Identifying the RNA Binding Sites on the MeCP2 Protein by Regional Deletion

Ishita M. Mohan

Molecular, Cellular, & Developmental Biology Departmental Honors Thesis
University of Colorado at Boulder

Defended on: April 12, 2023

Thesis Advisor:
Dr. Lee Niswander | Professor | Department of Molecular, Cellular, and Developmental Biology

Honors Council Representative:
Dr. Jennifer Martin | Teaching Associate Professor | Department of Molecular, Cellular, and Developmental Biology

Outside Reader:
Professor Ruth Heisler | Teaching Associate Professor | Integrative Physiology

Abstract

Long non-coding RNAs (LncRNAs) have been found to bind to the Methyl-CpG binding protein (MeCP2), which has been speculated to play a role in altering chromatin structure and regulating gene expression. Whether MeCP2 has other important functions related to RNA processing, the nature of the LncRNA-MeCP2 interaction and function, and how this may affect brain development is unknown. The goal of my project is to provide insight into important regions or domains of the MeCP2 protein involved in RNA binding. With this new information, a foundation for further in-depth research will be established, on the involvement of the MeCP2 protein in the pathogenesis of Rett Syndrome, through RNA binding and RNA regulation.

Table of Contents

List of Figures

Introduction

MeCP2 Protein:

The Methyl-CpG binding protein (MeCP2) is present in cells throughout the body and is particularly abundant in neuronal cells. MeCP2 is encoded on the X chromosome and is made up of ~500 amino acids. Two isoforms of the protein are encoded from the MeCP2 gene, through alternative splicing (Ta et al, 2022) (Figure 1).

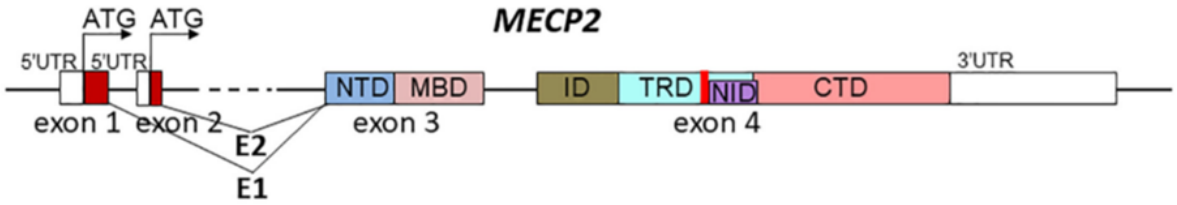

Figure 1. MeCP2 Protein Structure. MeCP2 can initiate translation within exon1 or exon 2. The resulting protein includes known domains of NTD (N-terminal Domain), MBD (Methyl Binding Domain), ID (Intervening Domain), TRD (Transcriptional Repression Domain), NID (NCoR Interaction Domain), and CTD (C-terminal Domain). From Good et al, 2021.

MeCP2 was first discovered in 1992 by Adrian Bird. MeCP2 was found to have transcriptional repression activity, due to MeCP2's ability to bind symmetrically methylated genomic DNA (5'-CpG-3'). CpG domains are composed of a cytosine nucleotide followed by a guanosine nucleotide, separated by a phosphate group that links the two together. CpG methylation within the genome, in turn, influences the ability to of MeCP2 to bind to DNA and alter chromatin structure (Figure 2).

Methylated DNA serves as an epigenetic mark, an alteration that does not change the genetic code, but influences the degree to which genes are transcribed. The mechanism involves adding a methyl group to the 5' end of the pyrimidine ring of cytosine, by the enzyme, DNA methyltransferase (Figure 2). Cytosines are therefore converted to 5-methylcytosines, and this modified DNA structure renders the DNA into a more compact form. These marks are

recognized and serve to recruit proteins, like MeCP2, involved in repression of the nearby gene.

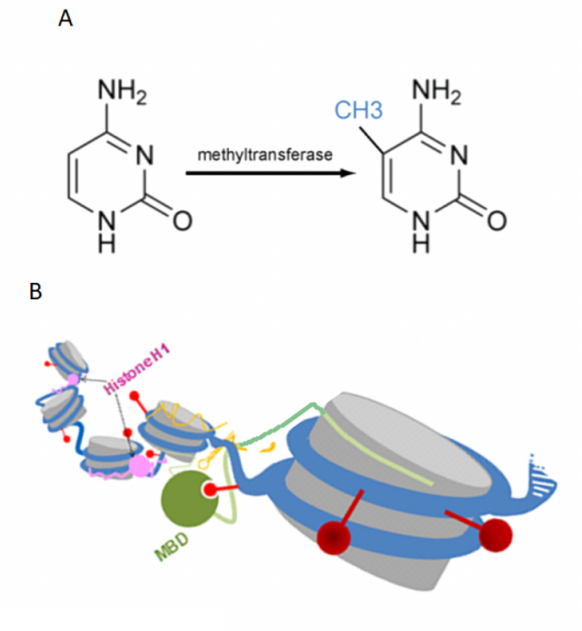

Figure 2. MeCP2 Protein Binds to Methylated Cytosine on DNA, via the MBD Domain. In the first panel (A), shown on the left is a normal cytosine nucleotide. After a methyl group is added by methyltransferase, the new conformation is 5-methylcytosine, shown on the right. (B) The panel below shows the MBD domain (green circle), binding to methylated DNA (red mark) around a histone. From Good et al, 2021.

MeCP2 not only binds to methylated DNA, but also binds to a partner protein within a complex of proteins called the nuclear receptor corepressor protein complex (NCoR), involved in silencing genes and repressing transcription. A key function of the MeCP2 protein is to act as a bridge between DNA molecules and the NCoR protein complex (Figure 3). There are two distinct domains of the MeCP2 gene and protein: the MBD (methyl binding domain) which binds to DNA (Figure 2), and the NID domain (NCoR interaction domain) domain which binds to NCoR (Figure 3). The MBD region spans amino acids 90-174, while the NID domain spans amino acids 271-322, and is separated by an intervening region (ID domain) from amino acids 174-219 (Figure 1). Mutations in the MBD region disrupt MeCP2's ability to bind to methylated DNA while mutations in the NID region disrupt MeCP2's ability to bind to NCoR. In both of these mutant scenarios, MeCP2 is unable to bridge the NCoR complex to DNA (Tillotson et al, 2017) (Figure 4).

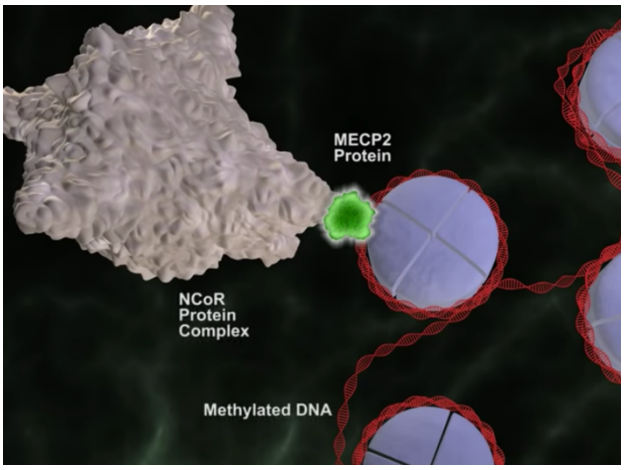

Figure 3. MeCP2 Binding Capacity. MeCP2 binds to methylated DNA through the MBD domain and NCoR through the NID domain. From Maxwell et al, 2013.

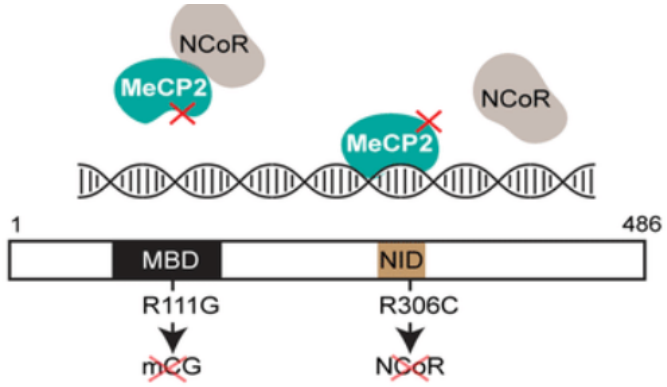

Figure 4. MBD and NID Mutations Affect MeCP2 Function. Mutations in MBD domain affect MeCP2's ability to bind to methylated DNA, while mutations in the NID domain affect MeCP2's ability to bind to the NCoR protein complex.
From Tillotson et al, 2017.

Rett Syndrome:

Rett Syndrome is a progressive neurological and developmental disorder caused by gene mutations, primarily in the MeCP2 protein (Amir et al, 1999). It is recognized in infancy and primarily affects females (1 in 10,000 births), due to its link to a paternal origin, inherited through the X chromosome (Good, Vincent, & Ausió, 2021). Symptoms include inability to communicate, loss of fine motor skills, decreased muscle tone, slow growth of the head, seizures, breathing issues, and disturbances in sleep (Kyle, Vashi, Justice, 2018).

More than 900 mutations in the MeCP2 gene have been identified in females with Rett Syndrome, which result in altered structures and function of the protein. Mutations of all types have been identified, including missense (single amino acid position is altered), nonsense (amino

acid is changed to encode a stop codon), and frameshift (reading frame of the gene is modified due to a nucleotide insertion or deletion). These mutations impair the regulation of gene expression in brain cells, disrupt alternative splicing of proteins critical for communication between neurons, and causes progressive neurological disorders, altered plasticity, and behavior (Good, Vincent, & Ausió, 2021).

The amino acids that code for MeCP2 are highly evolutionary conserved among vertebrates. However, the two primary domains where mutations significantly affect the integrity of MeCP2, and therefore lead to Rett Syndrome, include the MBD and NID domains. Mutations in other locations along the gene, do not seem to have a significant impact on phenotype expression (Tillotson et al, 2017). This data suggests that there must be important motifs located within the MBD and NID domains that play important roles in protein stability, like binding to other partner proteins, and perhaps RNAs, to influence transcriptional repression.

Long Non-Coding RNAs (LncRNAs):

A small percentage of the genome encodes protein-coding regions (~1%) while non-coding RNAs make up a large portion of transcripts of the mammalian genome. LncRNAs are composed of more than 200 nucleotides and do not code for any protein structures (Dunham et al, 2012). They are known to be important regulators in transcription within the nucleus by controlling nuclear proteins, and in the cytoplasm, they are important post-translational modifiers and play a role in mRNA stability (Yao, Wang, Y, & Chen, 2019).

Mass spectrometry results suggest that the MeCP2 protein interacts with partner proteins within the spliceosome (large protein-RNA complex that catalyzes removal of introns from nuclear pre-mRNA). Alternative splicing of RNA transcripts is critical to regulate gene

expression and some spliceosome proteins interact with MeCP2. Therefore, it has been deduced in past research, that MeCP2 functions within the spliceosome to regulate splicing. Along with spliceosome interactions, long non-coding RNAs (LncRNAs) have been found to bind to MeCP2 (Maxwell, 2013). This binding activity has been speculated to alter chromatin structure and regulate gene expression.

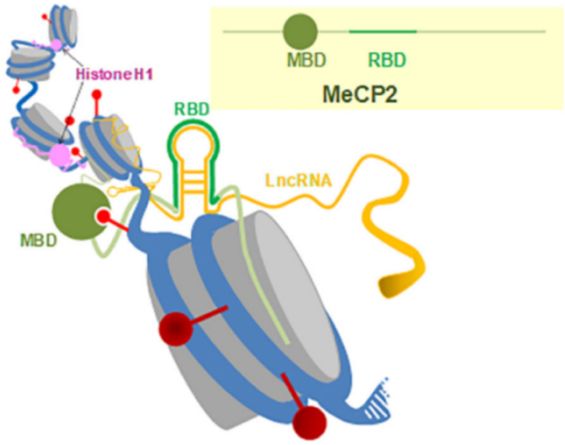

Figure 5. MeCP2, LncRNA Interactions. Through unique sequences or strucure, LncRNAs play different regulatory roles within the cell: modify chromatin structure and regulate transcription, control mRNA stability, bind, and regulate proteins. Shown is an example of how MeCP2 interacts with methylated DNA and LncRNA, through a hypothetical RNA Binding Domain (RBD). From Good et al, 2021.

These epigenetic modifiers influence gene expression through cis- or trans-factors and are tissue specific. They are highly dynamic during development, and therefore play a role in differential outcomes of gene expression in different tissues (Figure 5). Malfunctional LncRNA interactions suggest some sort of dysregulation of MeCP2 which leads to a nonfunctional protein, and therefore phenotypically presents as Rett Syndrome in some individuals (Good, Vincent, & Ausió, 2021).

For example, during neuronal development, the LncRNA *Evf2* recruits repressor protein, MeCP2, and a transcriptional activator, DLX1, to *Dlx5/6* DNA regulatory regions. The mechanisms of regulation due to the two potential binding proteins, with antagonizing functions in this region are unknown. Another LncRNA, Kcnq1ot1, acts as a control mechanism for DNA methylation, by recruiting DNMT1 (DNA methyltransferase), and increasing paternal-specific

CpG island methylation (Berghoff, 2013). The presence or absence of these LncRNAs in specific MeCP2 domains and how they regulate functions, are unknown.

Other important mechanisms that MeCP2 has in relation to RNA processing, the role underlying binding of different potential LncRNAs to MeCP2 and their interactions, and how these activities may ultimately affect brain development are unknown.

Project Objectives:

This project stems from recent findings that MeCP2 has the capability to bind and be regulated by LncRNAs. Based on the above known facts, the overarching question that underlies my research is: Which motifs of the MeCP2 protein is/are a major contributor for RNA binding and which amino acid residue/s are important for this RNA binding capacity? I propose to perform regional deletions and mutations of MeCP2 domains, to determine if binding of known RNAs expressed in neural cells is changed. I hope my results can provide insight into important domains on the MeCP2 gene involved in RNA binding.

Results

MeCP2 Domain Deletion Constructs for Structure-Function Studies:

The goal of this first set of experiments is to create specific domain deletions of the MeCP2 protein to test their role in LncRNA binding. For all MeCP2 constructs, a wildtype MeCP2_E1 (translational initiation in exon 1) plasmid was isolated and tagged with a green fluorescent protein (EGFP) reporter at the 5' end (MeCP2-GFP plasmid). To create the deletion in the NID domain, primers were created to introduce EcoRI restriction sites at the 5' and 3' ends of the NID domain (Figure 6).

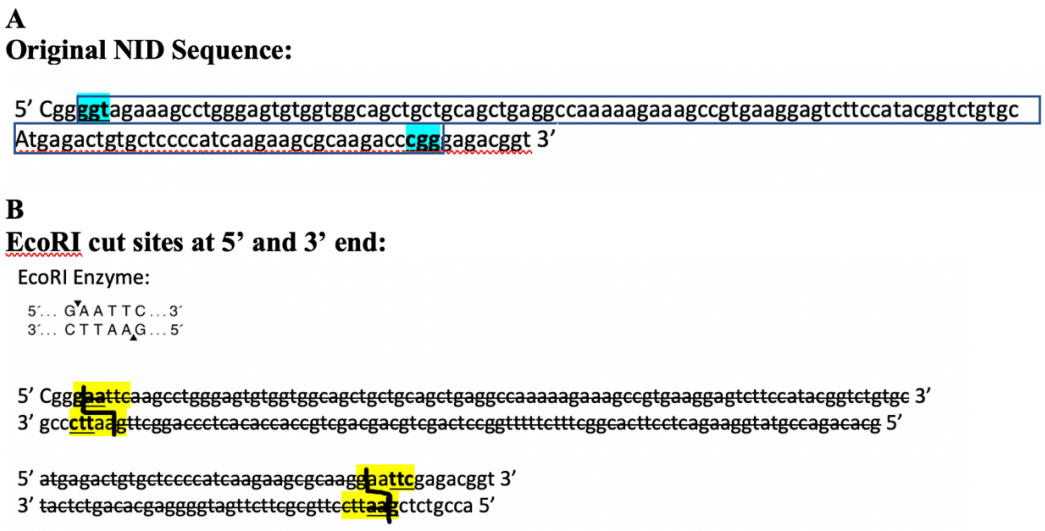

Figure 6. NID Deletion Plasmid Construction. (A) The original NID sequence within the MeCP2 DNA sequence is shown, with start and end codons highlighted in blue. The box represents the region of the domain that was deleted. (B) The EcoRI cut site, was added to the 5' and 3' ends of the NID domain (highlighted in yellow), to be targeted for enzyme digestion to create the regional NID Deletion. Details of the protocol are described in Materials and Methods section under "Plasmids and Transfection".

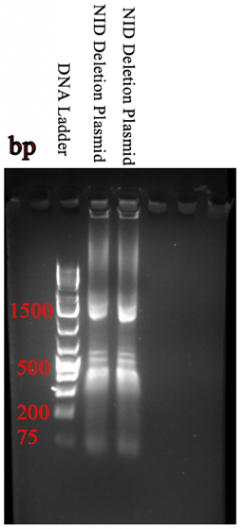

Figure 7. Confirmation of NID Deletion Plasmids. The gel shows successful transformation of NID deletion constructs.

Transformed plasmids were amplified with PCR (polymerase chain reaction) and confirmed with sequencing. Using the mutated plasmid, the restriction endonuclease enzyme, EcoRI, was used to digest the plasmid at either end of the double stranded DNA. The plasmid product was ligated in order to join the 5' and 3' ends, to create the final NID deleted plasmid (expected size: 1371 bp, Figure 7).

The plasmids were transformed and ran on a gel to check for quality and concentration (Figure 7). Large cultures were grown, and the plasmids extracted for future downstream experiments. Sequencing confirmed accurate deletion of the target NID domain, within the MeCP2-eGFP plasmid (Figure 8).

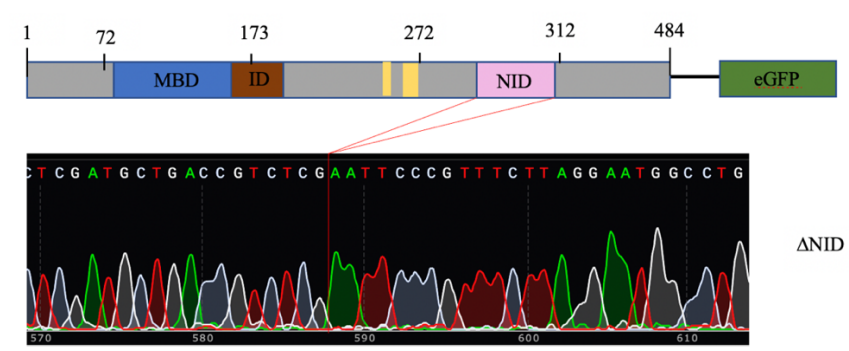

Figure 8. Sequence Data Confirming NID Deletion. Sequence data proved successful creation of the NID deletion plasmid construct. Red lines indicate the region of NID deletion within the sequence.

The same strategy was used to create other MeCP2 deletion constructs as seen in the partial MeCP2 full nucleotide sequence (Figure 9): deletions of the MBD and ID domains (Figure. 10), as well as lysine to alanine mutations in the K171-K175 and K177 region (Figure 11) were performed.

Figure 9. Partially Shown MeCP2 Full Sequence. Boxed sequences show nucleotide deletions and mutations to their corresponding domains. MBD is shown in blue, ID in brown, NID in pink, and 171/174/175 and 177 in yellow.

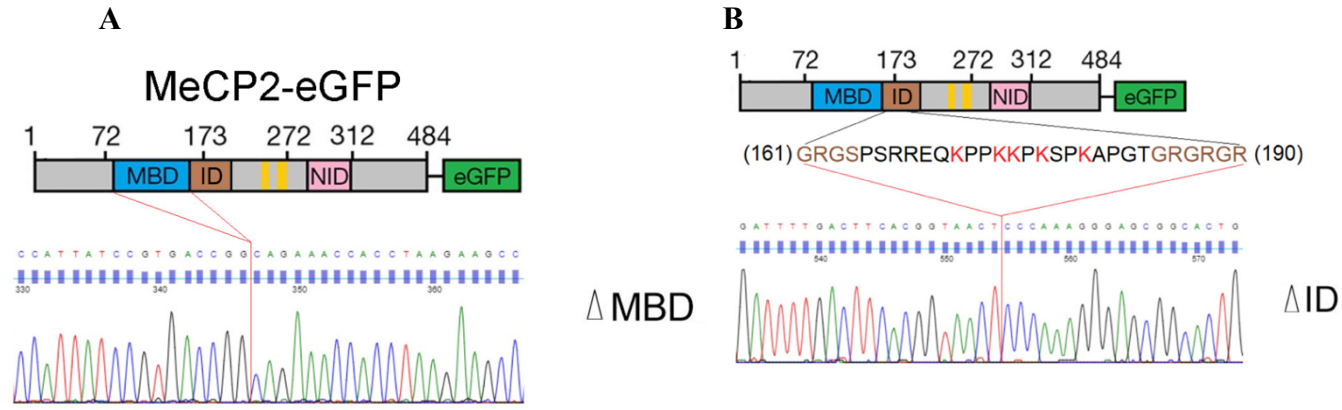

Figure 10. Sequence Data Confirming MBD and ID Domain Deletions. (A) Schematic diagram of MeCP2 protein. MBD is the defined cytosine-methylated DNA binding domain - we generated a full-length MeCP2 construct with MBD deletion. (B) ID (intervening domain) showing the thirty amino acid sequence containing the two RG repeat motifs and lysine (K)-rich region. RG repeat motifs are identified as RNA binding modules (Thandapani et al., 2013; Good, et al., 2021). We generated a full-length MeCP2 construct with ID deletion (thirty amino acids deletion). Panels show the design and sequencing trace data.

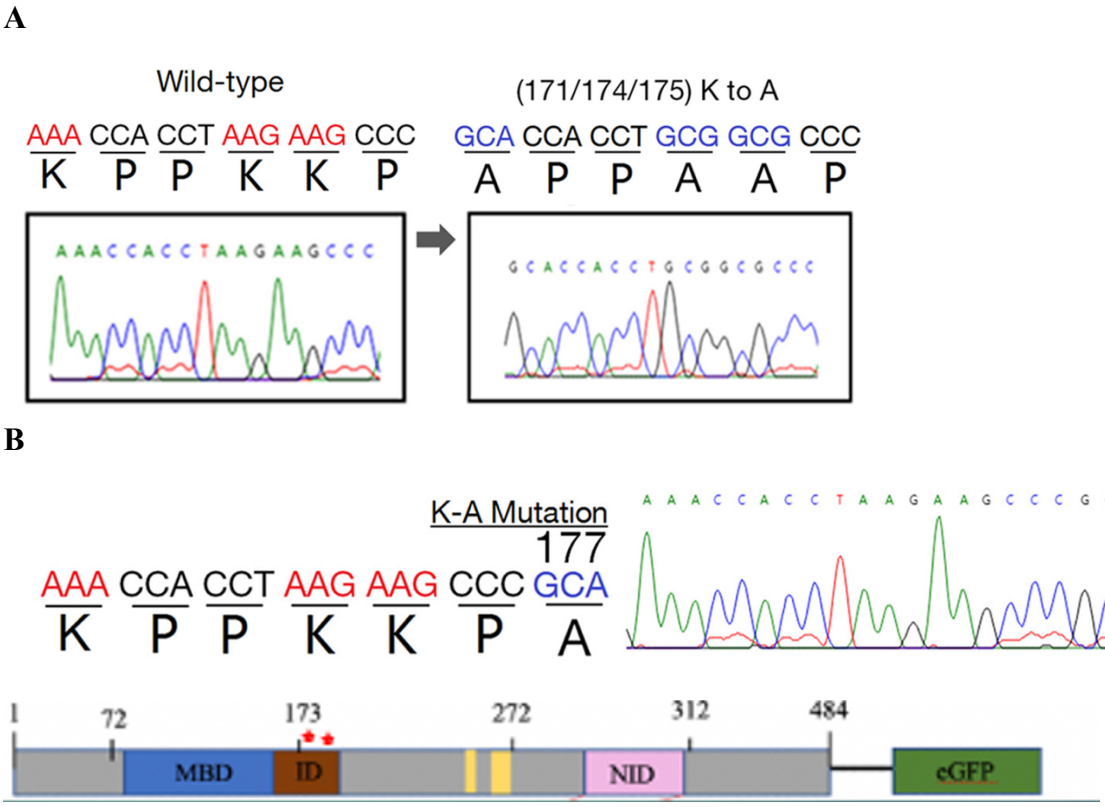

Figure 11. Sequence Data Confirming Mutations in K171-K175 and K177 Regions. (A) Schematic diagram of MeCP2 proteins. We generated a full-length MeCP2 construct with three lysine to alanine mutations in K171-K175 region (K171/174/175A). Panels show the design and sequencing trace data. (B) We generated a full-length MeCP2 construct with an individual mutation of lysine to alanine at residue K177. Panels show the design and sequencing trace data. The bottom image shows the lysine to alanine mutations, located within the ID domain.

Making Primers for Potential Known LncRNAs that Bind to MeCP2:

LncRNAs known to interact with MeCP2 based on RNA Immunoprecipitation (RIP) signal strength (Maxwell, 2013), were targeted for my experiments and include Malat1, Rncr3, Meg3, Rian, Meg9, Air, Neat1, Kcnq1ot1, and Adapt33. Primers for these LncRNAs were generated (Figure 12) using the NCBI website, where PCR product size was limited to ~80-200 nucleotides, and a T_m (primer melting temperature) between 57-63° Celsius. Most of these LncRNAs were targeted with primers designed to amplify the 5', middle, and 3' ends of the LncRNAs. Primers were then diluted to a final concentration of 10 μM and used for PCR of cDNA generated from total RNA from the NE-4C mouse embryonic neural stem cell line (see methods) to confirm correct product lengths.

LncRNA	Primer Forward	Primer Reverse	Product Length
	MALAT1:		
*** 5' end	TAATGGAGGCTGACCAGAGC	ACAAGGCCGACCTTCAAACT	88
*** middle	GGGATTGGGAAGCCCTAGTT	TGTCCTCATTTTGTCCACTGGT	94
*** 3' end	GGTTATGCCCAATTCGTCTT	AAATAGTAGGAGGTGTGTCGAC	53
	MEG3:		
*** 5' end	TTGGGGACAGTAATGGGGAC	TGGCCATCCCCATGAGAAAC	190
*** middle	GCTTCTCGAGGCCTGTCTAC	TGGAGAAGAGCGAGTCAGGA	89
	RIAN:		
*** 5' end	ATGAGAGCGTGGCTGGCATC	GTCTGTATCGTCCCTCCTTCTCT	87
*** middle	GGCTTTTTGGGCCAGGAGTA	TTACCCACACAGGGGAAAGC	120
***3' end	CACCCTGAAATCTGGGGTCC	GGCGGTCTCCGAAGTAAGTG	100
	MEG9:		
*** 5' end	ACCTCTGAGTCCCCAACTTCT	GTTCCTAAGCAAGGGCTGCTC	87
*** middle	GCCACTACTGTTCCCAAGGA	TCCTAGACCTTGCCCGATGA	85
	AIR:		
*** 5' end	CCTGCCGAGGCTTCAACATTA	CAGGATGTCTGCGTGGTAACT	90
*** 3' end	ACCTGCCAGGCCATTAGAAA	ACTTCTGGGGAAAGAAGTGTGT	94
	NEAT1:		
*** 5' end	AGGAGAAGCGGGGCTAAGTA	TAGGACACTGCCCCCATGTA	192
*** middle	GTCTTGGGCACGGCATTTTG	CTGCACTGCTACGACTCACA	233
*** 3' end	CGAACTTGGCCTCGAGAAGA	CTGGGGTCCCGATCTACAGT	84
	KCNQ1OT1:		
*** 5' end	CCCTGGCCTTTATCCTGCAA	ACCCCATGCATTGTGAGGAC	155
*** middle	AGCCTCCTGGGTCCATTGTA	GCATTGCCACAAAATGGGGT	131
*** 3' end	TTGCCGGGAAATGCAAACAC	TTCCCATCTACGTGGGCAAT	80
	ADAPT33:		
*** 5' end	CACGTTCTCGCTTCCATTCG	TTTGGCGCCATTCTCCAATC	200
*** middle	TTCTGAAGATGGGAGGTGGCT	TCTTCCAACGTCACGCATCT	80
*** 3' end	GGTGACCTCCGTTAATTCCTGG	AGAAGACAACTTTCAGGAGTCAGG	69

Figure 12. NCBI Generated Primers for LncRNAs Known to Bind to MeCP2. Table of different LncRNAs known to bind and interact with MeCP2. Forward and reverse primers were generated to amplify the 5', middle, and 3' ends of Malat1, Meg3, Rian, Meg9, Air, Neat1, Kcnq1ot1, and Adapt33, with expected product length listed beside each one.

After PCR amplification using the individual primer sets, the products were run on an agarose gel and separated by molecular size, to confirm accurate product lengths. This process was used to eliminate less efficient primers. Optimal primers were then selected for subsequent experiments. Figure 13 shows Malat1 primers run on a gel, confirming accurate 3', 5', and middle primer sizes.

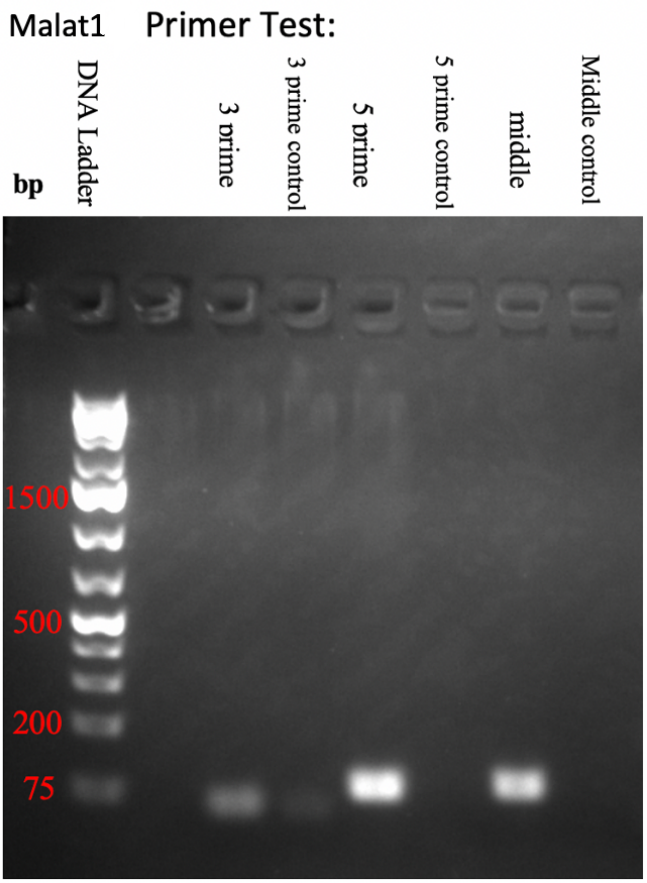

Figure 13. Malat1 3', Middle, and 5' Primer Gel Test. PCR and gel separation confirms accurate sized bands of 5', 3', and middle Malat1 primers. 3 prime lane confirms 53 bp band, 5 prime lane confirms 88 bp band, and middle lane confirms 94 bp band. Control lanes contained water rather than cDNA and acted as a negative control for each sample.

This process was repeated with each LncRNA primer set (not shown). Functional primers showed an adequate signal at the correct molecular location, relative to the DNA ladder. Less efficient primers either had a weak signal, noise, or inaccurate band locations, and were therefore excluded from further experiments (Figure 14).

Table with Working and Non-working Primer Pairs:

LncRNA	Status Based off Gel Running
MALAT1:	
*** 3' end	Good
*** 5' end	Good
*** middle	Good
MEG3:	
*** middle	VERY WEAK SIGNAL
*** 5' end	Good
RIAN:	
*** middle	Good
***3' end	NOISE
*** 5' end	Good
MEG9:	
*** middle	Good
*** 5' end	NO PRODUCT
AIR:	
*** 5' end	Good
*** 3' end	Good
NEAT1:	
*** 3' end	Good
*** middle	Good
*** 5' end	Good
KCNQ1OT1:	
*** 5' end	Good
*** middle	Good
*** 3' end	NOISE
ADAPT33:	
*** 5' end	Good
*** middle	Good
*** 3' end	Good

Figure 14. Working and Less Efficient Primers Based off PCR and Gel Separation. RNA isolated from proliferating NE-4C cells was used for PCR identification of accurate and less efficient primers for each targeted LncRNA. Gel running confirmed accurate and inaccurate primers to be used for downstream real time PCR experiments. Highlighted column on the left represents 3', 5', or middle end primers of the LncRNA. Cells highlighted in green on the right column indicate working primers confirmed by accurate band size and density, whereas unhighlighted cells represent non-working primers. Primers marked in green are appropriate for subsequent real-time PCR experiments.

Quantitative Real-Time PCR:

As MeCP2 plays vital roles in the developing nervous system, we used a mouse embryonic neural stem line, NE-4C, to study the binding of MeCP2 domain mutant proteins to brain-specific LncRNAs. NE-4C cells were grown in culture and transfected with plasmids expressing eGFP fused with wild-type and mutant MeCP2 with the help of Dr. Jing Zhang. Target LncRNAs that were bound to MeCP2 were extracted using RNA Immunoprecipitation (see methods).

After RNA Immunoprecipitation experiments, used to isolate RNA that was bound to different experimentally deleted MeCP2 protein constructs, cDNA was synthesized for subsequent analysis. Quantitative real-time PCR was performed with functional LncRNA primers. Input, RIP, and negative control IgG samples were analyzed, each from cells transfected with Wild Type, MBD deletion, ID deletion, lysine residues of 171-175 mutations, lysine residue of 177 mutation, and NID deletion, for a total testing of 18 samples per RNA target. Trials were repeated and analyzed using statistical analysis two tailed student t-tests, and visualized using the GraphPad Prism 9.0.0 (Figure 15). The results to date for two RNA targets, Malat1 and Rncr3 are shown in Figures 15 and 16. ID deletion and lysine-to-alanine mutated amino acids 171/174/175 constructs seemed to have the greatest effect on RNA binding, as seen with both Malat1 and Rncr3. The individual lysine to alanine 177 mutation more dramatically affected RNA binding for Rncr3 than for Malat1.

Malat1 5’ Quantitative Real-Time PCR

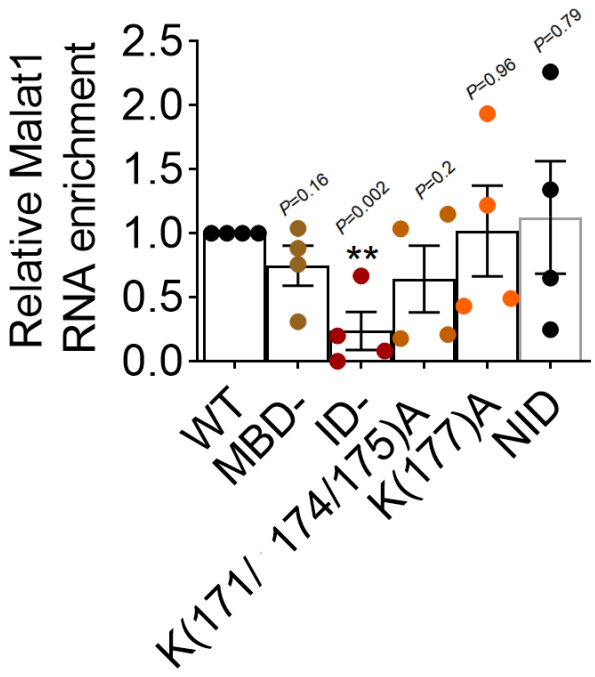

Figure 15. Malat1 5’ Quantitative Real-Time PCR Analysis. Malat1 5’ data is shown. MBD deletion did not seem to have a major effect on Malat1 binding. There was a mild effect of the lysine 171/174/175 mutation and the ID domain deletion had the greatest effect on Malat1 binding. NID and 177 deletion data shows too much variability for interpretation.

Rncr3 Quantitative Real-Time PCR Analysis

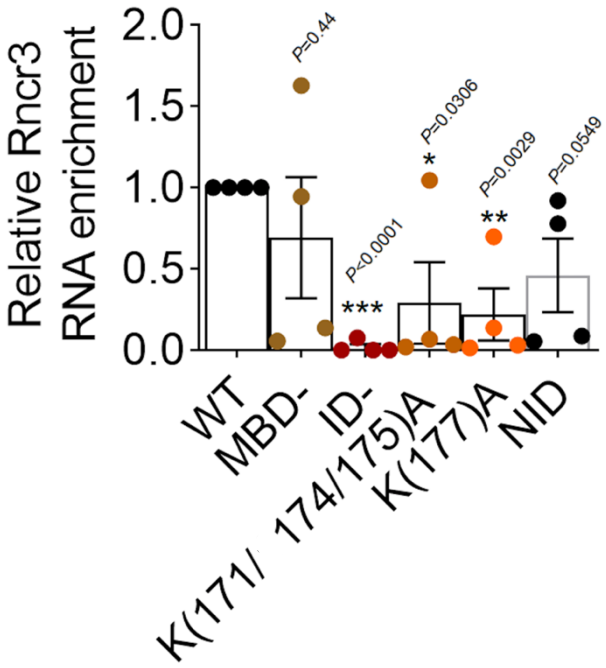

Figure 16. Rncr3 5’ Quantitative Real-Time PCR Analysis. Rncr3 5’ data is shown. MBD deletion did not seem to have a major effect on Rncr3 binding although the result was quite variable, followed by NID deletion. There was a significant effect due to mutation of lysines 171/174/175 and K177, and the ID domain deletion had the greatest effect on Rncr3 binding.

Discussion

Confirmation of Deletion Constructs and LncRNA Primers:

After site directed 5' and 3' mutations at either end of the NID domain to create EcoRI cut sites we performed restriction digests to create the NID deletion constructs. Deletions were confirmed by gel running and sequencing, to verify accurate mutations, along with all other deletions that were performed in this project (MBD, ID, K171-175, K177 – made by Dr. Jing Zhang).

Primers for target LncRNAs made using the NCBI website were confirmed by PCR analysis of RNA isolated from NE-4C cells and gel separation to check for correct product sizes. Working primers were identified to use for subsequent Real-Time PCR experiments with RIP samples.

MeCP2 Domains Contribute Different RNA Binding Activity:

Although MeCP2 is traditionally thought to bind methylated DNA, the field is only beginning to probe possible roles for MeCP2 in RNA regulation. Recent observations of MeCP2-RNA binding activity have expanded our view into the potential new regulatory mechanisms of MeCP2. These questions motivated my research project, to better understand which domains on the MeCP2 protein are major contributors to RNA binding. My data provide statistically significant evidence that different MeCP2 domains contribute differentially to MeCP2-RNA binding activity.

ID Domain:

When the ID domain was deleted, binding of target LncRNAs decreased significantly. This was seen in our Quantitative Real-Time PCR tests with Malat1 and Rncr3. The ID domain deletion showed the largest decrease in RNA binding activity compared to all other deletion constructs. This implies that the ID is a major contributor to MeCP2-LncRNA interactions. This suggests that mutations in this domain could lead to significant phenotypic changes.

MBD Domain:

On the other hand, the MBD domain does not seem to be a major contributor for MeCP2-RNA binding. This is because, for our MBD deletion construct, there was no significant decrease in LncRNA binding capacity compared to the wild type MeCP2 protein. Based on previous knowledge, the MBD is shown to be a major contributor for MeCP2 binding to methylated DNA. This suggests a separation of MeCP2-DNA binding function from MeCP2-RNA binding function, and compartmentalization of these activities to separate protein domains.

NID Domain:

The NID domain shows a partial contribution to RNA binding. However, some data was inconclusive and results are variable. Hence, a definite conclusion cannot be drawn from the data.

Speculation of Potential Competitive Binding Nature of MeCP2:

The complex nature of MeCP2 protein function is seen through these results, as MeCP2 binds to both DNA and RNA, and that MeCP2 uses different protein domains for differential binding. This suggests, there are different regulatory mechanisms that underlie MeCP2 function.

One possibility is that these activities might act together to bridge MeCP2 to both DNA and LncRNAs, simultaneously. Alternatively, I speculate that these DNA and RNA binding activities might be competitive such that when DNA is bound, RNA binding is suppressed and vice versa. The third possibility is that the domains may act independently. For example, if DNA is not methylated, MeCP2 would not be bound to DNA through the MBD domain but MeCP2 could still interact with LncRNAs to regulate gene expression. Depending on the gene, its epigenetic state, and chromatin context, all three of these possibilities could represent the complex interactions of MeCP2 in gene regulation.

Limitations of this Study:

My current data is limited to the LncRNAs tested – Malat1 and Rncr3. The next step would be to test all target genes using our LncRNA primers.

My data shows that the ID domain is important for binding to Malat1 and Rncr3. Moreover, it is possible that binding could be further sub-defined based on our data that specific lysine residues in the ID are involved in Rncr3 binding. Further studies will be needed to understand the basic binding mechanism of LncRNAs to MeCP2.

Future Studies:

The final research purpose, to check all potential binding regions of MeCP2 with a variety of LncRNAs will still need to be completed. My data is limited to two target genes (Malat1 and Rncr3). The remainder of the LncRNAs can readily be tested as we have RIP cDNA in hand and working primers. I speculate that since the ID domain had the most impact on RNA binding after testing the two target RNAs in this project, the ID domain might be consistently important for binding the remaining LncRNAs to be tested. However, the data may not be the same for all target RNAs which would pose interesting results and interpretations. One speculation from my current data is that different sub-domains of the ID will be important in LncRNA binding. Malat1 binding depends on the ID domain but mutation of the three conserved lysines within this domain had only a slight impact. In contrast, Rncr3 binding was dramatically affected by the ID deletion as well as even a single lysine to alanine mutation, suggesting that the lysine residues are important for binding Rncr3.

It will also be of interest to know whether alterations in these RNA binding activities play an important role in Rett Syndrome. It will also be of interest to test possible effects of mutations in the ID domain identified with Rett Syndrome. With better understanding of the mechanisms involved in RNA binding, we can see what pathways of neuronal development are affected and how regulation is controlled in these cells. The binding mechanisms would have to be studied more in depth to understand this relationship.

As we gain knowledge of MeCP2-RNA binding, I would hope that this knowledge could be used to inform clinical studies and viewed as a potential new strategy to prevent or rescue phenotypically presenting Rett Syndrome symptoms.

Materials and Methods

Cell Culture (performed by Ishita Mohan and Dr. Jing Zhang):

Mouse embryonic neural stem cell line NE-4C cells (ATCC #CRL-2925) were grown at 37°C and 5% CO2 under proliferation conditions in MEM (Invitrogen) supplemented with 5% FBS, 1×MEM nonessential amino acids (Invitrogen) and 1×GlutaMax (Invitrogen).

Quantitative Real-Time PCR:

For quantitative real-time PCR assays, total RNA was extracted from NE-4C cells using Quick-RNA TM Miniprep Kit (ZYMO RESEARCH). For the measurement of target transcripts, the total RNA was used for reverse transcription. cDNA was synthesized using random primer mix (NEB, S1330S) and 10 ul of purified RNA by First Strand cDNA Synthesis Kit (NEB, #M0253). Quantitative real-time PCR was performed with the LightCycler ® 480 Real-Time PCR System (Roche).

Primer Table for NID Deletion:

Primers	**Sequence - 5' to 3'**
NID 5' end - forward	gacccccaggccattcctaagaaacgg**gaattc**aagcctgggagtgtggtggcagctgct
NID 5' end - reverse	agcagctgccaccacactcccaggctt**gaattc**ccgtttcttaggaatggcctgggggtc
NID 3' end - forward	gtgctccccatcaagaagcgcaag**gaattc**gagacggtcagcatcgaggtcaag
NID 3' end - reverse	cttgacctcgatgctgaccgtctc**gaattc**cttgcgcttcttgatggggagcac

Primer Table for MBD Deletion (made by Dr. Jing Zhang):

Primers	**Sequence - 5' to 3'**
MBD 5' end - forward	cagcggcgctccattatccgtgaccgg**gaattc**atgtatgatgaccccaccttgc
MBD 5' end - reverse	gcaaggtggggtcatcatacat**gaattc**ccggtcacggataatggagcgccgctg
MBD 3' end - forward	gtaactgggagagggagcccctccagg**gaattc**cagaaaccacctaagaagccca
MBD 3' end - reverse	tgggcttcttaggtggtttctg**gaattc**cctggaggggctccctctcccagttac

Primer Table for ID Deletion (made by Dr. Jing Zhang):

Primers	**Sequence - 5' to 3'**
ID 5' end - forward	ccctaatgattttgacttcacggtaact**gaattc**gggagcccctccaggagagag
ID 5' end - reverse	ctctctcctggaggggctccc**gaattc**agttaccgtgaagtcaaaatcattaggg
ID 3' end - forward	ctccaggaactggcaggggtcgg**gaattc**cccaaagggagcggcactgggagacc
ID 3' end - reverse	ggtctcccagtgccgctccctttggg**gaattc**ccgacccctgccagttcctggag

Plasmids and Transfection:

For MeCP2-eGFP RIP experiments, the Lipofectamine™ 3000 Transfection system (Thermo Fisher; L3000001) was used to transfect NE-4C cells with plasmids expressing eGFP fused with wild-type MeCP2 (pEGFP-N1_MeCP2(WT), gift from Adrian Bird; Addgene plasmid # 110186; http://n2t.net/addgene:110186; RRID: Addgene_110186) or other MeCP2 mutants. Other MeCP2 mutations were: (1) MBD deletion, in which 78 amino acids were deleted encompassing glycine 92 to C-terminal glutamic acid 169 (NP_034918.1). (2) ID deletion in which 30 amino acid residues (GRGSPSRREQKPPKKPKSPKAPGTGRGRGR) of the ID

domain were deleted. (3) NID deletion in which 41 amino acid residues (GRKPGSVVAAAAAEAKKKAVKESSIRS VHETVLPIKKRKTR) of the NID domain were deleted (Primer Tables Shown Above). (4) Lysine to alanine mutations in K171-K175 region (K171/174/175A) in MeCP2 and individual K-to-A 177 mutation generated using QuikChange II Site-directed mutagenesis kit (Agilent, #200521). Mutations 2-4 were made by Dr. Jing Zhang.

RNA Immunoprecipitation (RIP) Assay (Performed by Dr. Jing Zhang):

RIP was adapted from the manufacturer's specifications of RNA ChIP-IT Magnetic Chromatin Immunoprecipitation Kit (Active Motif, #53024). A 10 cm dish (1-2 × 10 7 cells) of NE-4C cells were used per RIP experiment. Cells were cross-linked in 5 mL of 1% formaldehyde PBS solution for 10 min at room temperature with rotation. Cross-linking was stopped by 5 mL 125 mM glycine PBS solution and 10 min incubation with rotation. Cells were harvested and resuspended in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH8.0, 0.5 mM PMSF, 40 U RNase inhibitor, and 5 µL Protease Inhibitor Cocktail) on ice for 30 min. The total RNA in lysed material was sheared to between 100 and 500 nt using a sonicator and checked by gel electrophoresis (note the distance between exons 2/3 primers and exon 4 primers is 1337 nucleotides in the spliced transcript). Nuclear membrane and debris were pelleted by centrifugation at 14,000 rpm for 10 min. The sonicated lysed material was treated with DNase I before performing the IP experiment. To 10 µL Input sample was added 88 µL RNA IP Buffer, 0.5 µL RNase Inhibitors and 2 µL 5 M NaCl, (stored at -80°C). RIP reactions were performed by adding pre-washed protein G magnetic beads (25 µL per reaction), RIP buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH 8.0, and 167 mM NaCl) 15 µL, sheared

material (~100 µL), RNase inhibitor, Protease Inhibitor Cocktail, with the reaction components kept on ice during mix preparation. Anti-eGFP antibodies (4 µg per RIP reaction) were then added to a total reaction volume of 150 µL, and incubated on a rotator overnight at 4°C. RNA complex bound beads were washed four times with ice cold 200 µL Complete RNA-ChIP Wash Buffer 1 and two times with ice cold 200 µL Complete RNA-ChIP Wash Buffer 2 (50 mMTris, pH7.4, 1M NaCl, 1% NP-40, and 1% sodium deoxycgolate), removing as much supernatant as possible without disturbing the beads. Complete RNA-ChIP Elution Buffer (50 mMTris, pH7.4, 1M NaCl, 1% NP-40, 1% sodium deoxycholate, and 1M Urea) was added to the beads and the pellet resuspended by rotation for 30min at room temperature. A magnet was used to pellet the beads and the supernatant transferred into a fresh tube to which 2 µL 5M NaCl and 2 µL Proteinase K was added. The Input control sample was treated with 2 µL Proteinase K. The samples were incubated at 42°C for 1 hour to digest the protein and then incubated for 1.5 hour at 65°C to reverse the cross-link. RNA was purified with TRIzol (Sigma-Aldrich) and finally diluted with DEPC water. Three independent RIP experiments were performed for each analysis. The efficiency of RNA-protein IP assay to detect protein binding to the particular regions of Rncr3 RNA can be calculated using the following formula, which includes a comparison between input samples and RIP samples (antibody RIP or IgG RIP):

$$\% \text{ of input (recovery)} = 2\ (Ct\ input - Ct\ sample) * Fd \times 100\%$$

Here, Ct input and Ct sample are threshold values obtained from exponential phase of qPCR for a particular sample of protein; Fd is a dilution factor of the input RNA to balance the difference in the amounts of RIP samples and input RNA used for qPCR. Finally, relative enrichment of target RNA in the corresponding control group was normalized as 1.

Statistical Analysis:

Two-tailed Student's t-tests were used to analyze the experimental data. Values are represented as mean ± s.e.m.; $P < 0.05$ is considered as statistically significant. Statistical analysis was performed using SPSS (25.0.0) and GraphPad Prism 9.0.0 (121). GraphPad Prism 9.0.0 (121) was utilized for data visualization.

Acknowledgements

Most importantly, I would like to thank Dr. Niswander for providing me with the opportunity to learn and conduct this project at the Niswander Lab. The skills and knowledge I have acquired have been invaluable and the experience has inspired me to move further along in the field of Molecular, Cellular, and Developmental Biology. Further, I would like to extend my sincere thanks to senior research associate and mentor, Jing Zhang, for his countless hours of training and valuable advice throughout my time at the Niswander Lab. Finally, I would like to thank my thesis committee, Dr. Niswander, Dr. Jennifer Martin, and Professor Ruth Heisler, for their time to advise me on this project.

References

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics*, 23(2), Article 2. https://doi.org/10.1038/13810

Berghoff, E. G., Clark, M. F., Chen, S., Cajigas, I., Leib, D. E., & Kohtz, J. D. (2013). Evf2 (Dlx6as) lncRNA regulates ultraconserved enhancer methylation and the differential transcriptional control of adjacent genes. *Development*, 140(21), 4407–4416. https://doi.org/10.1242/dev.099390

Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B.-K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., … HudsonAlpha Institute, C., UC Irvine, Stanford group (data production and analysis). (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature, 489*(7414), Article 7414. https://doi.org/10.1038/nature11247

Good, K. V., Vincent, J. B., & Ausió, J. (2021). MeCP2: The Genetic Driver of Rett Syndrome Epigenetics. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.620859

Hewitt, J., & Xpress, M. (n.d.). *Decoding Rett syndrome: New pieces to the puzzle*. Retrieved March 29, 2023, from https://medicalxpress.com/news/2013-06-rett-syndrome-protein-surrenders-secrets.html

Kyle, S. M., Vashi, N., & Justice, M. J. (2018). Rett syndrome: A neurological disorder with metabolic components. *Open Biology*, *8*(2), 170216. https://doi.org/10.1098/rsob.170216

Maxwell, S. S., Pelka, G. J., Tam, P. P., & El-Osta, A. (2013). Chromatin context and ncRNA highlight targets of MeCP2 in brain. *RNA Biology*, 10(11), 1741–1757. https://doi.org/10.4161/rna.26921

Ta, D., Downs, J., Baynam, G., Wilson, A., Richmond, P., & Leonard, H. (2022). A brief history of MECP2 duplication syndrome: 20-years of clinical understanding. *Orphanet Journal of Rare Diseases*, *17*(1), 131. https://doi.org/10.1186/s13023-022-02278-w

Thandapani, P., O'Connor, T. R., Bailey, T. L., & Richard, S. (2013). Defining the RGG/RG motif. *Molecular Cell*, *50*(5), 613–623. https://doi.org/10.1016/j.molcel.2013.05.021

Tillotson, R., Selfridge, J., Koerner, M. V., Gadalla, K. K. E., Guy, J., De Sousa, D., Hector, R. D., Cobb, S. R., & Bird, A. (2017). Radically truncated MeCP2 rescues Rett syndrome-like neurological defects. *Nature*, 550(7676), 398–401. https://doi.org/10.1038/nature24058

Yao, R.-W., Wang, Y., & Chen, L.-L. (2019). Cellular functions of long noncoding RNAs. *Nature Cell Biology*, *21*(5), Article 5. https://doi.org/10.1038/s41556-019-0311-8