

INVITED REVIEW ARTICLE

# Regulation, diversity and function of *MECP2* exon and 3'UTR isoforms

Deivid Carvalho Rodrigues<sup>1,\*</sup>, Marat Mufteev<sup>1,2</sup> and James Ellis<sup>1,2</sup>

<sup>1</sup>Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, Toronto ON M5G 0A4, Canada and <sup>2</sup>Department of Molecular Genetics, University of Toronto, Toronto ON M5S 1A8, Canada

\*To whom correspondence should be addressed at: Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, Toronto ON M5G 0A4, Canada. Tel: +1 4168138169; Email: deivid.rodrigues@sickkids.ca

## Abstract

The methyl-CpG-binding protein 2 (*MECP2*) is a critical global regulator of gene expression. Mutations in *MECP2* cause neurodevelopmental disorders including Rett syndrome (RTT). *MECP2* exon 2 is spliced into two alternative messenger ribonucleic acid (mRNA) isoforms encoding *MECP2*-E1 or *MECP2*-E2 protein isoforms that differ in their N-termini. *MECP2*-E2, isolated first, was used to define the general roles of *MECP2* in methyl-deoxyribonucleic acid (DNA) binding, targeting of transcriptional regulatory complexes, and its disease-causing impact in RTT. It was later found that *MECP2*-E1 is the most abundant isoform in the brain and its exon 1 is also mutated in RTT. *MECP2* transcripts undergo alternative polyadenylation generating mRNAs with four possible 3'untranslated region (UTR) lengths ranging from 130 to 8600 nt. Together, the exon and 3'UTR isoforms display remarkable abundance disparity across cell types and tissues during development. These findings indicate discrete means of regulation and suggest that protein isoforms perform non-overlapping roles. Multiple regulatory programs have been explored to explain these disparities. DNA methylation patterns of the *MECP2* promoter and first intron impact *MECP2*-E1 and E2 isoform levels. Networks of microRNAs and RNA-binding proteins also post-transcriptionally regulate the stability and translation efficiency of *MECP2* 3'UTR isoforms. Finally, distinctions in biophysical properties in the N-termini between *MECP2*-E1 and E2 lead to variable protein stabilities and DNA binding dynamics. This review describes the steps taken from the discovery of *MECP2*, the description of its key functions, and its association with RTT, to the emergence of evidence revealing how *MECP2* isoforms are differentially regulated at the transcriptional, post-transcriptional and post-translational levels.

## Introduction

The biogenesis of eukaryotic messenger ribonucleic acid (mRNAs) in the cell nucleus involves processes ranging from transcription initiation to nuclear export. These include maturation steps, such as splicing where introns are removed and exons joined together, and cleavage and polyadenylation (pA) where the cleavage of the nascent RNA is followed by synthesis of a poly-A tail at its 3'-end. Both these mRNA maturation processes

are highly regulated across cell types and developmental stages (1,2). Fundamentally, splicing and pA are post-transcriptionally regulated by interactions between cis-elements in the primary mRNA and trans-acting factors, including RNA-binding proteins (RBPs) and non-coding RNAs (2,3), and have significant impact on the fate of the mRNA molecule in the cell.

The diversity of mRNA isoforms produced by any given gene is generated by the use of alternative exons and pA

Received: June 23, 2020. Revised: July 12, 2020. Accepted: July 14, 2020

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com  
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

sites. The presence of isoforms with diverse composition of exons serves the central role of increasing the genomic coding capacity, resulting in the greater proteomic diversity observed in mammals relative to other organisms with comparable number of genes (4). Alternative exon isoforms are the products of alternative splicing (AS) mechanisms that comprise, but are not limited to, inclusion or skipping of exons, use of alternative splice sites and intron retention (for a comprehensive review on AS mechanisms see (3)). Importantly, AS is a highly dynamic and context-dependent process and can generate mRNA isoforms with different untranslated regions (UTRs) or coding-sequences (CDS) (3).

Alternative 3'UTR isoforms allow distinct responses to post-transcriptional regulatory pressure due to the change in the number of regulatory cis-elements. For example, longer 3'UTR-bearing transcripts often have an increased likelihood of being bound by microRNAs (miRNAs) that target transcripts for destabilization or translation inhibition (5). These longer 3'UTRs also have more capacity to interact with regulatory RBPs to impact isoform-specific sub-cellular localization, stability, or translation efficiency (6–9). Alternative polyadenylation (APA) is a prevalent gene regulatory mechanism that creates transcripts of the same gene with distinct 3'-ends (for a comprehensive review on APA mechanisms see (10)). APA is highly regulated in response to environmental stimuli (7,11). During neurodevelopment, in particular, mRNAs with longer 3'UTRs accumulate through APA or by reduced RNA degradation of the long isoform (8,12).

In this review, we describe nearly 30 years of experimental findings that led to the characterization of the methyl-CpG-binding protein 2 (MECP2) mRNA isoforms, we discuss studies suggesting different protein isoform functions and summarize current knowledge regarding differential regulation of MECP2 exon and 3'UTR isoforms at the transcriptional, post-transcriptional and post-translational levels.

## Overview of Rett Syndrome and the MECP2 Gene Alternative Exon Isoforms

MECP2 is an X-linked gene encoding an abundant chromosomal protein that modulates transcription and translation globally and is crucial for normal brain development (13). Damaging mutations in the MECP2 gene are the primary cause of Rett syndrome (RTT) (14), a neurological disorder that occurs in one in 10–15 000 female births. After an initial window of normal development, girls acquire a variety of symptoms including microcephaly, autism, ataxia, loss of learned speech and purposeful hand movements, seizures, and hyperventilation (15). Duplications or triplications in the MECP2 locus in humans also lead to neurological deficits, implicating that MECP2 protein levels must be stringently regulated throughout neurodevelopment (16).

The MECP2 gene transcribes pre-mRNA encompassing four exons, which through AS mature into two transcripts encoding two protein isoforms. According to an agreed nomenclature (Rett Syndrome Research Foundation annual meeting, June 28–30, 2004), the alternative exon isoforms are named MECP2-E1 (skipping exon 2, protein encoded by exons 1, 3 and 4, 492 amino acids (aa)—previously identified as MECP2 $\alpha$  or B) and MECP2-E2 (including all 4 exons, protein encoded by exons 2, 3 and 4, 486 aa—previously identified as MECP2 $\beta$  or A) (Fig. 1). Customarily, splicing variants encode functionally diverse protein isoforms; however, in the case of MECP2-E1 and E2, the two protein isoforms only differ by 21 and 9 aa, respectively, in their N-Termini (Fig. 1).

## Discovery of the Mecp2-E2 Exon Isoform

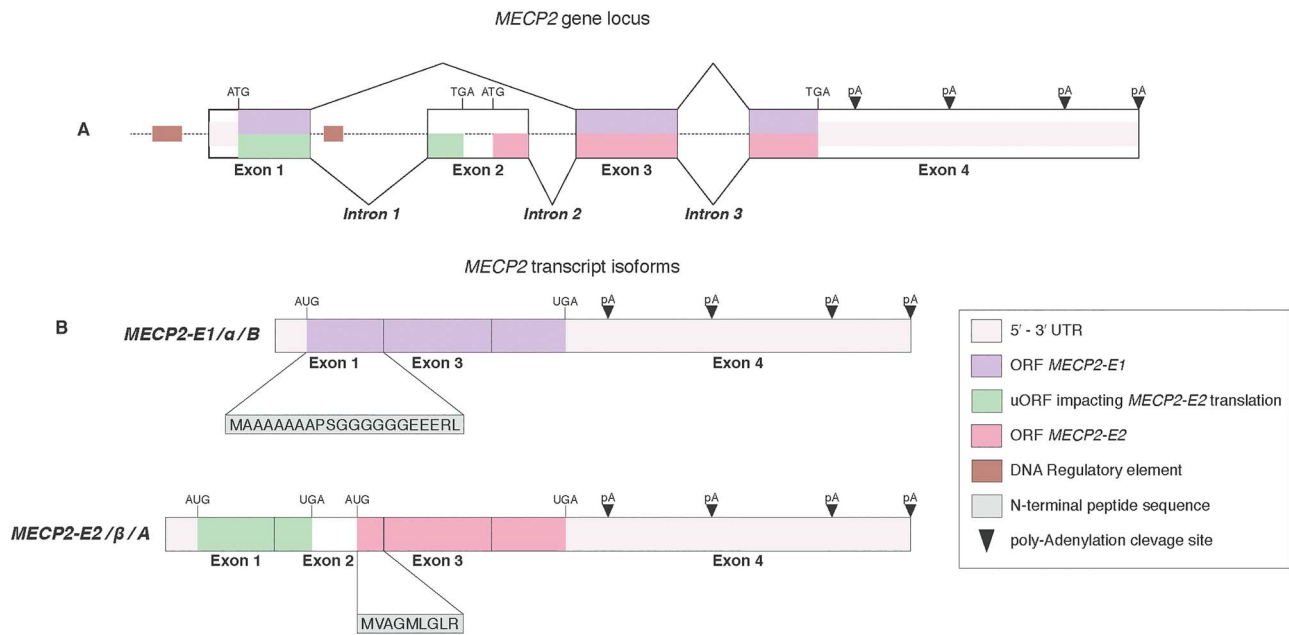
Mecp2 was described by Adrian Bird's group in 1992 as a protein that bound to deoxyribonucleic acid (DNA) probes containing a single methyl-CpG pair, distinguishing it from the Mecp1 complex that required 12 methyl-CpG pairs (17). A DNA probe derived from the partially sequenced 84-kDa protein isolated from rat brain lysates was used to screen a rat brain complementary DNA (cDNA) library. The resultant cDNA clone had an open reading frame (ORF) of 492-aa and its N-terminal sequence is diagnostic of the Mecp2-E2 protein isoform (Fig. 1). Detailed characterizations of the rat Mecp2 protein tissue distribution, chromatin association and its methyl-CpG binding domain were accomplished by the same group in 1992 and 1993 (18,19).

Using chromosome mapping approaches and the rat Mecp2 cDNA as probe (E2 isoform), the gene encoding Mecp2 in mouse was mapped to the X-chromosome (20), then found to be subject to X inactivation (21,22). The thorough physical mapping indicated a likely position of the human homolog (20), which was confirmed to be the Xq28 chromosome band (22,23). In the first characterization of the human MECP2 gene (22), the rat Mecp2 cDNA probe containing 80% of its CDS was used to screen a human skeletal muscle cDNA library. Multiple hits were overlapped to deduce the human MECP2 cDNA sequence with a single ORF encoding a putative 485-aa protein. When compared with the rat cDNA sequence, the human cDNA homolog lacked the sequence coding for the first 9-aa of the rat protein (MVAGMLGLR—Fig. 1). Incidentally, that truncation in the human cDNA clone maps exactly to the junction between exons 1 and 2, upstream of which, the two protein isoforms differ. Therefore, it is not possible to distinguish whether the first mapped human MECP2 cDNA clone was the E1 or E2 isoform. However, using the 5'-end of the rat cDNA (MECP2-E2), a search of human expressed sequence tag (EST) databases identified a clone (22) now known to be MECP2-E2.

## Defining Molecular Functions of Mammalian MECP2-E2 and *Xenopus* Mecp2-E1

The ability of MECP2 to bind methylated CpGs (17–19,24,25), and recruit transcription repressive complexes such as HDAC and mSin3 (24), in human and mouse were characterized using the MECP2-E2 isoform. Similar findings were reported by Alan Wolffe's group that in *Xenopus laevis*, Mecp2 was able to recruit histone deacetylases to repress transcription (25), and to stably associate with nucleosomal DNA (26). In both reports, the N-terminal sequence of the *Xenopus* Mecp2 cDNA used for recombinant protein and antibody production was similar to the human MECP2-E1 isoform (to date, no MECP2-E2 isoform has been reported for the *Xenopus* genera). Similarly, chicken Mecp2 was identified as being the previously reported attachment region binding protein (27). These authors used 5'-rapid amplification of cDNA ends (5' RACE) and all clones obtained showed a polyalanine tract in the N-terminal downstream of the first methionine, a signature of the Mecp2-E1 isoform. Together these results suggest that the two isoforms retain grossly similar functions across species.

Analysis of the MECP2-E2 coding region showed high conservation levels between human and several primate species, and most substitutions were found in non-critical regions of MECP2 protein. Interestingly, one Alu insert of 300-bp was found in intron 3 of two primate species, and a novel alternatively spliced transcript containing a region of the human-homologous intron 2 was found in a third species with the potential to code



**Figure 1.** Graphical description of the *MECP2* gene locus (A) and transcript isoforms (B). *MECP2* gene comprises four exons and three introns that are alternatively spliced to generate two transcript isoforms *MECP2-E1* and *MECP2-E2*. The methylation of DNA regulatory elements in the *MECP2* promoter, and first intron have been shown to differentially affect abundances of the splicing isoforms. *MECP2-E2* isoform also encodes an upstream ORF in its 5'UTR that affects translation of the main *MECP2* ORF. *MECP2* transcripts also undergo alternative polyadenylation generating transcripts with 3'UTRs that range from 130 to 8600 nucleotides. Model is not to scale.

for two putative polypeptides (28). Another study focusing on the analysis of the ordered and disordered regions of the *MECP2* protein suggested that these regions are stably conserved in chordates. Although insertions and deletions were more frequently found in the disordered regions, indicating that these are under higher evolution rates, the overall conservation pattern suggests that these regions could potentially harbor the typical functional roles of disorder protein structures including entropic chain activity or transient protein-binding sites (29).

### Experiments Linking the *MECP2-E2* Isoform to Disease in Patients and Mice

In the seminal publication by Amir *et al.* (14), from Huda Zoghbi's group, demonstrating that RTT is caused by mutations in *MECP2*, the primers used to amplify *MECP2* exons from a cohort of patients were based on the *MECP2-E2* isoform, thus leaving exon 1 out of the mutational analysis. Later, comparative sequence analysis of the *MECP2* locus in human and mouse led to the identification of the new 5'-exon (30). It was found to be transcribed in both species, but thought to be non-coding and extended the 5'UTR of *MECP2-E2*. The newly discovered upstream exon harbored a CpG island, believed to participate in the regulation of *MECP2* transcription, but today is known to code the polyalanine tract in the N-terminal of *MECP2-E1* protein isoform.

Subsequently, the *Mecp2*-null knockout mouse models generated by Adrian Bird (*Mecp2*<sup>tm1.1<sup>Bird</sup></sup>) (31), Rudolf Jaenisch (*Mecp2*<sup>tm1.1<sup>Jaen</sup></sup>) (32), Huda Zoghbi (*Mecp2*<sup>308</sup>) (33) and their colleagues demonstrated the critical role of *Mecp2* for RTT pathophysiology. Similar to the first *Mecp2*-knockout mouse generated in 1996 (34), all these models disrupt both isoforms but left the then recently discovered exon 1 and the upstream ORF (uORF) from the *Mecp2-E2* isoform intact. Although no functional role has been assigned to the peptide coded by *Mecp2-E2* uORF,

recent reports have suggested that such small peptides can have important biological roles (35,36). It is important to note that these valuable mouse models are well established and in common use for RTT phenotyping and rescue studies described later.

### Definition of the *MECP2-E1* Isoform in Mice and Patients

It was not until 12 years after the initial isolation and characterization of rat *Mecp2* that Kriaucionis and Bird (37) and Mnatzakanian *et al.* (38) added 'the additional complication' (as quoted in the former report) that *MECP2* undergoes AS generating two transcript isoforms and proteins with distinct N-termini. In the Kriaucionis report, a deep search into human and mouse EST databases revealed two categories of clones based on the presence or absence of exon 2. The transcript lacking exon 2 was first considered a non-coding mRNA as exon 2 contained the known translation start-codon. However, careful analysis revealed that exon 1 also contained an ATG in-frame with the rest of *MECP2* ORF. Therefore, even in the absence of exon 2, this new isoform encoded putative proteins of 501 and 498-aa in mice and humans, respectively. The authors named the new isoform lacking exon 2 as *MECP2 $\alpha$*  and the previously described transcript containing all four exons *MECP2 $\beta$* , but for clarity we will use the E1 and E2 nomenclature (Fig. 1). Sequence analysis showed that the new human and mouse *MECP2-E1* isoform shared the N-terminal polyalanine and polyglycine tracts with *Xenopus*, zebrafish and *Fugu* *Mecp2* sequences, indicating that *MECP2-E1* was the ancestral form of *MECP2* gene. In addition, transient transfection of plasmids encoding the full cDNA sequence of *MECP2-E1* or *MECP2-E2* isoforms showed that the latter translated at lower efficiencies, which was demonstrated to be caused by the presence of an uORF in its 5'UTR (Fig. 1).

The newly identified protein isoform also co-localized with the major satellite sequences (containing the majority of methylated DNA) in mouse cell nuclei, indicating that the different N-terminal of MECP2-E1 probably did not change the chromatin functions already described for mammalian MECP2-E2 and previously observed in *Xenopus* and chicken.

Above all, the discovery of a new coding exon 1 had implications for RTT studies. Mutations in exon 1 could potentially impact > 90% of the MECP2 protein content in cells, and as exon 1 had been previously described as a non-coding exon, it was excluded from mutational screening of RTT patients. Indeed, the analyses of MECP2 function in the context of RTT mutations (39) were done without the knowledge of a coding exon 1, and until 2002 (40) no mutations had been assigned to exon 1. In 2004, Mnatzakanian *et al.* (38) independently identified the translation potential starting in exon 1 of the new MECP2-E1 isoform and validated it in human and mouse tissue samples. Importantly, the authors described new mutations in the new CDS of exon 1 in a cohort of patients with typical RTT. Soon after, evidence accumulated that mutations in exon 1 of MECP2 and its promoter region were not common in RTT patients (41–43), an observation that still holds to today. In summary, MECP2-E1 variants are rare in RTT and all other mutations affect both isoforms.

### Consequences of MECP2-E2 Overexpression and Rescue

One way to begin to study the role of MECP2-E2 was to perform over-expression and rescue experiments in KO mice. Indeed, transgenic animals overexpressing excessive levels of Mesp2-E2 (in addition to the endogenous) displayed a range of severe motor and neurological dysfunctions, indicating that proper regulation of Mesp2 expression is critical for normal neurodevelopment (41,42). Likewise, it was shown that transient overexpression of the Mesp2-E2 isoform in mouse neurons cultured *in vitro* promoted enhanced axonal and dendritic outgrowth with significant gain in morphological complexity (43). The overexpression of MECP2-E2, but not E1, promoted cell death in healthy neurons (44). Although the mechanisms for specific MECP2-E2 neurotoxicity were not explored, the report showed that the neurotoxic effect was normally inhibited by direct interaction between MECP2-E2 and FOXG1 proteins, whose mutations can cause congenital RTT (45), through binding with the N-terminal 9-aa unique to MECP2-E2 protein. An implication of these studies is that the regulation of MECP2-E2 abundance needs tight control for proper brain function.

In parallel with the over-expression studies, new mouse models and rescues with Mesp2 were performed. Mesp2-mutant mice engineered by the Jaenisch group to re-express appropriate levels of Mesp2-E2 in neurons were phenotypically indistinguishable from wild-type (WT) animals, rendering a complete rescue (46). In a subsequent study by the same group, neuronal-specific re-activation of Mesp2-E2 postnatally prolonged lifespan and delayed the onset of neurologic symptoms in Mesp2-KO mice (41). The seminal work of Guy *et al.* (47) showed that endogenous re-expression of both isoforms post-symptomatically partially rescued phenotypes, and dramatically raised hopes that RTT could be treated in patients after symptoms developed.

Ultimately a high degree of functional similarity between the isoforms was shown after re-expressing each isoform separately in a Mesp2-null mice model (48). However, the rescue of motor specific phenotypes was not thoroughly concordant, as the MECP2-E1 isoform had a stronger performance, leaving open

the possibility that the two isoforms might, in fact, retain distinct functions *in vivo*. Additionally in 2012, a specific disruption of Mesp2-E2 isoform in mice showed that although RTT-related phenotypes were not observed, a survival disadvantage for embryos carrying the Mesp2-E2-null allele, and placental development defects were noticed (49). These studies suggested that mutations in the MECP2-E2 isoform may result in phenotypes that escape RTT diagnosis and that MECP2-E2 might perhaps keep specific functions distinct from MECP2-E1.

### MECP2-E1 Disease Models, Rescues and Isoform Function

A mouse model with specific disruption of the Mesp2-E1 protein was generated by altering the start-codon in exon 1. Despite an unexpected increase in Mesp2-E2 protein levels, the model recapitulated the neurological phenotypes reminiscent of RTT, confirming the strong relationship between MECP2-E1 and RTT (50). It is interesting to note that the start-codon in exon 1 is the same used by the uORF encoded in the 5'UTR of MECP2-E2 isoform. Therefore, by preventing translation of the uORF, one could expect that the translation efficiency of the main ORF (MECP2-E2) would increase, as observed in that study.

In an effort to establish causative effects of mutations in the MECP2-E1 isoform and RTT in humans, induced pluripotent stem cells (iPSCs) were isolated from the RTT patient with the first described 11-bp deletion and frameshift in the CDS of exon 1, sparing the MECP2-E2 isoform (51). iPSC-derived neurons from MECP2-E1 deleted lines exhibited decreased soma size, reduced dendritic complexity and altered electrophysiological parameters redolent of RTT pathophysiology and consistent with previous iPSC-derived neurons models of RTT with mutations affecting both mRNA isoforms (52–55). An intriguing observation was that although re-expression of a transgene MECP2-E1 at proper levels rescued the soma size defects, MECP2-E2 overexpression did not result in any phenotypic improvement across different expression ranges (51).

In 2014, the overexpression of both isoforms separately in neurons differentiated from the human SK-N-SH neuroblastoma line caused different genes to be differentially regulated with a modest overlap (56). Gene ontology analyses indicated a more diverse set of biological functions affected by the MECP2-E1, particularly on genes involved in neuronal development and function, whereas genes impacted by MECP2-E2 overexpression were involved in chromatin organization and transcriptional regulation.

Recently, it was shown that the N-terminal domains (NTD) of MECP2-E1, and E2 can modulate the DNA-binding kinetics of MECP2, neural depolarization response and induce changes in protein levels during circadian oscillations (57). Biophysical characterization of the NTD indicated that MECP2-E1 has slightly lower structural stability, diminished unfolding enthalpy and exhibits lower binding affinities for both methylated and unmethylated double-stranded DNA (dsDNA) than MECP2-E2. Thermodynamic binding profiles showed that the binding of MECP2-E1 isoform with dsDNA is mainly driven by specific interactions promoted by hydrogen bonds and electrostatic interactions, whereas the binding of MECP2-E2 with dsDNA is mainly driven by unspecific interactions such as hydrophobic desolvation and steric arrangements. The authors showed that Mesp2-E1 protein levels, but not Mesp2-E2, were dynamically regulated during the circadian cycle and after KCl-induced depolarization. In addition, 40 interacting proteins for the



Mecp2-E1 isoform and 7 for Mecp2-E2 were found using mouse whole brain lysates. Mecp2-E1 co-eluted proteins are highly enriched for microtubule-associated proteins, proteins related to mRNA processing, and chromatin regulation. Interestingly, functional network analysis suggests the participation of Mecp2-E2 in processes similar to those involving Mecp2-E1, but through the interaction with a different set of protein partners.

## Mechanistic Insights into Differential Gene Regulation of MECP2-E1 and E2 Isoforms

The first reports describing the existence of two MECP2 isoforms were also first to indicate they might be differentially regulated (37,38). MECP2-E1 transcript was at least 10-fold more abundant than MECP2-E2, and the two isoforms translated MECP2 protein with distinct efficiencies. Initially, it was assumed that mutations in exon 1 had no effect on the MECP2-E2 protein, as MECP2-E2 mRNA levels were observed to be unchanged in patients (38). Shortly after, it was shown that mutations in exon 1 were associated with a reduction in MECP2-E2 protein levels despite no change in MECP2-E2 mRNA levels, which was suggested to be caused by translational interference (58).

Using riboprobes for *in situ* hybridization of MECP2 mRNA isoforms in neonatal and postnatal mouse brain samples, spatiotemporal abundance variation of isoforms was observed, thus providing the first demonstration that the two mRNA isoforms are differentially regulated in the brain (59). Using an *in vitro* neurodevelopment system, a reciprocal pattern in the change of isoform abundances was observed with MECP2-E1 levels reduced at the beginning of the differentiation then increasing toward the end, and MECP2-E2 levels increased in the beginning to then decrease at the end of the differentiation protocol (60).

Methylation of CpG dinucleotides in regulatory elements (RE) within the promoter and intron 1 of MECP2 had been observed to negatively associate with total MECP2 levels (61–63). By altering the DNA methylation status, the authors observed a significant correlation between CpG methylation at the REs and the ratio of MECP2-E1/MECP2-E2 abundances. Interestingly, only MECP2-E1 was altered by artificial changes in DNA methylation suggesting that its effect is specific for MECP2-E1 regulation. These results were corroborated by other reports showing that CpG methylation at REs were also associated with the patterns of MECP2-E1 and E2 isoform abundances in different regions of mouse brains (64,65).

Recently, sex-dependent differential accumulation of MECP2 isoforms in the neurons and astrocytes of mouse brains was also shown (66). Female neurons and astrocytes accumulate similar levels of MECP2-E1 transcripts, whereas MECP2-E2 levels are higher in neurons relative to astrocytes. In contrast, in males both MECP2 isoforms are more abundant in neurons than astrocytes. These differences were shown to correlate, to some extent, with the CpG methylation of the same REs in the promoter and exon 1 of each male and female cell lines. These represent interesting evidence for differential regulation of MECP2 isoforms between cell types in a sex-specific fashion and support DNA methylation as a prevalent mechanism for the differential regulation. Also, although MECP2 is expressed at much lower levels in astrocytes relative to neurons, it has been reported that MECP2-null astrocytes negatively affect neurons in a non-cell-autonomous fashion (67,68). The mechanisms regulating MECP2 expression in astrocytes are likely to be relevant for RTT and warrant further research. It has also been reported that MECP2-E1, but not E2, can be pathogenically induced by insults to the

immune system such as the one in the autoimmune disease multiple sclerosis (69) or during pain experiences (70). No mechanistic insight into how these insults can impact isoform-specific abundance has been offered so far.

Post-translational regulation has also been reported for MECP2 with the demonstration that the two protein isoforms have different stability in human neuroblastoma-derived neurons. The authors showed that MECP2-E1 is at least 2-fold more stable than MECP2-E2, contributing to the distinct levels of both protein isoforms in the brain (50). In contrast, MECP2-E1 has a higher turn-over rate in neuronal systems (57). Post-translational regulation of MECP2 was also demonstrated using a forward genetics approach. The authors describe HIPK1/2 (a protein kinase) and PP2A (a protein phosphatase) as druggable regulators with strong effects on endogenous MECP2 protein levels without a significant impact on MECP2 mRNA levels (71). The MECP2 sequence used in the reporter gene to detect protein degradation in the screen was MECP2-E2.

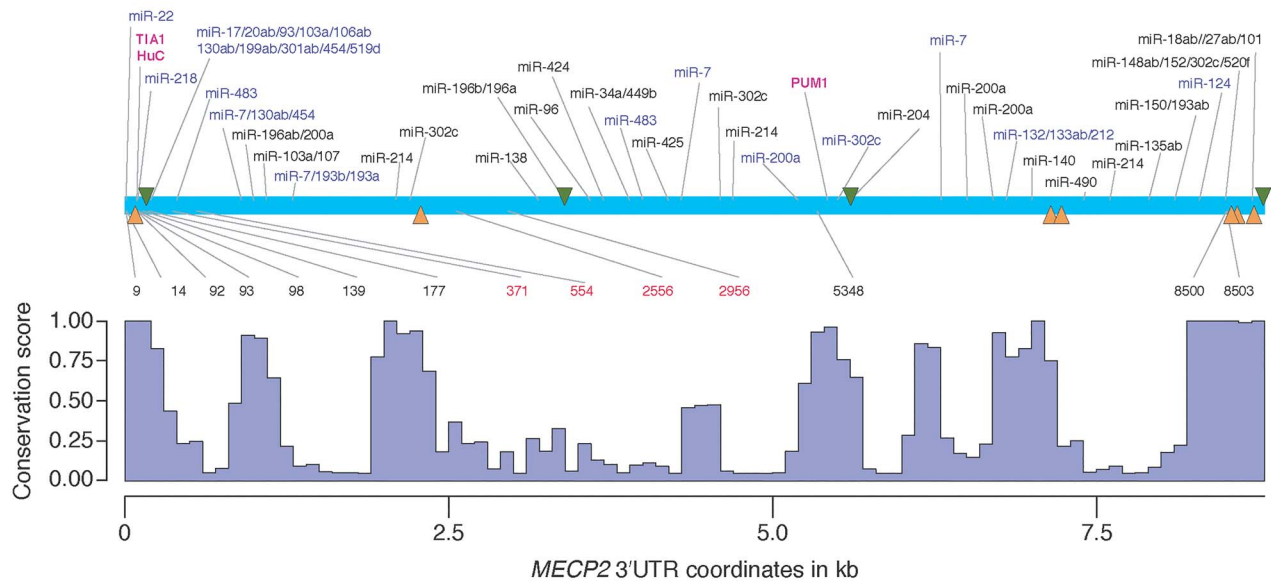
## Mapping MECP2 3'UTR Isoforms

MECP2 transcript maturation involves APA, generating transcripts with several 3'UTR lengths accumulating at variable abundances across cell types (8,72). Remarkably, the longest 3'UTR isoform is 8.6 kilobases (kb) long, positioning MECP2 in the group of genes with the longest 3'UTRs in the mammalian transcriptome. In addition, this long 8.6-kb 3'UTR is highly conserved across mammals (Fig. 2), indicating that positive selective pressure exists to preserve important RE.

In 1996, the first reported MECP2 northern blot of human tissue mRNA was performed using a probe common to both isoforms. Three transcripts with approximate sizes of 1.8, 7.5 and 9.5 kb, varying in abundance between tissues were detected, providing the first indication that MECP2 was expressed as multiple mRNA isoforms in humans (22). The set of detected transcripts was refined to 1.8, 5.0, 7.2, 10.1 kb after analysis of both fetal and adult human tissues (72) (Fig. 2). Later, exon 3 and 4 riboprobes detected three 3'UTR variant transcripts of 1.8, 5.4 and 10.2 kb in northern blot experiments, but an exon 2 riboprobe detected only two 3'UTR variants of 1.8, and 5.4 kb but not the long 10.2 kb transcript (59). This suggested that although both MECP2-E1 and -E2 isoforms were associated with the 1.8, 5.4-kb transcripts, there was a preferential association of MECP2-E1 with the longest 3'UTR isoform. Another report also indicated that the longest 3'UTR may be associated with MECP2-E1 isoform in pluripotent cells (8), although the mechanism coupling the AS and APA has not been explored. Recently, a high-throughput map of pA sites in human cells, including embryonic and neural stem cells, using 3'-end sequencing indicated the existence of 10 pA sites in MECP2 3'UTR further expanding the complexity of MECP2 post-transcriptional regulation (73).

## Expression of MECP2 3'UTR Isoforms

An important feature of the MECP2 3'UTR isoforms is that the longest 10.1-kb transcript is only enriched in the fetal brain and is replaced by the 5-kb transcript as the longest 3'UTR isoform in the adult brain (72). Although that drastic difference in 3'UTR length was anticipated to lead to distinct RNA degradation rates, the first measure of mRNA stability of both 1.8 and 10.1-kb transcripts showed that they had similar 3–4 h of half-life, within the accuracy of the technique used in Raji cells (30), which was contrasted with distinct half-life measurements made later (8).



**Figure 2.** Schematic representation of *MECP2* 3'UTR. Light blue rectangle: the longest *MECP2* 3'UTR isoform. Above are the approximate localizations of miRNA seed-sequences predicted for *MECP2* 3'UTR based on TargetScan using high stringency parameters (101). In blue and pink, miRNAs and RBPs demonstrated to regulate *MECP2* mRNA stability and/or translation efficiency, respectively. Below are approximate locations of mutations detected in RTT patients without alterations in the *MECP2* coding sequence. In red, mutations associated with decreased *MECP2* mRNA levels in patients (81). Green and orange arrowheads represent pA sites annotated by Coy et al. (72) and Wang et al. (73), respectively, the latter using 3'-READS assays from multiple cell types. Bottom graph: conservation degree of human *MECP2* 3'UTR compared to other 100 different vertebrate species shows areas of high similarities. Interestingly, most of these highly conserved sequences overlap with pA sites. The conservation score was calculated using the UCSC phastCons conservation score tool averaged in 100-nt bins (102).

*MECP2* expression in neural tissues follows two general patterns. An ontogenetic pattern (74), where *MECP2* appears first in the evolutionarily old structures like spinal cord and brain stem and only after in the more recent structures like cerebral cortex (75) and a laminar inside-out pattern within the brain regions, where Cajal–Retzius neurons become *MECP2* positive first (33). The net result of these patterns is the low *MECP2* level in the pre-natal stage, followed by its gradual increase at later stages (33). In the cerebral cortical layers III–V, *MECP2* abundance in the human brain gradually increases from 18-weeks post-conception and during the post-natal development. This gradual increase in *MECP2* levels coincides with the apparent age-dependent reduction in the abundance of the 10.1-kb 3'UTR isoform (76). However, if cells from the cerebral cortical layers III–V are split into *MECP2*<sub>high</sub> and *MECP2*<sub>low</sub> populations, the correlation between abundances of *MECP2* protein and the longest 3'UTR isoform disappears. The abundance of the longest 3'UTR isoform only decreases in the *MECP2*<sub>low</sub> population during the transition from fetal to juvenile brains, whereas for the *MECP2*<sub>high</sub> population the abundance of the longest 3'UTR isoform remains constant from infant to adult. In fact, in juvenile and adult samples, the relative abundance of the longest 3'UTR isoform is higher in the *MECP2*<sub>high</sub> than in *MECP2*<sub>low</sub> populations (77). This suggests that the correlation observed by Balmer et al., measuring the whole brain, was primarily driven by *MECP2*<sub>low</sub> population, the largest by cell number. As another observation, contrary to the cerebral cortex, *MECP2* abundance in the renal cortex remains constant during development (76).

Mouse *Mecp2* transcripts have lengths of 1.8, 5.4, 7.2 and 10.2 kb with abundance exhibiting spatiotemporal dependence. Starting from 12-weeks post-natal, the abundance of the 10.2-kb isoform in the cerebellum gradually increases, whereas the other isoforms show little variation. However, in the substantia nigra, basal ganglia, and occipital cortex, this increase is preceded by a

decrease of larger magnitude starting from E16.5. Furthermore, the 10.2-kb transcript isoform has minimal shifts in abundance in liver and lung. These data suggest the existence of a unique developmental program regulating *Mecp2* abundance profiles between visceral organs and brain (78).

### Regulation of 3'UTR Isoform Length by APA and RNA Destabilization

All these observed features of differential accumulation of 3'UTR isoforms in developing or adult brains are yet to be harmonized with a sufficiently strong mechanistic description. One proposed mechanism for the establishment of *MECP2* isoforms with different 3'UTRs is the tight pA regulation on premature mRNA, controlled by regulatory sequences surrounding pA sites. The pA site proximal to coding-sequence, corresponding to the short 3'UTR, has a canonical pA signal AAUAAA, conserved G-rich sequence and a suboptimal CstF binding element. In contrast, the distal pA site, corresponding to the long 3'UTR, has a non-canonical pA signal UAUAAA and an optimal CstF binding site, preceded by an upstream sequence element (USE). A comparison of the pA efficiencies between the proximal and distal pA sites in HeLa cells indicated that the distal pA site is at least 2-fold stronger than the proximal. Interestingly, a mutation in the conserved G-rich sequence causes a greater impact on pA site usage than mutation in the USE (79). The premature mRNA cleavage factor *NUDT21* (CFIm25) was also found to play a role in the regulation of *MECP2* 3'UTR isoforms. Elevated *NUDT21* increased the usage of the distal pA site of the *MECP2* transcripts, resulting in an enrichment of the inefficiently translated 10.1-kb long isoform in lymphoblastoid cells (80).

The longest 3'UTR of *MECP2* harbors a massive number of miRNA binding sites (81) (Fig. 2). In some instances, the *MECP2*-miRNA interaction forms a homeostatic feedback loop, and in

others forms a feedback loop between two distinct genes (82–85). In rat dorsal striatum, the overexpression of *miR-212* results in reduction of *Mecp2* protein levels. In addition, the overexpression of *miR-212* in rat PC12 cells leads to degradation of *Mecp2* long 3'UTR isoform (86). The non-conserved human-specific *miR-483-5p* was shown to regulate *MECP2* mRNA. Overexpression of *miR-483-5p* in hippocampal neurons of mouse with an additional copy of human *MECP2* rescued the mouse phenotypes to WT level. Notably, *miR-483-5p* is enriched in fetal brains then down-regulated postnatally, which might contribute to the dynamic regulation of short versus long *MECP2* 3'UTR isoforms. *miR-483-5p* also regulates multiple genes from the SIN3 and NCoR/SMRT complexes in which *MECP2* participates, implying a regulon type of relationship between these genes (87,88).

*MECP2* 3'UTR isoforms also undergo miRNA-mediated regulation in non-neural tissues (89–91). In human iPSCs, the pluripotent-specific miRNAs *miR-200a* and *miR-302c*, in cooperation with the RBP PUM1, actively degrade the 10.2-kb long mRNA, whereas the RBP TIA1 represses translation of the remaining *MECP2* transcripts (8). Mutations in *PUM1* are associated with adult-onset ataxia, and *PUM1* haploinsufficiency leads to developmental delay and seizures (92,93). *PUM1* overexpression in human iPSC-derived neurons decreased *MECP2* long 3'UTR isoform levels and caused dendrite complexity and soma size reductions (8). The long isoform is stabilized and accumulates in the human iPSC-derived cortical neurons, resulting in the upregulation of total *MECP2* mRNA levels. In addition, another RBP, HuC, outcompetes TIA1 for the same binding site increasing *MECP2* translation efficiency. The final result is that although total *MECP2* mRNA levels increase by ~2.5-fold, *MECP2* protein levels increase by at least 25-fold during *in vitro* neurodevelopment. In the same report, we showed that *MECP2* transcription rate remains unaltered in the neurodevelopment model; thus, the combination of miRNAs and RBPs is critical for the regulation of *MECP2* protein levels (8). We also observed that changes in abundance of the *MECP2* 3'UTR isoforms through differential transcript stabilization suggests that the global 3'UTR landscape can, in principle, be specified by mRNA stability rather than APA (8).

The introduction of CREB-induced *miR-132* into rat primary cortical neurons reduces *Mecp2* abundance without corresponding changes in mRNA levels, suggesting that *miR-132* regulates *MECP2* translation (83). In the *miR-132/212*-KO mice, the circadian rhythm of *Mecp2* is dysregulated in the circadian pacemaker neurons in the suprachiasmatic nucleus (94). The binding site for *miR-132* is specific for the long 3'UTR isoform, suggesting that its effect on translation rate is exclusive for the *MECP2* long 3'UTR, at least in cortical neurons. Changes in relative translation efficiency of the *MECP2* long versus short 3'UTR were independently corroborated by distinct high-throughput measurement of translation in hESC and hiPSC-derived neurons (9,95). Also, mRNA instability and inefficient translation were demonstrated for a gene therapy *MECP2* transgene cassette lacking the long 3'UTR and shown to be a barrier for recovery of *MECP2* protein levels to a physiological range in *MECP2* deficient neurons (96). These results highlight the breadth of post-transcriptional control on *MECP2* transcripts mediated by its multiple 3'UTR isoforms by the combination of miRNA and RBPs.

Mutations in the *MECP2* 3'UTR without significant alterations in the coding sequence have been reported in patients with neurological disorders (Fig. 2). Interestingly, some of these mutations have been correlated with reduction in total *MECP2* mRNA levels in blood cells of individuals with Autism Spectrum Disorder, suggesting that these mutations could affect *MECP2* regulation.

Some mutations reside in a region within 30 nt of the proximal pA site, potentially affecting its usage. At least one mutation has been associated with typical RTT, and two more were found in patients with atypical RTT (97,98). In addition, a deletion of almost the entire 3'UTR in one copy of duplicated *MECP2* was reported in a *MECP2* duplication syndrome patient with milder neurological symptoms, suggesting that *MECP2* duplication could be mitigated by the lack of its 3'UTR (99). Finally, a mutation found in *MECP2* 3'UTR was associated with reduced *MECP2* mRNA levels in peripheral blood cells. This mutation creates a de-novo binding-site for *miR-511* that was shown to suppress *MECP2* mRNA levels (100).

Altogether, these studies demonstrate that *MECP2* is under extensive post-transcriptional control and that its 3'UTR plays a pivotal role. Given the extreme length of *MECP2* 3'UTR, it is likely that more miRNAs and RBPs contribute to the developmental and homeostatic regulation of *MECP2*. This is an exciting aspect of *MECP2* biology and the discovery of new regulatory transacting factors overlapping with mutation sites in *MECP2* 3'UTR can provide important insights toward mechanisms of *MECP2* regulation and RTT.

## Conclusions

As a protein of such importance, and for which proper levels must be tightly controlled, some mechanisms that regulate overall *MECP2* gene expression and isoform-specific abundance remain uncharted. What is clear is that *MECP2* transcripts are under strong post-transcriptional regulation pressure, explaining the poor correlation between *MECP2* protein and transcript levels across multiple cell types in the brain of mice and humans. Several mechanisms for the fine regulation of *MECP2* protein levels have been identified including mRNA stability, translation efficiency control, and post-translational regulation. Because of the presence of a uORF, the *MECP2*-E2-encoding ORF is translated at a much lower efficiency than the *MECP2*-E1 counterpart. That feature likely contributes to the higher abundance of *MECP2*-E1 protein relative to E2. In addition, specific biophysical properties conferred by the different N-terminal of the two isoforms affecting protein stability might also contribute to the vast difference in abundance between the two protein isoforms.

Another important aspect of *MECP2* gene regulation refers to the APA, which generates transcripts with multiple 3'UTR lengths ranging from 0.1 to 8.6 kb. This positions *MECP2* mRNA among the genes with the longest 3'UTRs, further increasing its post-transcriptional pressure. Importantly, although the RefSeq database assigns the shortest (0.1 kb—RefSeq NM\_001110792) and longest (8.6 kb—RefSeq NM\_004992) 3'UTRs to the *MECP2*-E1 and *MECP2*-E2 isoforms, respectively, two independent reports actually suggest the opposite, at least for the cells studied (8,59).

At the moment, no strong evidence exists to explain the allocation of the multiple 3'UTR lengths with either *MECP2*-E1 or E2 isoforms. Therefore, a mechanism coupling the AS of exon 2 with the pA site usage is required. Long-reads sequencing, such as produced by Nanopore technology, generating reads covering exons 1 and 2 and the different 3'UTR lengths are required to ultimately discriminate whether there is a preferential pA site usage for each of the *MECP2* exon isoforms. Emerging evidence also correlates DNA methylation status of specific RE on the promoter and intron 1 of *MECP2* gene with the various abundances of *MECP2*-E1 and E2 isoforms in differentiating neural stem cells. However, it cannot be separated yet whether the changes in DNA methylation status or the sex of these cells affect splicing directly or indirectly through a post-transcriptional mechanism

acting on the stability of the mRNA isoforms that harbor the long 3'UTR.

Despite the existence of these transcriptional and intricate post-transcriptional mechanisms that fine-tune MECP2 protein levels, neurons cannot regularize MECP2 protein levels to compensate for the doubled gene dosage in MECP2 duplication syndrome, nor induce the MECP2-E2 protein isoform in RTT mutations affecting only the MECP2-E1. These upper and lower boundaries to fine regulation of MECP2 isoforms are overwhelmed in RTT and MECP2 duplication patients, and approaches to resolve these extraordinary expression differences could lead to improved therapeutic outcomes.

## Acknowledgements

We thank Rebecca S.F. Mok and Leah DeJong for comments on the manuscript. D.C.R. also thanks Maria and Sofia Rodrigues for inspiration at home, whereas this review was drafted during the 2020 coronavirus disease pandemic lockdown.

*Conflict of Interest statement.* None declared.

## Funding

Funding to J.E. was supported by the Canadian Institutes of Health Research (PJT-148746, PJT-168905) ERARE Team (ERT-161303); the Col. Harland Sanders Rett Syndrome Research Fund, University of Toronto; and the Ontario Brain Institute (POND Network).

## References

- Baralle, F.E. and Giudice, J. (2017) Alternative splicing as a regulator of development and tissue identity. *Nat. Rev. Mol. Cell Biol.*, **18**, 437–451.
- Mayr, C. (2016) Evolution and biological roles of alternative 3'UTRs. *Trends Cell Biol.*, **26**, 227–237.
- Ule, J. and Blencowe, B.J. (2019) Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol. Cell*, **76**, 329–345.
- Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
- Fabian, M.R. and Sonenberg, N. (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat. Struct. Mol. Biol.*, **19**, 586–593.
- Elkon, R., Ugalde, A.P. and Agami, R. (2013) Alternative cleavage and polyadenylation: extent, regulation and function. *Nat. Rev. Genet.*, **14**, 496–506.
- Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.*, **18**, 18–30.
- Rodrigues, D.C., Kim, D.-S., Yang, G., Zaslavsky, K., Ha, K.C.H., Mok, R.S.F., Ross, P.J., Zhao, M., Piekna, A., Wei, W. et al. (2016) MECP2 is post-transcriptionally regulated during human neurodevelopment by combinatorial action of RNA-binding proteins and miRNAs. *Cell Rep.*, **17**, 720–734.
- Rodrigues, D.C., Muftuev, M., Weatheritt, R.J., Djuric, U., Ha, K.C.H., Ross, P.J., Wei, W., Piekna, A., Sartori, M.A., Byres, L. et al. (2020) Shifts in ribosome engagement impact key gene sets in neurodevelopment and ubiquitination in Rett syndrome. *Cell Rep.*, **30**, 4179–4196.e11.
- Gruber, A.J. and Zolov, M. (2019) Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.*, **20**, 599–614.
- Sadek, J., Omer, A., Hall, D., Ashour, K. and Gallouzi, I.-E. (2019) Alternative polyadenylation and the stress response. *Wiley Interdiscip. Rev. RNA*, **10**, e1540.
- Miura, P., Shenker, S., Andreu-Agullo, C., Westholm, J.O. and Lai, E.C. (2013) Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res.*, **23**, 812–825.
- Tillotson, R. and Bird, A. (2019) The molecular basis of MeCP2 function in the brain. *J. Mol. Biol.*, **432**, 1602–1623.
- Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U. and Zoghbi, H.Y. (1999) Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.*, **23**, 185–188.
- Neul, J.L., Kaufmann, W.E., Glaze, D.G., Christodoulou, J., Clarke, A.J., Bahi-Buisson, N., Leonard, H., Bailey, M.E.S., Schanen, N.C., Zappella, M. et al. (2010) Rett syndrome: revised diagnostic criteria and nomenclature. *Ann. Neurol.*, **68**, 944–950.
- Chahrour, M. and Zoghbi, H.Y. (2007) The story of Rett syndrome: from clinic to neurobiology. *Neuron*, **56**, 422–437.
- Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell*, **69**, 905–914.
- Meehan, R.R., Lewis, J.D. and Bird, A.P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.*, **20**, 5085–5092.
- Nan, X., Meehan, R.R. and Bird, A. (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.*, **21**, 4886–4892.
- Quaderi, N.A., Meehan, R.R., Tate, P.H., Cross, S.H., Bird, A.P., Chatterjee, A., Herman, G.E. and Brown, S.D. (1994) Genetic and physical mapping of a gene encoding a methyl CpG binding protein, *Mecp2*, to the mouse X chromosome. *Genomics*, **22**, 648–651.
- Adler, D.A., Quaderi, N.A., Brown, S.D., Chapman, V.M., Moore, J., Tate, P. and Disteche, C.M. (1995) The X-linked methylated DNA binding protein, *Mecp2*, is subject to X inactivation in the mouse. *Mamm. Genome*, **6**, 491–492.
- D'Esposito, M., Quaderi, N.A., Ciccocicola, A., Bruni, P., Esposito, T., D'Urso, M. and Brown, S.D. (1996) Isolation, physical mapping, and northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2. *Mamm. Genome*, **7**, 533–535.
- Vilain, A., Apiou, F., Vogt, N., Dutrillaux, B. and Malfoy, B. (1996) Assignment of the gene for methyl-CpG-binding protein 2 (MECP2) to human chromosome band Xq28 by in situ hybridization. *Cytogenet. Cell Genet.*, **74**, 293–294.
- Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N. and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, **393**, 386–389.
- Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.*, **19**, 187–191.
- Chandler, S.P., Guschin, D., Landsberger, N. and Wolffe, A.P. (1999) The methyl-CpG binding transcriptional repressor MeCP2 stably associates with nucleosomal DNA. *Biochemistry*, **38**, 7008–7018.
- Weitzel, J.M., Buhrmester, H. and Strätling, W.H. (1997) Chicken MAR-binding protein ARBP is homologous to rat



- methyl-CpG-binding protein MeCP2. *Mol. Cell. Biol.*, **17**, 5656–5666.
28. Viana, M.C., Menezes, A.N., Moreira, M.A.M., Pissinatti, A. and Seuánez, H.N. (2015) MECP2, a gene associated with Rett syndrome in humans, shows conserved coding regions, independent Alu insertions, and a novel transcript across primate evolution. *BMC Genet.*, **16**, 77.
  29. Fahmi, M., Yasui, G., Seki, K., Katayama, S., Kaneko-Kawano, T., Inazu, T., Kubota, Y. and Ito, M. (2019, 2014) In Silico study of Rett syndrome treatment-related genes, MECP2, CDKL5, and FOXP1, by evolutionary classification and disordered region assessment. *Int. J. Mol. Sci.*, **15**, 23377–23388 20.
  30. Reichwald, K., Thiesen, J., Wiehe, T., Weitzel, J., Poustka, W.A., Rosenthal, A., Platzer, M., Strätling, W.H. and Kioschis, P. (2000) Comparative sequence analysis of the MECP2-locus in human and mouse reveals new transcribed regions. *Mamm. Genome*, **11**, 182–190.
  31. Guy, J., Hendrich, B., Holmes, M., Martin, J.E. and Bird, A. (2001) A mouse *Mecp2*-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.*, **27**, 322–326.
  32. Chen, R.Z., Akbarian, S., Tudor, M. and Jaenisch, R. (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nat. Genet.*, **27**, 327–331.
  33. Shahbazian, M., Young, J., Yuva-Paylor, L., Spencer, C., Antalffy, B., Noebels, J., Armstrong, D., Paylor, R. and Zoghbi, H. (2002) Mice with truncated MeCP2 recapitulate many Rett syndrome features and display hyperacetylation of histone H3. *Neuron*, **35**, 243–254.
  34. Tate, P., Skarnes, W. and Bird, A. (1996) The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nat. Genet.*, **12**, 205–208.
  35. Chen, J., Brunner, A.-D., Cogan, J.Z., Nuñez, J.K., Fields, A.P., Adamson, B., Itzhak, D.N., Li, J.Y., Mann, M., Leonetti, M.D. and Weissman, J.S. (2020) Pervasive functional translation of noncanonical human open reading frames. *Science*, **367**, 1140–1146.
  36. van Heesch, S., Witte, F., Schneider-Lunitz, V., Schulz, J.F., Adami, E., Faber, A.B., Kirchner, M., Maatz, H., Blachut, S., Sandmann, C.-L. et al. (2019) The translational landscape of the human heart. *Cell*, **178**, e29, 242–260.
  37. Kriaucionis, S. and Bird, A. (2004) The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucleic Acids Res.*, **32**, 1818–1823.
  38. Mnatzakanian, G.N., Lohi, H., Munteanu, I., Alfred, S.E., Yamada, T., MacLeod, P.J.M., Jones, J.R., Scherer, S.W., Schanen, N.C., Friez, M.J. et al. (2004) A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat. Genet.*, **36**, 339–341.
  39. Kudo, S., Nomura, Y., Segawa, M., Fujita, N., Nakao, M., Dragich, J., Schanen, C. and Tamura, M. (2001) Functional analyses of MeCP2 mutations associated with Rett syndrome using transient expression systems. *Brain Dev.*, **23**, S165–S173.
  40. Van den Veyver, I.B. and Zoghbi, H.Y. (2002) Genetic basis of Rett syndrome. *Ment. Retard. Dev. Disabil. Res. Rev.*, **8**, 82–86.
  41. Giacometti, E., Luikenhuis, S., Beard, C. and Jaenisch, R. (2007) Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. *Proc. Natl. Acad. Sci. U. S. A.*, **104**, 1931–1936.
  42. Collins, A.L., Levenson, J.M., Vilaythong, A.P., Richman, R., Armstrong, D.L., Noebels, J.L., David Sweatt, J. and Zoghbi, H.Y. (2004) Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum. Mol. Genet.*, **13**, 2679–2689.
  43. Jugloff, D.G.M., Jung, B.P., Purushotham, D., Logan, R. and Eubanks, J.H. (2005) Increased dendritic complexity and axonal length in cultured mouse cortical neurons overexpressing methyl-CpG-binding protein MeCP2. *Neurobiol. Dis.*, **19**, 18–27.
  44. Dastidar, S.G., Bardai, F.H., Ma, C., Price, V., Rawat, V., Verma, P., Narayanan, V. and D’Mello, S.R. (2012) Isoform-specific toxicity of *Mecp2* in postmitotic neurons: suppression of neurotoxicity by FoxG1. *J. Neurosci.*, **32**, 2846–2855.
  45. Ariani, F., Hayek, G., Rondinella, D., Artuso, R., Mencarelli, M.A., Spanhol-Rosseto, A., Pollazzon, M., Buoni, S., Spiga, O., Ricciardi, S. et al. (2008) FOXP1 is responsible for the congenital variant of Rett syndrome. *Am. J. Hum. Genet.*, **83**, 89–93.
  46. Luikenhuis, S., Giacometti, E., Beard, C.F. and Jaenisch, R. (2004) Expression of MeCP2 in postmitotic neurons rescues Rett syndrome in mice. *Proc. Natl. Acad. Sci. U. S. A.*, **101**, 6033–6038.
  47. Guy, J., Gan, J., Selfridge, J., Cobb, S. and Bird, A. (2007) Reversal of neurological defects in a mouse model of Rett syndrome. *Science*, **315**, 1143–1147.
  48. Kerr, B., Soto, C.J., Saez, M., Abrams, A., Walz, K. and Young, J.I. (2012) Transgenic complementation of MeCP2 deficiency: phenotypic rescue of *Mecp2*-null mice by isoform-specific transgenes. *Eur. J. Hum. Genet.*, **20**, 69–76.
  49. Itoh, M., Tahimic, C.G.T., Ide, S., Otsuki, A., Sasaoka, T., Noguchi, S., Oshimura, M., Goto, Y.-I. and Kurimasa, A. (2012) Methyl CpG-binding protein isoform MeCP2\_e2 is dispensable for Rett syndrome phenotypes but essential for embryo viability and placenta development. *J. Biol. Chem.*, **287**, 13859–13867.
  50. Yasui, D.H., Gonzales, M.L., Aflatooni, J.O., Crary, F.K., Hu, D.J., Gavino, B.J., Golub, M.S., Vincent, J.B., Carolyn Schanen, N., Olson, C.O. et al. (2014) Mice with an isoform-ablating *Mecp2* exon 1 mutation recapitulate the neurological deficits of Rett syndrome. *Hum. Mol. Genet.*, **23**, 2447–2458.
  51. Djuric, U., Cheung, A.Y.L., Zhang, W., Mok, R.S., Lai, W., Piekna, A., Hendry, J.A., Ross, P.J., Pasceri, P., Kim, D.-S. et al. (2015) MECP2e1 isoform mutation affects the form and function of neurons derived from Rett syndrome patient iPS cells. *Neurobiol. Dis.*, **76**, 37–45.
  52. Marchetto, M.C.N., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H. and Muotri, A.R. (2010) A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell*, **143**, 527–539.
  53. Cheung, A.Y.L., Horvath, L.M., Grafodatskaya, D., Pasceri, P., Weksberg, R., Hotta, A., Carrel, L. and Ellis, J. (2011) Isolation of MECP2-null Rett syndrome patient hiPS cells and isogenic controls through X-chromosome inactivation. *Hum. Mol. Genet.*, **20**, 2103–2115.
  54. Kim, K.-Y., Hysolli, E. and Park, I.-H. (2011) Neuronal maturation defect in induced pluripotent stem cells from patients with Rett syndrome. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 14169–14174.
  55. Li, Y., Wang, H., Muffat, J., Cheng, A.W., Orlando, D.A., Lovén, J., Kwok, S.-M., Feldman, D.A., Bateup, H.S., Gao, Q. et al. (2013) Global transcriptional and translational repression in human-embryonic-stem-cell-derived Rett syndrome neurons. *Cell Stem Cell*, **13**, 446–458.

56. Orlic-Milacic, M., Kaufman, L., Mikhailov, A., Cheung, A.Y.L., Mahmood, H., Ellis, J., Gianakopoulos, P.J., Minassian, B.A. and Vincent, J.B. (2014) Over-expression of either MECP2\_e1 or MECP2\_e2 in neuronally differentiated cells results in different patterns of gene expression. *PLoS One*, **9**, e91742.
57. Martínez de Paz, A., Khajavi, L., Martin, H., Claveria-Gimeno, R., Tom Dieck, S., Cheema, M.S., Sanchez-Mut, J.V., Moksa, M.M., Carles, A., Brodie, N.I. et al. (2019) MeCP2-E1 isoform is a dynamically expressed, weakly DNA-bound protein with different protein and DNA interactions compared to MeCP2-E2. *Epigenetics Chromatin*, **12**, 63.
58. Saxena, A., de Lagarde, D., Leonard, H., Williamson, S.L., Vasudevan, V., Christodoulou, J., Thompson, E., MacLeod, P. and Ravine, D. (2006) Lost in translation: translational interference from a recurrent mutation in exon 1 of MECP2. *J. Med. Genet.*, **43**, 470–477.
59. Dragich, J.M., Kim, Y.-H., Arnold, A.P. and Schanen, N.C. (2007) Differential distribution of the MeCP2 splice variants in the postnatal mouse brain. *J. Comp. Neurol.*, **501**, 526–542.
60. Liyanage, V.R.B., Zachariah, R.M. and Rastegar, M. (2013) Decitabine alters the expression of Mecp2 isoforms via dynamic DNA methylation at the Mecp2 regulatory elements in neural stem cells. *Mol Autism*, **4**, 46.
61. Adachi, M., Keefer, E.W. and Jones, F.S. (2005) A segment of the Mecp2 promoter is sufficient to drive expression in neurons. *Hum. Mol. Genet.*, **14**, 3709–3722.
62. Liu, J. and Francke, U. (2006) Identification of cis-regulatory elements for MECP2 expression. *Hum. Mol. Genet.*, **15**, 1769–1782.
63. Singh, J., Saxena, A., Christodoulou, J. and Ravine, D. (2008) MECP2 genomic structure and function: insights from ENCODE. *Nucleic Acids Res.*, **36**, 6035–6047.
64. Olson, C.O., Zachariah, R.M., Ezeonwuka, C.D., Liyanage, V.R.B. and Rastegar, M. (2014) Brain region-specific expression of MeCP2 isoforms correlates with DNA methylation within Mecp2 regulatory elements. *PLoS One*, **9**, e90645.
65. Liyanage, V.R.B., Zachariah, R.M., Davie, J.R. and Rastegar, M. (2015) Ethanol deregulates Mecp2/MeCP2 in differentiating neural stem cells via interplay between 5-methylcytosine and 5-hydroxymethylcytosine at the Mecp2 regulatory elements. *Exp. Neurol.*, **265**, 102–117.
66. Liyanage, V.R.B., Olson, C.O., Zachariah, R.M., Davie, J.R. and Rastegar, M. (2019, 2014) DNA methylation contributes to the differential expression levels of Mecp2 in male mice neurons and astrocytes. *Int. J. Mol. Sci.*, **15**, 23377–23388 20.
67. Liyo, D.T., Garg, S.K., Monaghan, C.E., Raber, J., Foust, K.D., Kaspar, B.K., Hirrlinger, P.G., Kirchhoff, F., Bissonnette, J.M., Ballas, N. et al. (2011) A role for glia in the progression of Rett's syndrome. *Nature*, **475**, 497–500.
68. Kahanovitch, U., Patterson, K.C., Hernandez, R. and Olsen, M.L. (2019, 2014) Glial dysfunction in MeCP2 deficiency models: implications for Rett syndrome. *Int. J. Mol. Sci.*, **15**, 23377–23388 20.
69. Khorshid Ahmad, T., Zhou, T., AlTaweel, K., Cortes, C., Lillico, R., Lakowski, T.M., Gozda, K. and Namaka, M.P. (2017, 2014) Experimental autoimmune encephalomyelitis (EAE)-induced elevated expression of the E1 isoform of methyl CpG binding protein 2 (MeCP2E1): implications in multiple sclerosis (MS)-induced neurological disability and associated myelin damage. *Int. J. Mol. Sci.*, **15**, 23377–23388 18.
70. Tochiki, K.K., Cunningham, J., Hunt, S.P. and Géranton, S.M. (2012) The expression of spinal methyl-CpG-binding protein 2, DNA methyltransferases and histone deacetylases is modulated in persistent pain states. *Mol. Pain*, **8**, 14.
71. Lombardi, L.M., Zaghlula, M., Sztainberg, Y., Baker, S.A., Klisch, T.J., Tang, A.A., Huang, E.J. and Zoghbi, H.Y. (2017) An RNA interference screen identifies druggable regulators of MeCP2 stability. *Sci. Transl. Med.*, **9**, eaaf7588.
72. Coy, J.F. (1999) A complex pattern of evolutionary conservation and alternative polyadenylation within the long 3'-untranslated region of the methyl-CpG-binding protein 2 gene (MeCP2) suggests a regulatory role in gene expression. *Hum. Mol. Genet.*, **8**, 1253–1262.
73. Wang, R., Nambiar, R., Zheng, D. and Tian, B. (2018) PolyA\_DB 3 catalogs cleavage and polyadenylation sites identified by deep sequencing in multiple genomes. *Nucleic Acids Res.*, **46**, D315–D319.
74. Cassel, S., Revel, M.-O., Kelche, C. and Zwiller, J. (2004) Expression of the methyl-CpG-binding protein MeCP2 in rat brain. An ontogenetic study. *Neurobiol. Dis.*, **15**, 206–211.
75. Mullaney, B.C., Johnston, M.V. and Blue, M.E. (2004) Developmental expression of methyl-CpG binding protein 2 is dynamically regulated in the rodent brain. *Neuroscience*, **123**, 939–949.
76. Balmer, D., Goldstine, J., Rao, Y.M. and LaSalle, J.M. (2003) Elevated methyl-CpG-binding protein 2 expression is acquired during postnatal human brain development and is correlated with alternative polyadenylation. *J. Mol. Med.*, **81**, 61–68.
77. Samaco, R.C., Nagarajan, R.P., Braunschweig, D. and LaSalle, J.M. (2004) Multiple pathways regulate MeCP2 expression in normal brain development and exhibit defects in autism-spectrum disorders. *Hum. Mol. Genet.*, **13**, 629–639.
78. Pelka, G.J., Watson, C.M., Christodoulou, J. and Tam, P.P.L. (2005) Distinct expression profiles of Mecp2 transcripts with different lengths of 3'UTR in the brain and visceral organs during mouse development. *Genomics*, **85**, 441–452.
79. Newnham, C.M., Hall-Pogar, T., Liang, S., Wu, J., Tian, B., Hu, J. and Lutz, C.S. (2010) Alternative polyadenylation of MeCP2: influence of cis-acting elements and trans-acting factors. *RNA Biol.*, **7**, 361–372.
80. Gennarino, V.A., Alcott, C.E., Chen, C.-A., Chaudhury, A., Gillentine, M.A., Rosenfeld, J.A., Parikh, S., Wheless, J.W., Roeder, E.R., Horovitz, D.D.G. et al. (2015) NUDT21-spanning CNVs lead to neuropsychiatric disease and altered MeCP2 abundance via alternative polyadenylation. *Elife*, **4**, 185.
81. McGowan, H. and Pang, Z.P. (2015) Regulatory functions and pathological relevance of the MECP2 3'UTR in the central nervous system. *Cell Regeneration*. doi: [10.1186/s13619-015-0023-x](https://doi.org/10.1186/s13619-015-0023-x).
82. Chen, Y., Shin, B.-C., Thamotharan, S. and Devaskar, S.U. (2014) Differential methylation of the micro-RNA 7b gene targets postnatal maturation of murine neuronal Mecp2 gene expression. *Dev. Neurobiol.*, **74**, 407–425.
83. Klein, M.E., Liyo, D.T., Ma, L., Impey, S., Mandel, G. and Goodman, R.H. (2007) Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nat. Neurosci.*, **10**, 1513–1514.
84. Lyu, J.-W., Yuan, B., Cheng, T.-L., Qiu, Z.-L. and Zhou, W.-H. (2016) Reciprocal regulation of autism-related genes MeCP2 and PTEN via microRNAs. *Sci. Rep.*, **6**, 20392.
85. Zhang, Y., Chen, M., Qiu, Z., Hu, K., McGee, W., Chen, X., Liu, J., Zhu, L. and Wu, J.Y. (2016) MiR-130a regulates neurite outgrowth and dendritic spine density by targeting MeCP2. *Protein Cell*, **7**, 489–500.
86. Im, H.-I., Hollander, J.A., Bali, P. and Kenny, P.J. (2010) MeCP2 controls BDNF expression and cocaine intake through

- homeostatic interactions with microRNA-212. *Nat. Neurosci.*, **13**, 1120–1127.
87. Han, K., Gennarino, V.A., Lee, Y., Pang, K., Hashimoto-Torii, K., Choufani, S., Raju, C.S., Oldham, M.C., Weksberg, R., Rakic, P. et al. (2013) Human-specific regulation of MeCP2 levels in fetal brains by microRNA miR-483-5p. *Genes Dev.*, **27**, 485–490.
  88. Keene, J.D. (2007) RNA regulons: coordination of post-transcriptional events. *Nat. Rev. Genet.*, **8**, 533–543.
  89. Zhao, H., Wen, G., Wen, G., Huang, Y., Yu, X., Chen, Q., Afzal, T.A., Luong, L.A., Zhu, J., Shu, Y. et al. (2015) MicroRNA-22 regulates smooth muscle cell differentiation from stem cells by targeting methyl CpG-binding protein 2. *Arterioscler. Thromb. Vasc. Biol.*, **35**, 918–929.
  90. Feng, Y., Huang, W., Wani, M., Yu, X. and Ashraf, M. (2014) Ischemic preconditioning potentiates the protective effect of stem cells through secretion of exosomes by targeting Mecp2 via miR-22. *PLoS One*, **9**, e88685.
  91. Kynast, K.L., Russe, O.Q., Möser, C.V., Geisslinger, G. and Niederberger, E. (2013) Modulation of central nervous system-specific microRNA-124a alters the inflammatory response in the formalin test in mice. *Pain*, **154**, 368–376.
  92. Gennarino, V.A., Singh, R.K., White, J.J., De Maio, A., Han, K., Kim, J.-Y., Jafar-Nejad, P., di Ronza, A., Kang, H., Sayegh, L.S. et al. (2015) Pumilio1 haploinsufficiency leads to SCA1-like neurodegeneration by increasing wild-type Ataxin1 levels. *Cell*, **160**, 1087–1098.
  93. Gennarino, V.A., Palmer, E.E., McDonell, L.M., Wang, L., Adamski, C.J., Koire, A., See, L., Chen, C.-A., Schaaf, C.P., Rosenfeld, J.A. et al. (2018) A mild PUM1 mutation is associated with adult-onset ataxia, whereas Haploinsufficiency causes developmental delay and seizures. *Cell*, **172**, 924–932.e11.
  94. Mendoza-Viveros, L., Chiang, C.-K., Ong, J.L.K., Hegazi, S., Cheng, A.H., Bouchard-Cannon, P., Fana, M., Lowden, C., Zhang, P., Bothorel, B. et al. (2017) miR-132/212 modulates seasonal adaptation and dendritic morphology of the central circadian clock. *Cell Rep.*, **19**, 505–520.
  95. Blair, J.D., Hockemeyer, D., Doudna, J.A., Bateup, H.S. and Floor, S.N. (2017) Widespread translational Remodeling during human neuronal differentiation. *Cell Rep.*, **21**, 2005–2016.
  96. Luoni, M., Giannelli, S., Indrigo, M.T., Niro, A., Massimino, L., Iannielli, A., Passeri, L., Russo, F., Morabito, G., Calamita, P. et al. (2020) Whole brain delivery of an instability-prone Mecp2 transgene improves behavioral and molecular pathological defects in mouse models of Rett syndrome. *Elife*, **9**, e52629.
  97. Fendri-Kriaa, N., Mkaouar-Rebai, E., Moalla, D., Belguith, N., Louhichi, N., Zemni, R., Slama, F., Triki, C., Fakhfakh, F. and Tunisian Network on Mental Retardation (2010) Mutational analysis of the MECP2 gene in Tunisian patients with Rett syndrome: a novel double mutation. *J. Child Neurol.*, **25**, 1042–1046.
  98. Santos, M., Yan, J., Temudo, T., Oliveira, G., Vieira, J.P., Fen, J., Sommer, S. and Maciel, P. (2008) Analysis of highly conserved regions of the 3'UTR of MECP2 gene in patients with clinical diagnosis of Rett syndrome and other disorders associated with mental retardation. *Dis. Markers*, **24**, 319–324.
  99. Hanchard, N.A., Carvalho, C.M., Bader, P., Thome, A., Omo-Griffith, L., Del Gaudio, D., Pehlivan, D., Fang, P., Schaaf, C.P., Ramocki, M.B. et al. (2012) A partial MECP2 duplication in a mildly affected adult male: a putative role for the 3' untranslated region in the MECP2 duplication phenotype. *BMC Med. Genet.*, **13**, 1–1.
  100. Tantra, M., Hammer, C., Kästner, A., Dahm, L., Begemann, M., Bodda, C., Hammerschmidt, K., Giegling, I., Stepniak, B., Castillo Venzor, A. et al. (2014) Mild expression differences of MECP2 influencing aggressive social behavior. *EMBO Mol. Med.* doi: [10.1002/emmm.201303744](https://doi.org/10.1002/emmm.201303744).
  101. Agarwal, V., Bell, G.W., Nam, J.-W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, 101.
  102. Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.