A BRAIN-DERIVED MECP2 COMPLEX SUPPORTS A ROLE FOR MECP2 IN RNA PROCESSING

BY

STEVEN WAYNE LONG

DISSERTATION

**Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Cell and Developmental Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2010**

Urbana, Illinois

Doctoral Committee:

Professor Andrew S. Belmont, Chair
Assistant Professor Peter L. Jones, Director of Research
Associate Professor Michel Bellini
Assistant Professor Christopher J Schoenherr
Assistant Professor Lin-Feng Chen

Abstract

Methyl CpG binding protein 2 (MeCP2) was originally characterized as a transcriptional repressor that preferentially bound methylated DNA, however, recent data indicates MeCP2 is a multifunctional protein. MeCP2 is now shown to associate with expressed genes as well as repressed genes indicating its gene regulatory function is context dependent. In addition, MeCP2 is involved in nuclear organization and proposed to regulate mRNA splicing. Mutations in *MECP2* are linked to the severe postnatal neurodevelopmental disorder Rett Syndrome (RTT). To further understand MeCP2 and potential roles in RTT pathogenesis, we have employed a biochemical approach to identify the MeCP2 protein complexes present in the mammalian brain. Here we show that MeCP2 exists in at least four biochemically distinct pools in the brain. We characterize one novel brain-derived MeCP2 complex that contains the splicing factor Prpf3 and Sdccag1. MeCP2 directly interacts with both Prpf3 and Sdccag1 *in vitro* independent of nucleic acids and certain RTT truncation disrupt the MeCP2-Prpf3-Sdccag1 complex. In addition, MeCP2 is localized to transcriptionally active *Xenopus* lampbrush chromosome loops and, both MeCP2 and Prpf3 associate *in vivo* with mRNAs of some genes thought to be regulated by MeCP2. This data supports a regulatory role for MeCP2 in mRNA biogenesis and suggests an additional mechanism for RTT pathophysiology.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION TO THE METHYL-CpG-BINDING PROTEIN MeCP2

DNA METHYLATION

A significant number of genes and a majority of non-protein coding DNA in a eukaryotic genome is silenced in any particular cell during the organism's life (Elgin and Grewal, 2003)**.** This 'silencing' is associated with the formation of heterochromatic domains and is influenced by covalent modifications of chromatin components. These epigenetic modifications are essential for normal development and can be re-established after each cell division to maintain stable cell identity (Bird, 2002; Jenuwein and Allis, 2001; Kouzarides, 2002; Turner, 2000). DNA methylation in the vertebrate genome is one such epigenetic mark generally associated with heterochromatin.

Methylation of vertebrate genomes are characterized by the post-replication addition of a methyl group to the 5 carbon of cytosine (5mC) in CpG dinucleotides resulting in symmetrically methylated CpG dinucleotides on each strand. The methylation reaction is carried out by members of the DNA methylatransferase (DNMT) family of proteins. Currently DNMT1, DNMT3A and DNMT3B are the only recognized metazoan DNMTs (recently DNMT2 was identified as a tRNA methyltransferase (Goll *et al.*, 2006). Members of this family contain variable N-terminal regulatory regions and conserved C-terminal domain that catalyzes the transfer of a methyl-group from S-adenosyl methionine (SAM) to the C5 position of cytosine.

The patterns of DNA methylation in a vertebrate genome are established through a sequence of developmental events (Razin and Shemer, 1995). For examples, in mice, pre-implantation embryos undergo a global demethylation event followed by *de novo* remethylation carried out by DNMT3A and DNMT3B, which are critical in determining somatic DNA methylation patterns (Hsieh, 1999; Jaenisch *et al.*, 1982; Okano *et al.*, 1998; Okano *et al.*, 1999). After these initial methylation patterns are established, they are maintained between cell generations by the maintenance methyltransferase DNMT1 (Bestor *et al.*, 1988; Leonhardt *et al.*, 1992). However, once established, DNA methylation patterns are not necessarily static and DNA demethylation can occur. One regulator of active demethylation, Gadd45a, has recently been identified and shown to act by promoting DNA repair and ultimately demethylation of methylated DNA (Barreto *et al.*, 2007; Rai *et al.*, 2008; Schmitz *et al.*, 2009). Interestingly, Gadd45a knockout mice do not result in site specific or global hypermethylation (Engel *et al.*, 2009), and the role of Gadd45a in DNA demethylation has been disputed (Jin *et al.*, 2008). Regardless, changes in developmental methylation patterns would suggest there are mechanisms controlling DNA methylation patterns in a specific manner.

Studies suggest DNA methylation is important in establishing and maintaining heterochromatic structures for gene regulation (Eden *et al.*, 2003; Gaudet *et al.*, 2003). Transient transfection and *in vitro* transcription assays consistently demonstrate that the repression of transcription is dependent on DNA methylation (Levine *et al.*, 1993; Murray and Grosveld, 1987). However, experiments comparing transcription rates of methylated versus unmethylated stably integrated transgenes during *Xenopus* embryo development suggest the repressive nature of DNA methylation may be developmentally

controlled (Appendix A). Two models have been proposed to explain the inhibitory effect of DNA methylation on transcription potential. The first involves the direct interference of transcription factors from their binding sites by the presence of CpG methylation (Figure 1.1). Research shows multiple proteins bind sequences containing CpG dinucleotides and some fail to bind when these sequences are methylated (Bell and Felsenfeld, 2000). The other model involves proteins with a high affinity to methyl-CpG and facilitates the formation of a chromatin environment unfavorable for active transcription (Fahrner *et al.*, 2002; Jones *et al.*, 1998) by in turn recruiting repressive chromatin modifiers such as histone deacetylases (HDACs) and histone methyltransfereases (HMTs) to change the chromatin structure (Boyes and Bird, 1991; El-Osta *et al.*, 2002; Fuks *et al.*, 2003) (Figure 1.2). This hypothesis is supported by the discovery of multiple methyl-CpG binding proteins.

METHYL-CpG-BINDING PROTEINS

Founding of the Methyl CpG Binding proteins (MBPs)

In 1989, Adrian Bird's lab was first to describe a nuclear activity that specifically interacted with CpG methylated DNA in-vitro, with no apparent requirement for a consensus binding sequence, and it was named MeCP (Meehan *et al.*, 1990; Meehan *et al.*, 1989). This was achieved using double strand synthetic oligonucleotides which had been methylated in-vitro with bacterial methtransferases. Competition of the MeCP activity binding to unmethylated DNA probes were detected only when the competitor DNA was methylated (Meehan *et al.*, 1989). This assay preferentially pulled down

protein from various tissues including: mouse brain, spleen, kidney, rat liver, and rabbit liver extracts. In the various extracts, it appeared MeCP bound the methyl CpG DNA templates without sequence specificity. However, further experimentation determined that MeCP was two distinct protein activities (named MeCP1 and MeCP2) with different requirements for methylated DNA (Meehan *et al.*, 1992). Meehan et al. identified that MeCP1 requires at least 12 symmetrically methylated CpGs while MeCP2 was able to bind a single methylated CpG pair. Considering MeCP2 was reported to be 100 times more abundant in adult somatic nuclei than MeCP1 (Meehan *et al.*, 1992), it was selected for further characterization and became the first MBP cloned and properly characterized. MeCP1 was finally identified and cloned in some years later, in 1999 (Ng *et al.*, 1999). The initial hypothesis for MBP function was that they normally bind methylated DNA in a chromatin context, leading to long-term transcriptional silencing. This idea was supported by results showing that digestion of rat brain nuclei with micrococcal nuclease released MeCP2, and that this chromatin "released" MeCP2 still retained the ability to bind a methylated DNA probe, but did not interact with unmethylated DNA (Meehan *et al.*, 1992). Furthermore, additional evidence was provided when transiently transfected constructs encoding a MeCP2 fused to the LacZ gene was expressed in mouse cell culture and similar localization to centromeric heterochromatin was observed (Lewis *et al.*, 1992). Interestingly, the MeCP2-LacZ fusion was unable to localize to centromeric heterochromatin in methyltransferase deficient mouse cells, indicating MeCP2 requires a methylated chromatin substrate for proper distribution (Lewis *et al.*, 1992). This early MBP hypothesis prompted the isolation of MeCP2's methyl CpG binding domain (MBD) by domain mapping, identifying amino acids 78 to 162 as the minimal region required to

bind methylated DNA (Nan *et al.*, 1993) (Figure 1.3). The identification of the first functional domain of MeCP2 allowed for bioinformatic analysis, which in turn led to identification of multiple relatives that share a related MBD (Cross *et al.*, 1997; Hendrich and Bird, 1998).

To date there are many proteins which have been identified as MBPs. The known proteins which have affinity to methylated CpGs include the methyl CpG binding protein family (MBD1, MBD2, MBD3, MBD4, and MeCP2), and a structurally unrelated Kaiso family (Kaiso, ZBTB4, and ZBTB38).

The Kaiso protein family

Kaiso, ZBTB4, and ZBTB38 are members of Kaiso protein family which lack the conserved MBD characteristic of MBPs. Instead, these proteins preferentially bind methylated DNA through zinc-finger domains (Filion *et al.*, 2006; Prokhortchouk *et al.*, 2001). Since the first Kaiso family member's discovery, it has been attributed to multiple important functions. Knockdown of Kaiso in *X. laevis* caused a premature activation of transcription in zygotes which resulted in apoptosis and developmental arrest (Ruzov *et al.*, 2004). This phenotype resembles knocking out specific DNA methyltransferases in *X. laevis* (Stancheva and Meehan, 2000), suggesting that Kaiso and DNA methylation-mediated repression mechanisms are partly responsible for repression of transcription before the mid-blastula transition. Interestingly, Kaiso knockout mice result in apparently normal development (Prokhortchouk *et al.*, 2006). Further functional studies in amphibian and mammal models suggest that the biological roles for Kaiso are not conserved (Prokhortchouk *et al.*, 2006; Ruzov *et al.*, 2004; Ruzov *et al.*, 2009). The two

other zinc-finger proteins, ZBTB4, and ZBTB38, are capable of binding methyl CpG DNA and have been shown to repress transcription in transient transfection assays (Filion *et al.*, 2006).

MBD1

MBD1 is the first named MBD protein identified in a bioinformatic search for MBD containing proteins (Hendrich and Bird, 1998). MBD1 functions as a transcriptional repressor both *in vivo* and *in vitro* and can bind both methylated as well as unmethylated DNA depending on its splice isoform (Fujita *et al.*, 2000; Ohki *et al.*, 1999). *MBD1*-null mice have no obvious developmental defects, however, they do have minor neural defects and reduced genomic stability leading to increased transcription of some retrotransposon elements, supporting a role for MBD1 in repression of methylated DNA (Zhao *et al.*, 2003). MBD1 associates with chromatin modifiers such as Suv39h1 and HP1, and is involved in DNA methylation-mediated transcriptional repression (Fujita *et al.*, 2003). Another study showed that MBD1 associated with the H3K9 methytransferase SETDB1 in HeLa cells (Sarraf and Stancheva, 2004). During S phase, the MBD1-SETDB1 complex is recruited to chromatin by the chromatin assembly factor CAF1 to establish new repressive H3K9 methyl marks. Furthermore, demethylation of specific promoters by 5-Azacitidine resulted in the disruption of the MBD1-SETDB1-CAF1 complex and diminished H3K9 methylation levels, resulting in activation of normally repressed genes (Sarraf and Stancheva, 2004). Interestingly, the interaction of MBD1 with other protein factors has been found to be negatively regulated by SUMOlation, providing a mechanism to control its repressive function (Lyst *et al.*, 2006).

MBD2

The second protein identified by MBD sequence homology in the study by Hendrich et al was aptly named MBD2. Experiments from Adrian Bird's group ultimately led to MBD2 as the protein responsible for MeCP1's ability to preferentially bind methylated DNA (Ng *et al.*, 1999). It has been since determined that MBD2 is capable of binding methylated CpGs in vitro and in vivo and is responsible for transcriptional repression (Boeke *et al.*, 2000; Ng *et al.*, 1999). It turned out MeCP1 consisted of MBD2 in a complex with the Mi-2/NuRD chromatin remodeling complex (Zhang *et al.*, 1999). Developmentally, *MBD2*-null mice are apparently normal and remain viable and fertile; however female *MBD2*-null mice fail to nurture their pups properly (Hendrich *et al.*, 2001). A connection of MBD2 to this observed behavior remains unclear.

MBD3

Mammalian MBD3, although identified by its conserved MBD and is overall highly similar to MBD2, lacks the ability to preferentially bind methylated DNA in vivo or in vitro (Hendrich and Bird, 1998; Saito and Ishikawa, 2002). Studies show mammalian MBD3 associates with the Mi-2/NuRD chromatin remodeling complex and its evolutionary conserved function as a transcriptional repressor remains intact (Saito and Ishikawa, 2002). MBD2 and MBD3 interactions with Mi2/NuRD are mutually exclusive and creates two distinct complexes (Denslow and Wade, 2007; Feng and Zhang, 2001). Despite a similar amino acid sequence and mutually exclusive interactions with

the Mi-2/NuRD complex, MBD2 and MBD3 do not appear to perform redundant functions. *MBD3*-null mice, in contrast to *MBD2*-null mice, are severely affected at day 8.5 and subsequently die before birth and *MBD3*-null ES cells fail to shut down the expression of multiple undifferentiated ES cell markers leading to compromised ability to differentiate (Kaji *et al.*, 2006; Kaji *et al.*, 2007). Interestingly, in contrast to mammals where both MBD2 and MBD3 interact with Mi-2/NuRD, the *Xenopus laevis* MBD3 homolog retains the ability to bind methylated DNA and is the only MBD protein that has been shown to be a subunit of Mi-2/NuRD (Wade *et al.*, 1999).

MBD4

MBD4, much like the other MBD protein family members, has the potential to function as a transcriptional repressor – as demonstrated by a series of in vitro repression assays (Kondo *et al.*, 2005). However, it appears that MBD4 functions mainly as a thymine glycosylase, acting as a DNA repair protein that targets sites of cytosine deamination (Hendrich *et al.*, 1999). In methylated genomes, the GpG dinucleotide is under-represented. This observation has been linked to spontaneous hydrolytic deamination of methylated cytosine which causes methylated CpG to TpG transitions, while non methylated CpG mutates to UpG (Bird, 1980). Experimental evidence indicates MBD4 can excise and repair both types of mutated nucleotides (Hendrich *et al.*, 1999). In line with this function, *MBD4*-null mice had a two to three times higher amount of mCpG to TpG mutation rate (Millar *et al.*, 2002; Wong *et al.*, 2002). The relatively mild mutator phenotype and apparently healthy mice in these studies suggest that other thymine glycosylases might carry out similar functions. Interestingly, one

recent study involving Gadd45 mediated active demethylation presented evidence that MBD4 may be involved in the process (Rai *et al.*, 2008).

MeCP2

Since the time initial experiments in Adrian Bird's lab identified MeCP2 as the founding member of methyl CpG binding proteins, multiple molecular roles have been proposed for the protein. Studies suggest that MeCP2 is involved in transcriptional repression, transcriptional activation, nuclear organization, and splicing. To further complicate the biology of MeCP2, new information about the MeCP2 gene itself has been reported. A second isoform, labeled the e1 isoform, differs only in the first exon by providing an alternate amino terminal region when compared to the previously characterized e2 isoform of MeCP2 (Mnatzakanian *et al.*, 2004) (Figure 1.3). The exact molecular context in which MeCP2 proteins functions in the proposed processes is not fully understood; however the data supporting the possible roles will be elaborated on further.

Transcriptional regulation by MeCP2

The MBD was the first functional domain isolated on MeCP2. Deletion studies showed MeCP2 requires the MBD for its high binding affinity specifically for methylated DNA *in vitro* and for localization to the highly methylated pericentromeric heterochromatin in cell culture (Nan *et al.*, 1993; Nan *et al.*, 1996). These data influenced the early driving hypothesis that MeCP2 binds methylated DNA in a chromatin context and function in long-term global silencing. These studies were

conducted using murine MeCP2 in mouse cell cultures. Further characterization of other MeCP2 homologs showed different preferences for methylated over unmethylated DNA. For example, the *Xenopus* MeCP2 homolog has a twenty fold higher affinity for methylated DNA when compared to unmethylated DNA (Fraga *et al.*, 2003), where as the preference for human MeCP2 for methylated versus unmethylated DNA is 2 - 3 fold (Nikitina *et al.*, 2007a). Later experimentation involving *in vitro* binding site selection would reveal a requirement of human MeCP2 for an A/T nucleotide rich sequence adjacent to the target CpG methylation site, indicating a basis for sequence specificity in addition to methylation for MeCP2 binding (Klose *et al.*, 2005).

A biological role of MeCP2 was first illustrated by showing the protein represses transcription on methylated promoters *in vitro* and in the cell (Nan *et al.*, 1997). Using a β-actin transient transfection transcription assay, various MeCP2 domains were fused to the Gal 4 DNA binding domain and tested for the ability to repress transcription (Nan *et al.*, 1997). This method resulted in the identification of a specific fragment between amino acids 205 and 310 which is required for transcriptional silencing, defining the second domain of MeCP2 – the transcriptional repression domain (TRD) (Figure 1.3). Further experimentation has let to identification of other conserved domains in MeCP2, for example, a nuclear localization signal (Nan *et al.*, 1996) and the WW domain binding region (WDR) thought to be involved in protein-protein interactions (Buschdorf and Stratling, 2004) (Figure 1.3).

The ability of MeCP2 to silence transcription and its localization to densely methylated heterochromatin further supports the idea that MeCP2 functions as a general DNA methylation dependent transcriptional repressor. One of the most compelling

models for MeCP2 mediated transcriptional repression is MeCP2 binding methylated DNA and subsequently the TRD of MeCP2 recruiting co-repressor complexes to create a chromatin environment unfavorable for transcription (Figure 1.2). This model, suggesting MeCP2 as a functional link between DNA methylation and transcriptional repression, is supported by the observation that the TRD binds the co-repressor mSin3A which recruits histone deacetylase 1 and 2 (Jones *et al.*, 1998; Nan *et al.*, 1998; Wade *et al.*, 1998). The MeCP2-Sin3A-HDAC interaction is illustrated in multiple studies. Jones et al. (1998) established that MeCP2 stably co-fractionates with the Sin3A-HDAC chromatin remodeling complex in *Xenopus laevis,* and that MeCP2 TRD mediated repression was relieved upon treatment with the histone deacetylase inhibitor TSA. Nan et al. (1998) found immunoprecipitation of MeCP2 from rat brain nuclear extract contained Sin3A and histone deacetylase activity. This idea is further supported by evidence that two MeCP2 target promoters, the brain-derived neurotrophic factor (*Bdnf*) and *xHairy2a,* are also bound by Sin3A (Chen *et al.*, 2003; Klose and Bird, 2003; Martinowich *et al.*, 2003; Stancheva *et al.*, 2003). These data indicate MeCP2 is targeted to methylated promoter DNA and result in strong transcriptional repression via post-translational modifications of histone tails (Jones *et al.*, 1998; Nan *et al.*, 1998). However, this model is not universally applicable to MeCP2 function due to the observation that, depending on the promoter in question, histone deacetylase independent transcriptional repression by the protein can occur (Yu *et al.*, 2000). Furthermore, a lack of global changes in histone tail modifications in *MeCP2*-null mice also argues against this as the sole functional model (Urdinguio *et al.*, 2007).

Additional co-repressors, such as NCoR and c-Ski, have also been identified as interacting directly with MeCP2 (Kokura *et al.*, 2001). Furthermore, Harikrishnan et al. (2005) demonstrated that that SWI/SNF chromatin-remodeling complex Brahma associates with MeCP2 in cell culture and is functionally linked with repression. The stability of co-repressor interactions with MeCP2 is controversial due to data from one study that shows the native protein is an elongated monomer that does not associate with anything in rat brain nuclear extracts or *Xenopus* oocyte extracts (Klose and Bird, 2004). Oddly, this study is in disagreement with that same lab's previous published and unretracted work showing MeCP2 stably interacts with Sin3A-HDAC in rat brain nuclear extracts (Nan *et al.*, 1998). Thus, the biochemical makeup of endogenous MeCP2 complexes is still not clear.

Although experimental evidence suggests MeCP2 might serve as a global transcriptional silencer (Jones *et al.*, 1998; Nan *et al.*, 1998), transcriptional profiling studies of *MeCP2*-null mice brains exhibit only subtle changes in gene expression. Of the few genes upregulated in the *MeCP2*-null mouse model, the largest change in expression was attributed to the *Irak1* gene – showing a twofold increase in expression (Tudor *et al.*, 2002). These findings were later confirmed in a microarray hybridization study using RNA from the cerebellum of MeCP2 mutant mice (Jordan *et al.*, 2007). Similarly, cultured human cell-lines with naturally occurring dysfunctional MeCP2 mutations showed only 49 upregulated and 21 downregulated targets by comparing global expression levels of potential MeCP2 target genes (Traynor *et al.*, 2002). Taken together, these data may suggest MeCP2 is not strictly involved in global transcriptional repression. In fact, recent work raised further questions about how MeCP2 functions at

the cellular level. Yasui et al. (2007) preformed a global epigenomic binding analysis of MeCP2 using a ChIP- (microarray) chip approach. Two novel insights into possible functions of MeCP2 were revealed by this work. The first is MeCP2 is often associated with actively transcribed genes, and not always with transcriptionally repressed genes. Secondly, the majority of MeCP2 binding sites are intergenic within the genome. Using the human SH-SY5Y neuroblastoma cell line as a source of MeCP2 bound chromatin for immunoprecipitations, the ChIP-chip analysis first focused on a 26.3Mb chromosomal loci containing known MeCP2 target genes. From this, a majority of MeCP2 (59%) was found to bind mostly non-methylated sites along intergenic spaces (Yasui *et al.*, 2007). Furthermore, 63% of MeCP2-bound promoters were found to be actively transcribed (Yasui *et al.*, 2007). Being contradictory to the traditional model of MeCP2 binding densely methylated promoters to repress transcription, the group compared genome-wide promoter methylation to MeCP2 binding by methylated DNA immunoprecipitation (MeDIP). The comparison revealed that only 2% of promoters with high levels of methylation were bound by MeCP2 (Yasui *et al.*, 2007). Importantly, the results obtained from the cell culture based experiments were reinforced by an *in vivo* model, further suggesting MeCP2 is associated with actively transcribed genes. Experiments profiling transcript expression levels of hypothalamus isolated from MeCP2 overexpressing mice or *MeCP2*-null mice and comparing them to wild type mice show that MeCP2 is not strictly associated with gene repression, but activate transcription on the majority of its targets (Chahrour *et al.*, 2008). Although indirect effects were considered, the presence of MeCP2 at some of the activated gene promoters was validated by ChIP and shown to be associated with the transcriptional activator CREB-1

(Chahrour *et al.*, 2008). These surprising results illustrate that the complete biological function of MeCP2 is not fully understood.

Aside from the global aspects of transcriptional regulation by MeCP2, its ability to regulate specific loci has been investigated. One extensively characterized target of MeCP2 is the brain-derived neurotrophic factor (BDNF) promoter III. Initial studies in mouse and rat neurons show MeCP2 dissociates from the BDNF promoter upon a calcium influx induced membrane depolarization, allowing transcription of BDNF to occur (Chen *et al.*, 2003; Martinowich *et al.*, 2003). ChIP experiments show the repression of BDNF is mediated though MeCP2 recruitment of the Sin3A-HDAC co-repressor complex, and depolarization disrupts chromatin remodeling activities required for repression of BDNF (Martinowich *et al.*, 2003). Furthermore, Western blot analysis of neuronal lysates revealed MeCP2 is phosphorylated by membrane depolarization, leading to dissociation from the BDNF promoter binding site (Chen *et al.*, 2003). Later work revealed multiple phosphorylation sites on neuronally derived MeCP2 (Tao *et al.*, 2009; Zhou *et al.*, 2006). MeCP2 can be phosphorylated at the serine 421 residue by a CaMKII kinase. This post-translational modification is responsible for MeCP2 dissociation and transcriptional activity of the BDNF promoter (Zhou *et al.*, 2006). Interestingly, phosphorylation at another residue, serine 80, can also modulate chromatin binding of MeCP2. In contrast to serine 421, phosphorylation of serine 80 is important for the association of MeCP2 with chromatin, and membrane depolarization of neurons by calcium influx induces dephosphorylation of this residue (Tao *et al.*, 2009).

Other functions of MeCP2

Affinity for DNA methylation and histone deacetylase dependent chromatin remodeling has been the dominant topic in the history of studying MeCP2. However, research over the years has expanded the proteins functional role beyond that of transcriptional regulation. The initial implication that MeCP2 may have a role in global chromatin architecture came from early research showing chicken MeCP2 (ARBP) has a high affinity for matrix attachment regions (MARs) (Weitzel *et al.*, 1997). Later, *in vitro* studies demonstrate binding of MeCP2 to chromatin templates and directly inducing chromatin compaction (Georgel *et al.*, 2003; Nikitina *et al.*, 2007a; Nikitina *et al.*, 2007b). Although there is an obvious preference for MeCP2 binding methylated linker DNA of chromatin templates (Ishibashi *et al.*, 2008; Nikitina *et al.*, 2007a), it is clear that MeCP2 is able interact with unmethylated chromatin templates. Using nucleosome arrays and electron microscopy, MeCP2 has been visualized binding chromatin fibers, in the absence of DNA methylation, ATP, or other chromatin remodelers, and directly compacting them into folded structures (Georgel *et al.*, 2003; Nikitina *et al.*, 2007b). During myogenic differentiation of tissue culture cells, MeCP2 can induce large scale rearrangements of heterochromatin (Brero *et al.*, 2005). Ectopic expression of fluorescently tagged MeCP2 was shown to mimic the effect, causing a dose-dependent clustering of chromocenters in the absence of differentiation (Brero *et al.*, 2005). Moreover, experiments utilizing ChIP coupled with chromosome conformation capture indicate the Dlx5-Dlx6 imprinted gene cluster requires MeCP2 to form chromatin loops on the imprinted allele (Horike *et al.*, 2005). Interestingly, in the absence of MeCP2, one

of the few genes upregulated in knock-out mice is *SATB1*, a protein known to bind specifically to MARs and mediate the formation of chromatin loops (Jordan *et al.*, 2007).

In addition to transcriptional regulation and nuclear architecture, MeCP2 function has also been linked to mRNA splicing (Jeffery and Nakielny, 2004; Young *et al.*, 2005). One *in vitro* study indicates that MeCP2, via the interdomain region between the MBD and TRD, can bind RNA directly with high affinity (Jeffery and Nakielny, 2004) (Figure 1.3). Furthermore, the binding of RNA is mutually exclusive from MeCP2 interacting with methylated DNA, and can do so independently of the MBD (Jeffery and Nakielny, 2004). In another study using HeLa cell extracts for coimmunoprecipitation experiments, epitope tagged MeCP2 was found to interact with the messenger ribonucleoprotein particle component YB-1 (Young *et al.*, 2005). The YB-1 protein is a multifunctional protein and considered to be the main mRNA packing protein. The MeCP2-YB-1 interaction is dependent on the presence of RNA, as MeCP2 coimmunoprecipitation treated with RNase failed to pull down YB-1 along with MeCP2 (Young *et al.*, 2005). It is unknown if the MeCP2-YB-1 complex is a protein-protein interaction stabilized by RNA or if it is a result of the two proteins interacting with the same mRNA transcript. Additionally, MeCP2 overexpression in cell culture can affect the alternative splicing of transiently transfected reporter minigenes (Young *et al.*, 2005). This information, along with a microarray splicing analysis of cerebral cortex mRNA isolated from MeCP2-null mice showing multiple aberrantly spliced genes, imply that MeCP2 may have a role as a splice site regulator (Young *et al.*, 2005).

MeCP2 and Rett syndrome

Mutations in *MeCP2* cause the majority (96%) of cases diagnosed as the severe neurodevelopmental disorder Rett syndrome (RTT) (Amir *et al.*, 1999). *MeCP2* is an X-linked gene and RTT occurs in approximately one out of every 10,000 females (Percy, 2002), however non-lethal mutations of the gene in males have been reported (Ravn *et al.*, 2003). Females with RTT develop normally after birth up to 6 to 18 months of age, after which they show regression, with deceleration of head growth, loss of speech and acquired motor skills, as well as severe mental retardation (Hagberg *et al.*, 1983). Of the 2000 reported mutations in *MeCP2* causing RTT, there are eight common mutational hotspots found in the MBD and TRD that result in loss of a functional protein (Bienvenu and Chelly, 2006). Taking into account both MeCP2 e1 and e2, mutations predicted to affect either isoform alone are rare yet known to cause RTT (Mnatzakanian *et al.*, 2004). Although both isoforms are highly expressed in the brain and MeCP2 e1 is more prevalent (Kriaucionis and Bird, 2004), and there is differential distribution of the two between the dorsal thalamus and hypothalamus in developing postnatal mouse brains (Dragich *et al.*, 2007). The functional significance of this is unknown, but suggests there may be important difference between the two isoforms.

Multiple mouse models of RTT have been developed using Cre-*lox* technology, each having different mutation type and phenotype severity. *MeCP2* knock out mice show a period of normal development followed by progressive neurological dysfunction leading to death by 10 weeks (Chen *et al.*, 2001; Guy *et al.*, 2001). *MeCP2* mutations restricted to neuronal lineages result in a phenotype indistinguishable from that of *MeCP2* knockout mice, indicating MeCP2 dysfunction in neurons is sufficient to cause

disease (Chen *et al.*, 2001; Guy *et al.*, 2001). Inactivation of *Mecp2* in post-mitotic neurons cause delayed onset of phenotypes similar to those of *MeCP2*-null mice, demonstrating that MeCP2 plays an essential role in post-mitotic neuronal function (Chen *et al.*, 2001). Conversely, expression of MeCP2 in post-mitotic neurons of *MeCP2* knockout mice show normal neurological function, indicating that MeCP2 plays no essential roles in early brain development and deficiency in non-neuronal tissues does not significantly influence disease progression (Luikenhuis *et al.*, 2004). However, this idea is being challenged by a recent study indicating the lack of MeCP2 specifically in glial cells contributes to RTT phenotypic neurons by an unknown secreted glial factor (Ballas *et al.*, 2009). Further functional studies found that mice carrying a duplicate copy of the gene and therefore overexpressing MeCP2 result in a severe postnatal, neural phenotype (Collins *et al.*, 2004). Taken all together, it appears MeCP2 must be tightly regulated in the brain for normal function.

MeCP2 has been implicated in gene regulation, splicing, and genome organization, yet how MeCP2 functions in the brain and the lack of a functional MeCP2 protein leads to disease is unknown. There is no solid evidence linking a target of MeCP2 to RTT, however, BDNF overexpression in MeCP2 knockout mice do show an extended life span (Chang *et al.*, 2006). Notably, *MeCP2*-null mice show aberrant RNA-splicing patterns; indicating lack of MeCP2 may not only affect transcriptional control (Young *et al.*, 2005). The complexity of RTT would suggest MeCP2 dysfunction is a combination of affects recognized as disease. Importantly, studies suggest that mice deficient in MeCP2 do not suffer irreversible damage, and that restoration of a functional MeCP2 can lead to a

reduction of neurological symptoms (Giacometti *et al.*, 2007; Guy *et al.*, 2007; Luikenhuis *et al.*, 2004).

There is a need to reveal the normal molecular mechanisms of MeCP2 function in the brain. Identification of interacting protein partners has proven many times over as a way to elucidate the function of proteins. Characterization of novel brain-specific MeCP2 protein complexes will yield insight into its molecular function and dysfunction in RTT. Here we biochemically characterized MeCP2 in adult rat brain and show that it exists in multiple biochemically distinct pools. One brain-derived MeCP2-complex shows a direct association with the pre-mRNA processing factor 3 (Prpf3), a known small nuclear ribonucleoprotein associated factor, and the serologically defined colon cancer antigen 1 (Sdccag1), a mediator of nuclear export. Furthermore, we find that MeCP2 and Prpf3 associates with mRNA *in-vivo*, further supporting the regulatory role of MeCP2 in mRNA splicing and providing another potential mechanism of pathogenesis in RTT.

REFERENCES

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.

Barreto, G., Schafer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Doderlein, G., Maltry, N., Wu, W., Lyko, F., and Niehrs, C. (2007) Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**: 671-675.

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

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* **203**: 971-983.

Bienvenu, T., and Chelly, J. (2006) Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. *Nat Rev Genet* **7**: 415-426.

Bird, A. (2002) 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.

Boeke, J., Ammerpohl, O., Kegel, S., Moehren, U., and Renkawitz, R. (2000) The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. *J Biol Chem* **275**: 34963-34967.

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

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.

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

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

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

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

Chen, 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* **302**: 885-889.

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

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.

Denslow, S.A., and Wade, P.A. (2007) The human Mi-2/NuRD complex and gene regulation. *Oncogene* **26**: 5433-5438.

Dragich, J.M., Kim, Y.H., Arnold, A.P., and Schanen, N.C. (2007) Differential distribution of the MeCP2 splice variants in the postnatal mouse brain. *J Comp Neurol* **501**: 526-542.

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

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

Elgin, S.C., and Grewal, S.I. (2003) Heterochromatin: silence is golden. *Curr Biol* **13**: R895-898.

Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E., Sapienza, C., Hoffman, B., and Liebermann, D.A. (2009) Conserved DNA methylation in Gadd45a(-/-) mice. *Epigenetics* **4**: 98-99.

Fahrner, J.A., Eguchi, S., Herman, J.G., and Baylin, S.B. (2002) Dependence of histone modifications and gene expression on DNA hypermethylation in cancer. *Cancer Res* **62**: 7213-7218.

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

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

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

Fujita, N., Shimotake, N., Ohki, I., Chiba, T., Saya, H., Shirakawa, M., and Nakao, M. (2000) Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. *Mol Cell Biol* **20**: 5107-5118.

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

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.

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

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.

Giacometti, E., Luikenhuis, S., Beard, C., and Jaenisch, R. (2007) Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. *Proc Natl Acad Sci U S A* **104**: 1931-1936.

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

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* **315**: 1143-1147.

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.

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.

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.

Hsieh, C.L. (1999) In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. *Mol Cell Biol* **19**: 8211-8218.

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

Jaenisch, R., Harbers, K., Jahner, D., Stewart, C., and Stuhlmann, H. (1982) DNA methylation, retroviruses, and embryogenesis. *J Cell Biochem* **20**: 331-336.

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.

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

Jin, S.G., Guo, C., and Pfeifer, G.P. (2008) GADD45A does not promote DNA demethylation. *PLoS Genet* **4**: e1000013.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**: 187-191.

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

Kaji, K., Caballero, I.M., MacLeod, R., Nichols, J., Wilson, V.A., and Hendrich, B. (2006) The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* **8**: 285-292.

Kaji, K., Nichols, J., and Hendrich, B. (2007) Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* **134**: 1123-1132.

Klose, R., and Bird, A. (2003) Molecular biology. MeCP2 repression goes nonglobal. *Science* **302**: 793-795.

Klose, R.J., and Bird, A.P. (2004) MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J Biol Chem* **279**: 46490-46496.

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

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

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

Kouzarides, T. (2002) Histone methylation in transcriptional control. *Curr Opin Genet Dev* **12**: 198-209.

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.

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

Levine, A., Yeivin, A., Ben-Asher, E., Aloni, Y., and Razin, A. (1993) Histone H1-mediated inhibition of transcription initiation of methylated templates in vitro. *J Biol Chem* **268**: 21754-21759.

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.

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., Nan, X., and Stancheva, I. (2006) Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. *Embo J* **25**: 5317-5328.

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

Meehan, R., Antequera, F., Lewis, J., MacLeod, D., McKay, S., Kleiner, E., and Bird, A.P. (1990) A nuclear protein that binds preferentially to methylated DNA in vitro may play a role in the inaccessibility of methylated CpGs in mammalian nuclei. *Philos Trans R Soc Lond B Biol Sci* **326**: 199-205.

Meehan, R.R., Lewis, J.D., McKay, S., Kleiner, E.L., and Bird, A.P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58**: 499-507.

Meehan, R.R., Lewis, J.D., and Bird, A.P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res* **20**: 5085-5092.

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

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

Murray, E.J., and Grosveld, F. (1987) Site specific demethylation in the promoter of human gamma-globin gene does not alleviate methylation mediated suppression. *Embo J* **6**: 2329-2335.

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.

Ng, H.H., Zhang, Y., Hendrich, B., Johnson, C.A., Turner, B.M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* **23**: 58-61.

Nikitina, T., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., Grigoryev, S.A., and Woodcock, C.L. (2007a) MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. *J Biol Chem* **282**: 28237-28245.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007b) Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Mol Cell Biol* **27**: 864-877.

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

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

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**: 247-257.

Percy, A.K. (2002) Rett syndrome. Current status and new vistas. *Neurol Clin* **20**: 1125-1141.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* **15**: 1613-1618.

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

Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008) DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**: 1201-1212.

Ravn, K., Nielsen, J.B., Uldall, P., Hansen, F.J., and Schwartz, M. (2003) No correlation between phenotype and genotype in boys with a truncating MECP2 mutation. *J Med Genet* **40**: e5.

Razin, A., and Shemer, R. (1995) DNA methylation in early development. *Hum Mol Genet* **4 Spec No**: 1751-1755.

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

Ruzov, A., Hackett, J.A., Prokhortchouk, A., Reddington, J.P., Madej, M.J., Dunican, D.S., Prokhortchouk, E., Pennings, S., and Meehan, R.R. (2009) The interaction of xKaiso with xTcf3: a revised model for integration of epigenetic and Wnt signalling pathways. *Development* **136**: 723-727.

Saito, M., and Ishikawa, F. (2002) The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J Biol Chem* **277**: 35434-35439.

Sarraf, S.A., and Stancheva, I. (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell* **15**: 595-605.

Schmitz, K.M., Schmitt, N., Hoffmann-Rohrer, U., Schafer, A., Grummt, I., and Mayer, C. (2009) TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Mol Cell* **33**: 344-353.

Stancheva, I., and Meehan, R.R. (2000) Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes Dev* **14**: 313-327.

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

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

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

Tudor, M., Akbarian, S., Chen, R.Z., and Jaenisch, R. (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc Natl Acad Sci U S A* **99**: 15536-15541.

Turner, B.M. (2000) Histone acetylation and an epigenetic code. *Bioessays* **22**: 836-845.

Urdinguio, R.G., Pino, I., Ropero, S., Fraga, M.F., and Esteller, M. (2007) Histone H3 and H4 modification profiles in a Rett syndrome mouse model. *Epigenetics* **2**: 11-14.

Wade, P.A., Jones, P.L., Vermaak, D., Veenstra, G.J., Imhof, A., Sera, T., Tse, C., Ge, H., Shi, Y.B., Hansen, J.C., and Wolffe, A.P. (1998) Histone deacetylase directs the dominant silencing of transcription in chromatin: association with MeCP2 and the Mi-2 chromodomain SWI/SNF ATPase. *Cold Spring Harb Symp Quant Biol* **63**: 435-445.

Wade, P.A., Gegonne, A., Jones, P.L., Ballestar, E., Aubry, F., and Wolffe, A.P. (1999) Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation. *Nat Genet* **23**: 62-66.

Weitzel, J.M., Buhrmester, H., and Stratling, W.H. (1997) Chicken MAR-binding protein ARBP is homologous to rat methyl-CpG-binding protein MeCP2. *Mol Cell Biol* **17**: 5656-5666.

Wong, E., Yang, K., Kuraguchi, M., Werling, U., Avdievich, E., Fan, K., Fazzari, M., Jin, B., Brown, A.M., Lipkin, M., and Edelmann, W. (2002) Mbd4 inactivation increases Cright-arrowT transition mutations and promotes gastrointestinal tumor formation. *Proc Natl Acad Sci U S A* **99**: 14937-14942.

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

Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S., and Zoghbi, H.Y. (2005) Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci U S A* **102**: 17551-17558.

Yu, F., Thiesen, J., and Stratling, W.H. (2000) Histone deacetylase-independent transcriptional repression by methyl-CpG-binding protein 2. *Nucleic Acids Res* **28**: 2201-2206.

Zhang, Y., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13**: 1924-1935.

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., Summers, R.G., Chun, J., Lee, K.F., and Gage, F.H. (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.N., Ho, H.Y., Schmidt, L., Chen, W.G., Lin, Y., Savner, E., Griffith, E.C., Hu, L., Steen, J.A., Weitz, C.J., and Greenberg, M.E. (2006) Brain-specific phosphorylation of MeCP2 regulates activity-dependent Bdnf transcription, dendritic growth, and spine maturation. *Neuron* **52**: 255-269.

FIGURES

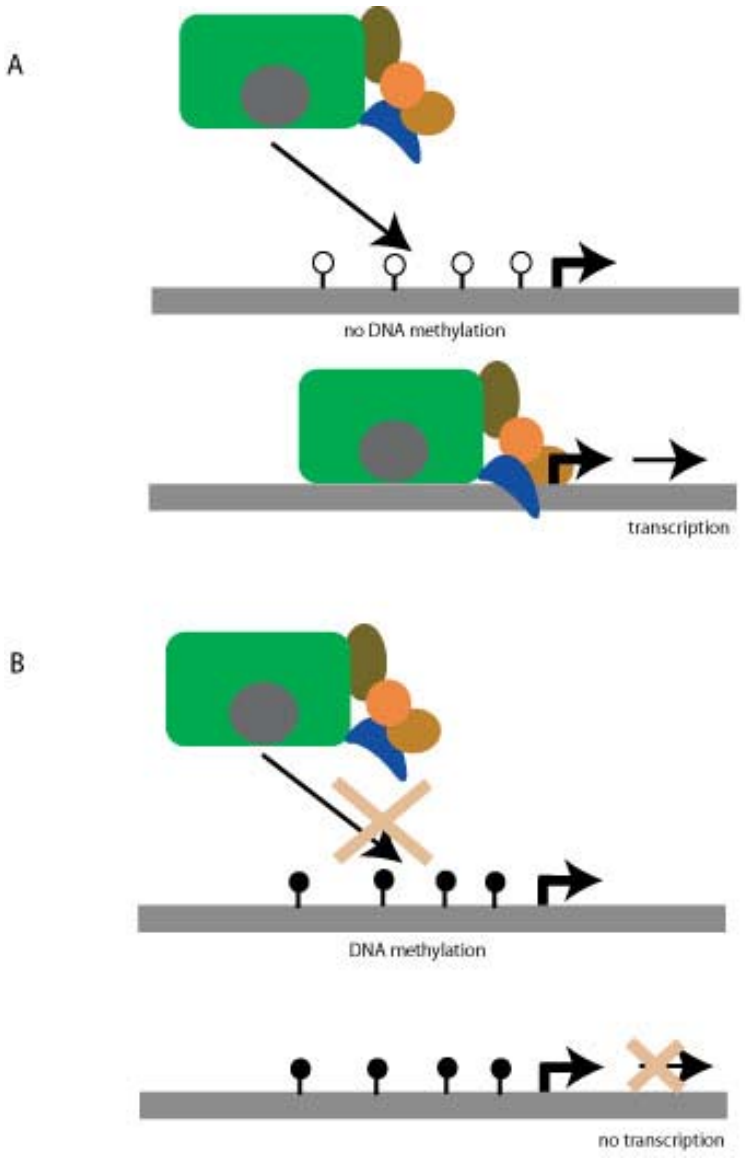

Figure 1.1: Schematic of model for DNA methylation mediated inhibition by direct interference. A) When unmethylated CpGs (open circles) are present at a target promoter, transcription factors are able to bind and initiate transcription. B) When a promoter contains CpGs that are methylated (closed circles), transcription factors are unable to bind, thus inhibiting transcription.

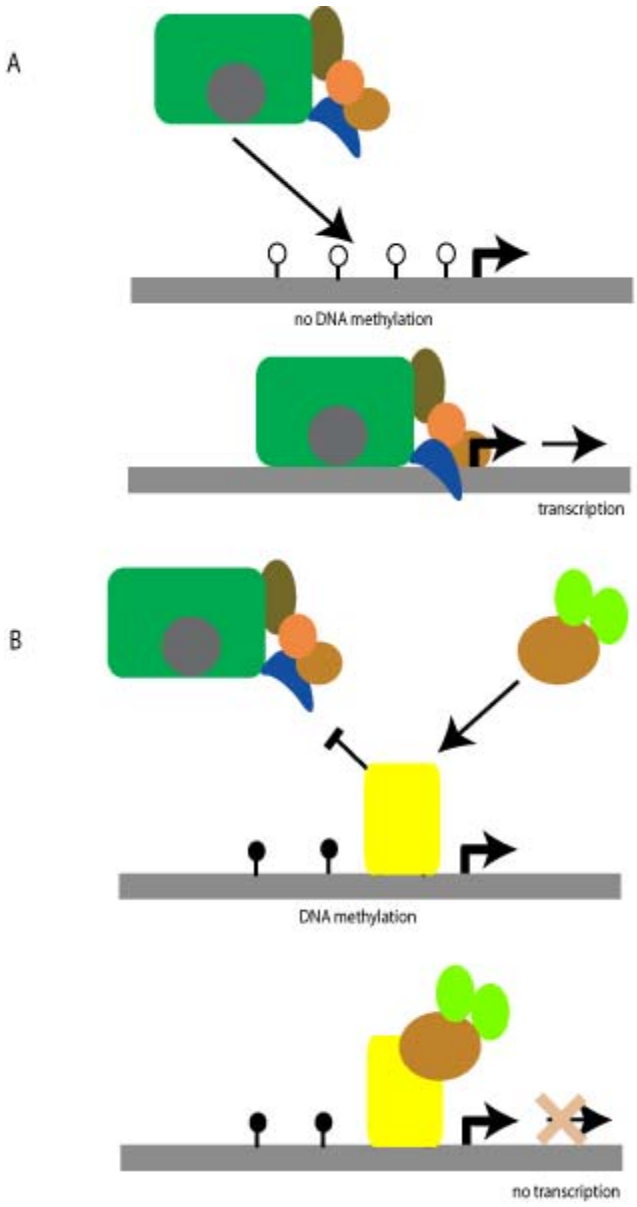

Figure 1.2: Schematic of model for DNA methylation mediated repression by recruitment of methyl-binding proteins (MBPs) and co-repressors. A) When unmethylated CpGs (open circles) are present at a target promoter, transcription factors are able to bind and initiate transcription. B) Promoter containing CpGs that are methylated (closed circles) allow for MBP binding and subsequent recruitment of repressive chromatin modifiers, thus inhibiting transcription.

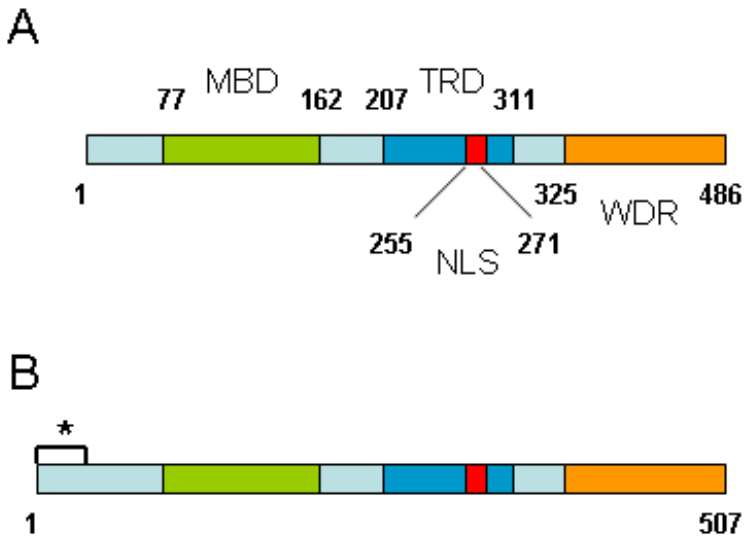

Figure 1.3: Schematic of MeCP2 domains. A) The commonly studies MeCP2 e2 isoform and its domains: MBD – methyl-binding domain; TRD – transcriptional repression domain; NLS – nuclear localization sequence; WDR - WW domain binding region. B) MeCP2 e1 isoform is identical to e2, except for an extra 5' exon providing an additional 21 amino acids (*) at the amino termini.

CHAPTER 2*

A BRAIN-DERIVED MeCP2 COMPLEX SUPPORTS A ROLE FOR MeCP2 IN RNA PROCESSING

*** Some of the data and writing is adapted from:**

Steven W. Long, Jenny Y. Y. Ooi, Peter M. Yau , and Peter L. Jones *A Brain-Derived MeCP2 Complex Supports a Role for MeCP2 in RNA Processing.* **NAR, 2010**

INTRODUCTION

MeCP2 was originally identified by its ability to preferentially bind double stranded DNA containing symmetrically methylated CpG dinucleotides and is the founding member of the methyl-CpG-binding domain (MBD) family of proteins (Hendrich and Bird, 1998; Meehan *et al.*, 1992). The first biological role for MeCP2 was illustrated by showing the protein interacts with methylated DNA *in vivo* and can repress transcription by association with a transcriptional co-repressor complex containing Sin3A and histone deacetylase (Jones *et al.*, 1998; Nan *et al.*, 1997; Nan *et al.*, 1998). In 1999 a genetic analysis identified mutations in *MECP2* as causal of Rett Syndrome (RTT), the first direct link between an epigenetic regulator and a human disease (Amir *et al.*, 1999). RTT is a severe postnatal neurodevelopmental disorder and one of the most common causes of mental retardation in females (Chahrour and Zoghbi, 2007). First described in 1966 by Andreas Rett (Rett, 1966), RTT is characterized by a period of apparently normal development from birth to 6-18 months followed by a regression of obtained language and motor skills (Chahrour and Zoghbi, 2007). RTT patients usually exhibit a deceleration of head growth, respiratory dysfunction, scoliosis, cognitive impairment, seizures, and social withdraw (Hagberg *et al.*, 1983; Rett, 1966). In addition to RTT,

numerous *MECP2* mutations have now been linked to a variety of additional disorders, including autism, Angelman syndrome, learning disabilities, and mental retardation syndromes (Carney *et al.*, 2003; Chahrour and Zoghbi, 2007; Christodoulou and Weaving, 2003; Lam *et al.*, 2000; Watson *et al.*, 2001; Ylisaukko-Oja *et al.*, 2005; Zoghbi, 2005).

Along with the discovery of MeCP2's association to human diseases, MeCP2 has also been reported to associate with myriad protein partners including Sin3A (Jones *et al.*, 1998; Nan *et al.*, 1998), c-REST and Suv39h1 (Lunyak *et al.*, 2002), c-Ski and N-CoR (Kokura *et al.*, 2001), Brm (Harikrishnan *et al.*, 2005), and HP1 (Agarwal *et al.*, 2007), all supporting a model of MeCP2 interacting with or being a stable component of co-repressor complexes, resulting in targeted transcriptional repression of methylated DNA through modification of the chromatin state. Despite this early focus on gene specific repression, more recent work has expanded MePC2's gene regulatory role beyond transcriptional repression; MeCP2 is implicated in transcriptional activation, genome-wide transcriptional silencing, mediating chromatin and nuclear architecture, and regulating pre-mRNA splicing as well (Chadwick and Wade, 2007; Chahrour *et al.*, 2008; Skene *et al.*; Yasui *et al.*, 2007; Young *et al.*, 2005). Genome-wide chromatin immunoprecipitation (ChIP) assays from cultured neuronal cell lines indicate MeCP2 binding resembles that of RNA polymerase II and MeCP2 recruits CREB to activate specific genes (Chahrour *et al.*, 2008; Yasui *et al.*, 2007). Overexpressed MeCP2 in cell culture leads to clustering of MeCP2-associated chromatin (Brero *et al.*, 2005), while *in vitro*, MeCP2 does not require any additional proteins to condense chromatin into higher order structures, illustrating a co-repressor independent function for MeCP2 in affecting chromatin structure (Georgel *et al.*, 2003; Nikitina *et al.*, 2007). Recently, a genome-

wide ChIP-bisulfite sequencing analysis found that in purified mouse brain neurons MeCP2 bound to methylated DNA sequences with high selectivity and maintained depressed global histone acetylation levels suggesting that in neurons MeCP2's main DNA associated function was as a genome-wide repressor of methylated DNA sequences (Skene *et al.*). In addition to binding methylated DNA, MeCP2 also binds multiple RNA species, including mRNA, and binding to RNA and methylated DNA are distinct, mutually exclusive interactions (Jeffery and Nakielny, 2004). Combined with the RNA-dependent association of MeCP2 with Y box-binding protein1 (YB-1), a component of messenger ribonucleoprotein (RNP) particles, to regulate splicing of reporter minigenes, MeCP2 is implicated as being directly involved in mRNA splicing (Young *et al.*, 2005). Overall, the biology of MeCP2 suggests a truly multifunctional protein (Hite *et al.*, 2009) with cell type specific functions and associations. The exact molecular process(es) disrupted in neurons by the pathogenic mutations in *MECP2* remains unclear.

Genetic studies in mice suggest that functional MeCP2 in neurons is essential for normal synapse formation and neuronal function during postnatal development and re-expression of MeCP2 in differentiated neurons alone rescues a RTT mouse model (Chen *et al.*, 2001; Guy *et al.*, 2001; Kishi and Macklis, 2004; Luikenhuis *et al.*, 2004; Shahbazian *et al.*, 2002b; Zoghbi, 2003). However, this idea is being challenged by a recent study indicating the lack of MeCP2 specifically in glial cells contributes to RTT phenotypic neurons by an unknown secreted glial factor (Ballas *et al.*, 2009). This discrepancy illustrates the need for more unbiased approaches in determining the molecular role of MeCP2 in RTT; thus, intact mammalian brain tissue would be the ideal source to study endogenous MeCP2 protein complexes. Here we use the power of

biochemistry to characterize MeCP2 in the mammalian brain and show that native MeCP2 protein purified from adult rat brain exists in multiple biochemically distinct pools, consistent with MeCP2 working as a multi-functional protein. We further characterize one brain-derived MeCP2-complex that contains Prpf3, a known spliceosome-associated protein (Wang *et al.*, 1997), as well as the Sdccag1 (Scanlan *et al.*, 1998), a mediator of nuclear export (Bi *et al.*, 2005). MeCP2 shows specific, direct interactions with Prpf3 and Sdccag1 and these interactions are disrupted by certain RTT mutations. In addition, we show that MeCP2 and Prpf3 co-associate *in vivo* with mRNAs from genes activated by MeCP2, further supporting the previously identified regulatory role of MeCP2 in mRNA biogenesis (Young *et al.*, 2005) and providing another potential mechanism disrupted during pathogenesis of RTT.

RESULTS

MeCP2 exists in at least four distinct protein pools in rat brain nuclei

A large-scale biochemical purification of endogenous MeCP2 protein from rat brain nuclear extract was performed to characterize the native MeCP2 in the mammalian brain. Whole rat brains were homogenized under non-denaturing conditions, the intact nuclei were purified by centrifugation through a sucrose cushion, and nuclear proteins were extracted under mild ionic conditions. Proteins were initially fractionated by strong anion exchange (MonoQ) column chromatography (Figure 2.1A**)** with the MeCP2 protein being tracked by western blotting with multiple MeCP2-specific antibodies (Figures 2.2), all providing virtually identical profiles.

The majority of soluble MeCP2 protein did not bind the MonoQ resin (Figure 2.1A, lane QF). To determine possibility of proteins associating with MeCP2 in the MonoQ flow through (QF MeCP2), complexity was reduced by multiple steps of liquid chromatography fractionation consisting of MonoS strong cation exchange chromatography, heparin affinity chromatography, and gel filtration through Superose6 (Figure 2.3). After gel filtration, western blotting and silver stain analysis indicates MeCP2, confirmed by spectrometry (Figure 2.4), does not co-fractionate with any other polypeptides (Figure 2.5), thus confirming a previous report (Klose and Bird, 2004).

A significant fraction (10%) of the brain-derived MeCP2 consistently interacted with the anion exchange resin, consistently eluting in two distinct peaks along a linear salt gradient, indicating multiple biochemically distinct pools of MeCP2 (Figure 2.1A). The MonoQ bound (QB) MeCP2 elution profile showed an initial broad peak (QB1/2) concentrated at 230mM NaCl and tailing to 370mM, suggesting multiple MeCP2 forms, with a minor yet distinct peak (QB3) centered at 450mM NaCl (Figure 2.1A). Since MeCP2 is a DNA binding protein and also known to interact with RNA, the potential of nucleic acids mediating the anionic association of MeCP2 with the cationic resin was addressed. The QB fractions were treated with benzonase nuclease to remove all DNA and RNA and the MonoQ chromatography was repeated. In each case, all of the nuclease-treated QB MeCP2 again bound the MonoQ resin and was released by step elution, indicating the interaction was both RNA and DNA independent (Figure 2.6A). Futhermore, MeCP2 is known to be phosphorylated (Tao *et al.*, 2009; Zhou *et al.*, 2006). To determine if phosphorylation of MeCP2 is responsible for the interaction with the MonoQ resin, QB fractions were treated with calf intestinal phosphotase (CIP) and the

MonoQ chromatography was repeated. All of the CIP-treated QB MeCP2 again bound the MonoQ resin and was released by step elution, indicating the interaction was not due to phosphorylation of MeCP2 (Figure 2.6B). Finally, the presence of MeCP2 in the MonoQ elution peak (QB1/2 and QB3) was confirmed by mass spectrometry (Figure 2.4).

The more abundant MonoQ bound pool of MeCP2 (QB1/2) was characterized for complexity and content (Figure 2.1A). The QB1/2 peak fractions were pooled and fractionated over the MonoS strong cation exchange resin, resolving into two peaks of MeCP2 protein (Figure 2.1B), the first eluting at 400mM NaCl (QB1) and the second eluting at 550mM NaCl (QB2), therefore indicating at least three biochemically distinct pools of MeCP2 exist in the initial QB fraction and four MeCP2 pools exist overall in rat brain nuclear extract (QF and QB1-3).

Brain-derived MeCP2 exists in a putative RNA processing complex

In order to identify potential MeCP2-interacting proteins, the more abundant MonoS bound pool of MeCP2 (QB2) was purified further by fractionation using heparin affinity chromatography and gel filtration through Superose6 (Figure 2.7). Western blotting and silver stain analysis of the final fractionation showed MeCP2 peaking with the apparent molecular weight of 600-700kDa and precisely co-purifying with six additional polypeptides (Figures 2.8 and 2.9). Therefore, all six MeCP2 co-purifying polypeptides were identified by mass spectrometry with significant coverage (Figure 2.10) as: Prpf3, Sdccag1, ATP-binding cassette 50, and 3 components of a translation initiation complex (Eif2 subunits 1, 2, and 3). Since a commercial antibody was available against Prpf3, western blot analysis of the size exclusion chromatography fractions were

carried out and show MeCP2 and Prpf3 precisely co-fractionate (Figure 2.8, bottom) confirming the silver-stain analysis (Figure 2.8 and 2.9).

With Prpf3 identified as a putative MeCP2-interacting protein, the MeCP2 fractionation scheme (Figure 2.7) was analyzed for Prpf3 by western blotting to determine if Prpf3 was present in all MeCP2 pools or specific to QB2 and similarly to determine if all Prpf3 in brain extracts associated with MeCP2. Not surprisingly, only a fraction of Prpf3 in brain extracts co-fractionated with MeCP2 when assaying the MonoQ separation profile (Figure 2.11A), however, all of the detectable Prpf3 overlapping with MeCP2 specifically co-fractionated with the QB2 pool of MeCP2 from the MonoS fractionation and not with the QB1 MeCP2 pool (Figure 2.11B). Thus, Prpf3 distinguished QB2 from other MeCP2 protein pools. Because Prpf3 is a known component of the spliceosome (Wang *et al.*, 1997) and MeCP2 has been shown to interact with another spliceosome-associated protein, YB-1 (Young *et al.*, 2005), the QB2 fractionation was screened for YB-1. However, YB-1 was not found by western blotting and YB-1 was absent from the mass spectrometry analysis, indicating YB-1 did not co-purify with QB2 MeCP2 under these conditions (data not shown).

To confirm that native MeCP2 and Prpf3 were in a complex *in vivo*, co-IP experiments were performed (Figure 2.12). In brain extracts, as determined by the biochemical fractionations described above, the vast majority of MeCP2 is not associated with Prpf3 and likewise, the majority of Prpf3 in brain is not associated with MeCP2 making it difficult to visualize any interaction by co-IP directly from crude nuclear extracts. Thus, two approaches were used; co-IPs from fractionated brain nuclear extract (Figure 2.12A) and co-IPs from crude extracts using tissue culture cells overexpressing

an epitope tagged version of MeCP2 (Figure 2.12B). Rat brain nuclear extract was prepared and fractionated over monoQ and monoS resins as described above for QB2. The peak MeCP2 and Prpf3 containing fractions were made to 100 mM NaCl in buffer A and used for co-IP experiments assayed by western blotting. Using the anti-MeCP2 antibody to IP, Prpf3 was specifically co-IP'ed from these fractions and the interaction between MeCP2 and Prpf3 was not dependent on nucleic acids (Figure 2.12A, lane 3 and 4). Alternatively, to show co-IP interaction between MeCP2 and Prpf3 from crude cell extracts, the murine hippocampal HT-22 cell line was used to generate a pool of cells stably expressing a HA-tagged version of human MeCP2. Immunofluorescence showed that HA-MeCP2 in these cell lines predominantly concentrated to DAPI-rich heterochromatic foci as expected for endogenous MeCP2 (Figure 2.13). Anti-HA antibodies were then used to specifically co-IP HA-MeCP2 and MeCP2-associated proteins from these cell lines. Western blotting showed that HA-MeCP2 specifically co-IP'ed a fraction of the endogenous mouse Prpf3 (Figure 2.12B, lane 3), indicating that Prpf3 and MeCP2 exist in a stable complex within these neuronal cells, and the interaction between MeCP2 and Prpf3 was not dependent on nucleic acids (Figure 2.12B, lane 4). We conclude that MeCP2 and Prpf3 exist in a native complex *in vivo*.

MeCP2 interacts directly with Prpf3 and Sdccag1 and independent of nucleic acids

To investigate which of the co-purifying candidate proteins directly interacted with MeCP2, the cDNAs for each identified protein was cloned and the recombinant proteins were tested for their ability to interact with MeCP2 *in vitro* (Figure 2.14). GST pull-down assays using a GST-MeCP2 fusion protein showed that rat Prpf3 specifically

interacted with full-length rat MeCP2, but not GST alone (Figure 2.14A, left panel), and reciprocally, rat GST-Prpf3 interacted with rat MeCP2 (Figure 2.14B). Furthermore, treatment of the GST-MeCP2/Prpf3 reaction with benzonase nuclease showed the interaction was not dependent on nucleic acids (Figure 2.14A, B). Similarly, rat Sdccag1 specifically interacted with rat GST-MeCP2 in a nucleic acid independent manner (Figure 2.14A, middle panel). Interestingly, neither rat GST-MeCP2 nor rat GST-Prpf3 interacted with any of the other four identified polypeptides supporting the specificity of the observed interactions of MeCP2, Prpf3 and Sdccag1 (Figure 2.14A, right panel and data not shown).

Prpf3 interacts with the MeCP2 MBD and transcriptional repression domain (TRD) and Sdccag1 interacts with the MeCP2 carboxyl terminal region

The regions of MeCP2 that mediate the direct interaction with Prpf3 were mapped by GST pull-down. A series of bacterially generated human GST-MeCP2 deletion mutants were tested for their ability to interact with *in vitro* synthesized human Prpf3 (Figures 2.15A and 2.16). Prpf3 retained the ability to interact with MeCP2 N-terminal deletions up through amino acid residue 195 as well as a C-terminal deletion lacking amino acid residues 309 – 486. However, Prpf3 did not interact with a MeCP2 deletion lacking amino acid residues 1 – 308, indicating that the TRD region is required for interaction (Figure 2.15A, upper panel and 2.16). Therefore, a fragment of MeCP2 that contains only amino acid residues 207-308, corresponding to the TRD domain, was tested and found to interact with Prpf3 with similar apparent affinity as full length MeCP2. Similarly, rat Sdccag1 was synthesized *in vitro* and subjected to the same series of GST-MeCP2 deletions as Prpf3 (Figure 2.15A, lower panel and 2.16). Sdccag1 was able to

interact with all of the N-terminal MeCP2 deletion proteins tested, but failed to interact with a MeCP2 deletion lacking the region C-terminal to the TRD. Thus, MeCP2's interaction domain with Sdccag1 resides between amino acid residues 309–486, adjacent with the Prpf3 interaction domain.

Considering most known RTT mutations reside in *MECP2*, it is likely that some of these mutations would also disrupt the interactions of Prpf3 and/or Sdccag1 with MeCP2. A series of bacterially generated human GST-MeCP2 RTT nonsense mutants were produced and tested for their ability to interact with *in vitro* synthesized human Prpf3 and rat Sdccag1 (Figure 2.15B and 2.16). The series of RTT mutations truncate MeCP2 ranging from amino acids 49 to 204. All of these RTT nonsense mutants lack the C-terminal region of MeCP2 and therefore disrupted its interaction with Sdccag1 as expected (Fig 15B, bottom panel and 16). The Prpf3 interaction was disrupted by RTT truncations at amino acids S49X, S68X and W104X as well. However, the Prpf3 interaction with MeCP2 was maintained with the RTT truncations Y141X, R168X, and S204X (Fig 15B, top panel and 16). This indicates there are two MeCP2 regions capable of interacting with Prpf3 flanking the inter-domain region between the MBD and TRD of MeCP2. Interestingly, this same inter-domain region of MeCP2 has been previously characterized as the RG domain and is required for MeCP2's RNA binding activity (Jeffery and Nakielny, 2004). Several common RTT point mutants, as well as additional RTT truncations, were similarly tested and found to have no impact on MeCP2's ability to interact with Prpf3 (Figure 2.15C). We conclude that MeCP2 contains two domains sufficient for interaction with Prpf3, one in the MBD between amino acids 104 and 141, and the second in the TRD between amino acids 207 and 294 and one domain for

MeCP2's interaction with Sdccag1, residing between amino acids 311 and 486 (Figure 2.15D). Therefore, any RTT mutation truncating MeCP2 at or before amino acid residue 104 will abolish the Prpf3 interaction, while any RTT truncation at or before amino acid residue 297 would disrupt the Sdccag1 interaction with MeCP2, all of which could affect MeCP2's role in RNA biogenesis.

MeCP2 interacts with specific mRNAs *in vivo*

Several lines of evidence implicate MeCP2 is associated with mRNA biogenesis *in vivo*; MeCP2 binds RNA *in vitro* (Jeffery and Nakielny, 2004), is part of a ribonucleoprotein complex with YB-1 in cell culture (Young *et al.*, 2005), MeCP2 knock-out mice show misspliced transcripts in the brain (Young *et al.*, 2005), it has been identified as a transcriptional activator at the majority of gene promoters it has been shown to regulate (Chahrour *et al.*, 2008), and here MeCP2 is shown to interact with the splicing factor Prpf3. To determine if MeCP2 interacts with the RNA transcripts of genes it regulates *in vivo*, RNA immunoprecipitations (RIP) were performed. Using anti-HA antibodies on HT-22 (HA-MeCP2) cell lysates, mRNAs for the MeCP2 regulated gene *Cdk10* were able to be specifically RIP'ed (Figure 2.17A). Assaying the IP's by RT-PCRs using oligonucleotide primers designed to amplify across exon junctions of *Cdk10* show MeCP2 is associated with the spliced form of the gene, while MeCP2 is not associated with the *Casc3* gene transcript (Figure 2.17A). The RIPs were RNase sensitive confirming that RNA and not DNA was IP'ed. Interestingly, RT-PCR for RIPs using primers for *Cdk10* show what appears to be pre-mRNA by size and RNase sensitivity; however, unlike the spliced forms these pre-mRNA products, while generally

consistent, were not observed in 100% of the RT-PCR analyses for *Cdk10* (Figure 2.17A top).

To investigate if Prpf3 was part of the HA-MeCP2/mRNA complex, a RIP and Re-RIP approach was implemented. Anti-HA antibodies were used to RIP HA-MeCP2/mRNA complexes from HT-22 (HA-MeCP2) cell lysates, the bound complexes were eluted from the HA antibodies intact using DTT to disrupt the IgG structure, and then anti-Prpf3 antibodies were used to Re-RIP. Therefore, any RNAs present in the Re-RIP must have been associated with both MeCP2 and Prpf3. RT-PCR analysis showed that RIP for HA-MeCP2 and subsequent Re-RIP for Prpf3 protein IP-ed mRNA for *Cdk10*, indicating that both proteins are in fact associated with this target mRNA (Figure 2.17B), supporting their being in a complex *in vivo*.

To further support MeCP2 RIP results, experiments were preformed to determine if MeCP2 interacts with transcriptionally active *Xenopus* lampbrush chromosome (LBC) spreads. Actively transcribed LBC loops from *Xenopus* oocytes are surrounded by an observable ribonucleoprotein (RNP) matrix, composed of RNA polymerase II and elongating transcripts packed together with processing factors that give rise to RNP fibrils (Austin *et al.*, 2009; Gall *et al.*, 1999) (Figure 2.18 - Phase Contrast). Staining of the lampbrush chromosomes with xNF7, a factor known to be associated with actively transcribed LBC loops (Beenders *et al.*, 2007), indicates active transcription on one isolated loop (Figure 2.18 - XNF7, white arrow). Microinjection of HA tagged human MeCP2 (hMeCP2-HA) mRNA into oocytes, and subsequent LBC spread shows that MeCP2 interacts with both the highly condensed axis (Figure 2.18 - MeCP2) and actively transcribed loops (Figure 2.18 - MeCP2, white arrow) by immunofluorescence.

Immunofluorescent staining of uninjected controls show antibodies are specifically staining hMeCP2-HA (Figure 2.19). This data further suggests a role for MeCP2 in mRNA processing.

DISCUSSION

MeCP2 has been characterized as a multifunctional protein using a variety of techniques and from numerous cellular contexts, however only the biochemical characteristics of the endogenous MeCP2 protein in the brain are relevant to RTT. Structural studies indicate that MeCP2 contains at least six structurally distinct domains (Adams *et al.*, 2007). Mutations generally manifested as point mutations resulting in a single missense or nonsense amino acid change have been identified in all domains of *MECP2* from RTT patients (www.rettsyndrome.org/), suggesting that all domains are critical for MeCP2 function (Hite *et al.*, 2009; Moore *et al.*, 2005). Although all these mutations produce clinical RTT pathology, certain mutations are more strongly associated with particular symptoms and disease severity suggesting that all RTT mutations in *MECP2* are not equal, with some potentially being more disruptive towards MeCP2's many functions (Jian *et al.*, 2005; Neul *et al.*, 2008). However, the underlying mechanisms of how these myriad mutations lead to RTT pathophysiology remain unclear due to a lack of understanding toward the complete scope of MeCP2's normal function in the brain. Here, we begin to reveal molecular mechanisms of MeCP2 function relevant to RTT through the identification of four distinct brain-derived MeCP2 protein pools, agreeing with the functional data proposing multiple roles for MeCP2 (Hite *et al.*, 2009).

Characterization of one of these complexes as containing MeCP2, Pprf3, Sdccag1, and mRNA strongly complements previously published data that MeCP2 is involved in mRNA splicing, and this activity is disrupted by certain RTT mutations (Young *et al.*, 2005).

We have found MeCP2 can associate with RNA by RIP and on transcriptionally active loops in *Xenopus* LBC spreads. MeCP2's association with Prpf3, a major component of the spliceosome, supports MeCP2 as having a role in modulating mRNA splicing; exactly what that role might be is still not clear. Interestingly, a recent study suggests that differential gene body methylation may play a role in transcript splicing (Laurent *et al.*). By identifying the genome wide methylation status in multiple human cell types, the group found that exons were more highly methylated than introns, and that there were sharp transitions of methylation at exon-intron boundaries (Laurent *et al.*). Based on what is known about MeCP2 and Prpf3 independently, potentially MeCP2's direct interaction with Prpf3 could function in splice site selection. Prpf3 is one of multiple associated proteins of the U4/U6.U5 tri-snRNP complex that is recruited to the splice site to form and stabilize the functional spliceosome, and is essential for pre-mRNA splicing (Brown and Beggs, 1992; Horowitz *et al.*, 1997; Lauber *et al.*, 1997; Wang *et al.*, 1997). Importantly, MeCP2 can bind RNA directly (Jeffery and Nakielny, 2004), and has been implicated as a regulator of alternative splicing in the HeLa and Neuro2A cell lines through an RNA-dependent interaction with YB-1 (Young *et al.*, 2005), a protein known to participate in splicing of mRNAs (Philips *et al.*, 1998; Stickeler *et al.*, 2001). While we were unable to detect YB-1 as a component of the MeCP2/Prpf3/Sdccag1 complex in brain, it is inefficient to inhibit the highly active and

stable RNases that would abolish any YB-1/RNA/MeCP2 interaction during purification so we cannot rule out that YB-1 could be part of a larger RNA-dependent alternative splicing complex. Nevertheless, a direct interaction of MeCP2 and Prpf3 further supports a role for MeCP2 in mRNA splicing regulation.

Sdccag1 was originally identified from colon cancer patients by a serological analysis of recombinant cDNA expression libraries (SEREX) (Scanlan *et al.*, 1998) and later identified as a tumor suppressor by its ability to cause cell cycle arrest in a non-small-cell lung cancer cell line (Carbonnelle *et al.*, 2001). The significance of finding Sdccag1 as part of the MeCP2-Prpf3 complex can only be implied due to the limited information on its function in vertebrates and *Drosophila*. Bioinformatically the Sdccag1 protein contains a predicted RNA-binding domain homologous to a eukaryotic small nuclear ribonucleoprotein (Marchler-Bauer *et al.*, 2009), suggesting the capacity to function in mRNA splicing. Functionally, the *Drosophila* Sdccag1 homolog Caliban, has been shown to interact with and mediate the nuclear export of the Prospero homeodomain transcription factor (Prox in mammals) and this interaction and function is conserved in mammalian cells (Bi *et al.*, 2005). This raises the possibility that the MeCP2/Prpf3/Sdccag1 complex may not only be involved in splicing mRNA, but also transporting mRNAs. Consistent with this model, MeCP2 has been shown to have both a nuclear and cytoplasmic localization in neuronal cell lines (Miyake and Nagai, 2007). Notably, *Mecp2*$^{308/Y}$ mice, which produce a truncated form of MeCP2 and reproduce many of the classical features of RTT (Moretti *et al.*, 2005; Shahbazian *et al.*, 2002a), have been shown to have multiple genes that are abnormally spliced in the brain (Young *et al.*, 2005). This suggests the C-terminal portion of MeCP2, which we have identified

as the putative Sdccag1 interaction domain, plays a critical role in regulating alternative splicing.

MeCP2 is present at the promoter of many actively transcribed genes (Yasui *et al.*, 2007), and activates a majority of genes it regulates in the mouse hippocampus (Chahrour *et al.*, 2008). MeCP2's ability to bind chromatin *in vivo* appears to be dynamically regulated by phosphorylation in the brain (Chen *et al.*, 2003; Tao *et al.*, 2009). Experimentally, MeCP2 is known to bind RNA *in vitro* with a similar affinity as to methylated DNA and that the two activities are mutually exclusive (Jeffery and Nakielny, 2004). Thus, MeCP2 at an active promoter or within the body of a transcribed gene would be well positioned spatially upon its release to interact with the transcripts of the genes it activates and influence splice site selection. Therefore, it is reasonable to propose that MeCP2 activated genes would also be targets for splicing regulation by a MeCP2 containing complex. Supporting this model, RNA transcripts from *Cdk10*, a gene positively regulated by MeCP2 binding at its promoter (Chahrour *et al.*, 2008) and abnormally spliced in *Mecp2*$^{308/Y}$ mice (Young *et al.*, 2005), were associated with both MeCP2 and Prpf3 *in vivo*.

Considering the range of mutations throughout *MECP2* in RTT, it is not surprising that many RTT mutations are within, or predicted to affect both the Prpf3 and Sdccag1 binding domains, disrupting the MeCP2/Prpf3/Sdccag1 complex and presumably MeCP2-mediated splicing regulation and mRNA transport. With MeCP2 emerging as a multifunctional protein and its biological role in RTT unclear, identifying biochemically distinct pools of Mecp2 in the brain containing novel MeCP2 interacting proteins is a valuable tool towards an understanding MeCP2 function and its dysfunction

in RTT. This study adds to the mounting evidence indicating that one such critical function of MeCP2 in the brain involves RNA biogenesis.

MATERIALS AND METHODS

Antibodies and western blot analysis

Protein samples were separated by SDS-PAGE and transferred to ECL nylon membrane (GE Healthcare) for western blotting by standard methods. For each experiment, the western blotting images presented are from the same exposure on the same piece of film, linearly adjusted for brightness in Adobe Photoshop. The anti-MeCP2 7-18 antibody is a rabbit polyclonal derived from bacterially expressed recombinant protein encoding amino acid residues 310 to 388 of human MeCP2e2 isoform. The anti-MeCP2 3998 antibody is a rabbit polyclonal derived from bacterially expressed recombinant protein encoding the full length human MeCP2e2 isoform. Antibodies were used at the following concentrations: anti-MeCP2 7-18, 1:2000; anti-MeCP2 3998 , 1:1000; anti-MeCP2 (Upstate 07-013) 1:1000; Anti-PRP3 (MBL D171-3) 1:2000; and Anti-HA High Affinity (Roche 14559100) at 1:1000.

Molecular biology and generation of plasmid constructs

RNA was purified from rat brain tissue or HeLa cells using Trizol Reagent (Invitrogen) per manufacturers' instructions. All cDNAs were generated using SuperScript III One-Step RT-PCR with Platinum Taq (Invitrogen). All PCRs were performed with Phusion polymerase (New England Biolabs) and cloned into pGEM-T

easy (Promega) for sequencing prior to sub-cloning into pGEX-5X1 (GE Life sciences), pCDNA 3.1 (Invitrogen), or the pCDNA 3.1 HA vector (Matzat *et al.*, 2008). All primers are listed in Table 1. All human MeCP2 constructs were generated from the human *MECP2E2* cDNA (NM_004992). To generate constructs for *in vitro* synthesized proteins, cDNAs were PCR amplified using primers listed in Table 1 and sub-cloned into the specified restriction sites of pCDNA 3.1. Rat cDNAs were sub-cloned between NotI and XhoI; the human Prpf3 cDNA clone was sub-cloned between EcoRI and XhoI; the human full-length MeCP2 was amplified from full-length cDNA and sub-cloned between NotI and XhoI. Constructs for bacterially generated GST fusion proteins were PCR amplified from full-length cDNAs, sub-cloned into the specified restriction sites of pGEX5-X1 using primers listed in Table 1. Rat GST-Prpf3 was sub-cloned between BamHI and XhoI. The rat and human GST-MeCP2 full-length, GST-MeCP2 deletions, and human GST-MeCP2 RTT cDNAs were sub-cloned between EcoRI and XhoI. The pCDNA3P HA-MeCP2 vector was created by PCR amplification of full-length human MeCP2, digestion and sub-cloning into the NotI and XhoI restriction sites of the pCDNA3.1 HA vector and subsequent digestion with NdeI and XhoI for cloning of the HA-MeCP2 fragment into the pCDNA3P puromycin vector.

Cell culture and immunofluorescent staining

HT-22 cells (a gift from Dr. Stephanie Ceman) were transfected with pCDNA 3.1P HA-MeCP2 using Fugene HD transfection reagent (Roche). Stable integrants were selected as pools in puromycin (1μg/ml), and maintained in Dulbecco's modified Eagle's medium (Biowhittaker) supplemented with 10% fetal bovine serum, glutamine,

antibiotics and puromycin. Immunofluorescent staining of stably transfected HA-MeCP2 Ht-22 cells was preformed essentially as previously described (Levesque *et al.*, 2001). In brief, cells grown on poly-d-lysine coated coverslips were washed in PBS, fixed with 4% formaldehyde and 4% glucose in PBS for 10 min at RT. Coverslips were washed 3 times in PBS and permeabilized with 0.2% TritonX-100 in PBS for 20 min at RT. Cells were then washed 3 times in PBS with 0.1% tween-20 and subsequently incubated in Block (2% BSA, 2% FBS, 0.1% Tween-20 in 1X PBS) for 1 hour at RT. Overnight incubation of coverslips in primary antibody (anti-HA High Affinity (Roche 14559100), 1:200 in Block) were preformed in a humidified box at 4°C. Cells were washed 3 times in PBS with 0.1% tween-20 and incubated with anti-RAT rhodamine (Jackson Labs) for 1 hour at RT. Coverslips were washed 3 times in PBS with 0.1% tween-20 and stained with DAPI, mounted.

Purification of endogenous MeCP2

Rat brain nuclei were isolated generally as described (Lewis *et al.*, 1992). Adult rat brains (n=200 total per preparation) were obtained (Pel-Freez Biologicals), thawed on ice, and homogenized with 14 strokes of a loose pestle in a 40ml Dounce homogenizer (Wheaton) in ice-cold HB (20mM Hepes pH 7.9, 25mM KCl, 0.5mM EGTA, 1mM EDTA, 2M sucrose, 10% glycerol, freshly added 0.5mM spermidine, 0.15mM spermine, 1μg/ml leupeptin, 1μg/ml pepstatin, 1μg/ml aprotinin). The homogenate (26ml/tube) was layered onto an 8ml cushion of HB and centrifuged in a SW28 rotor at 24,000 rpm for 1 hr at 4°C. Pelleted nuclei were resuspended in Buffer A (20mM Hepes pH 7.5, 1.5mM $MgCl_2$, 1mM EGTA, 10% glycerol, freshly added 0.5mM DTT, 1 μg/ml leupeptin,

1μg/ml pepstatin, 1μg/ml aprotinin) supplemented with 350mM NaCl (A-350) and extracted with rotation for 30 min at 4°C. Insoluble material was removed by centrifugation at 200,000 x g for 20min at 4°C. The supernatant was diluted with Buffer A to reduce the NaCl concentration below 200mM NaCl and used as a soluble protein source for chromatography.

All chromatography was preformed using an AKTA-FPLC (GE Healthcare) and FPLC columns (GE Healthcare) at 4°C in Buffer A with indicated concentrations of NaCl. Soluble protein (130 mg) was fractionated over a MonoQ10/10 column with bound protein eluted by a 20 column volume (cv) linear salt gradient from 100mM to 1000mM NaCl, collecting 1 cv fractions. Soluble protein (29 mg) from MonoQ flow through was fractionated over MonoS5/5 with bound protein eluted by a 20 column volume (cv) linear salt gradient from 100mM to 1000mM NaCl, collecting 1 cv fractions. MeCP2 containing fractions centered at 450 mM NaCl (QF) were combined and loaded onto a 1ml Heparin FastFlow column and eluted with a 10 cv linear gradient from 450mM to 1000mM, collecting 1 cv fractions. Pooled heparin column fractions containing MeCP2 were applied to a 110 ml superose 6 column and fractionated in buffer A-150 with 0.1% Trition-X100. Subsequently, 0.5 ml of each fraction was trichloriacetic acid-precipitated, subjected to SDS-PAGE, and used for Western blotting or silver staining (Sigma Aldrich – PROTSIL1-1KT). The silver-stained polypeptide corresponding by size to MeCP2 (and by western) was excised and analyzed by mass spectrometry. The MeCP2 containing fractions (QB1/2) peaking at 230mM NaCl were pooled (1.56 mg), diluted with buffer A and fractionated over MonoS5/5 with bound protein eluted by a 20 column volume (cv) linear salt gradient from 250mM to 1000mM NaCl, collecting 1 cv fractions.

The MeCP2 containing fractions peaking at 550 mM NaCl (QB2) were combined and loaded onto a 1ml Heparin FastFlow column and eluted with a 10 cv linear gradient from 450mM to 1000mM, collecting 1 cv fractions. Pooled heparin column fractions containing MeCP2 were applied to a 110 ml superose 6 column and fractionated in buffer A-150 with 0.1% Trition-X100. Subsequently, 0.5 ml of each fraction was trichloriacetic acid-precipitated, subjected to SDS-PAGE, and used for Western blotting or silver staining (Sigma Aldrich – PROTSIL1-1KT). Silver-stained polypeptide bands precisely cofractionating with MeCP2 (by western) were excised and analyzed by mass spectrometry.

For the nuclease treatment experiments, MonoQ eluted MeCP2 was dialyzed against buffer A-100 containing 1mM $CaCl_2$, 10mM $MgCl_2$ and without EGTA then either treated with or without 500 units of Benzonase nuclease (Sigma) for 30 min. at 37°C, and re-tested for ability to bind the MonoQ resin.

For the CIP treatment experiment, MonoQ eluted MeCP2 was dialyzed against buffer A-100 containing 1mM $CaCl_2$, 10mM $MgCl_2$ and without EGTA then either treated with 60 units of CIP (NEB) for 30 min. at 37°C, and re-tested for ability to bind the MonoQ resin.

Mass spectrometry

All mass spectrometry was carried out at the Protein Sciences Facility at the University of Illinois. FPLC purified fractions were separated by SDS-PAGE, visualized by mass spectrometry compatible silver staining and polypeptide bands were excised. Gel slices were destained (50% acetonitrile, 25 mM ammonium bicarbonate) then

digested in 25 µl of Sequencing Grade Trypsin (12.5 nanogram/microliter in 25 mM ammonium bicarbonate, G-Biosciences St. Louis, MO) using a CEM Discover Microwave Digestor (Mathews, NC) for 15 min at 55°C (60W). Digested peptides were extracted using 50% acetonitrile with 5% formic acid, dried in a Savant SpeedVac and suspended in 13 µl of 5% acetonitrile containing 0.1% formic acid with10 µl of sample used for mass spec analysis. The mass spectrometer used was Waters quadrupole time-of-flight mass spectrometer (Q-ToF) connected to a Waters nano-Acquity UPLC. The column used was Waters Atlantis C-18 (0.075 mm x 150 mm) with a flow rate of 250 nanoliters per minute. Peptides were eluted using a linear gradient of water/acetonitrile containing 0.1% formic acid (0-60% B) in 60 minutes. The mass spectrometer was set for data dependent acquisition, ms/ms was performed on the most abundant four peaks at any given time. Data analysis was performed using Waters Protein Lynx Global Server 2.2.5, Mascot (Matrix Sciences) and BLAST against NCBI NR database.

GST pull-down assay

Recombinant GST-fusion proteins were generated in *E. coli* BL-21(DE3) cells (Stratagene) according to manufacture's (GE Healthcare) protocol. *In vitro* transcribed/translated proteins radiolabeled with [^{35}S]methionine were produced using the T7 TnT Quick Coupled Transcription/Translation System (Promega). GST pull-down assays were preformed by incubation of Glutathione Sepharose 4B (GE Healthcare) bound GST-fusion proteins and *in vitro* synthesized candidate proteins in 1X PBS with 1mM DTT, 0.1% Tween-20, 1mM PMSF, 1µg/ml leupeptin, 1µg/ml pepstatin, and 1µg/ml aprotinin. Reactions were rotated overnight at 4°C, with or without 250 units of

Benzonease nuclease, washed 6 times in 1X PBS with 0.1% NP-40 with interacting [^{35}S]methionine labeled proteins separated by SDS-PAGE and visualized by autoradiography.

Co-immunoprecipitation (Co-IP)

Co-IPs from fractionated brain extracts were carried out as follows: Protein A Dynabeads (Invitrogen) were blocked with BSA and incubated with either anti-MeCP2 3998 or IgG in fractionated material, diluted to 100mM NaCl with Buffer A0, for 2 hrs at 4°C with rotation. IPs were washed three times with buffer A225, eluted with Laemmli buffer, separated by SDS-PAGE and subjected to Western blot analysis. The anti-MeCP2 3998 antibody is a rabbit polyclonal derived from bacterially expressed recombinant protein encoding the full length human MeCP2e2 isoform.

Co-IPs were carried out essentially as previously described (Young *et al.*, 2005). HT-22 cells (8 x 10^6) stably expressing HA-MeCP2 were lysed in 1ml of IPH buffer (50 mM Tris-HCl, pH 8.0, 150mM NaCl, 5mM EDTA, 0.5% NP-40, 1μg/ml leupeptin, 1μg/ml pepstatin, 1μg/ml aprotinin) at 4°C for 30min and debris removed by centrifugation for 15min at 16,000 x g at 4 °C. Lysates were precleared with protein A agarose then incubated for 4 hrs with anti-HA (Sigma, E6779) or irrelevant IgG with protein A agarose beads (Santa Cruz Biotechnologies) at 4°C with rotation. Nuclease treatment was preformed by re-suspension of protein bound beads with 125 units of Benzonase for 10 min at 37 °C. Precipitates were washed three times with IPH buffer, eluted with Laemmli buffer, separated by SDS-PAGE and subjected to Western blot analysis.

RNA immunopreciptation

RNA IP was preformed as previously described with modifications (Lin *et al.*, 2005). HT-22 cells (1 x 10^8) stably expressing HA tagged MeCP2 were collected and washed and resuspended in 4ml 1X PBS. 0.4 ml of crosslinking buffer (100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 50 mM HEPES, 11% formaldehyde) was added to resuspended cells and incubated at RT for 30 min with rocking. Quench with 0.22 ml of 2.5 M glycine pH 7.0 for 5 min at RT. Centrifuge at 1367 x g for 5 min at 4 °C to pellet and wash with 1X PBS. Lyse cells in 1 ml FA buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 1 μg/ml leupeptin, 1μg/ml pepstatin, 1μg/ml aprotinin, 100 units/ml RNasin [Promega]) by sonication (25% power, 50% duty) for 10 pulses of 30 sec with rest on ice. The lysed cells were treated with DNase I (Promega) by adding 25 mM $MgCl_2$, 5 mM $CaCl_2$, 6 μl RNasin and 200 units DNase I for 30 min at 37 °C. Extracts were cleared by centrifugation 15 min at 14,000 RPM in an Eppendorf 5415C centrifuge at 4 °C. 100 μl of cleared extract are diluted with 900 μl of ChIP dilution buffer (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40). Diluted extracts were incubated with either HA antibody or non-specific IgG with 4 μl RNasin, for 12 hours, rotating at 4 °C. 40 μl of ChIP buffer equilibrated protein A or G Dynabeads (Invitrogen) were then added for 1 hour rotating at 4 °C, then washed 3 x 10 min with wash buffer (50 mM Tris pH 7.4, 500 mM NaCl, 1% Triton X-100, 0.1% SDS, 100 units/ml RNasin). For Re-RIP experiments, beads were washed and bound immune-RNA complexes were released in 20mM dithiothreitol solution for 30 min at 37°C and resuspended in one volume of ChIP dilution buffer for ReRIP with antibody for PRPF3 and washed as before.

After final wash, beads were brought up in 200 µl of elution buffer (200 mM NaCl, 50 mM Tris pH 7.4, 20 µg Proteinase K) and digested for 1 hour at 42°C, and then cross-links reversed for 5 hours at 65°C. Samples were extracted with acid equilibrated (pH 4.8) phenol:chloroform (5:1) and ethanol precipitated. Precipitated material was resuspended in 50 µl DEPC-H_2O and 1 µl used for RT-PCR analysis with SuperScript III One-Step RT-PCR with Platinum Taq (Invitrogen) according to manufactures protocol. RT-PCR primers for mRNA targets Cdk10, Casc3 and Frg1 are listed in Table 1. Results were analyzed by electrophoresis on a 2% agarose gel.

Xenopus LBC spreads and immunofluorescent staining

Female adult frogs (*Xenopus laevis*) were anesthetized in 0.15% tricaine methanesulfonate (Sigma), and small fragments of ovary were surgically removed. Non-defolliculated oocytes were used for isolation of nuclei, and the nuclear spreads were performed as previously described (Patel *et al.*, 2008; Patel *et al.*, 2007). The samples were fixed with 2% paraformaldehyde in PBS plus 1 mM $MgCl_2$ for 1 hr at room temperature. After fixation, the nuclear spreads were rinsed in PBS and blocked with 0.5% bovine serum albumin (Sigma-Aldrich) plus 0.5% gelatin (from cold-water fish) in PBS for 10 min. Spreads were incubated with primary antibody, anti-HA antibody MAb 3F10 (Roche, Mannheim, Germany), used at a concentration of 20 ng/ µl, for 1 h at RT, washed for 30 min with two changes of PBS, incubated in secondary antibody, Alexa 488-labeled-goat anti-mouse IgG (Invitrogen, Carlsbad, CA), at a concentration of 2.5 µg/ml and incubated for 1 hr at RT, and washed again for 30 min with two changes of

PBS. Spreads were incubated with 1 μM Syto61 (Invitrogen) in PBS for 20 min at RT and briefly rinsed in PBS before mounting in 50% glycerol/PBS.

Microscopy

Fluorescent images were taken by fluorescence microscopy using an Olympus BX60 microscope equipped with a SpotRT monochrome model 2.1.1 camera and Spot Advanced software (Diagnostic Instruments, Sterling Heights, MI).

REFERENCES

Adams, V.H., McBryant, S.J., Wade, P.A., Woodcock, C.L., and Hansen, J.C. (2007) Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J Biol Chem* **282**: 15057-15064.

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.

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.

Austin, C., Novikova, N., Guacci, V., and Bellini, M. (2009) Lampbrush chromosomes enable study of cohesin dynamics. *Chromosome Res* **17**: 165-184.

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

Beenders, B., Jones, P.L., and Bellini, M. (2007) The tripartite motif of nuclear factor 7 is required for its association with transcriptional units. *Mol Cell Biol* **27**: 2615-2624.

Bi, X., Jones, T., Abbasi, F., Lee, H., Stultz, B., Hursh, D.A., and Mortin, M.A. (2005) Drosophila caliban, a nuclear export mediator, can function as a tumor suppressor in human lung cancer cells. *Oncogene* **24**: 8229-8239.

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, J.D., and Beggs, J.D. (1992) Roles of PRP8 protein in the assembly of splicing complexes. *EMBO J* **11**: 3721-3729.

Carbonnelle, D., Jacquot, C., Lanco, X., Le Dez, G., Tomasoni, C., Briand, G., Tsotinis, A., Calogeropoulou, T., and Roussakis, C. (2001) Up-regulation of a novel mRNA (NY-CO-1) involved in the methyl 4-methoxy-3-(3-methyl-2-butenoyl) benzoate (VT1)-induced proliferation arrest of a non-small-cell lung carcinoma cell line (NSCLC-N6). *Int J Cancer* **92**: 388-397.

Carney, R.M., Wolpert, C.M., Ravan, S.A., Shahbazian, M., Ashley-Koch, A., Cuccaro, M.L., Vance, J.M., and Pericak-Vance, M.A. (2003) Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatr Neurol* **28**: 205-211.

Chadwick, L.H., and Wade, P.A. (2007) MeCP2 in Rett syndrome: transcriptional repressor or chromatin architectural protein? *Curr Opin Genet Dev* **17**: 121-125.

Chahrour, M., and Zoghbi, H.Y. (2007) The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56**: 422-437.

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

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* **302**: 885-889.

Christodoulou, J., and Weaving, L.S. (2003) MECP2 and beyond: phenotype-genotype correlations in Rett syndrome. *J Child Neurol* **18**: 669-674.

Gall, J.G., Bellini, M., Wu, Z., and Murphy, C. (1999) Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol Biol Cell* **10**: 4385-4402.

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.

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.

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.

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., and El-Osta, A. (2005) Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat Genet* **37**: 254-264.

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

Hite, K.C., Adams, V.H., and Hansen, J.C. (2009) Recent advances in MeCP2 structure and function. *Biochem Cell Biol* **87**: 219-227.

Horowitz, D.S., Kobayashi, R., and Krainer, A.R. (1997) A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA* **3**: 1374-1387.

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.

Jian, L., Archer, H.L., Ravine, D., Kerr, A., de Klerk, N., Christodoulou, J., Bailey, M.E., Laurvick, C., and Leonard, H. (2005) p.R270X MECP2 mutation and mortality in Rett syndrome. *Eur J Hum Genet* **13**: 1235-1238.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**: 187-191.

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.

Klose, R.J., and Bird, A.P. (2004) MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J Biol Chem* **279**: 46490-46496.

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.

Lam, C.W., Yeung, W.L., Ko, C.H., Poon, P.M., Tong, S.F., Chan, K.Y., Lo, I.F., Chan, L.Y., Hui, J., Wong, V., Pang, C.P., Lo, Y.M., and Fok, T.F. (2000) Spectrum of mutations in the MECP2 gene in patients with infantile autism and Rett syndrome. *J Med Genet* **37**: E41.

Lauber, J., Plessel, G., Prehn, S., Will, C.L., Fabrizio, P., Groning, K., Lane, W.S., and Luhrmann, R. (1997) The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA* **3**: 926-941.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., and Wei, C.L. Dynamic changes in the human methylome during differentiation. *Genome Res* **20**: 320-331.

Levesque, L., Guzik, B., Guan, T., Coyle, J., Black, B.E., Rekosh, D., Hammarskjold, M.L., and Paschal, B.M. (2001) RNA export mediated by tap involves NXT1-dependent interactions with the nuclear pore complex. *J Biol Chem* **276**: 44953-44962.

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.

Lin, C., Yang, L., Yang, J.J., Huang, Y., and Liu, Z.R. (2005) ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol Cell Biol* **25**: 7484-7493.

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.

Lunyak, V.V., Burgess, R., Prefontaine, G.G., Nelson, C., Sze, S.H., Chenoweth, J., Schwartz, P., Pevzner, P.A., Glass, C., Mandel, G., and Rosenfeld, M.G. (2002) Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science* **298**: 1747-1752.

Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., and Bryant, S.H. (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* **37**: D205-210.

Matzat, L.H., Berberoglu, S., and Levesque, L. (2008) Formation of a Tap/NXF1 homotypic complex is mediated through the amino-terminal domain of Tap and enhances interaction with nucleoporins. *Mol Biol Cell* **19**: 327-338.

Meehan, R.R., Lewis, J.D., and Bird, A.P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res* **20**: 5085-5092.

Miyake, K., and Nagai, K. (2007) Phosphorylation of methyl-CpG binding protein 2 (MeCP2) regulates the intracellular localization during neuronal cell differentiation. *Neurochem Int* **50**: 264-270.

Moore, H., Leonard, H., de Klerk, N., Robertson, I., Fyfe, S., Christodoulou, J., Weaving, L., Davis, M., Mulroy, S., and Colvin, L. (2005) Health service use in Rett syndrome. *J Child Neurol* **20**: 42-50.

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.

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.

Neul, J.L., Fang, P., Barrish, J., Lane, J., Caeg, E.B., Smith, E.O., Zoghbi, H., Percy, A., and Glaze, D.G. (2008) Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology* **70**: 1313-1321.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007) Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Mol Cell Biol* **27**: 864-877.

Patel, S., Novikova, N., Beenders, B., Austin, C., and Bellini, M. (2008) Live images of RNA polymerase II transcription units. *Chromosome Res* **16**: 223-232.

Patel, S.B., Novikova, N., and Bellini, M. (2007) Splicing-independent recruitment of spliceosomal small nuclear RNPs to nascent RNA polymerase II transcripts. *J Cell Biol* **178**: 937-949.

Philips, A.V., Timchenko, L.T., and Cooper, T.A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**: 737-741.

Rett, A. (1966) [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. *Wien Med Wochenschr* **116**: 723-726.

Scanlan, M.J., Chen, Y.T., Williamson, B., Gure, A.O., Stockert, E., Gordan, J.D., Tureci, O., Sahin, U., Pfreundschuh, M., and Old, L.J. (1998) Characterization of human colon cancer antigens recognized by autologous antibodies. *Int J Cancer* **76**: 652-658.

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

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

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

Stickeler, E., Fraser, S.D., Honig, A., Chen, A.L., Berget, S.M., and Cooper, T.A. (2001) The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *EMBO J* **20**: 3821-3830.

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

Wang, A., Forman-Kay, J., Luo, Y., Luo, M., Chow, Y.H., Plumb, J., Friesen, J.D., Tsui, L.C., Heng, H.H., Woolford, J.L., Jr., and Hu, J. (1997) Identification and characterization of human genes encoding Hprp3p and Hprp4p, interacting components of the spliceosome. *Hum Mol Genet* **6**: 2117-2126.

Watson, P., Black, G., Ramsden, S., Barrow, M., Super, M., Kerr, B., and Clayton-Smith, J. (2001) Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein. *J Med Genet* **38**: 224-228.

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

Ylisaukko-Oja, T., Rehnstrom, K., Vanhala, R., Kempas, E., von Koskull, H., Tengstrom, C., Mustonen, A., Ounap, K., Lahdetie, J., and Jarvela, I. (2005) MECP2 mutation analysis in patients with mental retardation. *Am J Med Genet A* **132A**: 121-124.

Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S., and Zoghbi, H.Y. (2005) Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci U S A* **102**: 17551-17558.

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

Zoghbi, H.Y. (2003) Postnatal neurodevelopmental disorders: meeting at the synapse? *Science* **302**: 826-830.

Zoghbi, H.Y. (2005) MeCP2 dysfunction in humans and mice. *J Child Neurol* **20**: 736-740.

FIGURES & TABLES

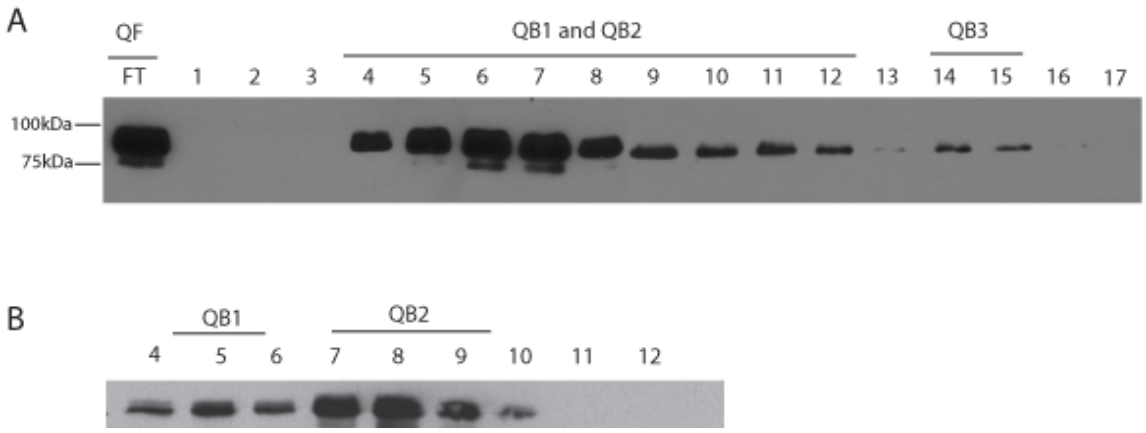

Figure 2.1: Brain-derived nuclear Mecp2 exists in multiple biochemically distinct pools. (A) Chromatographic separation of crude rat brain nuclear extract by strong anion exchange (MonoQ resin) results in three distinct pools of Mecp2 as indicated by western blot using the MeCP2 7-18 antibody. The majority of Mecp2 does not bind the column (QF), while the bound Mecp2 elutes in two peaks, at 230mM NaCl (QB1/2) and 450mM NaCl (QB3). (B) QB1/2 MeCP2 elutes in two peaks from the MonoS column, the first at 400mM NaCl (QB1) and the second eluting at 550mM NaCl (QB2).

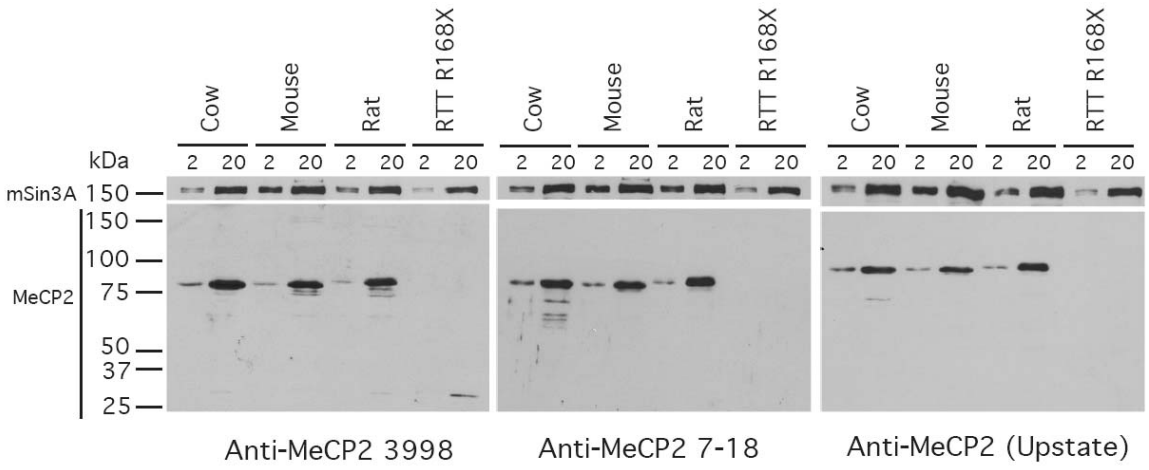

Figure 2.2: Anti-MeCP2 antibodies react specifically with MeCP2 from multiple protein sources. Antibodies against MeCP2 were used in western blotting of brain nuclear extracts from cow, mouse, and rat, and nuclear extracts from human lymphoblast expressing the RTT R168X truncation to show their specificities. The Jones lab generated anti-MeCP2 3998 (left panel) (Harikrishnan *et al.*, 2005), and anti-MeCP2 7-18 (center panel) were compared with an anti- MeCP2 antibody purchased from Upstate Biotech (right panel). All three antibodies react with a single polypeptide migrating at ~80 kDa, consistent with MeCP2's known migration on SDS-PAGE, in brain nuclear extracts (2 or 20 _g/lane) from cow, mouse, and rat. These antibodies show no reactivity to any polypeptides in the human MeCP2 RTT R168X extracts by western blot analysis. These data confirm that the anti-MECP2 7-17 antibody used in this study is specific for MeCP2. Load controls using an antibody for Sin3A are shown above.

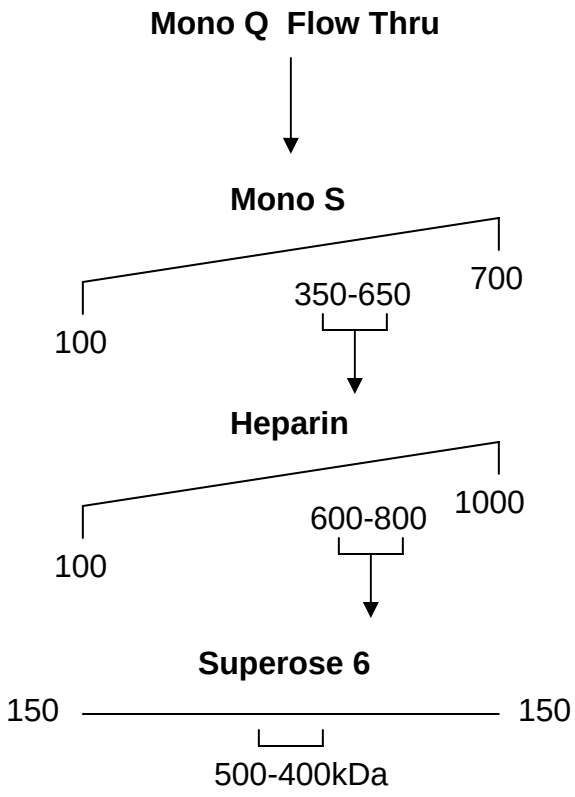

Figure 2.3: QF MeCP2 purification scheme. MonoQ flow through MeCP2 was purified using a four-step process including the MonoQ strong anion exchange resin, MonoS strong cation exchange resin, Heparin affinity resin and by Superose 6 gel filtration.

MeCP2 MonoQ flow-through (QF): Sequence coverage 47%

MVAGMLGLRKEKSEDQDLQGLKEKPLKFKKVKKDKKEDKEGKHEPLQPSAHHSAEPAEAGKAETSESSGSAPAVPEASASP
KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRG
SPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEKSPGKLLVKMPFQASPGGKGEGGGATTSAQ
VMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTRETVSIEVKEVVKPLLVS
TLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHHHAESPKAPMPLLPPPPPPEPQSSEDPISPPEPQDLSSSICK
EEKMPRAGSLESDGCPKEPAKTQPMVAAAATTTTTTTTTVAEKYKHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS

MeCP2 MonoQ bound 2 (QB2): Sequence coverage 23%

MVAGMLGLRKEKSEDQDLQGLKEKPLKFKKVKKDKKEDKEGKHEPLQPSAHHSAEPAEAGKAETSESSGSAPAVPEASASP
KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRG
SPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEKSPGKLLVKMPFQASPGGKGEGGGATTSAQ
VMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTRETVSIEVKEVVKPLLVS
TLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHHHAESPKAPMPLLPPPPPPEPQSSEDPISPPEPQDLSSSICK
EEKMPRAGSLESDGCPKEPAKTQPMVAAAATTTTTTTTTVAEKYKHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS

MeCP2 MonoQ bound 3 (QB3): Sequence coverage 33%

MVAGMLGLRKEKSEDQDLQGLKEKPLKFKKVKKDKKEDKEGKHEPLQPSAHHSAEPAEAGKAETSESSGSAPAVPEASASP
KQRRSIIRDRGPMYDDPTLPEGWTRKLKQRKSGRSAGKYDVYLINPQGKAFRSKVELIAYFEKVGDTSLDPNDFDFTVTGRG
SPSRREQKPPKKPKSPKAPGTGRGRGRPKGSGTGRPKAAASEGVQVKRVLEKSPGKLLVKMPFQASPGGKGEGGGATTSAQ
VMVIKRPGRKRKAEADPQAIPKKRGRKPGSVVAAAAAEAKKKAVKESSIRSVQETVLPIKKRKTRETVSIEVKEVVKPLLVS
TLGEKSGKGLKTCKSPGRKSKESSPKGRSSSASSPPKKEHHHHHHHAESPKAPMPLLPPPPPPEPQSSEDPISPPEPQDLSSSICK
EEKMPRAGSLESDGCPKEPAKTQPMVAAAATTTTTTTTTVAEKYKHRGEGERKDIVSSSMPRPNREEPVDSRTPVTERVS

Figure 2.4: Mass spectrometry peptide identifications (in red) for rat MeCP2 from the QF, QB2, and QB3 pools of MeCP2.

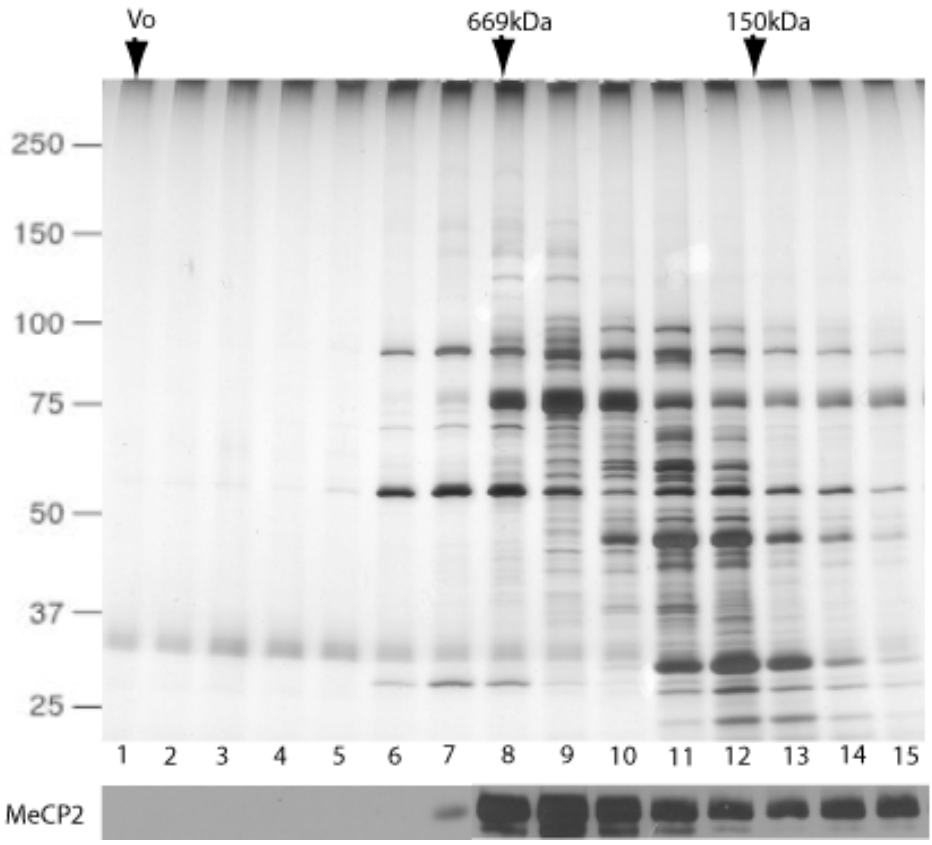

Figure 2.5: The QF pool of Mecp2 does not co-purify with other proteins. (Upper panel) Silver-stain analysis of the Superose 6 fractionation of the QF Mecp2 pool. (Lower panel) Western blot analysis of Superose 6 fractions showing Mecp2 protein peaks in fraction 9.

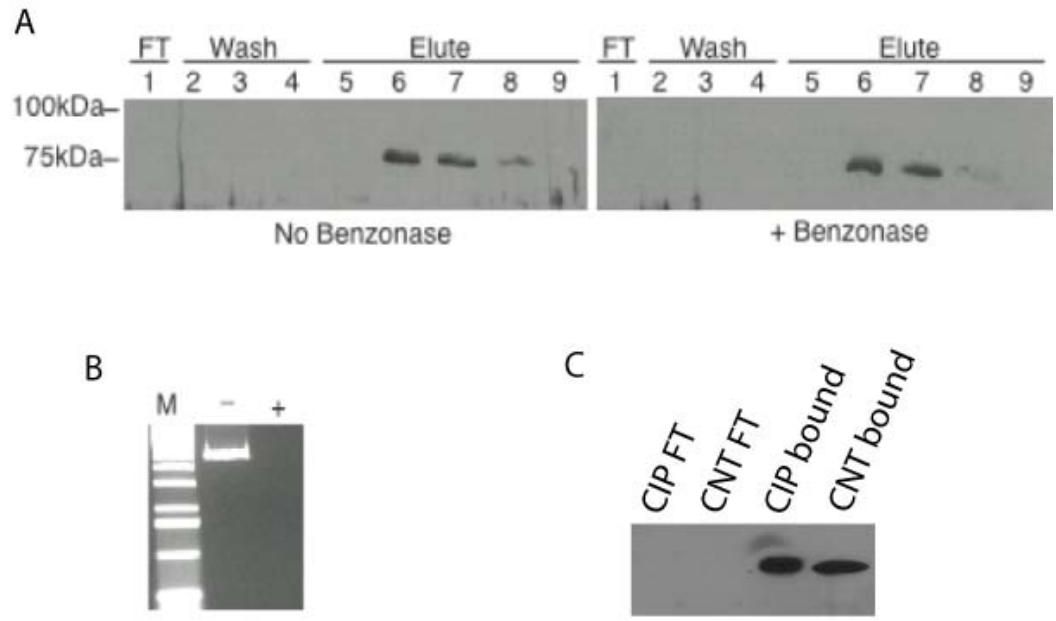

Figure 2.6: (A) MonoQ resin bound fractions of Mecp2 were treated with (+) or without Benzonase nuclease and tested for ability to re-bind the Mono Q. Western blot analysis for Mecp2 of MonoQ flow thru, wash, and 1000mM NaCl step elution shows Benzonase treatment does not affect binding of MeCP2 to the Mono Q column. (B) Plasmid spiked MonoQ fractions treated with benzonase (+) or untreated (-) in parallel served as controls for Benzonase treatment. (C) MonoQ resin bound fractions of Mecp2 were treated with or without calf intestinal phosphatase (CIP) and tested for ability to re-bind the Mono Q. Western blot analysis for Mecp2 of MonoQ flow through (FT), and column bound step fractions shows CIP treatment does not affect binding of MeCP2 to the Mono Q column.

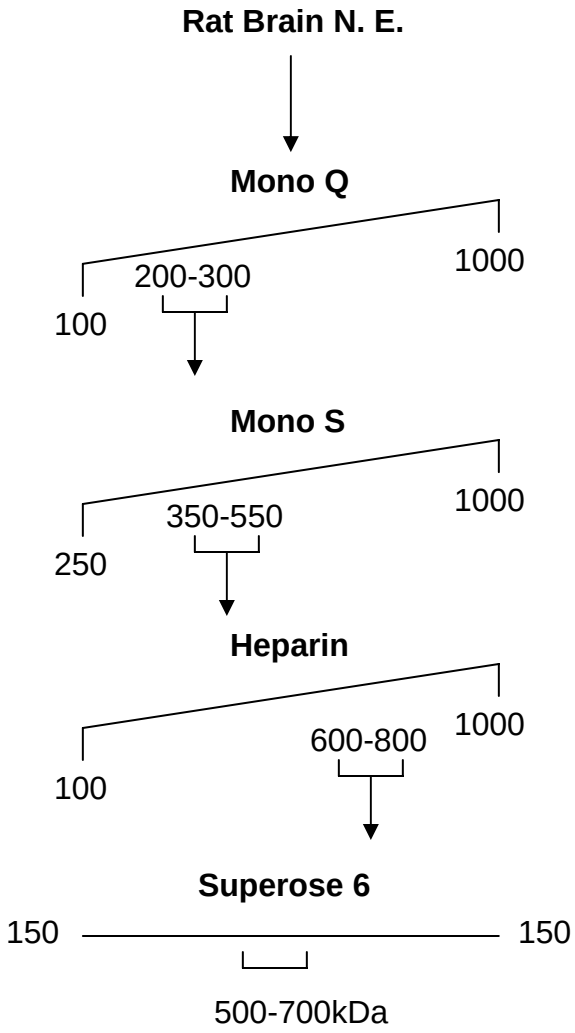

Figure 2.7: QB2 MeCP2 purification scheme. MeCP2 peak fractions were purified using a four-step process including the MonoQ strong anion exchange resin, MonoS strong cation exchange resin, Heparin affinity resin and by Superose 6 gel filtration.

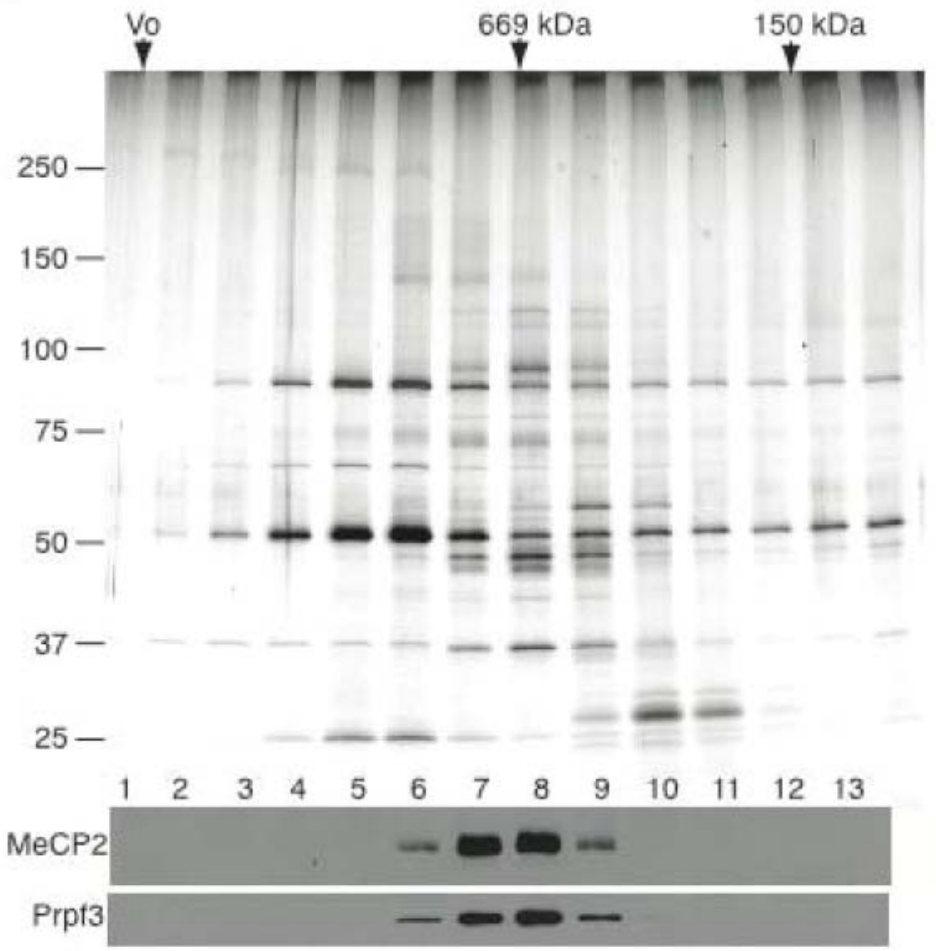

Figure 2.8: The QB2 pool of Mecp2 co-purifies with 6 candidate proteins including Prpf3. (Upper panel) Silver-stain analysis of the Superose6 fractionation of the QB2 Mecp2 pool. (Lower panel) Western blot analysis of Superose 6 fractions shows Mecp2 protein peaks in fractions 7 and 8, precisely co-fractionating with Prpf3 protein.

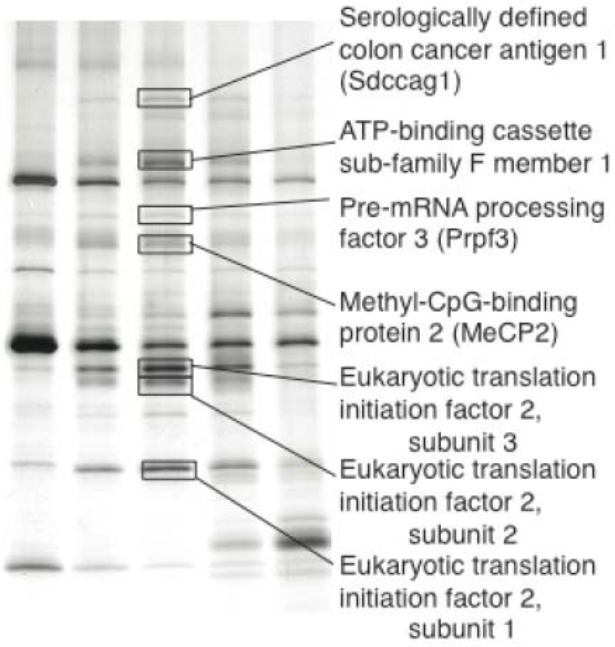

Figure 2.9: Identification of candidate Mecp2 complex proteins. (a) Polypeptides from silver stained Superose6 fractions were excised and identified by mass spectrometry (tandem LC MS/MS). Mecp2 and two other nuclear proteins, Sdccag1 and Prpf3 were identified as well as components of a translation initiation complex. (b) Mass spectrometry peptides identifying Mecp2, Prpf3, and Sdccag1.

EIF2S1 [eukaryotic translation initiation factor 2, subunit 1] Sequence Coverage: 33%
MPGLSCRFYQHKFPEVEDVVMVNVRSIAEMGAYVSLLEYNNIEGMILLSELSRRRIRSINKLIRIGRNECVVVIRVDKEKGYID
LSKRRVSPEEAIKCEDKFTKSKTVYSILRHVAEVLEYTKDEQLESLFQRTAWVFDDKYKRPGYGAYDAFKHAVSDPSILDSL
DLNEDEREVLINNINRRLTPQAVKIRADIEVACYGYEGIDAVKEALRAGLNCSTETMPIKINLIAPPRYVMTTTTLERTEGLSV
LNQAMAVIKEKIEEKRGVFNVQMEPKVVTDTDETELARQLERLERENAEVDGDDDAEEMEAKAED

EIF2S2 [eukaryotic translation initiation factor 2, subunit 2] Sequence Coverage: 30%
MSGDEMIFDPTMSKKKKKKKKPFMLDEEGDAQTEETQPSETKEVEPEPAEEKDVEADEEDSRKKDASDDLDDLNFFNQKK
KKKKTKKIFDIDEAEEAIKDVKIESDAQEPAEPEDDLDIMLGNKKKKKKNVKFPDEDEILEKDEALEDEDSKKDDGISFSNQT
GPAWAGSERDYTYEELLNRVFNIMREKNPDMVAGEKRKFVMKPPQVVRVGTKKTSFVNFTDICKLLHRQPKHLLAFLLAEL
GTSGSIDGNNQLVIKGRFQQKQIENVLRRYIKEYVTCHTCRSPDTILQKDTRLYFLQCETCHSRCSVASIKTGFQAVTGKRAQ
LRAKAN

EIF2S3 [eukaryotic translation initiation factor 2, subunit 3] Sequence Coverage: 37%
MAGGEAGVTLGQPHLSRQDLATLDVTKLTPLSHEVISRQATINIGTIGHVAHGKSTVVKAISGVHTVRFKNELERNITIKLGY
ANAKIYKLDDPSCPRPECYRSCGSSTPDEFPTDIPGTKGNFKLVRHVSFVDCPGHDILMATMLNGAAVMDAALLLIAGNESC
PQPQTSEHLAAIEIMKLKHILILQNKIDLVKESQAKEQYEQILAFVQGTVAEGAPIIPISAQLKYNIEVVCEYIVKKIPVPPRDFT
SEPRLIVIRSFDVNKPGCEVDDLKGGVAGGSILKGVLKVGQEIEVRPGIVSKDSEGKLMCKPIFSKIVSLFAEHNDLQYAAPGG
LIGVGTKIDPTLCRADRMVGQVLGAVGALPEIFTELEISYFLLRRLLGVRTEGDKKAAKVQKLSKNEVLMVNIGSLSTGGRVS
AVKADLGKIVLTNPVCTEVGEKIALSRRVEKHWRLIGWGQIRRGVTIKPTVDDD

PRP3 [PRP3 pre-mRNA processing factor 3 homolog] Sequence Coverage: 9%
MALSKRELDELKPWIEKTVKRVLGFSEPTVVTAALNCVGKGMDKKKAADHLKPFLDDSTLRFVDKLFEAVEEGRSSRHSKS
SSDRSRKRELKEVFGDDSEISKESSGVKKRRIPRFEEVEEEPEVIPGPPSESPGMLTKLQIKQMMEAATRQIEERKKQLSFISPPA
PQPKTPSSSQPERLPIGNTIQPSQAATFMNDAIEKARKAAELQARIQAQLALKPGLIGNANMVGLANLHAMGIAPPKVELKDQ
TKPTPLILDEQGRTVDATGKEVELTHRMPTLKANIRAVKREQFKQQLKEKPSEDMESNTFFDPRVSIAPSQRQRRTFKFHDKG
KFEKIAQRLRTKAQLEKLQAEISQAARKTGIHTSTRLALIAPKKELKEGDIPEIEWWDSYIIPNGFDLTEENPKREDYFGITNLV
EHPAQLNPPVDNDTPVTLGVYLTKKEQKKLRRQTRREAQKELQEKVRLGLTPPPEPKVRISNLMRVLGTEAVQDPTKVEAH
VRAQMAKRQKAHEEANAARKLTAEQRKVKKVKKLKEDISQGVHISVYRVRNLSNPAKKFKIEANAGQLYLTGVVVLHKD
VNVVVVEGGPKAQRKFKRLMLHRIKWDEQTSNTKGDDDEESDEEAVKKTNKCVLVWEGTAKDRSFGEMKFKQCPTENM
AREHFKKHGAEHYWDLALSESVLESTD

ABCF1 [ATP-binding cassette sub-family F member 1] Sequence Coverage: 32%
MPKGPKQQPPEPEWIGDGEGTSPADKVVKKGKKDKKTKKTFFEELAVEDKQAGEEEKLQKEKEQQQQQQQQKKKRDTRK
GRRKKDVDDDDDGDERVLMERLKQLSVPASDEEDEVPVPVPRGRKKAKGGNVFEALIQDESEEEKEEEEEKPVLKPAKPEK
NRINKAVAEEPPGLRNKKGKEEKSKGKAKNKPSATDSEGEDDEDMTKEKEPPRPGKDKDKKGAEQGSEEEKEEKEGEVKA
NDPYAHLSKKEKKKLKKQMDYERQVESLKAANAAENDFSVSQAEVSSRQAMLENASDIKLEKFSISAHGKELFVNADLYIV
AGRRYGLVGPNGKGKTTLLKHIANRALSIPPNIDVLLCEQEVVADETPAVQAVLRADTKRLRLLEEEKRLQGQLEQGDDTA
AEKLEKVYEELRATGAAAAEAKARRILAGLGFDPEMQNRPTQKFSGGWRMRVSLARALFMEPTLLMLDEPTNHLDLNAVI
WLNNYLQGWRKTLLIVSHDQGFLDDVCTDIIHLDTQRLHYYRGNYMTFKKMYQQKQKELLKQYEKQEKKLKELKAGGKS
TKQAEKQTKEVLTRKQQKCRRKNQDEESQDPPELLKRPREYTVRFTFPDPPPLSPPVLGLHGVTFGYEGQKPLFKNLDFGID
MDSRICIVGPNGVGKSTLLLLLTGKLTPTNGEMRKNHRLKIGFFNQQYAEQLHMEETPTEYLQRGFNLPYQDARKCLGRFGL
ESHAHTIQICKLSGGQKARVVFAELACREPDVLILDEPTNNLDIESIDALGEAINEYKGAVIVVSHDARLITETNCQLWVVEEQ
SVSQIDGDFDDYKREVLEALGEVMVNRPRD

SDCCAG1 [serologically defined colon cancer antigen 1] Sequence Coverage: 3%
MKTRFSTVDLRAVLAELNANLLGMRVNNVYDVDNKTYLIRLQKPDFKATLLLESGIRIYTTEFEWPKNMMPSSFAMKCRKH
LKSRRLVSAKQLGVDRIVDFQFGSDEAAYHLIIELYDRGNIVLTDYEYLILNILRFRTDEADDVKFAVRERYPIDHARAAEPLL
TLERLTEVIARAPRGELLKRVLNLLLPYGPALIEHCLIENGFSGNVKVDEKLESKDIEKILVCVQRAEDYLEKTANFNGKGYII
QKREVKPSLDANKPAEDILMYEEFHPFLFSQHLQCPYIEFESFDKEKQALKKLDNVXKDHENRLEALQQAQEIDKLKGELIE
MNLQIVDRAIQVVRSALANQIDWTEIGVIVKEAQAQGDHVASAIKELKLQTNHITMLLRNPYLLSEEEDGDGDGSIENSDAE
APKGKKKAKEQAAAEASEGQAAACRCGPQPVSLCQCQKVLXSXEVCCXKTTENCRSCXEGIQISREENKANLKRSTNSYFY
PKSKESVLVXEISVVYXFRELSHYRWSRSATEXDYCEKILNTRRHLCACXSSWSYQLCNXESNRRSHPSSDFDXSRHNGTLL
QRGLGCPCYHECLVGAPSSGIXNSTDRRVLNNWKLHDKRKKEFPSSFIPNDGVXLPFXGRXVLCLETSRXTKRQSAGXRHEN
IDKLHKXTHVRRNGTARRGRQQXRRDRGVMWNHGRSGTQDSGXSRGHCCSQWKRGTELXRWRSHQNSHERXRAHWXG
EGRGGRISXHHHXLVPSSVPKAPTETDSKRRIFXFKXQXITEPKTFVSQGEKRNEKEKAPMXLRGFRSDRRKGQRKRKCCAQ
XSXPEHKQKCGSWTANEKRPEEXNEKNEGKIQRPGXXRSXTYYEIVGICRFKQRRKGEERKERKNKRXTREKKPTETQRWT
AGFRCCXRNPVPSGVDSXLTRLCCGXATXXQGRTXSGSAGKXGKSIXLFDRATTSXRCTNVCYSNMCSLHHHDKLXIQSET
YSWSSEKGKSCKDSLEQFHALQRSNSKRKRLIPKCEGHRFTKKHSRESESVCPQSSARKKKI

Figure 2.10: Mass spectrometry peptide identifications (in red) for MeCP2 copurifying polypeptides from the QB2 pool of MeCP2.

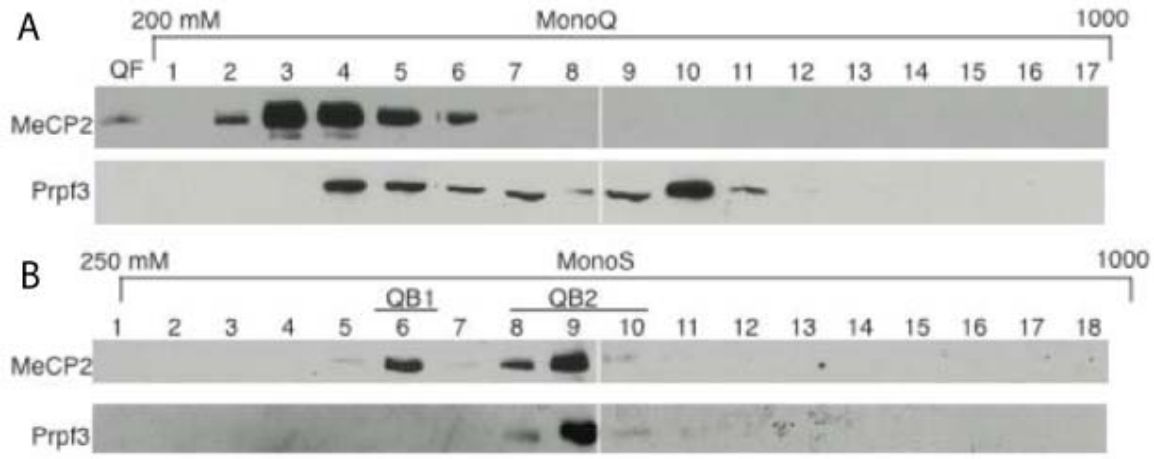

Figure 2.11: MeCP2 and Prpf3 precisely co-fractionation. (A) The MonoQ fractionation of brain-derived nuclear extract reveals an overlap of Prpf3 and Mecp2 yet pools of each protein do not co-fractionate and remain independent of the other. (B) Peak fractions of the QB1/2 fractionated over the MonoS resin show two distinct pools of MeCP2 peaking at 400 mM NaCl (QB1) and 550 mM NaCl (QB2) by western blot analysis. Corresponding MonoS fractions probed for Prpf3 protein shows precise co-fractionation with QB2 Mecp2.

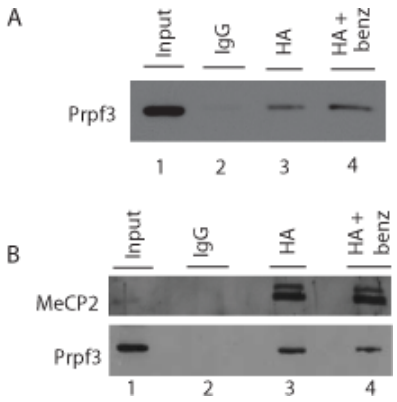

Figure 2.12: Co-IPs confirm that MeCP2 associates with Prpf3 *in vivo*. (A) Western blotting of a co-IP experiment from fractionated brain extract showing: a peak MeCP2 fraction (lane 1, 5% input), IgG IP (lane 2), anti-MeCP2 IP (lane 3), and anti-MeCP2 IP treated with benzonase (lane 4), demonstrates that PRPF3 interacts with MeCP2 in brain independent of nucleic acids. (B) Western blotting of a co-IP experiment from HA-MeCP2 transfected HT22 cells showing: whole cell extract (lane 1, 5% input), IgG IP (lane 2), anti-HA IP (lane 3) and anti-HA IP treated with benzonase (lane 4), demonstrates that PRPF3 interacts with MeCP2 cell culture independent of nucleic acids**.**

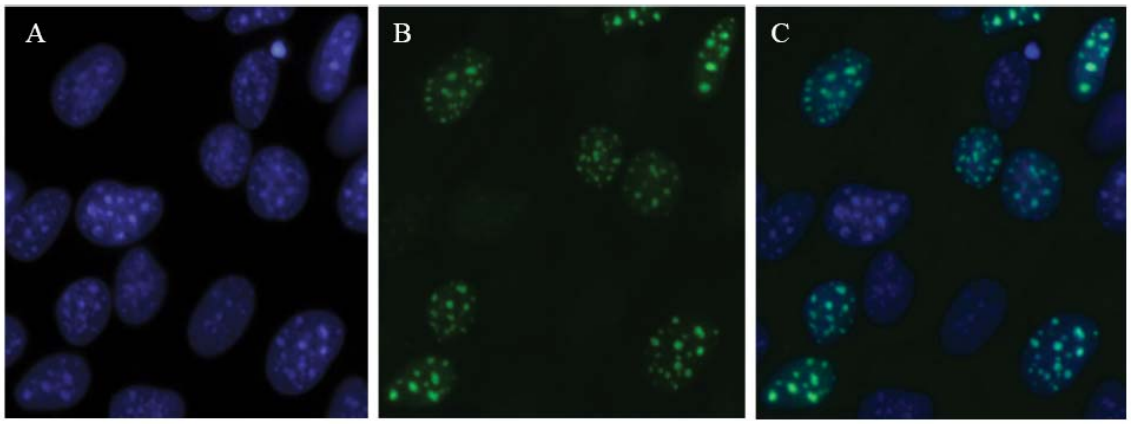

Figure 2.13: HT22 hippocampal neuronal cells stably transected with HA-MeCP2, visualized by A. DAPI staining and B. indirect immunofluorescence for HA-MeCP2 with C. images merged, show HA-MeCP2 localizes to heterochromatic foci.

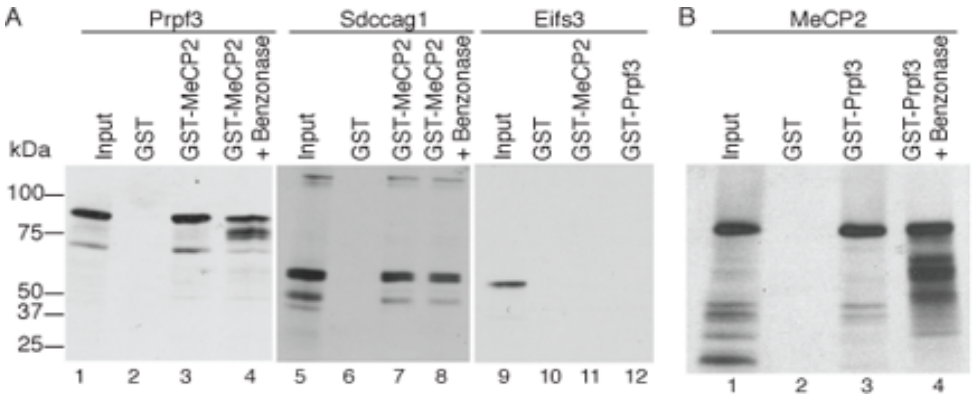

Figure 2.14: MeCP2 directly interacts with Prpf3 and Sdccag1. For all experiments, the GST-tagged protein was bacterially generated and purified and the visualized [^{35}S]-methionine labeled interacting proteins were generated by *in vitro* transcription and translation. (A) Prpf3 (left) and Sdccag1 (middle) interact directly with GST tagged MeCP2 but not GST alone. Eif2s3 (right) does not interact with either GST MeCP2 or GST Prpf3. Benzonase treatment indicates these interactions are independent of nucleic acids. (B) Reciprocally, MeCP2 interacts with GST tagged Prpf3 independent of nucleic acids.

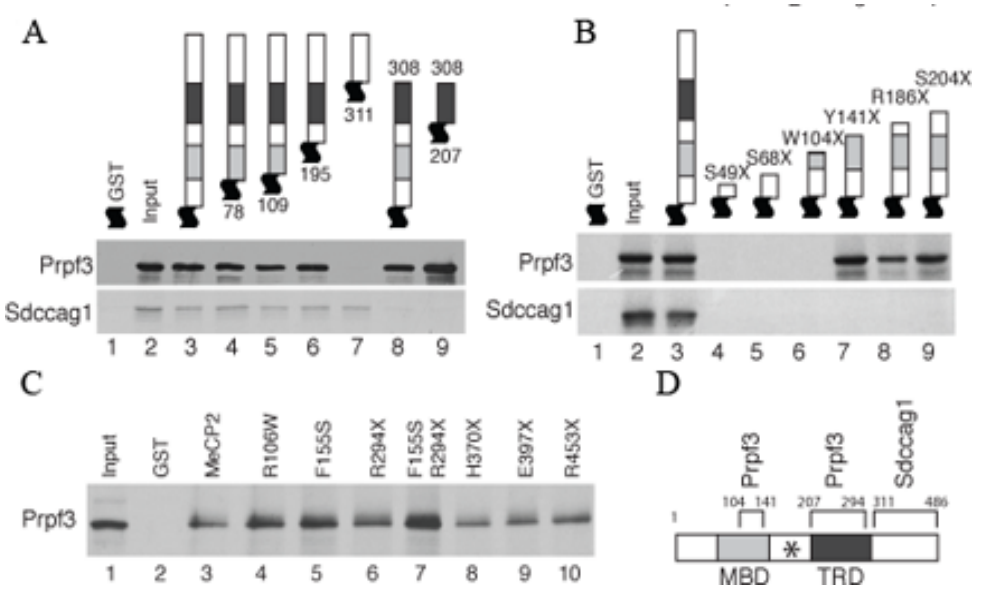

Figure 2.15: (A) GST-MeCP2 deletion constructs mapped the region of MeCP2 required for the direct interactions with Prpf3 or Sdccag1 *in vitro*. (B) GST tagged MeCP2 containing the indicated RTT nonsense mutations disrupt Prpf3 binding if truncations are prior to amino acid 104, but identify a second Prpf3 binding site on MeCP2 in the MBD between amino acids 104 and 141. All RTT truncations tested abolished Sdccag1 binding to MeCP2. (C) GST tagged MeCP2 containing the indicated RTT mutations residing outside of the Prpf3 interaction domains maintain the interaction with Prpf3. (D) Map of MeCP2e2 protein showing the Prpf3 and Sdccag1 interaction domains with boundary amino acid numbers. The MBD, TRD, and (*) RNA binding domain are indicated.

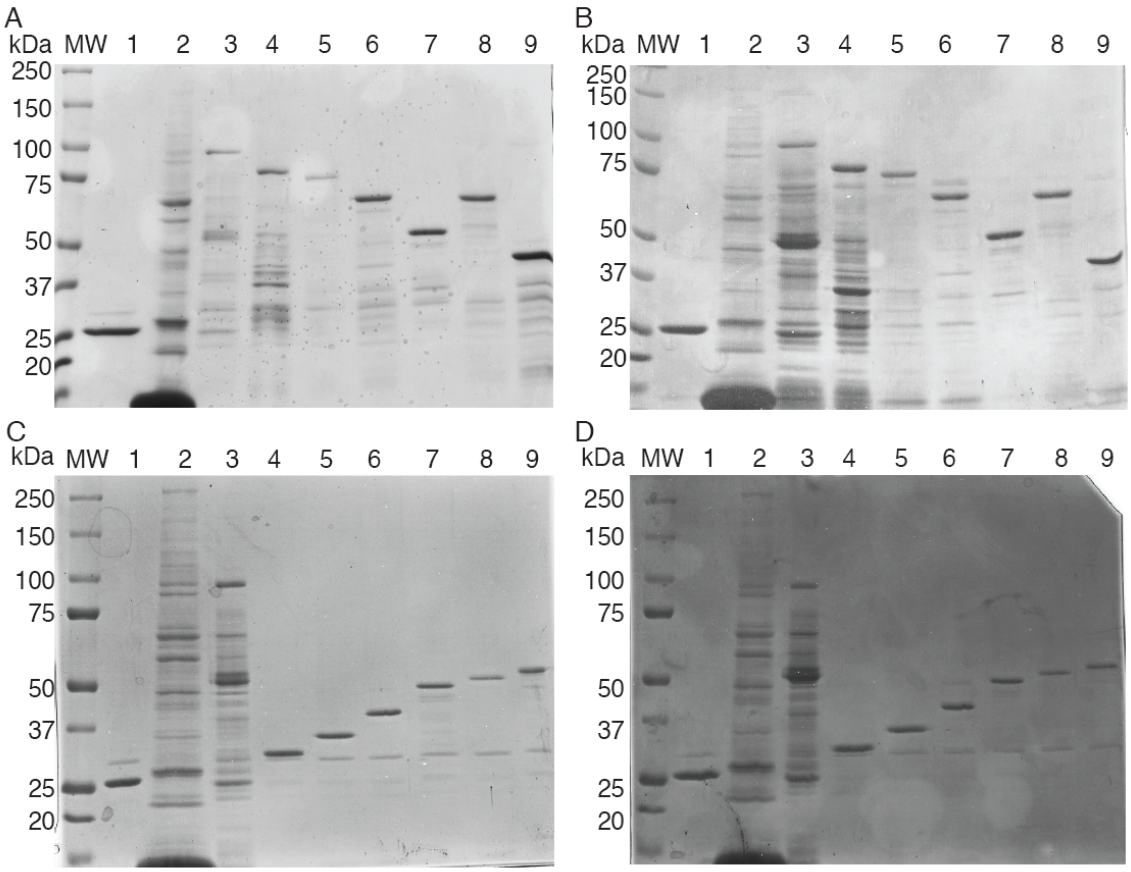

Figure 2.16: Coomassie brilliant blue stained SDS-PAGE gels from GST pull-down experiments. Staining shows equivalent levels of GST-fusion proteins in each pull-down reaction. A. GST-fusion proteins from Figure 2.15A top, Prpf3. B. GST-fusion proteins from Figure 2.15A bottom, Sdccag1. C. GST-fusion proteins from Figure 2.15B top, Prpf3. D. GST-fusion proteins from Figure 2.15B bottom, Sdccag1.

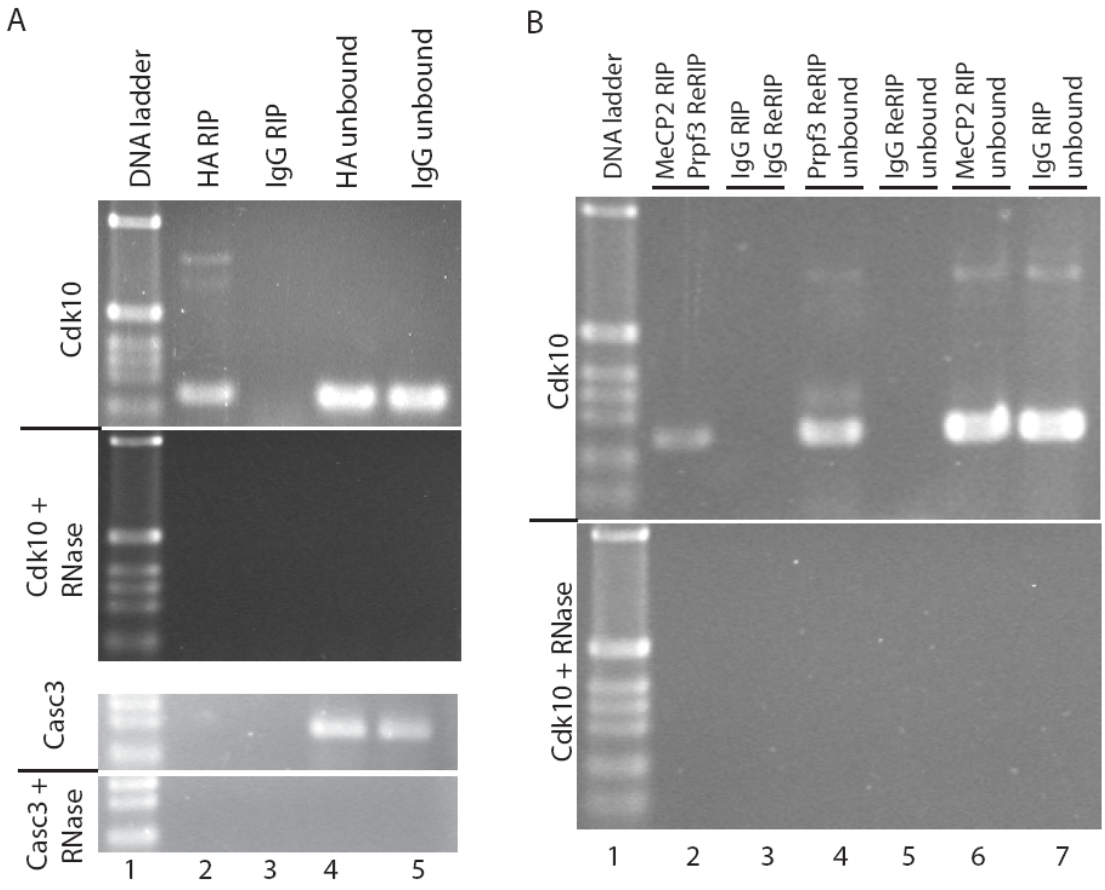

Figure 2.17: The MeCP2-Prpf3 complex interacts with mRNA *in vivo.* (A) RT-PCR analysis of a RIP from HA-MeCP2 stably transfected HT-22 cells indicates an association of HA-MeCP2 with Cdk10 mRNA (top, lane 2), but not Casc3 (bottom, lane 2). Control RIPs using normal rabbit serum (IgG) (top and bottom, lane 3) show no RT-PCR product. RT-PCRs of unbound RNA from RIPs indicate target mRNAs were present in all RIP samples (lanes 4 and 5). All RT-PCRs were RNase sensitive (+ RNase) confirming RNA and not DNA as being assayed. (B) An anti-HA RIP followed by anti-Prpf3 Re-RIP experiment from HA-MeCP2 HT-22 cells was assayed for Cdk10 mRNA by RT-PCR (lane 2). Control RIP and ReRIP experiments using normal rabbit serum (IgG) showed no product by RT-PCR (lane 3). Unbound mRNA was assayed by RT-PCR for the RIP (lanes 6 and 7) and reRIPs (lanes 4 and 5) to confirm the presence of the Cdk10 mRNA in the reactions. All RT-PCRs were sensitive to RNase treatment (lower panel) confirming the amplifications were from RNA and not DNA templates.

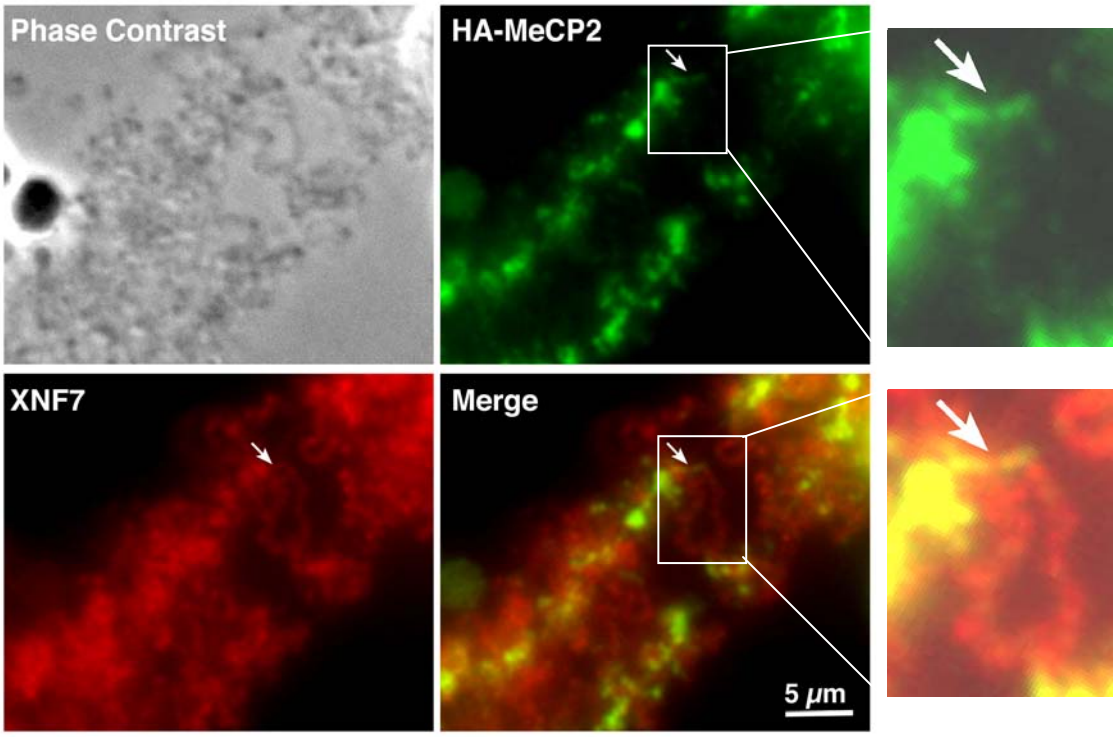

Figure 2.18: *Xenopus* lampbrush chromosomal distribution of hMeCP2. Transcripts coding for hMeCP2-HA were injected into stage IV oocytes and nuclear spreads were prepared. Endogenous xNF7 (red) stains actively transcribed DNA, including a highly extended chromatin loop (white arrow). MeCP2 (green) is detected extensively on heterochromatin rich chromosomal axes, as well as on transcriptionally active loops (white arrow). Insets show a close up of an isolated loop with MeCP2 staining and merge with xNF7. Scale bar represents 5 μm.

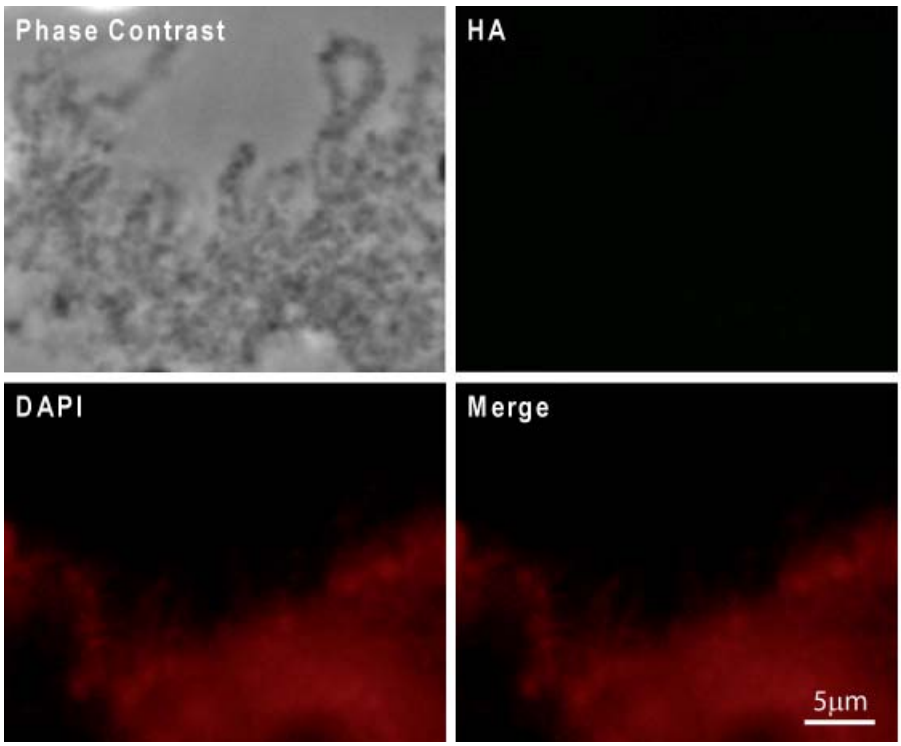

Figure 2.19: *Xenopus* lampbrush chromosome showing anti-HA staining pattern of uninjected controls. A phase contrast image and DAPI staining (red) show lampbrush chromosome axis and loops, while anti-HA staining (green) shows there is no detectable signal on chromatin of uninjected oocytes. Scale bar represents 5 µm.

Table 1: PCR and RT-PCR oligonucleotide primers

Primer	Plasmid Construct	Sequence (5'-3')
1	pGEM ratAbcf1	ATGCCGAAGGGTCCCAAGCAGC
2	pGEM ratAbcf1	TCAATCCCGAGGTCGGTTGAC
3	pGEM ratEif2s1	ATGCCGGGTCTAAGTTGTAGATTTTATCAAC
4	pGEM ratEif2s1	TTAATCTTCAGCTTTGGCTTCCATTTCTTCTG
5	pGEM ratEif2s2	ATGTCCGGGGACGAGATGATTTTGATCCTAC
6	pGEM ratEif2s2	TTAGTTAGCTTTGGCACGGAGCTG
7	pGEM ratEif2s3	ATGGCTGGGGGTGAGGCTG
8	pGEM ratEif2s3	TCAGTCATCATCTACAGTCGGCTTAATG
9	pGEM ratMeCP2	ATGGTAGCTGGGATGTTAGGGC
10	pGEM ratMeCP2	TCAGCTAACTCTCTCGGTCACG
11	pGEM ratPrpf3	ATGGCACTGTCTAAGCGGGAACTGGATG
12	pGEM ratPrpf3	CTACAGAGAACATGGCGCGTGA
13	pGEM ratSdccag1	ATGAAGACCCGCTTCAGCACTGTTGAC
14	pGEM ratSdccag1	CTATTTTCTTTTTACGTGCAGAAGATTGGG
15	pGEM human Prpf3	ATGGCACTGTCAAAGAGGGAGC
16	pGEM human Prpf3	TCAATCAGTGGACTCTAAC
17	pCDNA3.1 ratAbcf1	GCGGCCGCATGCCGAAGGGTCC
18	pCDNA3.1 ratAbcf1	CGCTCGAGTCAATCCCGAGGTCGG
19	pCDNA3.1 ratEif2s1	GCGGCCGCATGCCGGGTCTAAG
20	pCDNA3.1 ratEif2s1	CGCTCGAGTTAATCTTCAGCTTTGGCTTC
21	pCDNA3.1 ratEif2s2	GCGGCCGCATGTCCGGGGAC
22	pCDNA3.1 ratEif2s2	CGCTCGAGTTAGTTAGCTTTGGCACG
23	pCDNA3.1 ratEif2s3	GCGGCCGCATGGCTGGGGGTG
24	pCDNA3.1 ratEif2s3	CGCTCGAGTCAGTCATCATCTACAGTCG
25	pCDNA3.1 ratMeCP2	GCGGCCGCATGGTAGCTGGGATG
26	pCDNA3.1 ratMeCP2	CGCTCGAGTCAGCTAACTCTCTCGGTC
27	pCDNA3.1HA humanMeCP2	GCGGCCGCATGGTAGCTGGGATG
28	pCNA3.1HA humanMeCP2	CGCTCGAGTCAGCTAACTCTCTCGGTC
29	pCDNA3.1 ratPrpf3	GCGGCCGCATGGCACTGTCTAAGC
30	pCDNA3.1 ratPrpf3	CGCTCGAGTCACGCGCCATGTTCTC
31	pCDNA3.1 ratSdccag1	GCGGCCGCATGAAGAGCCGCT
32	pCDNA3.1 ratSdccag1	CGCTCGAGCTATTTCCTTTTTACGTTCAGAAG
33	pCDNA3.1 humanPrpf3	AGAATTCATGGCACTGTCAAAGAG
34	pCDNA3.1 humanPrpf3	CTCGAGTCAATCAGTGGACTCTAAC
35	pGEX5X ratPrpf3	AGAATTCATGGCACTGTCTAAGC
36	pGEX5X ratPrpf3	CTCGAGTCACGCGCCATGTTCTC
37	pGEX5X MeCP2 d1	AGAATTCGCCTCCCCCAAACAG
38	pGEX5X MeCP2 d1	CTCGAGTCAGCTAACTCTCTCGG
39	pGEX5X MeCP2 d2	AGAATTCAAGCAAAGGAAATCTGGC
40	pGEX5X MeCP2 d2	CTCGAGTCAGCTAACTCTCTCGG
41	pGEX5X MeCP2 d3	AGAATTCGGCACCACGAGACC
42	pGEX5X MeCP2 d3	CTCGAGTCAGCTAACTCTCTCGG
43	pGEX5X MeCP2 d4	AGAATTCACGGTCAGCATCGAG
44	pGEX5X MeCP2 d4	CTCGAGTCAGCTAACTCTCTCGG

(Table 1 continued)

45	pGEX5X MeCP2 d5	AGAATTCATGGTAGCTGGGATGTTAG
46	pGEX5X MeCP2 d5	CTCGAGGGTCTTGCGCTTCTTG
47	pGEX5X MeCP2 d6	AGAATTCGGCACCACGAGACC
48	pGEX5X MeCP2 d6	CTCGAGACCGAGGGTGGACAC
49	RT-PCR mouseFRG1 5’	AATTGCCCTGAAGTCTGGCTATGG
50	RT-PCR mouseFRG1 3’	CTTTCAATTTGGCTCTCCTGTCCAG
51	RT-PCR mouseCDK10 5’	GGCCAGGGATACCCAGACAG
52	RT-PCR mouseCDK10 3’	CCTCCGAGAAGGGTGTTGGC
53	RT-PCR mouseCASC3 5’	GATCCTCATCGTCATCCAAGTGC
54	RT-PCR mouseCASC3 5’	CTACAGTGAAGAGGAGAATTCCAAGGTG
55	pGEX5X MeCP2 S49X	AGAATTCGCCTCCCCCAAACAG
56	pGEX5X MeCP2 S49X	ACTCGAGTCATGGCTGCACGGGC
57	pGEX5X MeCP2 S68X	AGAATTCGCCTCCCCCAAACAG
58	pGEX5X MeCP2 S68X	ACTCGAGTCACCCTTCTGATGTCTC
59	pGEX5X MeCP2 W104X	AGAATTCGCCTCCCCCAAACAG
60	pGEX5X MeCP2 W104X	ACTCGAGTCAGCCTTCAGGCAGG
61	pGEX5X MeCP2 Y141X	AGAATTCGCCTCCCCCAAACAG
62	pGEX5X MeCP2 Y141X	ACTCGAGTCACGCAATCAACTCC
63	pGEX5X MeCP2 R168X	AGAATTCGCCTCCCCCAAACAG
64	pGEX5X MeCP2 R168X	ACTCGAGTCATCATCTCGCCGGGAGGG
65	pGEX5X MeCP2 S204X	AGAATTCGCCTCCCCCAAACAG
66	pGEX5X MeCP2 S204X	CTCGAGTCACGTGGCCGCCTTG
67	pGEX5X MeCP2 R106W	AGAATTCATGGTAGCTGGGATGTTAGGG
68	pGEX5X MeCP2 R106W	CTCGAGTCAGCTAACTCTCTCGGTC
69	pGEX5X MeCP2 F155S	AGAATTCATGGTAGCTGGGATGTTAGGG
70	pGEX5X MeCP2 F155S	CTCGAGTCAGCTAACTCTCTCGGTC
71	pGEX5X MeCP2 R294X	AGAATTCATGGTAGCTGGGATGTTAGGG
72	pGEX5X MeCP2 R294X	CTCGAGTCATCGGATAGAAGACTCCTTC
73	pGEX5X MeCP2 F155S/R294X	AGAATTCATGGTAGCTGGGATGTTAGGG
74	pGEX5X MeCP2 F155S/R294X	CTCGAGTCATCGGATAGAAGACTCCTTC
75	pGEX5X MeCP2 H370X	AGAATTCATGGTAGCTGGGATGTTAGGG
76	pGEX5X MeCP2 H370X	CTCGAGTCAGTGATGGTGGTGGTGC
77	pGEX5X MeCP2 E397X	AGAATTCATGGTAGCTGGGATGTTAGGG
78	pGEX5X MeCP2 E397X	CTCGAGTCACTCGGAGCTCTCGG
79	pGEX5X MeCP2 R453X	AGAATTCATGGTAGCTGGGATGTTAGGG
80	pGEX5X MeCP2 R453X	CTCGAGTCATCGGTGTTTGTACTTTTCTG

CHAPTER 3

INSIGHTS INTO THE MOLECULAR FUNCTION OF MECP2 AND RETT SYNDROME

CONCLUSIONS

In the past two decades we have witnessed an evolving role for MeCP2 function. MeCP2 was originally identified by its ability to bind methylated DNA and is the founding member of the methyl-CpG-binding domain (MBD) family of proteins (Hendrich and Bird, 1998; Meehan *et al.*, 1992). The proteins first biological function was repressing transcription by binding methylated DNA and recruiting transcriptional co-repressor complexs (Jones *et al.*, 1998; Nan *et al.*, 1997; Nan *et al.*, 1998). In 1999 a genetic link between MeCP2 and Rett Syndrome (RTT) was identified (Amir *et al.*, 1999). RTT is a severe postnatal neurodevelopmental disorder (Rett, 1966), characterized by a period of apparently normal development from birth to 6-18 months followed by a regression of obtained language and motor skills (Chahrour and Zoghbi, 2007). This severe disorder usually exhibits a deceleration of head growth, respiratory dysfunction, scoliosis, cognitive impairment, seizures, and social withdraw in patients (Hagberg *et al.*, 1983; Rett, 1966). In addition to RTT, numerous *MECP2* mutations have now been linked to a variety of other disorders, including autism, Angelman syndrome, learning disabilities, and mental retardation syndromes (Carney *et al.*, 2003; Chahrour and Zoghbi, 2007; Christodoulou and Weaving, 2003; Lam *et al.*, 2000; Watson *et al.*, 2001; Ylisaukko-Oja *et al.*, 2005; Zoghbi, 2005). The initial connection of MeCP2 to RTT is responsible for a large push among the research community to

understand the biology of MeCP2. Despite an early focus on gene specific repression, more recent work has expanded MePC2's regulatory role beyond transcriptional repression; MeCP2 is implicated in transcriptional activation, genome-wide transcriptional silencing, mediating chromatin and nuclear architecture, and regulating pre-mRNA splicing as well (Chadwick and Wade, 2007; Chahrour *et al.*, 2008; Skene *et al.*; Yasui *et al.*, 2007; Young *et al.*, 2005).

The identity of MeCP2 as a multifunctional protein is a result of using a variety of techniques from numerous cellular contexts, however only the biochemical characteristics of the endogenous MeCP2 protein in the brain are relevant to RTT. Many important protein-protein interactions have been identified by biochemical co-purification, notably, the first biological function for MeCP2 as a transcriptional repressor was established by biochemical purification of a MeCP2 complex containing Sin3A-HDAC from *Xenopus* oocytes (Jones *et al.*, 1998). Despite a large body of work suggesting MeCP2 interacts with many partners, the only published attempt to purify native MeCP2 complexes from mammalian brain resulted in a conclusion that MeCP2 purifies in a single biochemical pool that does not stably associate with other proteins (including itself) (Klose and Bird, 2004).

From work described in Chapter 2, we begin to reveal molecular mechanisms of MeCP2 function relevant to RTT through the identification of multiple distinct brain-derived MeCP2 protein pools, agreeing with the structural and functional data proposing multiple roles for MeCP2 (Hite *et al.*, 2009). A previous study indicates no brain derived MeCP2 protein interacts with MonoQ strong anion resin (Klose and Bird, 2004). This is a reasonable outcome for a basic protein (pI=10) unassociated with other factors. We

have confirmed that brain derived MeCP2 which is unable to bind the MonoQ (QF) is unassociated with other proteins. Recent *in vitro* work shows MeCP2 can form DNA-MeCP2-DNA bridges, as well as directly compact chromatin without DNA methylation, ATP, or other proteins (Georgel *et al.*, 2003; Nikitina *et al.*, 2007). What's more, a large portion of MeCP2 is observed at pericentric heterochromatin (Nan *et al.*, 1993), and MeCP2 has been shown to consolidate chromocenters during cellular differentiation (Brero *et al.*, 2005). Aside from extraction disrupted complex MeCP2, it is possible brain derived MeCP2 purified in the QF pool functions as a monomer in nuclear architecture.

In our lab, a distinguishing characteristic over other labs results is that we observe distinct pools of MeCP2 with differential anion resin affinities present in mammalian brain extracts. Furthermore, we found this interaction is not due to MeCP2 phosphorylation or nucleic acid mediated binding, suggesting the QB MeCP2 association is from negatively charged protein partners facilitating the interactions. We have further characterize one biochemically distinct pool of MeCP2 that is found to co-fractionate with Prpf3, a known spliceosome-associated protein (Wang *et al.*, 1997), Sdccag1 (Scanlan *et al.*, 1998), a mediator of nuclear export (Bi *et al.*, 2005), ATP-binding cassette 50, and 3 components of a translation initiation complex (Eif2 subunits 1, 2, and 3). We find MeCP2 specifically interacts with Prpf3 and Sdccag1, but not other copurifying candidates.

MeCP2's association with Prpf3, a major component of the spliceosome, supports MeCP2 as having a role in modulating mRNA splicing; exactly what that role might be is still not clear. Interestingly, a recent study suggests that differential gene body methylation may play a role in transcript splicing (Laurent *et al.*). By identifying the

genome wide methylation status in multiple human cell types, the group found that exons were more highly methylated than introns, and that there were sharp transitions of methylation at exon-intron boundaries (Laurent *et al.*). Based on what is known about MeCP2 and Prpf3 independently, potentially MeCP2's direct interaction with Prpf3 could function in splice site selection. Prpf3 is one of multiple associated proteins of the U4/U6.U5 tri-snRNP complex that is recruited to the splice site to form and stabilize the functional spliceosome, and is essential for pre-mRNA splicing (Brown and Beggs, 1992; Horowitz *et al.*, 1997; Lauber *et al.*, 1997; Wang *et al.*, 1997). Importantly, MeCP2 can bind RNA directly (Jeffery and Nakielny, 2004), and has been implicated as a regulator of alternative splicing in the HeLa and Neuro2A cell lines through an RNA-dependent interaction with YB-1 (Young *et al.*, 2005), a protein known to participate in splicing of mRNAs (Philips *et al.*, 1998; Stickeler *et al.*, 2001). While we were unable to detect YB-1 as a component of the MeCP2/Prpf3/Sdccag1 complex in brain, it is inefficient to inhibit the highly active and stable RNases that would abolish any YB-1/RNA/MeCP2 interaction during purification so we cannot rule out that YB-1 could be part of a larger RNA-dependent alternative splicing complex. Nevertheless, a direct interaction of MeCP2 and Prpf3 further supports a role for MeCP2 in mRNA splicing regulation.

MeCP2 is present at the promoter of many actively transcribed genes (Yasui *et al.*, 2007), and activates a majority of genes it regulates in the mouse hippocampus (Chahrour *et al.*, 2008). MeCP2's ability to bind chromatin *in vivo* appears to be dynamically regulated by phosphorylation in the brain (Chen *et al.*, 2003; Tao *et al.*, 2009). Experimentally, MeCP2 is known to bind RNA *in vitro* with a similar affinity as to methylated DNA and that the two activities are mutually exclusive (Jeffery and Nakielny,

2004). Thus, MeCP2 at an active promoter or within the body of a transcribed gene would be well positioned spatially upon its release to interact with the transcripts of the genes it activates and influence splice site selection. Therefore, it is reasonable to propose that MeCP2 activated genes would also be targets for splicing regulation by a MeCP2 containing complex. Supporting this model, RNA transcripts from *Cdk10*, a gene positively regulated by MeCP2 binding at its promoter (Chahrour *et al.*, 2008) and abnormally spliced in *Mecp2*$^{308/Y}$ mice (Young *et al.*, 2005), were associated with both MeCP2 and Prpf3 *in vivo*.

Sdccag1 was originally identified from colon cancer patients by a serological analysis of recombinant cDNA expression libraries (SEREX) (Scanlan *et al.*, 1998) and later identified as a tumor suppressor by its ability to cause cell cycle arrest in a non-small-cell lung cancer cell line (Carbonnelle *et al.*, 2001). The significance of finding Sdccag1 as part of the MeCP2-Prpf3 complex can only be implied due to the limited information on its function in vertebrates and *Drosophila*. Bioinformatically the Sdccag1 protein contains a predicted RNA-binding domain homologous to a eukaryotic small nuclear ribonucleoprotein (Marchler-Bauer *et al.*, 2009), suggesting the capacity to function in mRNA splicing. Functionally, the *Drosophila* Sdccag1 homolog Caliban, has been shown to interact with and mediate the nuclear export of the Prospero homeodomain transcription factor (Prox in mammals) and this interaction and function is conserved in mammalian cells (Bi *et al.*, 2005). Notably, *Mecp2*$^{308/Y}$ mice, which produce a truncated form of MeCP2 and reproduce many of the classical features of RTT (Moretti *et al.*, 2005; Shahbazian *et al.*, 2002), have been shown to have multiple genes that are abnormally spliced in the brain (Young *et al.*, 2005). This suggests the C-terminal

portion of MeCP2, which we have identified as the putative Sdccag1 interaction domain, plays a critical role in regulating alternative splicing.

It is tempting to speculate that the MeCP2/Prpf3/Sdccag1 complex may not only be involved in splicing mRNA, but also transporting mRNAs to the cytoplasm for translation. Consistent with this model, MeCP2 has been shown to have both a nuclear and cytoplasmic localization in neuronal cell lines (Miyake and Nagai, 2007). This idea raises the possibility that other co-purifying proteins identified may indeed be part of the MeCP2 complex. ABC50 has been shown to interact with the Eif2 complex (Tyzack *et al.*, 2000), and Eif2 is known to play an important role in the regulation of translation initiation (Pain, 1996; Price *et al.*, 1994). Although MeCP2 or Prpf3 did not directly interact with any of these proteins in our study, an untested Sdccag1 could function as a scaffolding protein, linking all of the identified proteins. Theoretically, this would result in a complex that could function in mRNA splicing, export of RNA to the cytoplasm, and ultimately translation regulation.

Structural studies indicate that MeCP2 contains at least six structurally distinct domains (Adams *et al.*, 2007). Mutations generally manifested as point mutations resulting in a single missense or nonsense amino acid change have been identified in all domains of *MECP2* from RTT patients, suggesting that all domains are critical for MeCP2 function (Hite *et al.*, 2009; Moore *et al.*, 2005). Although all these mutations produce clinical RTT pathology, certain mutations are more strongly associated with particular symptoms and disease severity suggesting that all RTT mutations in *MECP2* are not equal, with some potentially being more disruptive towards MeCP2's many functions (Jian *et al.*, 2005; Neul *et al.*, 2008). Considering the range of mutations

throughout *MECP2* in RTT, it is not surprising that many RTT mutations are within, or predicted to affect both the Prpf3 and Sdccag1 binding domains, disrupting the MeCP2/Prpf3/Sdccag1 complex and presumably MeCP2-mediated splicing regulation and mRNA transport.

Current literature suggests many roles for MeCP2. A hypothetical model using information available easily allows for integration of a MeCP2/Prpf3/Sdccag1 complex into suspected roles of MeCP2 (Figure 3.1). In this model, MeCP2 is associated with transcriptionally active promoters (Chahrour *et al.*, 2008; Yasui *et al.*, 2007), and at some targets, activation is dependent on neuronal stimulation leading to subsequent phosphorylation of MeCP2 (Chen *et al.*, 2003; Martinowich *et al.*, 2003; Zhou *et al.*, 2006). Phosphorylation results in release of MeCP2 from target promoters (Chen *et al.*, 2003; Zhou *et al.*, 2006) strategically localizing the protein spatially and temporally to interact with mRNA (Jeffery and Nakielny, 2004). Phosphorylation of MeCP2 and interaction with mRNA prevents MeCP2 from rebinding methylated chromatin (Chen *et al.*, 2003; Jeffery and Nakielny, 2004) and could allow for recruitment of splicing factors such as Prpf3, Sdccag1 and other spliceosomal components to modulate splicing (Young *et al.*, 2005). Interaction of a MeCP2/Prpf3/Sdccag1 complex with mRNA could also later function in mRNA transport to the cytoplasm, a known destination of MeCP2 in neurons (Miyake and Nagai, 2007), and possibly target ribosomes for translation.

With MeCP2 continuing to emerge as a multifunctional protein and its biological role in RTT unclear, identifying biochemically distinct pools of Mecp2 in the brain containing novel MeCP2 interacting proteins is a valuable tool towards an understanding MeCP2 function and its dysfunction in RTT. Work in Chapter 2 shows there are at least

four distinct pools of MeCP2 in mammalian brain. Interestingly, data suggests MeCP2 has the capacity to function in four ways: repression, activation, nuclear structure, and mRNA splicing. We have characterized two biochemically distinct pools of MeCP2, one confirming MeCP2 present as a monomer (possibly involved in nuclear structure), and another in a novel punitive mRNA splicing complex. The complexity of RTT likely stems from perturbation of multiple MeCP2 functions. However, these results add to the mounting evidence indicating that one such critical function of MeCP2 in the brain involves RNA biogenesis, and adds to a growing list of possible mechanisms perturbed in RTT.

REFERENCES

Adams, V.H., McBryant, S.J., Wade, P.A., Woodcock, C.L., and Hansen, J.C. (2007) Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J Biol Chem* **282**: 15057-15064.

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.

Bi, X., Jones, T., Abbasi, F., Lee, H., Stultz, B., Hursh, D.A., and Mortin, M.A. (2005) Drosophila caliban, a nuclear export mediator, can function as a tumor suppressor in human lung cancer cells. *Oncogene* **24**: 8229-8239.

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, J.D., and Beggs, J.D. (1992) Roles of PRP8 protein in the assembly of splicing complexes. *EMBO J* **11**: 3721-3729.

Carbonnelle, D., Jacquot, C., Lanco, X., Le Dez, G., Tomasoni, C., Briand, G., Tsotinis, A., Calogeropoulou, T., and Roussakis, C. (2001) Up-regulation of a novel mRNA (NY-CO-1) involved in the methyl 4-methoxy-3-(3-methyl-2-butenoyl) benzoate (VT1)-induced proliferation arrest of a non-small-cell lung carcinoma cell line (NSCLC-N6). *Int J Cancer* **92**: 388-397.

Carney, R.M., Wolpert, C.M., Ravan, S.A., Shahbazian, M., Ashley-Koch, A., Cuccaro, M.L., Vance, J.M., and Pericak-Vance, M.A. (2003) Identification of MeCP2 mutations in a series of females with autistic disorder. *Pediatr Neurol* **28**: 205-211.

Chadwick, L.H., and Wade, P.A. (2007) MeCP2 in Rett syndrome: transcriptional repressor or chromatin architectural protein? *Curr Opin Genet Dev* **17**: 121-125.

Chahrour, M., and Zoghbi, H.Y. (2007) The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56**: 422-437.

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

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* **302**: 885-889.

Christodoulou, J., and Weaving, L.S. (2003) MECP2 and beyond: phenotype-genotype correlations in Rett syndrome. *J Child Neurol* **18**: 669-674.

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.

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.

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

Hite, K.C., Adams, V.H., and Hansen, J.C. (2009) Recent advances in MeCP2 structure and function. *Biochem Cell Biol* **87**: 219-227.

Horowitz, D.S., Kobayashi, R., and Krainer, A.R. (1997) A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *RNA* **3**: 1374-1387.

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.

Jian, L., Archer, H.L., Ravine, D., Kerr, A., de Klerk, N., Christodoulou, J., Bailey, M.E., Laurvick, C., and Leonard, H. (2005) p.R270X MECP2 mutation and mortality in Rett syndrome. *Eur J Hum Genet* **13**: 1235-1238.

Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**: 187-191.

Klose, R.J., and Bird, A.P. (2004) MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J Biol Chem* **279**: 46490-46496.

Lam, C.W., Yeung, W.L., Ko, C.H., Poon, P.M., Tong, S.F., Chan, K.Y., Lo, I.F., Chan, L.Y., Hui, J., Wong, V., Pang, C.P., Lo, Y.M., and Fok, T.F. (2000) Spectrum of mutations in the MECP2 gene in patients with infantile autism and Rett syndrome. *J Med Genet* **37**: E41.

Lauber, J., Plessel, G., Prehn, S., Will, C.L., Fabrizio, P., Groning, K., Lane, W.S., and Luhrmann, R. (1997) The human U4/U6 snRNP contains 60 and 90kD proteins that are structurally homologous to the yeast splicing factors Prp4p and Prp3p. *RNA* **3**: 926-941.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., and Wei, C.L. Dynamic changes in the human methylome during differentiation. *Genome Res* **20**: 320-331.

Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., and Bryant, S.H. (2009) CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* **37**: D205-210.

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

Meehan, R.R., Lewis, J.D., and Bird, A.P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res* **20**: 5085-5092.

Miyake, K., and Nagai, K. (2007) Phosphorylation of methyl-CpG binding protein 2 (MeCP2) regulates the intracellular localization during neuronal cell differentiation. *Neurochem Int* **50**: 264-270.

Moore, H., Leonard, H., de Klerk, N., Robertson, I., Fyfe, S., Christodoulou, J., Weaving, L., Davis, M., Mulroy, S., and Colvin, L. (2005) Health service use in Rett syndrome. *J Child Neurol* **20**: 42-50.

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.

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., 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.

Neul, J.L., Fang, P., Barrish, J., Lane, J., Caeg, E.B., Smith, E.O., Zoghbi, H., Percy, A., and Glaze, D.G. (2008) Specific mutations in methyl-CpG-binding protein 2 confer different severity in Rett syndrome. *Neurology* **70**: 1313-1321.

Nikitina, T., Shi, X., Ghosh, R.P., Horowitz-Scherer, R.A., Hansen, J.C., and Woodcock, C.L. (2007) Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Mol Cell Biol* **27**: 864-877.

Pain, V.M. (1996) Initiation of protein synthesis in eukaryotic cells. *Eur J Biochem* **236**: 747-771.

Philips, A.V., Timchenko, L.T., and Cooper, T.A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**: 737-741.

Price, N.T., Francia, G., Hall, L., and Proud, C.G. (1994) Guanine nucleotide exchange factor for eukaryotic initiation factor-2. Cloning of cDNA for the delta-subunit of rabbit translation initiation factor-2B. *Biochim Biophys Acta* **1217**: 207-210.

Rett, A. (1966) [On a unusual brain atrophy syndrome in hyperammonemia in childhood]. *Wien Med Wochenschr* **116**: 723-726.

Scanlan, M.J., Chen, Y.T., Williamson, B., Gure, A.O., Stockert, E., Gordan, J.D., Tureci, O., Sahin, U., Pfreundschuh, M., and Old, L.J. (1998) Characterization of human colon cancer antigens recognized by autologous antibodies. *Int J Cancer* **76**: 652-658.

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

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

Stickeler, E., Fraser, S.D., Honig, A., Chen, A.L., Berget, S.M., and Cooper, T.A. (2001) The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. *EMBO J* **20**: 3821-3830.

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

Tyzack, J.K., Wang, X., Belsham, G.J., and Proud, C.G. (2000) ABC50 interacts with eukaryotic initiation factor 2 and associates with the ribosome in an ATP-dependent manner. *J Biol Chem* **275**: 34131-34139.

Wang, A., Forman-Kay, J., Luo, Y., Luo, M., Chow, Y.H., Plumb, J., Friesen, J.D., Tsui, L.C., Heng, H.H., Woolford, J.L., Jr., and Hu, J. (1997) Identification and characterization of human genes encoding Hprp3p and Hprp4p, interacting components of the spliceosome. *Hum Mol Genet* **6**: 2117-2126.

Watson, P., Black, G., Ramsden, S., Barrow, M., Super, M., Kerr, B., and Clayton-Smith, J. (2001) Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein. *J Med Genet* **38**: 224-228.

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

Ylisaukko-Oja, T., Rehnstrom, K., Vanhala, R., Kempas, E., von Koskull, H., Tengstrom, C., Mustonen, A., Ounap, K., Lahdetie, J., and Jarvela, I. (2005) MECP2 mutation analysis in patients with mental retardation. *Am J Med Genet A* **132A**: 121-124.

Young, J.I., Hong, E.P., Castle, J.C., Crespo-Barreto, J., Bowman, A.B., Rose, M.F., Kang, D., Richman, R., Johnson, J.M., Berget, S., and Zoghbi, H.Y. (2005) Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc Natl Acad Sci U S A* **102**: 17551-17558.

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

Zoghbi, H.Y. (2005) MeCP2 dysfunction in humans and mice. *J Child Neurol* **20**: 736-740.

FIGURES

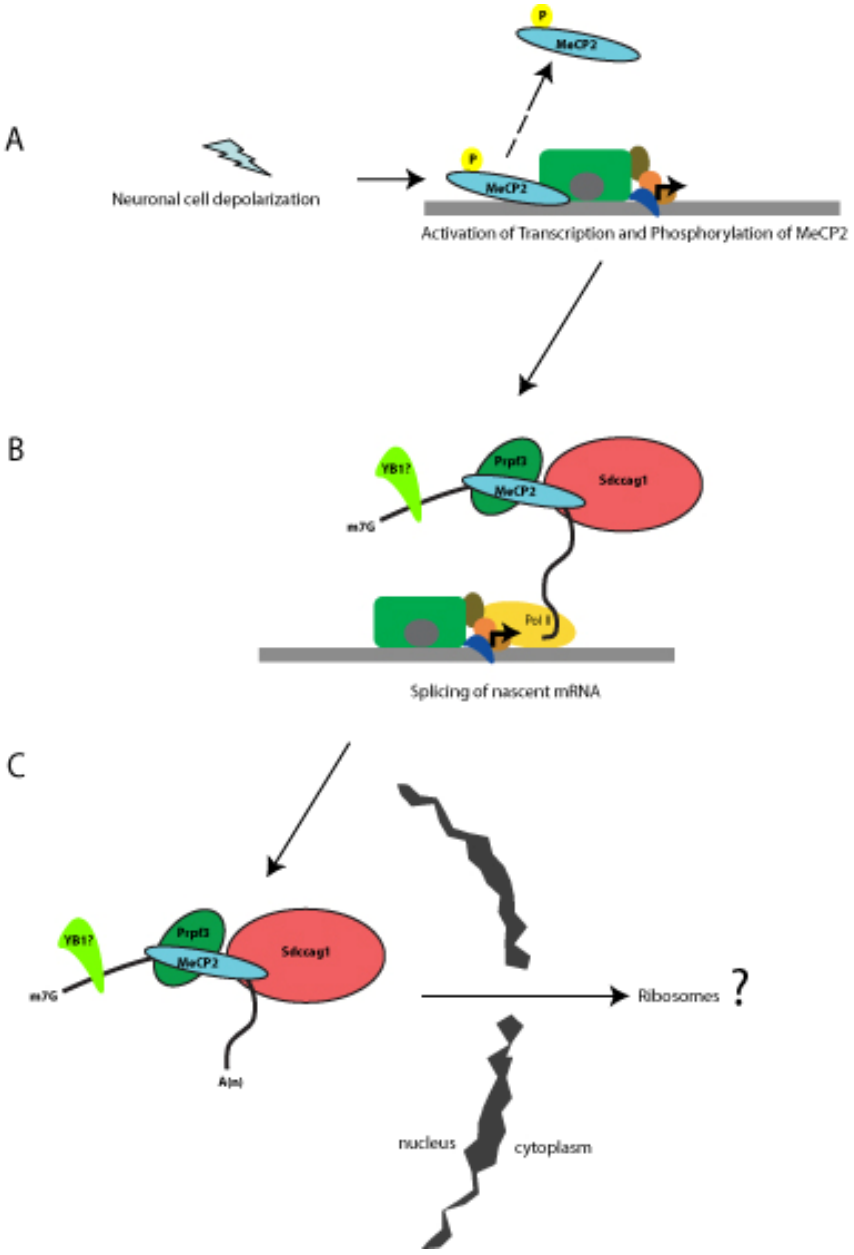

Figure 3.1: Model for the brain derived MeCP2 mRNA splicing complex. A) Neuronal depolarization results in transcriptional activation and subsequent MeCP2 phosphorylation. B) A MeCP2 containing splicing complex interacts with nascent mRNA. C) Sdccag1 mediated export of MeCP2 containing mRNA complex.

APPENDIX A*

DEVELOPMENTAL CONTROL OF DNA METHYLATION MEDIATED SILENCING

***Data is included in a manuscript by Gert Jan Veenstra, (*in preparation*).**

INTRODUCTION

Many genes and a majority of non-protein coding DNA in a eukaryotic genome is silenced in any particular cell during the organism's life (Elgin and Grewal, 2003). Transcriptional silencing is associated with the formation of heterochromatic domains and is influenced by covalent modifications of chromatin components. These modifications are essential for normal development and can be re-established after each cell division to maintain the stability of the genome (Bird, 2002; Jenuwein and Allis, 2001; Kouzarides, 2002; Turner, 2000). DNA methylation in the vertebrate genome is one such epigenetic mark generally associated with heterochromatic silencing.

DNA Methylation in vertebrates is characterized by the post-replication addition of a methyl group to the 5 carbon of cytosine (5mC) resulting in symmetrically methylated CpG dinucleotides carried out by a member of the DNA methylatransferase (DNMT) family of proteins. Family members catalyze the transfer of a methyl-group from S-adenosyl methionine (SAM) to the C5 position of cytosine.

DNA methylation in mammials is established on the genome through a sequence of developmental events (Razin and Shemer, 1995). Pre-implantation mouse embryos undergo a global demethylation event followed by *de novo* methylation by DNA methyltransferase 3A (DNMT3A) and DNMT3B, which is critical in determining

somatic DNA methylation patterns (Hsieh, 1999; Jaenisch *et al.*, 1982; Okano *et al.*, 1998; Okano *et al.*, 1999). After initial methylation patterns are established, they are maintained between cell generations by the maintenance methyltransferase DNMT1 (Bestor *et al.*, 1988; Leonhardt *et al.*, 1992).

Studies suggest DNA methylation is important in establishing and maintaining heterochromatic structures for gene regulation (Eden *et al.*, 2003; Gaudet *et al.*, 2003). Transient transfection and *in vitro* transcription assays consistently demonstrate that the repression of transcription is dependent on DNA methylation (Levine *et al.*, 1993; Murray and Grosveld, 1987). However, by comparing transcription rates of methylated and unmethylated stably integrated transgenes during *Xenopus* embryo development, we have evidence suggesting the repressive nature of DNA methylation may be developmentally controlled.

RESULTS

To determine the effects of DNA methylation on transcription during development of a vertebrate, the plasmid pGL3-CMV, containing a CMV promoter driving a luciferase gene, were differentially methylated *in vitro* with the bacterial methylase M.Sss I. Methylation status of the plasmids was tested by methyl-sensitive restriction digestion using BstUI after linearization (Figure A.1). Unmethylated plasmids are cleaved into several pieces, while methylated plasmid is comparable to the linear control, indicating it is fully methylated and not able to be cut by BstUI (Figure A.1). Linearized plasmids were subsequently used for *Xenopus* transgenesis.

The model organism *Xenopus laevis* allows for exogenous DNA to be stably incorporated into its genome (Amaya and Kroll, 1999). Transgenic *Xenopus* were produced with either methylated or unmethylated pGL3-CMV, and embryos developing normally were collected at: stages 9 (n = 45), stage 11 (n = 45), stage 15 (n = 28), or stage 21 (n = 20). RNA and DNA were purified from the various stages of unmethylated and methylated pGL3-CMV transgenic embryos and relative transcript levels were determined by reverse transcription quantitative PCR (RT-qPCR), normalized to integrated plasmid DNA. Stage 9 and 11 methylated pGL3-CMV embryos show substantial levels of active transcription compared, although lower than unmethylated pGL3-CMV embryos (Figure A.2). By stage 15, methylated pGL3-CMV embryos show a large drop in transcription compared to unmethylated pGL3-CMV embryos, and by stage 21 methylated pGL3-CMV transcript levels do not reach threshold, indicating it is strongly repressed (Figure A.2). We conclude DNA methylation associated repression of stably integrated plasmids is developmentally regulated.

DISCUSSION

The most common modification of vertebrate genomes, DNA methylation, is primarily associated with transcriptional repression (Ballestar and Wolffe, 2001). Global levels of DNA methylation vary not only between species, but different stages of embryonic development for some. In mice, embryos undergo global (but not total) demethylation before implantation, followed by *de novo* methylation to establish somatic DNA methylation patterns, possibly re-programming epigenetic mediated repression

(Hsieh, 1999; Jaenisch *et al.*, 1982; Okano *et al.*, 1998; Okano *et al.*, 1999). However, in *Xenopus* there is no genome-wide demethylation and re-methylation event (Veenstra and Wolffe, 2001). Oddly, genes with heavily methylated promoters have been found to be actively transcribed early in *Xenopus* embryo development, (Gert Jan Veenstra – personal communication). Initial studies suggest DNA methylation is important in establishing and maintaining heterochromatic structures for gene regulation (Eden *et al.*, 2003; Gaudet *et al.*, 2003). Mammalian cell culture experiments utilizing transient transfection and *in vitro* transcription assays consistently show repression of transcription is dependent on DNA methylation (Levine *et al.*, 1993; Murray and Grosveld, 1987). Also, before transcriptional senescence of mature eggs, *Xenopus* oocytes are able to strongly repress methylated plasmids (Kass *et al.*, 1997). Furthermore, depletion of DNA methyltransferases lead to premature onset of embryonic gene transcription in *Xenopus* embryos, indicating DNA methylation can be repressive during early development (Stancheva and Meehan, 2000). This raises questions as to why some heavily methylated promoters are observed active early in *Xenopus* embryo development. Is it merely due to unmethylated copies of the tetraploid *Xenopus* genome, or associated with developmentally regulated factors?

We have investigated the transcriptional activity of methylated versus unmethylated plasmid constructs stably integrated in developing *Xenopus* embryos. Completely methylated plasmids injected into *Xenopus* oocytes are strongly repressed after "chromatization" has taken place on the transient DNA molecules (Kass *et al.*, 1997). The *Xenopus* transgenesis procedure allows for exogenous DNA to be stably incorporated into the genome (Amaya and Kroll, 1999). Upon fertilization and

subsequent cell divisions, integrated transgenes likely contain the normal complement of chromatin factors found elsewhere in the genome. It is reasonable to think that transgene DNA methylated *in vitro* and subsequently integrated into the genome would be potently repressed as it is the case for injected plasmids. However, we find here that DNA methylation does not strongly silence transcription of the CMV promoter during early embryogenesis (stages 9 - 11), and we do not see complete transgene repression until neural tube formation begins (stages 15 – 21). This would suggest that although DNA methylation marks are present, the repressive effect associated is developmentally regulated. Xenopus does not have genomic imprinting and does not have genome wide demethylation events during development (Veenstra and Wolffe, 2001). This could suggest that genome wide demethylation events observed in mammals are strictly imprinting related. The complete role of DNA methylation remains a mystery, however, in time continuing research will help further understanding the complex nature of epigenetics and its related diseases.

MATERIALS AND METHODS

Plasmid constructs and *in vitro* methylation

CMV promoter of pEGFP (Clontech) was PCR amplified using primers: 5'-GGTACCTAGTTATTAATAGTAATCAATTAC-3' 5'-CTCGAGGTGGCGACCGGTAGCGCTAGCGGA-3', cloned into pGEM T-Easy (Promega), sequenced and cloned into KpnI and XhoI of pGL3 (Promega). Methyl pGL3-CMV was methylated *in vitro* with the bacterial methylase M.Sss I (New England

Biolabs) by adding 160 µM *S*-Adenosyl methionine (SAM) and 1U M.Sss I/µg plasmid and incubating for 1 hour at 37°C, followed by a second addition of M.Sss I and SAM and additional 1 hour at 37°C. Plasmid methylation status is tested by methyl-sensitive restriction digest using with 20U BstUI/µg linear pGL3-CMV for 1 hour at 60°C.

Generation of *Xenopus* transgenics

Procedure utilizing *Xenopus laevis* were purchased from Xenopus Express, Inc. and carried out in accordance with established UIUC IACUC approved protocols for animal welfare. *Xenopus* transgenesis was carried out essentially as described (Kroll and Amaya, 1996, Development 122(10):3173–3183), however, the sperm nuclei were not digested with restriction enzymes. Briefly, *Xenopus* high-speed egg extract (EXT) and sperm head nuclei were purified, and stored at −80°C in aliquots until use. The pGL3-CMV plasmid or methyl pGL3-CMV was linearized with Alw NI, purified, and incubated along with linear pEGFP-N1(100 ng each per experiment) and sperm nuclei for 5 min. EXT (10 µl) and 20 µl Sperm dilution buffer (SDB = 250 mM sucrose, 75 mM KCl, 0.5 mM spermidine, 0.2 mM spermine) were mixed and incubated at 65°C for 5 min, then centrifuged at 16,000 × g for 3 min to remove precipitate. The soluble fraction (6 µl) was diluted to 22 µl with SDB plus 10 mM $MgCl_2$. This activating solution was added to the nuclei/sperm mix and incubated for 15 min at room temperature. The swollen nuclei (with integrated transgenes) were gently added to 170 µl SDB and used for micro-injection at a rate of 0.586 µl/min using a microliter syringe pump (Harvard Apparatus). Normally developing embryos were collected at various stages, scored for percent transgenic by EGFP fluorecence and frozen in Trizol reagent (Invitrogen).

Quantitative reverse transcription PCR (qRT-PCR)

For each sample, total RNA and genomic DNA were purified from embryos using Trizol reagent per manufacture's protocol. RNA was then treated with 1U RQ1 DNase (Promega) per 1ug RNA for 30 min. at 37°C. cDNA synthesis was preformed using 1ug of total RNA, 50ng of random hexamer, and Superscript III (Invitrogen) per manufacture's suggested method. Relative transcript levels of integrated methylated pGL3-CMV luciferase to unmethylated pGL3-CMV luciferase construct was determined using 1ul of (1:20 diluted) cDNA (in triplicate), iQ SYBR Supermix, and luciferase transgene specific primers (5'-GAATCCATCTTGCTCCAACAC-3' and 5'-TTCGTCCACAAACACAACTC-3') on a Bio-Rad iCycler IQ machine. Relative transcript levels were normalized to integrated transgene copy number in genomic DNA. Experiments were preformed at least 3 times. Data analyses were preformed using the comparative Ct method and error bars are $\pm$ standard error of the mean.

REFERENCES

Amaya, E., and Kroll, K.L. (1999) A method for generating transgenic frog embryos. *Methods Mol Biol* 97: 393-414.

Ballestar, E., and Wolffe, A.P. (2001) Methyl-CpG-binding proteins. Targeting specific gene repression. *Eur J Biochem* 268: 1-6.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* 203: 971-983.

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

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

Elgin, S.C., and Grewal, S.I. (2003) Heterochromatin: silence is golden. *Curr Biol* 13: R895-898.

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

Hsieh, C.L. (1999) In vivo activity of murine de novo methyltransferases, Dnmt3a and Dnmt3b. *Mol Cell Biol* 19: 8211-8218.

Jaenisch, R., Harbers, K., Jahner, D., Stewart, C., and Stuhlmann, H. (1982) DNA methylation, retroviruses, and embryogenesis. *J Cell Biochem* 20: 331-336.

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

Kass, S.U., Landsberger, N., and Wolffe, A.P. (1997) DNA methylation directs a time-dependent repression of transcription initiation. *Curr Biol* 7: 157-165.

Kouzarides, T. (2002) Histone methylation in transcriptional control. *Curr Opin Genet Dev* 12: 198-209.

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

Levine, A., Yeivin, A., Ben-Asher, E., Aloni, Y., and Razin, A. (1993) Histone H1-mediated inhibition of transcription initiation of methylated templates in vitro. *J Biol Chem* 268: 21754-21759.

Murray, E.J., and Grosveld, F. (1987) Site specific demethylation in the promoter of human gamma-globin gene does not alleviate methylation mediated suppression. *Embo J* 6: 2329-2335.

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

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99: 247-257.

Razin, A., and Shemer, R. (1995) DNA methylation in early development. *Hum Mol Genet* 4 Spec No: 1751-1755.

Stancheva, I., and Meehan, R.R. (2000) Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes Dev* 14: 313-327.

Turner, B.M. (2000) Histone acetylation and an epigenetic code. *Bioessays* 22: 836-845.

Veenstra, G.J., and Wolffe, A.P. (2001) Constitutive genomic methylation during embryonic development of Xenopus. *Biochim Biophys Acta* 1521: 39-44.

FIGURES

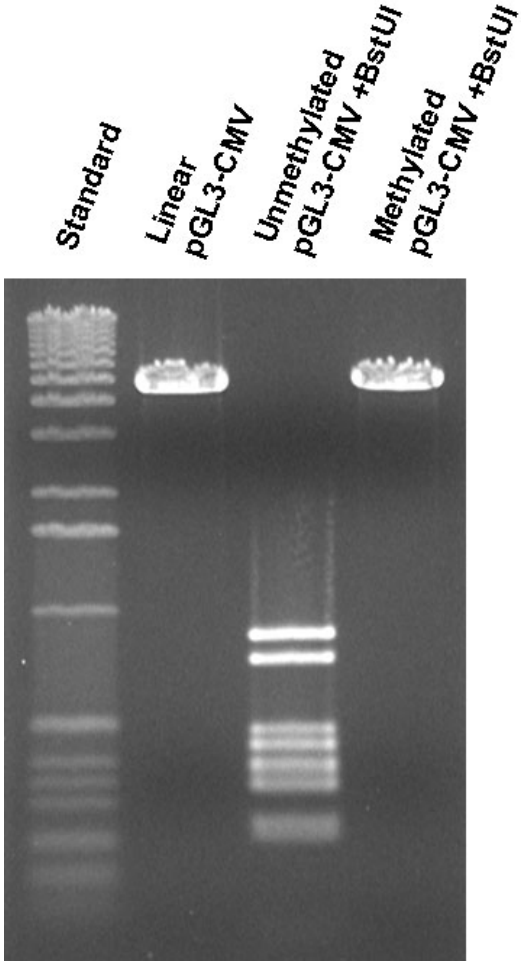

Figure A.1: Methylation status determination by methyl sensitive restriction digest. Methyl pGL3-CMV is completely methylated. Linear pGL3-CMV is sensitive to BstUI digestion, while linear methylated pGL3-CMV is not.

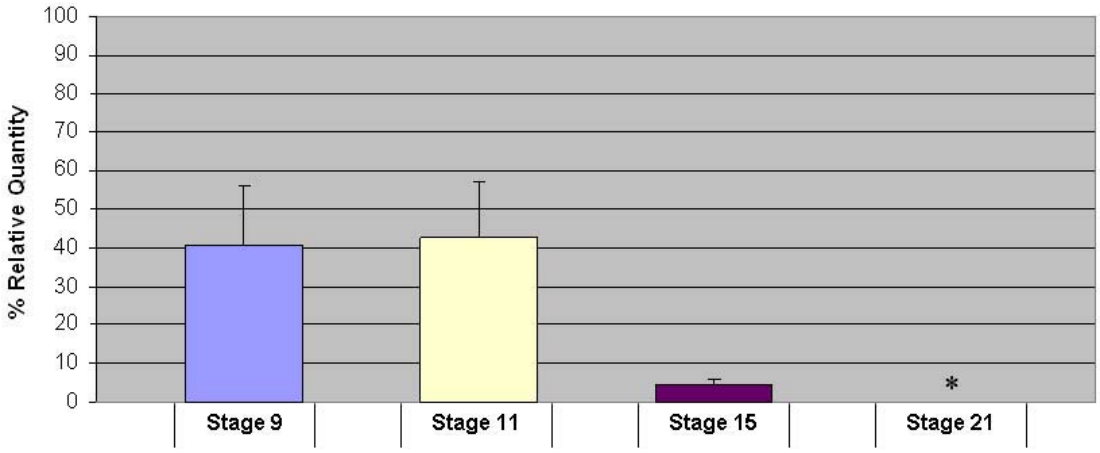

Figure A.2: Transcriptional activity of transgenically integrated methylated pGL3-CMV versus unmethylated pGL3-CMV promoters in *Xenopus laevis* during early embryogenesis. Completely methylated pGL3-CMV is transcriptionally active in transgenic *Xenopus* embryos during early development, but rapidly silenced compared to unmethylated pGL3-CMV. Percent relative quantity of methylated pGL3-CMV luciferase transcripts are normalized to genomic transgene copies and compared to unmethylated pGL3-CMV luciferase transcripts at a relative level of 100%; error bars are ±standard error of the mean. Methylated pGL3-CMV qRT-PCR samples that did not reach threshold levels are indicated by (*).

APPENDIX B*

TESTING THE EFFECTS OF FSHD CANDIDATE GENE EXPRESSION IN VERTEBRATE MUSCLE DEVELOPMENT

***Data and writing is adapted from:**

S.W. Long, R.D. Wuebbles, M.L. Hanel, and P.L. Jones. *Testing the effects of FSHD candidate gene expression on vertebrate muscle development* **(Int J Clin Exp Pathol, 2010;3(4):386-400)**

INTRODUCTION

FSHD is now recognized as one of the most prevalent forms of muscular dystrophy in adults (http://www.orpha.net). Prominent features of this myopathy are the progressive weakening of the skeletal muscles in the face, shoulder girdle, and the upper arms, and these muscular aspects are often combined (>50% of patients) with retinal vasculopathy (Fitzsimons *et al.*, 1987; Padberg *et al.*, 1995). The genetic lesion leading to the most prominent form of FSHD (FSHD1A), accounting for ~98% of FSHD patients, is an autosomal dominant contraction of the D4Z4 repeat array at chromosome 4q35 below 11 copies (van Deutekom *et al.*, 1993; Wijmenga *et al.*, 1992). This contraction leads to hypomethylation of the D4Z4 repeats, which has been proposed to lead downstream to the misregulation of one or more of the 4q35 localized genes including *FRG1*, *ANT1*, *FRG2*, *DUX4*, and *DUX4c* (van Overveld *et al.*, 2003). However, none of these candidate genes has consistently been shown to exhibit significantly altered RNA expression levels in affected FSHD muscle biopsies compared to unaffected controls (Dixit *et al.*, 2007; Gabellini *et al.*, 2002; Gabellini *et al.*, 2006; Jiang *et al.*, 2003; Klooster *et al.*, 2009; Rijkers *et al.*, 2004; Snider *et al.*, 2009; Winokur *et al.*, 2003b).

Multiple issues complicate these expression analyses including large differences within an affected muscle and potentially at the site of biopsy, the bias focusing on FSHD gene misexpression exclusively in the skeletal muscle lineage, and the potential that FSHD gene misexpression occurs during cell differentiation (Barro *et al.*, 2008; Bodega *et al.*, 2009; Morosetti *et al.*, 2007; Rijkers *et al.*, 2004). Thus, without knowing when and where in human muscle development gene misexpression leading to FSHD occurs, the cause of the FSHD pathophysiology has remained controversial.

To circumvent the ambiguity of RNA expression analyses, we have taken a developmental approach to the problem by first addressing the normal function of an FSHD candidate gene during development and then assaying the effect of overexpression of an FSHD candidate gene on vertebrate development. The system for these studies is the early development of *Xenopus laevis*. Our initial analysis focused on understanding the function and expression of one candidate gene, *frg1* (Hanel *et al.*, 2009; Wuebbles *et al.*, 2009). *FRG1* is a highly conserved gene of unknown function that is overexpressed in FSHD patient derived myoblasts undergoing myogenic differentiation (Bodega *et al.*, 2009). These studies found that frg1 is required for the normal development of the vertebrate musculature and vasculature (Hanel *et al.*, 2009; Wuebbles *et al.*, 2009). Consistent with a role in FSHD pathology, systemically elevated levels of frg1 led to phenotypes specifically in the vertebrate musculature and vasculature which strongly correlated to the two most common symptoms of FSHD, dystrophic muscle and increased angiogenesis (Hanel *et al.*, 2009; Wuebbles *et al.*, 2009). Thus, developmentally, FRG1 overexpression fits the criteria for being causal for FSHD pathology.

We have continued our analysis with three additional FSHD candidate genes, *DUX4*, *DUX4c*, and *PITX1* (Ansseau *et al.*, 2009; Dixit *et al.*, 2007; Kowaljow *et al.*, 2007). *DUX4* and *DUX4c* are encoded within open reading frames (ORFs) of different 4q35 D4Z4 repeat units within or near the FSHD deletion (Gabriels *et al.*, 1999; Wright *et al.*, 1993). Although D4Z4 repeat arrays exist in multiple loci in the genome (Winokur *et al.*, 1996), RNAs originating specifically from the 4q35 localized D4Z4/DUX4 and D4Z4/DUX4c loci are increased in certain FSHD patient-derived muscle cells (Ansseau *et al.*, 2009; Dixit *et al.*, 2007; Kowaljow *et al.*, 2007). A normal cellular or developmental role for the 4q35 DUX4 protein, if any, has not been described; however, expression of the currently accepted 4q35 derived DUX4 protein is highly toxic to all cells leading to a rapid onset of apoptosis (Bosnakovski *et al.*, 2008b; Kowaljow *et al.*, 2007). This apoptotic effect of DUX4 expression is postulated to be from direct competition with the regulatory targets of PAX3/PAX7 and is inhibited by elevated expression of PAX3 or PAX7 (Bosnakovski *et al.*, 2008b). Interestingly, in a cell culture system, DUX4 has been shown to bind the promoter and activate expression of *PITX1*, a non-4q35 localized FSHD candidate gene whose expression has been found to be upregulated in FSHD muscle, providing an alternative mechanism for DUX4-mediated pathology (Dixit *et al.*, 2007). *DUX4c*, located within a partial D4Z4 unit 42 kb proximal to the FSHD-associated D4Z4 array, is identical to DUX4 through their N-terminal double homeobox domains however they have differing C-terminal amino acid sequences (Wright *et al.*, 1993). DUX4c expression has been detected in muscle cells where it is proposed to act as a myogenic regulator and inhibitor of myoblast differentiation (Ansseau *et al.*, 2009; Bosnakovski *et al.*, 2008a).

In this study, we assayed the effects of expression of human DUX4 and DUX4c, as well as the *X. laevis* ortholog of *PITX1* on early vertebrate development, with particular attention to muscle growth and differentiation. We show that DUX4 expression and pitx1 overexpression both lead to massive cellular loss that is not muscle specific. With DUX4 in particular the cellular loss occurred at extremely low expression levels and was cell-type independent indicating that this protein is highly toxic to all vertebrate cells and this toxic effect was not specific to muscle. DUX4c expression did not lead to any observable change in muscle development or differentiation or changes in the expression of the myogenic regulators myf5 or myoD in *Xenopus*. Contradictory to what has been reported in cell culture, we found that both DUX4 and DUX4c significantly reduced expression levels of pitx1 transcripts in our animal model. Together with our previous studies on frg1, this presents the first analysis for direct comparison of the effects of expression of the main FSHD candidate genes in a developing vertebrate system.

RESULTS

Expression of DUX4 and pitx1 lead to cellular loss while DUX4c expression has minimal effect on development

The early development of *X. laevis* was used as a model system to determine the effects of FSHD candidate genes DUX4, DUX4c, and pitx1 expression levels on vertebrate development. In normal human tissues, DUX4 and DUX4c expression is undetectable and the proteins are neither required for nor involved in any known normal cellular function. The human 4q35 DUX4 and DUX4c genes are not conserved outside

of their double homeobox domains and *Xenopus* do not possess any orthologs. Therefore, we assayed the effects of the presence of DUX4 or DUX4c during muscle development compared to a background of no expression. Conversely, the pitx1 transcription factor, as was the case with frg1, is highly conserved between mammals and *Xenopus* in protein sequence, function, and developmental expression (Chang *et al.*, 2001). For this gene, we assayed the effect of increasing its expression during development as well as expression outside its normal developmental profile.

To determine the effects of DUX4 and DUX4c expression, and elevated levels of pitx1 during development, one side of early four cell-stage *X. laevis* embryos were injected with the corresponding mRNAs as well as mRNA encoding the *EGFP* marker mRNA. The developing embryos resulted in animals overexpressing the desired proteins on the injected side (confirmed by fluorescence) while co-developing a control, uninjected side. To study any tissue-specific effects on muscle, titration experiments with *DUX4*, *pitx1*, and *DUX4c* mRNA were performed and the numbers of abnormally developing embryos were observed. Abnormal development was scored as any abnormality observed during development, regardless of severity. Similar to previous data from cultured myoblasts, we found that our initial injection level of 100 pg *DUX4* mRNA was highly toxic to the embryos (Figure B.1A & B.1K). Embryos were arrested by stage 9 with an overall apoptotic appearance throughout the entire embryo. This severe phenotype occurred in all embryos injected with 100 ng (n = 13), 50pg (n = 26), and 25 pg (n = 19) examined, although developmental arrest occurred slightly later with less *DUX4* mRNA (Figure B.1A & B.1K). This severe apoptotic phenotype resulting in early developmental arrest was never observed in EGFP injected (n = 119) or uninjected

(n = 600) controls (Figure B.1J & B.1K). Unlike 100 pg, injection of 10 pg *DUX4* mRNA (n = 426) resulted in toxicity generally restricted to the injected side 72 hours post-injection (Figure B.1B). The heavy cellular loss in the injected side led to a small embryo and eventual arrest, likely due to gastrulation or neurulation defects caused by DUX4 expression. Further reducing the level to 1 pg *DUX4* mRNA (n = 166) resulted in a low percentage of animals which were able to progress through and complete closure of their neural tubes (Figure B.1C); however, all 1 pg *DUX4* injected embryos tested showed massive apoptosis by TUNEL staining (Figure B.1D). A further 2-fold reduction to 0.5 pg *DUX4* mRNA (n = 213) resulted in a significant increase of normal developing embryos, suggesting the effect of DUX4 is either cellular toxicity or nothing (Figure B.1E). Even in these embryos, the fluorescence of the EGFP appeared patchy (Figure B.1F), suggesting that though they appeared to have progressed through development normally, DUX4 had still led to cellular loss. Therefore, consistent with other systems, DUX4 expression is extremely cytotoxic in developing *Xenopus* (Bosnakovski *et al.*, 2008b; Kowaljow *et al.*, 2007).

DUX4 is capable of inducing *PITX1* expression in mouse C2C12 cell culture suggesting PITX1 misexpression as a possible mechanism for DUX4-mediated pathology (Dixit *et al.*, 2007). To determine the effect of aberrantly induced pitx1 expression, we bypassed DUX4 expression and microinjected mRNA encoding *Xenopus* pitx1 into embryos as above. Similar to injections of *DUX4* mRNA, we found that both 500 pg (n=272) and 300pg (n=122) of *pitx1* mRNA led to early embryo arrest while injection of 150 pg *pitx1* (n=254) led cellular loss specifically on the injected side leading to a curled phenotype and gastrulation or neurulation defects (Figure B.1G & B.1K). Injections of

50 pg *pitx1* mRNA (n=122) had no obvious effects on development (Figure B.1H & B.1K). Thus, while pitx1 overexpression is cytotoxic, developing embryos are much more tolerant of pitx1 overexpression than DUX4 expression suggesting that DUX4 cytotoxicity is not mediated through the activation of *pitx1*.

The analysis of the third FSHD candidate gene, DUX4c, produced results that were a stark contrast to the effects of DUX4 and pitx1. Microinjections of as much as 1ng of DUX4c mRNA (n=477) only lead to a modest increase over uninjected background levels of developmental abnormalities, yet still far fewer and less in severity than seen with a 2000-fold lower amount of *DUX4* mRNA (Figure B.1I & B.1K). Because DUX4c shares an identical double homeobox domain with DUX4, the fact that a 2000-fold increase in mRNA produced only minor developmental problems indicates that a strictly competitive interaction with Pax3 and Pax7 for DNA regulatory targets is not responsible for the DUX4 phenotype.

Cytotoxicity mediated by DUX4 and pitx1 is not muscle specific

The symptoms associated with FSHD are primarily muscular and often combined with a less prominent vascular component. Taking into account the generally accepted model whereby the FSHD1A deletion leads to an epigenetic upregulation of gene expression, one would expect the effect of a viable candidate gene's systemic expression to be primarily seen in tissues affected in FSHD as is the case with *frg1*. In order to determine if muscle was specifically affected by DUX4, pitx1, or DUX4c expression, injected embryos were analyzed for differentiated muscle and neurologic tissue by immunostaining with the 12/101 or NCAM antibodies, respectively (Figure B.2 & B.3).

Due to the severe loss of tissue in the 1 pg *DUX4* and 150 pg *pitx1* injected animals, the figures depict some of the best developing and staining animals, as most failed to form a complete neural tube closure and were consequently too curled to capture the staining pattern in photographs.

Immunostained animals were qualitatively determined to exhibit normal, depleted, or absent levels of immunostaining on the mRNA injected side (Figure B.2). Embryos scored as depleted included those exhibiting highly dispersed but significant staining, missing somites and somite disruptions, and a significantly thinner somite area. DUX4 injected embryos were first analyzed for affects on the developing muscle. Qualitative inspection of 1 pg *DUX4* injected embryos stained with 12/101 (n = 45, missing n = 22, depleted n = 23) indicate a severe loss of muscle tissue by the lack of 12/101 immunostaining specifically on the injected side of the embryo (Figure B.2A' and B.2B' compared to Figure B.2A and B.2B). However, transverse paraffin sectioning revealed these animals had a weak appearance of staining due to a build up of cellular debris between the 12/101 stained muscle near the notochord/neural tube and the lateral edge of the injected side (Figure B.2B and B.2F). The area of the myotome on the injected side was significantly decreased due to cellular loss. We find a large increase in the number of embryos with normal 12/101 staining when we inject 0.5 pg *DUX4* mRNA (n = 87, missing n = 27, depleted n = 31) and this increase in normally stained embryos corresponded well to the number of normally developing embryos (Figure B.2I' compared to B.2I, B.2J, Figure B.3).

Muscle is the primary affected tissue in FSHD. To determine if DUX4 was similarly cytotoxic to tissues unaffected in FSHD, neurons were assayed in 1 pg *DUX4*

injected embryos (n = 24). Immunostaining for NCAM, which at stage 34 of *Xenopus* development is confined to neural tissue, resulted in levels of missing (n = 13) and depleted (n = 11) NCAM staining similar to that of 12/101 (Figure B.3). In several animals the NCAM staining was specifically depleted in the area of the eye (Figure B.2D & B.2H) and transverse sections revealed that often the neural tube staining was depleted as well (Figure B.2C & B.2H). We conclude that the cellular loss mediated by DUX4 is not restricted to muscle.

Embryos injected with *pitx1* mRNA were similarly analyzed for muscle and neuronal defects. The injection of *pitx1* led to developmental abnormalities similar to those seen with DUX4 albeit at a much higher concentration of mRNA. As with the DUX4 injections, 150 pg *pitx1* mRNA injections (n = 64) led to depletion (n = 32) and loss (n = 31) of 12/101 immunostaining from the injected side (Figure B.2Q' compared to B.2Q, B.2R). This effect was dose dependent, with 50 pg *pitx1* mRNA (n = 71) causing almost no loss of 12/101 immunostaining (missing n = 2, depleted n = 11) (Figure B.2U' compared to B.2U, B.2V). Similarly, the 150 pg *pitx1* injection (n = 8) led to depleted (n = 6) or missing (n = 2) NCAM immunostaining (Figure B.2S & B.2T), while 50 pg *pitx1* injection (n = 19) did not (Figure B.2W & B.2X).

Despite the lack of any significant gross developmental defects (Figure B.1I), DUX4c injected embryos were assayed for muscular and neuronal alterations. Differences were expected in the myotome based on previous findings of DUX4c mediated inhibition of myoblast differentiation (Ansseau *et al.*, 2009; Bosnakovski *et al.*, 2008a). However, when stained with 12/101, 1 ng *DUX4c* mRNA injected embryos (n = 89) still displayed normal immunostaining (Figure B.2M' compared to B.2M, B.2N) with

abnormalities (missing n = 0, depleted n = 10) comparable to *EGFP* mRNA injected controls (n = 41, missing n = 0, depleted n = 3) (Figure B.2Y, B.2Y', Figure B.3). Similarly, the NCAM staining of DUX4c injected embryos (Figure B.2O & B.2P) had abnormalities (n = 14, missing n = 0, depleted n = 1) similar to NCAM stained *EGFP* mRNA injected controls (n = 15, missing n = 0, depleted n = 1) (Figure B.3). We conclude that DUX4c levels have no effect on the developing musculature or neurons.

DUX4 eliminates myoD, myf5, and pax3 expression profiles

DUX4 expression led to the loss of 12/101 immunostaining indicating differentiated muscle was degraded or missing. To determine if DUX4 expression leads to a loss of muscle cell precursors, *in situ* hybridizations with probes against *myoD*, *myf5*, or *pax3* were performed (Figure B.4). In stage 34 embryos from the 1 pg *DUX4* mRNA injections, the expression of all three of these markers was missing in a large majority (myoD n = 60/62; myf5 n = 12/20; pax3 n = 23/28) (Figure B.4A', B.4B', & B.4C' compared to B.4A, B.4B, & B.4C). Assaying earlier in development, stage 20 *DUX4* injected embryos were missing expression of *pax3* (n = 11/20) (Figure B.4G). Although the markers *myf5* and *pax3* were depleted, the entire tissue, including neural tissue was affected. When taken into consideration with our previous findings that the DUX4-meidated defects began at gastrulation (Figure B.1A) and we conclude that the DUX4-mediated cellular loss occurs prior to stage 20 and therefore is clearly not muscle specific.

Unlike DUX4, DUX4c had little to no observable effect on the expression of *myoD*, *myf5*, and *pax3* (myoD n = 5/31; myf5 n = 0/8; pax3 n = 0/30) (Figure B.4D',

B.4E', & B.4F' compared to B.4D, B.4E, & B.4F, respectively). Similarly, no change was observed in *pax3* staining at stage 20 in embryos injected with 1 ng *DUX4c* mRNA (n = 0/12) (Figure B.4H). Quantitative reverse transcription PCR (qRT-PCR) was used to confirm there is no change in transcript levels of the myogenic regulators myoD or myf5 when injected with 1 ng *DUX4c* mRNA. Neither stage 20 embryos (n = 30) nor stage 34 embryos (n = 20) have statistically significant differences in *myoD* or *myf5* transcript levels when compared to *EGFP* mRNA injected controls normalized to gapdh (Figure B.5). Thus, DUX4c has no effect on muscle precursors or myogenic regulators. Moreover, we found no significant changes of myogenic regulators by qRT-PCR on intact tissue from 1 pg *DUX4* stage 34 injected embryos, further suggesting DUX4 has a strictly apoptotic role (Figure B.5).

DUX4 and DUX4c expression reduce endogenous *pitx1* expression

FSHD affected muscle has elevated levels of *PITX1* transcript and in cell culture assays DUX4 activates transcription of *PITX1* (Dixit *et al.*, 2007). Endogenous *pitx1* expression was examined by *in situ* hybridization in 1 pg *DUX4* and 1 ng *DUX4c* injected stage 34 *X. laevis* embryos. During early *X. laevis* development (prior to hind-limb development), observable *pitx1* expression is confined to the cement and pituitary glands, and this expression pattern is not altered by expression of either DUX4 (Figure B.6A' and B.6A; n = 23) or DUX4c (Figure B.6B'and B.6B; n = 27). Although there is no statistically significant change in stage 20 embryos, surprisingly, analysis of stage 34 transcript levels by qRT-PCR showed that *pitx1* expression was actually reduced by

DUX4c and DUX4 when compared to EGFP controls (*p-value* < 0.05) (Figure B.6D). We conclude that neither DUX4 nor DUX4c expression activates *pitx1* expression in vivo.

DISCUSSION

Multiple candidate genes have been proposed as causal of FSHD pathophysiology based on selective upregulation in FSHD patient-derived tissues or cell lines (Dixit *et al.*, 2007; Gabellini *et al.*, 2002; Klooster *et al.*, 2009; Reed *et al.*, 2007; Rijkers *et al.*, 2004). However inconsistent results, likely due in part to variations within and between FSHD patient biopsies as well as culturing conditions, have made it unclear which gene(s) are misexpressed. Previously we used *X. laevis* as a developing vertebrate model to analyze the effects of altering the expression levels of *frg1*. In the current study, we have used this same system to similarly analyze the effects of three additional FSHD candidate genes, *DUX4, DUX4c* and *pitx1*, on vertebrate muscle development.

X. laevis is a well-defined model system for muscle development with many advantages including external development, cells grow and differentiate under normal growth and environmental conditions giving rise to all tissues, and gene expression is easily manipulated through microinjection or transgenesis. In addition, *Xenopus* and humans share high levels of conservation of tissue organization, developmental processes, genes, and proteins. For example, the conservation of some muscle and FSHD associated proteins between human and *Xenopus* are as follows: FRG1 (80% identity, 88% similar), PITX1 (77% identity, 84% similar), PAX3 (91% identity, 96% similar) including 100% conservation of the homeodomain (aa 220-277), PAX7 (88% identity,

94% similar) including 100% conservation of the homeodomain (aa 217-274), MYOD1 (65% identity, 75% similar), and MYF5 (69% identity, 84% similar), as determined by alignments using NCBI BLAST Alignp. In contrast, DUX4 and DUX4c belong to the DUX family of double homeobox domain proteins but do not have *Xenopus* orthologs. In fact, when considering the entire protein sequence, including that residing outside their homeodomains, they both completely lack evolutionary conservation and appear to be unique to humans. Even the most closely related DUX4 ORFs in mice (Bosnakovski *et al.*, 2009; Clapp *et al.*, 2007; Kawazu *et al.*, 2007; Wu *et al.*) show levels of sequence similarity (31% amino acid identity for the Duxbl protein aligned to DUX4 using the ClustalW function of BioEdit Sequence Alignment Editor software) far below what is expected for mouse to human conservation. Without any clear ortholog available, the human sequences for *DUX4* and *DUX4c* were used for these studies.

We have confirmed the toxic effect of DUX4 expression in the context of normal vertebrate development during stages of active myogensis. By differential staining of neuronal and muscle tissues in DUX4 injected *Xenopus* embryos, we observe massive cellular loss of both tissue types and thus toxicity to not be muscle specific. Moreover, DUX4 injected *X. laevis* embryos show heavy TUNEL staining, indicating the function of DUX4 to induce apoptosis is conserved in *Xenopus*. Interestingly we have identified an extremely low threshold level of *DUX4* mRNA (0.5 pg) required for developmental abnormalities in *Xenopus,* at which point embryos appear to be either apoptotic or normal on the injected side. Although "non-toxic levels" of DUX4 in myoblast cell culture has been shown to impair differentiation (Bosnakovski *et al.*, 2008b), we observed no change in myotome development unless it was accompanied with generalized tissue apoptosis.

This observation suggests that DUX4 has an "all or nothing" effect in *Xenopus*; we observe either severe developmental consequences or no effect on normal development of the organism.

The toxic effects of DUX4 are proposed to be due to the similarity of the homeodomains with those of myogenic regulators involved in development and regeneration such that DUX4 competes for their binding to regulatory sites (Bosnakovski *et al.*, 2008b). The DUX4 homeodomain sequences are similar to *ortodenticle* (*otx*) and *paired* (*prd*) classes of proteins (Bopp *et al.*, 1986; Finkelstein *et al.*, 1990; Hewitt *et al.*, 1994). Otx and Pax proteins represent two families containing otx and prd homeodomains, respectively. Otx1 and Otx2 function in the nervous system (Simeone *et al.*, 2002), while the Pax3 and Pax7 proteins have known functions in the development of skeletal muscle and Pax7 in the maintenance of muscle satellite cells (Buckingham, 2006). In support of the DUX4 competition model, MyoD, a well characterized target of Pax3 activation during development and Pax7 in adult myogenesis (Olguin and Olwin, 2004; Relaix *et al.*, 2006; Tajbakhsh *et al.*, 1997), was shown to be rapidly down regulated by DUX4 in a inducible C2C12 myoblast cell line system (Bosnakovski *et al.*, 2008b). Considering *Xenopus* pax3 and pax7 homeodomains are 100% conserved with human PAX3 and PAX7, if DUX4 is functioning as a competitor of both PAX3 and PAX7 target genes, and consequently an antipodal regulator of myogenic genes in myoblasts, we would expect this competition to also be conserved in our study. In phenotypic DUX4 injected embryos, we observed a reduced level of *myoD* and *myf5* staining by *in-situ* hybridization when compared to EGFP controls. Furthermore, DUX4 injected embryos showed that *pax3* transcripts, the upstream regulator of myoD, was

absent by stage 20 of *Xenopus* development indicating DUX4 toxicity precedes muscle development. These data lead us to conclude that the DUX4 injected *Xenopus* phenotype is likely due to massive apoptosis on the injected side of the individuals and not resulting from muscle cell specific competition for pax3/7 targets. Taking into account that an aberrant increase in apoptosis is not generally considered to be part of the muscle pathology in FSHD (Sandri *et al.*, 2001; Winokur *et al.*, 2003a), this data could be consistent with DUX4 having a role in FSHD muscle pathology provided DUX4 is either only expressed at very low levels if at all under normal conditions and is only overexpressed in the muscle cell precursors of FSHD patients and not any other cells.

DUX4c, located within a truncated and inverted D4Z4 repeat located just centromeric from the FSH1A locus, has been shown to be up-regulated in FSHD (Ansseau *et al.*, 2009; Wright *et al.*, 1993). The gene encodes an ORF identical to DUX4 except for differing in the last 82 amino acids which are substituted with 32 unrelated amino acids. Interestingly, the C-terminal substitution leaves DUX4c with the exact homeodomains found in DUX4. This, in theory, would enable DUX4c to interact with all of the same genetic targets of DUX4. Therefore, considering 100% conservation of pax3/7 homeodomains from human to *Xenopus,* expression of DUX4c should also compete with the pax3 and pax7 for myogenic target genes and thus lead to myogenic abnormalities in *Xenopus*. Interestingly, 1ng *DUX4c* mRNA injections (2000 fold over DUX4 threshold levels) produce only a slight increase in abnormal *Xenopus* development, further indicating the DUX homeodomains do not compete with pax3 or pax7 for myogenic target genes. In two previous studies, DUX4c has been shown to inhibit myoblast differentiation and down-regulate MyoD (Ansseau *et al.*, 2009; Bosnakovski *et*

al., 2008a). Oddly, both studies investigate effects of DUX4c on Myf5 expression in identical myoblast cell lines and find opposite results; one finding Myf5 is down-regulated (Bosnakovski *et al.*, 2008a), while the other shows an up regulation of Myf5 (Ansseau *et al.*, 2009). Interestingly, differences between FSHD and control DUX4c levels were only observed in myotubes, after the effects on myoblast differentiation would have passed (Ansseau *et al.*, 2009). We observed no obvious changes in staining patterns for myoD or myf5 in DUX4c injected *Xenopus* embryos by in-situ hybridization and mRNA levels are not significantly different from that of EGFP injected controls by qRT-PCR. This study on the effects of DUX4c on myogenic regulators in a vertebrate going through muscle development leads us to conclude that DUX4c expression has no overt effects on muscle development and is not consistent with DUX4c expression having a role in FSHD pathology.

The FSHD candidate PITX1 is a member of the paired family of homeodomain transcription factors (Tremblay *et al.*, 2000). Multiple studies focused on PITX1 shows it is involved in specification of hind limb identity, as well as left-right symmetry (Cole *et al.*, 2003; Lanctot *et al.*, 1999; Logan and Tabin, 1999; Shapiro *et al.*, 2004; Szeto *et al.*, 1999; Tanaka *et al.*, 2005). It was recently shown that DUX4 could activate transient expression of a reporter gene fused to the *PITX1* promoter as well as the endogenous *PITX1* gene in transfected C2C12 cells (Dixit *et al.*, 2007). Although the sequence of the *pitx1* promoter is unknown in *X. laevis*, the fact that DUX4 maintains its characteristic ability to induce apoptosis suggests it is interacting with its conserved targets. Taking into consideration that DUX4 and DUX4c contain identical homeodomains we tested their potential to regulate *pitx1*. At stage 34 pitx1 is characteristically expressed at high

levels in the cement gland (Chang *et al.*, 2001; Hollemann and Pieler, 1999; Schweickert *et al.*, 2001). We found no obvious increase of *pitx1* in DUX4 or DUX4c injected embryos. Interestingly, we did see a statistically significant (*p-value* <0.05) decrease in *pitx1* mRNA when DUX4c or DUX4 injected stage 34 tadpoles were subjected to qRT-PCR. We conclude that neither DUX4 nor DUX4c induce *pitx1* in *Xenopus* and DUX4-mediated apoptosis is not mediated through activation of *pitx1*. To directly test the effects of overexpression of pitx1, we circumvented the issues related to DUX4 expression by directly increasing pitx1 through microinjection and determine the effects on *Xenopus* development. As with DUX4, we observed severe, general (not muscle cell specific) developmental abnormalities when pitx1 was overexpressed in *Xenopus*, agreeing with previous reports (Schweickert *et al.*, 2001). Like DUX4, these abnormalities likely arise from the induction of apoptosis, as increased pitx1 expression has been shown to directly lead to increased p53 expression and cellular loss (Liu and Lobie, 2007). This result is consistent with PITX1 playing a role in FSHD assuming it is only overexpressed in muscle lineages of FSHD patients.

At this point, the mechanism of FSHD pathophysiology remains unknown. In total we tested the effects of systemic overexpression of four FSHD candidate genes on vertebrate development in our *Xenopus* system. We have found systemic overexpression of DUX4c has little effect while DUX4 and pitx1 produce a general cytotoxicity to all cell types in developing embryos. FSHD is likely an epigenetic disorder but it is not known if the cause of FSHD is a misregulation of a gene specifically restricted to skeletal muscle and its precursors or if there is a global misregulation of a gene with skeletal muscle myogenesis being specifically susceptible. Only if it is the former, and DUX4

and PITX1 were exclusively overexpressed in skeletal muscle precursors, could they have a role in FSHD pathology. We know no mechanism whereby DUX4 or PITX1 cytotoxicity could produce the vasculature phenotype strongly associated with FSHD. In respect to DUX4c we conclude it likely has no role in FSHD pathology, however, it is possible multiple candidates including DUX4c could function together to produce a synergistic effect ultimately resulting in FSHD-like pathology. This compares poorly to FRG1 from our previous studies where systemic overexpression of frg1 could recapitulate both major symptoms of FSHD in *Xenopus*, dystrophic muscle and increased angiogenesis (Hanel *et al.*, 2009; Wuebbles *et al.*, 2009). Taken together, the functional and phenotypic data point to *FRG1* as the most likely candidate whose misexpression, either systemically or specifically during myogenesis, leads to FSHD pathology.

MATERIALS AND METHODS

Frog husbandry

Adult *X. laevis* were purchased from *Xenopus* Express. All procedures were carried out in accordance with established UIUC IACUC approved protocols for animal welfare.

Plasmid constructs and RNA production

The vectors pCIneo DUX4 and pCIneo DUX4c were generously provided by Dr. Alexandra Belayew (Ansseau *et al.*, 2009; Gabriels *et al.*, 1999). The plasmids for *EGFP, myoD*, *pax3*, and *myf5* RNA have been previously described (Hanel *et al.*, 2009).

The *pitx1* cDNA was produced by RT-PCR using primers 5' gtgattgaccatggattcctttaaagg 3' and 5' tcaactgttatattggcaagcattgag 3', cloned into pGEM T-Easy (Promega) and sequenced. The cDNA was subcloned into the EcoRI and XbaI sites of pcDNA3.1 (Invitrogen). Production of *EGFP* mRNA was performed as previously described (Hanel *et al.*, 2009). For DUX4, DUX4c and pitx1 mRNA, constructs were linearized and capped mRNA was generated using T7 RNA polymerase and the mMessage mMachine kit (Ambion, Inc).

Xenopus embryo injections

In vitro fertilized embryos were generated as described (Hanel *et al.*, 2009). Embryos were microinjected after completion of the two cell stage, as indicated by the beginning of the second cleavage, in 1X MMR with 3% Ficoll and incubated at 19°C. Between 3-6 hours after injection, embryos were transferred to 0.1X MMR with 3% Ficoll. After 24-36 hours embryos were either peeled and fixed for stage 18-22 embryos or cultured in 0.1X MMR until the desired stage. After neural tube closure all injected embryos were sorted based on left, right or bilateral fluorescence. *DUX4* mRNA was injected at 500pg, 250pg, 100pg, 10pg, 1pg, and 0.5pg along with 500pg *EGFP* mRNA. *DUX4c* mRNA was injected at 1 ng along with 500pg *EGFP*. *pitx1* was injected at 150pg and 50pg along with 500pg *EGFP*. Control *EGFP* mRNA injections were performed at 500pg.

TUNEL assay

TUNEL staining of whole-mount *Xenopus* embryos was carried out using a protocol adapted from Hensey and Gautier (Hensey and Gautier, 1997). All procedures were carried out at room temperature unless noted otherwise. Embryos were fixed for 1 hr. in MEMFA, (100 mM MOPS (pH 7.4), 2 mM EGTA, 1 mM MgSO4, 4% formaldehyde). Embryos were washed in methanol 2 x 30 min. and stored in methanol at -20°C. For rehydration, half of methanol was replaced with PBS and washed 5 x 5min. The embryos were washed with PBT (0.2% Tween-20 in PBS), 2 x 15 min., followed by 2 x 15 min. washes in PBS. Embryo pigment was removed by treatment for 1-2 hours in 1% H_2O_2, 5% Formamide, and 0.5X SSC under bright light, and washed 3 x 15 min. in PBS. Embryos were transferred to terminal deoxynucleotidyl transferase, (TdT), buffer (Invitrogen) and washed for 30 min. End labeling was carried out overnight in TdT buffer containing 0.5 mM digoxygenin-dUTP (Roche Diagnostics), and 150 U/ml TdT (Invitrogen). Embryos were then washed 2 x 1 hr. in PBS/1 mM EDTA, at 65°C, followed by 4 x 1 hr. in PBS. Detection and chromogenic reaction was carried out as previously described (Harland, 1991). Embryos were viewed and stored following rehydration in 1X PBS.

***In situ* hybridizations**

Embryos were staged according to Nieuwkoop and Faber (Nieuwkoop and Faber, 1994), fixed 1-2 hrs in MEMFA, washed 2 x 30 min in 100% methanol and stored in 100% methanol at -20°C until use. The *EGFP*, *Xenopus myoD*, *pax3*, and *myf5* antisense probes generated as previously described (Hanel *et al.*, 2009). The *pitx1* probe was

generated by linearizing pGEM *pitx1* with SalI and using T7 RNA polymerase transcription to generate digoxigenin (DIG) -11-UTP (Roche Diagnostics) antisense RNA probes. *In situ* hybridizations were performed according to standard methods (Harland, 1991) and detected with alkaline phosphatase (AP) linked anti-DIG antibody (Roche Diagnostics) and the chromogenic substrates BCIP (5-Bromo-4-chloro-3-indolyl phosphate, toluidine salt) and NBT (Nitro blue tetrazolium chloride) (Roche Diagnostics). Embryos were refixed overnight in Bouin's fixative, followed by washing in 70% ethanol/30% PBS-Tween 0.1%, and pigment was removed by treatment for 1-2 hours in 1% H_2O_2, 5% Formamide, and 0.5X SSC under bright light. Embyos were then washed in methanol 10 minutes and transferred to 1mM EDTA in PBS or glycerol for analysis and photography.

Immunohistochemistry

Embryos were staged and fixed as above, rehydrated in PBS-DT (1%DMSO, 1% Tween-20) and washed for 15 min in PBS-DT. Samples were blocked in 0.1M glycine, 2% milk, 1% BSA, 1% Tween-20 and 1% DMSO for 4 hours at room temperature or overnight at 4°C. Primary antibodies were diluted in blocking solution as follows: Skeletal muscle marker (12/101) diluted 1:3 or NCAM (4d) diluted 1:20 were incubated with embryos overnight at 4°C and detected using a HRP secondary (GE Healthcare) with a DAB staining kit from (Roche Diagnostics). For paraffin sectioning, tadpoles immunostained for 12/101 or NCAM were dehydrated through an EtOH series, placed in 50/50 EtOH/Xylene for 10 minutes, washed twice with 100% Xylene, emebedded in paraffin, positioned, and sectioned using a microtome.The 12/101 monoclonal antibody,

developed by J.P. Brockes, and the NCAM 4d monoclonal antibody, developed by U. Rutishauser, were obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained at The University of Iowa, Department of Biological Sciences, Iowa City, IA 52242.

qRT-PCR

For each sample, total RNA was purified from 10 pooled embryos using Trizol reagent (Invitrogen) per manufacture's protocol. RNA was then treated with 1U RQ1 DNase (Promega) per 1ug RNA for 30 min. at 37°C. cDNA synthesis was preformed using 1ug of total RNA, 50ng of random hexamer, and Superscript III (Invitrogen) per manufacture's suggested method. Relative transcript levels were determined using 1ul of (1:20 diluted) cDNA (in triplicate), iQ SYBR Supermix, and gene specific primers (myoD: 5' TGCCAAGAGTCCAGATTTCC 3', 5' CAGGTCTTCAAAGAAACTCATGTC 3'; myf5: 5' GCTTATCTAGTATTGTGGATCGG 3', 5' CTGGTTTGTTGGGTGTAAGG 3'; pitx1: 5' CATGAGCAGAAGTGATTGAC 3', 5' GTAAAGTGAGTCCTTTGTCTCC 3'; gapdh: 5' GGTGAAGGTTGGAATTAACGG 3', 5' GATCAGCTTGCCATTCTCAG 3') on a Bio-Rad iCycler IQ machine. Experiments were preformed at least 3 times. Data analyses were preformed using the comparative Ct method and error bars are $\pm$ standard error of the mean. Changes were determined using the two tailed student's t-test and considered significantly different at a *P-value* <0.05.

REFERENCES

Ansseau, E., Laoudj-Chenivesse, D., Marcowycz, A., Tassin, A., Vanderplanck, C., Sauvage, S., Barro, M., Mahieu, I., Leroy, A., Leclercq, I., Mainfroid, V., Figlewicz, D., Mouly, V., Butler-Browne, G., Belayew, A., and Coppee, F. (2009) DUX4c is up-regulated in FSHD. It induces the MYF5 protein and human myoblast proliferation. *PLoS One* **4**: e7482.

Barro, M., Carnac, G., Flavier, S., Mercier, J., Vassetzky, Y., and Laoudj-Chenivesse, D. (2008) Myoblasts from affected and non affected FSHD muscles exhibit morphological differentiation defects. *J Cell Mol Med*.

Bodega, B., Ramirez, G.D., Grasser, F., Cheli, S., Brunelli, S., Mora, M., Meneveri, R., Marozzi, A., Mueller, S., Battaglioli, E., and Ginelli, E. (2009) Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC Biol* **7**: 41.

Bopp, D., Burri, M., Baumgartner, S., Frigerio, G., and Noll, M. (1986) Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of Drosophila. *Cell* **47**: 1033-1040.

Bosnakovski, D., Lamb, S., Simsek, T., Xu, Z., Belayew, A., Perlingeiro, R., and Kyba, M. (2008a) DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. *Exp Neurol*.

Bosnakovski, D., Xu, Z., Gang, E.J., Galindo, C.L., Liu, M., Simsek, T., Garner, H.R., Agha-Mohammadi, S., Tassin, A., Coppee, F., Belayew, A., Perlingeiro, R.R., and Kyba, M. (2008b) An isogenetic myoblast expression screen identifies DUX4-mediated FSHD-associated molecular pathologies. *EMBO J* **27**: 2766-2779.

Bosnakovski, D., Daughters, R.S., Xu, Z., Slack, J.M., and Kyba, M. (2009) Biphasic myopathic phenotype of mouse DUX, an ORF within conserved FSHD-related repeats. *PLoS One* **4**: e7003.

Buckingham, M. (2006) Myogenic progenitor cells and skeletal myogenesis in vertebrates. *Curr Opin Genet Dev* **16**: 525-532.

Chang, W., KhosrowShahian, F., Chang, R., and Crawford, M.J. (2001) xPitx1 plays a role in specifying cement gland and head during early Xenopus development. *Genesis* **29**: 78-90.

Clapp, J., Mitchell, L.M., Bolland, D.J., Fantes, J., Corcoran, A.E., Scotting, P.J., Armour, J.A., and Hewitt, J.E. (2007) Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am J Hum Genet* **81**: 264-279.

Cole, N.J., Tanaka, M., Prescott, A., and Tickle, C. (2003) Expression of limb initiation genes and clues to the morphological diversification of threespine stickleback. *Curr Biol* **13**: R951-952.

Dixit, M., Ansseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., Sauvage, S., Matteotti, C., van Acker, A.M., Leo, O., Figlewicz, D., Barro, M., Laoudj-Chenivesse, D., Belayew, A., Coppee, F., and Chen, Y.W. (2007) DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc Natl Acad Sci U S A* **104**: 18157-18162.

Finkelstein, R., Smouse, D., Capaci, T.M., Spradling, A.C., and Perrimon, N. (1990) The orthodenticle gene encodes a novel homeo domain protein involved in the development of the Drosophila nervous system and ocellar visual structures. *Genes Dev* **4**: 1516-1527.

Fitzsimons, R.B., Gurwin, E.B., and Bird, A.C. (1987) Retinal vascular abnormalities in facioscapulohumeral muscular dystrophy. A general association with genetic and therapeutic implications. *Brain* **110 (Pt 3)**: 631-648.

Gabellini, D., Green, M.R., and Tupler, R. (2002) Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* **110**: 339-348.

Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R., Green, M.R., and Tupler, R. (2006) Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* **439**: 973-977.

Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., and Belayew, A. (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* **236**: 25-32.

Hanel, M.L., Wuebbles, R.D., and Jones, P.L. (2009) Muscular dystrophy candidate gene FRG1 is critical for muscle development. *Dev Dyn* **238**: 1502-1512.

Harland, R.M. (1991) In situ hybridization: an improved whole-mount method for Xenopus embryos. *Methods Cell Biol* **36**: 685-695.

Hensey, C., and Gautier, J. (1997) A developmental timer that regulates apoptosis at the onset of gastrulation. *Mech Dev* **69**: 183-195.

Hewitt, J.E., Lyle, R., Clark, L.N., Valleley, E.M., Wright, T.J., Wijmenga, C., van Deutekom, J.C., Francis, F., Sharpe, P.T., Hofker, M., and et al. (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum Mol Genet* **3**: 1287-1295.

Hollemann, T., and Pieler, T. (1999) Xpitx-1: a homeobox gene expressed during pituitary and cement gland formation of Xenopus embryos. *Mech Dev* **88**: 249-252.

Jiang, G., Yang, F., van Overveld, P.G., Vedanarayanan, V., van der Maarel, S., and Ehrlich, M. (2003) Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum Mol Genet* **12**: 2909-2921.

Kawazu, M., Yamamoto, G., Yoshimi, M., Yamamoto, K., Asai, T., Ichikawa, M., Seo, S., Nakagawa, M., Chiba, S., Kurokawa, M., and Ogawa, S. (2007) Expression profiling of immature thymocytes revealed a novel homeobox gene that regulates double-negative thymocyte development. *J Immunol* **179**: 5335-5345.

Klooster, R., Straasheijm, K., Shah, B., Sowden, J., Frants, R., Thornton, C., Tawil, R., and van der Maarel, S. (2009) Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. *Eur J Hum Genet* **17**: 1615-1624.

Kowaljow, V., Marcowycz, A., Ansseau, E., Conde, C.B., Sauvage, S., Matteotti, C., Arias, C., Corona, E.D., Nunez, N.G., Leo, O., Wattiez, R., Figlewicz, D., Laoudj-Chenivesse, D., Belayew, A., Coppee, F., and Rosa, A.L. (2007) The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein. *Neuromuscul Disord* **17**: 611-623.

Lanctot, C., Moreau, A., Chamberland, M., Tremblay, M.L., and Drouin, J. (1999) Hindlimb patterning and mandible development require the Ptx1 gene. *Development* **126**: 1805-1810.

Liu, D.X., and Lobie, P.E. (2007) Transcriptional activation of p53 by Pitx1. *Cell Death Differ* **14**: 1893-1907.

Logan, M., and Tabin, C.J. (1999) Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. *Science* **283**: 1736-1739.

Morosetti, R., Mirabella, M., Gliubizzi, C., Broccolini, A., Sancricca, C., Pescatori, M., Gidaro, T., Tasca, G., Frusciante, R., Tonali, P.A., Cossu, G., and Ricci, E. (2007) Isolation and characterization of mesoangioblasts from facioscapulohumeral muscular dystrophy muscle biopsies. *Stem Cells* **25**: 3173-3182.

Nieuwkoop, P.D., and Faber, J. (1994) *Normal table of Xenopus laevis (Daudin) : a systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis.* New York: Garland Pub.

Olguin, H.C., and Olwin, B.B. (2004) Pax-7 up-regulation inhibits myogenesis and cell cycle progression in satellite cells: a potential mechanism for self-renewal. *Dev Biol* **275**: 375-388.

Padberg, G.W., Brouwer, O.F., de Keizer, R.J., Dijkman, G., Wijmenga, C., Grote, J.J., and Frants, R.R. (1995) On the significance of retinal vascular disease and hearing loss in facioscapulohumeral muscular dystrophy. *Muscle Nerve* **2**: S73-80.

Reed, P.W., Corse, A.M., Porter, N.C., Flanigan, K.M., and Bloch, R.J. (2007) Abnormal expression of mu-crystallin in facioscapulohumeral muscular dystrophy. *Exp Neurol* **205**: 583-586.

Relaix, F., Montarras, D., Zaffran, S., Gayraud-Morel, B., Rocancourt, D., Tajbakhsh, S., Mansouri, A., Cumano, A., and Buckingham, M. (2006) Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells. *J Cell Biol* **172**: 91-102.

Rijkers, T., Deidda, G., van Koningsbruggen, S., van Geel, M., Lemmers, R.J., van Deutekom, J.C., Figlewicz, D., Hewitt, J.E., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2004) FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *J Med Genet* **41**: 826-836.

Sandri, M., El Meslemani, A.H., Sandri, C., Schjerling, P., Vissing, K., Andersen, J.L., Rossini, K., Carraro, U., and Angelini, C. (2001) Caspase 3 expression correlates with skeletal muscle apoptosis in Duchenne and facioscapulo human muscular dystrophy. A potential target for pharmacological treatment? *J Neuropathol Exp Neurol* **60**: 302-312.

Schweickert, A., Deissler, K., Blum, M., and Steinbeisser, H. (2001) Pitx1 and Pitx2c are required for ectopic cement gland formation in Xenopus laevis. *Genesis* **30**: 144-148.

Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter, D., and Kingsley, D.M. (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**: 717-723.

Simeone, A., Puelles, E., and Acampora, D. (2002) The Otx family. *Curr Opin Genet Dev* **12**: 409-415.

Snider, L., Asawachaicharn, A., Tyler, A.E., Geng, L.N., Petek, L.M., Maves, L., Miller, D.G., Lemmers, R.J., Winokur, S.T., Tawil, R., van der Maarel, S.M., Filippova, G.N., and Tapscott, S.J. (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum Mol Genet* **18**: 2414-2430.

Szeto, D.P., Rodriguez-Esteban, C., Ryan, A.K., O'Connell, S.M., Liu, F., Kioussi, C., Gleiberman, A.S., Izpisua-Belmonte, J.C., and Rosenfeld, M.G. (1999) Role of the Bicoid-related homeodomain factor Pitx1 in specifying hindlimb morphogenesis and pituitary development. *Genes Dev* **13**: 484-494.

Tajbakhsh, S., Rocancourt, D., Cossu, G., and Buckingham, M. (1997) Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. *Cell* **89**: 127-138.

Tanaka, M., Hale, L.A., Amores, A., Yan, Y.L., Cresko, W.A., Suzuki, T., and Postlethwait, J.H. (2005) Developmental genetic basis for the evolution of pelvic fin loss in the pufferfish Takifugu rubripes. *Dev Biol* **281**: 227-239.

Tremblay, J.J., Goodyer, C.G., and Drouin, J. (2000) Transcriptional properties of Ptx1 and Ptx2 isoforms. *Neuroendocrinology* **71**: 277-286.

van Deutekom, J.C., Wijmenga, C., van Tienhoven, E.A., Gruter, A.M., Hewitt, J.E., Padberg, G.W., van Ommen, G.J., Hofker, M.H., and Frants, R.R. (1993) FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum Mol Genet* **2**: 2037-2042.

van Overveld, P.G., Lemmers, R.J., Sandkuijl, L.A., Enthoven, L., Winokur, S.T., Bakels, F., Padberg, G.W., van Ommen, G.J., Frants, R.R., and van der Maarel, S.M. (2003) Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat Genet* **35**: 315-317.

Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., and et al. (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* **2**: 26-30.

Winokur, S.T., Bengtsson, U., Vargas, J.C., Wasmuth, J.J., Altherr, M.R., Weiffenbach, B., and Jacobsen, S.J. (1996) The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum Mol Genet* **5**: 1567-1575.

Winokur, S.T., Barrett, K., Martin, J.H., Forrester, J.R., Simon, M., Tawil, R., Chung, S.A., Masny, P.S., and Figlewicz, D.A. (2003a) Facioscapulohumeral muscular dystrophy (FSHD) myoblasts demonstrate increased susceptibility to oxidative stress. *Neuromuscul Disord* **13**: 322-333.

Winokur, S.T., Chen, Y.W., Masny, P.S., Martin, J.H., Ehmsen, J.T., Tapscott, S.J., van der Maarel, S.M., Hayashi, Y., and Flanigan, K.M. (2003b) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol Genet* **12**: 2895-2907.

Wright, T.J., Wijmenga, C., Clark, L.N., Frants, R.R., Williamson, R., and Hewitt, J.E. (1993) Fine mapping of the FSHD gene region orientates the rearranged fragment detected by the probe p13E-11. *Hum Mol Genet* **2**: 1673-1678.

Wu, S.L., Tsai, M.S., Wong, S.H., Hsieh-Li, H.M., Tsai, T.S., Chang, W.T., Huang, S.L., Chiu, C.C., and Wang, S.H. Characterization of genomic structures and expression profiles of three tandem repeats of a mouse double homeobox gene: Duxbl. *Dev Dyn*.

Wuebbles, R.D., Hanel, M.L., and Jones, P.L. (2009) FSHD region gene 1 (FRG1) is crucial for angiogenesis linking FRG1 to facioscapulohumeral muscular dystrophy-associated vasculopathy. *Dis Model Mech* **2**: 267-274.

FIGURES

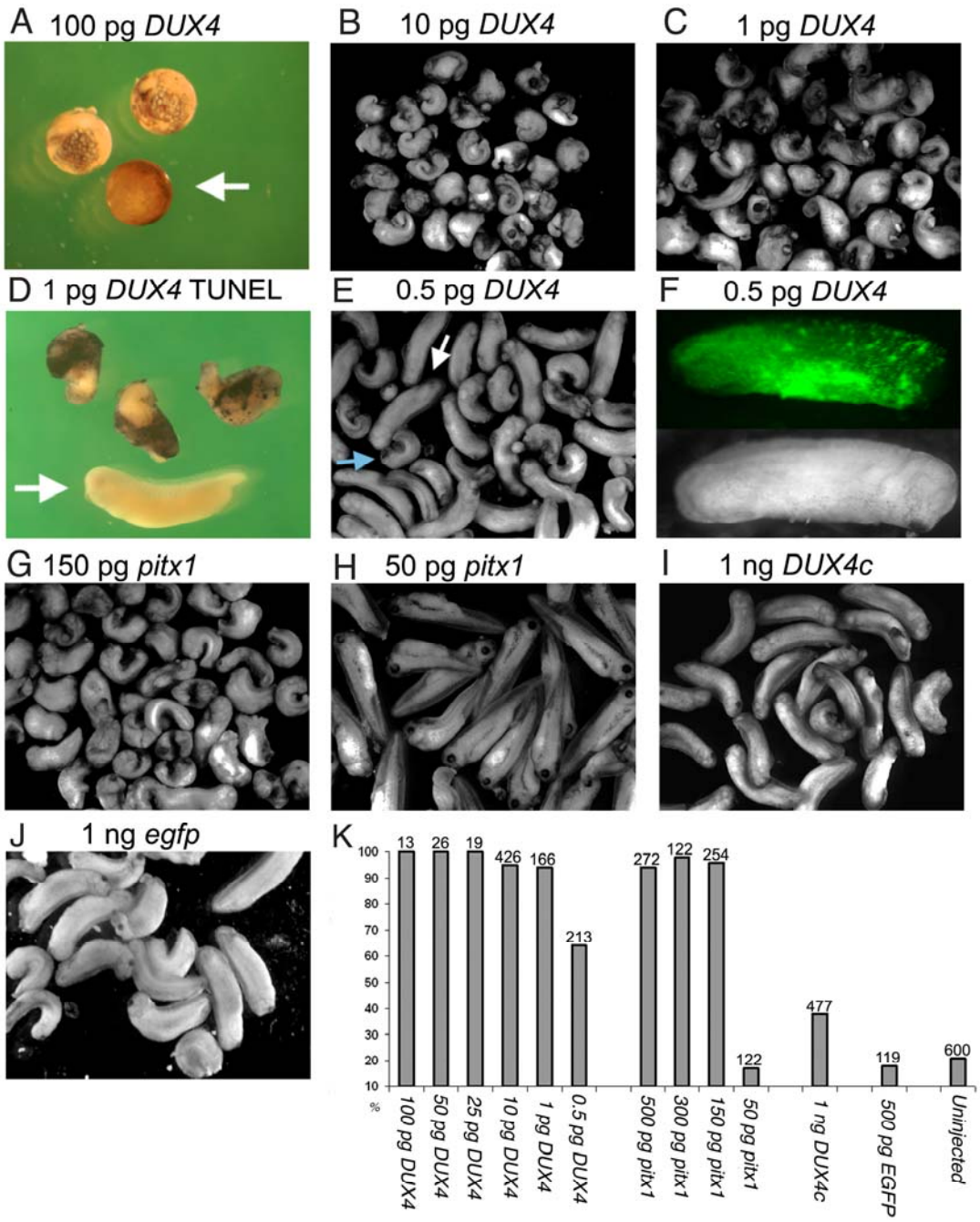

Figure B.1: DUX4 and pitx1 induce developmental abnormalities**.** A) Injection of 100 pg *DUX4* mRNA leads to early developmental arrest (stage 9) and apoptosis compared to control embryos (arrow). B and C) Injection of10 pg and 1 pg of *DUX4* mRNA lead to developmental abnormalities on only the injected side of embryos (arrow head).

(Figure B.1 continued) D) TUNEL assay on 1 pg DUX4 mRNA injection shows apoptosis in embryos (blue staining) compared to control injections (arrow). E) Injection of 0.5 pg *DUX4* mRNA allows some normal development (white arrow) compared with abnormal developing animals (blue arrow). F) Normally developing 0.5 pg *DUX4* mRNA injected embryo is indicated by expression of co-injected EGFP (top) compared with the uninjected, non-fluorescing side (bottom, shown as a bright field image). G) Injection of 150 pg *pitx1* mRNA results in developmental abnormalities on the injected side (arrow head). H) Embryos injected with 50 pg *pitx1* mRNA are developmentally normal (compared with G). I and J) The effect of 1ng DUX4c mRNA injection (I) is similar to EGFP controls (J). K) Summary of percent abnormal embryos observed in mRNA injection experiments with number of animals analyzed for each set.

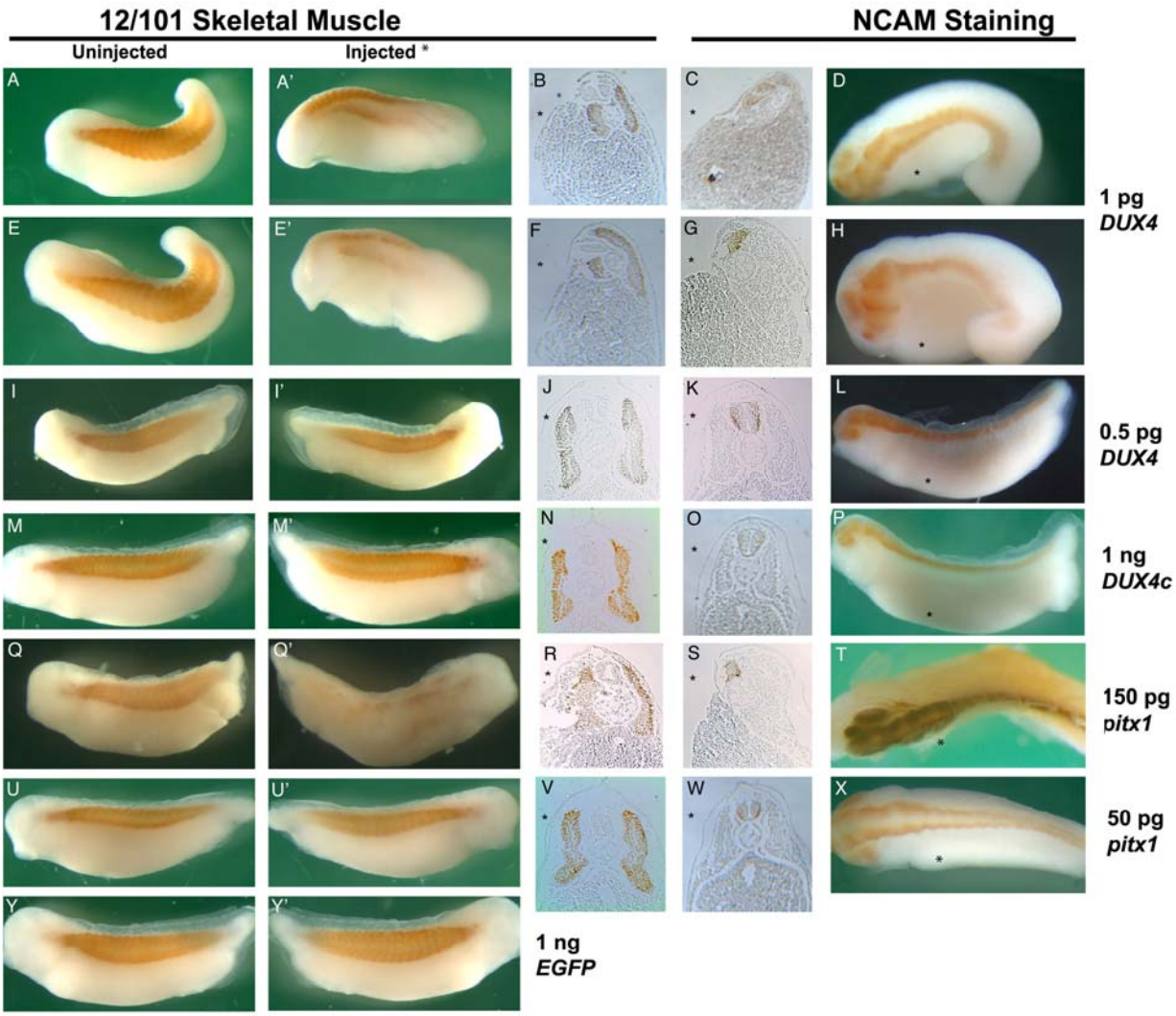

Figure B.2: Expression of DUX4 and pitx1 but not DUX4c affect muscle and neural tissue development. A-H) *DUX4* (1 pg) injected embryos show depletion of 12/101 (A, B, E, and F) and NCAM (C, D, G, and H) immunostaining and cellular loss specifically on the mRNA injected side (*) in wholemount (A, A', D, E, E', and H) and transverse sections (B, C, F, and G). Wholemount (I, L) and transverse sectioned (J, K) *DUX4* (0.5 pg) injected embryos have normal 12/101 (I, J) and NCAM (K, L) staining and myotome development on the injected side (*). Wholemount (M, P) and transverse sectioned (N, O) *DUX4c* (1 ng) injected embryos have normal 12/101 (M, N) and NCAM (O, P) staining and myotome development on the injected side (*). Wholemount (Q, T) and transverse sectioned (R, S) *pitx1* (150 pg) injected embryos show depletion of 12/101 (Q, R) and NCAM (S, T) immunostaining, abnormal myotome development and cellular loss specifically on the injected side (*). Wholemount (U, X) and transverse sectioned (V, W) U-X) *pitx1* (50 pg) injected embryos show no depletion of 12/101 (U, V) and NCAM (W, X) immunostaining and exhibit normal myotome development with no cellular loss on the injected side (*). Wholemount (Y) 1 ng *EGFP* (1 ng) injected embryos have normal 12/101 staining when compared to the uninjected side.

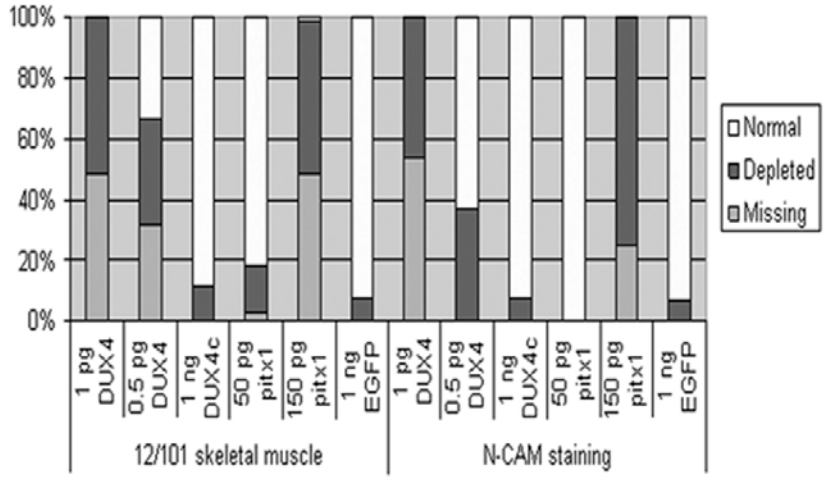

Figure B.3: Summary of DUX4, DUX4c, and pitx1 mediated affects**.** Graphic summary of results from Figure 2, showing percent of animals with missing, depleted, or normal 12/101 and NCAM immunostaining.

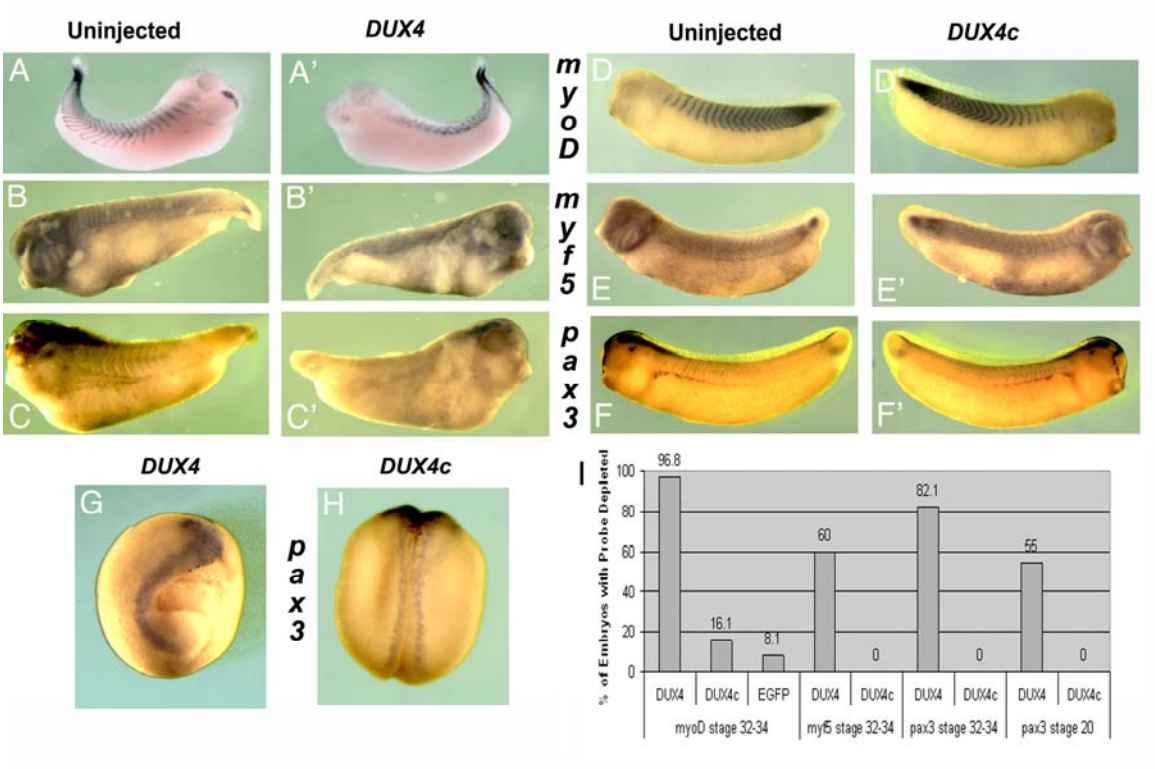

Figure B.4: Effects of DUX4 and DUX4c on *myoD*, *myf5*, and *pax3* mRNA expression patterns. A-F) Stage 32-34 embryos were assessed for observable defects produced by DUX4 (A-C) and DUX4c (D-F) by *in-situ* hybridization for myogenic regulators. Injection of *DUX4* (1 pg) mRNA into developing embryos decreased myoD (A, A'), myf5 (B, B"), and pax3 (C, C') staining specifically on the injected side due to cellular loss. Injection of *DUX4c* (1 ng) mRNA into embryos produces no observable changes in myoD (D, D'), myf5 (E, E'), or pax3 (F, F') staining intensities when compared to the uninjected side. G) Stage 20 embryos injected with *DUX4* (1 pg) mRNA show depletion of pax3 staining before skeletal muscle development. H) Stage 20 embryos injected with *DUX4c* (1 ng) mRNA does not affect pax3 staining. I) Summary graph showing the percentage of observed *in situ* probe depleted embryos injected with DUX4 and DUX4c.

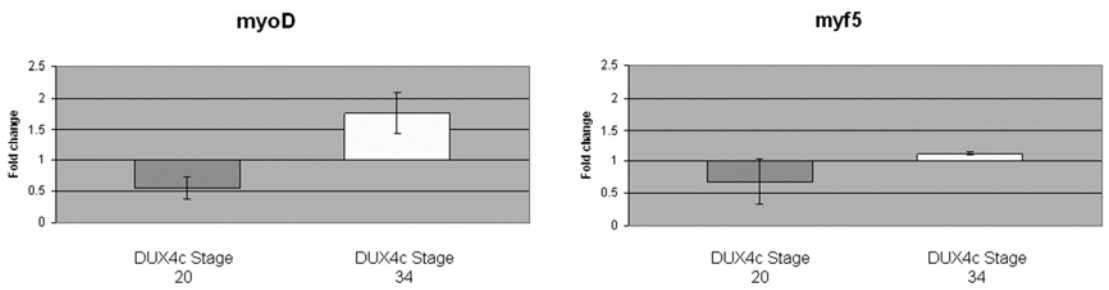

Figure B.5: Systemic DUX4c or DUX4 does not alter global levels of *myoD* or *myf5*. Levels of mRNA were quantified by qRT-PCR for *myoD* and *myf5* in *DUX4c* (1 ng) or DUX4 (1 pg) injected embryos. Fold change is normalized to *gapdh* levels and compared to *EGFP* mRNA at a relative level of 1; error bars are ± standard error of the mean.

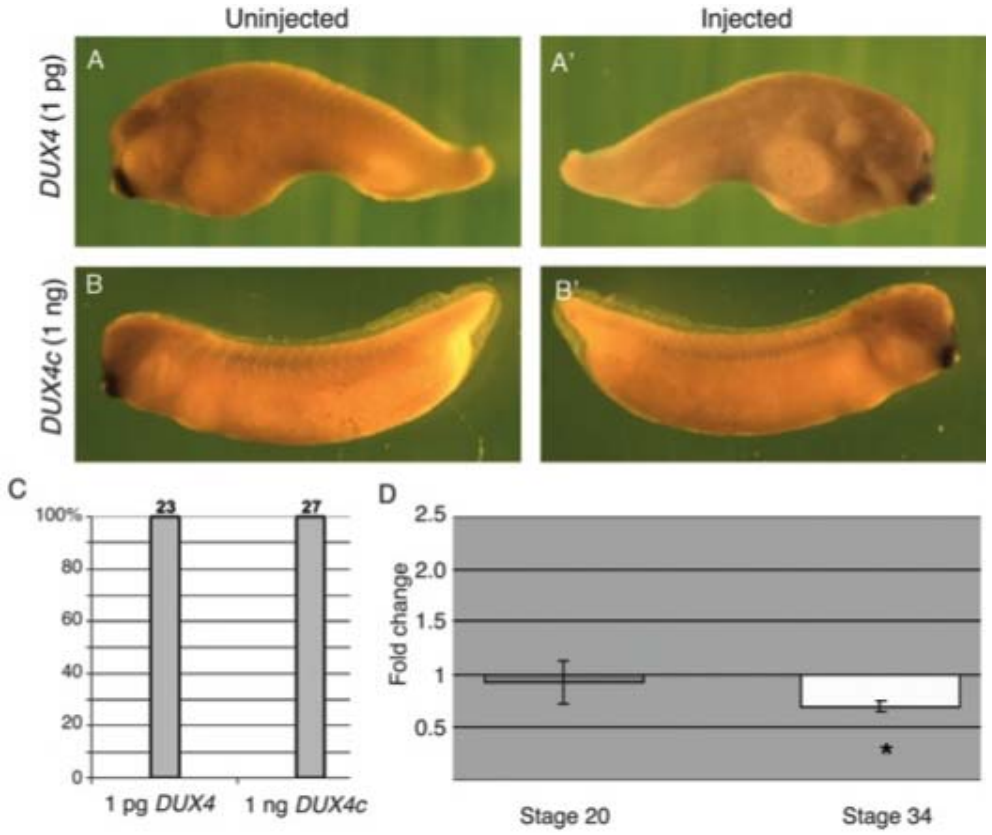

Figure B.6: Neither DUX4 nor DUX4c can induce *pitx1* expression. A-B) *In situ* hybridizations for *pitx1* indicate that neither embryos injected with *DUX4* nor *DUX4c* mRNA upregulate pitx1. C) Graphic summary showing none of the embryos tested by *in situ* hybridization had obvious upregulation of *pitx1*, as illustrated by percent embryos without increased probe, with numbers of animals assayed. D) Quantification of *pitx1* mRNA in *DUX4c* (1 ng) or *DUX4* injected embryos as determined by qRT-PCR. Fold change is normalized to *gapdh* and compared to *EGFP* at a relative level of 1; error bars are ±standard error of the mean. Significance (p-value < 0.05) is indicated by (*).

APPENDIX C*

FACIOSCAPULOHUMERAL MUSCULAR DYSTROPHY (FSHD) REGION GENE 1 (FRG1) IS A DYNAMIC NUCLEAR AND SARCOMERIC PROTEIN

***Data and writing is adapted from:**

Meredith L. Hanel, Chia-Yun Jessica Sun, Steven W. Long, Takako I. Jones, Derek Milner, and Peter L. Jones *Facioscapulohumeral muscular dystrophy (FSHD) region gene 1 (FRG1) is a dynamic nuclear and sarcomeric protein* **(Differentiation, 2010, submitted)**

Chia-Yun Jessica Sun, Silvana van Koningsbruggen, Steven W. Long, Michel Bellini, Lyne Levesque, William M. Brieher, Silvère M. van der Maarel, and Peter L. Jones *FSHD region gene 1 (FRG1) is a dynamic RNA-associated, actin bundling protein* **(Manuscript in preparation)**

INTRODUCTION

Facioscapulohumeral muscular dystrophy (FSHD) is one of the most common muscular dystrophies (incidence of 1:7,500 – 1:14,000), and third most common overall (Lunt and Harper, 1991; Orphanet, 2009), although its cause is still not clear. The disease is generally late onset, producing progressive weakening of the skeletal muscles particularly in the face, shoulder girdle, and upper arms (Padberg, 1982; Pandya *et al.*, 2008). FSHD1A (OMIM 158900), the most common form of FSHD (~98% of all cases), is associated with a contraction of the large D4Z4 tandem repeat array at chromosome 4q35 from as many as 150 repeats in the unaffected population to between 1 and 11 copies of the D4Z4 repeat unit in FSHD patients (Lunt *et al.*, 1995; Wijmenga *et al.*, 1992). Presumably, this deletion leads to epigenetic changes and subsequently upregulation of FSHD1A genes that ultimately leads to the pathology (de Greef *et al.*, 2008). Multiple candidate genes have been proposed to lead to FSHD pathology based in part on their position with D4Z4 repeats (Gabriels *et al.*, 1999; Snider *et al.*, 2009; van Deutekom *et al.*, 1996; Wijmenga *et al.*, 1993), differences in FSHD patient expression compared to unaffected controls (Ansseau *et al.*, 2009; Bodega *et al.*, 2009; Bosnakovski

et al., 2008; Gabellini *et al.*, 2006; Rijkers *et al.*, 2004; Winokur *et al.*, 2003), and phenotypes when overexpressed in animal models (Gabellini *et al.*, 2006; Hanel *et al.*, 2009; Liu *et al.*; Wuebbles *et al.*, 2009; Wuebbles, 2010). This study focuses on the FSHD candidate gene *FRG1* (FSHD region gene 1), encoding a highly conserved protein of unknown cellular function in human muscle (Grewal *et al.*, 1998).

FRG1, located 125kb centromeric to the FSHD1A deletion, was one of the early candidate genes for FSHD (van Deutekom *et al.*, 1996), yet expression studies have failed to find significant FRG1 misexpression in FSHD patient derived muscle cells and biopsies (Arashiro *et al.*, 2009; Klooster *et al.*, 2009; Masny *et al.*; Osborne *et al.*, 2007). Furthermore, there is a lack of understanding towards FRG1's normal cellular function. Studies using *Xenopus* as a model for vertebrate development found frg1 was widely expressed early and throughout development, showing elevated levels in vascular tissues and developing muscles (Hanel *et al.*, 2009; Wuebbles *et al.*, 2009). Morphilino knockdown and overexpression studies show frg1 is necessary for in development of the musculature and vasculature (Hanel *et al.*, 2009; Wuebbles *et al.*, 2009). Developmental analysis of the *C. elegans* FRG1 ortholog (FRG-1) showed a conserved expression profile throughout development (Liu *et al.*). Much like *Xenopus*, increased FRG-1 levels in *C. elegans* produced developmental defects in the musculature.

There is little known in regards to FRG1's precise function. Overexpression in cell culture originally suggested FRG1 is a nuclear/nucleolar protein (van Koningsbruggen *et al.*, 2004; van Koningsbruggen *et al.*, 2007), and has been identified as a component of the spliceosome (Rappsilber *et al.*, 2002), suggesting it is involved in RNA biogenesis. However, work in *C. elegans* show that the endogenous FRG-1 is both

a nuclear and cytoplasmic protein, and localized to muscle attachment sites (Liu *et al.*). Furthermore, FRG-1 has F-actin bundling activity consistent with its localization to muscle attachment sites (Liu *et al.*). While previous studies provide potential insight into FRG-1's function in human muscle development and disease, it is not known how these results translate to the human condition and potentially FSHD.

We have found endogenous human FRG1 is prominently expressed in muscle and vasculature, the two tissues affected in FSHD. Furthermore, we have identified three subcellular pools of FRG1: in the nucleus, cytoplasm, and on sarcomere. Interestingly, the distribution of FRG1 during human skeletal muscle myogenesis changes dramatically (Hanel et al. Differentiation, submitted). Furthermore, we find FRG1 is associated with RNA and cytoplasmic localization is dependent upon the integrity of the actin cytoskeleton (Sun et al. manuscript in preparation). Here we show FRG1 is a dynamic protein by shuttling between the nucleus and cytoplasm, suggesting that subcellular pools of FRG1 and its ability to bind RNA and bundle actin may be linked.

RESULTS

FRG1 is a nuclear-cytoplasmic shuttling protein

The endogenous FRG1 is localized in both the nucleus and cytoplasm. Nuclear shuttling assays were performed (Figure C.1) to determine if these two pools of FRG1 were linked. Murine C2C12 cells, easily identifiable by their DNA-dense nuclear foci, were transfected with a plasmid expressing epitope tagged HA-FRG1 and allowed to accumulate HA-FRG1 overnight. Cycloheximide (CHX) was added to the culture media

to block translation and the cells were fused with non-transfected HeLa cells, readily identifiable by their DNA poor nucleoli, in continued presence of CHX, and HA-FRG1 localization was monitored over time by immunocytochemistry (ICC) probing for HA. Thus, any HA signal in the HeLa cells represents FRG1 protein synthesized in the C2C12 cells. Within two hours of starting the fusion FRG1 synthesized in a C2C12 cell (Figure C.1A-D, white arrow) had begun to accumulate in the nuclei and concentrate the in nucleoli of a fused HeLa cell (Figure C.1A-D, blue arrow). This nuclear import of FRG1 was more evident at three hours (Figure C.1E-H) and at four hours appeared to have reached equilibrium (Figure C.1I-L). As the amount of cytoplasmic HA-FRG1 is almost undetectable, we deduce that much of the HeLa nuclear HA-FRG1 came from the C2C12 nuclear FRG1 and conclude that FRG1 shuttles between the nucleus and cytoplasm.

FRG1 shuttling is dependent on RNA polymerase II-mediated transcription

FRG1 has been shown to shuttle between the nucleus and cytoplasm (Figure C.1). Considering FRG1's association with RNA and interaction with TAP (Sun et al. manuscript in preparation), FRG1's nuclear shuttling was investigated further. HeLa cells, readily identified by Hoechst staining as containing large DNA-poor nucleoli (Figure C.2, blue arrows), transfected and expressing HA-FRG1 were fused with non-transfected C2C12 cells, easily identified by the multiple DNA-bright foci (Figure C.2, white arrows), in the presence of translation blocking levels of cycloheximide (CHX). FRG1 shuttling is illustrated by the synthesized in the HeLa cells being imported into the nuclei of non-transfected C2C12 cells (Figure C.2A-D). Treatment with Leptomycin B (LMB) to block the CRM1/exportin1 nuclear export pathway had no

noticeable effect on FRG1 shuttling (Figure C.2E-H compared with C.A-D, CHX alone) indicating FRG1 nuclear export was through a CRM1 independent pathway. Treatment with low levels of Act D such that RNA polymerase I was selectively inhibited similarly had no apparent effect on FRG1 shuttling or nucleolar localization (Figure C.2I-L); however, treatment with a higher dose of ActD to inhibit both RNA polymerase I and II transcription resulted in FRG1 being excluded from nucleoli and appeared consistently to be preferentially accumulated into the nuclei of recipient C2C12 cells (Figure C.2M-P). These data suggest FRG1 nuclear export is somehow linked to RNA polymerase II-mediated transcription.

DISSCUSION

Work preformed in our lab has sought to further understand the molecular function of FRG1 protein. It has been discovered that endogenous FRG1 is localized to both the nucleus and cytoplasm, and this localization changes throughout myoblast differentiation (Hanel et al. Differentiation, submitted). We also find FRG1 is associated with RNA and the cytoplasmic localization is dependent upon the integrity of the actin cytoskeleton (Sun et al. manuscript in preparation). Experiments have been preformed to determine if the endogenous FRG1 localization to both the nucleus and cytoplasm are linked (Figure C.1 and C.2). Previous cell culture studies using epitope-tagged FRG1 transgenes characterized FRG1 as near exclusively nuclear with strong nucleolar and nuclear speckle concentrations implicating FRG1 in RNA biogenesis (van Koningsbruggen *et al.*, 2004; van Koningsbruggen *et al.*, 2007). Although we do not

observe endogenous FRG1 strictly in the nucleus, it does accumulate in the nucleoli during myotube formation, and interacts with RNA, supporting a role in RNA biogenesis (Hanel et al. Differentiation, submitted; Sun et al. manuscript in preparation). When HeLa cells are HA-FRG1 recipient cells in our nuclear shuttling assays, the transiently expressed FRG1 is almost exclusively in the recipient nuclei and specifically in the nucleoli (Figure C.1), despite the endogenous FRG1 showing both cytoplasmic and nuclear staining (Hanel et al. Differentiation, submitted). Furthermore, nuclear shuttling appears to somehow be linked to RNA polymerase II-mediated transcription. These data indicate that the majority of overexpressed FRG1 protein is nuclear and preferentially nucleolar. This raises the question of how is the exogenous or overexpressed FRG1 different from the endogenously regulated FRG1? It is interesting to note that different cell types showed different ratios of nuclear to cytoplasmic FRG1 with undifferentiated and fully differentiated muscle cells showing the greatest amount in the cytoplasm. Since exogenous or overexpressed FRG1 preferentially accumulates in the nucleus, potentially a certain cell-type specific level of endogenous FRG1 is capable of being actively maintained in the cytoplasm (FRG1 is ~29kDa) at any one time and any increases in FRG1 protein levels result in default FRG1 nuclear localization. Combined with the dynamics shown by the nuclear shuttling assays, this would predict that eventually the exogenous or overexpressed FRG1 would show the similar cytoplasmic staining albeit with the intense nuclear staining, dependent on the turnover rate for the endogenous cytoplasmic retained FRG1. This is in fact exactly what was seen in the *C. elegans* study on FRG-1; the overexpressed epitope tagged FRG-1 intensely localized to the nuclei yet over time, faint but detectable FRG-1 localization was seen in the body-wall muscle

attachment sites (Liu *et al.*). We suggest that in our nuclear shuttling assays and published overexpression studies; the overexpressed FRG1 is actively shuttling between the nucleus and cytoplasm (possibly transporting RNA) but is visualized exclusively in the nuclei because it is not being readily retained in the cytoplasm. Conversely, the endogenous FRG1 is stably maintained in the cytoplasm awaiting a signal to release it to the nucleus. This cytoplasmic retention model is supported by the dramatic change in endogenous FRG1 localization to the nucleus in myoblasts upon stimulation of myogenic differentiation (Hanel et al. Differentiation, submitted). This model further predicts that even small changes in FRG1 levels would alter its subcellular distribution, aberrantly increasing its levels in the nucleus. Active cytoplasmic retention of FRG1 likely involves interaction with other proteins to anchor it. Recently we showed that FRG1 is a bona fide F-actin binding and bundling protein (Liu *et al.*), further supporting a cytoplasmic role for FRG1.

Overall, our lab has shown FRG1, a protein that is critical for muscle and vascular development, is a dynamic nuclear and cytoplasmic protein that can bind RNA, bundle actin, and localize to the sarcomeric. Furthermore, our studies place FRG1 as the only current FSHD candidate gene whose product is directly linked to the skeletal muscle contractile apparatus.

MATERIALS AND METHODS

Nuclear shuttling assay

The assay was carried out essentially as described (Kawamura *et al.*, 2002). The HA-FRG1 expression plasmid was generated by subcloning the human FRG1 coding sequence into pcDNA3.1 HA (Matzat *et al.*, 2008). Murine C2C12 cells (~60% confluent) were transfected with pcDNA3.1HA-FRG1 using TransIT-LT1 reagent (Mirus Bio LLC, Madison, WI) and allowed to grow for 24 hrs. The cells were removed by trypsinization, washed with PBS, plated on poly-L-lysine coated coverslips ($1 \times 10^5/cm^2$) and allowed to adhere for 2 hrs before non-transfected HeLa cells were overlayed ($5 \times 10^4/cm^2$) onto the transfected C2C12 cells for 3 hours. The co-cultures were incubated with 100μg/ml Cycloheximide (CHX) for 15 min to stop translation, and the cells were fused by adding 50% (wt/vol) polyethylene glycol 4000 in DMEM for 2 min. The fusions were immediately washed with DMEM and then incubated with 100μg/ml CHX for 2, 3, or 4 hrs followed by ICC analysis. The cells were fixed with 4% formaldehyde (FA) in PBS for 15 min, immunostained with HA monoclonal antibody clone 3F10 (1:100) (Roche) as described below, and co-stained stained with Hoechst 33342 (5 μg/ml in PBS).

ICC staining

HeLa, C2C12, and MDSC, were fixed in 4% FA in PBS and HSMM were fixed in 2% FA in PBS, for 15 min at room temperature (RT). After fixation, cells were permeablized with 0.25% Triton-X 100 in PBS for 10 min on ice, and subsequently

blocked with 2% BSA in PBS for 30 min at RT. Primary antibody incubations were carried out at RT for 1 hr up to overnight at 4°C and secondary antibody incubations were for 40 min at RT**.**

Microscopy

Fluorescence mages were taken by fluorescence microscopy using an Olympus BX60 microscope equipped with a SpotRT monochrome model 2.1.1 camera and Spot Advanced software (Diagnostic Instruments, Sterling Heights, MI).

REFERENCES

Ansseau, E., Laoudj-Chenivesse, D., Marcowycz, A., Tassin, A., Vanderplanck, C., Sauvage, S., Barro, M., Mahieu, I., Leroy, A., Leclercq, I., Mainfroid, V., Figlewicz, D., Mouly, V., Butler-Browne, G., Belayew, A., and Coppee, F. (2009) DUX4c is up-regulated in FSHD. It induces the MYF5 protein and human myoblast proliferation. *PLoS One* **4**: e7482.

Arashiro, P., Eisenberg, I., Kho, A.T., Cerqueira, A.M., Canovas, M., Silva, H.C., Pavanello, R.C., Verjovski-Almeida, S., Kunkel, L.M., and Zatz, M. (2009) Transcriptional regulation differs in affected facioscapulohumeral muscular dystrophy patients compared to asymptomatic related carriers. *Proc Natl Acad Sci U S A* **106**: 6220-6225.

Bodega, B., Ramirez, G.D., Grasser, F., Cheli, S., Brunelli, S., Mora, M., Meneveri, R., Marozzi, A., Mueller, S., Battaglioli, E., and Ginelli, E. (2009) Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC Biol* **7**: 41.

Bosnakovski, D., Lamb, S., Simsek, T., Xu, Z., Belayew, A., Perlingeiro, R., and Kyba, M. (2008) DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation. *Exp Neurol*.

de Greef, J.C., Frants, R.R., and van der Maarel, S.M. (2008) Epigenetic mechanisms of facioscapulohumeral muscular dystrophy. *Mutat Res* **647**: 94-102.

Gabellini, D., D'Antona, G., Moggio, M., Prelle, A., Zecca, C., Adami, R., Angeletti, B., Ciscato, P., Pellegrino, M.A., Bottinelli, R., Green, M.R., and Tupler, R. (2006) Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1. *Nature* **439**: 973-977.

Gabriels, J., Beckers, M.C., Ding, H., De Vriese, A., Plaisance, S., van der Maarel, S.M., Padberg, G.W., Frants, R.R., Hewitt, J.E., Collen, D., and Belayew, A. (1999) Nucleotide sequence of the partially deleted D4Z4 locus in a patient with FSHD identifies a putative gene within each 3.3 kb element. *Gene* **236**: 25-32.

Grewal, P.K., Todd, L.C., van der Maarel, S., Frants, R.R., and Hewitt, J.E. (1998) FRG1, a gene in the FSH muscular dystrophy region on human chromosome 4q35, is highly conserved in vertebrates and invertebrates. *Gene* **216**: 13-19.

Hanel, M.L., Wuebbles, R.D., and Jones, P.L. (2009) Muscular dystrophy candidate gene FRG1 is critical for muscle development. *Dev Dyn* **238**: 1502-1512.

Kawamura, H., Tomozoe, Y., Akagi, T., Kamei, D., Ochiai, M., and Yamada, M. (2002) Identification of the nucleocytoplasmic shuttling sequence of heterogeneous nuclear ribonucleoprotein D-like protein JKTBP and its interaction with mRNA. *J Biol Chem* **277**: 2732-2739.

Klooster, R., Straasheijm, K., Shah, B., Sowden, J., Frants, R., Thornton, C., Tawil, R., and van der Maarel, S. (2009) Comprehensive expression analysis of FSHD candidate genes at the mRNA and protein level. *Eur J Hum Genet* **17**: 1615-1624.

Liu, Q., Jones, T.I., Tang, V.W., Brieher, W.M., and Jones, P.L. Facioscapulohumeral muscular dystrophy region gene-1 (FRG-1) is an actin-bundling protein associated with muscle-attachment sites. *J Cell Sci* **123**: 1116-1123.

Lunt, P.W., and Harper, P.S. (1991) Genetic counselling in facioscapulohumeral muscular dystrophy. *J Med Genet* **28**: 655-664.

Lunt, P.W., Jardine, P.E., Koch, M.C., Maynard, J., Osborn, M., Williams, M., Harper, P.S., and Upadhyaya, M. (1995) Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD). *Hum Mol Genet* **4**: 951-958.

Masny, P.S., Chan, O.Y., de Greef, J.C., Bengtsson, U., Ehrlich, M., Tawil, R., Lock, L.F., Hewitt, J.E., Stocksdale, J., Martin, J.H., van der Maarel, S.M., and Winokur, S.T. Analysis of allele-specific RNA transcription in FSHD by RNA-DNA FISH in single myonuclei. *Eur J Hum Genet* **18**: 448-456.

Matzat, L.H., Berberoglu, S., and Levesque, L. (2008) Formation of a Tap/NXF1 homotypic complex is mediated through the amino-terminal domain of Tap and enhances interaction with nucleoporins. *Mol Biol Cell* **19**: 327-338.

Orphanet (2009) *Prevalence of rare diseases: Bibliographic data*.: Orphanet Report Series: Rare Diseases collection.

Osborne, R.J., Welle, S., Venance, S.L., Thornton, C.A., and Tawil, R. (2007) Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy. *Neurology* **68**: 569-577.

Padberg, G.W. (1982) Facioscapulohumeral Disease. Leiden, the Netherlands: Leiden University.

Pandya, S., King, W.M., and Tawil, R. (2008) Facioscapulohumeral dystrophy. *Phys Ther* **88**: 105-113.

Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12**: 1231-1245.

Rijkers, T., Deidda, G., van Koningsbruggen, S., van Geel, M., Lemmers, R.J., van Deutekom, J.C., Figlewicz, D., Hewitt, J.E., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2004) FRG2, an FSHD candidate gene, is transcriptionally upregulated in differentiating primary myoblast cultures of FSHD patients. *J Med Genet* **41**: 826-836.

Snider, L., Asawachaicharn, A., Tyler, A.E., Geng, L.N., Petek, L.M., Maves, L., Miller, D.G., Lemmers, R.J., Winokur, S.T., Tawil, R., van der Maarel, S.M., Filippova, G.N., and Tapscott, S.J. (2009) RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum Mol Genet* **18**: 2414-2430.

van Deutekom, J.C., Lemmers, R.J., Grewal, P.K., van Geel, M., Romberg, S., Dauwerse, H.G., Wright, T.J., Padberg, G.W., Hofker, M.H., Hewitt, J.E., and Frants, R.R. (1996) Identification of the first gene (FRG1) from the FSHD region on human chromosome 4q35. *Hum Mol Genet* **5**: 581-590.

van Koningsbruggen, S., Dirks, R.W., Mommaas, A.M., Onderwater, J.J., Deidda, G., Padberg, G.W., Frants, R.R., and van der Maarel, S.M. (2004) FRG1P is localised in the nucleolus, Cajal bodies, and speckles. *J Med Genet* **41**: e46.

van Koningsbruggen, S., Straasheijm, K.R., Sterrenburg, E., de Graaf, N., Dauwerse, H.G., Frants, R.R., and van der Maarel, S.M. (2007) FRG1P-mediated aggregation of proteins involved in pre-mRNA processing. *Chromosoma* **116**: 53-64.

Wijmenga, C., Hewitt, J.E., Sandkuijl, L.A., Clark, L.N., Wright, T.J., Dauwerse, H.G., Gruter, A.M., Hofker, M.H., Moerer, P., Williamson, R., and et al. (1992) Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat Genet* **2**: 26-30.

Wijmenga, C., Winokur, S.T., Padberg, G.W., Skraastad, M.I., Altherr, M.R., Wasmuth, J.J., Murray, J.C., Hofker, M.H., and Frants, R.R. (1993) The human skeletal muscle adenine nucleotide translocator gene maps to chromosome 4q35 in the region of the facioscapulohumeral muscular dystrophy locus. *Hum Genet* **92**: 198-203.

Winokur, S.T., Chen, Y.W., Masny, P.S., Martin, J.H., Ehmsen, J.T., Tapscott, S.J., van der Maarel, S.M., Hayashi, Y., and Flanigan, K.M. (2003) Expression profiling of FSHD muscle supports a defect in specific stages of myogenic differentiation. *Hum Mol Genet* **12**: 2895-2907.

Wuebbles, R.D., Hanel, M.L., and Jones, P.L. (2009) FSHD region gene 1 (FRG1) is crucial for angiogenesis linking FRG1 to facioscapulohumeral muscular dystrophy-associated vasculopathy. *Dis Model Mech* **2**: 267-274.

Wuebbles, R.D., Long, S.W., Hanel, M.L., and Jones, P.L. (2010) Testing the effects of FSHD candidate gene expression in vertebrate muscle development. *Int J Clin Exp Pathol* **3**: 386-400.

FIGURES

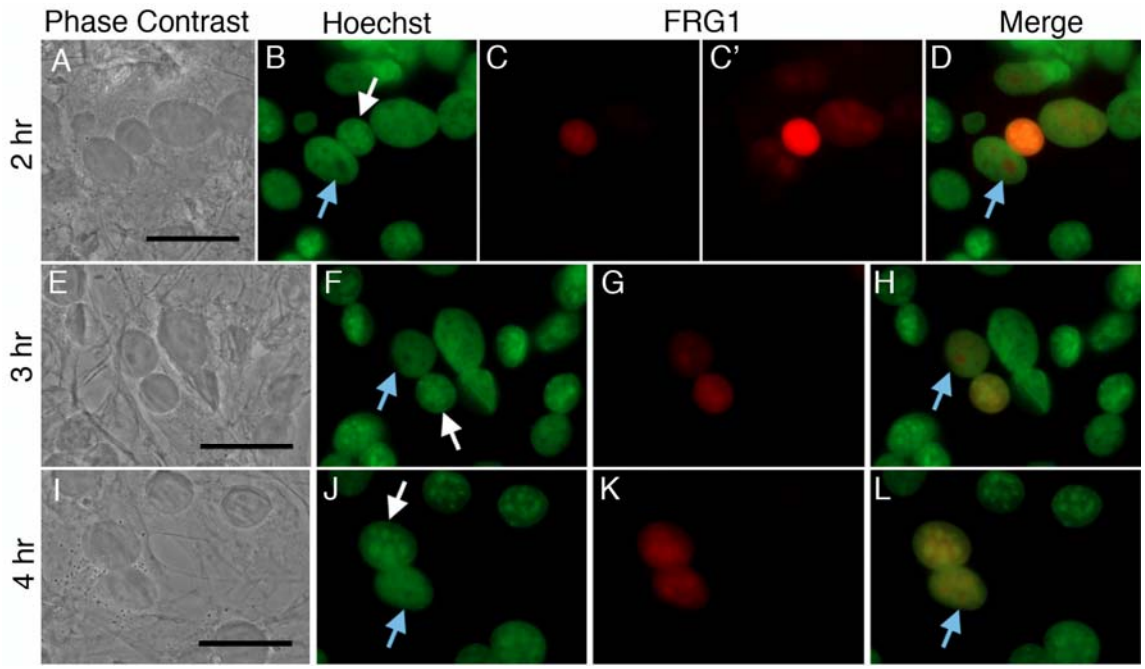

Figure C.1: FRG1 shuttles between the nucleus and cytoplasm. (A-L) Murine C2C12 cells, morphologically distinguished by their DNA-bright foci (white arrow), expressing HA-FRG1 (red) and treated with CHX were fused with HeLa cells, distinguished by their DNA-poor nucleoli (blue arrows) in the presence of CHX. (A-D) Two hours into the fusion process FRG1 translated in the C2C12 cells begins to localize in the HeLa cell nuclei (C', longer exposure of C) and specifically the nucleoli (D, blue arrows). This translocation of FRG1 from C2C12 to HeLa nuclei is more evident at 3 hours (E-H) and at 4 hours (I-L), appearing to have reached equilibrium between the two cell type nuclei (K). Hoechst 33342 staining (green) identified nuclei. Bars = 10 μm

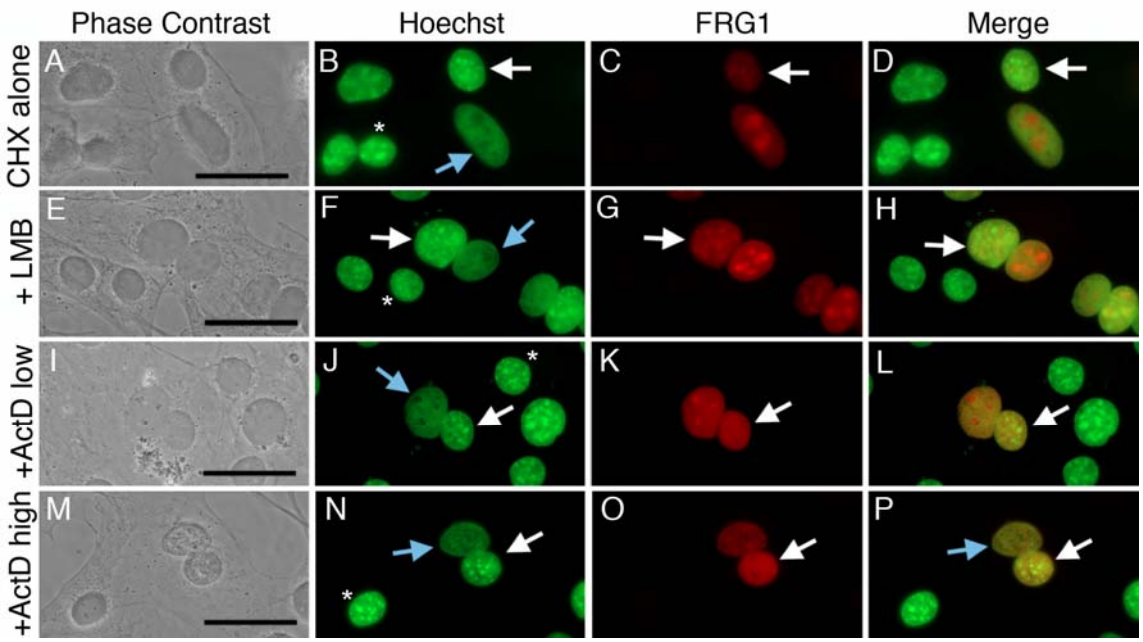

Figure C.2: FRG1 nuclear-cytoplasmic shuttling is independent of LMB but affected by inhibition of RNA polymerase II-mediated transcription. HeLa cells, distinguished by clear DNA-poor regions (blue arrow), expressing HA-FRG1 were fused to murine C2C12 cells distinguished by DNA-rich foci (white arrow) in the presence of the translation inhibitor CHX for 3 hours with or without additional treatments described. Unfused C2C12 cells (examples indicated by *) do not express HA-FRG1. (A-D) HA-FRG1 expressed in HeLa cells localizes to C2C12 nuclei 3 hours after inducing fusion, showing that FRG1 shuttles between the nucleus and the cytoplasm. Additional cell fusion experiments were carried out in the presence of LMB (E-H), low levels of ActD (I-L) to shut down RNA polymerase I mediated transcription, and high levels (M-P) of ActD to shut down RNA polymerase I and II mediated transcription, respectively. LMB or low does of ActD treatment do not affect FRG1 shuttling. Under high ActD concnetrations, HA-FRG1 accumulated in the recipient C2C12 cells, and was excluded from the HeLa nucleoli. Bars = 10 μm.