Chromatin context and ncRNA highlight targets of MeCP2 in brain

Scott S. Maxwell[1], Gregory J. Pelka[5], Patrick P. L.Tam[5,6] and Assam El-Osta[1,2,3,4,7]

[1]Epigenetics in Human Health and Disease Laboratory.
[2]Epigenomic Profiling Facility.
Baker IDI Heart and Diabetes Institute, The Alfred Medical Research and Education Precinct, Melbourne, Victoria, 3004, Australia.
[3]Department of Pathology, The University of Melbourne, Parkville, Victoria, Australia.
[4]Faculty of Medicine, Nursing and Health Sciences, Monash University, Victoria, Australia.
[5]Embryology Unit, Children's Medical Research Institute, Sydney, Australia.
[6]Sydney Medical School, University of Sydney, Sydney, Australia.

[7]Corresponding Author;

Assam El-Osta
Epigenetics Laboratory, BakerIDI
75 Commercial Road, Melbourne, Victoria, Australia, 3004
Email: assam.el-osta@bakeridi.edu.au
Telephone: 613-8532-1389
Fax: 613-8532-1100

Scott Maxwell <Scott.Maxwell@bakeridi.edu.au>
Gregory Pelka <gpelka@gmail.com>
Patrick Tam <PTam@cmri.org.au>
Assam El-Osta <Assam.El-Osta@bakeridi.edu.au>

Running Title: Contextual targeting of MeCP2 function in brain

Keywords: MeCP2, chromatin, gene expression, DNA methylation, spliceosome, ncRNA

Abstract

The discovery that Rett syndrome (RTT) is caused by mutation of the methyl-CpG-binding-protein MeCP2 provided a major breakthrough in understanding the neurodevelopmental disorder and accelerated MeCP2 research. However, gene regulation by MeCP2 is complicated. The current consensus for MeCP2 remains as a classical repressor complex, with major emphasis on its role in methylation-dependent binding and repression. However, recent evidence indicates additional regulatory roles, suggesting non-classical mechanisms in gene activation. This has opened the field of MeCP2 research and suggests that the gene targets may not be the usual suspects, that is, dependent only on DNA methylation. Here we examine how binding in chromatin fractions and sequence preference may confer MeCP2 functionality, and connect relevant pathways in an active genome. Through evidence indicating MeCP2 spliceosome interaction, we discovered broad MeCP2 enrichment of the transcriptome while our focus toward long non-coding RNA (lncRNA) revealed MeCP2 association with RNCR3. Our data may indicate an as yet unappreciated role between lncRNA and MeCP2. We hypothesize that ncRNA may mediate chromatin-remodelling events by interacting with MeCP2, thereby conferring changes in gene expression. We consider that these results may suggest new mechanisms of gene regulation conferred by MeCP2 and its interactions upon chromatin structure and gene function.

Sequencing data has been submitted to NCBI Gene Expression Omnibus (GEO) under the accession number GSE38324. This data is currently not public and may be viewed using the reviewer link;

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=hnwzxwkaukysaly&acc=GSE38324

Introduction

The functional role of MeCP2 has become increasingly complex since its discovery.[1, 2]. The first designation of MeCP2 as a global repressor was initially established through methylated CpG di-nucleotide affinity and the protein's ability to reduce gene expression, either alone or as part of a complex.[3, 4] There are at least two categories of methylation mediated gene suppression. DNA methylation can suppress gene expression by directly interfering with the binding of sequence-specific transcription factors.[5] The alternative indirect mechanism involves methyl-CpG dependent proteins and the best characterized is MeCP2.[3, 4] In recent years, there has been a huge increase in the number of characterized repression mechanisms and only a few of them, including MeCP2 can easily be placed into these two categories. Any strict classification of regulation is further complicated by the complexity and diversity of protein-protein interactions involved in the regulation of transcription. Many aspects are likely to contribute to MeCP2 functionality, including A/T nucleotide enhanced binding[6], high intrinsic disorder[7] and domain interaction, encompassing the recruitment of co-factors.[8] Beyond MeCP2's association with co-repressors, several interactions were thought to be incongruous at first. The association of MeCP2 with the chromatin remodeling complex factors BRM and ATRX[9, 10], CREB1[11], RNA and splicing factors[12-14] are not exclusively repressive and therefore more indicative of a dynamic and multifunctional methyl-CpG binding protein.

Mutation of MeCP2 in humans is most notably linked to RTT, a severe autism spectrum disorder.[15] The predominant neurological phenotype is consistent with high expression levels of MeCP2 in neurons and its effect on nucleosome assembly.[16, 17] More specifically, reduced histone H1 expression in neurons has been attributed to MeCP2 competition for its nucleosome site.[17, 18] However, the contribution and targeting of differential gene regulation by MeCP2 still remains poorly characterized. Genome wide analyses of MeCP2 have revealed a broad binding distribution following patterns of CpG methylation.[17, 19] Interestingly, the specific binding preference for methylated CpG is only 2-3 fold greater than its unmethylated variant *in vitro*.[20] A review of gene expression analysis in autism spectrum disorders and Rett Syndrome studies suggest the involvement of an abnormal immune response, oxidative stress, altered mitochondrial function, and deregulated protein synthesis but again provide limited explanation as to why MeCP2 may affect these pathways.[21-25]

Non-coding RNAs (ncRNAs) categorized under either short or long classes, can manipulate epigenetic regulatory pathways.[26, 27] Notably, small interfering RNAs (siRNAs) have been indicated to mediate sequence-specific formation of chromatin by guiding the activity of conserved RNAi

machinery components.[28] This example, and other mechanisms of ncRNA mediated chromatin assembly are able to regulate transcription[29, 30] in the absence of DNA methylation.[31] Perhaps the most prominent example of ncRNA regulation in mammals involves the long ncRNA (lncRNA) Xist, which induces silencing of one of the two X chromosomes in females.[32] The role of ncRNAs that mediate regulatory control in the brain are yet to be characterized, however, recent discoveries indicate that ncRNAs can mediate regulatory events with specific roles that include the regulation of gene expression[33], as well as chromatin structure and function.[31, 34-37] Here, our studies extend the understanding of MeCP2 functionality in the brain. Using massive parallel sequencing technologies and context analysis we specifically examine chromatin and sequence attributes, highlighting potential targeting associations to relevant pathways, and explore the association of MeCP2 with ncRNA. We consider the findings presented here indicate a new role for MeCP2 interacting with ncRNAs to mediate chromatin remodeling changes and gene regulatory events in the brain.

Results

Chromatin fractionation reveals MeCP2 binding characteristics in the active genome

The genome is comprised of a range of varying chromatin structures with diverse functions. Broadly, chromatin has either active and open euchromatic or inactive and compact heterochromatic properties, which can be fractionated due to differential resistance to Micrococcal Nuclease (MNase) digestion.[38, 39] Euchromatin is typically represented in the first fraction, 'S1', as its open structure is readily digestible by MNase. Heterochromatin in contrast has higher resistance to MNase and is represented in a subsequent 'S2' fraction.[38, 39] MeCP2 is largely localized to heterochromatic regions consistent with high levels of methylated DNA and is required for regulating chromatin structure and remodeling changes during neuronal maturation.[40-42] In order to better understand MeCP2 binding we performed chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-Seq) experiments using sonicated chromatin from whole mouse brain, as well as MNase fractionated and enriched S1 euchromatin from whole mouse brain and the cerebellum. In order to account for sonication and digestion biases that may occur due to nucleosome positioning and other factors[43, 44], all ChIP-Seq data was analyzed relative to input controls. We designed analysis probes around the transcription start site (TSS) of all genes (MM9 assembly) +/- 1000 bp, reflecting the averaged peak distribution of CpG di-nucleotides (**Figure 1A**). Binding of MeCP2 relative to input material using sonicated chromatin from whole male mouse brain is shown in **Figure 1B**. The MeCP2 enrichment of TSS regions increases with CpG di-nucleotide content in S1 euchromatin, as compared to sonicated whole chromatin (**Figure 1C**). Indeed, this enrichment is evident in the cerebellum of MNase digested chromatin that represents mono-nucleosome fractions (**Figure 1D**), and further emphasized in the S1 di-nucleosome derived fractions (**Figure 1E**). Interestingly, there appears to be a critical content threshold of 50 CpG di-nucleotides per 2kbp TSS region. When exceeding this level, MeCP2 enrichment occurs to a greater extent in S1 fractions and when below, shows enrichment in sonicated whole chromatin. These results highlight relative differences in MeCP2 enrichment between the selective (S1) and non-selective (sonication) techniques, revealing greater association with higher CpG content regions in the active genome as compared with viewing the bulk genome overall.

There is a close correspondence of CpG methylation correlated with transcriptional suppression in the exiting literature.[45] MeCP2 binds strongly to methylated DNA and the predominant model has been a classical co-repressor complex serving to recruit regulatory determinants to suppress gene expression. However, recent experimental evidence has questioned whether the classical view of MeCP2 is an accurate reflection of its role in mediating changes in chromatin structure and transcriptional activity.[46] A histogram of TSS region CpG content highlights a distribution bias

against content of approximately 50 CpGs per region of which corresponds with greatest methylation capture (**Figure 2A**). In examining the methylation profile of all TSS regions, we observe lower methylation levels above the threshold of 50 CpGs (**Figure 2B**). Coincidently, this bias point is the same threshold level that separates MeCP2 enrichment (**Figure 1B-E**). As the presence of an un-methylated CpG island near a TSS is commonly associated with actively expressed genes, the MeCP2 enrichment of S1 di-nucleosome material with higher CpG content may indicate chromatin binding that could be associated with gene expression.

MeCP2 CG-AT4 interaction is enhanced in active chromatin fractions

MeCP2 DNA binding is enhanced by the presence of at least four consecutive A or T nucleotides in proximity to the CpG di-nucleotide or CG-AT4.[6] To investigate the effect of CpG/AT proximity, we extrapolated 200 bp of sequence for 100,000 ChIP-Seq reads for both input and bound samples and quantified CG-n-AT4 enrichment for 'n' representing up to 16 random nucleotides. MeCP2 displayed almost identical enrichment patterns across the sonicated, mono- and di-nucleosome data sets, with peaks between 0 to 4, 9 to 11, and 15 random nucleotides (**Figure 3A**). The strongest trend was identified in the di- nucleosome data of cerebellum, while whole brain sonicated material was the weakest (**Figure 3A**). We also quantified enrichment according to CpG di-nucleotide or CG-n-AT4 motif frequency, with 'n' of '0 to 15'. Greater MeCP2 enrichment was found with increasing CG-AT4 frequency compared with the same number of CpG di-nucleotides (**Figure 3B-C**). These results suggest a preference for CG-AT4 sequences and highlight differential overall contribution to MeCP2 binding between chromatin environments. To examine the relationship between CG-AT4 content and gene expression, we performed comparative RNA-Seq expression analysis of male wild type and MeCP2 null mouse cerebellum and examined expression by CG-AT4 content within TSS regions. We then considered the ratio of gene expression differences as potentially activated versus potentially repressed between samples comparatively. Binning genes according to CG-AT4 content revealed an overall reduction of the expression ratio between wild type samples as CG-AT4 content increased (**Figure 3D**). This suggests that gene expression is more consistent between samples as CG-AT4 content increases. In contrast, the comparison between wild type and MeCP2-null cerebellum samples showed a greater proportion of genes were potentially activated in wild type with increasing CG-AT4 content (**Figure 3E**). These trends were also seen if genes were separated into groups with TSS CpG content above or below the 50 CpGs threshold (**Figures 3D-E)**. Hence, we can infer from these data that the binding of MeCP2 and

interaction with CG-AT4 over these specific regions may explain, in part, the differences observed with gene expression.

Methylation, CG-AT4 TSS content and MeCP2 differential regulation

The above results indicate the interaction of MeCP2 with active chromatin fractions, which is consistent prior microarray analysis showing functional association are associated with active gene expression.[46] To further explore this idea, we employed chromatin immunoprecipitation and methylation sequencing (Methyl-Seq) to compare methyl-CpG profiles and MeCP2 binding across TSS regions. Methylated DNA enrichments were prepared using a Methylminer kit (Invitrogen) with stepwise 350mM (lower methylation density) and 2000mM (higher methylation density) salt gradient elutions to interrogate differences associated with methyl-CpG content. In the high salt extractions, only the high CG-AT4 content profile was consistent with methylation-dependent repression (**Figure 4A**). The high CG-AT4 content profile of MeCP2 repressed genes indicate stronger methylation levels in comparison to activated genes. This was in contrast to the low CG-AT4 content group in which methylation of MeCP2 repressed genes was lower than the activated gene profile (**Figure 4B**). This indicates that an increasing number of methylated CG-AT4s in the TSS region resemble the “classical” repressive behavior of MeCP2. However, when we analyzed the density ratio of high verses low salt fractions for CpG methylation, a pattern of comparatively higher methylation densities that are up- and down-stream of the transcription start sites of repressed genes is revealed, while the activated profiles indicate peak methylation density downstream of the TSS (**Figure 4C**). Although a similar pattern may be observed between the high and low CG-AT4 content groups, with respect to relative density, a prominent difference remains in the high CG-AT4 group (**Figure 4D**).

While a clear relationship between differential gene expression and MeCP2 binding is evident in TSS, the relationship between CG-AT4 content on di-nucleosomes is less clear (**Figure 4E**). For example, inspection of the relationships between high and low CG-AT4 MeCP2 binding profiles on differentially regulated genes is less readily apparent on di-nucleosomes (**Figure 4F**). However, while MeCP2 binding profiles between differentially regulated genes show subtle variations, what is readily apparent upstream of the TSS is the lack of MeCP2 di-nucleosome enrichment which is consistent with nucleosome free positioning. These data illustrate a relationship between CG-AT4 content and MeCP2 binding status with specific regulation of gene expression.

CG-AT4 frequency and MeCP2 gene targets

The enrichment of MeCP2 at active sites of chromatin and the relationship with high CG-AT4 content suggests that gene targets in the brain share these genomic features that regulate MeCP2 mediated gene expression. To investigate this we used a bioinformatics approach and screened CG-AT4 content in both mouse and human genomes as a potential marker for MeCP2 interaction. Sequences were extracted for regions 1Kbp +/- of the TSS, while counting CG and CG-AT4 motif content. Approximately 3000 of the highest or lowest genes according to TSS motif content were analyzed for gene pathway association (**Table. 1**). KEGG pathways analysis of high CpG content genes associated largely with cancer, while low CG or CG-AT4 content genes link to olfactory and compound metabolic pathways. However, while there were clear relationships with CG content, there were also strong relationships readily apparent when quantified for the occurrence of CG-AT4 motifs. For example, inspection of our large datasets demonstrates that for high CG-AT4 content genes, there are significant relationships of pathway analysis demonstrating links to neurodegenerative disease and pathways associated with MeCP2 (**Table. 1**). These included oxidative phosphorylation, Huntington's, Alzheimer's and Parkinson's disease, which share several genes relating to mitochondrial function and encode respiratory complex subunits. Previously, mitochondrial abnormalities relating to loss of MeCP2 function were linked to respiratory complex III by the UQCRC1 subunit.[25] Significant relationships of this type were evident and while a clear association between CG-AT4 content and gene pathways implicated in neurodegenerative pathways is prominent, the potential interaction of MeCP2 with spliceosome components at the gene level was intriguing. Recent experimental results by Young *et al* have highlighted MeCP2 involvement with RNA, its splicing and interaction with RNA proteins.[14] The ribosomal pathway may also be linked to RTT and neurodegenerative disease through aberrant protein synthesis associated with these conditions.[24, 47] A comparison of high CG-AT4 motif content with high MeCP2 di-nucleosome binding showed that a number of pathways are shared between the differentially selected genes. Once again, ribosome, spliceosome and neurodegenerative pathways including Huntington's disease were identified (**Table. 2**). Additionally, comparison of high CG-AT4 content genes from either high or low MeCP2 binding groups revealed that the previously observed pathways are associated only with high MeCP2 enrichment (**Table. 3**). These results illustrate that CG-AT4 sequences in context to chromatin have the potential to regulate gene expression by targeting MeCP2 to the genome.

MeCP2 co-factor interaction highlights the spliceosome

Our experiments have examined MeCP2 functionality in terms of genome-wide and sequence-specific binding. The enrichment of MeCP2 at CG-AT4 sequences and gene pathways associated with ribosome and spliceosome as well as predominant binding of MeCP2 with di-nucleosomes suggest interactions with RNA networks. To examine further the relationship between MeCP2 and its involvement with RNA, soluble nuclear extracts from whole mouse brains were prepared, followed by endogenous MeCP2 co-immuno precipitation (Co-IP) with and without RNase treatment. The enriched protein samples were then processed for protein identification by mass spectrometry (MS). Fewer proteins were detected within the RNase treated sample (**Figure 5A**) and in agreement with the RNA dependency of large MeCP2 complexes.[14] Analysis of both the RNase treated and untreated samples identified proteins having DNA and RNA dependent functions. Consistent with the above results, MS protein fingerprinting also showed a predominant association of spliceosomal components, reiterating the association of MeCP2 with RNA (**Table. 2**). Western analysis of MeCP2 Co-IP material confirmed the interaction with the spliceosome component PRP8, the RNA dependent interactions of TOP2b and DHX9, as indicated by MS, as well as interaction with BRM and BRG1 (**Figure 5B**). These results clearly highlight a RNA dependent class of MeCP2 interactions.

Functional correlation suggests greater roles between MeCP2 and RNA

The identification of MeCP2 protein interacting with RNA-based proteins was surprising given that previous studies focused on regulatory complexes associated with gene transcription.[48] Given the demonstration that MeCP2 interacts with RNA processing proteins, we screened for MeCP2 associated RNA using a modified RNA-ChIP-Seq protocol (RIP-Seq). MeCP2-immunopurified complexes from fixed, sonicated wild type mouse cerebellum were DNase treated, followed by massively parallel sequencing of enriched RNA transcripts. RIP experiments were also performed using MeCP2-null cerebellum but did not yield detectable levels of enriched RNA for use in library generation. The sequencing data indicates that MeCP2 associates with many RNAs, with the strongest signals originating from those genes with greatest expression. Similar functional groups were found in pathway analysis of high RIP-Seq signal transcripts as observed in analyses of both high di-nucleosome ChIP binding and high CG-AT4 TSS content (**Table. 4**). However, pathway similarity was lost if the analysis of RIP-Seq was performed relative to expression level (**Table. 4**). Since RNA binding domains exist outside of the methyl-CpG binding domain[12] and given the high expression in the brain, binding might be proportional to mRNA levels. While specific interactions with RNA have not been associated with MeCP2, preference for double stranded RNA has been

reported.[12] We therefore directed our focus toward long non-coding RNAs (lncRNAs) as they have been implicated with chromatin structure.[49] While a number of lncRNAs are likely to interact with MeCP2, we selected the high RIP-Seq signal transcripts RNCR3 and MALAT1 for validation (**Figure 6A-B**). Using qRT-PCR relative to GAPDH and input, we observed higher relative enrichment of RNCR3 compared to MALAT1, with greatest enrichment over RNCR3 exons 2 and 3 (**Figure 6C**). While there is a clear relationship between MeCP2 associated with the exons of RNCR3, the relationship between MeCP2 and RNCR3-124a, which encodes the mir124a, was less evident indicating that the interaction appears to be predominantly exonic with sequence specificity. In contrast this interaction is significantly reduced for MALAT1 in the cerebellum, which suggests the higher RIP-Seq signal for Malat1 may be attributed to greater transcript copy number. Next, we considered the possibility that lncRNA binding to MeCP2 could differentially regulate RNCR3 and MALAT1 gene expression. Comparative expression between wild type and MeCP2 null male mouse cerebellum revealed greater expression of RNCR3 exons 2 and 3 compared to exon 4 in null tissue, but unremarkable change in MALAT1 (**Figure 6D**). The variation in RNCR3 exonic expression in the absence of functional MeCP2, suggests abnormal transcription or splicing events between exons two to three and the forth exon of the RNCR3 transcript. Hence, it is likely that MeCP2 is required for the normal regulation of these processes. These data are consistent with MeCP2 associated lncRNA interactions and further implicate MeCP2 with RNCR3.

Discussion

The data presented here identifies a role for MeCP2 in the regulation of gene expression that highlights physiological targets in the brain. The binding of MeCP2 to CG-AT4 is enhanced in active chromatin fractions and supporting a model of regulation that include the interactions of long non-coding RNAs. The identification of MeCP2 association with the splicesome at multiple levels was also surprising given that previous studies have demonstrated interactions with DNA binding transcription factors and core machinery that serve to alter chromatin structure and regulate gene expression.

The first glimpse into the mechanism of methylation mediated silencing by MeCP2 came in 1998 from the seminal papers from the groups of Adrian Bird and Alan Wolffe, when they reported the three-way connection between DNA methylation and gene silencing was associated by the modification of chromatin.[3, 4] The discovery that MeCP2 binds a co-repressor complex that includes mSin3A and histone deacetylases paved the way to a molecular mechanism between MeCP2 and chromatin.[50, 51] A large body of evidence followed these reports exploring the control of gene expression by MeCP2.[52, 53] Coordinated by DNA methylation and given its preference for the 5-methyl-cytosine moiety, MeCP2 was largely associated with transcriptional repression and chromatin remodeling.[9, 10] Although specific MeCP2 targets have been sought, recent studies[17, 19] and our current data illustrate more widespread genome-wide interactions. While CG-AT4 enhanced binding has been considered for some time[6], CG-AT4 frequency has not been examined for targeting of MeCP2 to the genome. In this regard, we show a link of MeCP2 activity to neurodegenerative disease, energy production, protein synthesis and RNA splicing pathways, which are molecular processes thought to be involved in RTT and autism pathophysiology.[22-25] The specific regulation of genes by MeCP2 requires additional refinement, but we find that CG-AT4 frequency and methylation status can account for some of the variability. Commonality between MeCP2 di-nucleosome enrichment and high CG-AT4 pathways suggest MeCP2 competition with histone H1 for linker DNA in brain. In support of this, histone H1 was shown to be enriched exclusively in di-nucleosome material from Alzheimer's patients.[54] This study clearly implicated H1 with neurodegenerative disease, while highlighting the importance of chromatin context. The specific affinity of MeCP2 for the chromatin template over a region may define its ability to compete with histone H1, which in turn (de)stablilises chromatin structure, affecting gene expression. Indeed, the effectiveness of MeCP2 to alter chromatin structure was recently shown in mouse models expressing mutations at amino acids 270 and 273 of the MeCP2 protein.[55] This mutation resembles the severe phenotype seen in Rett individuals, the authors of the study identifying a 250 amino acid stretch homologous to HMGA1that encodes three AT-hook like

domains that alter ATRX localization. These results suggest the AT-hook of MeCP2 is important to maintain chromatin structure.

The multifaceted DNA binding properties of MeCP2 were also recently expanded to 5-hydroxymethyl-CpG (5hmC) sites in the brain to further highlight its interaction with euchromatin could regulate genome wide changes in gene expression.[56] Although the current consensus for MeCP2 remains as a "classical" repressor complex with the major emphasis of research on its well characterized role with respect to methylation-dependent binding[57], recent studies now show additional regulatory roles that are more suggestive of a "non-classical mechanism" of regulation mediating gene activation.[11, 46] An elegant solution to the long standing relationship that MeCP2 recognizes unmethylated DNA and regulates euchromatin was recently shown in the brain.[56] The high levels of 5hmC in neuronal genomes, specifically in euchromatin of mature cerebellar cells was associated with MeCP2 binding which was disrupted by the Rett mutation, R133C. Interestingly, 5hmC enrichment at active genes were depleted for 5mC at these regions which were consistent with strong affinity of MeCP2 at these sites. These findings now opens the field of MeCP2 research and suggests that the gene targets may not be the usual suspects, i.e. dependent on DNA methylation . Clearly, these recent experimental findings demonstrate how MeCP2 recognizes the genome and is targeted to chromatin is as important as the gene under regulation.[58, 59] Indeed, the identification of MeCP2 enrichment at 5hmC sites and its relationship with active euchromatin suggests the reevaluation of its binding properties. The binding of MeCP2 on 5mC and 5hmC with similar affinities could explain gene activation as well as transcriptional suppression in animal models that either lack or overexpress MeCP2.[11, 60]

Consistent with recent published data, we observed broad genomic enrichment[17], but also identified unique patterns of MeCP2 binding between different genomic fractions. While a clear relationship between differential gene expression and MeCP2 binding is evident at gene promoters and between high and low CG-AT4 content this relationship becomes less readily apparent when considering MeCP2 enrichment of di-nucleosomes. Although neurodegenerative pathways are immediately intriguing for MeCP2 function, pathways of spliceosome and ribosome were most prominent. This would suggest a model of regulation mediated by MeCP2 in the regulation of RNA networks and could account for MeCP2 interaction with the RNA binding protein YB-1.[14] Our mass spectrometry results also implicate the participation of MeCP2 with RNA regulation and the spliceosome. Alternative splicing of RNA transcripts in the brain is critical to regulating gene expression patterns and neuronal plasticity.[61] We identified several proteins by mass spectrometry associated with RNA remodelling and the spliceosome such as DHX-9 and PRP-8, respectively. The ATP-dependent RNA helicase, DHX9 functions in the unwinding of double-stranded RNA and DNA-RNA

complexes.[62] The DHX-9 gene belongs to a family of DEAH-box RNA-helicases characterized by the Asp-Glu-Ala-His motif and is implicated in RNA metabolism and transcriptional activation.[63] We found that DHX-9 is an RNA-dependent MeCP2-interacting protein by mass spectrometry and this was validated by protein co-immunoprecipitation in nuclear extracts. Given the abundance of MeCP2 in the brain, its enrichment with heterochromatin and involvement in maintaining chromatin structure, we hypothesize the interaction with DHX-9 could also regulate RNA and DNA-RNA chromatin complexes. Although the mechanism by which MeCP2 cooperates with RNA helicases remains to be determined, this seems probable based on mass spec-protein identification. We also found MeCP2 interacting with members of the DEAD-box family characterized by the conserved motif Asp-Glu-Ala-Asp such as DDX-proteins.[64] For example, the ATP-dependent RNA helicase, DDX5 is involved in alternative regulation of mRNA splicing mediated by DNA- and chromatin-binding factors.[65] A central component of the spliceosome, the PRP-8 protein is a component of U2- and U12-dependent spliceosomes and essential in pre-mRNA splicing events.[66] PRP-proteins also interact with SWI/SNF enzymes to coordinate pre-mRNA splicing with chromatin remodelling events involved in transcriptional regulation.[67] The relevance of these findings to MeCP2 suggests a possible role in the regulation of chromatin structure associated by interacting with components implicated with RNA splicing.

Alternative splicing is observed in higher eukaryotes to diversify transcriptomes and the serine/arginine (SR) splicing factors are engaged in regulating neuronal gene expression. Long non-coding RNAs such as MALAT1 interact with pre-mRNA splicing factors to regulate alternative splicing.[68, 69] In this study, we show interacting RNAs by screening for MeCP2 using a modified RIP-Seq technique. MeCP2-immunopurified complexes identified RNA transcripts enriched for several lncRNAs, including MALAT1 and RNCR3. Although a direct mechanism by which lncRNAs could regulate MeCP2 binding remains to be determined, we speculate that ncRNAs such as RNCR3 guide MeCP2 to chromatin to regulate gene structure and function, and that dsRNAs may also act to buffer MeCP2 chromatin binding through displacement. To our knowledge, this is perhaps the first demonstration of context preference as well as the association of MeCP2 with lncRNA that may mediate targeting of regulatory functions. Recently, an RNCR3-/- mouse model was created to study MIR-124A sourced from the RNCR3 lncRNA.[70] Intriguingly, RNCR3-/- mice display hind limb clasping, reduced brain weight and other characteristics that are reminiscent of those seen in MeCP2 null mice, which model the clinical features of RTT.[70] Furthermore, both MeCP2 and RNCR3 are required in normal GABAergic neuron function and development.[71, 72] Given the high capacity of ncRNAs to form secondary structures and the ability of double stranded RNA to displace MeCP2 from bound methylated DNA[12], this previously unknown connection to

RNCR3 offers an alternative mechanism to the long held view of CpG methylation-mediated repression by MeCP2. In addition to its participation in RNA splicing, the molecular interaction with ncRNAs extends the mechanism of gene regulation by targeting MeCP2 to chromatin, beyond the context of CpG methylation and CG-AT4. The association of ncRNA provides an new entry point in understanding the MeCP2 regulatory mechanisms as well as providing some insight of how ncRNA assembly may guide MeCP2 to chromatin. The identification of MeCP2 gene targets could be achieved by examining contextual relationships of chromatin and their role in the function of lncRNAs and other RNA molecules.

Based on the findings presented here the experimental data allow us to reassess MeCP2 mediated gene regulation. We hypothesize MeCP2 interacting ncRNAs may mediate chromatin remodeling changes and gene regulatory events in the brain. Although we do not fully appreciate the relative importance of the spliceosome proteins and the MeCP2 interacting long non-coding RNAs, the results presented allow reinterpretation of long-standing questions relevant to our understanding of gene regulation as well as providing insight for MeCP2 gene targets. Not only is this of direct therapeutic relevance to Rett syndrome, but will also be important for our understanding of new mechanisms of gene regulation conferred by MeCP2 interacting ncRNAs that influence chromatin structure and function.

Methods and Materials

Animals

Brain tissue for sequencing experiments was obtained from wild type (mixed 129/C57BL6 background) or symptomatic MeCP2 Null (*Mecp2*^{tm1Tam}) male mice.[73] Mass spectrometry analysis was performed using brains collected from swiss-webster mice (Pel Freeze Biologicals).

Antibodies

MeCP2 (9317, Sigma), Prpf8 (ab51366, Abcam), DHX9 (NB110-40579, Novus Bio), Histone H1 (05-457, Millipore), TOP2b (NB100-40842, Novus Bio), BRM (ab15597, Abcam), BRG-1 (Dr. Said Sif).

Soluble Nuclear Protein Extraction

All buffers and steps were kept on ice unless otherwise stated. All buffers contained 1x complete protease inhibitor cocktail tablets (Roche Diagnostics). Nuclear extracts were prepared from whole mouse brains (Pel-Freez Biologicals). Five brains were dounce homogenized in PBS and pelleted (2 minutes, 1500 g). The homogenised tissue was resuspended in 10 mL low-salt buffer (10 mM HEPES, 15 mM KCl, 0.2 mM EDTA, 0.1 mM EGTA, 0.5 mM DDT, 0.5 mM spermidine, 0.15 mM spermine, 0.075% NP40) for 30 min at 4°C. Nuclei were isolated by mixing 1 part sucrose buffer (50 mM HEPES, 10 mM KCl, 0.2 mM EDTA, 0.1 mM EGTA, 0.5 mM DDT, 0.5 mM spermidine, 0.15 mM spermine, 75% sucrose) with 2 parts low-salt homogenate and centrifugation (30 sec, 8000 g). The isolated nuclei pellet was resuspended with 2x the pellet volume of high-salt buffer (50 mM HEPES, 450 mM KCl, 0.2 mM EDTA, 2 mM EGTA, 0.5 mM DDT, 1 mM MgCl, 25% gylcerol), and incubated 5 minutes at 4˚C. Soluble nuclear protein extract (supernatant) was recovered by centrifugation for 15 min at 18,000 g, then snap frozen and stored at -80°C.

Co – Immunoprecipitation

For each immunoprecipitation, RNase treated (100µg RNase A (Sigma) for 30 minutes at 4˚C unless otherwise stated) or non RNase treated soluble nuclear extract was used. A 150 µL nuclear extract, adjusted to contain 1% BSA, was precleared twice (2 x 45 min), using 20 µL packed volume protein A Sepharose beads (GE Healthcare) each time. Precleared extracts were incubated for 1 hour with MeCP2 antibody (Sigma), before adding 20 µl packed volume of protein A Sepharose beads blocked with 1% BSA and a further 2 h incubation at 4°C. Bead complexes were washed 5 times with washing solution (1 x PBS, 1% Triton X-100), before boiling in 40 µl elution buffer (0.1 M glycine pH 2.5, 50 mM DTT). Precipitated proteins were size fractionated by SDS PAGE, and alternately processed for mass spectrometry or transferred to PVDF membrane for examination by western blotting using standard techniques.

Mass Spectrometry

SDS PAGE gels of MeCP2 co-IP proteins were stained with coomassie blue R-250 (Amresco). Protein bands were excised and initially cleaned by sonication in 10 mM Tris for 5 minutes. Standard techniques for in-gel protein digestion followed, involving sequential dehydration (50% acetonitrile, 25 mM ammonium bicarbonate) and rehydration (100mM ammonium bicarbonate) treatments to further clean and remove coomassie staining from gel pieces. Cysteine residues were reduced (10mM DTT, 100 mM ammonium bicarbonate, 60˚C for 1 hour) and carbamidomethylated using iodoacetamide (55 mM, 20 minutes). Treated gel pieces were further washed (100 mM ammonium bicarbonate) and finally with water before drying by speed vac (Thermo Scientific). Dried gel pieces were hydrated with digestion solution (0.05 µg / mL trypsin (Promega) in 25 mM ammonium bicarbonate) on ice for 30 minutes, excess digest solution was removed by washing with and final suspension in 10 µL digest buffer (25 mM ammonium bicarbonate). Trypsin digest of proteins continued for 3 hours at 37˚C, and left at 4˚C overnight. A 0.6 µL aliquot of MALDI matrix (α-cyano-4-hydroxy cinnamic acid, 3 mg/ml in acetone/ethanol (1:2, v/v)) was spotted to an anchorchip target plate (Bruker Daltonics) with 0.5 µL acidified protein digest (0.1 % TFA), and examined by MALDI-TOF using an Autoflex II mass spectrometer, Flex and biotools analysis software (Bruker Daltonics). The mass range (900-3,000 mhz) of the mass spectrometer was internally calibrated using autolytic peaks of trypsin. Peptide mass lists of digests were compared to the protein database SwissProt using the mascot search algorithm (Matrix Science, http://www.matrixscience.com). Typical search parameters were as follows: mass tolerance, 0.25 Da; missed cleavages, 2; enzyme, trypsin; fixed modifications, carbamidomethylation; variable

modification, Oxidation (M); taxonomy, Mus musculus. Protein ID matches with probability based Mowse scores of at least 82 were recorded ($p<0.0001$).

Chromatin -Sonication

The frozen brain of one mouse was homogenized in 5 mL of 1% formaldehyde PBS solution using a 10 mL dounce (Wheaton). Cross-linking of the homogenate solution continued for a total of 10 minutes at room temperature (RT) with rotation from the start of homogenization (Note: within the last 2 minutes the homogenate was centrifuged (2 minutes, 1500 g, RT)). Formaldehyde cross-linking was halted at exactly 10 minutes by addition of 5 mL 125 mM glycine PBS solution to the pellet and a further 10 minute incubation at RT with rotation. The quenched solution (200 µL) was aliquoted into 5 tubes (1.5 mL, Eppendorf), pelleted and re-suspended in 250 µL lysis buffer (1% SDS, 10mM EDTA, and 50mM Tris-HCl pH8.0). Lysed material was sheared to between 100 and 500 bp using a waterbath sonicator (Diagenode). For each aliquot 200 µL of cleared sonicate was collected and added to 1.8 mL ChIP dilution buffer (0.01 % SDS, 1.1 % Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH8.0, and 167 mM NaCl) with 40 µg RNase A (Sigma) and incubated for 30 minutes at 4˚C with rotation and finally pooled.

Chromatin -MNase Fractionation

All buffers and steps were kept on ice unless otherwise stated. The brain or cerebellum of one mouse was isolated and homogenized in 5 mL of 1% formaldehyde PBS solution using a 10 mL dounce (Wheaton). Cross-linking of the homogenate solution continued for a total of 10 minutes at room temperature (RT) with rotation from the start of homogenization (Note: within the last 2 minutes the homogenate was centrifuged (2 minutes, 1500 g, RT)). Formaldehyde cross-linking was halted at exactly 10 minutes by addition of 5 mL 125 mM glycine PBS solution to the pellet and a further 10 minute incubation at RT with rotation. The formaldehyde quenched pellet was isolated (2 minutes, 1500 g, RT), and 5 mL of nuclear isolation buffer (15 mM Tris-HCL pH 7.5, 60 mM KCl, 0.5 M sucrose, 0.25 mM EDTA, 0.125 mM EGTA, 1 mM DTT, 0.5 mM PMSF, 0.15 mM spermidine, 0.15 mM spermine, 0.2 % NP-40, and 1X complete proteinase inhibitor (Roche)) was added and incubated for 5 minutes at 4˚C with rotation. The solution was equally divided between four 1.5 mL tubes (Eppendorf) and the nuclear pellet isolated by centrifuge (1 min, 5000 g, 4˚C). Each pellet was washed briefly with 200 µL MNase digestion buffer (20 mM TRIS-HCl, 70mM NaCl, 20 mM KCl, 5mM MgCl, 3 mM CaCl, 0.5 mM PMSF, and 1X complete proteinase inhibitor (Roche)) and centrifuged (1min, 4000 g, 4˚C). Each washed pellet was re-suspended in 400 µL of MNase digestion buffer and rested on ice, before adding 250 units micrococcal nuclease (Worthington) and allowing digestion for 8 minutes. Digestion was stopped by addition of 16 µL

MNase stop solution (125 mM EDTA, 125mM EGTA). The digest was centrifuged (5 minutes, 5000 g, 4˚C) and the fixed s1 chromatin fraction (supernatant) from each tube collected and pooled. RNA was digested from collected S1 samples by RNase treatment at the rate of 100 µg RNase A (Sigma) per mL of sample for 30 minutes at 4˚C with rotation.

Chromatin Immunoprecipitation

ChIP was performed on both sonicated and MNase fractionated chromatin. All buffers and steps are kept on ice unless otherwise stated. A 40 µL aliquot was retained as Input material. For MNase fractionated chromatin 100 µL sample was aliquoted into each of 16 tubes (1.5 mL, Eppendorf), along with 20 µL Protein A Dynabeads (Invitrogen) and 400 µL ChIP dilution buffer (0.01 % SDS, 1.1 % Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH8.0 and 167 mM NaCl). For sonicated material 500 µL diluted sample was added to each of 16 tubes with 20 µL Protein A Dynabeads (Invitrogen). For each experiment, an additional set of 16 tubes were set up in parallel containing 500 µL ChIP dilution buffer and 20 µL Protein A Dynabeads, with 2.5 µg of antibody added to 8 tubes only. All tubes were placed at 4˚C with rotation for 1 hour allowing pre-clearing of sample and pre-binding of antibody to beads. The pre-cleared chromatin was separated from dynabeads using a magnetic rack (Invitrogen) and transferred to the parallel set of either pre-bound or no-antibody bead containing tubes of which solution was discarded. Antibody and no-antibody control tubes were then left over night to at 4˚C with rotation in order to precipitate chromatin. The following day chromatin bound beads were washed twice consecutively with 0.5 mL each time of the following buffers in order; ChIP dilution buffer, Low Salt buffer (0.1 % SDS, 1% Triton X-100, 2 mM EDTA, 20mM Tris-HCl pH8.0 and 150 mM NaCl), High Salt buffer (0.1 % SDS, 1% Triton X-100, 2 mM EDTA, 20mM Tris-HCl pH8.0 and 500 mM NaCl), LiCl buffer (0.25 M LiCl, 1 % Tergitol, 1 % deoxycholic acid, 1 mM EDTA and 10 mM Tris-HCl pH 8.0) and TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). After washes, beads were transferred to a new set of tubes using 0.5 mL of TE buffer with 0.01 % SDS and the solution discarded. To each tube of antibody or no-antibody control beads, 100 µL of RT ChIP elution buffer (20 mM Tris-HCl pH 7.5, 5 mM EDTA, 50 mM NaCL and 1% SDS) and 40 µg protinase K (P4850, Sigma) was added. A360 µL aliquot of ChIP elution buffer and 120 µg protinase K was added to the retained input material of 40 µL. The elutions of ChIP material and Input were incubated at 62˚C in a thermomixer (Eppendorf) at 1400 rpm for 2 hours to reverse cross-links. Beads were then removed and ChIP samples pooled into Antibody bound, No-Antibody control and Input. An equal volume of Phenol: Chloroform: Isoamyl Alcohol was added to each sample, then vortexed and rotated at RT for 10 minutes before centrifugation (15 minutes, 18000 g, RT). DNA was purified from the aqueous phase using the

Nucleospin extract II kit (Macherey-Nagel) as per manufacturer instructions. Purified DNA was used to make genome sequencing libraries or in RT-PCR.

RNA Immunoprecipitation

The frozen cerebellum of one mouse was homogenized in 5 mL of 1% formaldehyde PBS solution using a 10 mL dounce (Wheaton). Cross-linking of the homogenate solution continued for a total of 10 minutes at room temperature (RT) with rotation from the start of homogenization (Note: within the last 2 minutes the homogenate was centrifuged (2 minutes, 1500 g, RT)). Formaldehyde cross-linking was halted at exactly 10 minutes by addition of 5 mL 125 mM glycine PBS solution to the pellet and a further 10 minute incubation at RT with rotation. The quenched solution (200 µL) was aliquoted into 5 tubes (1.5 mL, Eppendorf), pelleted and re-suspended in 300 µL RNA lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl pH8.0, 0.5 mM PMSF, 60 units RNase inhibitor (Applied Biosystems) and 1X complete proteinase inhibitor (Roche)). Lysed material was sheared to between 100 and 500 bp using a waterbath sonicator (Diagenode). For each aliquot 250 µL of cleared sonicate was collected and added to 1.8 mL ChIP dilution buffer (0.01 % SDS, 1.1 % Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH8.0, and 167 mM NaCl), 200 µL was retained as Input material. The diluted material (500 µL) was added to each of 4 tubes with 20 µL Protein A Dynabeads (Invitrogen). For each experiment, an additional set of 4 tubes were set up in parallel containing 500 µL ChIP dilution buffer and 20 µL Protein A Dynabeads, with 2.5 µg of MeCP2 antibody added to 2 tubes only. All tubes were placed at 4˚C with rotation for 1 hour allowing pre-clearing of sample and pre-binding of antibody to beads. The pre-cleared chromatin was separated from dynabeads using a magnetic rack (Invitrogen) and transferred to the parallel set of either pre-bound or no-antibody bead containing tubes of which solution was discarded. Antibody and no-antibody control tubes were then incubated 2 hours at 4˚C with rotation to precipitate RNA complexes. RNA complex bound beads were washed twice consecutively with 0.5 mL each time of the following buffers in order; ChIP dilution buffer, Low Salt buffer (0.1 % SDS, 1% Triton X-100, 2 mM EDTA, 20mM Tris-HCl pH8.0 and 150 mM NaCl), High Salt buffer (0.1 % SDS, 1% Triton X-100, 2 mM EDTA, 20mM Tris-HCl pH8.0 and 500 mM NaCl), LiCl buffer (0.25 M LiCl, 1 % Tergitol, 1 % deoxycholic acid, 1 mM EDTA and 10 mM Tris-HCl pH 8.0) and TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). After washes, beads were transferred to a new set of tubes using 0.5 mL of TE buffer with 0.01 % SDS, and the solution discarded. To each tube of antibody or no-antibody control beads, 100 µL of RIP elution buffer (20 mM Tris-HCl pH 7.5, 5 mM EDTA, 50 mM NaCL, 1% SDS, 20 units RNase inhibitor and 40 µg protinase K (P4850, Sigma) was added. Forty units RNase inhibitor and 80 µg protinase K was added to the 200 µL of diluted input material. The elution samples and input were incubated at 62˚C in a thermomixer (Eppendorf) at

1400 rpm for 2 hours to reverse cross-links. Beads were then removed and ChIP samples pooled into Antibody bound, No-Antibody control and Input. An equal volume of Chloroform was added to each sample, then vortexed and rotated at RT for 10 minutes before centrifugation (15 minutes, 18000 g, RT). RNA was purified from the aqueous phase using the RNAeasy Minelute kit with DNase pre-digestion (Qiagen) as per manufacturer instructions. RNA was eluted in 30 µL nuclease free water, and used to make RNA sequencing libraries or in RT-PCR analysis.

Methylated DNA Enrichment

Male mouse cerebellum DNA (100 µg) was sonicated to an average length of approximately 300bp (Diagenode Bioruptor). Methylated DNA was then enriched using a Methylminer kit (Invitrogen) according to the manufactures protocol.[74] Briefly, the kit employs a biotin labeled methyl binding domain from the human MBD2 protein to capture methylated DNA. When the captured DNA is eluted over an increasing stepwise salt gradient, DNA with fewer to higher methyl groups are separated. Following capture of sample DNA, salt washes were performed at 160mM before collection of the 350 mM NaCl and 2000 mM NaCl elutions sequentially. Completely methylated and unmethylated control duplexes are included with the Methylminer kit and were used in parallel with each sample processed to validate the effectiveness and functionally of the enrichment. Both duplexes are 80bp in length and contain either 8 methylated or 9 unmethylated CpGs. Sequencing libraries were constructed using the sequential 350 mM and 2000 mM sample elutions.

Sequencing library preparation

ChIP and Methylminer enriched sequencing libraries were prepared as per the ChIP-Seq library preparation kit protocol using single end adaptors (Illumina). The following changes were made to the protocol. “Variation to starting material”- sequencing libraries were constructed using 15 ng of sample DNA from Methylminer, ChIP bound and ChIP input material. “Variation to 2% gel size selection of adaptor ligated material” - 200 bp to 300 bp selection range was excised and collected for Methylminer material. For ChIP material two size selections were taken, one selection range of 200 bp to 300 bp representing a largely mono-nucleosome fraction and another selection of 300 bp to 400 bp to represent material up to di-nucleosome in length. Gel size selection was based on the 1 Kb plus ladder (Invitrogen). PCR amplified libraries were diluted to 10 nM and stored at -20˚C in readiness for cluster generation and sequencing.

RNA-Seq

Total RNA was extracted from wild type and MeCP2-null male mouse cerebellum using TRIZOL reagent (Invitrogen) and purified by column combined DNase digestion (Qiagen). Purified RNA

was quantified by a qubit fluorometer (Invitrogen) and 5 µg per sample used in construction of sequencing libraries according to the mRNA-sequencing kit protocol (Illumina). PCR amplified libraries were diluted to 10 nM and stored at -20˚C in readiness for cluster generation and sequencing.

RIP-Seq (RNA-ChIP-Seq)

Purified RIP RNA (18 µL) was used in construction of sequencing libraries according to the NEBnext mRNA sample prep master mix set 1 protocol (New England Biolabs). PCR amplified libraries were diluted to 10 nM and stored at -20˚C in readiness for cluster generation and sequencing.

Massively parallel sequencing

Sequencing libraries were initially seeded to a flow cell and cluster amplified using the cluster kit version 4 protocol and cluster station system (Illumina). Massively parallel sequencing was performed using a Genome Analyser IIx (Illumina), using the 36-cycle version 4 sequencing kit protocol (Illumina). Sequence and quality data were extracted from cluster cycle imaging using Pipeline v1.6 software (Illumina) and aligned to the mouse genome (July 2007 NCBI37/mm9 assembly) using the Burrows-Wheeler Aligner.[75] Sequencing data has been submitted to NCBI Gene Expression Omnibus (GEO) under the accession number GSE38324.

GEO data is currently not public and may be viewed using the reviewer link; http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=hnwzxwkaukysaly&acc=GSE38324

Bioinformatics Tools

Genomic feature trends were analyzed with SeqMonk software package. DNA sequence of genomic regions and extended sequencing reads was extracted using Galaxy bioinformatics server tools. Sequence was quantified for CpG content or the regular expression describing CG-AT4 as CG(n ≤ 15)(A/T ≥ 4) where ‘n’ represents number of random nucleotides, unless otherwise stated. Functional annotation of gene groups was carried out using DAVID bioinformatics resources.[76, 77]

RT-PCR

RNA was reverse transcribed using a high capacity cDNA reverse transcriptase Kit (Applied Biosystems). PCR amplification was performed using a 7500 Fast Real-Time PCR System (Applied Biosystems). Five pmoles of forward and reverse primer, 2 µL cDNA template and FAST SYBR® Green Master Mix (Roche) were mixed to a final volume of 13 µl with nuclease free water. RT-PCR cycle parameters were as follows- incubation at 95°C for 10 min, followed by 50 cycles of - 95°C for 3 s and 60°C for 30 s. RIP enrichments were adjusted by input, and expressed relatively to GAPDH, while target gene expression levels were calculated as fold change normalized to GAPDH. Primers were as follows; GAPDH Forward 5'-TGA AGC AGG CAT CTG AGG G -3', GAPDH Reverse 5'- CGA AGG TGG AAG AGT GGG AG -3', *RNCR3* e2-3 Forward 5'-GGG GTT CCG GGA CAA AGC GT -3', *RNCR3* e2-3 Reverse 5'- CTC GCT TAG CTC AGA GGT CCG CA -3', *RNCR3* e2-I Forward 5'- ACG CTG CCC AAC AGT ACC CG -3', *RNCR3* e2-I Reverse 5'- CAC CCC ACA GGG TTG GGG GA -3', *RNCR3* 124a Forward 5'-AAG GCC TCT CTC TCC GTG TT -3', *RNCR3* 124a Reverse 5'- CTC AGA CAG CCC CAT TCT TG -3', *RNCR3* e4 Forward 5'- CCT CCT TGG GGC TGT GGC TG -3', *RNCR3* e4 Reverse 5'-GAG CGA TGG GGA GAG GCC CA-3', *MALAT1* mas2 Forward 5'- TGA GGC TGG GGA GTG TTC CAG T -3', *MALAT1* mas2 Reverse 5'- GAC GGG GTT CAA GTC CCT GCG -3', *MALAT1* e1 Forward 5'- GCC ACA CAG GAA GGC TCC GC -3', *MALAT1* e1 Reverse m 5'- GAG GGG TGA GGT GGG CGC TA -3', *MALAT1* mas1 Forward 5'- GGC TGG GGA GTG TTC CAG TGA -3', *MALAT1* mas1 Reverse 5'- TGG TGG CTG GCA CTC CTG GT -3'.

Sources of Funding:

The authors acknowledge grant and fellowship support from the National Health and Medical Research Council (NHMRC). SM was supported by a Postgraduate Research Scholarship and GP by a Biomedical (Peter Doherty) Fellowship. PPLT is a Senior Principal Research Fellow and AE-O is a Senior Research Fellow of the NHMRC. Our work is supported in part by the Victorian Government's Operational Infrastructure Support Program and Mr. James Fairfax.

Acknowledgements:

We sincerely thank Dr. Said Sif for provision of Brg1 antibody and give special thanks to Dr. Harikrishnan Kaipananickal, Antony Kaspi and Dr. Mark Ziemann for their assistance with next generation sequencing and bioinformatic analysis, and Professor John Christodoulou for assistance with mouse models.

Conflict of Interest:

None declared

Disclosures:

None

References:

1. Nan X, Campoy FJ, Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. Cell 1997; 88:471-81.

2. Meehan RR, Lewis JD, Bird AP. Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res 1992; 20:5085-92.

3. Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet 1998; 19:187-91.

4. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 1998; 393:386-9.

5. Iguchi-Ariga SM, Schaffner W. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. Genes Dev 1989; 3:612-9.

6. Klose RJ, Sarraf SA, Schmiedeberg L, McDermott SM, Stancheva I, Bird AP. DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. Mol Cell 2005; 19:667-78.

7. Adams VH, McBryant SJ, Wade PA, Woodcock CL, Hansen JC. Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. J Biol Chem 2007; 282:15057-64.

8. Ghosh RP, Nikitina T, Horowitz-Scherer RA, Gierasch LM, Uversky VN, Hite K, Hansen JC, Woodcock CL. Unique physical properties and interactions of the domains of methylated DNA binding protein 2. Biochemistry 2010; 49:4395-410.

9. Harikrishnan KN, Chow MZ, Baker EK, Pal S, Bassal S, Brasacchio D, Wang L, Craig JM, Jones PL, Sif S, et al. Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. Nat Genet 2005; 37:254-64.

10. Nan X, Hou J, Maclean A, Nasir J, Lafuente MJ, Shu X, Kriaucionis S, Bird A. Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proc Natl Acad Sci U S A 2007; 104:2709-14.

11. Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, Zoghbi HY. MeCP2, a key contributor to neurological disease, activates and represses transcription. Science 2008; 320:1224-9.

12. Jeffery L, Nakielny S. Components of the DNA methylation system of chromatin control are RNA-binding proteins. J Biol Chem 2004; 279:49479-87.

13. Buschdorf JP, Stratling WH. A WW domain binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. J Mol Med 2004; 82:135-43.

14. Young JI, Hong EP, Castle JC, Crespo-Barreto J, Bowman AB, Rose MF, Kang D, Richman R, Johnson JM, Berget S, et al. Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. Proceedings of the National Academy of Sciences of the United States of America 2005; 102:17551-8.

15. Chahrour M, Zoghbi HY. The story of Rett syndrome: from clinic to neurobiology. Neuron 2007; 56:422-37.

16. Shahbazian MD, Antalffy B, Armstrong DL, Zoghbi HY. Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. Hum Mol Genet 2002; 11:115-24.

17. Skene PJ, Illingworth RS, Webb S, Kerr AR, James KD, Turner DJ, Andrews R, Bird AP. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol Cell 2010; 37:457-68.

18. Ghosh RP, Horowitz-Scherer RA, Nikitina T, Shlyakhtenko LS, Woodcock CL. MeCP2 binds cooperatively to its substrate and competes with histone H1 for chromatin binding sites. Mol Cell Biol 2010; 30:4656-70.

19. Cohen S, Gabel HW, Hemberg M, Hutchinson AN, Sadacca LA, Ebert DH, Harmin DA, Greenberg RS, Verdine VK, Zhou Z, et al. Genome-Wide Activity-Dependent MeCP2 Phosphorylation Regulates Nervous System Development and Function. Neuron 2011; 72:72-85.

20. Fraga MF, Ballestar E, Montoya G, Taysavang P, Wade PA, Esteller M. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. Nucleic Acids Res 2003; 31:1765-74.

21. Bienvenu T, Chelly J. Molecular genetics of Rett syndrome: when DNA methylation goes unrecognized. Nat Rev Genet 2006; 7:415-26.

22. Lintas C, Sacco R, Persico AM. Genome-wide expression studies in Autism spectrum disorder, Rett syndrome, and Down syndrome. Neurobiol Dis 2010.

23. Leoncini S, De Felice C, Signorini C, Pecorelli A, Durand T, Valacchi G, Ciccoli L, Hayek J. Oxidative stress in Rett syndrome: Natural history, genotype, and variants. Redox Rep 2011; 16:145-53.

24. Kelleher RJ, 3rd, Bear MF. The autistic neuron: troubled translation? Cell 2008; 135:401-6.

25. Kriaucionis S, Paterson A, Curtis J, Guy J, Macleod N, Bird A. Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. Mol Cell Biol 2006; 26:5033-42.

26. Matzke MA, Birchler JA. RNAi-mediated pathways in the nucleus. Nat Rev Genet 2005; 6:24-35.

27. Zaratiegui M, Irvine DV, Martienssen RA. Noncoding RNAs and gene silencing. Cell 2007; 128:763-76.

28. Mattick JS, Amaral PP, Dinger ME, Mercer TR, Mehler MF. RNA regulation of epigenetic processes. Bioessays 2009; 31:51-9.

29. Kim DH, Villeneuve LM, Morris KV, Rossi JJ. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nat Struct Mol Biol 2006; 13:793-7.

30. Han J, Kim D, Morris KV. Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. Proc Natl Acad Sci U S A 2007; 104:12422-7.

31. Ting AH, Schuebel KE, Herman JG, Baylin SB. Short double-stranded RNA induces transcriptional gene silencing in human cancer cells in the absence of DNA methylation. Nat Genet 2005; 37:906-10.

32. Jeon Y, Sarma K, Lee JT. New and Xisting regulatory mechanisms of X chromosome inactivation. Curr Opin Genet Dev 2012; 22:62-71.

33. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. Proc Natl Acad Sci U S A 2008; 105:716-21.

34. Li LC, Okino ST, Zhao H, Pookot D, Place RF, Urakami S, Enokida H, Dahiya R. Small dsRNAs induce transcriptional activation in human cells. Proc Natl Acad Sci U S A 2006; 103:17337-42.

35. Politz JC, Zhang F, Pederson T. MicroRNA-206 colocalizes with ribosome-rich regions in both the nucleolus and cytoplasm of rat myogenic cells. Proc Natl Acad Sci U S A 2006; 103:18957-62.

36. Hwang HW, Wentzel EA, Mendell JT. A hexanucleotide element directs microRNA nuclear import. Science 2007; 315:97-100.

37. Place RF, Li LC, Pookot D, Noonan EJ, Dahiya R. MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc Natl Acad Sci U S A 2008; 105:1608-13.

38. Hamid QA, Thanumalayan S, Parnaik VK. An improved method to distinguish micrococcal nuclease sensitivity of chromatin. J Biochem Biophys Methods 1996; 33:59-64.

39. Huang SY, Garrard WT. Electrophoretic analyses of nucleosomes and other protein-DNA complexes. Methods Enzymol 1989; 170:116-42.

40. Singleton MK, Gonzales ML, Leung KN, Yasui DH, Schroeder DI, Dunaway K, LaSalle JM. MeCP2 is required for global heterochromatic and nucleolar changes during activity-dependent neuronal maturation. Neurobiology of disease 2011; 43:190-200.

41. Nan X, Tate P, Li E, Bird A. DNA methylation specifies chromosomal localization of MeCP2. Mol Cell Biol 1996; 16:414-21.

42. Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 1992; 69:905-14.

43. Choi JK. Contrasting chromatin organization of CpG islands and exons in the human genome. Genome Biol 2010; 11:R70.

44. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. Genome Res 2009; 19:1732-41.

45. Miranda TB, Jones PA. DNA methylation: the nuts and bolts of repression. J Cell Physiol 2007; 213:384-90.

46. Yasui DH, Peddada S, Bieda MC, Vallero RO, Hogart A, Nagarajan RP, Thatcher KN, Farnham PJ, Lasalle JM. Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A 2007; 104:19416-21.

47. Ricciardi S, Boggio EM, Grosso S, Lonetti G, Forlani G, Stefanelli G, Calcagno E, Morello N, Landsberger N, Biffo S, et al. Reduced AKT/mTOR signaling and protein synthesis dysregulation in a Rett syndrome animal model. Hum Mol Genet 2011; 20:1182-96.

48. Guy J, Cheval H, Selfridge J, Bird A. The role of MeCP2 in the brain. Annual review of cell and developmental biology 2011; 27:631-52.

49. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet 2009; 10:155-9.

50. Kaludov NK, Wolffe AP. MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. Nucleic Acids Res 2000; 28:1921-8.

51. El-Osta A, Kantharidis P, Zalcberg JR, Wolffe AP. 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 2002; 22:1844-57.

52. Chen WG, Chang Q, Lin Y, Meissner A, West AE, Griffith EC, Jaenisch R, Greenberg ME. Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. Science 2003; 302:885-9.

53. Martinowich K, Hattori D, Wu H, Fouse S, He F, Hu Y, Fan G, Sun YE. DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. Science 2003; 302:890-3.

54. McLachlan DR, Lewis PN, Lukiw WJ, Sima A, Bergeron C, De Boni U. Chromatin structure in dementia. Ann Neurol 1984; 15:329-34.

55. Baker SA, Chen L, Wilkins AD, Yu P, Lichtarge O, Zoghbi HY. An AT-Hook Domain in MeCP2 Determines the Clinical Course of Rett Syndrome and Related Disorders. Cell 2013; 152:984-96.

56. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell 2012; 151:1417-30.

57. Kass SU, Pruss D, Wolffe AP. How does DNA methylation repress transcription? Trends Genet 1997; 13:444-9.

58. Bayles R, Harikrishnan KN, Lambert E, Baker EK, Agrotis A, Guo L, Jowett JB, Esler M, Lambert G, El-Osta A. Epigenetic modification of the norepinephrine transporter gene in postural tachycardia syndrome. Arteriosclerosis, thrombosis, and vascular biology 2012; 32:1910-6.

59. Harikrishnan KN, Bayles R, Ciccotosto GD, Maxwell S, Cappai R, Pelka GJ, Tam PP, Christodoulou J, El-Osta A. Alleviating transcriptional inhibition of the norepinephrine slc6a2 transporter gene in depolarized neurons. J Neurosci 2010; 30:1494-501.

60. Ben-Shachar S, Chahrour M, Thaller C, Shaw CA, Zoghbi HY. Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. Hum Mol Genet 2009; 18:2431-42.

61. Li Q, Lee JA, Black DL. Neuronal regulation of alternative pre-mRNA splicing. Nature reviews Neuroscience 2007; 8:819-31.

62. Hartman TR, Qian S, Bolinger C, Fernandez S, Schoenberg DR, Boris-Lawrie K. RNA helicase A is necessary for translation of selected messenger RNAs. Nature structural & molecular biology 2006; 13:509-16.

63. Aratani S, Fujii R, Oishi T, Fujita H, Amano T, Ohshima T, Hagiwara M, Fukamizu A, Nakajima T. Dual roles of RNA helicase A in CREB-dependent transcription. Molecular and cellular biology 2001; 21:4460-9.

64. Abdelhaleem M, Maltais L, Wain H. The human DDX and DHX gene families of putative RNA helicases. Genomics 2003; 81:618-22.

65. Dardenne E, Pierredon S, Driouch K, Gratadou L, Lacroix-Triki M, Espinoza MP, Zonta E, Germann S, Mortada H, Villemin JP, et al. Splicing switch of an epigenetic regulator by RNA helicases promotes tumor-cell invasiveness. Nature structural & molecular biology 2012; 19:1139-46.

66. Kuhn AN, Reichl EM, Brow DA. Distinct domains of splicing factor Prp8 mediate different aspects of spliceosome activation. Proceedings of the National Academy of Sciences of the United States of America 2002; 99:9145-9.

67. Dellaire G, Makarov EM, Cowger JJ, Longman D, Sutherland HG, Luhrmann R, Torchia J, Bickmore WA. Mammalian PRP4 kinase copurifies and interacts with components of both the U5 snRNP and the N-CoR deacetylase complexes. Molecular and cellular biology 2002; 22:5141-56.

68. Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdren L, Coulpier F, et al. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. The EMBO journal 2010; 29:3082-93.

69. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, et al. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Molecular cell 2010; 39:925-38.

70. Sanuki R, Onishi A, Koike C, Muramatsu R, Watanabe S, Muranishi Y, Irie S, Uneo S, Koyasu T, Matsui R, et al. miR-124a is required for hippocampal axogenesis and retinal cone survival through Lhx2 suppression. Nat Neurosci 2011; 14:1125-34.

71. Mercer TR, Qureshi IA, Gokhan S, Dinger ME, Li G, Mattick JS, Mehler MF. Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. BMC Neurosci 2010; 11:14.

72. Chao HT, Chen H, Samaco RC, Xue M, Chahrour M, Yoo J, Neul JL, Gong S, Lu HC, Heintz N, et al. Dysfunction in GABA signalling mediates autism-like stereotypies and Rett syndrome phenotypes. Nature 2010; 468:263-9.

73. Pelka GJ, Watson CM, Radziewic T, Hayward M, Lahooti H, Christodoulou J, Tam PP. Mecp2 deficiency is associated with learning and cognitive deficits and altered gene activity in the hippocampal region of mice. Brain 2006; 129:887-98.

74. De Meyer T, Mampaey E, Vlemmix M, Denil S, Trooskens G, Renard JP, De Keulenaer S, Dehan P, Menschaert G, Van Criekinge W. Quality evaluation of methyl binding domain based kits for enrichment DNA-methylation sequencing. PLoS One 2013; 8:e59068.

75. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25:1754-60.

76. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009; 4:44-57.

77. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009; 37:1-13.

Table and figure captions

Figure 1. MeCP2 enrichment of TSS regions in chromatin fractions. Frequency profile plot. (**A**) Plot shows CpG di-nucleotide frequency over all gene TSS regions +/- 2000bp. Scatter plots illustrate MeCP2 enrichment of genomic regions designed to the TSS +/- 1000bp of all genes (MM9 assembly). Individual scatter plots illustrate MeCP2 enrichment relative to input material using sonicated chromatin from whole male mouse brain (**B**), Mono-nucleosome selected material from S1 fractionated whole male mouse brain (**C**), and either Mono (**D**) or Di-nucleosome selected material from S1 fractionated male mouse cerebellum (**E**). CpG di-nucleotides and ChIP-Seq read bases occurring in each genomic region were totaled and the enrichment value calculated as Log2 (MeCP2 base count/Input base count).

Figure 2. TSS region CpG content and methylation distribution. (**A**) Histogram of TSS (+/- 1000 bp) region frequency by CpG content for all genes. (**B**) Scatter plot of relative methylation level in male mouse cerebellum by TSS region CpG content according to methyl-miner enrichment.

Figure 3. MeCP2 motif enrichment in genome fractions. Plots (**A to C**) indicate MeCP2 fold enrichment of ChIP-Sequencing fragments compared to input for CG.n.AT4 content where there are ‘n’ number of random nucleotides (**A**), CpG content (**B**), and CG-AT4 content (**C**). Reads (100000) were randomly sampled from sonicated and mono-nucleosome whole brain, mono and di-nucleosome input and MeCP2 enriched ChIP-Seq libraries. The sampled read sequences were extended to 200bp and quantified for motif feature content. MeCP2 fold enrichment was determined by the ‘MeCP2 enriched reads count’ divided by ‘input reads count’ for sequence reads binned by feature motif content. Plots (**D and E**) illustrate the ratio of genes (Activated in wild type : Repressed in wild type) between wild type samples represented in absolute value (**D**), and between wild type and MeCP2 null samples (**E**). The gene ratio was determend by RNA-Seq analysis of male mouse cerebellum, with grouping according to the CG-AT4 content of regions desgined to the TSS +/- 1000bp of all genes. Plots represent genes with TSS region CG content of below 50 occurences (red), above 50 occurences (blue), or combined CG content genes (black)

Figure 4. Methylation levels of TSS regions with high CG-AT4 content link MeCP2 differential expression. Traces illustrate Methyl-Miner and MeCP2 ChIP data from cerebellum, over genomic regions aligned to gene TSS +/- 2000bp. Individual regions were filtered to have at least 50 CpG di-nucleotides and divided into high (>18 times) and low (less than 7 times) of CG-AT4 content. These regions were further subdivided into groups that were either activated or repressed in wild-type male mouse cerebellum compared to MeCP2-Null according to mRNA-Seq expression analysis. (**A**) and (**B**) display the averaged methylation profiles of TSS region subgroups of total high salt (2000mM) methylminer elution data, while (**C**) and (**D**) shows methylation profiles of high salt relative to low salt elution (Number 2000mM reads divided by 350mM reads). (**E**) and (**F**) display MeCP2 fold enrichment of Di-nucleosome material over input over TSS region subgroups. Profiles of genes activated in wild-type are shown in 'green', and repressed in wild-type as 'red'.

Figure 5. MeCP2 Co-immunoprecipitation confirms M/S identified proteins and shows RNA dependance. Soluable mouse brain nuclear extract was pre-treated with either 0, 10, 30 or 100 µg / mL RNase A before Co-immunoprecipitation using endogenous MeCP2. The precipitated proteins were resolved by SDS-PAGE for mass spectrometry analysis and western blot confirmation. Proteins listed in (**A**) were identifiable through mass spectrometry analysis of MeCP2 Co-IP material with or without RNase treatment as indicated. Western blot confirmations (**B**) were carried out for mass spectrometry identified proteins where effective antibodies were available. These proteins included TOP2b-DNA topoisomerase, DHX9 RNA helicase and PRP8 RNA splicing factor. Western blots were also examined for the SWI/SNF components BRM and BRG-1, previously shown to interact with MeCP2.

Figure 6. MeCP2 RIP sequencing of LncRNAs and assessment of *RNCR3* and *MALAT1*.

(A) List of MeCP2 RIP captured LncRNAs ranked by RIP-sequencing reads per 1000 bp of transcript length. LncRNAs having either no RIP reads or reads that were indistinguishable from the enrichment of a overlapping genomic transcript were not considered in this list. (**B**) Expression and RIP primer sets were designed over *RNCR3* and *MALAT1*, chromosomes 14 and 19 respectively as shown: TSS indicated by directional arrows, gene structure shows either red or blue exon regions according to direction. Overlapping primer sets occur in yellow. Positioning of micro RNAs are marked in the gene structure. (**C**) MeCP2 RIP fold enrichment adjusted to input and relative to *GAPDH*, error bars indicate SEM. (**D**) MeCP2 Expression shows distribution, mean and standard deviation of RT-PCR differential expression analysis between wild type and MeCP2 null mouse cerebellum. RIP data were analyzed by t-test compared to *GAPDH*, and expression data analyzed by ANOVA. Significant p values are as indicated.

Table 1. Gene pathway analysis of motif content in TSS regions.

Gene regions +/- 1000bp of the TSS were quantified for the occurrence of CG or CG-AT4 motifs for both mouse (MM9 assembly) and Human (HG18 assembly) genomes. Approximately 3000 of the highest CG-AT4 content ranked genes were submitted to pathways analysis.

Table 2. Gene pathway analysis: high CG-AT4 content vs high MeCP2 Di-nucleosome binding.

Gene regions +/- 1000bp of the TSS were quantified for the occurrence of CG-AT4 motifs in the mouse (MM9 assembly) genome, as well as the ChIP-Seq binding ratio of MeCP2 vs Input of the Di-nucleosome fraction. The 3000 highest CG-AT4 content or MeCP2 binding ratio ranked genes were submitted to pathway analysis for comparison.

Table 3. Gene pathway analysis: low vs high MeCP2 Di-nucleosome binding.

Gene regions +/- 1000bp of the TSS were quantified for the occurrence of CG-AT4 motifs in the mouse (MM9 assembly) genome, as well as the ChIP-Seq binding ratio of MeCP2 vs Input of the Di-nucleosome fraction. Genes were separated into two groups relative MeCP2 ChIP-Seq binding data including- Low MeCP2 binding (MeCP2:Input ratio < 1) and High MeCP2 binding (MeCP2:Input ratio >1). The 3000 highest CG-AT4 content genes were selected from each group and subjected to pathway analysis.

Table 4. High and Low RIP signal pathways analysis.

The highest and lowest RIP signal transcripts adjusted for transcript length and having at least 40 combined RIP-Seq reads in total per transcript (1500 each) were analysed for functional pathways association. Pathways are also shown for high and low RIP signals relative to expression.

Table 1. Gene pathway analysis of motif content in TSS regions.

Gene regions +/- 1000bp of the TSS were quantified for the occurrence of CG or CG-AT4 motifs for both mouse (MM9 assembly) and Human (HG18 assembly) genomes. Approximately 3000 of the highest CG-AT4 content ranked genes were submitted to pathways analysis.

Mouse Pathways MM9 assembly							
P Val	High CG	P Val	Low CG	P Val	High CGAT4	P Val	Low CGAT4
3E-21	Wnt signaling pathway	9E-191	Olfactory transduction	1E-18	Ribosome	1E-95	Olfactory transduction
5E-19	Pathways in cancer	4E-10	Retinol metabolism	5E-17	Spliceosome	6E-08	Linoleic acid metabolism
1E-13	Chronic myeloid leukemia	8E-07	Linoleic acid metabolism	1E-06	Huntington's disease	6E-08	Retinol metabolism
2E-13	Colorectal cancer	1E-06	Metabolism of xenobiotics by cytochrome P450	2E-06	Cell cycle	4E-06	Metabolism of xenobiotics by cytochrome P450
2E-13	Melanogenesis	1E-06	Taste transduction	3E-06	RNA degradation	9E-06	Arachidonic acid metabolism
4E-13	Neurotrophin signaling pathway	2E-06	Drug metabolism	1E-05	Oxidative phosphorylation	3E-05	Drug metabolism
5E-12	Adherens junction	2E-03	Steroid hormone biosynthesis	2E-05	Parkinson's disease	2E-02	Taste transduction
5E-11	MAPK signaling pathway	4E-03	Arachidonic acid metabolism	6E-05	Pyrimidine metabolism	3E-02	Steroid hormone biosynthesis
3E-10	TGF-beta signaling pathway	3E-02	Drug metabolism	2E-04	RNA polymerase	8E-02	Cytosolic DNA-sensing pathway
1E-09	Insulin signaling pathway			4E-03	Homologous recombination		
1E-09	ErbB signaling pathway			4E-03	Basal transcription factors		
3E-09	Prostate cancer			4E-03	Alzheimer's disease		
9E-09	Axon guidance			7E-03	Purine metabolism		
2E-08	Chemokine signaling pathway			2E-02	Proteasome		
4E-08	Non-small cell lung cancer			2E-02	p53 signaling pathway		
5E-08	Glioma			2E-02	DNA replication		
6E-08	Basal cell carcinoma			3E-02	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis		
9E-08	Ubiquitin mediated proteolysis			4E-02	Base excision repair		
1E-07	Regulation of actin cytoskeleton			5E-02	Glioma		
5E-07	Endometrial cancer			6E-02	Bladder cancer		
6E-07	Acute myeloid leukemia			6E-02	Aminoacyl-tRNA biosynthesis		
1E-06	Hedgehog signaling pathway			6E-02	Nucleotide excision repair		
1E-06	Small cell lung cancer			6E-02	Mismatch repair		
2E-06	Tight junction			7E-02	Prostate cancer		
2E-06	Pancreatic cancer			8E-02	Chronic myeloid leukemia		
Human Pathways HG18 assembly							
9E-16	Wnt signaling pathway	6E-24	Olfactory transduction	2E-17	Ribosome	3E-04	Retinol metabolism
1E-14	Pathways in cancer	2E-10	Retinol metabolism	1E-15	Spliceosome	1E-03	Arachidonic acid metabolism
1E-11	Adherens junction	2E-09	Drug metabolism	1E-06	Cell cycle	2E-03	Olfactory transduction
4E-11	Melanogenesis	8E-09	Autoimmune thyroid disease	6E-06	Systemic lupus erythematosus	5E-03	Cytokine-cytokine receptor
2E-10	Neurotrophin signaling pathway	9E-09	Metabolism of xenobiotics by cytochrome P450	5E-04	Mismatch repair	2E-02	Chemokine signaling
4E-09	MAPK signaling pathway	8E-07	Regulation of autophagy	7E-04	Protein export	3E-02	Calcium signaling pathway
5E-09	Colorectal cancer	2E-06	Steroid hormone biosynthesis	1E-03	Huntington's disease	3E-02	Toll-like receptor signaling
5E-08	TGF-beta signaling pathway	5E-06	Antigen processing and presentation	1E-03	DNA replication	3E-02	Neuroactive ligand-receptor interaction
6E-08	Basal cell carcinoma	6E-06	Drug metabolism	2E-03	Oxidative phosphorylation	3E-02	Natural killer cell mediated cytotoxicity
9E-08	Axon guidance	2E-05	Complement and coagulation cascades	2E-03	Homologous recombination	3E-02	NOD-like receptor signaling pathway
3E-06	Focal adhesion	6E-05	Cytosolic DNA-sensing	4E-03	Aminoacyl-tRNA biosynthesis	4E-02	Fc epsilon RI signaling
4E-06	Regulation of actin cytoskeleton	1E-04	RIG-I-like receptor signaling	4E-03	Proteasome	5E-02	GnRH signaling pathway
7E-06	Hedgehog signaling pathway	1E-04	Cytokine-cytokine receptor	7E-03	Alzheimer's disease	5E-02	Cytosolic DNA-sensing
1E-05	GnRH signaling pathway	2E-04	Toll-like receptor signaling	7E-03	Parkinson's disease	5E-02	PPAR signaling pathway
2E-05	Long-term potentiation	2E-04	Starch and sucrose metabolism	2E-02	RNA degradation	6E-02	RIG-I-like receptor signaling pathway
3E-05	Renal cell carcinoma	3E-04	Porphyrin and chlorophyll metabolism	3E-02	beta-Alanine metabolism	8E-02	Jak-STAT signaling pathway
5E-05	Oocyte meiosis	5E-04	Ascorbate and aldarate metabolism	3E-02	Basal transcription factors	9E-02	Primary immunodeficiency
1E-04	Prostate cancer	6E-04	Pentose and glucuronate interconversions	5E-02	Non-homologous end-joining		
1E-04	Chronic myeloid leukemia	6E-04	Androgen and estrogen metabolism	6E-02	Pyrimidine metabolism		
1E-04	Melanoma	1E-03	Natural killer cell mediated cytotoxicity	6E-02	Selenoamino acid metabolism		
2E-04	Calcium signaling pathway	8E-03	ABC transporters	8E-02	Cysteine and methionine metabolism		
2E-04	ErbB signaling pathway	2E-02	Jak-STAT signaling pathway	8E-02	Colorectal cancer		
2E-04	Glioma	3E-02	Linoleic acid metabolism				
2E-04	Tight junction	5E-02	Primary immunodeficiency				
3E-04	Type II diabetes mellitus	8E-02	Fatty acid metabolism				

Table 2. Gene pathway analysis: high CG-AT4 content vs high MeCP2 Di-nucleosome binding.

Gene regions +/- 1000bp of the TSS were quantified for the occurrence of CG-AT4 motifs in the mouse (MM9 assembly) genome, as well as the ChIP-Seq binding ratio of MeCP2 vs Input of the Di-nucleosome fraction. The 3000 highest CG-AT4 content or MeCP2 binding ratio ranked genes were submitted to pathway analysis for comparison.

P Value	Pathways - High CG-AT4	P Value	Pathways - MeCP2 Di-nuclesome binding
1.9E-18	Ribosome	4.5E-11	Ribosome
2.3E-17	Spliceosome	5.6E-05	RNA polymerase
1.1E-06	Cell cycle	1.7E-04	Pyrimidine metabolism
1.1E-06	Huntington's disease	1.8E-04	Cell cycle
4.5E-06	Oxidative phosphorylation	2.4E-04	Neurotrophin signaling pathway
7.6E-06	Parkinson's disease	5.8E-04	Purine metabolism
9.0E-06	RNA degradation	7.7E-04	Lysosome
1.6E-04	Pyrimidine metabolism	3.9E-03	Huntington's disease
3.0E-04	RNA polymerase	4.8E-03	ErbB signaling pathway
1.0E-03	Proteasome	5.3E-03	Oocyte meiosis
3.3E-03	Alzheimer's disease	9.0E-03	Circadian rhythm
6.0E-03	Homologous recombination	1.1E-02	Oxidative phosphorylation
6.6E-03	Basal transcription factors	1.3E-02	Spliceosome
1.4E-02	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	1.9E-02	Ubiquitin mediated proteolysis
1.6E-02	Purine metabolism	2.6E-02	Mismatch repair
2.5E-02	Mismatch repair	2.7E-02	Parkinson's disease
2.9E-02	DNA replication	3.0E-02	Prion diseases
3.1E-02	p53 signaling pathway	3.8E-02	Cardiac muscle contraction
3.2E-02	Aminoacyl-tRNA biosynthesis	4.0E-02	Insulin signaling pathway

Table 3. Gene pathway analysis: low vs high MeCP2 Di-nucleosome binding.

Gene regions +/- 1000bp of the TSS were quantified for the occurrence of CG-AT4 motifs in the mouse (MM9 assembly) genome, as well as the ChIP-Seq binding ratio of MeCP2 vs Input of the Di-nucleosome fraction. Genes were separated into two groups relative MeCP2 ChIP-Seq binding data including- Low MeCP2 binding (MeCP2:Input ratio < 1) and High MeCP2 binding (MeCP2:Input ratio >1). The 3000 highest CG-AT4 content genes were selected from each group and subjected to pathway analysis.

	Pathways		**Pathways**
P Value	**Low MeCP2 binding – high CG-AT4**	**P Value**	**High MeCP2 binding - high CG-AT4**
1.8E-08	Neuroactive ligand-receptor interaction	5.3E-24	Ribosome
2.4E-03	Complement and coagulation cascades	1.6E-18	Spliceosome
1.1E-02	Cell adhesion molecules (CAMs)	4.2E-08	Huntington's disease
3.6E-02	Adipocytokine signaling pathway	2.3E-07	Oxidative phosphorylation
3.6E-02	Calcium signaling pathway	2.9E-07	RNA degradation
4.6E-02	ABC transporters	4.2E-07	Parkinson's disease
4.7E-02	TGF-beta signaling pathway	1.7E-05	Pyrimidine metabolism
4.7E-02	Apoptosis	3.0E-04	Alzheimer's disease
		3.2E-04	DNA replication
		4.7E-04	Cell cycle
		4.9E-04	RNA polymerase
		4.9E-04	Homologous recombination

Table 4. High and Low RIP signal pathways analysis.

The highest and lowest RIP signal transcripts adjusted for transcript length and having at least 40 combined RIP-seq reads in total per transcript (1500 each) were analysed for functional pathways association. Pathways are also shown for high and low RIP signals relative to expression.

High RIP signal		High RIP signal relative to expression	
P Value	**Term**	**P Value**	**Term**
9.22E-08	Oxidative phosphorylation	5.98E-04	Systemic lupus erythematosus
5.97E-07	Alzheimer's disease	1.22E-02	Complement and coagulation cascades
2.17E-06	Parkinson's disease	2.61E-02	Calcium signaling pathway
2.17E-06	Huntington's disease	3.10E-02	Primary immunodeficiency
9.59E-05	Spliceosome	3.13E-02	MAPK signaling pathway
5.34E-04	Cardiac muscle contraction	3.39E-02	Fructose and mannose metabolism
2.65E-03	Aldosterone-regulated sodium reabsorption	5.53E-02	Spliceosome
3.68E-03	Neurotrophin signaling pathway	5.85E-02	ECM-receptor interaction
5.47E-03	Ribosome		
9.50E-03	Lysosome		
1.71E-02	Glycolysis / Gluconeogenesis		
2.04E-02	Long-term potentiation		
2.83E-02	Gap junction		
3.70E-02	Circadian rhythm		
3.72E-02	MAPK signaling pathway		
4.12E-02	Glycerophospholipid metabolism		
5.40E-02	Glycosphingolipid biosynthesis		
Low RIP signal		**Low RIP signal relative to expression**	
6.75E-12	Focal adhesion	2.71E-04	Purine metabolism
6.37E-09	ECM-receptor interaction	7.39E-04	Valine, leucine and isoleucine degradation
7.36E-07	Pathways in cancer	8.46E-04	Lysosome
7.31E-06	Small cell lung cancer	2.67E-03	Long-term depression
1.43E-05	Neuroactive ligand-receptor interaction	4.44E-03	Glycine, serine and threonine metabolism
3.43E-04	Pancreatic cancer	1.17E-02	Proteasome
4.95E-04	Axon guidance	1.26E-02	Oxidative phosphorylation
1.01E-03	Dilated cardiomyopathy	1.30E-02	Propanoate metabolism
1.04E-03	Regulation of actin cytoskeleton	1.53E-02	Melanogenesis
1.80E-03	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	2.73E-02	Gap junction
2.19E-03	ABC transporters	3.11E-02	Biosynthesis of unsaturated fatty acids
2.53E-03	Calcium signaling pathway	3.14E-02	Parkinson's disease
4.47E-03	Hypertrophic cardiomyopathy (HCM)	4.24E-02	Renal cell carcinoma
4.76E-03	Cytokine-cytokine receptor interaction	4.24E-02	Long-term potentiation
6.99E-03	Non-small cell lung cancer	5.27E-02	Citrate cycle (TCA cycle)
8.58E-03	Purine metabolism	5.46E-02	Pyrimidine metabolism
1.60E-02	TGF-beta signaling pathway	5.67E-02	Vascular smooth muscle contraction
2.28E-02	Hedgehog signaling pathway	5.87E-02	Sphingolipid metabolism
2.47E-02	B cell receptor signaling pathway		
4.67E-02	Adherens junction		
4.77E-02	Prostate cancer		
5.22E-02	Vascular smooth muscle contraction		