp2*-nulls or mice carrying RTT-causing mutations, these symptoms do not progress and survival is not affected. The backcrossed cohort was scored weekly or biweekly for one year. The average score of these animals rose slowly to 2.4 at 8 weeks of age, but plateaued at this level (fluctuating around ~2.5). Individual scores throughout this scoring period are mostly below 4.5. The wild-type littermates in this cohort were comparable to those in the ΔN-EGFP and ΔNC-EGFP cohorts described above, mostly only receiving low scores for activity and gait. The higher scores obtained by the ΔNIC-EGFP mice are due to more severe gait defects and partial hind-limb clasping. Several animals

regularly obtained gait scores of 1.5 or 2 from ~8 weeks. Partial clasping (score of 1) was detected from 7 weeks of age; and in four out of ten animals, this progressed to more severe hind-limb clasping (score of 2) from 15, 24, 32 and 44 weeks. One ΔNIC-EGFP animal in this cohort was found dead at 43 weeks of age. The cause of its death is unknown as its symptoms were very mild (never exceeding a score of 2.5), but it had lost 15% of its bodyweight (not quite reaching the 20% threshold where the animal must be culled).

An outbred cohort of ten ΔNIC-EGFP mice and one wild-type littermate was also scored. In this cohort, all ten ΔNIC-EGFP animals survived until one year of age. Their scores increased very slowly over the first 36 weeks, then plateaued around an average of 3.5-4 (Fig. 5.2.3Aii). The slower progression of symptoms in this cohort was due to the later onset of hind-limp clasping: ~18 weeks compared to ~7 weeks. Later, their scores were slightly elevated compared to the backcrossed cohort, due to higher activity and gait scores from 28 weeks of age and higher general condition scores from 34 weeks. It is difficult to say whether this is due to more severe defects in these categories or a drift in the scoring criteria over the course of this study. Activity and general condition (mostly age-related piloerection) are particularly prone to subjectivity. The presence of the *CMV-Cre* transgene in some individuals had no effect on the scores in this cohort (Fig. S10iii).

Both scoring cohorts were weighed over the same period. Backcrossed ΔNIC-EGFP mice were ~40% lighter than their wild-type littermates - significant at all time points from 5 weeks of age, analysed using t-tests (Fig. 5.2.3Bi). The two growth curves were compared using repeated measures ANOVA, giving high significance (**** $p<0.0001$). The weight phenotype was also observed on the outbred background (Fig. 5.2.3Bii), but as there was only one wild-type littermate in this cohort, statistical analysis is not possible. It is important to note that the growth curve of this wild-type animal is comparable to the other wild-types on this background (i.e. wild-types in the ΔNC-EGFP outbred cohort). The weights of the outbred cohort were not affected by the presence of the *CMV-Cre* transgene in some individuals (Fig. S10iv).

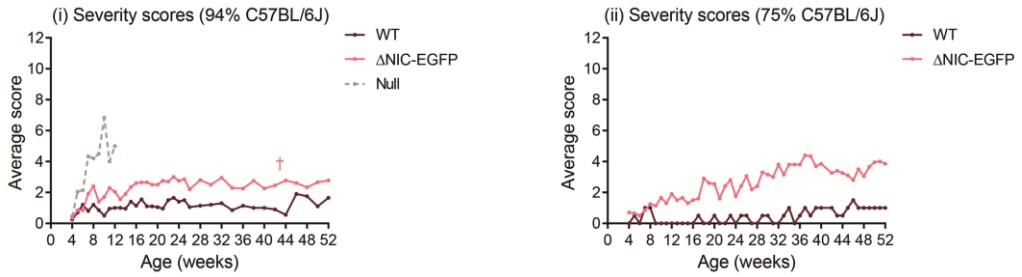


Figure 5.2.3A: ΔNIC-EGFP mice display mild RTT-like symptoms
Scoring of backcrossed (i) and outbred (ii) hemizygous male ΔNIC-EGFP mice and their wild-type (WT) littermates. Graphs show mean values. Backcrossed (94% C57BL/6J) mice were scored weekly from 4-26 weeks, and every 2 weeks thereafter up to one year. WT n=10, ΔNIC-EGFP n=10. Outbred (75% C57BL/6J) mice were scored weekly from 4-52 weeks (excluding week 41). WT n=1, ΔNIC-EGFP n=10. † denotes the time point where one ΔNIC-EGFP mouse from the backcrossed cohort was found dead, aged 43 weeks. Backcrossed null scoring data was obtained by Jacky Guy (Brown, Selfridge et al., 2016). Null animals: n=12. (Deaths of null animals are not shown.)

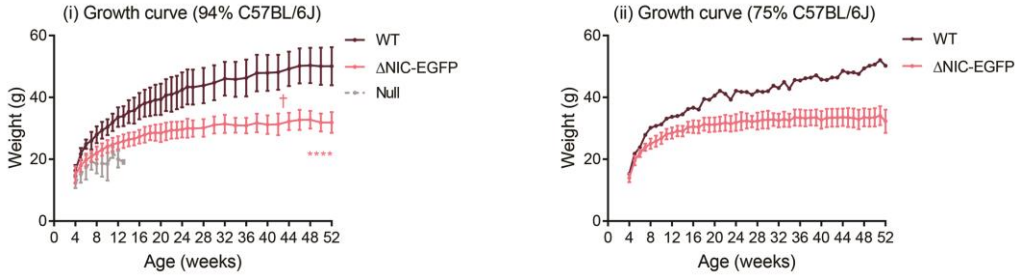


Figure 5.2.3B: ΔNIC-EGFP mice are lighter than their wild-type littermates
Growth curves of backcrossed (i) and outbred (ii) hemizygous male ΔNIC-EGFP mice and their wild-type (WT) littermates. Graphs show mean and standard deviation. Backcrossed (94% C57BL/6J) mice were weighed weekly from 4-26 weeks, and every 2 weeks thereafter up to one year. WT n=10, ΔNIC-EGFP n=10. Outbred (75% C57BL/6J) mice were weighed weekly from 4-52 weeks (excluding week 41). WT n=1, ΔNIC-EGFP n=10. † denotes the time point where one ΔNIC-EGFP mouse from the backcrossed cohort was found dead, aged 43 weeks. WT and ΔNIC growth curves were compared using repeated measures ANOVA: backcrossed cohort **** $p<0.0001$, outbred cohort N/A. Backcrossed null weight data was obtained by Jacky Guy (Brown, Selfridge et al., 2016). Null animals: n=20. (Deaths of null animals are not shown.)

5.3 Behavioural analysis using a series of standard tests

A second cohort of backcrossed mice for each line underwent a series of behavioural tests. The chosen tests are commonly used for mouse models of Rett syndrome, detecting differences in activity, anxiety and motor skills (Goffin et al., 2012; Pitcher et al., 2015; Brown, Selfridge et al., 2016). All three mouse lines underwent the same behavioural analysis: Elevated Plus Maze on day 1; Open Field test on day 2; Hanging Wire test on day 3; and Accelerating Rotarod on days 6-9 (one day of training followed by three days of trials). Prior to the tests, the animals had not been handled except for husbandry. This is particularly important for the Elevated Plus Maze and Open Field test, which monitor exploratory behaviour. These tests are therefore done at the beginning of the experimental regime as the Hanging Wire and Accelerating Rotarod tests involve handing by the experimenter. The Accelerating Rotarod is a very tiring task, so is carried out at the end of the series of behavioural characterisation assays. This analysis was performed when the animals were 20-21 weeks of age.

5.3.1 ΔN-EGFP mice show no phenotype in any test

The Open Field test is used to measure activity and anxiety. It consists of a square arena measuring 50 by 50 cm, which is evenly lit and littered with fresh wood chippings. The animals are left to explore this area for 20 minutes each. Activity is assessed by total distance travelled and time immobile. An immobile episode is defined as lack of movement for a minimum of 10 seconds. ΔN-EGFP mice show no phenotype compared to their wild-type littermates using either of these measures (Fig. 5.3.1A). To assess anxiety, the arena is divided into nine equal squares and the time spent in the middle square is calculated. Wild-type mice tend to explore the perimeter of their enclosure and avoid this middle section (Fig. 5.3.1B). RTT models tend have decreased anxiety, which can be measured by the increase in time spent in this middle region. An alternative measurement for anxiety is Centre:Total ratio of distance travelled, used by Shahbazian et al. (2002b), McGill et al. (2006) and Samaco et al. (2008). The ΔN-EGFP mice show no anxiety phenotype in this test

using either method of measurement (Fig. 5.3.1C). The way that animals behave in the arena can change over the course of the test, with differences between genotypes only apparent in some segments (Shahbazian et al., 2002b; Samaco et al., 2008). For this experiment, reanalysis of the data divided into two segments of 10 minutes gives the same results (Fig. 5.3.1A,C).

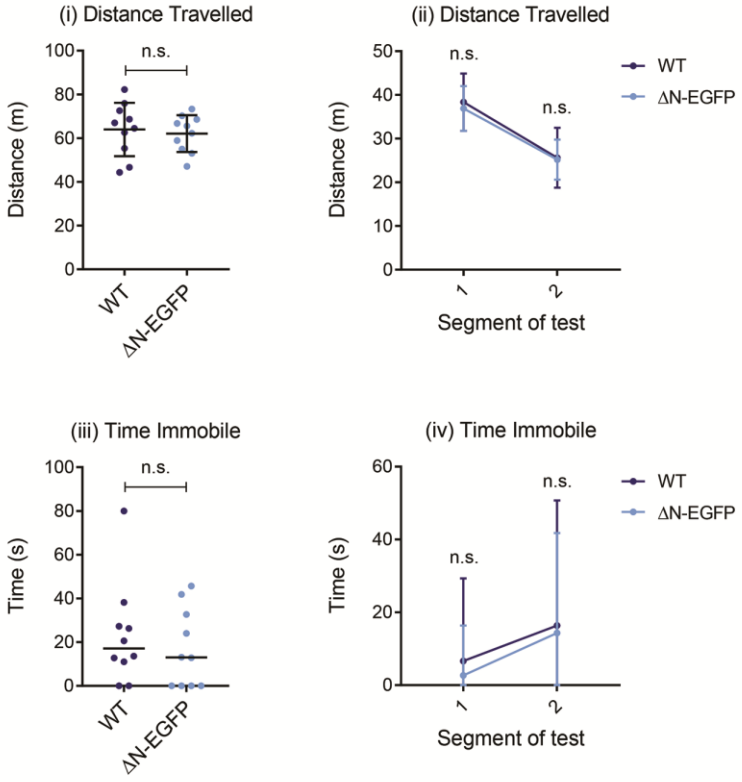


Figure 5.3.1A: ΔN-EGFP mice have no activity phenotype in the Open Field test
(i-iv) Results of the Open Field test performed on ΔN-EGFP mice and their wild-type (WT) littermates. The data was analysed using AnyMaze software. WT n=10; ΔN-EGFP n=10. (i) Total distance travelled. Graph shows individual values and the group mean and standard deviation. Not significant p=0.691 (t-test). (ii) Distance travelled in the two 10 minute segments of the test. Graph shows group mean and SD. Not significant: segment 1 p=0.587, segment 2 p=0.870 (t-tests). (iii) Time spent immobile. Graph shows individual values and the group median. Not significant: p=0.988 (KS test). (iv) Time spent immobile in the two 10 minute segments of the test. Graph shows group median and range. Not significant: segment 1 p=0.988, segment 2 p=0.988 (KS tests). *Note: one immobile episode may be split between the two segments.*

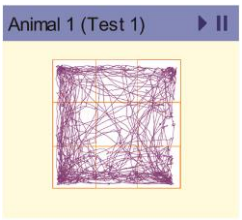


Figure 5.3.1B: Open Field plot
Example track of a wild-type mouse from the Open Field test

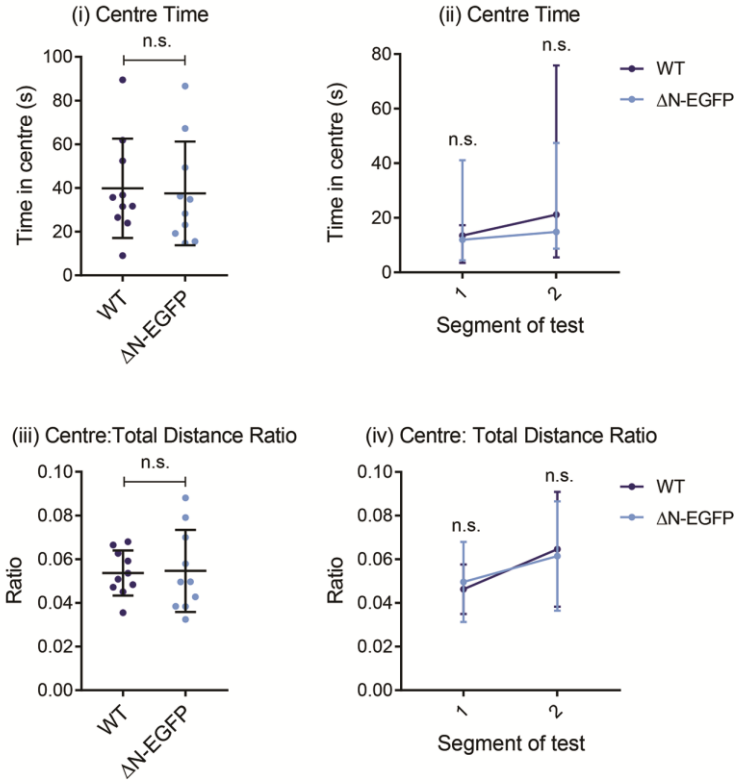


Figure 5.3.1C: ΔN-EGFP mice have no anxiety phenotype in the Open Field Test
(i-iv) Results of the Open Field test performed on ΔN-EGFP mice and their wild-type (WT) littermates. The data was analysed using AnyMaze software. WT n=10; ΔN-EGFP n=10. (i) Centre time: total time spent in the central region. Graph shows individual values and the group mean and standard deviation. Not significant: p=0.822 (t-test). (ii) Time spent in the centre in the two 10 minute segments of the test. Graph shows group median and range. Not significant: segment 1 p=0.759, segment 2 p=0.759 (KS tests). (iii) Centre:Total distance ratio. Graph shows individual values and the group mean and standard deviation. Not significant: p=0.891 (t-test). (iv) Centre:Total distance ratio in the two 10 minute segments of the test. Graph shows group mean and SD. Not significant: segment 1 p=0.634, segment 2 p=0.790 (t-tests).

The Elevated Plus maze is a more commonly used test for anxiety. It consists of a cross-shaped maze with two open well-lit arms, two closed darker arms and a small central region between them. The animals are left to explore this maze for 15 minutes each. Wild-type mice tend to spend the vast majority of time in the darker arms and avoid the anxiety-provoking open arms (Fig. 5.3.1D). RTT mouse models display a decreased anxiety phenotype in the test by spending more time in the open arms than their wild-type littermates. Interestingly, several RTT models show significantly decreased anxiety is this test but not the Open Field test (Brown, Selfridge et al., 2016 and unpublished data), suggesting that this test is more stringent. The ΔN-EGFP mice behave the same as their wild-type littermates in this test (Fig. 5.3.1D).

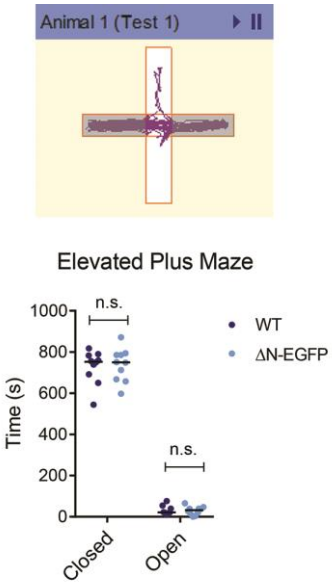


Figure 5.3.1D: ΔN-EGFP mice have no anxiety phenotype in the Elevated Plus Maze
Upper: Example track of a wild-type mouse, which spends more time in the closed arms (shaded).
Lower: Results of the Elevated Plus Maze performed on ΔN-EGFP mice and their wild-type (WT) littermates. The times spent in the open and closed arms were determined for each genotype using AnyMaze software. The graph shows individual values and group medians. WT n=10; ΔN-EGFP n=10. No difference was found between the ΔN-EGFP mice and their WT littermates using Kolmogorov-Smirnov tests: closed arms p=0.988, open arms p=0.759.

The Hanging Wire test measures muscle strength and motor coordination. Animals are placed on a horizontal wire with their forepaws, and the time taken to bring a hind paw to the wire is recorded. Animals are given a maximum of 30 seconds to complete this task. Animals that take longer or fall off the wire are given the maximum score of 30. The test was performed three times for each animal and the mean of the three tests was calculated. RTT mouse models show a reduced performance in this test (Brown, Selfridge et al., 2016). The ΔN-EGFP mice show no phenotype in this test (Fig. 5.3.1E). As seen with other lines (presented below), there

is considerable variation between individuals when this test is performed on older animals, making it difficult to reach statistical significance.

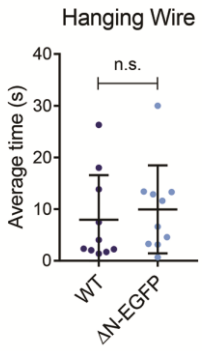


Figure 5.3.1E: ΔN-EGFP mice have no muscle strength phenotype in the Hanging Wire test
Results of the Hanging Wire test performed on ΔN-EGFP mice and their wild-type (WT) littermates. The time taken to bring their hind paw to a horizontal wire was recorded, up to a maximum score of 30 seconds. The graph shows the mean of three tests for each animal and the group mean and standard deviation. WT n=10; ΔN-EGFP n=10. No difference was found between the ΔN-EGFP mice and their WT littermates: p=0.606 (t-test).

Accelerating Rotarod is a test of motor coordination and learning. This test is performed over four days. On the first day, the animals are accustomed to the apparatus in a short training session where they must stay on the track for 30 seconds at its lowest speed (4 rpm). On the subsequent three days, each animal undergoes four trials where the speed is gradually increased from 4 rpm to 40 rpm over 5 minutes. The time taken to fall (latency) is recorded and the average time for each day is calculated. Wild-type mice tend to display motor learning in this test, with increasing latency to fall over the three day period. RTT models show a reduced performance in this test, which becomes significant on day 2 or 3 (Brown, Selfridge et al., 2016). The ΔN-EGFP mice did not perform significantly differently from their wild-type littermates on any of the three days of this test (Fig. 5.3.1F, analysed by KS tests). They do, however, have a trend towards reduced performance by day 3 and did not display learning in this test, unlike their wild-type littermates (analysed using Freidman tests to determine variation over time: WT * p=0.018; ΔN-EGFP n.s. p=0.187).

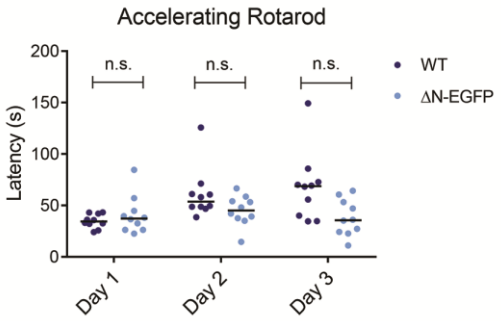


Figure 5.3.1F: ΔN-EGFP mice have no motor deficit on the Accelerating Rotarod
Results of the Accelerating Rotarod performed on ΔN-EGFP mice and their wild-type (WT) littermates. The time taken to fall off the rotating track was recorded, and the mean was taken for four trials for each day. The graph shows mean values for each individual animal and the group medians. WT n=10; ΔN-EGFP n=10. No difference was found between the ΔN-EGFP mice and their WT littermates for any day using Kolmogorov-Smirnov tests: day 1 p=0.759, day 2 p=0.401, day 3 p=0.055. Learning was determined using Friedman tests to determine variation over time: WT * p=0.018, ΔN-EGFP not significant p=0.187.

5.3.2 ΔNC-EGFP mice may have a weak anxiety phenotype

Behavioural analysis of the ΔNC-EGFP backcrossed cohort using the Open Field test showed that these mice have no activity phenotype compared to their wild-type littermates (Fig. 5.3.2A). However, these mice displayed a weak anxiety phenotype, spending more time in the central region of the arena (Fig. 5.3.2B; * p=0.020, t-test). This difference occurred during the first 10 minute segment (* p=0.036, t-test). This phenotype was below the threshold of significance using the alternative measure of anxiety, the Centre:Total distance ratio over the whole test (p=0.089, t-test) and in the two 10 minute segments (segment 1: p=0.137, segment 2: p=0.460, t-tests).

The weak decreased anxiety phenotype illustrated in the Open Field test, was not found in the potentially more stringent Elevated Plus Maze. In this test, ΔNC-EGFP mice resemble their wild-type littermates, both spending very little time in the anxiety-provoking open arms of the maze (Fig. 5.3.2C).

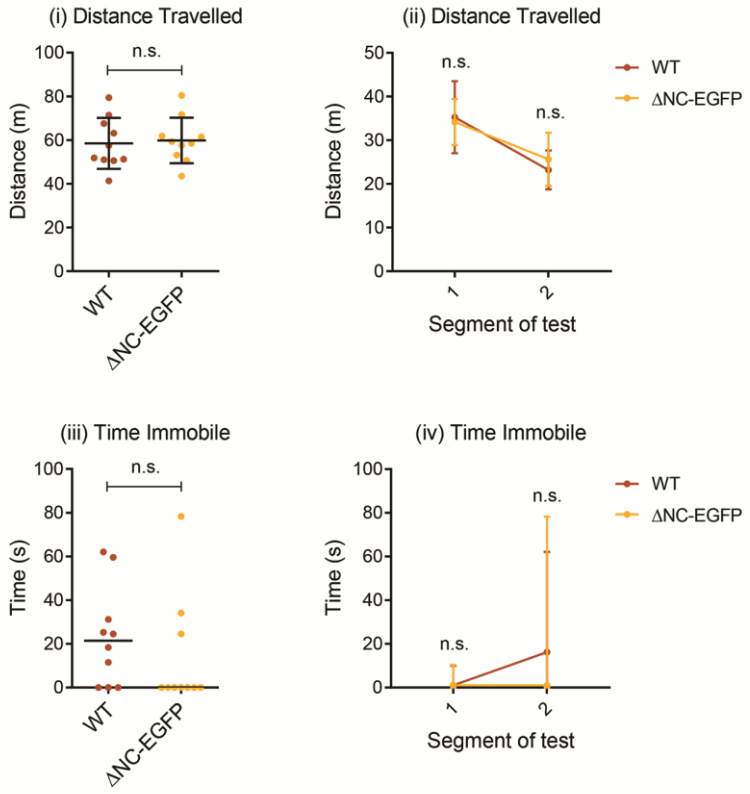


Figure 5.3.2A: ΔNC-EGFP mice have no activity phenotype in the Open Field Test

(i-iv) Results of the Open Field test performed on ΔNC-EGFP mice and their wild-type (WT) littermates. The data was analysed using AnyMaze software. WT n=10; ΔNC-EGFP n=10. (i) Total distance travelled. Graph shows individual values and the group mean and standard deviation. Not significant p=0.791 (t-test). (ii) Distance travelled in the two 10 minute segments of the test. Graph shows group mean and SD. Not significant: segment 1 p=0.724, segment 2 p=0.320 (t-tests). (iii) Time spent immobile (immobile episodes are >10s). Graph shows individual values and the group median. Not significant: p=0.401 (KS test). (iv) Time spent immobile in the two 10 minute segments of the test. Graph shows group median and range. Not significant: segment 1 p=>0.999, segment 2 p=0.401 (KS tests). *Note: one immobile episode may be split between the two segments.*

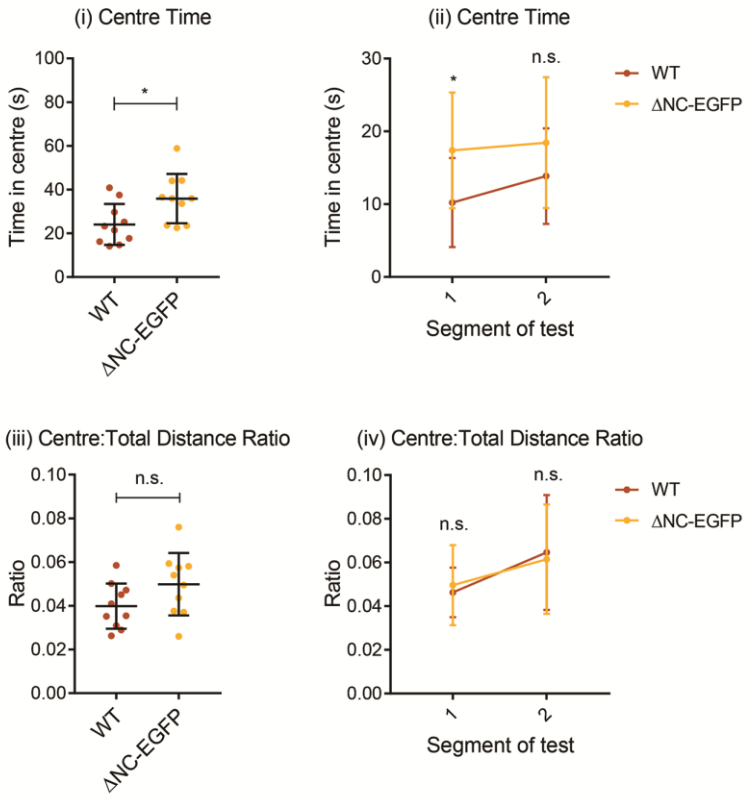


Figure 5.3.2B: ΔNC-EGFP mice may have a mild anxiety phenotype in the Open Field Test

(i-iv) Results of the Open Field test performed on ΔNC-EGFP mice and their wild-type (WT) littermates. The data was analysed using AnyMaze software. WT n=10; ΔNC-EGFP n=10. (i) Centre time: total time spent in the central region. Graph shows individual values and the group mean and standard deviation. * p=0.020 (t-test). (ii) Time spent in the centre in the two 10 minute segments of the test. Graph shows group mean and SD. Only segment 1 is significantly different: * p=0.036, segment 2: n.s. p=0.208 (t-tests). (iii) Centre:Total distance ratio. Graph shows individual values and the group mean and standard deviation. Not significant: p=0.089 (t-test). (iv) Centre:Total distance ratio in the two 10 minute segments of the test. Graph shows group mean and SD. Not significant: segment 1 p=0.137, segment 2 p=0.460 (t-tests).

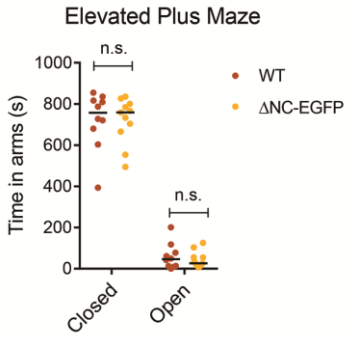


Figure 5.3.2C: ΔNC-EGFP mice have no anxiety phenotype in the Elevated Plus Maze Results of the Elevated Plus Maze performed on ΔNC-EGFP mice and their wild-type (WT) littermates. The times spent in the open and closed arms were determined for each genotype using AnyMaze software. The graph shows individual values and group medians. WT n=10; ΔNC-EGFP n=11. No difference was found between the ΔNC-EGFP mice and their WT littermates using Kolmogorov-Smirnov tests: closed arms p=0.950, open arms p=0.932. (Open arms have a normal distribution so can also be analysed by t-test: p=0.510.)

Muscle strength and motor co-ordination were analysed using the Hanging Wire test and Accelerating Rotarod, respectively. The ΔNC-EGFP mice did not perform significantly differently from their wild-type littermates in either test (Fig. 5.3.2D-E). However, there is a trend to reduced performance in the Hanging Wire test that could be caused by their increased weight compared to their wild-type siblings (Fig. 5.3.2Dii). Notably, in this cohort, neither the wild-type nor the ΔNC-EGFP mice displayed learning on the Accelerating Rotarod (analysed by Freidman test: p=0.353 for wild-type, p=0.549 for ΔNC-EGFP).

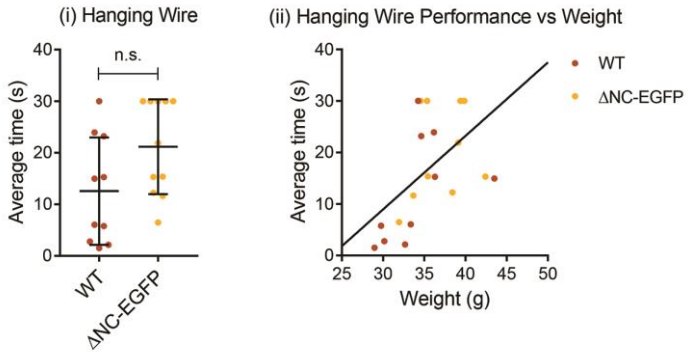


Figure 5.3.2D: legend on next page

Figure 5.3.2D: ΔNC-EGFP mice have no muscle strength phenotype in the Hanging Wire test
(i) Results of the Hanging Wire test performed on ΔNC-EGFP mice and their wild-type (WT) littermates. The time taken to bring their hind paw to a horizontal wire was recorded, up to a maximum score of 30 seconds. The graph shows the mean of three tests for each animal and the group mean and standard deviation. WT n=10; ΔNC-EGFP n=11. No difference was found between the ΔNC-EGFP mice and their WT littermates: p=0.058 (t-test).
(ii) Correlation between Hanging wire performance and body weight. Regression line for all animals (genotypes combined) R^2 * p=0.011.

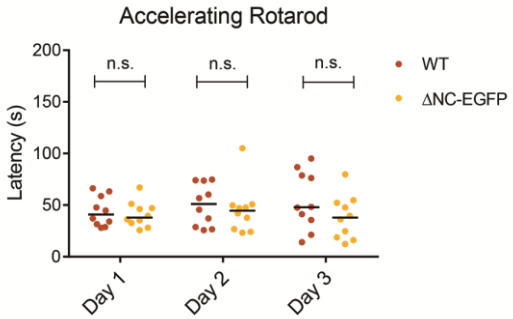


Figure 5.3.2E: ΔNC-EGFP mice have no motor deficit on the Accelerating Rotarod
Results of the Accelerating Rotarod performed on ΔNC-EGFP mice and their wild-type (WT) littermates. The time taken to fall off the rotating track was recorded, and the mean was taken for four trials for each day. The graph shows mean values for each individual animal and the group medians. WT n=10; ΔN-EGFP n=10. No difference was found between the ΔNC-EGFP mice and their WT littermates for any day using Kolmogorov-Smirnov tests: day 1 p=0.988, day 2 p=0.401, day 3 p=0.759. Neither genotype showed learning, analysed using Friedman tests to determine variation over time: WT p=0.353, ΔN-EGFP p=0.549. *Note: one ΔNC-EGFP animal couldn't complete the task and was excluded from the experiment.*

5.3.3 ΔNIC-EGFP mice have slightly decreased anxiety and reduced motor co-ordination compared to their wild-type littermates

Behavioural analysis of a backcrossed ΔNIC-EGFP cohort was carried out at the same age as the other two lines (20-21 weeks). At this age, the backcrossed scoring cohort had an average score of 2.50-2.75. They displayed gait abnormalities and partial hind-limb clasping but no activity phenotype was observed. To investigate the

activity of this knock-in line further, the distance travelled and time spent immobile during the Open Field test were determined using a separate cohort for behavioural testing. The results show that these mice have no activity phenotype compared to their wild-type littermates (Fig. 5.3.3A).

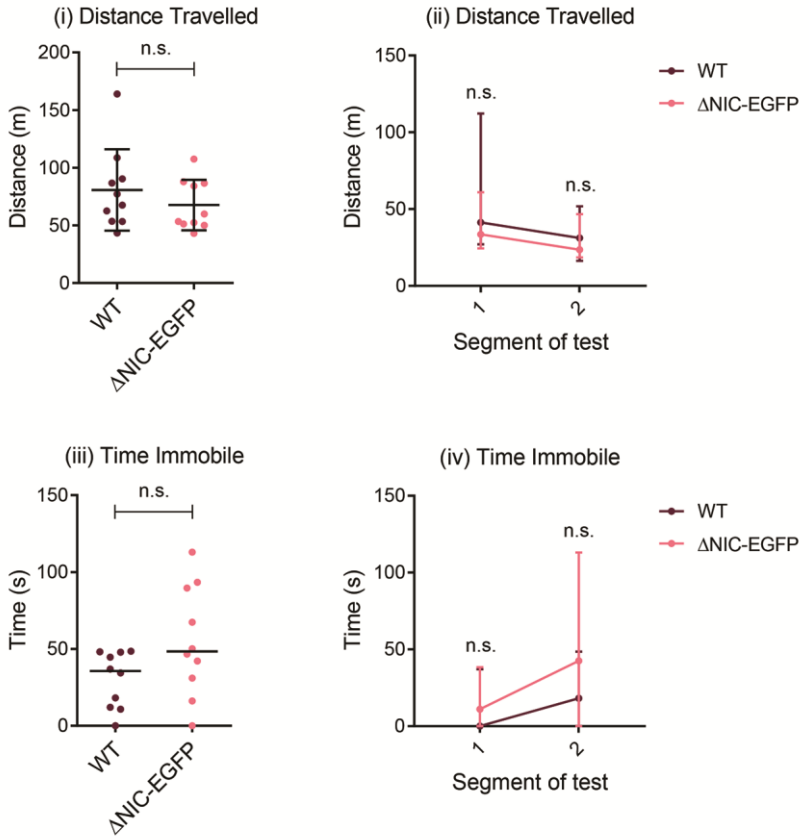


Figure 5.3.3A: ΔNIC-EGFP mice have no activity phenotype in the Open Field Test
(i-iv) Results of the Open Field test performed on ΔNIC-EGFP mice and their wild-type (WT) littermates. The data was analysed using AnyMaze software. WT n=10; ΔNIC-EGFP n=10. (i) Total distance travelled. Graph shows individual values and the group mean and standard deviation. Not significant p=0.333 (t-test). (ii) Distance travelled in the two 10 minute segments of the test. Graph shows group median and range. Not significant: segment 1 p=0.988, segment 2 p=0.401 (KS tests). (iii) Time spent immobile (immobile episodes are >10s). Graph shows individual values and the group median. Not significant: p=0.164 (KS test), or p=0.067 (t-test). (iv) Time immobile in the two 10 minute segments of the test. Graph shows group median and range. Not significant: segment 1: p=0.759, segment 2: p=0.401 (KS tests). *Note: one immobile episode may be split between the two segments.*

As with the other two lines described above, anxiety was tested using both the Open Field test and the Elevated Plus Maze. No anxiety phenotype was found in the Open Field test using either method of measurement: the time spent in the central region (Centre Time) or the Centre:Total Distance Ratio (Fig. 5.3.3B). Separation of the two 10 minute segments of the tests did not affect this result. The ΔNIC-EGFP mice did show a decreased anxiety phenotype in the Elevated Plus Maze (Fig. 5.3.3C). Mice usually display a decreased anxiety phenotype by a reduced avoidance of the open arms and indeed the ΔNIC-EGFP animals tend to spend longer in the open arms than their wild-type littermates, although this was just below the threshold of significance ($p=0.055$, KS test or $p=0.056$, t-test). They did, however, spend a significantly reduced amount of time in the closed arms (** $p=0.003$, KS test or t-test). Further analysis shows that these mice spent significantly longer than their wild-type siblings in the open areas of the maze once the open central region was taken into account (Open arms + Centre ** $p=0.009$ analysed by t-test, or just below the significance threshold analysed by KS test $p=0.055$). The longer time spent in the central region (** $p=0.008$, t-test or * $p=0.015$, KS test) explains why the time spent in the closed arms but not the time spent in the open arms is significantly different. To determine whether the difference in time spent in the open regions was due to increased exploration in these regions or reduced mobility, the distance travelled was analysed. Firstly, the mice showed no difference in the total distance travelled (Fig. 5.3.3Cii), consistent with the results of the Open Field test. Analysis of the distance ratios for each region gives identical results to those obtained with the data on the amount of time spent in the regions of the maze, showing that there was no difference in their activity between the regions (Fig. 5.3.3Ciii). A complication in the interpretation of these results is that although the central region is ‘open’, it must be crossed to switch between all arms. This makes its inclusion in the anxiety-provoking open areas debatable. There is no difference in the total number of arm (closed + open) entries (Fig. 5.3.3Civ), indicating that the increased time and distance travelled in the central region was due to increased exploration of this region rather than increased switching between arms. Another difference between the central region and the open arms that complicates the interpretation of this result is that only the open arms have a sudden drop (i.e. a risk of falling). The related Elevated Zero Maze does not have this

problem, so may have given more informative results. Nevertheless, I believe there is sufficient evidence to conclude that the ΔNIC-EGFP mice have a mild but significant decreased anxiety phenotype.

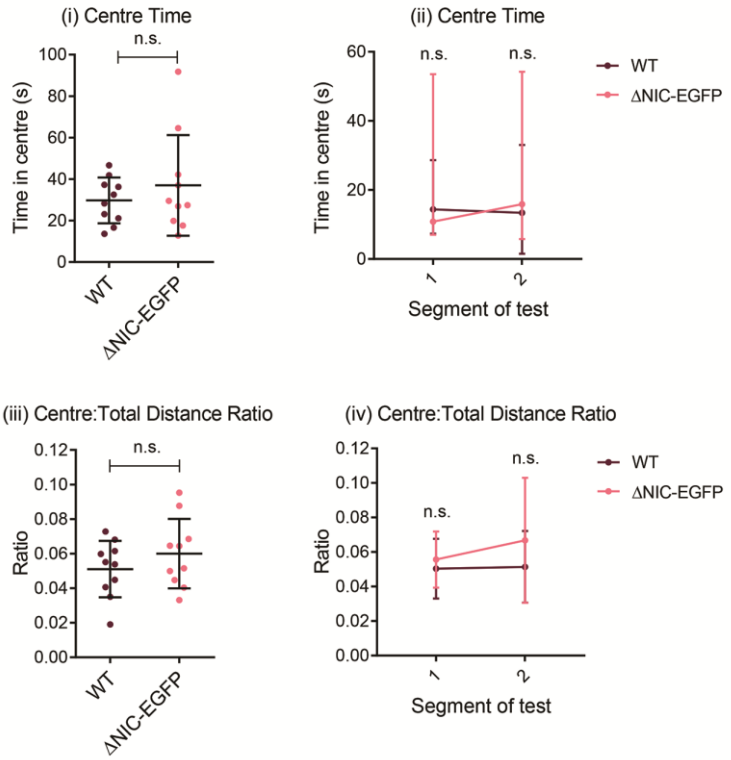


Figure 5.3.3B: ΔNIC-EGFP mice have no anxiety phenotype in the Open Field Test
(i-iv) Results of the Open Field test performed on ΔNIC-EGFP mice and their wild-type (WT) littermates. The data was analysed using AnyMaze software. WT n=10; ΔNIC-EGFP n=10. (i) Centre time: total time spent in the central region. Graph shows individual values and the group mean and standard deviation. Not significant p=0.402 (t-test). (ii) Time spent in the centre in the two 10 minute segments of the test. Graph shows group median and range. Not significant: segment 1 p=0.988, segment 2 p=0.988 (KS tests). (iii) Centre:Total distance ratio. Graph shows individual values and the group mean and standard deviation. Not significant: p=0.289 (t-test). (iv) Centre:Total distance ratio in the two 10 minute segments of the test. Graph shows group mean and SD. Not significant: segment 1 p=0.494, segment 2 p=0.259 (t-tests).

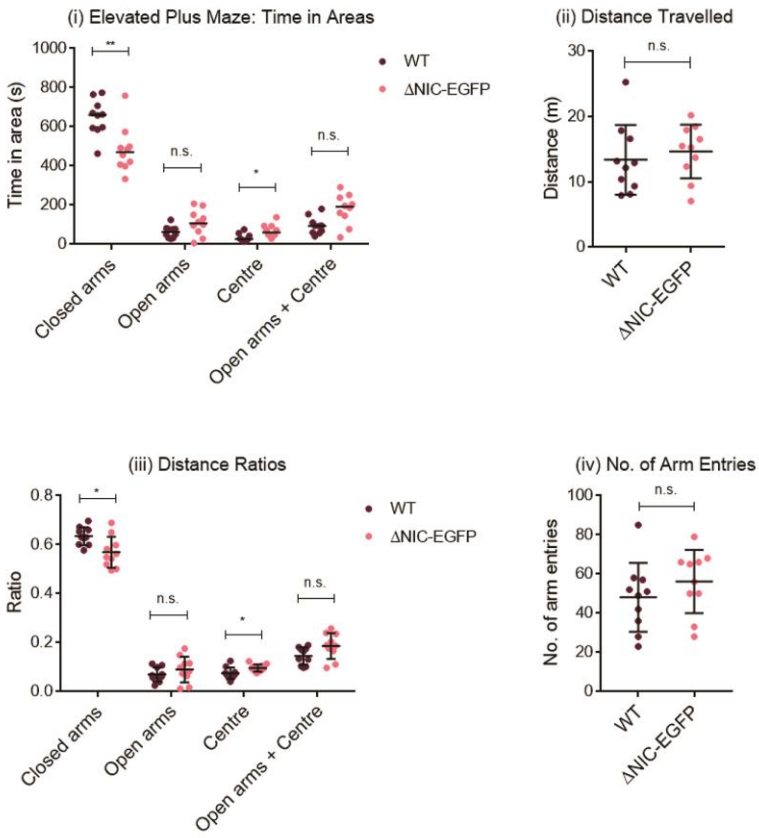


Figure 5.3.3C: ΔNIC-EGFP mice display an anxiety phenotype in the Elevated Plus Maze

(i-iv) Results of the Elevated Plus Maze performed on ΔNIC-EGFP mice and their wild-type (WT) littermates. The data was analysed using AnyMaze software. WT n=10; ΔNIC-EGFP n=10. (i) Time spent in the different areas of the maze. The graph shows individual values and group medians. All datasets fitted a normal curve so can be analysed by t-tests, however KS tests were also carried out for better comparison to results for the other two knock-in lines. The graphs are therefore labelled with symbols reflecting the p values obtained in the KS tests. Results of statistical analysis: Closed arms ** p=0.003 (KS test) ** p=0.003 (t-test), Open arms n.s. 0.055 (KS test) n.s. 0.056 (t-test), Centre * p=0.015 (KS test) ** p=0.008 (t-test), Open arms + Centre n.s. p=0.055 (KS test) ** p=0.009 (t-test). (ii) Total distance travelled. Graph shows individual values and the group mean and standard deviation. Not significant p=0.560 (t-test). (iii) Distance ratios: distance travelled in each area as a proportion of the total distance each individual travelled. Graph shows individual values and the group mean and standard deviation. Statistical analysis using t-tests: Closed arms * p=0.011, Open arms n.s. p=0.304, Centre * p=0.026, Open arms + Centre n.s. p=0.054. (iv) Number of arm entries (Closed + Open). Graph shows individual values and the group mean and standard deviation. Not significant: p=0.303, t-test.

Finally, the ΔNIC-EGFP behavioural cohort underwent motor testing using the Hanging Wire and Accelerating Rotarod tests. As the mice display gait defects at this age, I expected them to have reduced motor abilities in these tests. No difference was observed in the Hanging Wire test (Fig. 5.3.3D; p=0.960, t-test). As seen with the other mouse lines, there was a lot of variation between individuals, which is common when this test is performed on older animals making it difficult to obtain a significant result. The ΔNIC-EGFP mice did, however, display a striking motor defect on the Accelerating Rotarod (Fig. 5.3.3E). Their reduced performance compared to their wild-type siblings became significant on day 3 (** p=0.003, KS test). This result was due to their worsening performance throughout the test (** p=0.003, Friedman test), not just lack of learning. As seen in the ΔNC-EGFP cohort, the wild-type littermates did not display learning in this experiment (p=0.601, Friedman test). The data presented here includes three ΔNIC–EGFP animals that failed training: i.e. they were unable to walk on the rod at 4 rpm for 30 seconds within a 5 minute training period. They managed a maximum time in the trials of 15, 23 and 33 seconds. Exclusion of these three animals does not affect the result of this experiment: ** p=0.010 (KS test) for day 3 (graph not shown).

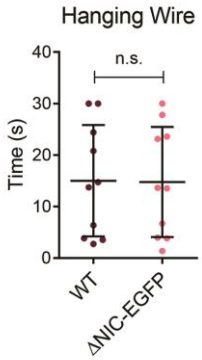


Figure 5.3.3D: ΔNIC-EGFP mice have no muscle strength phenotype in the Hanging Wire test
Results of the Hanging Wire test performed on ΔNIC-EGFP mice and their wild-type (WT) littermates. The time taken to bring their hind paw to a horizontal wire was recorded, up to a maximum score of 30 seconds. The graph shows the mean of three tests for each animal and the group mean and standard deviation. WT n=10; ΔNIC-EGFP n=10. No difference was found between the ΔNIC-EGFP mice and their WT littermates: p=0.960 (t-test).

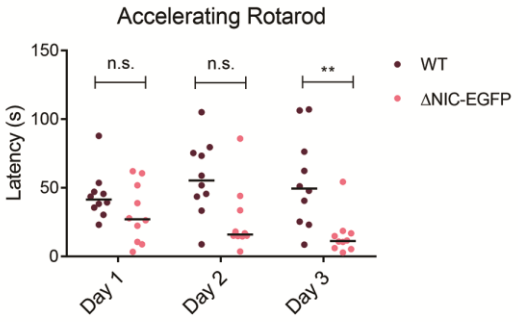


Figure 5.3.3E: ΔNIC-EGFP mice have a motor deficit on the Accelerating Rotarod
Results of the Accelerating Rotarod performed on ΔNIC-EGFP mice and their wild-type (WT) littermates. The time taken to fall off the rotating track was recorded, and the mean was taken for four trials for each day. The graph shows mean values for each individual animal and the group medians. WT n=10; ΔNIC-EGFP n=10. The ΔNIC-EGFP mice have reduced ability, which becomes significant on day 3: day 1 n.s. p=0.164, day 2 n.s. p=0.055, day 3 ** p=0.003 (KS tests). The WT mice did not display learning, analysed using the Friedman test to determine variation over time: p=0.601. The ΔNIC-EGFP show significant worsening performance over the course of this test, ** p=0.003 (Friedman test).

5.4 Discussion

The three novel truncated *Mecp2* knock-in lines all underwent extensive phenotypic characterisation. Hemizygous males were used for the analysis as they provide a more sensitive assay for MeCP2 function *in vivo*. Firstly, cohorts of knock-in animals and their wild-type littermate controls were scored for overt RTT-like symptoms over the course of a year using an established system (Guy et al., 2007). Secondly, separate cohorts were analysed using a series of standard behavioural tests at 20-21 weeks of age that assess activity, anxiety, muscle strength and motor function. Mice expressing the least severely truncated protein, ΔN-EGFP (which is only missing the N-terminus), display no phenotype in any of these assays. The second truncation, ΔNC-EGFP (which is missing both the N- and C-termini), also displayed no overt RTT-like symptoms throughout the year-long scoring period, and had no activity or motor phenotype. They did, however, present a very mild but significant decreased anxiety phenotype in the Open Field test, assessed by spending more time in the central region. This result was not backed up by their performance

in the Elevated Plus Maze, where no anxiety phenotype was detected. Lastly, these animals have a mild but significant increased body weight phenotype, which only became apparent after backcrossing onto a C57BL/6J background. Mice expressing the third and most severely truncated protein, ΔNIC-EGFP (which consists solely of the MBD and the NID), displayed mild RTT-like symptoms and were lighter than wild-type controls but had a normal lifespan and were fertile. Phenotypic scoring detected gait abnormalities and hind-limb clasping in these animals. Consistent with gait defects, they displayed poor motor coordination on the Accelerating Rotarod. Conversely, no differences in muscle strength were detected in the Hanging Wire test. These animals displayed a decreased anxiety phenotype, which was apparent in the Elevated Plus Maze but not the Open Field test. Lastly, no activity phenotype was detected in the Open Field test, consistent with observations during scoring. Overall, the N- and C-termini, which together make up 46% of the full-length MeCP2 protein sequence, are broadly dispensable for its function *in vivo*. Additional deletion of the Intervening region leads to reduced protein levels, weaker binding to the NCoR/SMRT co-repressor complex (see Chapter 4) and mild RTT-like symptoms.

5.4.1 The use of untagged wild-type littermate controls instead of WT-EGFP mice

In this study, all three novel lines were maintained separately and bred with wild-type C57BL/6J males (using heterozygous knock-in females). This means that the wild-type littermates used as controls in the phenotypic characterisation carry the endogenous *Mecp2* allele, which is not tagged with EGFP. At the early stages of this project, I considered breeding these lines with the pre-existing WT-EGFP mice, so that the resulting litters would contain EGFP-tagged truncation mutants and EGFP-tagged WT controls. This would fit with the standard practice of using the same epitope tags on all mutant and wild-type proteins that are to be compared. This is required in most cases to provide the most stringent comparator possible when determining differences between wild-type and mutant proteins. In this study, however, the 'bridge' hypothesis predicts an absence of a phenotype in the novel knock-in mice, so the most stringent test for normal behaviour is required. The WT-

EGFP mice are broadly normal phenotypically, validating the choice of EGFP as an epitope tag for this study, but they are leaner than their wild-type littermates and have a mild but significant defect in the Hanging Wire test. (I was aware of these results at the early stages of my project. They were later published in Brown, Selfridge et al., 2016.) The untagged wild-type mice therefore provide the more stringent comparator for 'normality'.

5.4.2 Both the N- and C-termini are dispensable for normal MeCP2 function in vivo

As mentioned above, the results of the phenotypic characterisation of the ΔN-EGFP and ΔNC-EGFP mouse lines strongly suggest that both the N- and C-termini of MeCP2 are not required for its function *in vivo*. The complete absence of RTT-like phenotypes in the ΔN-EGFP mice is apparent as they are indistinguishable from their wild-type littermates in all tests. One possible exception is the fact that the wild-type animals displayed learning on the Accelerating Rotarod and the ΔN-EGFP individuals did not. While this test is frequently used to assess motor learning (Collins et al., 2004; Samaco et al., 2008; Goffin et al., 2012), only the wild-type animals in the ΔN-EGFP cohort displayed learning in this study – the wild-type controls in the ΔNC-EGFP and ΔNIC-EGFP cohorts did not. A possible explanation for this is that this test is usually carried out using younger animals.

Interpretation of the results of the phenotypic characterisation of the ΔNC-EGFP knock-in mice is more difficult. These mice are healthy and fertile with a normal lifespan and display no overt RTT-like symptoms over the entirety of the year-long scoring period. Furthermore, no activity phenotype or motor defects were detected in these animals. They do, however, display slightly decreased anxiety in the Open Field test (but not the Elevated Plus Maze) and have a mild but significant increased body weight phenotype. Decreased anxiety is a common feature of mouse models of RTT. It is frequently assessed using the Elevated Plus Maze (Brown, Selfridge et al., 2016; Guy, unpublished) or the similar Elevated Zero Maze (Goffin et al., 2012; Pitcher et al., 2015) and more rarely using the Open Field test (Samaco et al., 2008; Selfridge and Brown, unpublished). To my knowledge, this is the only *Mecp2*-

mutant line to display an anxiety phenotype in the Open Field test but resemble wild-type controls in Elevated Plus Maze. In contrast, two RTT models (T158M-EGFP and R133C-EGFP) display decreased anxiety in the Elevated Plus Maze but not the Open Field test (Brown, Selfridge et al., 2016 and their unpublished data), suggesting that the Elevated Plus Maze is a more stringent test of anxiety.

The backcrossed ΔNC-EGFP mice also displayed a mild but significant increased body weight phenotype. Mouse models of RTT frequently have altered body weight, with some loss-of-function mutations causing an increase in body weight, e.g. the transgenic lines expressing R270X-EGFP, G273X-EGFP and R306-EGFP (Baker et al., 2013; Heckman et al., 2014). A decrease in body weight is more common, as seen in multiple *Mecp2*-mutant lines: T158A, T158M-EGFP, R133C-EGFP, R306C-EGFP [knock-in], R255X, P225R and P322L (Goffin et al., 2012; Brown, Selfridge et al., 2016; Pitcher et al., 2015; Guy, unpublished). Body weight phenotypes in *Mecp2*-mutant mice have been shown to be affected by genetic background as nulls have a reduced weight phenotype on a C57BL/6 background, but become obese when crossed with 129 animals (Guy et al., 2001). Genetic background may also account for the difference in weight phenotypes reported for the two independent R306C-EGFP lines (Heckman et al., 2014; Brown, Selfridge et al., 2016). Consistent with these observations, the outbred (75% C57BL/6J) ΔNC-EGFP scoring cohort did not display a weight phenotype. These weight phenotypes are therefore unlikely to have any biological significance related to RTT pathophysiology. They do, however, need to be taken into account when interpreting the results of some behavioural tests. Of the tests used here, only the Hanging Wire test is likely to be affected by body weight. This is illustrated by the strong inverse correlation between body weight and performance in this test by the ΔNC-EGFP cohort (both ΔNC-EGFP and wild-type animals). Despite the slightly heavier body weight of the ΔNC-EGFP individuals, their reduced performance in this test did not reach the threshold of significance.

5.4.3 Partial loss of the NID is the likely cause of the RTT-like phenotypes displayed by the published 308/y mice

A truncated mouse line (308/y) expressing MeCP2 residues 1-308 has been extensively characterised using multiple behavioural tests. These mice display several of the features associated with RTT, including increased anxiety, impaired motor function, seizures and decreased survival (90% at one year of age) (Shahbazian et al., 2002b; Moretti et al., 2005; McGill et al., 2006). The NCoR/SMRT interaction domain (NID) was mapped to residues 285-309, with peptides ending in residue 308 showing reduced binding (Lyst et al., 2013). If the 'bridge' hypothesis is correct, the phenotype displayed by these mice is a result of reduced ability of MeCP2 1-308 to bind the NCoR/SMRT complex, which could be rescued by the addition of one or more amino acids to restore NID function. The ΔNC-EGFP mice produced in this study and the pre-existing ΔC (residues 1-312) mice (Lagger, unpublished) test this hypothesis as they both have an additional four amino acids (309-312) completing the NID. The lack of a discernible phenotype in the ΔNC-EGFP mice strongly supports this hypothesis.

Characterisation of the ΔC mice also suggests that the C-terminus of MeCP2 after the NID is broadly dispensable, though the results are less clear. The ΔC mice underwent the same phenotypic characterisation as the mice produced in this study, after backcrossing onto the same genetic background (94% C57BL/6J). Briefly, these animals display no activity or anxiety phenotype in the Open Field test or the Elevated Plus Maze; and show normal performance in the Hanging Wire test, indicating normal muscle strength. They do, however, show a mild but significant motor learning defect on the Accelerating Rotarod, and mild breathing problems were recorded in the majority of animals from 20 weeks of age. Like the ΔNC-EGFP animals they have increased body weight, but in the ΔC cohort the two heaviest animals (out of a cohort of ten) died during the course of the year-long scoring experiment: one at 40 weeks and one at 48 weeks (Lagger and Selfridge, unpublished). Although both the ΔNC-EGFP and ΔC mice show a mild but

significant phenotype in one behavioural test (the ΔNC-EGFP mice have decreased anxiety in the Open Field test and the ΔC mice have reduced motor learning on the Accelerating Rotarod), the fact that these results are not reproduced by the other mouse line weakens their significance. More surprising is the reduced survival and mild breathing abnormalities in the ΔC line given that these mice could be considered biochemically more 'normal': the protein is not fused to an EGFP epitope tag and it is expressed at wild type levels in the brain (Lagger, unpublished). Therefore, the negligible phenotype of the ΔNC-EGFP knock-in mice may be due to a positive effect on protein function by either fusion with the EGFP tag or its slightly higher expression level.

5.4.4 Additional deletion of the Intervening region results in mild RTT-like symptoms: should these mice be considered a model of Rett syndrome?

Mice expressing ΔNIC-EGFP, the most severely truncated of the three MeCP2 mutant proteins in this study, displayed mild RTT-like symptoms: gait abnormalities, hind-limb clasping, decreased anxiety (in the Elevated Plus Maze) and motor dysfunction (on the Accelerating Rotarod). Despite this, they are fertile and have a normal lifespan. Although one animal in the scoring cohort was found dead at 43 weeks, it had only developed very mild symptoms up to that point. Its cause of death is therefore likely to be unrelated to the genetic mutation.

To determine whether these mice should be classed as a model of RTT syndrome, I compared my results to characterisation of an *Mecp2*-mutant model expressing a RTT-causing mutation at the milder end of the severity spectrum, R133C. The R133C mutation is a common cause of a milder form of classical Rett syndrome in patients (Cuddapah et al., 2014), and has been shown to cause a milder phenotype in a series of EGFP-tagged mutant mouse models (Brown, Selfridge et al., 2016). R133C-EGFP mice have been characterised using the same methods after backcrossing onto a C57BL/6J background, facilitating direct comparison between these mice and the ΔNIC-EGFP mice. Hemizygous male R133C-EGFP mice display many of the features of RTT, including reduced survival, gait abnormalities, hind-

limb clasping, tremors and hypoactivity. Their cumulative scores increase to an average of ~5 by 32 weeks when they plateau. Behavioural characterisation was performed at 8-10 weeks (Brown, Selfridge et al., 2016) and at 20 weeks (Brown, unpublished), the latter being the same age as the ΔNIC-EGFP in this study. R133C-EGFP mice of both ages show decreased anxiety in the Elevated Plus Maze and reduced muscle strength in the Hanging Wire test. They do not, however, display altered activity or anxiety in the Open Field test or affected performance on the Accelerating Rotarod. Overall, the phenotype of these animals is more severe than that of the ΔNIC-EGFP mice as the truncated mice do not have reduced survival, have lower cumulative scores (Fig. S11) and do not develop tremors.

Another useful comparator is a novel mouse model of RTT, expressing MeCP2 with the P225R mutation. Notably, this is one of the three 'unexplained' RTT-causing mutations that lie outside the MBD and NID (see section 1.6.2), and it is located in the Intervening region that has been deleted in the ΔNIC-EGFP mice but retained in the other two truncated proteins. There are an insufficient number of patients with this mutation to determine where it fits on the disease severity scale, but the phenotype displayed by the P225R knock-in mice is milder than the other mouse models. These animals have a median survival of 51 weeks (longer than 42 weeks for the R133C-EGFP mice). They do, however, display RTT-like symptoms including hind-limb clasping, gait abnormalities and tremors from an early age, with their average cumulative scores increasing to ~5-6 by 32 weeks. Behavioural characterisation of these mice reveals hypoactivity, decreased anxiety, and reduced muscle strength at 12 weeks of age (Guy and Selfridge, unpublished). Unfortunately, the younger age of testing limits the comparison to the ΔNIC-EGFP for these assays. Nevertheless, it is clear that have a more severe phenotype than the ΔNIC-EGFP mice presented in this study (Fig. S11). I therefore conclude that the ΔNIC-EGFP mice should not be considered to be a model of Rett syndrome.

5.4.5 Concluding remarks

In summary, the absence of RTT-like symptoms in the ΔN-EGFP and ΔNC-EGFP mice strongly suggests that both the N-terminal region before the MBD and the long C-terminus after the NID are dispensable for MeCP2 function *in vivo*. This result is striking as together these regions make up 46% of the MeCP2 protein sequence. Despite their length, these regions contain relatively few of the mapped MeCP2 interaction domains and post-translational modifications (Fig. S7 and Table S2). These regions contain the binding sites for the heterochromatin protein HP1ɣ (Agarwal et al., 2007), miRNA processing complex component DGCR8 (Cheng et al., 2014a), and the majority of the mapped splicing factor interaction sites (Buschdorf and Strätling, 2004; Young et al., 2005). The ΔNC-EGFP mice could therefore be used to uncouple these interactions and their associated functions with the RTT phenotype. This is discussed further in Chapter 7.

Mice expressing the shortest truncation, ΔNIC-EGFP, show mild RTT-like symptoms but their phenotype is much less severe than all existing mouse models carrying mutations that cause Classical RTT in patients. Therefore, I do not consider them to be a model of Rett syndrome. Their symptoms may be a result of three factors: lower protein levels in the brain, reduced ability to recruit the NCoR/SMRT co-repressor complex to chromatin *in vivo*, or to loss of a function that requires the Intervening region. Further work (discussed in Chapter 7) is needed to distinguish between these three factors: the first and second of which are consistent with the 'bridge' hypothesis; and the third, which would disprove it.

Chapter 6 – Induced activation of ΔNIC-EGFP after onset of symptoms reverses the *Mecp2*-null phenotype

6.1 Introduction and aims

Previous work has shown that induced activation of wild-type MeCP2 after the onset of symptoms in *Mecp2*-deficient mice leads to recovery of neurological function in both hemizygous males and heterozygous females (Guy et al., 2007). This result was a major milestone in the RTT field as it showed that Rett syndrome is caused by defective neurological function rather than being a neurodevelopmental disease. Importantly, this was a proof-of-principle experiment that showed reintroduction of functional MeCP2 into RTT patients had the potential to relieve symptoms of the disease. This led to gene therapy studies using *Mecp2*-mutant mice (Garg et al., 2013; Gadalla et al., 2013; reviewed by Katz et al., 2016).

In order to determine whether the Methyl-CpG Binding Domain (MBD) and NCoR/SMRT Interaction Domain (NID) are sufficient for MeCP2 to restore function to mature neurons in *Mecp2*-deficient mice, I repeated the genetic activation experiment described in Guy et al. (2007) using the ΔNIC-EGFP allele. A second reason for this experiment was that the packaging capacity of viral vectors used for gene therapy limits the choice of regulatory elements that can be included. Reducing the length of the *Mecp2/MECP2* cDNA sequence in these vectors would potentially allow for better control of its expression. MeCP2 dosage is extremely important as too much MeCP2 results in *MECP2* Duplication syndrome (van Esch et al., 2005). This chapter describes the results of the genetic activation experiment where the ΔNIC-EGFP allele was activated after symptom onset in *Mecp2*-deficient mice.

6.2 Generation of STOP mice and method to activate the ΔNIC-EGFP allele

6.2.1 Breeding program to retain the NeoSTOP cassette and characterisation of the STOP mice

The targeting vector used to create all three knock-in ES cell lines (Fig. 4.2.1A) was the same one that was used to create the null-like 'STOP' mice by Guy et al. (2007). This vector introduced a neomycin resistance gene and transcriptional stop cassette flanked by *loxP* sites ('floxed') into intron 2 of the *Mecp2* gene. To create the lines that express the truncated MeCP2 alleles, the chimaeras containing the targeted ES cells were bred with *CMV-Cre* homozygous females leading to the removal of the cassette. In contrast, for this experiment, I wanted to produce STOP mice that do not express the truncated MeCP2 protein (ΔNIC-EGFP) but have the potential for this gene to be activated. These animals retained the floxed NeoSTOP cassette, which silences transcription of the ΔNIC-EGFP gene. This STOP allele is activatable upon recombination with Cre recombinase, which excises the floxed NeoSTOP cassette. To produce STOP mice, some of the ΔNIC-EGFP chimaeras were bred with wild-type C57BL/6J females (Fig. 6.2.1A, Table 4.3.1). As before, the first hemizygous males were in the N=2 generation (75% C57BL/6J). These animals have very low levels of protein that was only just detectable by western blotting (Fig. 6.2.1Bi). This was more easily quantified using flow cytometry analysis of nuclei obtained from whole brains: 2.7% of WT-EGFP and 5.7% of ΔNIC-EGFP levels (Fig. 6.2.1Bii). Very similar values were obtained for neuronal (high NeuN) nuclei: 2.6% and 6.4% of WT-EGFP and ΔNIC-EGFP levels, respectively. A follow-up paper to the original genetic reversal study estimated the protein level in the STOP(WT) mice to be around 2.5% of wild-type levels (Robinson et al., 2012), which would be around 1.6% of WT-EGFP levels. The higher levels observed here in the STOP(ΔNIC-EGFP) mice is surprising given the lower levels of ΔNIC-EGFP protein in the knock-in mice. This result may be due to the higher sensitivity of flow cytometry analysis compared to western blotting, used by Robinson et al. (2012).

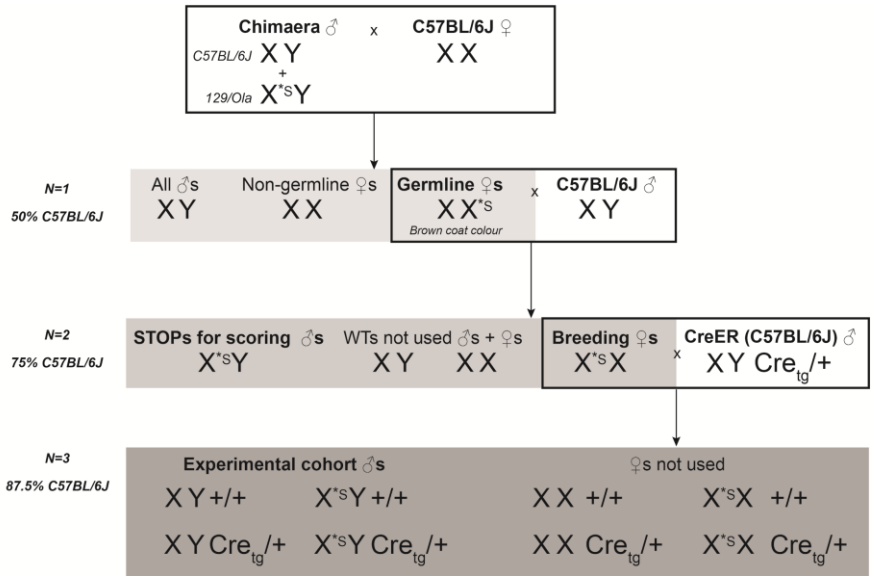


Figure 6.2.1A: Mouse breeding programme to produce MeCP2-deficient mice with the potential to activate ΔNIC-EGFP
Diagram of crosses and genotypes of offspring produced at each stage. The knocked-in truncated allele is denoted by *, the NeoSTOP cassette by 'S', CreER^{T} by 'Cre_{tg}'. Breeding pairs are indicated by black rectangles. Each generation is labelled by the number of times is has been backcrossed (N) and the resulting percentage of C57BL/6J background.

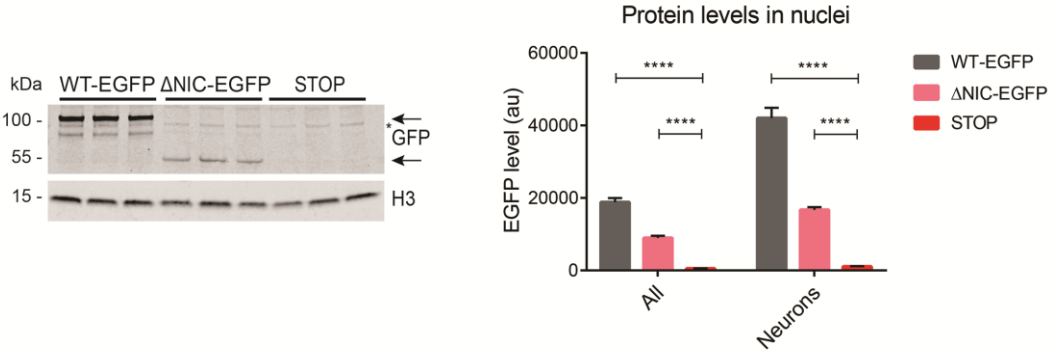


Figure 6.2.1B: STOP mice express very low levels of ΔNIC-EGFP protein
(Left) Western blot of crude whole brain extract derived from hemizygous male WT-EGFP, ΔNIC-EGFP and STOP mice (three biological replicates of each). MeCP2 was detected with a GFP antibody. Histone H3 was used as a loading control. 'm' =size marker, * = non-specific band. (Right) Nuclei were purified from the brains of WT-EGFP, ΔNIC-EGFP (see Fig. 4.4.1C) and STOP mice (three biological replicates of each genotype) and stained with anti-NeuN-Alexa Fluor 647. Protein level was determined by EGFP fluorescence in all nuclei (n=50,000) and the neuronal (high NeuN) subpopulation (n>8,000). Graph shows mean and standard deviation (au = arbitrary units).Statistical analysis was performed using t-tests: **** p<0.0001.

As expected, hemizygous male STOP mice developed RTT-like symptoms, resembling *Mecp2*-nulls. Mice were scored from 5 weeks of age, where symptoms were already apparent (Fig. 6.2.1Ci). The average score increased rapidly, similar to *Mecp2*-null animals, with symptoms resulting in premature death. The majority of STOP mice died between 6 and 18 weeks of age, although two individuals lived for 32 and 45 weeks (Fig. 6.2.1Cii). The somewhat increased lifespan of the STOP animals compared to nulls is consistent with previous observations with the STOP(WT) mice (Guy et al., 2007; Robinson et al., 2012). The slight amelioration of the null phenotype in both STOP lines is likely to be due to the presence of low levels of MeCP2 protein. Although, it should be noted that the null data used for comparison was obtained from backcrossed mice – which is likely to contribute to the difference observed since nulls are fitter on an outbred background.

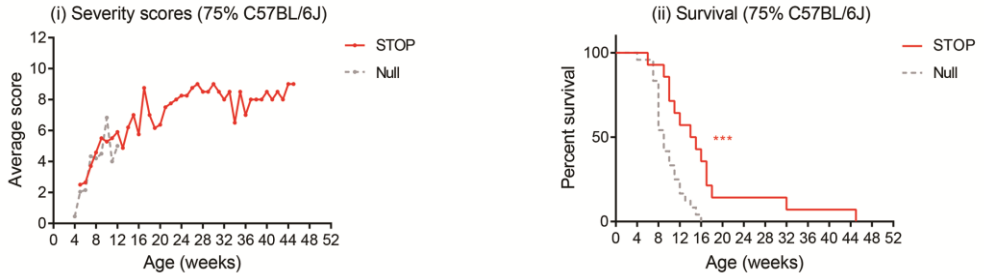


Figure 6.2.1C: STOP mice resemble *Mecp2*-nulls
Scoring (i) and survival (ii) curves for hemizygous male STOP(ΔNIC-EGFP) mice. STOP mice were scored from 5 weeks of age. STOP animals: scoring n=22, survival n=14. (Backcrossed) null scoring and survival data was obtained by Jacky Guy (Brown, Selfridge et al., 2016). Null animals: scoring n=12, survival n=24. STOP(ΔNIC-EGFP) mice have significantly increased survival *** p=0.0008 (Mantel-Cox test).

6.2.2 Generation of mice with activatable ΔNIC-EGFP

In order to achieve inducible activation of the ΔNIC-EGFP gene after the onset of symptoms, the *CreER*^{T} transgene was introduced into this line. This gene expresses Cre recombinase fused to a modified oestrogen receptor (ER^{T}). The resulting protein is localised in the cytoplasm until injected Tamoxifen allows it to translocate into the

nucleus, where it can excise the NeoSTOP cassette. This system was used by Guy et al. (2007), who showed that it is robust and non-leaky. Heterozygous *STOP/+* females were bred with males that were heterozygous for the *CreER^T* transgene. This cross produces males of four genotypes: wild-type, wild-type CreER^T, STOP, and STOP CreER^T (Fig. 6.2.1A). All four genotypes were included in the reversal experiment, in which all individuals received a series of Tamoxifen injections (see below). The STOP CreER^T individuals carry both the ΔNIC-EGFP allele with the floxed NeoSTOP cassette and the *CreER^T* transgene. They therefore develop symptoms before the successful excision of the NeoSTOP cassette upon Tamoxifen injection. The other three genotypes serve as controls: the wild-types (*CreER^T* positive and negative) to check that the Tamoxifen injections do not result in toxicity and the STOP animals to prove that reversal of symptoms is dependent on excision by CreER^T.

6.2.3 Injection regime to activate the ΔNIC-EGFP allele

It has previously been shown that rapid activation of *Mecp2* causes toxicity (Guy et al., 2007). This study and the follow up study (Robinson et al., 2012) used more gradual injection regimes to successfully reverse symptoms in MeCP2-deficient mice. I used a similar regime of two weekly injections followed by five daily injections, each at a dose of 100 μg Tamoxifen/g weight. This procedure was started when the animals reached a symptom score of 2 or 3, when they were aged 12-17 weeks of age. At this time point, they would expect to survive for another four weeks. The STOP(ΔNIC-EGFP) mice in this study developed symptoms more rapidly than the STOP(WT) animals, with eight out of nine STOP CreER^T individuals reaching a score of two or more at 6 weeks of age (Table 6.2.3). Seven of these animals had developed a tremor and only one animal (number 5) had a lower score, with only gait abnormalities. Because of this somewhat earlier onset, I chose to begin the injection regime at this point for all STOP CreER^T animals and the wild-type and STOP controls. Due to Cre toxicity (reviewed by Sharma and Zhu, 2014), the STOP animals that did not contain the CreER^T transgene had slightly lower scores than the CreER^T positive mice (Fig. 6.3.1i). This did not affect the validity of

the STOP animals to serve as controls in this experiment as their symptoms continue to progress in spite of Tamoxifen injection. If, however, the opposite had been true, the STOP animals would have more advanced symptoms that might reach a stage that is no longer reversible had the *CreER^T* allele been present. In this case, these animals would have required Tamoxifen injections from an earlier age (before their symptoms had progressed too far) to serve as proper controls.

Mouse	**Total**	Activity	Gait	Clasping	**Tremor**	Breathing	General condition
1 (MGS19)	**2**	1	0	0	**1**	0	0
2 (MGS20)	**3**	1	1	0	**1**	0	0
3 (MGS36)	**4**	1	1	0	**1**	1	0
4 (MGS49)	**3.5**	1	1.5	0	**1**	0	0
5 (MGS52)	**1**	0	1	0	**0**	0	0
6 (MGS6)	**2.5**	0.5	1	0	**1**	0	0
7 (MGS21)	**3**	0.5	1.5	0	**0**	1	0
8 (MGS33)	**2.5**	1	1	0	**0.5**	0	0
9 (MGS50)	**4**	1	2	0	**1**	0	0

Table 6.2.3: Scores of the nine STOP CreER^T mice at 6 weeks of age, the day before the first Tamoxifen injection

6.3 Successful phenotypic reversal after activation of ΔNIC-EGFP

6.3.1 Reversed mice exhibited reduced symptoms and lived for the duration of the study

As mentioned above, the STOP CreER^T and STOP mice had developed symptoms before the Tamoxifen injection regime was started. The injections halted the symptom progression in the STOP CreER^T mice, shown by the plateau in the score (Fig. 6.3.1i). By 12 weeks of age, their symptoms had reduced to almost the same levels as the wild-type controls. The scores of these animals remain around 2, similar to the ΔNIC-EGFP knock-in mice. All nine STOP CreER^T mice survived the duration of the experiment (28 weeks of age). The STOP mice that did not have the *CreER^T* transgene also developed symptoms from 4 weeks; and, as expected, these were not reversed by the Tamoxifen injections. Instead their symptoms progressed, resulting in premature death between 8 and 26 weeks of age (Fig. 6.3.1ii). The wild-type control animals (both *CreER^T* positive and negative) displayed no RTT-like symptoms before the injections began, but later received low scores (mostly for

activity and gait) for the duration of the experiment. This relatively mild effect may be due to the Tamoxifen (or its oil solvent, which appeared to build up around the testes of some animals). Overall, this experiment shows that the severely truncated protein, ΔNIC-EGFP, is able to restore neuronal function in the STOP mice, reversing RTT-like symptoms.

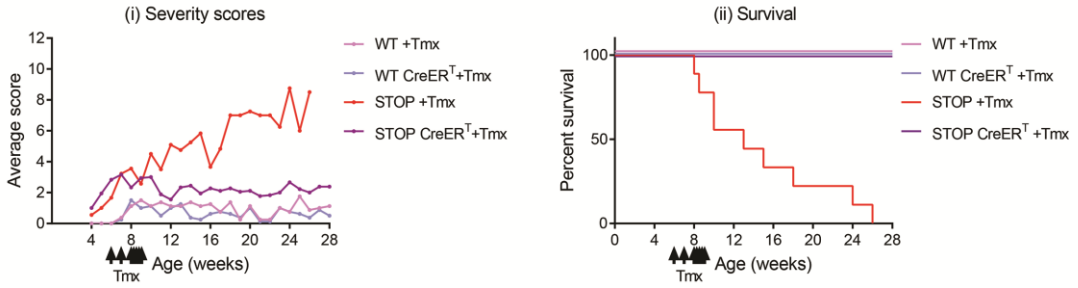


Figure 6.3.1 Activation of ΔNIC-EGFP reverses RTT-like symptoms and prevents premature death in STOP CreER^T mice
Scoring (i) and survival (ii) curves for WT, WT CreER^T, STOP and STOP CreER^T mice, all injected with Tamoxifen (denoted by '+Tmx'). The time points of the seven Tamoxifen injections are shown by arrows on both graphs. 100% of WT, WT CreER^T and STOP CreER^T mice survived the duration of the experiment (28 weeks). WT n=4, WT CreER^T n=4, STOP n=9, STOP CreER^T n=9.

6.3.2 Biochemical analysis shows high rates of recombination and protein expression in the brains of reversed mice

The activation of ΔNIC-EGFP upon Tamoxifen injection can be analysed at both the DNA and protein level. Southern blot analysis using genomic DNA extracted from the brains of the reversed mice at the end of the experiment (28 weeks of age) shows the percentage of cells in which Cre-mediated recombination had occurred, excising the NeoSTOP cassette (Fig. 6.3.2A). The average recombination frequency was 83.1% (range 73-90%). This is comparable to the original study, where the average recombination frequency was ~80% (Guy et al., 2007).

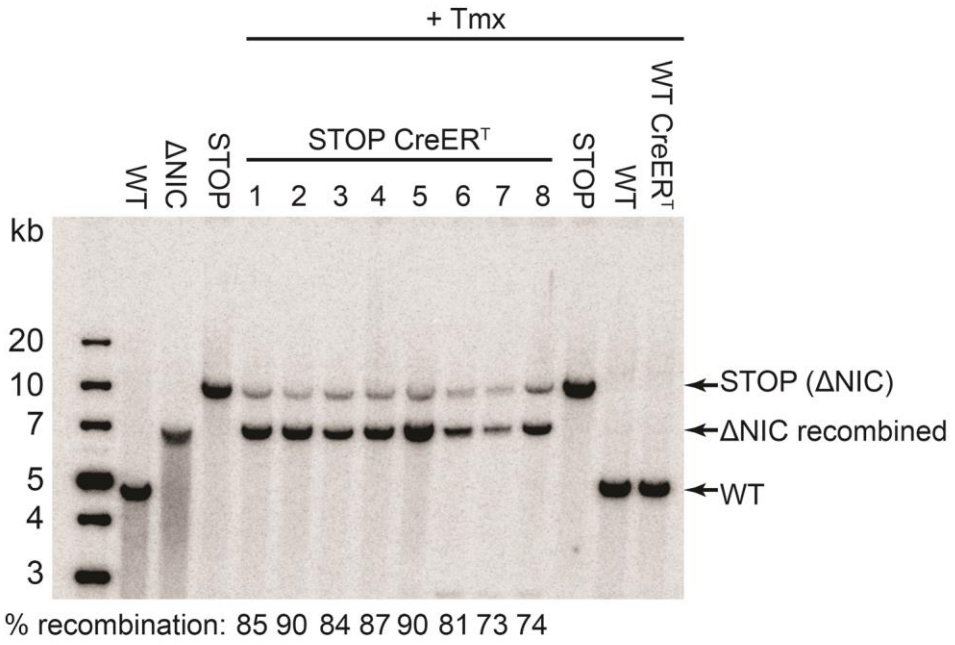


Figure 6.3.2A High levels of recombination occurred in all the Tamoxifen-treated STOP CreER^T mice
Southern blot using Bsu36I restriction enzyme (see Fig 4.3.2C for a map of restriction sites). Brain tissue from eight out nine STOP CreER^T + Tmx mice was included in this analysis. The % of Cre-mediated recombination in each is given below. The STOP + Tmx mouse shows that no recombination could occur without the CreER^T transgene. Other samples were included for reference. *Note: the ninth STOP CreER^T + Tmx mouse was perfused with paraformaldehyde to be used for microscopy (not yet carried out).*

To determine if these high rates of recombination restore gene expression, the protein levels in these eight STOP CreER^T mice were analysed by western blotting (n=5) or flow cytometry (n=3). Western blot analysis using whole brain crude extract shows high levels of ΔNIC-EGFP protein in all five STOP CreER^T animals, which was not significantly different from the protein levels of the ΔNIC-EGFP mice (Fig. 6.3.2B).

The protein levels in the other three reversed STOP CreER^T individuals were analysed instead by flow cytometry, using the fluorescence of the EGFP tag. Again, there was no significant difference between protein levels of STOP CreER^T mice and the ΔNIC-EGFP (Fig. 6.3.2C). Staining the nuclei with a NeuN antibody fused to Alexa Fluor 647 enabled the neuronal subpopulation to be analysed separately.

Again, the results were very similar but just over the threshold of significance: 87% of ΔNIC-EGFP levels, * p=0.016. These results agree with the level of Cre-mediated recombination of the genomic DNA in the STOP CreER^T mice.

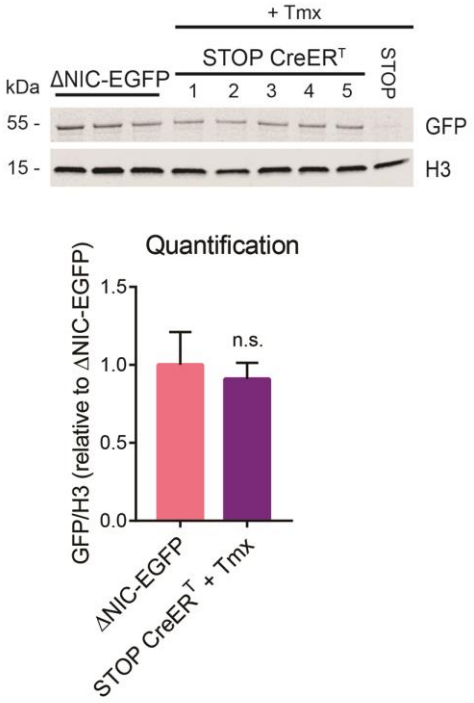


Figure 6.3.2B Cre-mediated recombination restored protein levels in STOP CreER^T mice
Western blot analysis using whole brain crude extract from five CreER^T mice and one STOP mouse (all injected with Tamoxifen), compared to the ΔNIC-EGFP knock-in line (n=3). The quantification shows the mean percentage protein level of the five CreER^T individuals, relative to the ΔNIC-EGFP controls. (Lysates were run on two replicate blots, values from which were combined in the quantification.) The protein level in the STOP CreER^T mice was not significantly different from the ΔNIC-EGFP controls (t-test, p=0.437). *'+Tmx' denotes Tamoxifen injected animals.*

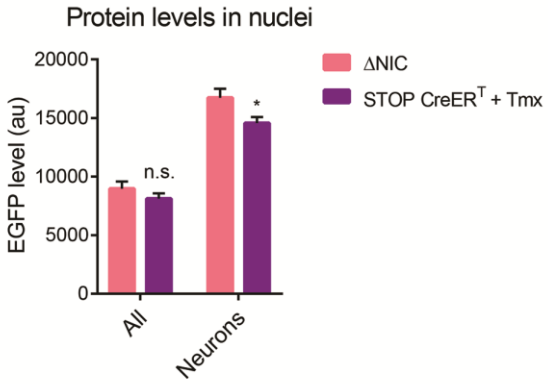


Figure 6.3.2C Cre-mediated recombination restored protein levels in the neurons of STOP CreER^T mice
Nuclei were purified from the brains of STOP CreER^T mice (n=3, numbers 6-8) and ΔNIC-EGFP controls (n=3, see Fig. 4.4.1C) and stained with NeuN-Alexa Fluor 647. Protein level was determined by EGFP fluorescence in all nuclei (n=50,000) and the neuronal (high NeuN) subpopulation (n>8,000). Statistical analysis was performed using t-tests: all nuclei not significant p=0.128, neuronal nuclei * p=0.016.

Overall, the reversal of RTT-like symptoms by activation of the ΔNIC-EGFP allele was successful with high levels of recombination and protein expression, reflecting the results of the original study (Guy et al., 2007).

6.4 Discussion

Genetic activation of ΔNIC-EGFP in MeCP2-deficient 'STOP' male mice after the onset of neurological symptoms successfully reversed the disease. High levels Cre-mediated recombination resulting in protein levels similar to ΔNIC-EGFP mice were obtained in all eight Tamoxifen-injected STOP CreER^T animals that were analysed. In the future, the extent of activation of ΔNIC in specific brain regions or cell types could be analysed by microscopy using brain slices. The brain of the ninth STOP CreER^T animal in this cohort was perfused with paraformaldehyde to be used for this. Analysis of ΔNIC-EGFP protein levels in brain slices should reflect the results obtained by Southern blotting, western blotting and flow cytometry. The use of cell-type specific markers for immunofluorescence will enable the determination of ΔNIC-EGFP levels in different cell types.

The design of this experiment was almost identical to the initial study by Guy et al. (2007) who demonstrated successful disease reversal upon activation of the wild-type protein. The experiment presented here differs only in two respects. Firstly, before the activation, the STOP(ΔNIC-EGFP) mice develop symptoms more rapidly than the previously described STOP(WT) mice. Both lines express low levels of MeCP2 protein, causing their phenotype to be slightly milder than *Mecp2*-nulls. The reduced stability and NCoR/SMRT complex binding activity of ΔNIC-EGFP (presented in Chapter 4) is consistent with this having a lesser effect on the STOP phenotype. The earlier symptom onset had no implications on the suitability of the STOP(ΔNIC-EGFP) mice for this experiment, it simply meant that the regime of Tamoxifen injections had to begin earlier to reflect their phenotypic scores: at 6 weeks rather than 12-17 weeks. The second difference between these studies is that the protein introduced here, ΔNIC-EGFP, does not fully prevent RTT-like symptoms

in constitutively-expressing knock-in mice (see Chapter 5). Therefore, complete reversal was not expected in these mice. Indeed, after Tamoxifen injections, the average scores of the reversed animals decreased and stabilised to ~2, similar to constitutively-expressing ΔNIC-EGFP mice of the same age. It is surprising, however, that the scores of the reversed mice in the original study by Guy et al. (2007) also stabilise around 2. To investigate whether reversal by ΔNIC-EGFP is as efficient as the wild-type protein, this experiment must be repeated with both genotypes intermixed and scored blind.

In conclusion, introduction of ΔNIC-EGFP into MeCP2-deficient mice after symptom onset led to a dramatic reversal of the RTT phenotype. This experiment shows that the MeCP2 activities that are absent from the truncated ΔNIC-EGFP protein are not required to restore function to mature MeCP2-deficifient neurons. These results suggest that this truncated allele is suitable for gene therapy trials.

Chapter 7 – Further Work

7.1 Introduction

This study investigated whether the ability of MeCP2 to recruit the NCoR/SMRT co-repressor complex to chromatin was sufficient for its function *in vivo*. This activity requires two domains of MeCP2: the MBD and the NID. I have produced knock-in mouse lines expressing truncated alleles of MeCP2 that lack the other three regions of the protein: the N-terminus before the MBD, the Intervening region between the MBD and the NID, and the C-terminus after the NID. These regions were deleted in a step-wise manner in increasingly short truncated proteins (Fig. 7.1). My results indicate that both the N- and C-terminal regions of MeCP2 (which together make up 46% of the protein sequence) are dispensable for MeCP2 function. Additional deletion of the Intervening region between the MBD and the NID results in the presence of mild RTT-like symptoms, which may be due to reduced protein stability and/or reduced ability to bind the NCoR/SMRT co-repressor complex. This most severely truncated protein is nevertheless able to reverse the phenotype of MeCP2-deficient mice when activated after the onset of symptoms.

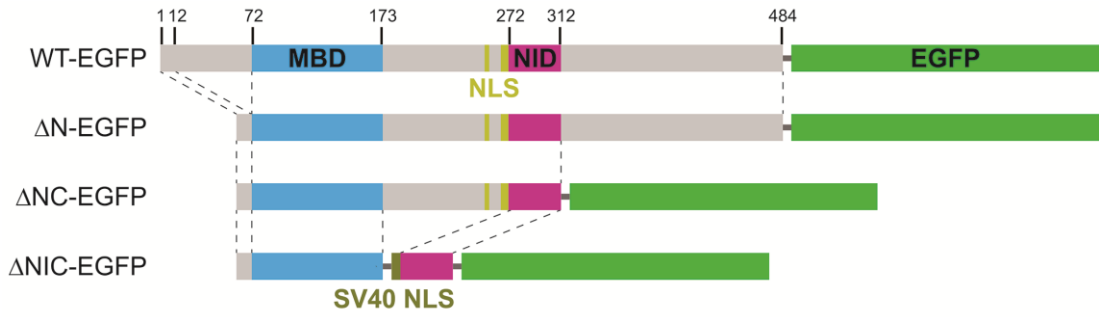


Figure 7.1: Stepwise truncations of MeCP2 protein
(Copy of Fig. 3.1) Schematic diagram of the truncated proteins designed in this study, compared to the full length protein (WT-EGFP). Drawn to scale (e2 isoforms). Linker sequences are shown as dark grey bars. Dotted lines denote corresponding regions between the proteins.

7.2 Further characterisation of the ΔN-EGFP and ΔNC-EGFP knock-in mice

The three knock-in mouse lines (ΔN-EGFP, ΔNC-EGFP and ΔNIC-EGFP) produced in this study underwent extensive phenotypic characterisation. The methods used recorded the presence of overt RTT-like symptoms over the course of a year, and tested for activity, anxiety, muscle strength and motor coordination at 20-21 weeks of age (all presented in Chapter 5). While the results obtained for the ΔN-EGFP and ΔNC-EGFP mice strongly suggest that both the N- and C-termini may be dispensable for MeCP2 function *in vivo*, further behavioural testing may be required. It is clearly impossible to categorically prove the complete absence of a phenotype in these animals so instead I will briefly mention some tests of particular relevance. All of these tests have been used to detect more subtle phenotypes in *Mecp2*-mutant mice that carry milder mutations (i.e. mutations that do not cause RTT in patients). The well characterised mouse line expressing a C-terminally truncated allele 308/y (consisting of residues 1-308) displayed abnormal social interaction and nesting behaviour (Shahbazian et al., 2002b; Moretti et al., 2005). To better determine whether the addition of the extra four C-terminal amino acids (residues 309-312) included in the ΔNC-EGFP protein are sufficient to rescue the phenotype of the 308/y mice, these behaviours should be tested. The C-terminal region contains a well-studied activity-dependent phosphorylation site (Ser421). Mice expressing the phospho-abolishing S421A mutant protein do not display any neurological symptoms except a reduced preference for novelty in the '3 Chamber Test' (Cohen et al., 2011). Additional mutation of a nearby activity-dependent phospho-site (S424A) results in improved spatial memory in the Morris Water Maze and improved learning and memory in the context-dependent fear conditioning test (Li et al., 2011). Analysis of these behavioural paradigms in the ΔNC-EGFP mice will determine whether loss of the entire C-terminal region has less of a phenotypic effect than mutation of these phospho-sites. Finally, abnormal electrophysiology has been reported in all three of these mouse models (Shahbazian et al., 2002b; Cohen et al., 2011; Li et al., 2011) and also in mice expressing the A140V mutation, which causes milder intellectual disability in patients (Ma et al., 2014). It is therefore important to investigate this in the ΔN-EGFP and ΔNC-EGFP mice.

As previously discussed, both increased and decreased body weight phenotypes are frequently reported in *Mecp2*-mutant mouse models. These are affected by genetic background so may not be of biological importance (Guy et al., 2001). Strikingly, mouse models of RTT consistently have reduced brain size, even if their body weight is increased (Baker et al., 2013; Heckman et al., 2014). Analysis of brain weights in the novel mice produced in this study is therefore an important part of their characterisation. I would not predict there to be a difference as this phenotype is not present in the 308/y animals (Shahbazian et al., 2002b).

7.3 Investigating the cause of the mild RTT-like phenotype in the ΔNIC-EGFP mice

Mice expressing the shortest of the three truncated proteins, ΔNIC-EGFP, exhibit mild RTT-like symptoms including gait abnormities, hind-limb clasping, impaired motor coordination and decreased anxiety. Biochemical characterisation of all three novel mouse lines shows that this protein differs from ΔN-EGFP and ΔNC-EGFP in two unexpected ways: ΔNIC-EGFP protein is present at lower level and has reduced binding to HDAC3 (the catalytic component of the NCoR/SMRT complex) in the brain. A negative phenotypic effect of one or both of these factors could be responsible for the neurological phenotype displayed by the ΔNIC-EGFP mice. Alternatively, their phenotype could be due to loss of an important function that requires the deleted Intervening region of MeCP2. Only the last of these three possibilities would contradict the 'bridge' hypothesis.

7.3.1 Could reduced ΔNIC-EGFP protein levels be responsible for the phenotype in these mice?

MeCP2 protein levels in the brains of the knock-in mice were analysed by western blotting and flow cytometry. Both techniques show that ΔNIC-EGFP levels are decreased compared to WT-EGFP controls, but to different extents. The higher and more reliable estimate (obtained by flow cytometry) estimates the ΔNIC-EGFP to be

48% of WT-EGFP levels in nuclei prepared from whole brains. Given the slightly elevated expression of the WT-EGFP protein (Brown, Selfridge et al., 2016), this could be as high as 76% of the endogenous untagged wild-type protein. This level is higher than that in the hypomorphic mice (floxed mice), which have 50% as much as wild-type controls (Samaco et al., 2008). As the hypomorphic mice have a milder phenotype than the ΔNIC-EGFP mice (Kerr et al., 2008; Cheval, Guy et al., 2012), reduced protein level is unlikely to be the sole cause of their mild RTT-like symptoms. Reduced proteins levels may, however, increase the severity of their phenotype. This could be investigated by overexpressing ΔNIC-EGFP from the *Tau* promoter, using the system developed by Luikenhuis et al. (2004).

7.3.2 Could reduced 'bridge' function be responsible for the phenotype in the ΔNIC-EGFP mice?

Immunoprecipitation of these proteins from brain nuclear extracts indicated that binding affinity to the NCoR/SMRT complex component, HDAC3, is greatly reduced in the ΔNIC-EGFP truncated protein. This contrasts the earlier result obtained by transiently expressing these proteins in HeLa cells. Further analysis of NCoR/SMRT complex binding in the brain using this immunoprecipitation assay is required: particularly determining whether the same result is obtained for other members of the complex. However, difficulties in quantifying the amounts of these differently-sized proteins by western blotting may make quantification of their NCoR/SMRT binding affinities by this technique impossible. An alternative method of determining 'bridge' function is quantification of the TBL1X-mCherry recruitment assay (described in section 3.3.3). As shown in Fig. 3.3.3, the truncated proteins were able to bring TBL1X-mCherry to heterochromatin, but the percentage of cells in which this occurs was not calculated. Previous studies have shown this assay to be quantitative: wild-type MeCP2 recruits TBL1X-mCherry to heterochromatin in 55-70% of cells and the NID mutant, R306C, recruits TBL1X-mCherry in 0% of cells (Lyst et al., 2013). Strikingly, recent analysis of the two 'unexplained' proline mutants (P225R and P322L), showed that although they are both localised at pericentromeric heterochromatin when overexpressed in NIH-3T3

cells and can bind endogenous NCoR/SMRT complex components in HEK293 cells, they are only able to recruit TBL1X-mCherry in 20-30% of cells (Guy and Alexander-Howden, unpublished). Therefore, quantification analysis of this assay may uncover defects in 'bridge' function in the truncated proteins that I have missed.

More quantitative analysis of the ability of ΔNIC-EGFP to form the other end of the 'bridge', i.e. binding to methylated DNA, is also required. The method used to assess DNA binding in this study (localisation of MeCP2 mutant proteins to heterochromatic foci in mouse cells) has been used to shown defective binding of a large number of MBD-mutants (Table S3; e.g. Kudo et al., 2003). As several RTT-causing MBD mutants showed normal localisation at heterochromatic foci, their binding abilities have been investigated further using Fluorescence Recovery After Photobleaching (FRAP) (Schmeideberg et al., 2009). This study showed that wild-type MeCP2 had a residence time on heterochromatin of 15 seconds, which was decreased dramatically in MBD mutants. Importantly, only those mutants with residence times of less than ~5 seconds showed diffuse nuclear staining upon fixation - making FRAP more informative than localisation. FRAP analysis could therefore be used to further investigate the methylated DNA binding ability of ΔNIC-EGFP. Alternative methods of analysis include a variety of *in vitro* techniques using purified proteins, which enable a dissociation constant (K_D) to be calculated. The most commonly used technique, electrophoretic mobility shift assay (EMSA), may be unsuitable for this as full-length MeCP2 gets trapped in the wells of the gels so an N-terminal fragment (residues 1-205) is often used (Klose et al., 2005). It may be possible to use this technique to compare the binding of ΔNC and ΔNIC (without the EGFP tag), but other *in vitro* techniques such as Surface Plasmon Resonance (used by Mellén et al., 2012) may be more informative.

Determining the methylated DNA binding abilities of ΔNIC-EGFP in neurons may be more difficult. The presence of the EGFP tag in the knock-in mice expressing the truncated MeCP2 proteins permits FRAP analysis using brain slices or cortical neurons derived from these mice. Alternatively, the knock-in ES cells produced in this study could be differentiated into neurons in culture and used for FRAP analysis.

FRAP has been used successfully to measure protein dynamics in *ex vivo* brain slices (e.g. Tønnesen et al., 2014) and cultured neurons (e.g. Caron et al., 2014). Unfortunately, however, preliminary studies quantifying MeCP2 dynamics at heterochromatic foci in cultured ES cell-derived neurons found that the foci moved in the Z plane (after photobleaching), where they could not be tracked – making measurement of recovery impossible (unpublished observations from the Bird lab). An alternative strategy was used by Goffin et al. (2012) to show reduced binding of the MBD mutant, T158A, in the brains of knock-in mice. They showed that this mutant protein was more readily extracted at lower NaCl concentrations than the wild-type protein, indicating weaker binding to DNA. As discussed in section 4.4.3, ΔNIC-EGFP was more readily extracted than the longer MeCP2 mutants using 150mM NaCl and benzonase. Further analysis is required to determine its extraction efficiency under a variety of NaCl concentrations in the absence of benzonase. As percentage extraction is quantified by calculating the amount of protein extracted at each concentration as a fraction of the total (i.e. with 500mM NaCl and benzonase), its quantification would not require comparison between the differently sized truncation proteins, making western blotting suitable for this assay.

7.3.3 Is it possible to design a minimal MeCP2 protein with full stability and bridge activity

In order to properly determine whether NCoR/SMRT complex recruitment activity is sufficient for MeCP2 function *in vivo*, I would need to alter the design of ΔNIC-EGFP to restore full stability and bridge activity. Although low ΔNIC-EGFP protein levels despite high mRNA levels indicates that this protein is unstable, this would first have to be confirmed. Goffin et al. (2012) demonstrated instability of the MBD mutant, T158A, by treating cultured cortical neurons derived from knock-in mice with cycloheximide, which inhibits new protein translation. They show significantly reduced protein levels in T158A knock-in cells compared to wild-type controls after six hours of cycloheximide. Importantly, although this assay uses western blotting for quantification of protein levels, normalisation of protein levels at each time point to untreated samples makes it suitable to compare the stability of proteins of different

sizes. If a difference is found in this assay, it might be useful to determine how ΔNIC-EGFP is degraded in order to design a more stable version. A recent study treated T158M-expressing cell lines with inhibitors of several degradation pathways: calpain (with calpeptin), autophagy (with 3-methyladenine), lysosome (with ammonium chloride) and proteasome (with MG132 or lactacyctin). They identified the proteasome as the predominant pathway, with greatest elevation of T158M levels in MG132- and lactacystin-treated samples (Lamonica et al., 2017). This assay could be used to determine the pathway responsible for ΔNIC-EGFP degradation in ΔNIC-EGFP-expressing cells, e.g. cortical neurons derived from knock-in mice.

Thambirajah et al. (2009) describe two predicted PEST sequences in MeCP2, a common feature of intrinsically disordered proteins. These sequences are proline/glutamic acid/serine/threonine rich and promote degradation by the 26S ubiquitin proteasome system (UPS) after phosphorylation of serine residues a subsequent poly-ubiquitination on nearby lysines (reviewed by Ausió et al., 2014). The two predicted PEST sequences in MeCP2 consist of the N-terminal residues 61-82 (PEST1) and the C-terminal residues 377-412 (PEST2). Notably, part of PEST1 and all of PEST2 is missing from both ΔNC-EGFP and ΔNIC-EGFP. I used the ePESTfind predictor tool (http://emboss.bioinformatics.nl/cgi-bin/emboss/epestfind) to look for predicted PEST sequences in these truncated proteins. Although this tool found the two predicted PEST sequences in the full-length sequences, no PEST sequences were found in either of the truncated proteins. Therefore, this is unlikely to be the pathway responsible for ΔNIC-EGFP degradation.

Given that the levels of both ΔN-EGFP and ΔNC-EGFP proteins are similar to WT-EGFP in the brains of knock-in mice, it is most likely that replacement of the Intervening domain with the exogenous linker and SV40 NLS is responsible for protein instability. Structural information of this region in both the wild-type and truncated proteins could uncover any mis-folding in ΔNIC-EGFP. Novel mass spectrometry-based methods that analyse crosslinked peptides can be used to determine intramolecular interactions in MeCP2, which had previously been impossible as it is so highly disordered. Surprisingly, recent analysis has identified

multiple crosslinks in part of the Intervening region: residues 212-237 (Connelly and Belsom, unpublished). This sequence is highly conserved in evolution (Fig. S2) and contains the 'unexplained' RTT-causing mutation, P225R, that causes instability (Guy, unpublished). Structural analysis of ΔNIC-EGFP (ideally compared to ΔNC-EGFP) is needed to determine how deletion of this region affects its structure.

It is also important that an alternative minimal MeCP2 protein has full bridge activity. I have already discussed methods of determining whether this is reduced in ΔNIC-EGFP (see section 7.3.2). If reduced activity is detected in one or more of these assays, they could be used to determine whether full activity is restored in alternative designs. Some of these assays (e.g. FRAP and SPR) require transiently-expressed or purified MeCP2 protein so could easily be used to test multiple designs. Other assays require the *Mecp2* alleles to be stably expressed – which is more labour intensive. Although I have discussed assays that use brain tissue from knock-in mice, it is likely that ES cell-derived neurons could also be used.

It is yet to be determined whether ΔNC-EGFP has mildly decreased NCoR/SMRT complex binding activity *in vivo* – as suggested by the slight reduction in HDAC3 that was co-immunoprecipitated from ΔNC-EGFP brains (compared to WT-EGFP and ΔN-EGFP samples). Further investigation of the bridge activity of ΔNC-EGFP protein (using the assays described in section 7.3.2) is needed to determine whether deletion of the Intervening region is solely responsible for reduced activity in ΔNIC-EGFP. If ΔNC-EGFP has reduced activity, it may be necessary to alter which residues in the N- and C-terminal regions are removed. Alternative epitope tags can also be used. If, however, ΔNC-EGFP does not have reduced activity, only the intervening linker sequence would need to be changed. The NID used in this study was extended at the N-terminal end to include a predicted α-helix (Fig. S4; see Chapter 3). This was almost certainly unnecessary as recent analysis of NCoR/SMRT complex binding by MeCP2 has reduced the minimal NID sequence to 298-309 (Kruusvee et al., 2017). New designs could therefore use a shorter NID sequence. One possible reason for decreased bridge activity in ΔNIC-EGFP is steric hindrance between the MBD and the NID as a result of bringing them closer

together. This could be overcome by replacing the Intervening region with a flexible linker sequence of similar length. While such a design would no longer infer such a great advantage for Gene Therapy due to its increased sequence length (see section 7.5.1), its use would still robustly test the bridge hypothesis as potentially dispensable native MeCP2 sequences have been removed.

7.4 Uncoupling the RTT phenotype from other proposed functions of MeCP2

Given the absence of RTT-like symptoms in ΔN-EGFP and ΔNC-EGFP hemizygous male mice, these lines could be used as a vital tool to uncouple some of the proposed interactions and their associated molecular phenotypes from disease. Although the shorter of the two truncations (ΔNC-EGFP) is missing half of the full-length amino acid sequence, only a few of the mapped binding sites are deleted. The N-terminus contains the full binding site for HP1ɣ (Agarwal et al., 2007) making the loss of this interaction likely in these truncation proteins. Deletion of the N-terminal region may also affect binding to FOXG1 (Dastidar et al., 2012). This region contains five identified post-translational modifications, of which only one (pSer13) has been studied (Yasui et al., 2014). Deletion of the C-terminus is highly likely to abolish binding to the splicing factors, FBP11 and HYPC (Buschdorf and Strätling, 2004), the miRNA processing component, DGCR8 (Cheng et al., 2014a), and the kinase, CDKL5 (Mari et al., 2005), as all four interactions have been shown to be lost by removal of C-terminal fragments. Additionally, the C-terminally truncated protein consisting of residues 1-308 was shown to only interact very weakly with a third splicing factor, YB-1 (Young et al., 2005). Further analysis is required to determine whether the addition of the extra four amino acids present in ΔNC-EGFP are sufficient to restore full YB-1 binding. The C-terminal region contains six identified post-translational modifications, of which only two (pSer421 and pSer424) have been characterised *in vivo* (Tao et al., 2009; Cohen et al., 2011; Li et al., 2011). Their phenotypes are discussed above in section 7.2. Lastly, the extreme C-terminus has been proposed to be required for intramolecular interactions that are regulated by Ser80 phosphorylation (Cheng et al., 2014a; Williams et al., 2016).

7.4.1 Confirming the loss of these interaction partners in the N- and C-terminally truncated proteins

Three functions stand out that could be affected by the deletion of both termini: complete loss of the ability to regulate miRNA processing (via DGCR8), and partial loss of alternative splicing regulation (via FBP11, HYPC and YB-1) and chromatin condensation (via HP1ɣ). To investigate whether the truncated proteins can perform these functions, I would first confirm their inability to bind the relevant MeCP2 interaction partners using immunoprecipitation experiments. These could be performed with brain nuclear extracts (using WT-EGFP brains as a positive control) to determine binding abilities in neurons; and using transient overexpression in HeLa cells, enabling additional truncations to be used as controls (for example, sequences matching the ones used in the original mapping of these interactions).

7.4.2 Uncoupling miRNA processing dysfunction from the RTT phenotype

If binding to the miRNA processing complex is lost in ΔNC-EGFP but not ΔN-EGFP, as predicted by its mapping to the C-terminus (Cheng et al., 2014a), miRNA processing defects would be predicted in the ΔNC-EGFP mice but not the ΔN-EGFP mice. Genome-wide miRNA processing defects could be analysed by RNA-sequencing of the non-coding RNA fraction and then comparing both these lines to the previous data obtained using null mice (Cheng et al., 2014a; Tsujimura et al., 2015). This experiment could uncouple miRNA processing defects (if they are present in ΔNC-EGFP mice) with RTT-like symptoms. A smaller scale experiment could focus on miR-199a, a miRNA reported to be processed by MeCP2 that regulates cell size via the AKT/mTOR pathway (Tsujimura et al., 2015). Levels of primary, precursor and mature miR-199a RNAs could be analysed by quantitative PCR to determine whether it is normally transcribed (primary transcript) but inefficiently processed (precursor and mature RNAs) in brains of the ΔNC-EGFP mice (as observed in nulls).

7.4.3 Uncoupling alternative splicing dysfunction from the RTT phenotype

Deletion of the C-terminus is likely to abolish binding to three of splicing factors that interact with full-length MeCP2: FBP11, HYPC and YB-1 (Buschdorf and Strätling, 2004; Young et al., 2005). As described above, loss of these interactions would first have to be confirmed. The 308/y mice could also be used as a comparator as this protein has been shown to have greatly reduced binding to YB-1 (Young et al., 2005) and is missing the other two interaction domains. mRNA-sequencing analysis could be used to determine whether the splicing defects observed in the null (Li et al., 2016) and 308/y (Young et al., 2005) mice are also present in the ΔNC-EGFP mice but not the ΔN-EGFP mice. This experiment could uncouple splicing defects (if they are present in ΔNC-EGFP mice) with RTT-like symptoms.

7.4.4 Uncoupling loss of HP1ɣ binding from the RTT phenotype

The least well-studied of these three functions is the role of the heterochromatin protein, HP1ɣ, when complexed to MeCP2. This interaction has been shown to require the deleted N-terminal region (Agarwal et al., 2007) so is likely to be abolished in ΔN-EGFP. If so, the myoblast system used by Brero et al (2005) could be used to determine whether ΔN-EGFP has reduced chromatin condensation activity compared to the WT-EGFP protein. This is measured by a reduced number of chromocentres due to their clustering. Loss of HP1ɣ binding by MeCP2 may result in lower levels of HP1ɣ localisation, and as a result reduced H3K9 trimethylation, at heterochromatin (Agarwal et al., 2007). This method may identify differences in chromatin remodelling activity between these proteins but it would tell us very little about how chromatin structure is affected in the brain by the loss of HP1ɣ binding. Although the novel 3D imaging technique (ChromATin) developed by Linhoff et al. (2015) detected increased chromatin density in *Mecp2*-null neurons in female *Mecp2+/-* brains, it is yet to be determined whether it can detect potentially subtler differences in neurons expressing mutant MeCP2 proteins that cause milder phenotypes, e.g. those that retain MBD function. It may be that this technique is not sensitive enough to detect a change in chromatin condensation in ΔN-EGFP mice,

which is predicted to be subtle given that all other domains in MeCP2 that have been attributed to chromatin remodelling activity are retained in this protein.

7.4.5 Uncoupling loss of intramolecular interactions from the RTT phenotype

It has long been considered that MeCP2 is highly unstructured (Ghosh et al., 2011). This belief is now being challenged with increased evidence of intramolecular interactions (Cheng et al., 2014a; Connelly and Belsom, unpublished). Additionally, changes in MeCP2 conformation upon S80 dephosphorylation have be proposed to regulate its activity, with intramolecular binding rendering it less active (Cheng et al., 2014a). It is therefore likely that loss of these interactions due to deletion of required regions of MeCP2 results in a constitutively active state. Cheng et al. (2014a) demonstrate the intramolecular interaction using two independent methods. Firstly, they showed direct interaction between N- and C-terminal fragments of MeCP2 by co-immunoprecipitation. As the division between these fragments is residue 305/306, this method is unsuitable for determining whether intramolecular binding is lost in proteins truncated after residue 312 (the C-terminal fragment would only consist of residues 306-312). The second strategy they used could, however, prove to be informative. They used FRET (Fluorescent Resonance Energy Transfer) between CFP (cyan fluorescent protein) fused to the N-terminus and YFP (yellow fluorescent protein) fused to the C-terminus to detect interactions between the two ends of transiently-expressed proteins in cultured cells. Expression in mouse cortical neurons enabled them to determine the effects of Ser80 dephosphorylation upon stimulation with KCl. They show that stimulation leads to an increase in intramolecular binding (FRET efficiency). This technique could be used to determine whether similar conformational changes occur upon neuronal stimulation in the truncated proteins. Their C-terminal EGFP tag could be utilised as part of the fluorescent protein pair in this analysis, combined with an N-terminal mCherry tag (Albertazzi et al., 2009). Loss of the activity-dependent intramolecular interaction in ΔNC-EGFP protein would uncouple this regulatory mechanism with RTT-like symptoms.

7.4.6 Preservation of transcriptional regulation in the phenotypically normal ΔN-EGFP and ΔNC-EGFP mice?

As well as uncoupling potentially less important roles, it is important to demonstrate correct gene repression via NCoR/SMRT complex recruitment in ΔN-EGFP and ΔNC-EGFP mice. The regions of MeCP2 protein retained in both truncations include all of the mapped sites for transcriptional repressors, which all lie within the historically defined Transcriptional Repression Domain (TRD). As the binding sites of the two transcriptional activators (CREB1 and MYCN) have not been mapped, it is impossible to predict whether the truncated proteins retain these interactions. If either, all the interactions with transcriptional repressors and activators are retained, or NCoR/SMRT recruitment is sufficient for proper MeCP2-mediated transcriptional regulation, then it could be predicted that gene expression patterns are unaltered in ΔN-EGFP and ΔNC-EGFP mice. RNA-sequencing analysis using tissue samples from these mouse lines would determine whether this is the case. Analysis of this data should pay particular attention to the regulation of long genes and genes that have potential phenotypic consequences.

A possible caveat to the interpretation of this experiment would be if defective miRNA processing or alternative splicing had downstream effects on mRNA transcript levels. Altered levels of mature miRNAs could affect the transcript levels of their target genes and mis-spliced transcripts could have altered stability. In this case, analysis of nascent transcription could be performed. Given the proposed role of MeCP2 in the regulation of transcriptional elongation in gene bodies (Kinde et al., 2016; Cholewa-Waclaw and Shah, unpublished), analysis of transcriptional elongation is likely to be most informative. Either Global nuclear Run On-sequencing (GRO-seq) or Native Elongating Transcript sequencing (NET-seq) could be used for this. GRO-seq involves the purification of nuclei, from which lysates are used for *in vitro* transcription with BrdUTP. The resulting nascent transcripts are purified with an antibody that recognises BrdUTP and then sequenced (Core et al., 2008). NET-seq is a newer technique that analyses nascent transcripts that have been transcribed *in vivo*. The chromatin fraction is obtained from nuclei and treated with

MNase before immunoprecipitation of elongating RNA polymerase II (phosphorylated at Ser2 in the C-terminal domain) using a phospho-specific antibody. The associated RNAs are then sequenced (Nojima et al., 2015). So far, only GRO-seq has been used for the analysis of nascent transcription in tissues (mouse liver by Fang et al., 2014). Analysis of nascent RNA transcription in *Mecp2*-mutant mouse models may be of particular value to the field as these techniques also detect spurious unstable transcripts, which are missed in standard RNA-seq experiments due to their rapid degradation. Skene et al. (2010) observed that loss of MeCP2 results in elevated transcription from repeated regions (analysed by qPCR) that can be detected in nuclear RNA but not total RNA preparations. It is possible that there are more global defects in spurious or abortive transcription that have so far been overlooked. A final complication is differentiating between a MeCP2-mediated global effect of transcription via recruitment of transcriptional repressors/activators or by altering global chromatin structure via recruitment of chromatin remodelling enzymes. HP1ɣ is the only protein involved in chromatin remodelling whose interaction is thought to be lost in the truncated proteins, ΔN-EGFP and ΔNC-EGFP.As HP1ɣ is found mainly in constitutive heterochromatin rather than gene-rich regions, loss of HP1ɣ-binding is likely to have minimal effects of transcription.

7.5 Therapeutic advantages of truncating MeCP2

7.5.1 The use of the truncated Mecp2 alleles in Gene Therapy

While the primary reason for investigating the functionality of truncated MeCP2 proteins in this study was to test the bridge hypothesis, the restricted packaging capacity of viral vectors used for Gene Therapy meant that reducing the essential coding sequence length could be a therapeutic benefit. To be useful for Gene Therapy, truncated MeCP2 has to be sufficient not only to prevent the onset of neurological symptoms (as seen in the knock-in mice that constitutively express these proteins), but it is also able to reverse these symptoms. I therefore used the Cre/*loxP* system designed by Guy et al. (2007) to induce activation of the shortest

truncated allele, ΔNIC-EGFP, in symptomatic MeCP2-deficient male mice, resulting in successful symptom reversal and prevention of premature death. This result suggests that recruitment of the NCoR/SMRT complex to chromatin is the key function of MeCP2 required for restoring function to mature MeCP2-deficient neurons. It is highly likely that similar results would be obtained using the other two less severe truncations: ΔN-EGFP and ΔNC-EGFP.

A recent Gene Therapy study using *Mecp2*-null male mice by our collaborators at the University of Glasgow (Stuart Cobb's lab) optimised conditions to improve transduction efficiency and symptom reversal. These are the best results to date, delaying premature death by ~26 weeks (Table 1.3.7; Gadalla et al., 2017). One of the key parameters in this study was the use of self-complementary AAV vectors, which have half the packaging capacity than other vectors (e.g. single stranded AAV vectors) – 2.2 kb of exogenous DNA. As the full-length *MECP2* cDNA sequence is 1.5 kb long, only 700 bp remained for regulatory elements. They developed a vector (Fig. 7.5.1) that included a 426 bp fragment of the mouse *Mecp2* promoter and selected regions of the 3'UTR (including conserved miRNA recognition sites). In order to explore the therapeutic potential of the truncated MeCP2 proteins, they must first be tested using the same optimised conditions developed by Gadalla et al. (2017). For this, the large C-terminal EGFP tag must be removed and replaced with a C-terminal Myc tag, consistent with the wild-type vector. Therefore, this experiment would determine the functionality of these proteins without EGFP. If viral delivery of these alleles can successfully prevent neurological symptoms in the *Mecp2*-null mice, further experiments can be performed using alternative regulatory elements that may better control dosage and improve symptomatic rescue. These regulatory elements can utilities the extra packaging capacity that results from shortening the coding sequence (Fig. 7.5.1).

A major concern for Gene Therapy is toxicity from overexpression: resulting in neurological dysfunction in the brain or liver toxicity if viral vectors preferentially infect the liver. The study by Gadalla et al. (2017) found that there is an extremely narrow dosage window between inefficient rescue and overexpression-induced

toxicity. Importantly, the permanence of Gene Therapy treatment means that too high dosages of virally-delivery DNA cannot be removed. Excitingly, the recent development of post-translational regulatory mechanisms could mean that MeCP2 dosage could be controlled at the protein level. Two examples of these mechanisms are the SMASh tag and the Destabilisation Domain (DD) of FKBP12, both of which can be fused to the protein of interest at either terminus, allowing induction of degradation (SMASh tag) or stabilisation (DD-FKBP12) by drugs that work in a dose-dependent manner (reviewed by Hannah and Zhou, 2015; Auffenberg et al., 2016). These systems would therefore allow the levels of MeCP2 protein to be modulated by increasing or decreasing the dose of these drugs. Another possible concern for Gene Therapy is that rapid introduction of MeCP2 was found to be toxic even in asymptomatic animals (Guy et al., 2007). All Gene Therapy studies have used a single viral injection, as the immune response makes repeated treatment impossible (Ciesielska et al., 2013). Contrary to the findings by Guy et al. (2007), three separate Gene Therapy studies reported symptomatic rescue in animals treated at 4-6 weeks of age (Gadalla et al., 2013; Garg et al., 2013; Matagne et al., 2017). A possible explanation for this discrepancy is the low transduction efficiency in the brains of these mice (2-25%) compared to the high recombination rate using the Cre/*loxP* system (~80%). It is therefore important to keep this observation in mind when optimising transduction efficiencies of Gene Therapy vectors. The use of regulatory tags may provide a mechanism for avoiding toxicity caused by to rapid reintroduction: by slowly increasing the stability of protein produced from virally-delivered genes. These post-translational regulatory mechanisms are of particular relevance my findings as the SMASh and DD-FKBP12 tags are ~300 bp and ~110 bp long, respectively, so would only fit in scAAV vectors in the coding sequence length is reduced.

The full stability of ΔNC-EGFP and absence of a discernible phenotype in the knock-in mice suggests that ΔNC would be a better choice than ΔNIC to use in Gene Therapy experiments. Although slightly longer than ΔNIC, it removal of the N- and C-terminal regions would create an extra ~700 bp of space for regulatory sequences. Unexpectedly, preliminary experiments in the Cobb lab using their optimised

conditions have found that ΔNIC provided better phenotypic rescue than ΔNC (Gadalla, Hector and Cobb, unpublished). This was particularly surprising as removal of the EGFP tag is likely to further reduced stability of ΔNIC. One possible explanation is that its decreased stability and/or activity reduces the risk of toxic overexpression. While MeCP2 protein dosage is widely regarded as crucial for neuronal function (reviewed by Chao and Zoghbi, 2012), it is important to note that having normal levels of MeCP2 is only 50% of cells (as in the case for RTT patients) causes much more severe symptoms in (female) mouse models than (male) hypomorphic mice that have ~50% of normal protein levels in all cells (Guy et al., 2001; Chen et al., 2001; Samaco et al., 2008; Kerr et al., 2008). Due to the lack of control when DNA is introduced by viral delivery, it may be that using a less stable/active protein for Gene Therapy results in a larger percentage of cells that lie within an optimal window (e.g. between 0.5 and 1.5-fold native MeCP2 activity). This is likely to be particularly important when treating female patients due to mosaicism. Thorough analysis of protein transduction efficiencies and the resulting MeCP2 expression levels in different cell types is therefore of vital importance when optimising *MECP2* Gene Therapy.

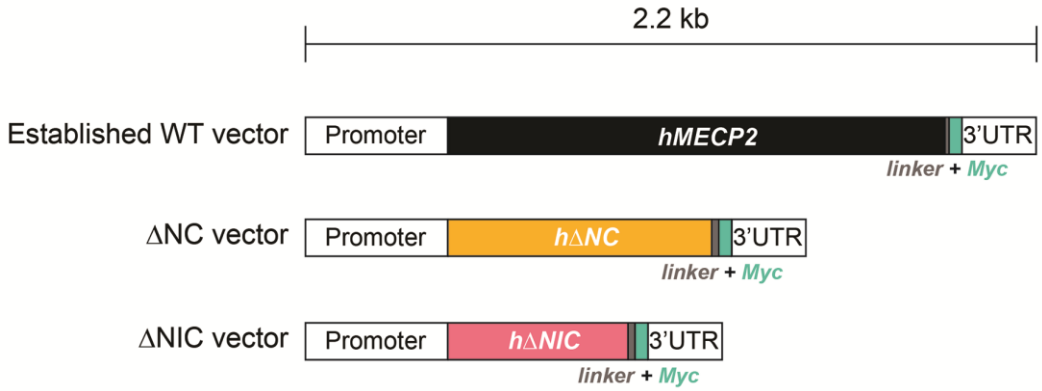


Figure 7.5.1: Gene Therapy vectors containing truncated *MECP2* alleles
Schematic of the vector developed by Gadalla et al. (2017) containing the human *MECP2* cDNA sequence with a C-terminal Myc tag, a 426 bp fragment of the mouse *Mecp2* promoter and selected elements from the *Mecp2* 3'UTR. The full-length cDNA sequence can be replaced by humanised versions of the ΔNC and ΔNIC alleles (replacing the EGFP tag with the Myc tag after the GSSGSSG linker sequence). This creates space for additional regulatory elements within the 2.2 kb capacity of scAAV vectors.

7.5.2 The use of the truncated Mecp2 alleles in Protein Replacement Therapy

Unlike for Gene Therapy, reducing the sequence length of MeCP2 has no direct advantages for Protein Replacement Therapy. These truncated proteins may, however, have altered properties that make them more suitable. The main obstacle to overcome in Protein Replacement Therapy trials is crossing the blood brain barrier (BBB) and entering neuronal cells. Further work is needed to test the transduction efficiency of these proteins with different tags (e.g. the TAT sequence). Altering the design of these proteins (e.g. by removing the EGFP tag or the extreme N-terminal sequences) may also improve transduction.

7.6 Concluding remarks

The mice produced in this study provide a valuable tool for determining whether the multiple roles attributed to MeCP2 are relevant for Rett syndrome pathophysiology. Analysis of their molecular phenotypes is likely to facilitate uncoupling of several protein interactions and their associated roles from the disease symptoms exhibited by *Mecp2*-null and other loss-of-function mice. Molecular analysis of these mice alongside those expressing the NID mutant, R306C-EGFP (Brown, Selfridge et al., 2016), would be particularly useful for furthering our understanding of the importance of NCoR/SMRT complex recruitment. This would allow differentiation between the molecular consequences of only disrupting NCoR/SMRT complex binding and retaining this interaction but abolishing progressively more of the other functions of MeCP2. The use of novel technologies to investigate nascent transcription in the brains of mice expressing truncated/mutant MeCP2 is a priority in furthering our understanding of the role of MeCP2 in regulating transcriptional elongation.

Identification of the key function of MeCP2 could be of therapeutic benefit. The most obvious advantage is the use of shortened alleles in Gene Therapy, allowing for larger regulatory elements within the limited capacity of scAAV viral vectors. The

truncated proteins may also be beneficial for Protein Replacement Therapy, if their altered properties lead to better production and/or transduction efficiency.

References

Adler, D.A., Quaderi, N.A., Brown, S.D.M., Chapman, V.M., Moore, J., Tate, P., and Disteche, C.M. (1995). The X-linked methylated DNA binding protein, Mecp2, is subject to X inactivation in the mouse. Mamm. Genome *6*, 491–492.

Agarwal, N., Hardt, T., Brero, A., Nowak, D., Rothbauer, U., Becker, A., Leonhardt, H., and Cardoso, M.C. (2007). MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. Nucleic Acids Res. *35*, 5402–5408.

Agarwal, N., Becker, A., Jost, K.L., Haase, S., Thakur, B.K., Brero, A., Hardt, T., Kudo, S., Leonhardt, H., and Cardoso, M.C. (2011). MeCP2 Rett mutations affect large scale chromatin organization. Hum. Mol. Genet. *20*, 4187–4195.

Albertazzi, L., Arosio, D., Marchetti, L., Ricci, F., and Beltram, F. (2009). Quantitative FRET analysis with the E0GFP-mCherry fluorescent protein pair. Photochem. Photobiol. *85*, 287–297.

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

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

Armstrong, D., Dunn, K., Antalffy, B., and Trivedi, R. (1995). Selective denritic alterations in the cortex of Rett Syndrome. J. Neuropathol. Exp. Neurol. *54*, 195–201.

Asaka, Y., Jugloff, D.G.M., Zhang, L., Eubanks, J.H., and Fitzsimonds, R.M. (2006). Hippocampal synaptic plasticity is impaired in the Mecp2-null mouse model of Rett syndrome. Neurobiol. Dis. *21*, 217–227.

Auffenberg, E., Jurik, A., Mattusch, C., Stoffel, R., Genewsky, A., Namendorf, C., Schmid, R.M., Rammes, G., Biel, M., Uhr, M., et al. (2016). Remote and reversible inhibition of neurons and circuits by small molecule induced potassium channel stabilization. Sci. Rep. *6*, 19293.

Ausió, J., Paz, A.M. de, and Esteller, M. (2014). MeCP2: the long trip from a chromatin protein to neurological disorders. Trends Mol. Med. *20*, 487–498.

Babbio, F., Castiglioni, I., Cassina, C., Gariboldi, M., Pistore, C., Magnani, E., Badaracco, G., Monti, E., and Bonapace, I. (2012). Knock-down of methyl CpG-binding protein 2 (MeCP2) causes alterations in cell proliferation and nuclear lamins expression in mammalian cells. BMC Cell Biol. *13*, 19.

Baker, S.A., Chen, L., Wilkins, A.D., Yu, P., Lichtarge, O., and Zoghbi, H.Y. (2013). An AT-hook domain in MeCP2 determines the clinical course of Rett syndrome and related disorders. Cell *152*, 984–996.

Baker, S.A., Lombardi, L.M., and Zoghbi, H.Y. (2015). Karyopherin alpha 3 and Karyopherin alpha 4 Mediate Nuclear Import of Methyl-CpG Binding Protein 2. J. Biol. Chem. *290*, 22485–22493.

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.

Balmer, D., Arredondo, J., Samaco, R.C., and LaSalle, J.M. (2002). MECP2 mutations in Rett syndrome adversely affect lymphocyte growth, but do not affect imprinted gene expression in blood or brain. Hum. Genet. *110*, 545–552.

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

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res. *21*, 381–395.

Barth, T.K., and Imhof, A. (2010). Fast signals and slow marks: the dynamics of histone modifications. Trends Biochem. Sci. *35*, 618–626.

Bauman, M.L., Kemper, T.L., and Arin, D.M. (1995). Microscopic observations of the brain in Rett syndrome. Neuropediatrics *26*, 105–108.

Becker, A., Allmann, L., Hofstätter, M., Casà, V., Weber, P., Lehmkuhl, A., Herce, H.D., and Cardoso, M.C. (2013). Direct Homo- and Hetero-Interactions of MeCP2 and MBD2. PLoS One *8*, e53730.

Becker, A., Zhang, P., Allmann, L., Meilinger, D., Bertulat, B., Eck, D., Hofstaetter, M., Bartolomei, G., Hottiger, M., Schreiber, V., et al. (2016). Poly(ADP-ribosyl)ation of Methyl CpG Binding Domain Protein 2 Regulates Chromatin Structure. J. Biol. Chem. *291*, 4873–4881.

Bellacosa, A., and Drohat, A.C. (2015). Role of base excision repair in maintaining the genetic and epigenetic integrity of CpG sites. DNA Repair (Amst). *32*, 33–42.

Bergo, A., Strollo, M., Gai, M., Barbiero, I., Stefanelli, G., Sertic, S., Gigli, C.C., Di Cunto, F., Kilstrup-Nielsen, C., and Landsberger, N. (2015). Methyl-CpG binding protein 2 (MeCP2) localizes at the centrosome and is required for proper mitotic spindle organization. J. Biol. Chem. *290*, 3223–3237.

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

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

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

Blackledge, N.P., Rose, N.R., and Klose, R.J. (2015). Targeting Polycomb systems to regulate gene expression: modifications to a complex story. Nat. Rev. Mol. Cell Biol. *16*, 643–649.

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., and Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. Proc. Natl. Acad. Sci. U. S. A. *100*, 7480–7485.

Bracaglia, G., Conca, B., Bergo, A., Rusconi, L., Zhou, Z., Greenberg, M.E., Landsberger, N., Soddu, S., and Kilstrup-Nielsen, C. (2009). Methyl-CpG-binding protein 2 is phosphorylated by homeodomain-interacting protein kinase 2 and contributes to apoptosis. EMBO Rep. *10*, 1327–1333.

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

Brown, K., Selfridge, J., Lagger, S., Connelly, J., Sousa, D. De, Webb, S., Guy, J., Merusi, C., Koerner, M. V, and Bird, A. (2016). The molecular basis of variable phenotypic severity among common missense mutations causing Rett syndrome. Hum. Mol. Genet. *25*, 558–570.

Buchovecky, C.M., Turley, S.D., Brown, H.M., Kyle, S.M., McDonald, J.G., Liu, B., Pieper, A. a, Huang, W., Katz, D.M., Russell, D.W., et al. (2013). A suppressor screen in Mecp2 mutant mice implicates cholesterol metabolism in Rett syndrome. Nat. Genet. *45*, 1013–1020.

Buschdorf, J.P., and Strätling, W.H. (2004). A WW domain binding region in methyl-CpG-binding protein MeCP2: Impact on Rett syndrome. J. Mol. Med. *82*, 135–143.

Caron, N.S., Hung, C.L., Atwal, R.S., and Truant, R. (2014). Live cell imaging and biophotonic methods reveal two types of mutant huntingtin inclusions. Hum. Mol. Genet. *23*, 2324–2338.

Cartron, P.-F., Nadaradjane, A., Lepape, F., Lalier, L., Gardie, B., and Vallette, F.M. (2013). Identification of TET1 Partners That Control Its DNA-Demethylating Function. Genes Cancer *4*, 235–241.

Chahrour, M., Jung, S.Y., Shaw, C., Zhou, X., Wong, S.T.C., Qin, J., and Zoghbi, H.Y. (2008). MeCP2, a key contributor to neurological disease, activates and represses transcription. Science (80-.). *320*, 1224–1229.

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

Chao, H.-T., and Zoghbi, H.Y. (2012). MeCP2: only 100% will do. Nat. Neurosci. *15*, 176–177.

Chao, H.-T., Chen, H., Samaco, R.C., Xue, M., Chahrour, M., Yoo, J., Neul, J.L., Gong, S., Lu, H.-C., Heintz, N., et al. (2010). Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. Nature *468*, 263–269.

Cheadle, J.P. (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, L.A., Baker, S.A., Shaw, C.A., Li, W., and Zoghbi, H.Y. (2015). MeCP2 binds to non-CG methylated DNA as neurons mature, influencing transcription and the timing of onset for Rett syndrome. Proc. Natl. Acad. Sci. *112*, 5509–5514.

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

Chen, W.G., Chang, Q., Lin, Y., Meissner, A., West, A.E., Griffith, E.C., Jaenisch, R., and Greenberg, M.E. (2003). Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science (80-.). *302*, 885–889.

Cheng, T.-L., Wang, Z., Liao, Q., Zhu, Y., Zhou, W.-H., Xu, W., and Qiu, Z. (2014a). MeCP2 suppresses nuclear microRNA processing and dendritic growth by regulating the DGCR8/Drosha complex. Dev. Cell *28*, 547–560.

Cheng, J., Huang, M., Zhu, Y., Xin, Y.J., Zhao, Y.K., Huang, J., Yu, J.X., Zhou, W.H., and Qiu, Z. (2014b). SUMOylation of MeCP2 is essential for transcriptional repression and hippocampal synapse development. J. Neurochem. *128*, 798–806.

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.

Ciesielska, A., Hadaczek, P., Mittermeyer, G., Zhou, S., Wright, F., Bankiewicz, K., and Forsayeth, J. (2013). Cerebral Infusion of AAV9 Vector-encoding Non-self Proteins Can Elicit Cell-mediated Immune Responses. Mol. Ther. *21*, 158–166.

Clayton-Smith, J., Watson, P., Ramsden, S., and Black, G.C.M. (2000). Somatic mutation in MECP2 as a non-fatal neurodevelopmental disorder in males. Lancet *356*, 830–832.

Cohen, S., Gabel, H.W., Hemberg, M., Hutchinson, A.N., Sadacca, L.A., Ebert, D.H., Harmin, D.A., Greenberg, R.S., Verdine, V.K., Zhou, Z., et al. (2011). Genome-Wide Activity-Dependent MeCP2 Phosphorylation Regulates Nervous System Development and Function. Neuron *72*, 72–85.

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

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (80-.). *322*, 1845–1848.

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

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.

Cukier, H.N., Rabionet, R., Konidari, I., Rayner-Evans, M.Y., Baltos, M.L., Wright, H.H., Abramson, R.K., Martin, E.R., Cuccaro, M.L., Pericak-Vance, M.A., et al. (2010). Novel variants identified in methyl-CpG-binding domain genes in autistic individuals. Neurogenetics *11*, 291–303.

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

Dahl, C., Grønbæk, K., and Guldberg, P. (2011). Advances in DNA methylation: 5-hydroxymethylcytosine revisited. Clin. Chim. Acta *412*, 831–836.

Dastidar, S.G., Bardai, F.H., Ma, C., Price, V., Rawat, V., Verma, P., Narayanan, V., and D'Mello, S.R. (2012). Isoform-Specific Toxicity of Mecp2 in Postmitotic Neurons: Suppression of Neurotoxicity by FoxG1. J. Neurosci. *32*, 2846–2855.

Deaton, A., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Dhawan, S., Georgia, S., Tschen, S. ing, Fan, G., and Bhushan, A. (2011). Pancreatic Beta Cell Identity Is Maintained by DNA Methylation-Mediated Repression of Arx. Dev. Cell *20*, 419–429.

Dintilhac, A., and Bernués, J. (2002). HMGB1 interacts with many apparently unrelated proteins by recognizing short amino acid sequences. J. Biol. Chem. *277*, 7021–7028.

Du, F., Nguyen, M.V.C., Karten, A., Felice, C.A., Mandel, G., and Ballas, N. (2016). Acute and crucial requirement for MeCP2 function upon transition from early to late adult stages of brain maturation. Hum. Mol. Genet. *25*, 1690–1702.

Du, Q., Luu, P.-L., Stirzaker, C., and Clark, S.J. (2015). Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics *7*, 1051–1073.

Ebert, D.H., Gabel, H.W., Robinson, N.D., Kastan, N.R., Hu, L.S., Cohen, S., Navarro, A.J., Lyst, M.J., Ekiert, R., Bird, A.P., et al. (2013). Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. Nature *499*, 341–345.

ePESTfind Predictor Tool (http://emboss.bioinformatics.nl/cgi-bin/emboss/epestfind)/ May 2017

Escamilla-Del-Arenal, M., Da Rocha, S.T., and Heard, E. (2011). Evolutionary diversity and developmental regulation of X-chromosome inactivation. Hum. Genet. *130*, 307–327.

Van Esch, H. (2011). MECP2 duplication syndrome. Mol. Syndromol. *2*, 128–136.

Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L.R., Ge, J., et al. (2005). Duplication of the MECP2 Region Is a Frequent Cause of Severe Mental Retardation and Progressive Neurological Symptoms in Males. Am. J. Hum. Genet. *77*, 442–453.

Exome Aggregation Consortium (ExAC), Cambridge, MA (http://exac.broadinstitute.org)/ September 2016

Fang, B., Everett, L.J., Jager, J., Briggs, E., Armour, S.M., Feng, D., Roy, A., Gerhart-Hines, Z., Sun, Z., and Lazar, M.A. (2014). Circadian enhancers coordinate multiple phases of rhythmic gene transcription in vivo. Cell *159*, 1140–1152.

Fendri-Kriaa, N., Mkaouar-Rebai, E., Moalla, D., Belguith, N., Louhichi, N., Zemni, R., Slama, F., Triki, C., and Fakhfakh, F. (2010). Mutational analysis of the MECP2 gene in Tunisian patients with Rett syndrome: a novel double mutation. J. Child Neurol. *25*, 1042–1046.

Forlani, G., Giarda, E., Ala, U., Di Cunto, F., Salani, M., Tupler, R., Kilstrup-Nielsen, C., and Landsberger, N. (2010). The MeCP2/YY1 interaction regulates ANT1 expression at 4q35: novel hints for Rett syndrome pathogenesis. Hum. Mol. Genet. *19*, 3114–3123.

Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suñer, D., Cigudosa, J.C., Urioste, M., Benitez, J., et al. (2005). Epigenetic differences arise during the lifetime of monozygotic twins. Proc. Natl. Acad. Sci. U. S. A. *102*, 10604–10609.

Free, A., Wakefield, R.I.D., Smith, B.O., Drydenll, D.T.F., Barlow, P.N., and Bird, A.P. (2001). DNA Recognition by the Methyl-CpG Binding Domain of MeCP2. J. Biol. Chem. *276*, 3353–3360.

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

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.

Gadalla, K.K.E., Bailey, M.E.S., and Cobb, S.R. (2011). MeCP2 and Rett syndrome: reversibility and potential avenues for therapy. Biochem. J. *439*, 1–14.

Gadalla, K.K.E., Bailey, M.E.S., Spike, R.C., Ross, P.D., Woodard, K.T., Kalburgi, S.N., Bachaboina, L., Deng, J. V, West, A.E., Samulski, R.J., et al. (2013). Improved survival and reduced phenotypic severity following AAV9/MECP2 gene transfer to neonatal and juvenile male Mecp2 knockout mice. Mol. Ther. *21*, 18–30.

Gadalla, K.K.E., Vudhironarit, T., Hector, R.D., Sinnett, S., Bahey, N.G., Bailey, M.E.S., Gray, S.J., and Cobb, S.R. (2017). Development of a Novel AAV Gene Therapy Cassette with Improved Safety Features and Efficacy in a Mouse Model of Rett Syndrome. Mol. Ther. - Methods Clin. Dev. *5*, 180–190.

Garg, S.K., Lioy, D.T., Cheval, H., McGann, J.C., Bissonnette, J.M., Murtha, M.J., Foust, K.D., Kaspar, B.K., Bird, A., and Mandel, G. (2013). Systemic delivery of MeCP2 rescues behavioral and cellular deficits in female mouse models of Rett syndrome. J. Neurosci. *33*, 13612–13620.

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

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

Ghosh, R.P., Nikitina, T., Horowitz-scherer, R.A., Gierasch, L.M., Uversky, V.N., Hite, K., Hansen, J.C., and Christopher, L. (2011). Unique physical properties and interactions of the domains of methylated DNA binding protein 2 (MeCP2). Biochemistry *49*, 4395–4410.

Goffin, D., Allen, M., Zhang, L., Amorim, M., Wang, I.-T.J., Reyes, A.-R.S., 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*, 274–283.

Gonzales, M.L., Adams, S., Dunaway, K.W., and LaSalle, J.M. (2012). Phosphorylation of distinct sites in MeCP2 modifies cofactor associations and the dynamics of transcriptional regulation. Mol. Cell. Biol. *32*, 2894–2903.

Grasshoff, U., Bonin, M., Goehring, I., Ekici, A., Dufke, A., Cremer, K., Wagner, N., Rossier, E., Jauch, A., Walter, M., et al. (2011). De novo MECP2 duplication in two females with random X-inactivation and moderate mental retardation. Eur. J. Hum. Genet. *19226*, 507–512.

Gregory, R.I., Randall, T.E., Johnson, C. a, Khosla, S., Hatada, I., Neill, L.P.O., Bryan, M., Feil, R., and Turner, B.M. (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.

Guarda, A., Bolognese, F., Bonapace, I.M., and Badaracco, G. (2009). Interaction between the inner nuclear membrane lamin B receptor and the heterochromatic methyl binding protein, MeCP2. Exp. Cell Res. *315*, 1895–1903.

Le Guezennec, X., Vermeulen, M., Brinkman, A.B., Hoeijmakers, W. a M., Cohen, A., Lasonder, E., and Stunnenberg, H.G. (2006). MBD2 / NuRD and MBD3 / NuRD , Two Distinct Complexes with Different Biochemical and Functional Properties. Mol. Cell. Biol. *26*, 843–851.

Guo, J.U., Su, Y., Shin, J.H., Shin, J., Li, H., Xie, B., Zhong, C., Hu, S., Le, T., Fan, G., et al. (2014). Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. Nat. Neurosci. *17*, 215–222.

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.

Guy, J., Gan, J., Selfridge, J., Cobb, S., and Bird, A. (2007). Reversal of neurological defects in a mouse model of Rett syndrome. Science (80-.). *315*, 1143–1147.

Guy, J., Cheval, H., Selfridge, J., and Bird, A. (2011). The role of MeCP2 in the brain. Annu. Rev. Cell Dev. Biol. *27*, 631–652.

Haas, R.H. (1988). The history and challenge of Rett syndrome. J. Child Neurol. *3 Suppl*, S3–S5.

Hagberg, B. (1985). Rett's Syndrome : Prevalence and Impact on Progressive Severe Mental Retardation in Girls. Acta Paediatr. Scand. *74*, 405–408.

Hagberg, B., Aicardi, J., Dias, K., and 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., Goutières, F., Hanefeld, F., Rett, a, and Wilson, J. (1985). Rett syndrome: criteria for inclusion and exclusion. Brain Dev. *7*, 372–373.

Hagberg, G., Stenbom, Y., and Witt Engerström, I. (2001). Head growth in Rett syndrome. Brain Dev. *23*, S227–S229.

Hannah, J., and Zhou, P. (2015). Methods: A small-molecule SMASh hit. Nat. Chem. Biol. *11*, 637–638.

Harikrishnan, K.N., Chow, M.Z., Baker, E.K., Pal, S., Bassal, S., Brasacchio, D., Wang, L., Craig, J.M., Jones, P.L., Sif, S., et al. (2005). Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat. Genet. *37*, 254–264.

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.

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 *eLife2014*, e02676.

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

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

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

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

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

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262–1278.

Huttlin, E.L., Jedrychowski, M.P., Elias, J.E., Goswami, T., Rad, R., Beausoleil, S.A., Villén, J., Haas, W., Sowa, M.E., and Gygi, S.P. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. Cell *143*, 1174–1189.

Hwang, C.K., Kim, C.S., Kim, D.K., Law, P., Wei, L., and Loh, H.H. (2010). Up-regulation of the mu-opioid receptor gene is mediated through chromatin remodeling and transcriptional factors in differentiated neuronal cells. Mol. Pharmacol. *78*, 58–68.

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

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

Jellinger, K., and Seitelberger, F. (1986). Neuropathology of Rett Syndrome. Am. J. Med. Genet. *Supplement*, 259–288.

Jentarra, G.M., Olfers, S.L., Rice, S.G., Srivastava, N., Homanics, G.E., Blue, M., Naidu, S., and Narayanan, V. (2010). Abnormalities of cell packing density and dendritic complexity in the MeCP2 A140V mouse model of Rett syndrome/X-linked mental retardation. BMC Neurosci. *11*, 19.

Jiang, M., Ash, R.T., Baker, S.A., Suter, B., Ferguson, A., Park, J., Rudy, J., Torsky, S.P., Chao, H.-T., Zoghbi, H.Y., et al. (2013). Dendritic arborization and spine dynamics are abnormal in the mouse model of MECP2 duplication syndrome. J. Neurosci. *33*, 19518–19533.

Jørgensen, H.F., Adie, K., Chaubert, P., and Bird, A.P. (2006). Engineering a high-affinity methyl-CpG-binding protein. Nucleic Acids Res. *34*, e96.

Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A Short Amino Acid Sequence Able to Sepcify Nuclear Location. Cell *39*, 499–509.

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

Kankirawatana, P., Leonard, H., Ellaway, C., Scurlock, J., Mansour, A., Makris, C.M., Dure IV, L.S., Friez, M., Lane, J., Kiraly-Borri, C., et al. (2006). Early progressive encephalopathy in boys and MECP2 mutations. Neurology *67*, 164–166.

Katz, D.M., Bird, A., Coenraads, M., Gray, S.J., Menon, D.U., Philpot, B.D., and Tarquinio, D.C. (2016). Rett Syndrome: Crossing the Threshold to Clinical Translation. Trends Neurosci. *39*, 100–113.

Kernohan, K.D., Jiang, Y., Tremblay, D.C., Bonvissuto, A.C., Eubanks, J.H., Mann, M.R.W., and Bérubé, N.G. (2010). ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. Dev. Cell *18*, 191–202.

Kernohan, K.D., Vernimmen, D., Gloor, G.B., and Bérubé, N.G. (2014). Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. Nucleic Acids Res. *42*, 8356–8368.

Kerr, A.M., and Stephenson, J.B.P. (1985). Clinical Topics Rett's syndrome in the west of Scotland. Br. Med. J. *291*, 579–582.

Kerr, B., Alvarez-Saavedra, M., Sáez, M. a, Saona, A., and Young, J.I. (2008). Defective body-weight regulation, motor control and abnormal social interactions in Mecp2 hypomorphic mice. Hum. Mol. Genet. *17*, 1707–1717.

Kerr, B., Soto C, J., Saez, M., Abrams, A., Walz, K., and Young, J.I. (2012). Transgenic complementation of MeCP2 deficiency: phenotypic rescue of Mecp2-null mice by isoform-specific transgenes. Eur. J. Hum. Genet. *20*, 69–76.

Khoshnan, A., and Patterson, P.H. (2012). Elevated IKKalpha accelerates the differentiation of human neuronal progenitor cells and induces MeCP2-dependent BDNF expression. PLoS One *7*, e41794.

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

Kinde, B., Wu, D.Y., Greenberg, M.E., and Gabel, H.W. (2016). DNA methylation in the gene body influences MeCP2-mediated gene repression. Proc. Natl. Acad. Sci. *113*, 15114–15119.

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

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

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

Klose, R.J., Sarraf, S. a, Schmiedeberg, L., McDermott, S.M., Stancheva, I., and Bird, A.P. (2005). DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Mol. Cell *19*, 667–678.

Klungland, A., and Robertson, A.B. (2016). Oxidized C5-methyl cytosine bases in DNA: 5-Hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine. Free Radic. Biol. Med. http://dx.doi.org/10.1016/j.freeradbiomed.2016.11.

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

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

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science (80-.). *324*, 929–930.

Kruusvee, V., Lyst, M.J., Taylor, C., Tarnauskaitė, Ž., Bird, A.P., and Cook, A.G. (2017). Structure of the MeCP2–TBLR1 complex reveals a molecular basis for Rett syndrome and related disorders. Proc. Natl. Acad. Sci. 17007311114.

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.

Kuo, L.J., and Yang, L.-X. (2008). Gamma-H2AX - a novel biomarker for DNA double-strand breaks. In Vivo (Brooklyn). *22*, 305–309.

Lagger, S., Connelly, J.C., Schweikert, G., Webb, S., Selfridge, J., Ramsahoye, B.H., Yu, M., He, C., Sanguinetti, G., Sowers, L.C., et al. (2017). MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. PLOS Genet. *13*, e1006793.

Lalonde, R., and Strazielle, C. (2011). Brain regions and genes affecting limb-clasping responses. Brain Res. Rev. *67*, 252–259.

Lamonica, J.M., Kwon, D.Y., Goffin, D., Fenik, P., Johnson, B.S., Cui, Y., Guo, H., Veasey, S., and Zhou, Z. (2017). Elevating expression of MeCP2 T158M rescues DNA binding and Rett syndrome – like phenotypes. J. Clin. Invest. *127*, 1889–1904.

Lawson-Yuen, A., Liu, D., Han, L., Jiang, Z.I., Tsai, G.E., Basu, A.C., Picker, J., Feng, J., and Coyle, J.T. (2007). Ube3a mRNA and protein expression are not decreased in Mecp2R168X mutant mice. Brain Res. *1180*, 1–6.

Lek, M., Karczewski, K.J., Minikel, E. V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, a (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905–914.

Li, H., Yamagata, T., Mori, M., Yasuhara, A., and Momoi, M.Y. (2005). Mutation analysis of methyl-CpG binding protein family genes in autistic patients. Brain Dev. *27*, 321–325.

Li, H., Zhong, X., Chau, K.F., Williams, E.C., and Chang, Q. (2011). Loss of activity-induced phosphorylation of MeCP2 enhances synaptogenesis, LTP and spatial memory. Nat. Neurosci. *14*, 1001–1008.

Li, R., Dong, Q., Yuan, X., Zeng, X., Gao, Y., Chiao, C., Li, H., Zhao, X., Keles, S., Wang, Z., et al. (2016). Misregulation of Alternative Splicing in a Mouse Model of Rett Syndrome. PLoS Genet. *12*, 1–25.

Li, Y., Wang, H., Muffat, J., Cheng, A.W., Orlando, D.A., Kwok, S., Feldman, D.A., Bateup, H.S., Gao, Q., Mitalipova, M., et al. (2013). Global transcriptional and translational repression in human embryonic stem cells-derived Rett Syndrome neurons. Cell Stem Cell *13*, 446–458.

Lin, Q., Jo, D., Gebre-Amlak, K.D., and Ruley, H.E. (2004). Enhanced cell-permeant Cre protein for site-specific recombination in cultured cells. BMC Biotechnol. *4*, 25.

Linhoff, M.W., Garg, S.K., and Mandel, G. (2015). A High-Resolution Imaging Approach to Investigate Chromatin Architecture in Complex Tissues. Cell *163*, 246–255.

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 (80-.). *341*, 1237905.

Lombardi, L.M., Baker, S.A., and Zoghbi, H.Y. (2015). MECP2 disorders : from the clinic to mice and back. J. Clin. Invest. *125*, 2914–2923.

Long, S.W., Ooi, J.Y., Yau, P.M., and Jones, P.L. (2011). A brain-derived MeCP2 complex supports a role for MeCP2 in RNA processing. Biosci Rep *31*, 333–343.

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

Lyst, M.J., and Bird, A. (2015). Rett syndrome: a complex disorder with simple roots. Nat. Rev. Genet. *16*, 261–274.

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.

Lyst, M.J., Connelly, J., Merusi, C., and Bird, A. (2016). Sequence-specific DNA binding by AT-hook motifs in MeCP2. FEBS Lett. *590*, 2927–2933.

Lyu, J.-W., Yuan, B., Cheng, T.-L., Qiu, Z.-L., and Zhou, W.-H. (2016). Reciprocal regulation of autism-related genes MeCP2 and PTEN via microRNAs. Sci. Rep. *6*, 20392.

Ma, L.Y., Wu, C., Jin, Y., Gao, M., Li, G.H., Turner, D., Shen, J.X., Zhang, S.J., Narayanan, V., Jentarra, G., et al. (2014). Electrophysiological phenotypes of MeCP2 A140V mutant mouse model. CNS Neurosci. Ther. *20*, 420–428.

Makałowski, W., and Boguski, Mark, S. (1998). Evolutionary parameters of the transcribed mammalian genome : An analysis of 2,820 orthologous rodent and human sequences. Proc. Natl. Acad. Sci. *95*, 9407–9412.

Mann, J., Oakley, F., Akiboye, F., Elsharkawy, a, Thorne, a W., and Mann, D. a (2007). Regulation of myofibroblast transdifferentiation by DNA methylation and MeCP2: implications for wound healing and fibrogenesis. Cell Death Differ. *14*, 275–285.

Mari, F., Azimonti, S., Bertani, I., Bolognese, F., Colombo, E., Caselli, R., Scala, E., Longo, I., Grosso, S., Pescucci, C., et al. (2005). CDKL5 belongs to the same molecular pathway of MeCP2 and it is responsible for the early-onset seizure variant of Rett syndrome. Hum. Mol. Genet. *14*, 1935–1946.

Martin Caballero, I., Hansen, J., Leaford, D., Pollard, S., and Hendrich, B.D. (2009). The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for

mouse embryogenesis, but playa redundant function in neural differentiation. PLoS One *4*, e4315.

Masuyama, T., Matsuo, M., Jing, J.J., Tabara, Y., Kitsuki, K., Yamagata, H., Kan, Y., Miki, T., Ishii, K., and Kondo, I. (2005). Classic Rett syndrome in a boy with R133C mutation of MECP2. Brain Dev. *27*, 439–442.

Matagne, V., Ehinger, Y., Saidi, L., Borges-Correia, A., Barkats, M., Bartoli, M., Villard, L., and Roux, J.-C. (2017). A codon-optimized Mecp2 transgene corrects breathing deficits and improves survival in a mouse model of Rett syndrome. Neurobiol. Dis. *99*, 1–11.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Cell Res. *23*, 1256–1269.

McFarland, K.N., Huizenga, M.N., Darnell, S.B., Sangrey, G.R., Berezovska, O., Cha, J.H.J., Outeiro, T.F., and Sadri-Vakili, G. (2014). MeCP2: A novel huntingtin interactor. Hum. Mol. Genet. *23*, 1036–1044.

McGill, B.E., Bundle, S.F., Yaylaoglu, M.B., Carson, J.P., Thaller, C., and Zoghbi, H.Y. (2006). Enhanced anxiety and stress-induced corticosterone release are associated with increased Crh expression in a mouse model of Rett syndrome. Proc. Natl. Acad. Sci. U. S. A. *103*, 18267–18272.

McGowan, H., and Pang, Z.P. (2015). Regulatory functions and pathological relevance of the MECP2 3’UTR in the central nervous system. Cell Regen. *4*.

McGraw, C.M., Samaco, R.C., and Zoghbi, H.Y. (2011). Adult neural function requires MeCP2. Science (80-.). *333*, 186.

Meehan, R., Lewis, J.D., and Bird, A.P. (1992). Characterization of MECP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. *20*, 5085–5092.

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.

Mellén, 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.

Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. Science (80-.). *297*, 403–405.

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

Moog, U., Smeets, E.E.J., Van Roozendaal, K.E.P., Schoenmakers, S., Herbergs, J., Schoonbrood-Lenssen, A.M.J., and Schrander-Stumpel, C.T.R.M. (2003). Neurodevelopmental disorders in males related to the gene causing Rett syndrome in females (MECP 2). Eur. J. Paediatr. Neurol. *7*, 5–12.

Moretti, P., Bouwknecht, J.A., Teague, R., Paylor, R., and Zoghbi, H.Y. (2005). Abnormalities of social interactions and home-cage behavior in a mouse model of Rett syndrome. Hum. Mol. Genet. *14*, 205–220.

Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmiller, V., Michalakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. Angew. Chemie - Int. Ed. *49*, 5375–5377.

Muotri, A.R., Marchetto, M.C., Coufal, N.G., Oefner, R., Yeo, G., Nakashima, K., and Gage, F.H. (2010). L1 retrotransposition in neurons is modulated by MeCP2. Nature *468*, 443–446.

Murphy, D.M., Buckley, P.G., Das, S., Watters, K.M., Bryan, K., and Stallings, R.L. (2011). Co-localization of the oncogenic transcription factor MYCN and the DNA methyl binding protein MeCP2 at genomic sites in neuroblastoma. PLoS One *6*, e21436.

Nagai, K., Miyake, K., and Kubota, T. (2005). A transcriptional repressor MeCP2 causing Rett syndrome is expressed in embryonic non-neuronal cells and controls their growth. Dev. Brain Res. *157*, 103–106.

Naidu, S., Kitt, C.A., and Dean, F. (1988). Research on Rett Syndrome : Strategy and Preliminar Results. J. Child Neurol. *3*, S78–S86.

Nan, X., Meehan, R.R., and Bird, A. (1993). Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res. *21*, 4886–4892.

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

Nan, X., Campoy, F.J., and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell *88*, 471–481.

Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature *393*, 386–389.

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

Neul, J.L. (2012). The relationship of Rett syndrome and MECP2 disorders to autism. Transl. Res. *14*, 253–262.

Neul, J.L., Kaufmann, W.E., Glaze, D.G., Christodoulou, J., Clarke, A.J., Bahi-Buisson, N., Leonard, H., Bailey, M.E.S., Schanen, N.C., Zappella, M., et al. (2010).

Rett syndrome: Revised diagnostic criteria and nomenclature. Ann. Neurol. *68*, 944–950.

Nguyen, M.V.C., Du, F., Felice, C. a., Shan, X., Nigam, a., Mandel, G., Robinson, J.K., and 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*, 10021–10034.

Nikitina, T., Ghosh, R.P., Horowitz-Scherer, R. a., Hansen, J.C., Grigoryev, S. a., and Woodcock, C.L. (2007). MeCP2-Chromatin Interactions Include the Formation of Chromatosome-like Structures and Are Altered in Mutations Causing Rett Syndrome. J. Biol. Chem. *282*, 28237–28245.

Nojima, T., Gomes, T., Grosso, A.R.F., Kimura, H., Dye, M.J., Dhir, S., Carmo-Fonseca, M., and Proudfoot, N.J. (2015). Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. Cell *161*, 526–540.

Nomura, Y., Segawa, M., and Hasegawa, M. (1984). Rett syndrome--clinical studies and pathophysiological consideration. Brain Dev. *6*, 475–486.

O'Donovan, K.J., and Darnell, R.B. (2001). Neuronal signaling through alternative splicing: some exons CaRRE... Sci. STKE *2001*, pe2.

Oberoi, J., Fairall, L., Watson, P.J., Yang, J.-C., Czimmerer, Z., Kampmann, T., Goult, B.T., Greenwood, J.A., Gooch, J.T., Kallenberger, B.C., et al. (2011). Structural basis for the assembly of the SMRT/NCoR core transcriptional repression machinery. Nat. Struct. Mol. Biol. *18*, 177–184.

Pandey, S., Jr, G.E.S., Malyarchuk, S., Calhoun, T.N., and Pruitt, K. (2015). A novel MeCP2 acetylation site regulates interaction with ATRX and HDAC1. Genes Cancer *6*, 408–421.

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

Pelka, G.J., Watson, C.M., Christodoulou, J., and Tam, P.P.L. (2005). Distinct expression profiles of Mecp2 transcripts with different lengths of 3'UTR in the brain and visceral organs during mouse development. Genomics *85*, 441–452.

Percy, A.K. (1988). Research in Rett Syndrome: Past, Present, and Future. J. Child Neurol. *3*, S72–S75.

Pitcher, M.R., Herrera, J.A., Buffington, S.A., Kochukov, M.Y., Merritt, J.K., Fisher, A.R., Schanen, N.C., Costa-Mattioli, M., and Neul, J.L. (2015). Rett syndrome like phenotypes in the R255X Mecp2 mutant mouse are rescued by MECP2 transgene. Hum. Mol. Genet. *24*, 2662–2672.

PHD Secondary Structure Prediction (https://npsa-prabi.ibcp.fr/NPSA/npsa_phd.html)/ October 2013 and October 2016

Pradeepa, M.M. (2017). Causal role of histone acetylations in enhancer function. Transcription *8*, 40–47.

Quaderi, N.A., Meehan, R.R., Tate, P.H., Cross, S.H., Bird, A.P., Chatterjee, A., Herman, G.E., and Brown, S.D.M. (1994). Genetic and Physical Mapping of a Gene Encoding a Methyl CpG Binding Protein, MeCP2, to the Mouse X Chromosome. Genomics *22*, 648–651.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, a P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. U. S. A. *97*, 5237–5242.

Ran, F.A., Cong, L., Yan, W.X., Scott, D. a., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., et al. (2015). In vivo genome editing using Staphylococcus aureus Cas9. Nature *520*, 186–190.

Reardon, W., Donoghue, V., Murphy, A.M., King, M.D., Mayne, P.D., Horn, N., and Birk Møller, L. (2010). Progressive cerebellar degenerative changes in the severe

mental retardation syndrome caused by duplication of MECP2 and adjacent loci on Xq28. Eur. J. Pediatr. *169*, 941–949.

Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. Nature *447*, 425–432.

RettBase: Rett Syndrome Variation Database (RettBASE: http://mecp2.chw.edu.au)/ September 2016

Ricciardi, S., Boggio, E.M., Grosso, S., Lonetti, G., Forlani, G., Stefanelli, G., Calcagno, E., Morello, N., Landsberger, N., Biffo, S., et al. (2011). Reduced AKT/mTOR signaling and protein synthesis dysregulation in a Rett syndrome animal model. Hum. Mol. Genet. *20*, 1182–1196.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG Islands genome wide. Mol. Cell *55*, 347–360.

Ringrose, L., and Paro, R. (2007). Polycomb/Trithorax response elements and epigenetic memory of cell identity. Development *134*, 223–232.

Robinson, L., Guy, J., McKay, L., Brockett, E., Spike, R.C., Selfridge, J., De Sousa, D., Merusi, C., Riedel, G., Bird, A., et al. (2012). Morphological and functional reversal of phenotypes in a mouse model of Rett syndrome. Brain *135*, 2699–2710.

Roloff, T.C., Ropers, H.H., and Nuber, U.A. (2003). Comparative study of methyl-CpG-binding domain proteins. BMC Genomics *4*, 1.

Ross, P.D., Guy, J., Selfridge, J., Kamal, B., Tanner, E., Gillingwater, T.H., Jones, R.A., Loughrey, C.M., Mccarroll, C.S., Adrian, M.E.S.B., et al. (2016). Exclusive expression of MeCP2 in the nervous system distinguishes between brain and peripheral Rett syndrome-like phenotypes. Hum. Mol. Genet. *25*, 4389–4404.

Royant, A., and Noirclerc-Savoye, M. (2011). Stabilizing role of glutamic acid 222 in the structure of Enhanced Green Fluorescent Protein. J. Struct. Biol. *174*, 385–390.

Rugg-Gunn, P.J., Cox, B.J., Ralston, A., and Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc. Natl. Acad. Sci. U. S. A. *107*, 10783–10790.

Salomão Schwartzman, J., Zatz, M., Vasquez, L. dos R., Gomes, R.R., Koiffmann, C.P., Fridman, C., and Otto, P.G. (1999). Rett Syndrome in a Boy with a 47,XXY Karyotype. Am. J. Hum. Genet. *64*, 1781–1784.

Samaco, R.C., Nagarajan, R.P., Braunschweig, D., and LaSalle, J.M. (2004). Multiple pathways regulate MeCP2 expression in normal brain development and exhibit defects in autism-spectrum disorders. Hum. Mol. Genet. *13*, 629–639.

Samaco, R.C., Fryer, J.D., Ren, J., Fyffe, S., Chao, H.T., Sun, Y., Greer, J.J., Zoghbi, H.Y., and Neul, J.L. (2008). A partial loss of function allele of Methyl-CpG-binding protein 2 predicts a human neurodevelopmental syndrome. Hum. Mol. Genet. *17*, 1718–1727.

Samaco, R.C., Mandel-Brehm, C., McGraw, C.M., Shaw, C. a, McGill, B.E., and Zoghbi, H.Y. (2012). Crh and Oprm1 mediate anxiety-related behavior and social approach in a mouse model of MECP2 duplication syndrome. Nat. Genet. *44*, 206–211.

Sansom, O.J., Berger, J., Bishop, S.M., Hendrich, B., Bird, A., and Clarke, A.R. (2003). Deficiency of Mbd2 suppresses intestinal tumorigenesis. Nat. Genet. *34*, 145–147.

Schanen, N.C., Dahle, E.J., Capozzoli, F., Holm, V.A., Zoghbi, H.Y., and 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.

Schanen, N.C., Kurczynski, T.W., Brunelle, D., Woodcock, M.M., Dure, L.S., and Percy, A.K. (1998). Neonatal encephalopathy in two boys in families with recurrent Rett syndrome. J. Child Neurol. *13*, 229–231.

Schick, J.A., Seisenberger, C., Beig, J., Bürger, A., Iyer, V., Maier, V., Perera, S., Rosen, B., Skarnes, W.C., Wurst, W., et al. (2016). CRISPR-Cas9 enables conditional mutagenesis of challenging loci. Sci. Rep. *6*, 32326.

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

Schneider, M.F., Wettengel, J., Hoffmann, P.C., and Stafforst, T. (2014). Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. Nucleic Acids Res. *42*, 1–9.

Schüle, B., Armstrong, D.D., Vogel, H., Oviedo, A., and Francke, U. (2008). Severe congenital encephalopathy caused by MECP2 null mutations in males: Central hypoxia and reduced neuronal dendritic structure. Clin. Genet. *74*, 116–126.

Seisenberger, S., Peat, J.R., Hore, T. a, Santos, F., Dean, W., and Reik, W. (2013). Reprogramming DNA methylation in the mammalian life cycle: building and breaking epigenetic barriers. Philos. Trans. R. Soc. London *368*, 20110330.

Shah, R.R., Cholewa-waclaw, J., Davies, F.C.J., Paton, K.M., Heard, E., Abbott, C.M., and Bird, A.P. (2016). Efficient and versatile CRISPR engineering of human neurons in culture to model neurological disorders. Wellcome Open Res. *1*.

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

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

Sharma, S., and Zhu, J. (2014). Immunologic Applications of Conditional Gene Modification Technology in the Mouse. Curr. Protoc. Immunol. *105*, 1–13.

Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. PLoS Genet. *9*.

Silberstein, J., Falk, R., Houwink-manville, I., Ellaway, C., Raffaele, L.S., and Engerstr, I.W. (2001). Occurrence of Rett Syndrome Boys. J. Child Neurol. *16*, 333–338.

Sirianni, N., Naidu, S., Pereira, J., Pillotto, R.F., and Hoffman, E.P. (1998). Rett Syndrome: Confirmation of X-Linked Dominant Inheritance, and Localization of the Gene to Xq28. Am. J. Hum. Genet. *63*, 1552–1558.

Skene, P.J., Illingworth, R.S., Webb, S., Kerr, A., James, K.D., Turner, D.J., Andrews, R., and Bird, A.P. (2010). Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol. Cell *37*, 457–468.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic readers for 5-(Hydroxy)methylcytosine and its oxidized derivatives. Cell *152*, 1146–1159.

Stefanelli, G., Gandaglia, A., Costa, M., Cheema, M.S., Di Marino, D., Barbiero, I., Kilstrup-Nielsen, C., Ausió, J., and Landsberger, N. (2016). Brain phosphorylation of MeCP2 at serine 164 is developmentally regulated and globally alters its chromatin association. Sci. Rep. *6*, 28295.

Subbanna, S., Nagre, N.N., Shivakumar, M., Umapathy, N.S., Psychoyos, D., and Basavarajappa, B.S. (2014). Ethanol induced acetylation of histone at G9a exon1 and G9a-mediated histone H3 dimethylation leads to neurodegeneration in neonatal mice. Neuroscience *258*, 422–432.

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.

Suter, B., Treadwell-Deering, D., Zoghbi, H.Y., Glaze, D.G., and Neul, J.L. (2014). MECP2 Mutations in People without Rett Syndrome. J. Autism Dev. Disord. *44*, 703–711.

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

Sztainberg, Y., Chen, H., Swann, J.W., Hao, S., Tang, B., Wu, Z., Tang, J., Wan, Y.-W., Liu, Z., Rigo, F., et al. (2015). Reversal of phenotypes in MECP2 duplication mice using genetic rescue or antisense oligonucleotides. Nature *528*, 123–126.

Szulwach, K.E., Li, X., Smrt, R.D., Li, Y., Luo, Y., Lin, L., Santistevan, N.J., Li, W., Zhao, X., and Jin, P. (2010). Cross talk between microRNA and epigenetic regulation in adult neurogenesis. J. Cell Biol. *189*, 127–141.

Szulwach, K.E., Li, X., Li, Y., Song, C.-X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat. Neurosci. *14*, 1607–1616.

Tai, D.J.C., Liu, Y.C., Hsu, W.L., Ma, Y.L., Cheng, S.J., Liu, S.Y., and Lee, E.H.Y. (2016). MeCP2 SUMOylation rescues Mecp2-mutant-induced behavioural deficits in a mouse model of Rett syndrome. Nat. Commun. *7*.

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

Tarquinio, D.C., Hou, W., Neul, J.L., Kaufmann, W.E., Glaze, D.G., Motil, K.J., Skinner, S.A., Lee, H.S., and Percy, A.K. (2015). The Changing Face of Survival in Rett Syndrome and MECP2-Related Disorders. Pediatr. Neurol. *53*, 402–411.

Tate, P.H., and Bird, A.P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Curr. Opin. Genet. Dev. *3*, 226–231.

Tate, P., Skarnes, W., and Bird, A. (1996). The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. Nat. Genet. *14*, 353–356.

The Rett Syndrome Diagnostic Criteria Work Group (1988). Diagnostic Criteria for Rett Syndrome. Ann. Neurol. *23*, 425–428

Tong, D., Zhao, L., He, K., Sun, H., Cai, D., and Ni, L. (2016). MECP2 promotes the growth of gastric cancer cells by suppressing miR-338-mediated antiproliferative effect. Oncotarget *7*, 34845–34859.

Tønnesen, J., Katona, G., Rózsa, B., and Nägerl, U.V. (2014). Spine neck plasticity regulates compartmentalization of synapses. Nat. Neurosci. *17*, 678–685.

Trappe, R., Laccone, F., Cobilanschi, J., Meins, M., Huppke, P., Hanefeld, F., and Engel, W. (2001). MECP2 Mutations in Sporadic Cases of Rett Syndrome Are Almost Exclusively of Paternal Origin. Am. J. Hum. Genet. *68*, 1093–1101.

Tsujimura, K., Irie, K., Nakashima, H., Egashira, Y., Fukao, Y., Fujiwara, M., Itoh, M., Uesaka, M., Imamura, T., Nakahata, Y., et al. (2015). miR-199a Links MeCP2 with mTOR Signaling and Its Dysregulation Leads to Rett Syndrome Phenotypes. Cell Rep. *12*, 1887–1901.

Tweedie-Cullen, R.Y., Reck, J.M., and Mansuy, I.M. (2009). Comprehensive mapping of post-translational modifications on synaptic, nuclear, and histone proteins in the adult mouse brain. J. Proteome Res. *8*, 4966–4982.

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.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-behn, F., Cross, M.K., Williams, B. a, Stamatoyannopoulos, J. a, Crawford, G.E., et al. (2013). Dynamic DNA methylation across diverse human cell lines and tissues Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. *23*, 555–567.

Vilain, A., Apiou, B., Malfoy, F., Vogt, N., and Dutrillaux, B. (1996). Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to human chromosome band Xq28 by in situ hybridization. Cytogenet. Genome Res. *74*, 292–294.

Villa, R., Morey, L., Raker, V. a, Buschbeck, M., Gutierrez, A., De Santis, F., Corsaro, M., Varas, F., Bossi, D., Minucci, S., et al. (2006). The methyl-CpG binding protein MBD1 is required for PML-RARalpha function. Proc. Natl. Acad. Sci. U. S. A. *103*, 1400–1405.

Villard, L. (2007). MECP2 mutations in males. J. Med. Genet. *44*, 417–423.

Voigt, P., LeRoy, G., Drury, W.J., Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193.

Voigt, P., Tee, W.W., and Reinberg, D. (2013). A double take on bivalent promoters. Genes Dev. *27*, 1318–1338.

Wada, R., Akiyama, Y., Hashimoto, Y., Fukamachi, H., and Yuasa, Y. (2010). miR-212 is downregulated and suppresses methyl-CpG-binding protein MeCP2 in human gastric cancer. Int. J. Cancer *127*, 1106–1114.

Williams, A.A., Mehler, V.J., Mueller, C., Vonhoff, F., White, R., and Duch, C. (2016). Apoptotic Activity of MeCP2 Is Enhanced by C-Terminal Truncating Mutations. PLoS One *11*, e0159632.

Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M.C., Lipkin, M., et al. (2002). Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. Proc. Natl. Acad. Sci. U. S. A. *99*, 14937–14942.

Wood, K.H., and Zhou, Z. (2016). Emerging Molecular and Biological Functions of MBD2, a Reader of DNA Methylation. Front. Genet. *7*, 1–14.

Wood, K.H., Johnson, B.S., Welsh, S.A., Lee, J.Y., Cui, Y., Krizman, E., Brodkin, E.S., Blendy, J.A., Robinson, M.B., Bartolomei, M.S., et al. (2016). Tagging methyl-CpG-binding domain proteins reveals different spatiotemporal expression and supports distinct functions. Epigenomics *8*, 455–473.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell *148*, 816–831.

Xue, J., Wijeratne, S.S.K., and Zempleni, J. (2013). Holocarboxylase synthetase synergizes with methyl CpG binding protein 2 and DNA methyltransferase 1 in the transcriptional repression of long-terminal repeats. Epigenetics *8*, 504–511.

Yang, J.J., Liu, L.P., Tao, H., Hu, W., Shi, P., Deng, Z.Y., and Li, J. (2016a). MeCP2 silencing of LncRNA H19 controls hepatic stellate cell proliferation by targeting IGF1R. Toxicology *359-360*, 39–46.

Yang, Y., Kucukkal, T.G., Li, J., Alexov, E., and Cao, W. (2016b). Binding Analysis of Methyl-CpG Binding Domain of MeCP2 and Rett Syndrome Mutations. ACS Chem. Biol. *11*, 2706–2715.

Yasui, D.H., Gonzales, M.L., Aflatooni, J.O., Crary, F.K., Hu, D.J., Gavino, B.J., Golub, M.S., Vincent, J.B., Schanen, N.C., Olson, C.O., et al. (2014). Mice with an isoform-ablating Mecp2exon 1 mutation recapitulate the neurologic deficits of Rett syndrome. Hum. Mol. Genet. *23*, 2447–2458.

Yazdani, M., Deogracias, R., Guy, J., Poot, R.A., Bird, A., and Barde, Y. (2012). Disease Modeling Using Embryonic Stem Cells: MeCP2 Regulates Nuclear Size and RNA Synthesis in Neurons. Stem Cells *30*, 2128–2139.

Yu, F., Thiesen, J., and Strätling, W.H. (2000). Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2. Nucleic Acids Res. *28*, 2201–2206.

Zeev, B. Ben, Yaron, Y., Schanen, N.C., Wolf, H., Brandt, N., Ginot, N., Shomrat, R., and Orr-Urtreger, A. (2002). Rett syndrome: clinical manifestations in males with MECP2 mutations. J. Child Neurol. *17*, 20–24.

Zhao, X., Ueba, T., Christie, B.R., Barkho, B., McConnell, M.J., Nakashima, K., Lein, E.S., Eadie, B.D., Willhoite, A.R., Muotri, A.R., et al. (2003). Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. Proc. Natl. Acad. Sci. U. S. A. *100*, 6777–6782.

Zhou, Z., Hong, E.J., Cohen, S., Zhao, W. ning, Ho, H. yi H., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., et al. (2006). Brain-Specific Phosphorylation of MeCP2 Regulates Activity-Dependent Bdnf Transcription, Dendritic Growth, and Spine Maturation. Neuron *52*, 255–269.

Ziller, M.J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C.B., Bernstein, B.E., Lengauer, T., et al. (2011). Genomic distribution and Inter-Sample variation of Non-CpG methylation across human cell types. PLoS Genet. *7*.

Zocchi, L., and Sassone-Corsi, P. (2012). SIRT1-mediated deacetylation of MeCP2 contributes to BDNF expression. Epigenetics *7*, 695–700.

Zoghbi, H. (1988). Genetic Aspects of Rett Syndrome. J. Child Neurol. *3*, S76–S78.

Unpublished work by members of Adrian Bird's lab: Beatrice Alexander-Howden, Justyna Cholewa-Waclaw, John Connelly, Jacky Guy, Sabine Lagger, Matthew Lyst, Martha Koerner, Jim Selfridge. Collaborations with: Adam Belsom and Juri Rappsilber (University of Edinburgh); Kamal Gadalla, Ralph Hector and Stuart Cobb lab (University of Glasgow).

Appendix

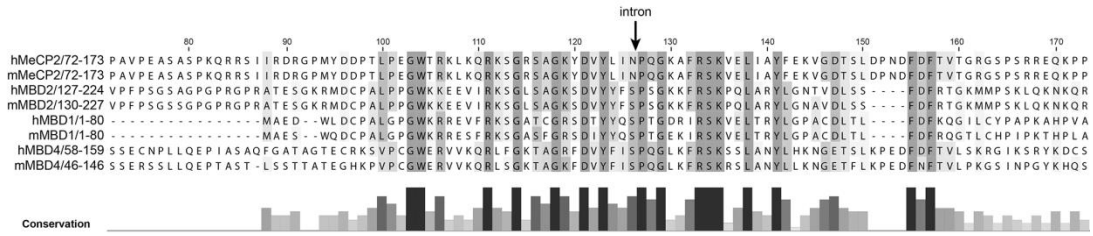


Figure S1: Protein sequence alignment of the MBD domains
Protein sequence alignment of the MBDs in all four members of the MBD protein family (human and mouse) using ClustalWS, shaded according to BLOSUM62 score. The position of the intron in all genes encoding these proteins is indicated by the arrow above. Conservation is shown below. Residue numbers correspond to that of human/ mouse MeCP2 (e2 isoforms). *The aligned region shown spans the length of the extended MBD of MeCP2 used in the truncated protein, ΔNIC-EGFP (see Chapter 3).*

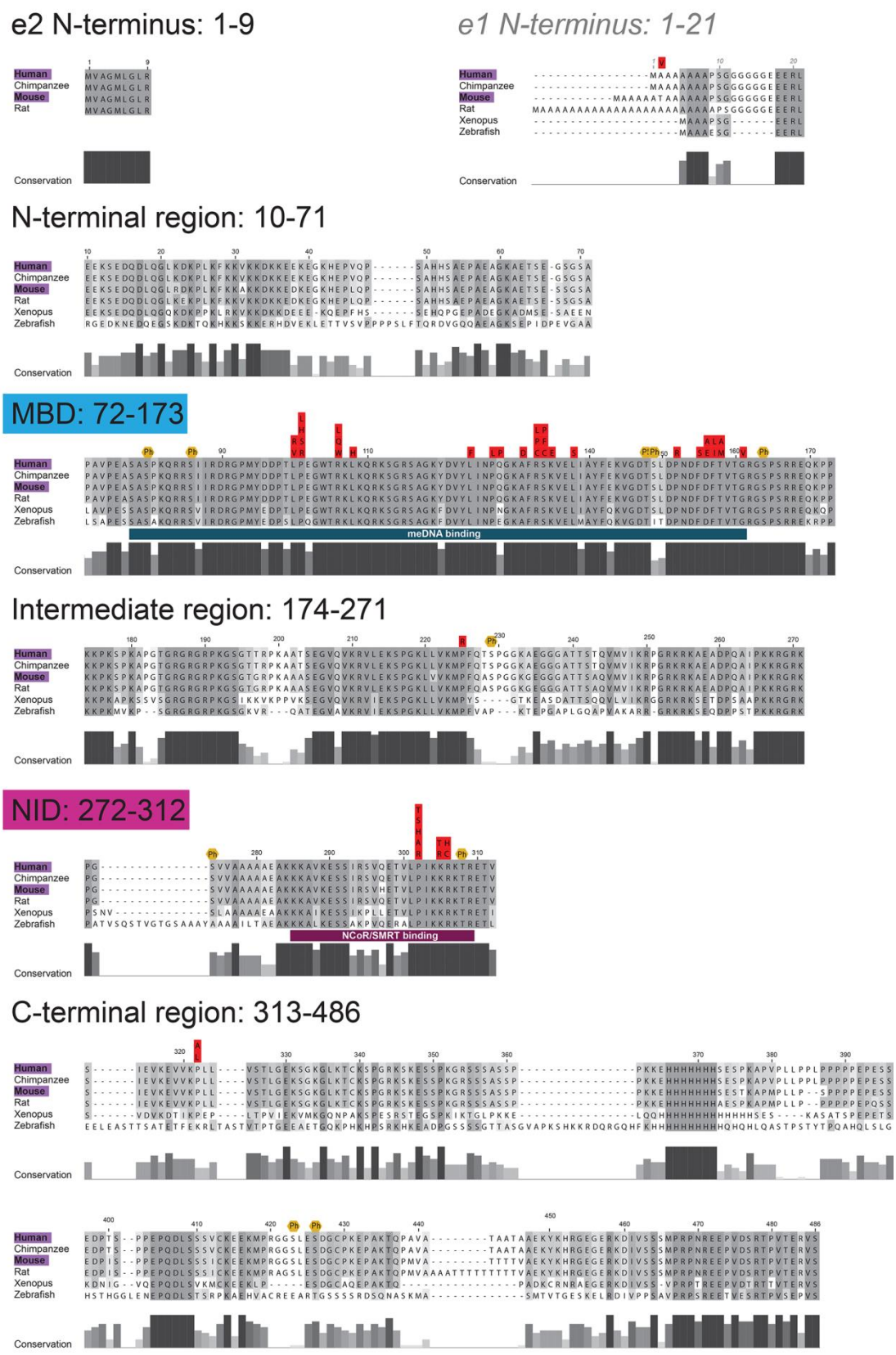


Figure S2: legend on next page

Figure S2: Alignment of MeCP2 protein sequences
Alignment of MeCP2 protein sequences using mammalian e1 isoforms and the sole isoform in Xenopus and zebrafish. The different N-terminus of the mammalian e2 isoform is also shown (top left). Sequences were aligned using ClustalWS and shaded according to BLOSUM62 score. Annotated with (from top): RTT-missense mutations (red); activity-dependent phosphorylation sites (orange); minimal Methyl-CpG binding domain (MBD) 78-162 (Nan et al., 1993); and NCoR/SMRT interaction domain, 285-309 (Lyst et al., 2013). Residue numbers correspond to human isoform e2. The alternative e1 extreme N-terminus is numbered according to the human protein (grey and italic). *The alignment is separated into the regions defined in this study (see Chapter 3).*

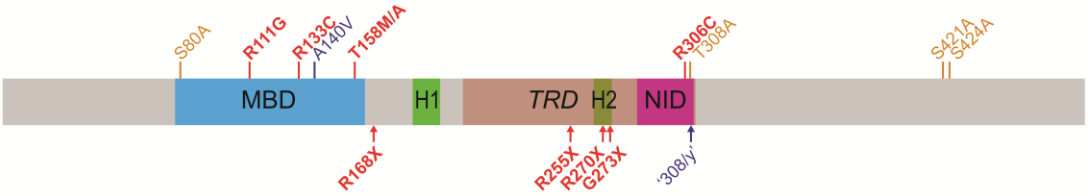


Figure S3: Mutations in MeCP2 expressed by knock-in/transgenic mice
Schematic diagram of MeCP2 protein (e2) annotated with missense (above) and truncating (below) mutations expressed by knock-in/transgenic mouse models. RTT-causing mutations are shown in red, milder mutations are shown in blue and phospho-abolishing mutations are shown in orange. These mouse models are discussed in section 1.4 and 1.5. MeCP2 is annotated with several domains: the Methyl-CpG binding domain (MBD, residues 78-162); AT-hook 1 (H1, residues 184-195); AT-hook 2 (H2, residues 265-272); the Transcriptional repression domain (TRD, residues 207-310); and the NCoR/SMRT Interaction Domain (NID, residues 285-309). These domains are discussed in section 1.5.

Table S1A: Characterisation of *Mecp2*-null mice (ND=not determined)

MeCP2 allele, strain background and reference(s)	**Protein level** (relative to wild-type)	**Summary of phenotype (age of death, fertility, weight, brain size)**	**Activity phenotype?** (OF=Open Field)	**Anxiety phenotype?** (EPM=Elevated Plus Maze; EZM=Elevated zero maze; OF=Open Field)	**Impaired motor function?**	**Other (memory; electrophysiology; social interaction)**
Null C57BL/6 or F1 C57BL/6; 129 (Guy et al., 2001)[1] Mixed: 129, C57BL/6 and BALB/c (Chen et al., 2001)[2] C57BL/6 (Goffin et al., 2012)[3] C57BL/6 (Asaka et al., 2006)[4]	None[1] Very small amount of protein with internal deletion (exon 3)[2]	Median survival: 9 weeks. Infertile (internal testes)[1]. Altered body weight: lighter (C57BL/6)[1] or obese (outbred)[1,2]. Reduced brain size[2].	Less active (in home cage) *at 9 weeks*[3].	Decreased anxiety[3] (EZM) *at 9 weeks*[3].	Yes (Accelerating Rotarod) *at 9 weeks*[3].	Abnormal electrophysiology: impaired (long term potentiation) LTP[4]. Abnormal EEG *at P90*[3].

Table S1B: Characterisation of mice with MBD mutations (ND=not determined)

MeCP2 allele, strain background and reference(s)	**Protein level** (relative to wild-type)	**Summary of phenotype (age of death, fertility, weight, brain size)**	**Activity phenotype?** (OF=Open Field)	**Anxiety phenotype?** (EPM=Elevated Plus Maze; EZM=Elevated zero maze; OF=Open Field)	**Impaired motor function?**	**Other (memory; electrophysiology; social interaction)**
T158A C57BL/6 (Goffin et al., 2012)	~0.5 fold	Median survival: 16 weeks. Slightly delayed growth between 4 and 8 weeks. Reduced brain size *(pre-symptomatic P30 and post-symptomatic P90)*	Less active (in home cage; OF) *at 9 weeks and 11 weeks, respectively*	Decreased anxiety (EZM) *at 9 weeks*	Yes (Accelerating Rotarod) *at 9 weeks*	Learning/memory defects (context- and cue- dependant fear conditioning) *at 10 weeks*. Abnormal EEG *at 90 days*. Seizes >*5 weeks*.
T158M-EGFP C57BL/6J (Brown, Selfridge et al., 2016) *unpublished	0.5 fold	Median survival: 13 weeks. Lighter body weight.	No phenotype (OF)* *at 8-10 weeks*	Decreased anxiety (EPM) *at 8-10 weeks*	Yes (Accelerating Rotarod, Hanging wire) *at 8-10 weeks*	ND
R133C-EGFP C57BL/6J (Brown, Selfridge et al., 2016) *unpublished	~ 1 fold	Median survival: 42 weeks. Lighter body weight.	No phenotype (OF)* *at 8-10 weeks*	Decreased anxiety (EPM) *at 8-10 weeks*	Yes (Hanging wire; no difference in Accelerating Rotarod) *at 8-10 weeks*	ND
R111G-EGFP F1 129SvEvTac; FVB (Heckman et al., 2014)	~1 fold (human transgene on *Mecp2 null* background)	Median survival: 11 weeks.	ND	ND	ND	ND

Table S1C: Characterisation of mice with R306C mutations (ND=not determined)

MeCP2 allele, strain background and reference(s)	**Protein level** (relative to wild-type)	**Summary of phenotype (age of death, fertility, weight, brain size)**	**Activity phenotype?** (OF=Open Field)	**Anxiety phenotype?** (EPM=Elevated Plus Maze; EZM=Elevated zero maze; OF=Open Field)	**Impaired motor function?**	**Other (memory; electrophysiology; social interaction)**
R306C Mixed 129; C57BL/6 (Lyst et al., 2013) *unpublished	1 fold	44% died before 18.5 weeks. No weight phenotype. Reduced brain weight.	Less active (OF) *at 15-18 weeks*	ND	Yes (Accelerating Rotarod) *at 15-18 weeks*	ND
R306C-EGFP C57BL/6J (Brown, Selfridge et al., 2016)	1.5 fold	Median survival: 30 weeks. Lighter body weight.	Less active (OF)* *at 8-10 weeks*	Decreased anxiety (EPM; OF*) *at 8-10 weeks*	Yes (Accelerating Rotarod, Hanging wire) *at 8-10 weeks*	ND
R306C-EGFP F1: 129SvEvTac; FVB (Heckman et al., 2014)	~1 fold (human transgene on *Mecp2 null* background)	Median survival: 18 weeks. Mildly over weight. Reduced brain weight	Less active (OF) *at 11 weeks*	Increased anxiety (light/dark box) *at 5 weeks* and (OF: % time in centre) *at 11 weeks*	Yes (parallel rod footslip) *at 5 and 11 weeks*, (but not Accelerating Rotarod) *at 11 weeks*	Learning/memory defects (context- and cue- dependant fear conditioning) *at 5 and 11 weeks.* Poor nesting *at 5 weeks*

Table S1D: Characterisation of mice with nonsense mutations (ND=not determined)

MeCP2 allele, strain background and reference(s)	**Protein level** (relative to wild-type)	**Summary of phenotype (age of death, fertility, weight, brain size)**	**Activity phenotype?** (OF=Open Field)	**Anxiety phenotype?** (EPM=Elevated Plus Maze; EZM=Elevated zero maze; OF=Open Field)	**Impaired motor function?**	**Other (memory; electrophysiology; social interaction)**
R168X Mixed 129SvJ; 129S6/SvEv (Lawson-Yuen et al., 2007)	Very low levels (not quantified)	Average survival: 86 days (12 weeks). Body weight diverges over time (failure to thrive or excessive gain).	ND	ND	ND	ND
R255X Mixed 129S6; C57BL/6 (Pitcher et al., 2015)	undetectable	Median survival: 61 days (9 weeks). Underweight at weaning but normal by 6 weeks. Reduced brain weight.	ND	Decreased anxiety (EPM) *at 6 weeks*	Yes (Accelerating Rotarod; Dowel test; Parallel Bars) *at 6 weeks*	Learning defect (contextual fear conditioning) *at 6 weeks* Abnormal electrophysiology (LTP) in *ex vivo* brain slices *at 6-7 weeks*
R270X-EGFP F1 FVB;129SvEv (Baker et al., 2013)	~1 fold (human transgene on *Mecp2 null* background)	Median survival: 85 days (12 weeks). Obese. Reduced brain weight >*7 weeks*)	ND	ND	ND	ND
R273X-EGFP F1 FVB;129SvEv (Baker et al., 2013)	~1 fold (human transgene on *Mecp2 null* background)	Median survival: 201 days (29 weeks). Heavier body weight. Reduced brain weight >*7 weeks*)	ND	ND	ND	ND

Table S1E: Characterisation of mice with mild mutations (ND=not determined)

MeCP2 allele, strain background and reference(s)	**Protein level** (relative to wild-type)	**Summary of phenotype (age of death, fertility, weight, brain size)**	**Activity phenotype?** (OF=Open Field)	**Anxiety phenotype?** (EPM=Elevated Plus Maze; EZM=Elevated zero maze; OF=Open Field)	**Impaired motor function?**	**Other (memory; electrophysiology; social interaction)**
308/y (1-308) F1 C57BL/6J; 129/SvEv (Shahbazian et al., 2002b)[1] 129/SvEv (Moretti et al., 2005)[2] F1 C57BL/6J; 129/SvEv (McGill et al., 2006)[3]	Not quantified[1]. 0.5 fold (Lagger, unpublished)	10% die before one year. Fertile. No body weight phenotype. Normal brain weight[1].	Less active (OF) *at 3&5 months*[1] Hypoactivity in dark phase, hyperactivity in light phase *at 8-11 weeks*[2].	Increased anxiety (OF *at 3&5 months*[1] *and 4 months*[3]; EPM *at 4 months*[3]; and light-dark box *at 4 months*[3]).	Yes (vertical pole test; wire suspension test; dowel test; no difference in Accelerating Rotarod unless grips were removed; no difference in grip strength) *at 3&5 months*[1].	No learning/memory defects (conditioned fear test; Morris water maze) *at 3&5 months*[1]. Abnormal electrophysiology (EEG) and seizures[1]. Abnormal social interactions (tube test[1,2]; interaction test with partition[2]; but not in resident-intruder paradigm[1,2]) *at 3&5 months*[1] at *30 weeks*[2]. Abnormal nest building/usage *at 8-11 weeks*[2].
A140V Mixed 129X1/S1; C57BL/6 (Jentarra et al., 2010)[1] C57BL/6 (Ma et al., 2014)[2]	ND; mRNA levels as WT[1].	Normal lifespan. No body weight phenotype.	ND	ND	ND	No seizures/ tremors[1]. Abnormal electrophysiology[2].

Table S1F: Characterisation of mice altered MeCP2 levels (ND=not determined)

MeCP2 allele, strain background and reference(s)	**Protein level** (relative to wild-type)	**Summary of phenotype (age of death, fertility, weight, brain size)**	**Activity phenotype?** (OF=Open Field)	**Anxiety phenotype?** (EPM=Elevated Plus Maze; EZM=Elevated zero maze; OF=Open Field)	**Impaired motor function?**	**Other (memory; electrophysiology; social interaction)**
Hypomorph (floxed) F1 129S6/SvEvTac; C57BL/6J (Samaco et al., 2008)[1] F1 129S1/SvImJ; B6/CBA (Kerr et al., 2008)[2]	0.5 fold on average[1] 0.3-0.9 fold, depending on brain region[2]	Normal lifespan[1]. Fertile[2]. Heavier body weight (fat accumulation)[2], not for F1 129S6/SvEvTac; FVB/N[1].	No phenotype (OF) *at 12 weeks*[1,2].	Decreased anxiety (OF) on 129S6/SvEvTac; FVB/N background *at 12 weeks*[1]. No phenotype (EPM; OF) *at 12 weeks*[2].	Yes (Accelerating Rotarod; wire suspension test, not F1 129S6/SvEvTac; FVB/N) *at 11-12 weeks*[1]. (dowel test) *at 12 weeks*[2].	Learning defect (fear conditioning task) *at 22 weeks*[1]. Reflex testing: no difference *at 12 weeks*[2]. Olfaction/motivation to eat; no difference *at 12 weeks*[2]. Difference in social novelty but not social interaction *at 12 weeks*[2]. Social interaction (partition test) *at 20 weeks*[1]. Deficient nest building *at 19 weeks*[1].
Overexpression FVB (Collins et al., 2004)[1] F1: FVB/N; C57BL/6 (Sztainberg et al., 2015)[2]	~2 fold (human transgene on WT background)	~30% die before 1 year (from 18 weeks). Reduced fertility[1]. No weight phenotype[1].	No phenotype (OF) *at 10 or 20 weeks*[1]. Hypoactivity (OF) *at 17-18 weeks*[2].	No phenotype (light/dark box; OF) *at 10 or 20 weeks*[1]. Increased anxiety (OF; EPM) *at 17-18 weeks*[2].	Improved motor learning (Accelerating Rotarod) *at 10 (but not 6) weeks of age*[1].	Enhanced Learning/memory (context- but not cue-dependant fear conditioning) *at 20 weeks*[1]. Enhanced synaptic plasticity (LTP); abnormal EEG, seizures *at ≥ 20 weeks*[1]. Abnormal social behaviour *at 18-19 weeks*[2].

Table S1G: Characterisation of mice phospho-mutations (ND=not determined)

MeCP2 allele, strain background and reference(s)	**Protein level** (relative to wild-type)	**Summary of phenotype (age of death, fertility, weight, brain size)**	**Activity phenotype?** (OF=Open Field)	**Anxiety phenotype?** (EPM=Elevated Plus Maze; EZM=Elevated zero maze; OF=Open Field)	**Impaired motor function?**	**Other (memory; electrophysiology; social interaction)**
S80A Mixed C57BL/6J; 129/Sv (Tao et al., 2009)	Not quantified	Normal lifespan. Slightly overweight. Normal brain weight.	ND	ND	Yes (dark cycle running wheel) *age not stated*	ND
T308A C57BL/6 (Ebert et al., 2013)	1 fold	Normal lifespan. Normal body weight. Reduced brain weight.	ND	ND	Yes (accelerating Rotarod) *at 13-15 weeks*	Lower seizure threshold (upon pentylenetetrazol injection) *at 14-16 weeks*
S421A C57BL/6 (Cohen et al., 2011)	1 fold	ND	No phenotype (running wheel) *at 20-24 weeks*	No phenotype (EZM; Light-dark box) *age not stated*	Slight defect (Rotarod at a constant speed of 24 rpm, but not Accelerating Rotarod) *age not stated*	Reduced preference for novelty: both mice and objects (3 chamber test) *at 8-12 weeks* No difference in spatial learning/memory (Morris water maze) *age not stated* Abnormal electrophysiology (increased mIPSC amplitude)
S421A and S424A Mixed C57BL/6J; 129/Sv (Tao et al., 2009)	Not quantified	Normal lifespan. No weight phenotype. Increased brain weight.	ND	ND	Improved motor function (dark cycle running wheel) *age not stated*[1]	ND
S421A and S424A-FLAG C57BL/6 (Li et al., 2011)	1 fold	Normal lifespan. Fertile.	No phenotype (OF) *at 10-12 weeks*	No phenotype (EPM; OF) *at 10-12 weeks*	No phenotype (Accelerating Rotarod) *at 10-12 weeks*	Improved spatial memory (Morris water maze) *at 10-12 weeks* Enhanced Learning/memory (context-but not cue- dependant fear conditioning) *at 10-12 weeks* Enhanced synaptic plasticity (LTP)

Table S1H: Characterisation of mice 'unexplained' mutations (ND=not determined)

MeCP2 allele, strain background and reference(s)	**Protein level** (relative to wild-type)	**Summary of phenotype (age of death, fertility, weight, brain size)**	**Activity phenotype?** (OF=Open Field)	**Anxiety phenotype?** (EPM=Elevated Plus Maze; EZM=Elevated zero maze; OF=Open Field)	**Impaired motor function?**	**Other (memory; electrophysiology; social interaction)**
P225R C57BL/6J (Guy and Selfridge, unpublished)	0.22 fold	Median survival: 51 weeks (death from 30 weeks). Lighter body weight.	Less active (OF) *at 12 weeks*	Decreased anxiety (EPM;OF) *at 12 weeks*	Yes (Hanging Wire but not Accelerating Rotarod) *at 12 weeks*	ND
P322L C57BL/6J (Guy and Selfridge, unpublished)	0.03 fold	Survival as nulls. Lighter body weight.	No phenotype (OF) *at 6 weeks*	Decreased anxiety (EPM;OF) *at 6 weeks*	Yes (Accelerating Rotarod and Hanging Wire) *at 6 weeks*	ND

Site (activity dependant?)	PTM	Enzyme(s) responsible	RTT	ExAC	Conserved?
Lys12[6]	ac			N (♀)	Zebrafish E
Ser13[6,9]	phos				Zebrafish D
Lys22[11]	ac	removed by SIRT1			Mouse R
Ser68[5]	phos				Xenopus A, Zebrafish V
Ser70[5]	phos				Xenopus E, Zebrafish A
Ser80[1,2,3,5]	phos	HIPK2 kinase			yes
Lys82[6]	ac			R (♀)	yes
Ser86[8]	phos				yes
Lys119[6]	ac				yes
Tyr120[12]	phos				Xenopus F, Zebrafish F
Lys130[6]	ac				yes
Lys135[6]	ac	removed by SIRT1	E		yes
Thr148(/Ser149)[2]	phos				yes
(Thr148/)Ser149[2,6]	phos				Zebrafish I
Thr160[4]	phos			S (♀)	yes
Ser164[2,13]	phos				yes
Ser166[413]	phos				yes
Lys171[11]	ac	removed by SIRT1		Q (♀)	Zebrafish R
Ser178[13]	phos			C (♀)	Xenopus A, Zebrafish M
Lys200[11]	ac	removed by SIRT1			Xenopus P, Zebrafish -
Lys223[10]	SUMO				yes
Ser229[1,2,6]	phos			L	Xenopus -, Zebrafish P
Lys233[6]	ac				Xenopus T
Lys249[6]	ac				Zebrafish R
Lys256[6]	ac	removed by SIRT1			yes
Lys271[6]	ac	removed by SIRT1			yes
Ser274[4,6,8]	phos				Zebrafish A
Lys289[11]	ac	removed by SIRT1			yes
Lys305(/307)[6]	ub		R, T		yes
(Lys305/)307[6]	ub				yes
Thr308[8]	phos				yes
Lys321[6]	ac				Zebrafish E
Ser399[401][2,6]	phos			H/A	Xenopus G, Zebrafish G
Lys412[Lys414][14]	SUMO				yes
Ser421[423][1,2,8]	phos	CamKII kinase		P	Xenopus -, Zebrafish E
Ser424[Ser426][2,5]	phos				Zebrafish T
Lys447[Lys449][7]	ac	added by p300; removed by SIRT1			yes

Table S2: Post-translational modifications in MeCP2. Includes all the published post-translational modifications (PTMs) and the responsible enzymes identified in neurons or brain tissue**.** Activity dependent modifications are highlighted in grey. Ser80 is dephosphorylated upon neuronal activity (highlighted in red). Columns 4 and 5 list RTT-causing mutations and neutral variants at these sites, respectively.

Column 6 lists different amino acids at these sites in MeCP2 homologs in chimpanzee, mouse, rat, xenopus and zebrafish. Amino acids standardly used as PTM- abolishing mutations are shown in red, and as mimetic mutations in green. Phosphorylation sites preserved by mutation to another phosphorylatable amino acid (S/T/Y) are shown in blue.

References: (1) Zhou et al., 2006; (2) Tao et al., 2009; (3) Bracaglia et al., 2009; (4) Tweedie-Cullen et al., 2009; (5) Huttlin et al., 2010; (6) Gonzales et al., 2012; (7) Zocchi and Sassone-Corsi, 2012; (8) Ebert et al., 2013; (9) Yasui et al., 2014; (10) Cheng et al., 2014b; (11) Pandely et al., 2015; (12) Bergo et al., 2015; (13) Stefanelli et al., 2016; (14) Tai et al., 2016.

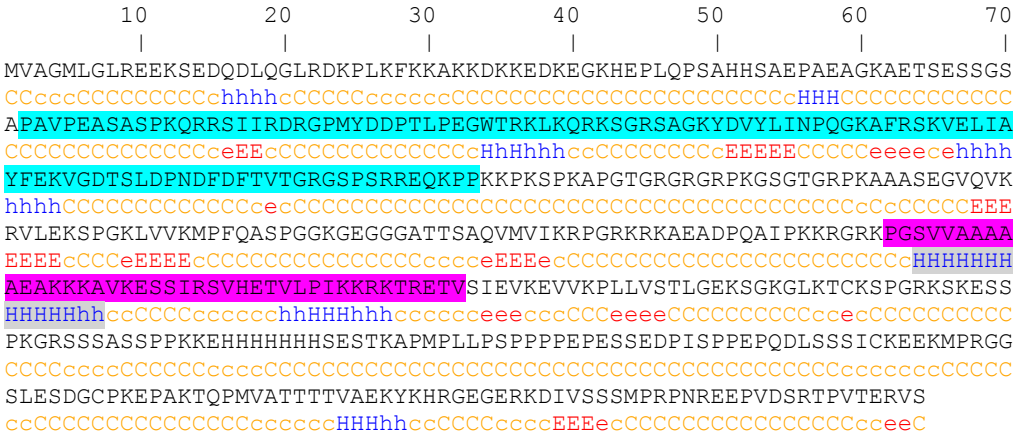
Mutation	Localisation to PHC?	Binding in EMSA/SPR?
L100V	Intermediate[3,5,6]	Yes (mCpG and hmC)[7]; Partial[11]
L100R		
P101R	Foci[3,6]	
P101S	Intermediate[3,6]	
P101H	Foci (more strongly)[3,6]	
P101L	Intermediate[3,6]	
R106W	Diffuse[3,5,6]	No[1,2,11]
R106Q	Diffuse[3]	No[11]
R106L		
L108H		
R111G	Diffuse[3,5]; in mice[9]	No[2,11]
L124F	Diffuse[3]	
P127L	Foci[3,6]	
Q128P		
A131D		
R133C	Intermediate[3,5,6]; in mice[10]	No[1,2,10,11]; to mCpG but not hmC[7] or mCpA[10]
R133P		
R133L	Diffuse[3,6]	
S134C	Intermediate[3,6]	Yes (mCpG and hmC)[7]; Partial[11]
S134F		Very weak (mCpG and hmC)[7]
S134P		
K135E	Diffuse[3]	
L138S		
P152R	Intermediate[3,6]	Partial1[1]
F155S	Diffuse[3,6]	Partial[2,11]; No[1]
D156E	Diffuse[3]	Partial[11]
D156A		
F157L		
F157I		
T158M	Diffuse[3,5,6]; in mice[10]	Yes[2]; No[4]; Partial[1,11]
T158A	Diffuse[3,6]	No[4]; Partial[11]
G161V		

Table S3: Missense mutations in the MBD of MeCP2 that cause Classical Rett syndrome often disrupt binding to methylated DNA

The DNA-binding abilities of MBD mutants have been investigated using several methods: localisation to pericentromeric heterochromatin (PCH) in mouse cells; and binding methylated-DNA probes in *in vitro* binding assays (Electrophoretic mobility shift assay, EMSA; Surface Plasmon Resonance, SPR).

References: 1. Ballestar et al., 2000; 2. Free et al., 2001; 3. Kudo et al., 2003; 4. Ho et al., 2008; 5. Schmiedeberg et al., 2009; 6. Agarwal et al., 2011; 7. Mellén et al., 2012; 8. Goffin et al., 2012; 9. Heckman et al., 2014; 10. Brown, Selfridge et al., 2016; 11. Yang et al., 2016b.

PHD Secondary Structure prediction of mouse MeCP2 (isoform e2): Oct 2013



PHD Secondary Structure prediction of mouse MeCP2 (isoform e2): Oct 2016

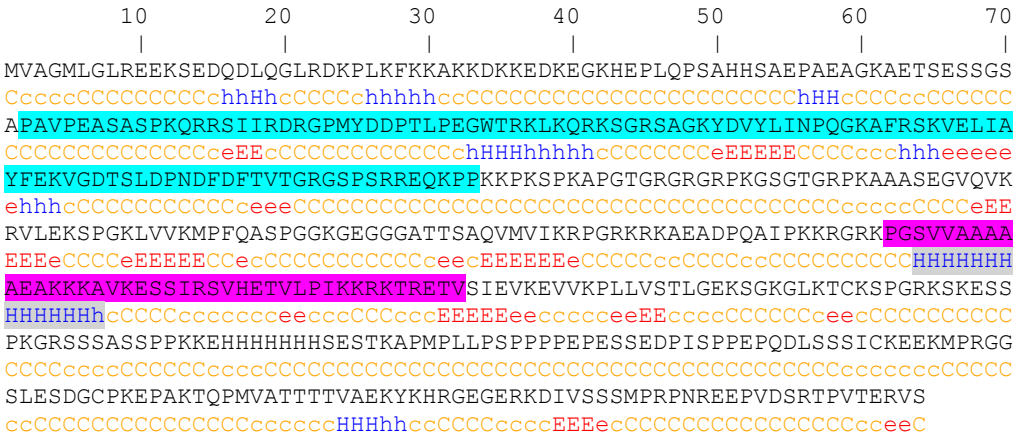


Figure S4: Secondary structure prediction of mouse MeCP2 e2 isoform
PHD software (https://npsa-prabi.ibcp.fr/NPSA/npsa_phd.html) was used to predict the secondary structure around the NID (residues 272-312, highlighted in pink) for the design of the truncated proteins. Predicted alpha helices are denotes by the letter 'H' and beta strands by the letter 'E', with capitalisation reflecting higher confidence. Although, the predicted secondary structure has changed since I designed these proteins in October 2013, the predicted alpha helix at the N-terminal end of the NID (residues 274-287, highlighted in grey) is still present. The NID used in this study was extended at the N-terminus to retain this predicted helix. The MBD is highlighted in blue.

Alignment of the truncated MeCP2 proteins

Key: extreme N terminus MBD linkers SV40 NLS NID EGFP

```
WT-EGFP/1-748    MAAAAATAAAAAAPSGGGGGGEEERLEEKSEDQDLQGLRDKPLKFKKAKKDKKEDKEGKH
ΔN-EGFP/1-689    MAAAAATAAAAAAPSGGGGGGEEERLEEK------------------------------
ΔNC-EGFP/1-516   MAAAAATAAAAAAPSGGGGGGEEERLEEK------------------------------
ΔNIC-EGFP/1-432  MAAAAATAAAAAAPSGGGGGGEEERLEEK------------------------------

WT-EGFP/1-748    EPLQPSAHHSAEPAEAGKAETSESSGSAPAVPEASASPKQRRSIIRDRGPMYDDPTLPEG
ΔN-EGFP/1-689    ----------------------------PAVPEASASPKQRRSIIRDRGPMYDDPTLPEG
ΔNC-EGFP/1-516   ----------------------------PAVPEASASPKQRRSIIRDRGPMYDDPTLPEG
ΔNIC-EGFP/1-432  ----------------------------PAVPEASASPKQRRSIIRDRGPMYDDPTLPEG

WT-EGFP/1-748    WTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRG
ΔN-EGFP/1-689    WTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRG
ΔNC-EGFP/1-516   WTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRG
ΔNIC-EGFP/1-432  WTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRG

WT-EGFP/1-748    SPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEKSPGKLVVK
ΔN-EGFP/1-689    SPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEKSPGKLVVK
ΔNC-EGFP/1-516   SPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEKSPGKLVVK
ΔNIC-EGFP/1-432  SPSRREQKPP----------------------------------------------------

WT-EGFP/1-748    MPFQASPGGKGEGGGATTSAQVMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEA
ΔN-EGFP/1-689    MPFQASPGGKGEGGGATTSAQVMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEA
ΔNC-EGFP/1-516   MPFQASPGGKGEGGGATTSAQVMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEA
ΔNIC-EGFP/1-432  ----------------------------------GSSGSSGPKKKRKVPGSVVAAAAAEA

WT-EGFP/1-748    KKKAVKESSIRSVHETVLPIKKRKTRETVSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPG
ΔN-EGFP/1-689    KKKAVKESSIRSVHETVLPIKKRKTRETVSIEVKEVVKPLLVSTLGEKSGKGLKTCKSPG
ΔNC-EGFP/1-516   KKKAVKESSIRSVHETVLPIKKRKTRETV------------------------------
ΔNIC-EGFP/1-432  KKKAVKESSIRSVHETVLPIKKRKTRETVG-----------------------------

WT-EGFP/1-748    RKSKESSPKGRSSSASSPPKKEHHHHHHHSESTKAPMPLLPSPPPPEPESSEDPISPPEP
ΔN-EGFP/1-689    RKSKESSPKGRSSSASSPPKKEHHHHHHHSESTKAPMPLLPSPPPPEPESSEDPISPPEP
ΔNC-EGFP/1-516   ------------------------------------------------------------
ΔNIC-EGFP/1-432  ------------------------------------------------------------

WT-EGFP/1-748    QDLSSSICKEEKMPRGGSLESDGCPKEPAKTQPMVATTTTVAEKYKHRGEGERKDIVSSS
ΔN-EGFP/1-689    QDLSSSICKEEKMPRGGSLESDGCPKEPAKTQPMVATTTTVAEKYKHRGEGERKDIVSSS
ΔNC-EGFP/1-516   ------------------------------------------------------------
ΔNIC-EGFP/1-432  ------------------------------------------------------------

WT-EGFP/1-748    MPRPNREEPVDSRTPVTERVSCKDPPVATMVSKGEELFTGVVPILVELDGDVNGHKFSVS
ΔN-EGFP/1-689    MPRPNREEPVDSRTPVTERVSCKDPPVATMVSKGEELFTGVVPILVELDGDVNGHKFSVS
ΔNC-EGFP/1-516   ----------------------GSSGSSGMVSKGEELFTGVVPILVELDGDVNGHKFSVS
ΔNIC-EGFP/1-432  ----------------------GSSGSSGMVSKGEELFTGVVPILVELDGDVNGHKFSVS

WT-EGFP/1-748    GEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPE
ΔN-EGFP/1-689    GEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPE
ΔNC-EGFP/1-516   GEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPE
ΔNIC-EGFP/1-432  GEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPE

WT-EGFP/1-748    GYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
ΔN-EGFP/1-689    GYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
ΔNC-EGFP/1-516   GYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
ΔNIC-EGFP/1-432  GYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV

WT-EGFP/1-748    YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKD
ΔN-EGFP/1-689    YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKD
ΔNC-EGFP/1-516   YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKD
ΔNIC-EGFP/1-432  YIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKD

WT-EGFP/1-748    PNEKRDHMVLLEFVTAAGITLGMDELYK
ΔN-EGFP/1-689    PNEKRDHMVLLEFVTAAGITLGMDELYK
ΔNC-EGFP/1-516   PNEKRDHMVLLEFVTAAGITLGMDELYK
ΔNIC-EGFP/1-432  PNEKRDHMVLLEFVTAAGITLGMDELYK
```

Figure S5: ClustalWS alignment of the truncated proteins (e1 isoforms)

cDNA sequences of the truncated MeCP2 alleles

Key: extreme N terminus MBD linkers SV40 NLS NID EGFP

WT-EGFP

ATGGCCGCCGCTGCCGCCACCGCCGCCGCCGCCGCCGCCGCCGCCGAGCGGAGGAGGAGGAGGAGGCGAGGAGGAGAGACTGGAGGAAA
AGTCAGAAGACCAGGATCTCCAGGGCCTCAGAGACAAGCCACTGAAGTTTAAGAAGGCGAAGAAAGACAAGAAGGAGGACAAAGA
AGGCAAGCATGAGCCACTACAACCTTCAGCCCACCATTCTGCAGAGCCAGCAGAGGCAGGCAAAGCAGAAACATCAGAAAGCTCA
GGCTCTGCCCCAGCAGTGCCAGAAGCCTCGGCTTCCCCCAAACAGCGGCGCTCCATTATCCGTGACCGGGGACCTATGTATGATG
ACCCCACCTTGCCTGAAGGTTGGACACGAAAGCTTAAACAAAGGAAGTCTGGCCGATCTGCTGGAAAGTATGATGTATATTTGAT
CAATCCCCAGGGAAAAGCTTTTCGCTCTAAAGTAGAATTGATTGCATACTTTGAAAAGGTGGGAGACACCTCCTTGGACCCTAAT
GATTTTGACTTCACGGTAACTGGGAGAGGGAGCCCCTCCAGGAGAGAGCAGAAACCACCTAAGAAGCCCAAATCTCCCAAAGCTC
CAGGAACTGGCAGGGGTCGGGGACGCCCCAAAGGGAGCGGCACTGGGAGACCAAAGGCAGCAGCATCAGAAGGTGTTCAGGTGAA
AAGGGTCCTGGAGAAGAGCCCTGGGAAACTTGTTGTCAAGATGCCTTTCCAAGCATCGCCTGGGGGTAAGGGTGAGGGAGGTGGG
GCTACCACATCTGCCCAGGTCATGGTGATCAAACGCCCTGGCAGAAAGCGAAAAGCTGAAGCTGACCCCCAGGCCATTCCTAAGA
AACGGGGTAGAAAGCCTGGGAGTGTGGTGGCAGCTGCTGCAGCTGAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTCCATACGGTC
TGTGCATGAGACTGTGCTCCCCATCAAGAAGCGCAAGACCCGGGAGACGGTCAGCATCGAGGTCAAGGAAGTGGTGAAGCCCCTG
CTGGTGTCCACCCTTGGTGAGAAAAGCGGGAAGGGACTGAAGACCTGCAAGAGCCCTGGGCGTAAAAGCAAGGAGAGCAGCCCCA
AGGGGCGCAGCAGCAGTGCCTCCTCCCCACCTAAGAAGGAGCACCATCATCACCACCATCACTCAGAGTCCACAAAGGCCCCCAT
GCCACTGCTCCCATCCCCACCCCCACCTGAGCCTGAGAGCTCTGAGGACCCCATCAGCCCCCCTGAGCCTCAGGACTTGAGCAGC
AGCATCTGCAAAGAAGAGAAGATGCCCCGAGGAGGCTCACTGGAAAGCGATGGCTGCCCCAAGGAGCCAGCTAAGACTCAGCCTA
TGGTCGCCACCACTACCACAGTTGCAGAAAAGTACAAACACCGAGGGGAGGGAGAGCGCAAAGACATTGTTTCATCTTCCATGCC
AAGGCCAAACAGAGAGGAGCCTGTGGACAGCCGGACGCCCGTGACCGAGAGAGTTAGCTGTAAGGATCCACCGGTCGCCACCATG
GTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCG
TGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTG
GCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAG
TCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGT
TCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGA
GTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAAC
ATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC
ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGC
CGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA

ΔN-EGFP

ATGGCCGCCGCTGCCGCCACCGCCGCCGCCGCCGCCGCCGCCGAGCGGAGGAGGAGGAGGAGGCGAGGAGGAGAGACTGGAGGAAA
AGCCAGCAGTGCCAGAAGCCTCGGCTTCCCCCAAACAGCGGCGCTCCATTATCCGTGACCGGGGACCTATGTATGATGACCCCAC
CTTGCCTGAAGGTTGGACACGAAAGCTTAAACAAAGGAAGTCTGGCCGATCTGCTGGAAAGTATGATGTATATTTGATCAATCCC
CAGGGAAAAGCTTTTCGCTCTAAAGTAGAATTGATTGCATACTTTGAAAAGGTGGGAGACACCTCCTTGGACCCTAATGATTTTG
ACTTCACGGTAACTGGGAGAGGGAGCCCCTCCAGGAGAGAGCAGAAACCACCTAAGAAGCCCAAATCTCCCAAAGCTCCAGGAAC
TGGCAGGGGTCGGGGACGCCCCAAAGGGAGCGGCACTGGGAGACCAAAGGCAGCAGCATCAGAAGGTGTTCAGGTGAAAAGGGTC
CTGGAGAAGAGCCCTGGGAAACTTGTTGTCAAGATGCCTTTCCAAGCATCGCCTGGGGGTAAGGGTGAGGGAGGTGGGGCTACCA
CATCTGCCCAGGTCATGGTGATCAAACGCCCTGGCAGAAAGCGAAAAGCTGAAGCTGACCCCCAGGCCATTCCTAAGAAACGGGG
TAGAAAGCCTGGGAGTGTGGTGGCAGCTGCTGCAGCTGAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTCCATACGGTCTGTGCAT
GAGACTGTGCTCCCCATCAAGAAGCGCAAGACCCGGGAGACGGTCAGCATCGAGGTCAAGGAAGTGGTGAAGCCCCTGCTGGTGT
CCACCCTTGGTGAGAAAAGCGGGAAGGGACTGAAGACCTGCAAGAGCCCTGGGCGTAAAAGCAAGGAGAGCAGCCCCAAGGGGCG
CAGCAGCAGTGCCTCCTCCCCACCTAAGAAGGAGCACCATCATCACCACCATCACTCAGAGTCCACAAAGGCCCCCATGCCACTG
CTCCCATCCCCACCCCCACCTGAGCCTGAGAGCTCTGAGGACCCCATCAGCCCCCCTGAGCCTCAGGACTTGAGCAGCAGCATCT
GCAAAGAAGAGAAGATGCCCCGAGGAGGCTCACTGGAAAGCGATGGCTGCCCCAAGGAGCCAGCTAAGACTCAGCCTATGGTCGC
CACCACTACCACAGTTGCAGAAAAGTACAAACACCGAGGGGAGGGAGAGCGCAAAGACATTGTTTCATCTTCCATGCCAAGGCCA
AACAGAGAGGAGCCTGTGGACAGCCGGACGCCCGTGACCGAGAGAGTTAGCTGTAAGGATCCACCGGTCGCCACCATGGTGAGCA
AGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGG
CGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC
CTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCA
TGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGG
CGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAAC
TACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGG
ACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCT
GAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATC
ACTCTCGGCATGGACGAGCTGTACAAGTAA

Figure S6: cDNA sequences of truncated MeCP2 alleles (e1 isoforms) –
continued on next page

ΔNC-EGFP

ATGGCCGCCGCTGCCGCCACCGCCGCCGCCGCCGCCGCCGCCGAGCGGAGGAGGAGGAGGAGGCGAGGAGGAGAGACTGGAGGAAA
AGCCAGCAGTGCCAGAAGCCTCGGCTTCCCCCAAACAGCGGCGCTCCATTATCCGTGACCGGGGACCTATGTATGATGACCCCAC
CTTGCCTGAAGGTTGGACACGAAAGCTTAAACAAAGGAAGTCTGGCCGATCTGCTGGAAAGTATGATGTATATTTGATCAATCCC
CAGGGAAAAGCTTTTCGCTCTAAAGTAGAATTGATTGCATACTTTGAAAAGGTGGGAGACACCTCCTTGGACCCTAATGATTTTG
ACTTCACGGTAACTGGGAGAGGGAGCCCCTCCAGGAGAGAGCAGAAACCACCTAAGAAGCCCAAATCTCCCAAAGCTCCAGGAAC
TGGCAGGGGTCGGGGACGCCCCAAAGGGAGCGGCACTGGGAGACCAAAGGCAGCAGCATCAGAAGGTGTTCAGGTGAAAAGGGTC
CTGGAGAAGAGCCCTGGGAAACTTGTTGTCAAGATGCCTTTCCAAGCATCGCCTGGGGGTAAGGGTGAGGGAGGTGGGGCTACCA
CATCTGCCCAGGTCATGGTGATCAAACGCCCTGGCAGAAAGCGAAAAGCTGAAGCTGACCCCCAGGCCATTCCTAAGAAACGGGG
TAGAAAGCCTGGGAGTGTGGTGGCAGCTGCTGCAGCTGAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTCCATACGGTCTGTGCAT
GAGACTGTGCTCCCCATCAAGAAGCGCAAGACCCGGGAGACGGTCGGGAGCTCCGGCAGTTCTGGAATGGTGAGCAAGGGCGAGG
AGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGA
GGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACC
ACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAG
GCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCT
GGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGC
CACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCG
TGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCA
GTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGC
ATGGACGAGCTGTACAAGTAA

ΔNIC-EGFP

ATGGCCGCCGCTGCCGCCACCGCCGCCGCCGCCGCCGCCGCCGAGCGGAGGAGGAGGAGGAGGCGAGGAGGAGAGACTGGAGGAAA
AGCCAGCAGTGCCAGAAGCCTCGGCTTCCCCCAAACAGCGGCGCTCCATTATCCGTGACCGGGGACCTATGTATGATGACCCCAC
CTTGCCTGAAGGTTGGACACGAAAGCTTAAACAAAGGAAGTCTGGCCGATCTGCTGGAAAGTATGATGTATATTTGATCAATCCC
CAGGGAAAAGCTTTTCGCTCTAAAGTAGAATTGATTGCATACTTTGAAAAGGTGGGAGACACCTCCTTGGACCCTAATGATTTTG
ACTTCACGGTAACTGGGAGAGGGAGCCCCTCCAGGAGAGAGCAGAAACCACCTGGATCCAGTGGCAGCTCTGGGCCCAAGAAAAA
GCGGAAGGTGCCTGGGAGTGTGGTGGCAGCTGCTGCAGCTGAGGCCAAAAAGAAAGCCGTGAAGGAGTCTTCCATACGGTCTGTG
CATGAGACTGTGCTCCCCATCAAGAAGCGCAAGACCCGGGAGACGGTCGGGAGCTCCGGCAGTTCTGGAATGGTGAGCAAGGGCG
AGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGG
CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG
ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCG
AAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACAC
CCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAAC
AGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCA
GCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCAC
CCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTC
GGCATGGACGAGCTGTACAAGTAA

Figure S6: cDNA sequences of truncated MeCP2 alleles (e1 isoforms) – continued from previous page

Protein (e1)	% of FL MeCP2 protein sequence	Predicted size (kDa)
WT-EGFP	100%	81
ΔN-EGFP	88%	75
ΔNC-EGFP	54%	56
ΔNIC-EGFP	34%	47

Table S4: Basic information about MeCP2 truncated proteins
Note: MeCP2 runs higher than expected on SDS gels

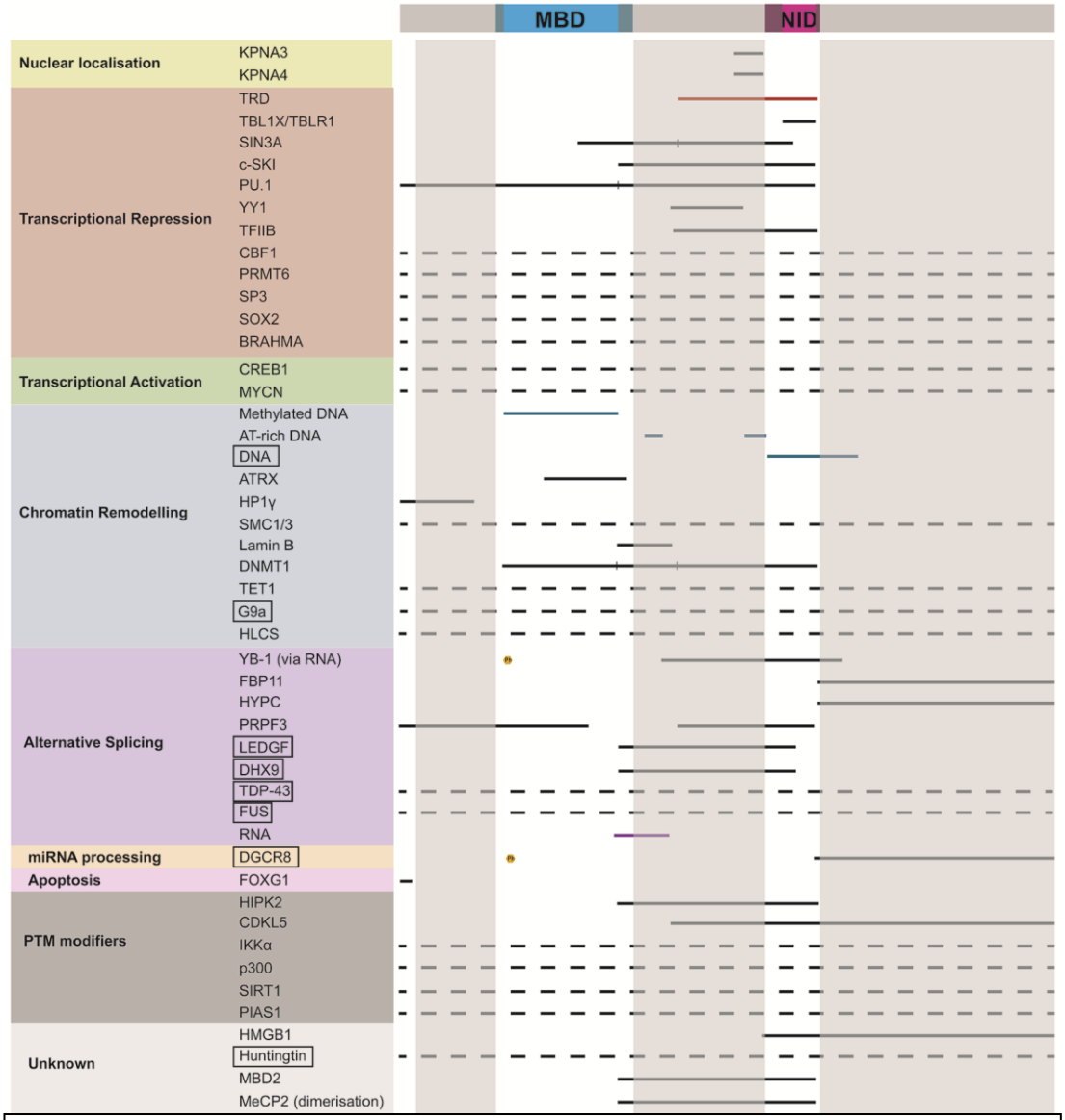


Figure S7: Binding sites of MeCP2 interaction partners retained in the truncated proteins
As Fig. 1.5, but with MBD and NID domains extended to the regions defined in this study (extensions shown in a darker shade). MBD residues 72-173 and NID residues 173-312. Shaded grey areas show the deleted regions. Proteins whose interactions were identified after the design of the truncated proteins are boxed.

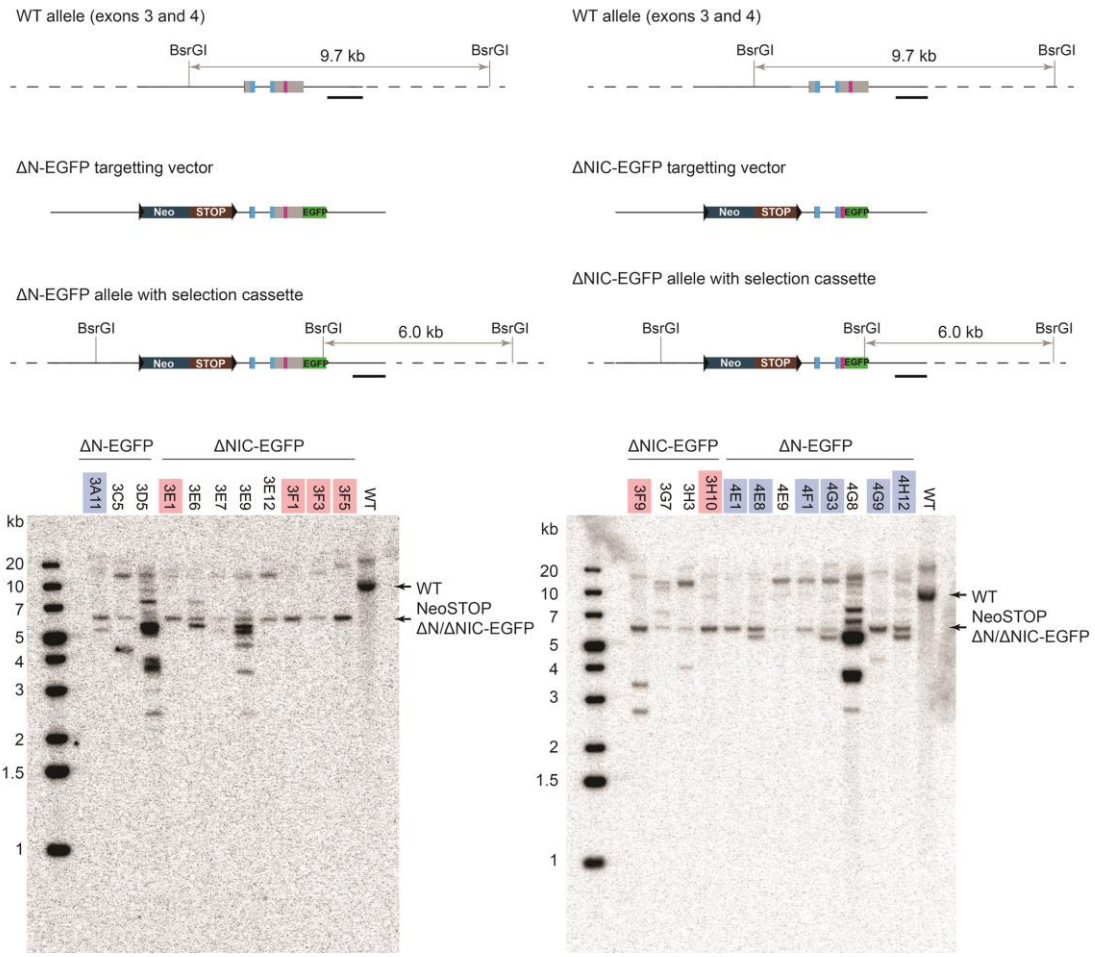


Figure S8A: Preliminary Southern blot screen of ΔN-EGFP and ΔNIC-EGFP clones

The upper portion shows the ΔN-EGFP and ΔNIC-EGFP targeting strategies with maps of the restriction sites of the enzyme used in Southern analysis, BsrGI. An internal probe, recognising the 3' homology arm was used, drawn as a black line. The clones taken forward for further analysis are highlighted (blue for ΔN-EGFP, pink for ΔNIC-EGFP).

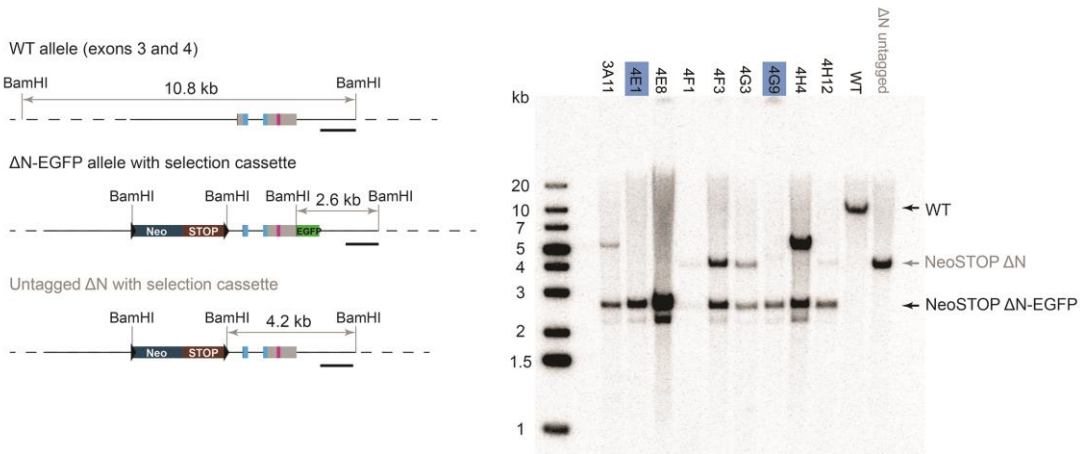


Figure S8B: Southern blotting shows two successfully targeted ΔN-EGFP clones

Screening of potential positive clones by Southern blotting. The left hand portion shows WT and ΔN-EGFP sequences with maps of the restriction sites of the enzyme used in Southern analysis, BamHI. The untagged ΔN allele is produced by internal recombination. The untagged ΔN clone 1A2 (produced during ΔNC-EGFP targeting) was included as a control in this analysis. An internal probe, recognising the 3' homology arm was used, drawn as a black line. The positive clones, 4E1 and 4G9, are highlighted in blue.

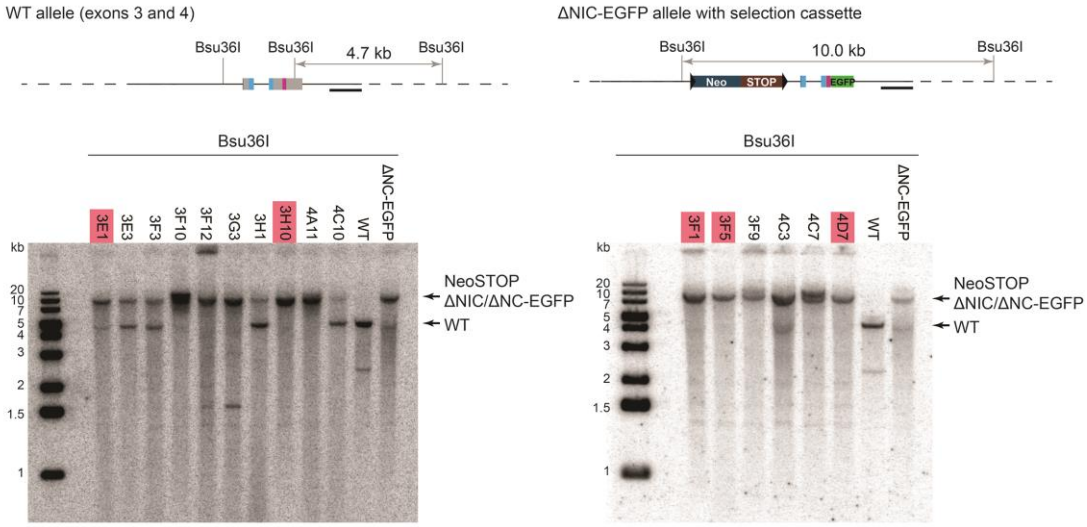


Figure S8C: Southern blotting shows five successfully targeted ΔNIC-EGFP clones

Screening of potential positive clones by Southern blotting. The top portion shows WT and ΔNIC-EGFP genomic DNA sequences with maps of the restriction sites of the enzyme used in Southern analysis, Bsu36I. An internal probe, recognising the 3' homology arm was used, drawn as a black line. The positive clones: 3E1, 3F1, 3F5, 3H10 and 4D7, are highlighted in darker pink.

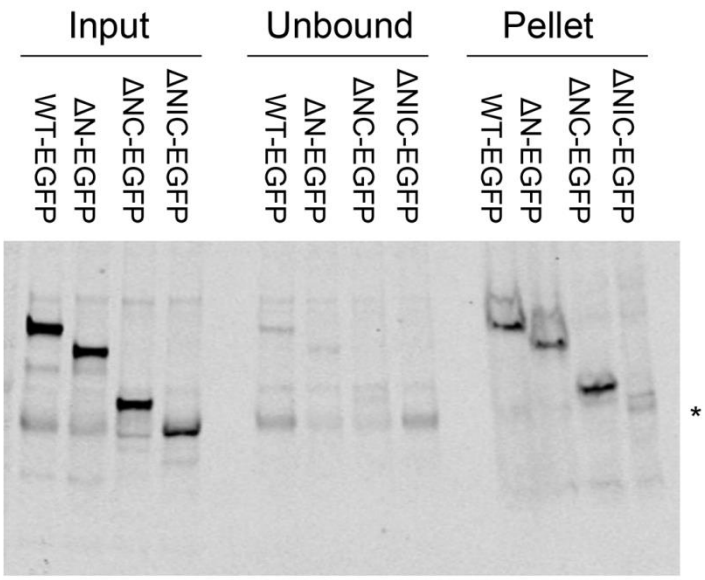


Figure S9: Making nuclear extracts with knock-in mouse brains for immunoprecipitation analysis
Following the MeCP2 truncation proteins through the steps of the nuclear extraction protocol. Nuclear extracts (inputs) are used for immunoprecipitation, the pellet at this stage is discarded. After incubation with GFP-TRAP beads, the remaining MeCP2 in the unbound supernatant was determined. Analysis by western blotting using a GFP antibody to detect MeCP2 truncation proteins. * denotes the presence of IgG from small amounts of blood in these samples, detected by the mouse secondary antibody.

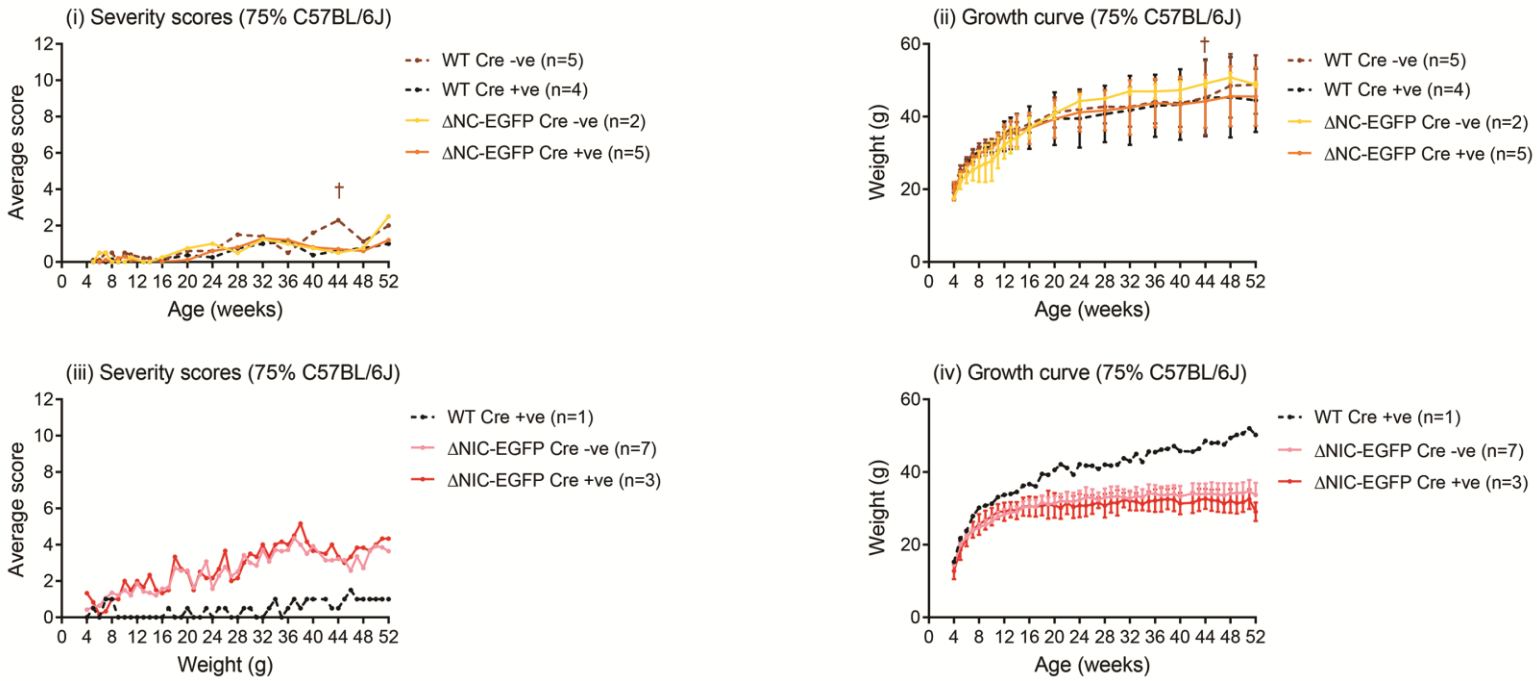


Figure S10: The *CMV-Cre* transgene does not affect the phenotypic score or weight of animals in the outbred cohorts
(i) The average scores of ΔNC-EGFP mice and their wild-type littermates (as shown in Fig. 5.2.2Aii) separated into Cre negative and positive groups. † denotes the time point where one wild-type Cre -ve was culled due to weight loss, aged 10 weeks. (ii) The weights of these animals (as shown in Fig. 5.2.2Bii) separated into Cre negative and positive groups. Graph shows mean and standard deviation. The presence of Cre has no significant effect (analysed by repeated measures ANOVA): WT Cre –ve vs Cre +ve n=0.633, ΔNC-EGFP Cre –ve vs Cre +ve n=0.949. (iii) The average scores of ΔNIC-EGFP mice and their wild-type littermates (as shown in Fig. 5.2.3Aii) separated into Cre negative and positive groups. (ii) The weights of these animals (as shown in Fig. 5.2.3Bii) separated into Cre negative and positive groups. Graph shows mean and standard deviation. The presence of Cre has no significant effect (analysed by repeated measures ANOVA): ΔNIC-EGFP Cre –ve vs Cre +ve n=0.482. *Note: larger sample sizes are required for proper statistical analysis.*

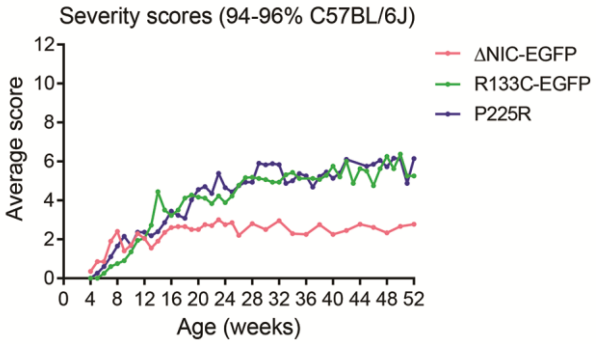


Figure S11: ΔNIC-EGFP phenotype is less severe than the mildest RTT models, R133C-EGFP and P225R

Scoring of backcrossed hemizygous male ΔNIC-EGFP mice (this study) compared to R133C-EGFP mice (Brown, Selfridge et al., 2016) and P225R mice (Guy, unpublished). Graphs show mean values. All cohorts were scored from 4-52 weeks of age. ΔNIC-EGFP n=10, R133C-EGFP n=10 and P225R n=19.