Published online 27 September 2008 *Nucleic Acids Research, 2008, Vol. 36, No. 19* ***6035–6047***
doi:10.1093/nar/gkn591

SURVEY AND SUMMARY

MECP2 genomic structure and function: insights from ENCODE

Jasmine Singh[1,2], Alka Saxena[1], John Christodoulou[3] and David Ravine[1,2,*]

[1]Western Australian Institute for Medical Research, Centre for Medical Research, University of Western Australia, [2]School of Medicine and Pharmacology, University of Western Australia and [3]Western Sydney Genetics Program, Children's Hospital at Westmead (Sydney), University of Sydney, Australia

Received April 27, 2008; Revised September 2, 2008; Accepted September 3, 2008

ABSTRACT

***MECP2*, a relatively small gene located in the human X chromosome, was initially described with three exons transcribing RNA from which the protein MeCP2 was translated. It is now known to have four exons from which two isoforms are translated; however, there is also evidence of additional functional genomic structures within *MECP2*, including exons potentially transcribing non-coding RNAs. Accompanying the recognition of a higher level of intricacy within *MECP2* has been a recent surge of knowledge about the structure and function of human genes more generally, to the extent that the definition of a gene is being revisited. It is timely now to review the published and novel functional elements within *MECP2*, which is proving to have a complexity far greater than was previously thought.**

INTRODUCTION

Anatomy of a typical gene

To perform their biological role, most well-characterized genes are expressed as protein, although it is now recognized that genes also transcribe a diverse array of non-coding RNAs (1). A typical eukaryotic gene, in addition to coding regions and protein domains, has a number of functional elements that play roles in transcription, post-transcriptional modification and translation. Exons, introns and 5′- and 3′-untranslated regions (UTRs) define the 'skeleton' of the gene, but a gene is also associated with sequences influencing splicing, regulatory elements such as promoters, enhancers, suppressors and CpG islands, at least one polyadenylation signal in the 3′-UTR, and transcription and translation initiation and termination sites. In addition to recognized functional elements, many genes also contain blocks of genomic sequence deep within introns that are highly conserved across species. The preservation of these sequence blocks strongly suggests that they have important functions that are yet to be defined.

Upgraded concept of a eukaryotic gene

Until recently, insights into the roles of the multiple components within a gene, particularly the many sequence-specific functional elements in non-coding areas, have been limited. This is now changing quickly with many new insights emerging from the pilot project phase of the Encyclopaedia Of DNA Elements (ENCODE) project (2). ENCODE aims to produce a functional annotation of the human genome, beginning with a targeted ~1% of the genome, including *MECP2*, for the pilot project. ENCODE is generating a vast quantity of new information about the transcriptional profiles of sequences within the targeted regions in different tissues and cell lines. The ENCODE output is also defining regions likely to contain functional regulatory elements based on features of chromatin structure, such as histone modifications and DNase I hypersensitivity sites, as well as binding sites of regulatory proteins. Measures of the extent of evolutionary conservation and secondary-structure predictions also help to highlight regions likely to be functionally relevant (2).

The ENCODE findings have provoked a fundamental reassessment of the concept of a gene (3,4). The term 'gene' was first coined by Johannsen in 1909 (5) to encapsulate Gregor Mendel's notion of a unit of heredity. Prompted particularly by studies of the segregation of mutations in *Drosophila*, which gave rise to the hypothesis that genes are structured as discrete elements organized in a linear array (6), the concept of a gene as a calculating unit transformed over time to the notion of a material unit with both structure and function. Recognition that gene mutations may cause inborn errors of metabolism strengthened the idea of a gene as a functional unit, or 'one gene, one enzyme' (7). The concept of a gene being a blueprint for a protein was strengthened in the 1950s when it was demonstrated that heritable information is stored in

*To whom correspondence should be addressed. Tel: +61 –8 9224 0349; Fax: +61 8 9224 0322; Email: ravine@waimr.uwa.edu.au

DNA and the 'central dogma' was developed to provide the crucial link between the physical structure of a gene within DNA and its functional protein products. In the last decade the view of genes as discrete, linearly arranged loci that generally encode a protein has evolved further with ongoing revelations regarding the structure, organization and functions of genes (3,4,8). Now ENCODE and others (9) have demonstrated that transcription is far more extensive than previously anticipated. Transcripts from the same and opposite strands often overlap, making it difficult to define discrete boundaries separating genes from one another. Furthermore, many novel transcribed regions have been detected in intergenic and intronic sequences. Recognition that many newly identified transcripts, including those from protein-coding loci, have tissue-specific expression but limited coding potential has spawned the now fast-growing interest in regulatory non-protein-coding RNAs (2).

Novel features identified within *MECP2* by the pilot phase of ENCODE help illustrate these emerging insights. As shown in Figure 1, 5′ rapid amplification of cDNA ends (5′-RACE) has extended the known 5′-terminal of *MECP2* further upstream, and also identified transcripts that connect downstream genes with the 3′-UTR of *MECP2*, or with genes upstream of *MECP2* (10; http://genome.ucsc.edu/). Novel transcribed regions identified by 5′-RACE appear to be expressed in a tissue-specific manner, which raises the spectre of a diversity of previously unrecognized tissue-specific functions. Such findings across the ENCODE regions have prompted suggestions that a gene should be defined not at the DNA level, but rather at the level of its functional products, whether RNA or protein (3,4).

MECP2 IN RELATION TO OTHER HUMAN GENES

There is great diversity in the size and complexity of human genes. The dystrophin gene, for example, has 79 exons spanning over 2.3 million base pairs (11). To begin examining the structural features of a gene using the example of dystrophin would indeed be a daunting task. By contrast, the *MECP2* gene with only four published exons (12) offers a more tractable opportunity to summarize the currently discernible functional components of a multi-exonic gene. In addition, as an X-linked gene subject to X inactivation (13–15), the impact of mutations within *MECP2* is generally easier to characterize than mutations in most autosomal genes. It is becoming increasingly evident, however, that *MECP2* has a higher level of complexity than originally believed. *MECP2* has some remarkable features, most notably the extent of its 3′-UTR (12,16), which is one of the longest known in the human genome. It also contains several polyadenylation signals that give rise to transcripts of different lengths with different expression patterns (12,14,16,17). The second intron of *MECP2*, traversing almost 60 000 nucleotides, is atypically long and contains several regions with notable evolutionary conservation (12; http://genome.ucsc.edu/). Alternative transcripts in addition to those reported in the literature have been identified by the ENCODE

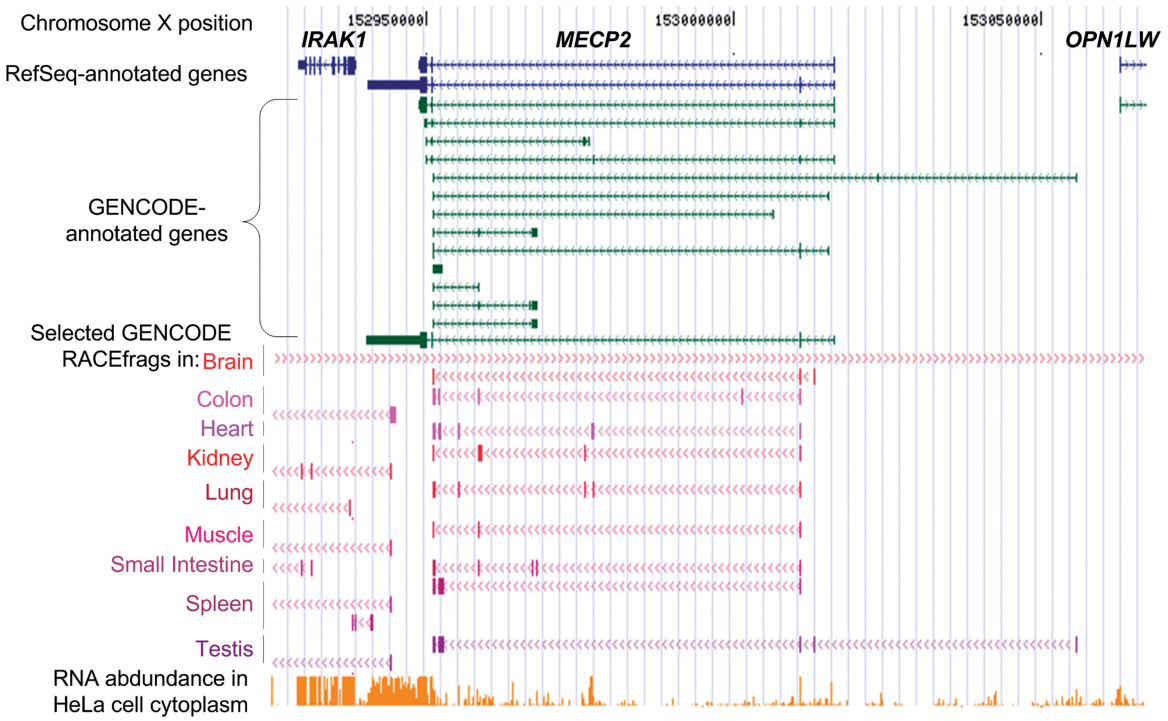

Figure 1. ENCODE data reveals increased transcriptional complexity within *MECP2*. This screenshot from the March 2006 assembly of the UCSC Genome Browser (http://genome.ucsc.edu/) compares the previously annotated *MECP2* gene structure (RefSeq-annotated genes) with the GENCODE annotation (v3.1, March 2007). The tracks 'GENCODE RACEfrags' show the mapped locations of 5′-RACE products from a range of tissues (83). Also displayed is the Affymetrix data representing the genomic source and relative abundance of RNA transcripts detected in the cytoplasm of HeLa cells.

Consortium and are now being deposited in genome databases.

The emerging details from ENCODE are yielding more insight into the function and regulation of *MECP2* expression, as well as pinpointing genomic sites where previously unrecognized disease-causing mutations could occur. For these reasons it is timely to conduct a detailed reassessment of the *MECP2* genomic sequence. We integrate here the data made available by the ENCODE project with the previously published features of the *MECP2* gene. Insights gathered from this relatively small human gene have the potential to yield lessons that may have relevance to larger, more complex genes.

BIOLOGICAL FUNCTIONS OF MECP2

MeCP2 was first identified as a protein capable of binding to methylated DNA (18), although it was soon noted to have a weaker affinity for unmethylated DNA as well (19,20). MeCP2 differs from its other methyl-DNA-binding family members (21,22), by its specific and high-affinity binding to symmetrically methylated CpG dinucleotides (18). More recently, it has been recognized that MeCP2 requires an A/T run of four or more bases adjacent to the methyl-CpG for efficient DNA binding (23). As DNA methylation was known to be an important mechanism for regulating gene expression (24) and MeCP2 had been shown to localize to heterochromatin (18,25), it seemed likely, and was soon confirmed, that MeCP2 could function as a transcriptional repressor (26). Although initially described as a global transcriptional repressor, a limited number of targets of MeCP2-mediated regulation have now been identified (Table 1). The most well-established mechanism by which MeCP2 is able to repress transcription involves the recruitment of co-repressor complexes containing histone deacetylases that are able to alter chromatin structure (27–29). However, as trichostatin A, a histone deacetylase inhibitor, does not entirely relieve the transcriptional repression conferred by MeCP2 (27,28), it is evident that MeCP2 must also be able to regulate transcription independently of HDACs, and several such HDAC-independent mechanisms have been proposed. Kaludov and Wolffe (30) demonstrated that MeCP2 is still able to repress transcription *in vitro* in the absence of histones by interacting with transcription factor IIB (TFIIB) to prevent assembly of the pre-initiation complex. MeCP2 may also interact with other chromatin-modifying enzymes or complexes, including histone methyltransferase (31), Brm and SWI/SNF (32) and ATRX (33). MeCP2 may also be able to directly influence chromatin structure. MeCP2 binds to methylated linker DNA in nucleosome arrays (34) and induces chromatin compaction *in vitro* by mediating nucleosome–nucleosome interactions. Interestingly, MeCP2 also mediates conformational changes in naked DNA *in vitro*, including cross-linking of two or more DNA molecules (35–37). Consistent with its role as a chromatin-binding protein, MeCP2 was shown to be capable of binding to four-way DNA junctions, a property that is shared by several other proteins including HP1, HMGB1 and the SWI/SNF complex (38). The interaction of MeCP2 with chromatin is transient and dynamic, as demonstrated by rapid recovery of fluorescence after photobleaching (39). It is also proposed that MeCP2 may be involved in the formation of a chromatin loop at the *Dlx5* and *Dlx6* genes that results in silencing of these genes. Horike *et al.* (40) identified a MeCP2-binding site near *Dlx5* and *Dlx6*, and observed that transcript levels from each of these genes were elevated 2-fold in the frontal cortex of *Mecp2*-null mice when compared to wild type. They went on to perform chromatin conformation capture (3C) assays and concluded that normal silencing of *Dlx5* and *Dlx6* was the result of a chromatin loop that was not detected in *Mecp2*-null cells, suggesting that MeCP2 mediated formation of this loop. While MeCP2 may be able to form such structures, at least *in vitro*, Schule *et al.* (41) have been unable to detect a reproducible or significant increase in expression of *Dlx5* and *Dlx6* in the frontal cortex of *Mecp2*-null mice. As a result, the role of MeCP2 in the silencing of these genes is now in doubt, although MeCP2-binding sites within the *DLX5*/*DLX6* locus have been confirmed by independent chromatin-immunoprecipitation (ChIP) assays (42).

A recent study of MeCP2-binding sites within a human neuronal cell line has also provided evidence that MeCP2

Table 1. MeCP2 target genes

Gene	Function	References
xHairy2a	Neuronal repressor	(56)
BDNF/*Bdnf*	Long term potentiation	(81,90)
IGF2	Cell proliferation	(91)
MPP1	Signal transduction	(91)
UBE3A/*Ube3a*	Proteolysis	(92,93)
GABRB3/*Gabrb3*	GABA receptor subunit	(92)
DLX5/*Dlx5*	Transcription factor	(40)
Dlx6	Transcription factor	(23,40)
Fkbp5	Hormone signalling	(94)
Sgk1	Ion channel activation	(94)
ID1/*Id1, ID2*/*Id2, ID3*/*Id3*	Transcriptional regulation	(95)
Uqcrc1	Mitochondrial respiratory complex subunit	(59)
Crh	Anxiety and stress response	(96)
IGFBP3/*Igfbp3*	Hormone signalling	(97)
FXYD1/*Fxyd1*	Na^+/K^+-ATPase activity	(98,99)
Reln	Cell signalling	(99)
Gtl2/*Meg3*	Non-coding RNA	(99)
JUNB	Early response gene, Oncogene	(42)
RNASEH2A	Ribonucleotide cleavage from RNA–DNA complex	(42)
Sst	Hormone signalling	(43)
Oprk1	Opioid receptor	(43)
Gamt	Methyltransferase	(43)
Gprin1	Neurite formation	(43)
Mef2c	Myogenesis	(43)
A2bp1	Splicing	(43)
Creb1	Transcriptional co-activator	(43)

The table includes genes shown to bind MeCP2/xMeCP2 directly and with expression that is altered in the presence of dysfunctional MeCP2/xMeCP2. It should be noted, however, that independent studies have not confirmed MeCP2-mediated regulation of *DLX5*/*Dlx5* and *Dlx6* (41) and *Ube3a* (100). Furthermore, there are some inconsistencies in the results of the groups identifying *UBE3A*/*Ube3a* (92,93) and *FXYD1*/*Fxyd1* (98,99) as MeCP2 targets. *Sst*, *Oprk1*, *Gamt*, *Gprin1* and *Creb1* are reported to be activated by MeCP2 (43).

function extends beyond transcriptional repression mediated by binding to methylated promoters (42). In this analysis, ChIP assays revealed that CpG islands and transcriptionally silent promoters constitute only a minority of MeCP2-binding sites. Instead, a large number of MeCP2-binding sites are intergenic, consistent with a role for MeCP2 in long-range chromatin modifications. Another interesting observation from this analysis was a correlation between MeCP2 and RNA polymerase II binding sites, which suggests that MeCP2 often binds to transcriptionally active promoters. In this setting, MeCP2 may be functioning to downregulate rather than completely silence gene expression from these promoters, as appears to be the case for *JUNB* (42). Alternatively, Chahrour *et al.* (43) have recently proposed that MeCP2 can function as a transcriptional activator. This hypothesis arose from observations that a large number of genes were downregulated in the hypothalami of *Mecp2*-null mice but upregulated in transgenic *Mecp2*-overexpressing mice relative to wild-type controls, a result that contradicts the idea of MeCP2 as a transcriptional repressor. Chahrour *et al.* (43) went on to confirm binding of MeCP2 to the promoter regions of four of these activated genes. Furthermore, they showed that MeCP2 co-immunoprecipitates the co-activator CREB-1 and that both MeCP2 and CREB-1 binding were enriched at the *Sst* promoter, one of the presumed targets of MeCP2-mediated activation. Thus, it appears that MeCP2 may be more correctly referred to as a transcriptional modulator than a transcriptional repressor. If MeCP2 is indeed able to activate as well as repress transcription, it will be interesting to identify which regions of the MeCP2 protein are involved in the recruitment of transcriptional co-activators and how MeCP2 is able to specifically recruit co-repressors in some instances and co-activators in others.

Beyond a role in transcriptional regulation, MeCP2 has also been implicated in alternative splicing as part of an RNA-protein complex (44,45), and genome-wide alterations in RNA splicing have been observed in mice with MeCP2 dysfunction (45). The recent delineation of at least six biochemically distinct domains within MeCP2 (Figure 2), along with the recognition that MeCP2 has a disordered tertiary structure potentially allowing different combinations of domains to act together for specific functions (46), provide a strengthened framework for understanding the various structural elements within this multifunctional protein.

MECP2 DYSFUNCTION

MeCP2 shows complex spatial and temporal patterns of distribution, and both up- and down-regulation of MeCP2 may have disastrous consequences. Mutations in *MECP2* resulting in loss-of-function are implicated in several neurological disorders including Rett Syndrome (RTT), which is a severe neurodevelopmental disorder

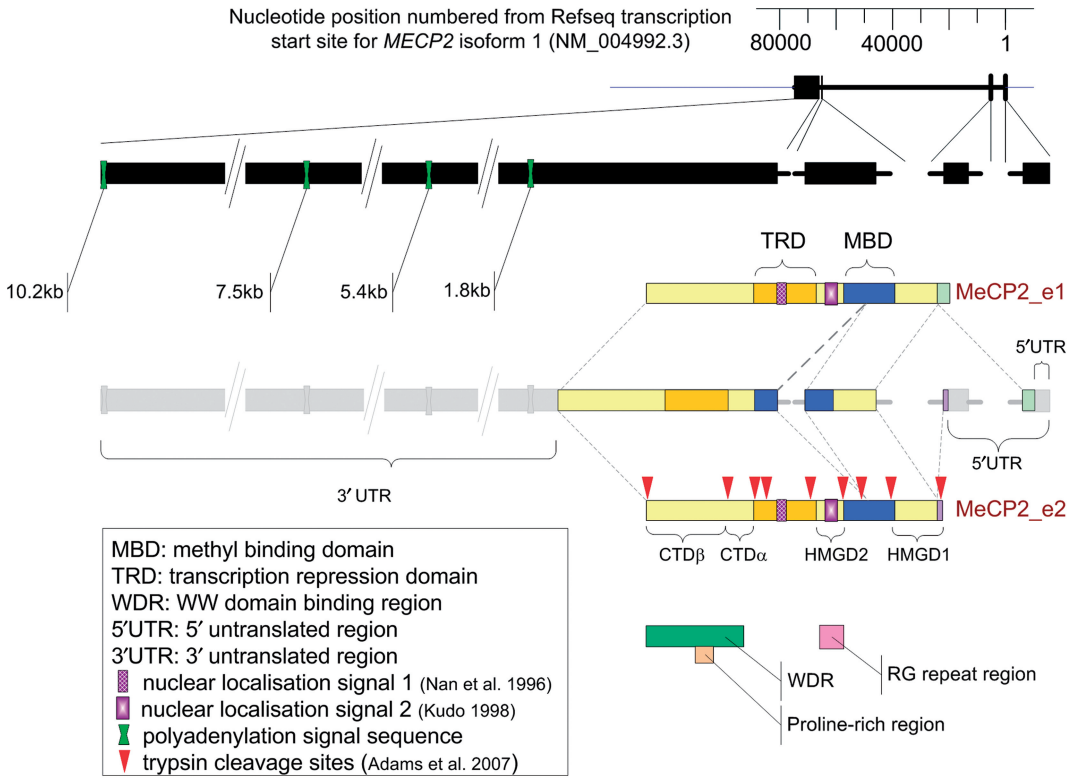

Figure 2. *MECP2* coding regions and functional protein domains. The four annotated exons of the *MECP2* gene are magnified and their coding portions are shown in colour, with the non-coding portions represented in grey. The MeCP2_e1 and MeCP2_e2 isoforms are also displayed. The 3′-UTR, with four alternative polyadenylation signals, and the 5′-UTR of each transcript are also shown. The positions of the well-characterized MBD, TRD and nuclear localization signals of the MeCP2 protein as well as the more-recently identified WDR and RG repeat region are as labelled. The distinct tertiary structures and functional domains defined by trypsin cleavage sites (46) are also indicated.

predominantly affecting girls. RTT is characterized typically by a period of seemingly normal development followed by the onset of decelerating developmental milestone achievements and brain growth, loss of purposeful hand movements and language skills, seizures and respiratory dysfunction (47). Mutations in *MECP2* are also implicated in a minority of cases presenting clinically as Angelman syndrome, autistic spectrum disorder, nonspecific mental retardation and neonatal encephalopathy. Decreased MeCP2 expression, in the absence of coding region mutations, has also been observed in patients with RTT and other neurodevelopmental disorders such as autism, Angelman syndrome, Prader–Willi syndrome, Down syndrome and attention-deficit hyperactivity disorder (48,49).

More recently it has been established that an increase in *MECP2* dosage is also highly detrimental to normal central nervous system function. *Mecp2* over-expression in mice is associated with seizures, hypoactivity and ataxia (50). In humans, duplications of the *MECP2* gene are now also recognized among males with a phenotype of severe mental retardation, hypotonia, recurrent respiratory infections, absence of speech development, seizures and spasticity (51).

Consistent with the brain-specific phenotype associated with *MECP2* dysfunction, MeCP2 is hypothesized to regulate genes involved in neuronal maturation and in maintaining synaptic plasticity. Several neuronal-related genes have been observed to be dysregulated in the brains of *Mecp2*-mutant mice (43,52,53). *Bdnf*, which has been identified as a direct target of MeCP2 (Table 1), is particularly interesting given its role in neuronal development and plasticity, as well as the phenotypic overlap between mice lacking *Bdnf* or *Mecp2* (54,55). In *Xenopus*, xMeCP2 regulates *xHairy2a* expression and is involved in cell-fate decisions. *xHairy2a* is one of a family of genes involved in repressing neuronal differentiation, and xMeCP2-null embryos show upregulated *xHairy2a* and fewer differentiated neurons. It is hypothesized that xMeCP2 may help reduce *xHairy2a* expression by recruiting the Sin3A-containing SMRT complex, until activation of the *Notch-Delta* signalling pathway causes SMRT-bound xMeCP2 to dissociate from the *xHairy2a* promoter region (56). By contrast, MeCP2 in mice does not appear to influence early cell-fate decisions (57), although it does still recruit co-repressors and co-activators to genes relevant to neuronal function (43,55).

THE MECP2 PROTEIN

Two MeCP2 isoforms are recognized, MeCP2_e1 (e1) and MeCP2_e2 (e2). The e1 isoform is 498 amino acids in length and encoded by exons 1, 3 and 4, whereas the e2 isoform is 486 amino acids long and encoded by exons 2, 3 and 4 [Figure 2; refs. (58,59)]. Upon identifying the MeCP2_e1 isoform, Mnatzakanian *et al.* (58) observed that it may be of greater clinical relevance than the e2 isoform, as it is the sole isoform in non-mammalian vertebrates and its transcript is 10 times more abundant in the human adult brain than that of e2. Furthermore, while no RTT-associated mutations have been characterized within exon 2, which is unique to e2, Mnatzakanian *et al.* (58) were the first to report pathogenic mutations in exon 1, which encodes the unique N-terminal to e1. It was initially assumed that such exon 1 mutations had no effect on the e2 protein isoform, as e2 transcript levels were observed to be unchanged (58). However, it has since been demonstrated that exon 1 mutations may be associated with a reduction in e2 protein levels despite the presence of e2 transcripts, possibly through translational interference (60). Thus the question remains of whether MeCP2_e2 is truly redundant, or whether it may have some unique function that, when disrupted, contributes to the RTT phenotype. In support of the latter possibility, Dragich *et al.* (61) have demonstrated that, in addition to the two isoforms having amino termini with unique physiochemical properties, the mouse *Mecp2e1* and *Mecp2e2* transcripts show distinct expression patterns within different brain regions and developmental stages, and show different preferences for alternative polyadenylation sites within the 3′-UTR (61).

The domain responsible for specific binding to methylated CpGs (methyl-binding domain or MBD) has been localized to an 85-amino-acid region encoded within exons 3 and 4, which are present in both isoforms of MeCP2 [Figure 2; ref. (20)]. The MBD has been shown to be essential for binding of MeCP2 to heterochromatin (25), as well as to unmethylated four-way DNA junctions (38).

The residues on the N-terminal side of the MBD have an amino acid composition very similar to the high mobility group (HMG) proteins, which are examples of proteins in which absence of secondary and tertiary structures permits different combinations of domains to perform different functions. On the basis of this structure, Adams *et al.* (46) refer to this N-terminal domain as HMGD1. The N-terminal region has recently been shown to interact with the repressive chromatin regulator heterochromatin protein 1 (HP1). This is consistent with observations of simultaneously increased concentrations of MeCP2 and HP1α and HP1γ within heterochromatin during myogenic differentiation, however MeCP2 also binds to HP1β which does not associate with heterochromatin (62). The N-terminal portion, including HMGD1 and MBD, as well as the C-terminal region have also been shown to have an association with histone H3 methyltransferase activity (31).

The 104-amino-acid transcription repression domain (TRD) encoded within exon 4 (26) interacts with histone deacetylases (HDACs) and transcriptional co-repressors Sin3A (27,28), c-Ski and N-CoR (29). HDAC1 and HDAC2 are histone deacetylases that combine with transcriptional repressors such as mSin3A to form a co-repressor complex. The TRD also directly interferes with assembly of the pre-initiation complex through interactions with transcription factor IIB (TFIIB), a component of the basal transcriptional machinery (30). The maintenance DNA methyltransferase, Dnmt1, also associates directly with MeCP2 at the TRD, and immunoprecipitated MeCP2-Dnmt1 complexes show DNA methyltransferase activity to hemimethylated DNA, implicating MeCP2 in the maintenance of methylation patterns

during DNA replication (63). The TRD, as well as neighbouring sequences, are necessary for the interaction of MeCP2 with Y box binding protein 1 (YB1). YB1, which is an RNA-binding protein, has a number of functions, including a role in alternative splicing and MeCP2, when bound to YB1, influences splice site selection (45). More recently, it has been demonstrated that the TRD is also capable of interacting with DNA *in vitro* (46).

One of two identified nuclear localization signals (NLS) is embedded within the TRD [see Figure 2; ref. (25)]. The other NLS, located in HMGD2, between the MBD and TRD, was identified by expressing human MeCP2 in *Drosophila melanogaster* cells, and has not been shown to be functional in mammalian cells (64).

Jeffery and Nakielny (44) identified an RG (arginine-glycine) repeat region in HMGD2 between amino acids 160 and 200 (Figure 2). RG repeat regions are known to mediate RNA-protein interactions, and MeCP2 was shown to bind to mRNA and double-stranded siRNA independent of its MBD, suggesting that MeCP2 may have activity as part of a selective RNA–protein complex. Further evidence in support of this claim comes from *in vivo* experiments documenting an RNA-mediated interaction between MeCP2 and YB1 (45).

In vivo, MeCP2 interacts with chromatin via one or more chromatin interaction domains (35,36,38,65). A chromatin-interacting domain is likely to be positioned within the C-terminal, as truncated MeCP2 mutants also show impaired binding to chromatin (36,65) and induce different changes in chromatin architecture *in vitro* compared to the wild-type MeCP2 (36). The tertiary structural regions predicted by protease digestion mapping suggest that there are two biochemically distinct C-terminal regions, CTDα and CTDβ [Figure 2; ref. (46)]. CTDα may contribute to the recognition of methylated DNA in chromatin (36), while CTDβ encodes the WW-domain-binding region [WDR; ref. (66)]. WW domains are characterized by two tryptophan (W) residues separated by 20–22 amino acids and facilitate protein–protein interactions by recognizing proline-rich motifs (67). WW domains are found in proteins such as the neural Wiskott–Aldrich syndrome protein (N-WASP), and formin-binding protein 11 (FBP11). N-WASP is involved in actin polymerization and cell locomotion, while FBP11 is a splicing-factor involved in mouse limb development and also known to interact with the Huntingtin protein implicated in Huntington disease (66,68). FBP11 has also been shown to interact with MeCP2, and although the precise role of the MeCP2 WDR *in vivo* remains unclear, mutations reducing WW-binding activity have been identified in cases of RTT as well as mild to moderate mental retardation (66).

MECP2 EXPRESSION

Expression of MeCP2 is high in the brain, specifically in neurones rather than glia, and increases progressively during mouse embryonic and human post-natal brain development (69–71). Immunohistochemical staining has revealed that the timing of increasing MeCP2 expression is correlated with the pattern of maturation of different brain regions and cell types (69), suggesting that MeCP2 may play a role in neuronal maturation (48,52,69,71,72). Such a function is consistent with well-recognized features of RTT such as apparently normal early development and reduced brain size with decreased dendritic branching (52,69). On the other hand, the recent finding by Guy *et al*. (73) that in the *Mecp2*-null mouse, at least, the neurological abnormalities appear to be reversible, strengthens that MeCP2 plays a role in the maintenance of neuronal plasticity. The devastating neurofunctional consequences of inappropriate up- and down-regulation of MeCP2 reveal that a more detailed understanding of the regulatory elements influencing MeCP2 expression could yield more insights into the pathogenic processes prompted by dysregulation of MeCP2 expression.

REGULATORY ELEMENTS

Liu and Francke (74) recently examined conserved elements throughout the *MECP2* gene and its neighbouring regions for regulatory activity, and identified four enhancer and two silencer elements, as well as a conserved fragment spanning 1080 nt immediately upstream of exon 1, which incorporates the *MECP2* core promoter and at least one positive and two negative regulatory elements (Figure 3). The identified regulatory elements show cell-specific activity differences. All four enhancers and two silencers were able to interact with nuclear proteins in gel-shift assays, and the enhancer elements contain predicted binding sites for brain-specific transcription factors. Furthermore, three of the enhancers and both of the silencers were shown to interact in *cis* with the core promoter, providing further support for their potential role in transcriptional regulation (74).

Promoter and 5′-UTR

Contrary to the classical description of 'TATA-dependent' core promoters, most core promoters are now believed to be characterized by a high GC content and multiple transcription start sites (TSSs) clustered together over a stretch of 50–100 nt (2,75,76). The *MECP2* core promoter (Figure 3), as identified by Liu and Francke (74), is embedded within a CpG island (12), and a region rich with regulatory factor binding sites (http://genome.ucsc.edu/). Methylation levels within the promoter region have been shown to correlate inversely with MeCP2 expression, and significantly increased methylation has been observed in male autism cases relative to controls (49). Cap analysis gene expression (CAGE) has revealed a cluster of predicted TSSs upstream of the coding region of exon 1, as well as several Sp1 binding sites. TSSs or TSS clusters, when considered collectively, show distinct patterns of histone modifications, chromatin accessibility and regulatory factor binding (2,77). Consistent with these results, ChIP data from the Ludwig Institute demonstrates that the regions upstream and downstream of the *MECP2* TSS cluster are enriched for RNA polymerase II, TAF1 and the activating histone modifications histone 3 lysine 4 dimethylation (H3K4me2) and histone 3 lysine 4

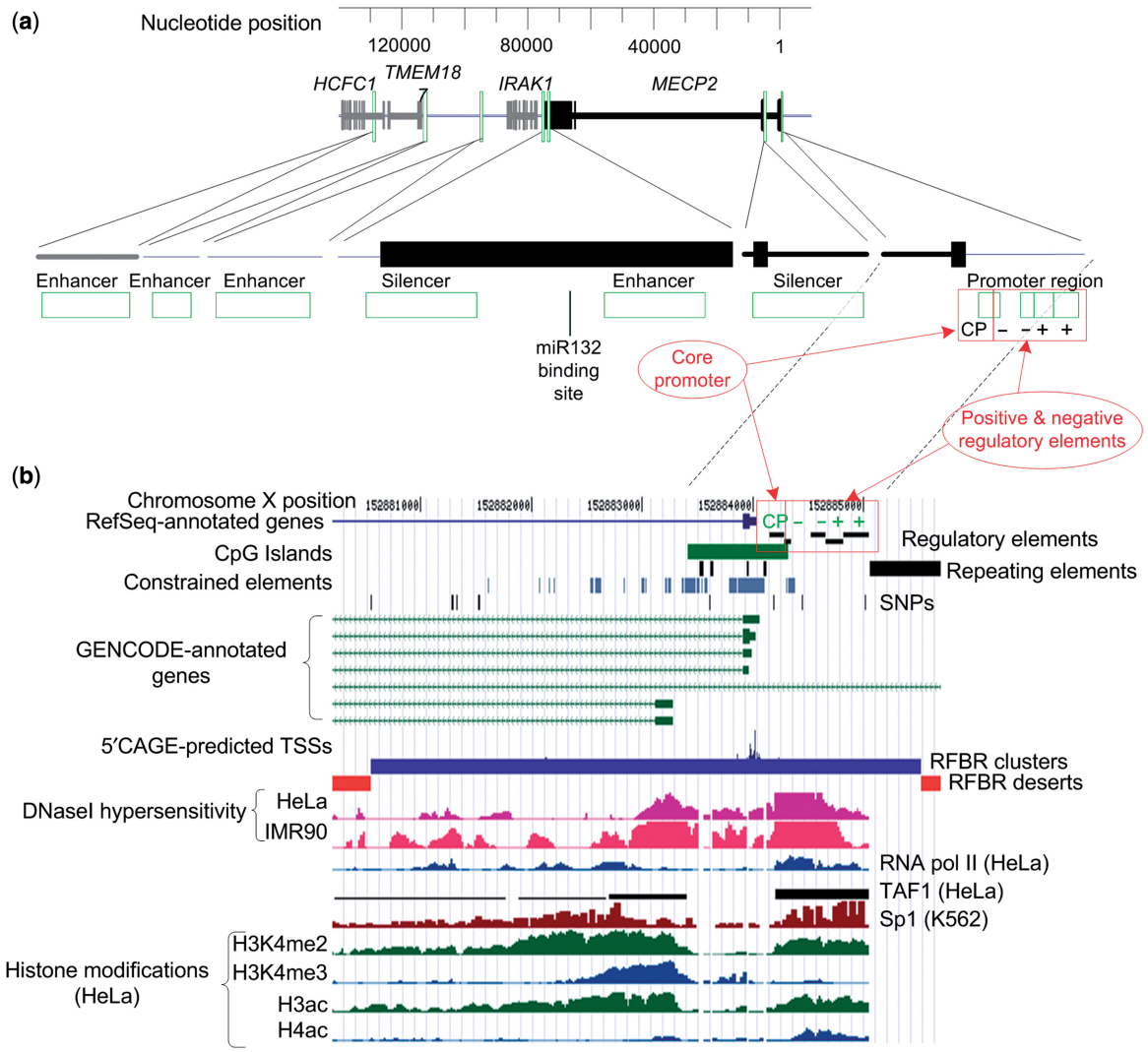

Figure 3. *MECP2* regulatory elements. (**a**) *MECP2* regulatory elements include the *cis*-regulatory elements identified by Liu and Francke (74) and the TargetScanS-identified binding site for miR-132, which has recently been shown to repress *MECP2* translation (82). (**b**) A UCSC Genome Browser screenshot (http://genome.ucsc.edu/, May 2004 assembly) of a 5500-bp region encompassing the promoter and exon 1 is also included. This displays the *cis*-regulatory elements identified in (a) above; CpG islands; repeating elements identified by RepeatMasker; constrained elements defined by ENCODE; single nucleotide polymorphisms (SNPs) from dbSNP build 126 (March 2006 assembly); GENCODE gene annotations (v3.1 March 2007); RIKEN CAGE-predicted transcription start sites (TSSs) on the minus strand; and Yale Regulatory Factor Binding Region (RFBRs) clusters and deserts. DNase I hypersensitivity sites are shown for HeLa (epithelioid carcinoma) and IMR90 (fibroblast) cells, as identified by Duke/NHGRI. The binding of RNA polymerase II (RNA pol II) and TATA-binding protein-associated factor 1 (TAF1); as well as the locations of histone modifications, was determined by ChIP studies performed by Ludwig Institute/UCSD. The Sp1-binding sites were identified by Stanford in K562 (chronic myeloid leukaemia) cells.

trimethylation (H3K4me3). Histone 3 acetylation (H3ac) is also seen downstream of the TSSs. Furthermore, these ChIP results confirm that the core promoter identified by Liu and Francke (74) binds RNA polymerase II and TAF1 in HeLa cells (Figure 3), as expected for a site upon which the transcription initiation complex is assembled. Polymorphisms in gene promoters can disrupt sequence-specific interactions with regulatory molecules, resulting in altered gene expression. A presumably rare polymorphism within the *MECP2* promoter region has been reported in one autistic female (49), and several other polymorphisms, including two within the core promoter, have also been documented (http://genome.ucsc.edu/).

3′-UTR

3′-UTRs of RNA transcripts are known to have a number of cytoplasmic functions including involvement in post-transcriptional regulation, mRNA stability and mRNA localization (78). The 3′-UTR of *MECP2* has attracted particular attention because it is one of the longest 3′-UTRs yet documented. Comparison of the mouse and human 3′-UTR genomic sequences reveals a striking pattern of highly conserved regions [see Figure 4; refs. (12,16)], a finding that has been confirmed by the more stringent measures of evolutionary sequence conservation defined by the ENCODE Consortium (79; http://genome.ucsc.edu/). It extends over 8.5 kb and contains multiple

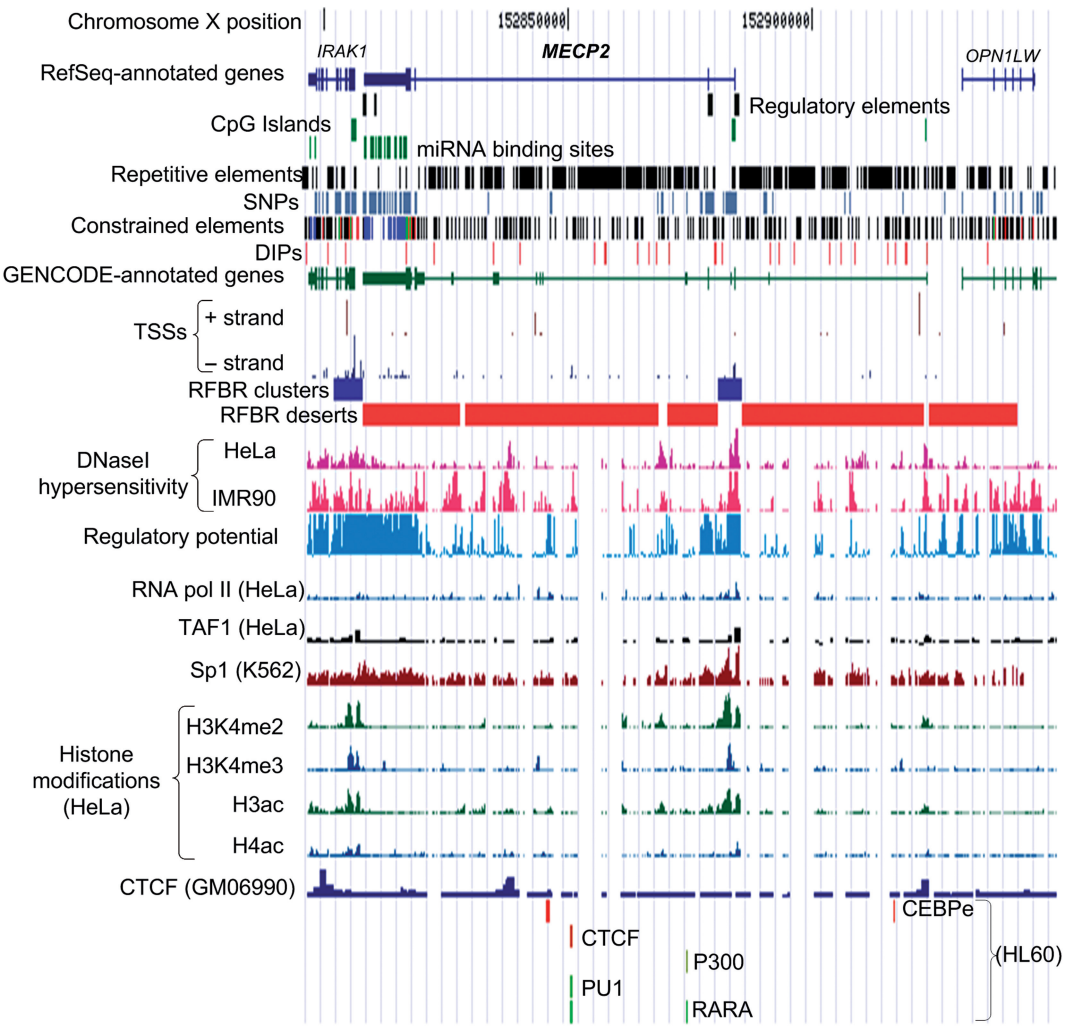

Figure 4. Features within the *MECP2* genomic region and adjacent genes. A UCSC Genome Browser screenshot of a 160 000-nt region encompassing *MECP2* and its immediate neighbouring genes (http://genome.ucsc.edu/, May 2004 assembly) is displayed. The tracks displaying regulatory elements, CpG islands, repetitive elements, constrained elements, SNPs, GENCODE-annotated genes, TSSs, RFBR clusters and deserts, DNase I hypersensitivity, regulatory potential; RNA pol II, TAF1, Sp1 and the histone modifications are all as defined in Figure 3. Also shown are TargetScanS-predicted miRNA-binding sites (March 2006 assembly) deletion-insertion polymorphisms (DIPs) defined by NHGRI, regulatory potential as predicted by ESPERR, CCCTC-binding factor (CTCF) binding sites identified by Sanger in GM06990 (lymphoblastoid) cells and binding sites for CCAAT/enhancer binding protein epsilon (CEBPe), CTCF, P300, PU1 and retinoic acid receptor, alpha (RARA) identified by Affymetrix in HL60 (promyelocytic leukaemia) cells.

polyadenylation signals (Figure 2) associated with alternative transcripts ≈1.8 kb, ≈5.4 kb, ≈7.5 kb and ≈10.2 kb in length (12,14,16,69). The shortest *MECP2* transcript (1.8 kb) uses the canonical 5′ – AATAAA – 3′ polyadenylation signal, which is usually found 10–30 nucleotides upstream of the cleavage site and is bound by cleavage and polyadenylation specificity factor [CPSF; ref. (80)]. The longest transcript (10.2 kb) instead uses the less-efficient sequence 5′ – TATAAA – 3′ (12,16). Both polyadenylation signal sequences are followed by a GT-rich region (16), which is known to bind cleavage stimulation factor (CstF), involved in the polyadenylation and 3′-end cleavage of pre-mRNAs (80).

The alternative *MECP2* transcripts show quantitative differences in different tissues (12,14,16,17,69) as well as in different stages of mouse embryonic development (16) and human post-natal brain development (48,71). Given the unique expression patterns of each transcript, it is likely that the 3′-UTR and the alternative transcripts have biological significance. The 3′-UTR was hypothesized to play a role in transcript stability, as it contains conserved regions predicted to influence RNA folding (16), as well as an AT-rich sequence (12,16) known to influence mRNA stability (81). However, a significant difference in the half lives of *MECP2* long and short transcripts has not been observed (12), and no simple relationship has been found between alternative transcripts and protein levels across a range of tissues (69) or within individual cells (48). Although a quantitative reduction in the 10.2-kb transcript is seen in parallel with increasing MeCP2 during post-natal brain development (71), this inverse relationship between levels of MeCP2 protein and long 3′-UTR usage does not hold at a cellular level, where neurones with abundant MeCP2

protein have increased levels of the 10.2-kb transcript (17,48).

In addition to protein–protein and protein–DNA interactions, gene expression may also be regulated by RNA–protein, RNA–RNA and RNA–DNA interactions. One mechanism by which 3′-UTRs are known to participate in post-transcriptional regulation is through interactions with proteins or microRNAs (miRNAs). miRNAs are small RNA molecules that may be derived from intronic regions, transposons and processed pseudogenes, as well as exons, and are known to suppress translation by binding to mRNA (1). The *MECP2* 3′-UTR contains 52 TargetScanS-predicted miRNA-binding sites (http://genome.ucsc.edu/). Observing that microRNA 132 (miR132) is highly expressed in the brain, Klein *et al.* (82) have recently proposed that miR132, which interacts with the 3′-UTR of the long *MECP2* transcript (Figure 3), may function as a key regulator of MeCP2 expression. miR132 is induced by BDNF and CREB-1, which are both targets of MeCP2-mediated gene regulation (Table 1), suggesting that miR132 may be involved in feedback loops regulating MeCP2 protein levels (82).

Other recently defined features within the *MECP2* genomic region

Until recently, transcripts from the *MECP2* gene locus were limited to a 76-kb interval positioned in the 125-kb gap between the neighbouring genes, encoding the interleukin-1 receptor (*IRAK1*) and a long-wavelength sensitive opsin (*OPN1LW*). The interval to O*PN1LW*, which is transcribed from the opposite strand, is 50 kb and the interval to the annotated exon 1 of *IRAK1* (Gencode v3.1, March 2007), which is transcribed in the same direction as *MECP2*, is ~2 kb. RACE experiments conducted as part of ENCODE have identified a novel 5′exon in testis that is located ~40-kb upstream of the *MECP2* promoter (83), and greatly reduces the size of the intergenic interval at the telomeric end of *MECP2* (Figure 4). Interestingly, the region surrounding the newly identified distal exon contains a CpG island, binds RNA polymerase II and TAF1, and is enriched for the histone modifications H3K4me2, H3K4me3 and H3ac, although to a lesser extent than the region surrounding the currently annotated exon 1 of *MECP2*.

ENCODE has also provided a large resource of comparative genomics data that can assist in the identification of additional *MECP2* functional elements. 4.9% of the nucleotides in the ENCODE regions were identified as constrained by at least two (out of a possible three) programs when using at least two of three multiple sequence alignments (2,79). DNase I hypersensitivity is a marker of regions likely to regulate transcriptional activation, and the region around the newly identified upstream exon displays DNaseI hypersensitivity. It is also predicted by the ESPERR (Evolutionary and Sequence Pattern Extraction through Reduced Representation) program to be regulatory, but only small segments are conserved across species. This is consistent with the observation across the ENCODE regions that although over 50% of non-coding functional elements overlap constrained elements, the proportion of constrained nucleotides within these elements is <10%. This compares with a constraint of almost 70% among bases within coding regions (2,79). In light of the extensive additional transcription revealed by ENCODE, it is also interesting to note the binding of RNA polymerase II and TAF1 to several intronic regions corresponding to the 5′-termini of non-coding exons.

It is notable that the novel transcribed regions identified within introns of *MECP2* also do not have the high evolutionary constraint seen in coding exons, and several of these, as well as a surprisingly large proportion of non-coding functional elements identified throughout other ENCODE regions, do not show evidence of evolutionary constraint in this stringent analysis. Margulies *et al.* (79) suggest that this discrepancy may be explained by possibilities such as elements performing functions that offer no evolutionary advantage, or not requiring constraint at the primary sequence level. Another alternative, which they suggest, is that these elements may have evolved more recently and so may be conserved across primates but not other mammals (79). By contrast, all but one of the *cis*-regulatory elements identified by Liu and Francke (74) overlap these stringently defined constrained elements. It should be noted also that a large proportion of constrained sequences and predicted regulatory regions within *MECP2* do not overlap any of the functional elements that have been described to date, providing an indication of the extent of the pool of uncharacterized functional elements.

Of the novel GENCODE-annotated *MECP2* transcripts, only one has an open reading frame annotated in the UCSC genome browser (http://genome.ucsc.edu), making it likely that the others may function as regulatory ncRNAs. Transcription of ncRNAs from the mammalian genome is an increasingly recognized phenomenon (2,84), and ncRNAs are being implicated in diverse cellular processes, including regulation of various levels of gene expression, regulation of localization and function of other molecules and in the targeting of generic enzymes to specific genomic loci (1). Several recently discovered ncRNAs that have not been well-characterized functionally, have been shown to have regulated expression patterns, suggesting that many ncRNAs have important functions that are yet to be determined (85–88). MeCP2 has been reported to interact with RNA (44), and tissue-specific non-coding transcripts from *MECP2* or other genes may be able to regulate MeCP2 expression or function in a tissue-specific manner. In view of these insights, it seems possible that mutations disrupting the secondary structures of regulatory ncRNAs from *MECP2* may result in disturbed MeCP2 expression and present clinically with a similar phenotype to those patients who have *MECP2* coding-region mutations. As MeCP2 is implicated in neurodevelopmental disorders, the novel transcribed region within intron 1 (Figure 1) that has been detected in brain is of particular interest with regard to disease-causing capacity. Further knowledge of the cellular localization, molecular interactions and qualitative or quantitative differences between RTT-patients and unaffected controls may provide additional insights

into the biological functions and clinical relevance of these newly discovered ncRNAs.

Other DNase I hypersensitivity sites have a distribution within *MECP2* that varies between different cell lines (Figure 4), suggesting the existence of cell-specific regulatory elements. Similar to the chromatin signature patterns observed at TSSs or TSS clusters, the DNase I hypersensitivity sites located downstream from the major TSS cluster display enrichment for H3ac. However, absence of H3K4me3 or RNA polymerase II enrichment, as well as the presence of H4ac, distinguishes these from the TSS-related DNase I hypersensitivity sites (2,77).

CCCTC-binding factor (CTCF) is a transcription factor known to be required for insulator function (89). A CTCF-binding site is located near the most 5′-terminal of the novel *MECP2* transcripts identified by ENCODE, while another is positioned downstream within the *IRAK1* gene. Another CTCF-binding site was reported deep within intron 2 by Kim *et al*. (89) in fibroblasts, and confirmed by the Sanger Institute in a lymphoblastoid cell line (http://genome.ucsc.edu/). This intronic CTCF-binding site displays DNaseI hypersensitivity in the lymphoblastoid cell line, as well as in HeLa cells and fibroblasts. Using the HL60 (promyelocytic leukaemia) cell line, Affymetrix identified another intronic CTCF-binding site that also binds PU1 and the alpha chain of the retinoic acid receptor (RARA; Figure 4; http://genome.ucsc.edu/).

While ENCODE has provided a tremendous amount of information concerning potential regulatory elements within *MECP2*, further work is still needed to meaningfully interpret and integrate these datasets. From the perspective of human disease arising from *MECP2* expression abnormalities, which is a predominantly neuronal phenotype, a particular limitation of ENCODE is that chromatin immunoprecipitation results are derived from non-neuronal cell lines. While the mechanisms that regulate *MECP2* expression within central nervous system tissues may differ from those in the cell lines represented to date, ENCODE provides a shortlist of candidate regulatory regions. The challenge now is to identify which of these regulatory elements are functional in neuronal cells, and whether any of these can assist in the development of new diagnostic or management strategies for patients with dysregulated *MECP2* expression or function.

CONCLUSION

Since the discovery that many RTT cases are due to mutations in *MECP2*, our knowledge of the structure and function of genomic elements within this relatively small gene has advanced significantly. It is now appreciated that *MECP2* probably has several biological functions and mechanisms of action, including the recruitment of chromatin-modifying enzymes. Evidence of precise spatial and temporal regulation of an increasing diversity of expression patterns of the two MeCP2 protein isoforms underlines the importance of the emerging details of epigenetic signatures across the *MECP2* genomic region. In this regard, it is also interesting to note the numerous recently identified highly conserved non-coding regions, including those within the long 3′-UTR and intron 2. The discovery of transcripts arising from novel intronic exons, as well as from what had been regarded as the upstream intergenic region, is particularly exciting in view of the growing recognition of the existence and functional significance of ncRNAs. Such advances in our understanding of gene biology may aid the process of characterizing precisely how *MECP2* dysfunction at the molecular and cellular level leads to neurological phenotypes.

ACKNOWLEDGEMENTS

We would like to thank Paula Moolhuijzen, Dr David S Dunn and Professor Matthew Bellgard of the Centre of Comparative Genomics at Murdoch University for offering their bioinformatic expertise and the use of their resources. We appreciate also the helpful critical comments from an anonymous reviewer.

FUNDING

Robert Vandongen Vacation Scholarship awarded by the Royal Perth Hospital Medical Research Foundation to J.S. A.S. is supported by an NH & MRC Biomedical postdoctoral fellowship 404132.

Conflict of interest statement. None declared.

REFERENCES

1. Mattick,J.S. and Makunin,I.V. (2006) Non-coding RNA. Hum. Mol. Genet., 15 (Spec No. 1), R17–R29.2.
2. The ENCODE Project Consortium. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
3. Gerstein,M.B., Bruce,C., Rozowsky,J.S., Zheng,D., Du,J., Korbel,J.O., Emanuelsson,O., Zhang,Z.D., Weissman,S. and Snyder,M. (2007) What is a gene, post-ENCODE? History and updated definition. *Genome Res*., **17**, 669–681.
4. Gingeras,T.R. (2007) Origin of phenotypes: genes and transcripts. *Genome Res*., **17**, 682–690.
5. Johannsen,W. (1909) *Elemente der exakten erblichkeitslehre : Deutsche Wesentlich erweiterte ausgabe in fünfundzwanzig vorlesuengen*.G. Fisher Jena.
6. Morgan,T.H., Sturtevant,A.H., Muller,H.J. and Bridges,C.B. (1915) *The Mechanism of Mendelian Heredity*. New York, Holt Rinehart & Winston Holt Rinehart & Winston, New York.
7. Beadle,G.W. and Tatum,E.L. (1941) Genetic control of biochemical reactions in neurospora. *Proc. Natl Acad. Sci. USA*, **27**, 499–506.
8. Kapranov,P., Willingham,A.T. and Gingeras,T.R. (2007) Genome-wide transcription and the implications for genomic organization. *Nat. Rev. Genet*., **8**, 413–423.
9. Cheng,J., Kapranov,P., Drenkow,J., Dike,S., Brubaker,S., Patel,S., Long,J., Stern,D., Tammana,H., Helt,G. *et al*. (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–1154.
10. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res*., **12**, 996–1006.
11. Tennyson,C.N., Klamut,H.J. and Worton,R.G. (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet*., **9**, 184–190.
12. Reichwald,K., Thiesen,J., Wiehe,T., Weitzel,J., Poustka,W.A., Rosenthal,A., Platzer,M., Stratling,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.
13. 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.
14. D'Esposito,M., Quaderi,N.A., Ciccodicola,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.
15. Carrel,L. and Willard,H.F. (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, **434**, 400–404.
16. Coy,J.F., Sedlacek,Z., Bachner,D., Delius,H. and Poustka,A. (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.
17. Pelka,G.J., Watson,C.M., Christodoulou,J. and Tam,P.P. (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.
18. 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.
19. 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.
20. 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.
21. Hendrich,B., Abbott,C., McQueen,H., Chambers,D., Cross,S. and Bird,A. (1999) Genomic structure and chromosomal mapping of the murine and human Mbd1, Mbd2, Mbd3, and Mbd4 genes. *Mamm. Genome*, **10**, 906–912.
22. Roloff,T.C., Ropers,H.H. and Nuber,U.A. (2003) Comparative study of methyl-CpG-binding domain proteins. *BMC Genomics*, **4**, 1.
23. Klose,R.J., Sarraf,S.A., Schmiedeberg,L., McDermott,S.M., Stancheva,I. and Bird,A.P. (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell*, **19**, 667–678.
24. Bird,A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
25. Nan,X., Tate,P., Li,E. and Bird,A. (1996) DNA methylation specifies chromosomal localization of MeCP2. *Mol. Cell. Biol.*, **16**, 414–421.
26. Nan,X., Campoy,F.J. and Bird,A. (1997) MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell*, **88**, 471–481.
27. 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.
28. 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.
29. Kokura,K., Kaul,S.C., Wadhwa,R., Nomura,T., Khan,M.M., Shinagawa,T., Yasukawa,T., Colmenares,C. and Ishii,S. (2001) The Ski protein family is required for MeCP2-mediated transcriptional repression. *J. Biol. Chem.*, **276**, 34115–34121.
30. Kaludov,N.K. and Wolffe,A.P. (2000) MeCP2 driven transcriptional repression in vitro: selectivity for methylated DNA, action at a distance and contacts with the basal transcription machinery. *Nucleic Acids Res.*, **28**, 1921–1928.
31. Fuks,F., Hurd,P.J., Wolf,D., Nan,X., Bird,A.P. and Kouzarides,T. (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J. Biol. Chem.*, **278**, 4035–4040.
32. Harikrishnan,K.N., Chow,M.Z., Baker,E.K., Pal,S., Bassal,S., Brasacchio,D., Wang,L., Craig,J.M., Jones,P.L., Sif,S. *et al.* (2005) Brahma links the SWI/SNF chromatin-remodeling complex with MeCP2-dependent transcriptional silencing. *Nat. Genet.*, **37**, 254–264.
33. Nan,X., Hou,J., Maclean,A., Nasir,J., Lafuente,M.J., Shu,X., Kriaucionis,S. and Bird,A. (2007) Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proc. Natl Acad. Sci. USA*, **104**, 2709–2714.
34. Ishibashi,T., Thambirajah,A.A. and Ausio,J. (2008) MeCP2 preferentially binds to methylated linker DNA in the absence of the terminal tail of histone H3 and independently of histone acetylation. *FEBS Lett.*, **582**, 1157–1162.
35. Georgel,P.T., Horowitz-Scherer,R.A., Adkins,N., Woodcock,C.L., Wade,P.A. and Hansen,J.C. (2003) Chromatin compaction by human MeCP2. Assembly of novel secondary chromatin structures in the absence of DNA methylation. *J. Biol. Chem.*, **278**, 32181–32188.
36. Nikitina,T., Shi,X., Ghosh,R.P., Horowitz-Scherer,R.A., Hansen,J.C. and Woodcock,C.L. (2007) Multiple modes of interaction between the methylated DNA binding protein MeCP2 and chromatin. *Mol. Cell Biol.*, **27**, 864–877.
37. Nikitina,T., Ghosh,R.P., Horowitz-Scherer,R.A., Hansen,J.C., Grigoryev,S.A. and Woodcock,C.L. (2007) MeCP2-chromatin interactions include the formation of chromatosome-like structures and are altered in mutations causing Rett syndrome. *J. Biol. Chem.*, **282**, 28237–28245.
38. Galvao,T.C. and Thomas,J.O. (2005) Structure-specific binding of MeCP2 to four-way junction DNA through its methyl CpG-binding domain. *Nucleic Acids Res.*, **33**, 6603–6609.
39. Kumar,A., Kamboj,S., Malone,B.M., Kudo,S., Twiss,J.L., Czymmek,K.J., LaSalle,J.M. and Schanen,N.C. (2008) Analysis of protein domains and Rett syndrome mutations indicate that multiple regions influence chromatin-binding dynamics of the chromatin-associated protein MECP2 in vivo. *J. Cell Sci.*, **121**, 1128–1137.
40. Horike,S., Cai,S., Miyano,M., Cheng,J.F. and Kohwi-Shigematsu,T. (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.*, **37**, 31–40.
41. Schule,B., Li,H.H., Fisch-Kohl,C., Purmann,C. and Francke,U. (2007) DLX5 and DLX6 expression is biallelic and not modulated by MeCP2 deficiency. *Am. J. Hum. Genet.*, **81**, 492–506.
42. Yasui,D.H., Peddada,S., Bieda,M.C., Vallero,R.O., Hogart,A., Nagarajan,R.P., Thatcher,K.N., Farnham,P.J. and Lasalle,J.M. (2007) Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc. Natl Acad. Sci. USA*, **104**, 19416–19421.
43. Chahrour,M., Jung,S.Y., Shaw,C., Zhou,X., Wong,S.T., Qin,J. and Zoghbi,H.Y. (2008) MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science*, **320**, 1224–1229.
44. Jeffery,L. and Nakielny,S. (2004) Components of the DNA methylation system of chromatin control are RNA-binding proteins. *J. Biol. Chem.*, **279**, 49479–49487.
45. Young,J.I., Hong,E.P., Castle,J.C., Crespo-Barreto,J., Bowman,A.B., Rose,M.F., Kang,D., Richman,R., Johnson,J.M., Berget,S. *et al.* (2005) Regulation of RNA splicing by the methylation-dependent transcriptional repressor methyl-CpG binding protein 2. *Proc. Natl Acad. Sci. USA*, **102**, 17551–17558.
46. Adams,V.H., McBryant,S.J., Wade,P.A., Woodcock,C.L. and Hansen,J.C. (2007) Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. *J. Biol. Chem.*, **282**, 15057–15064.
47. Christodoulou,J. (Updated 25 January 2008) *MECP2*-Related disorders. In: GeneReviews at GeneTests: Medical Genetics Information Resource (database online) Copyright, University of Washington, Seattle. 1993–2008. Available at: http://www.genetests.org (7 February 2008, data last accessed).
48. Samaco,R.C., Nagarajan,R.P., Braunschweig,D. and LaSalle,J.M. (2004) Multiple pathways regulate MeCP2 expression in nomal brain development and exhibit defects in autism spectrum disorders. *Hum. Mol. Genet.*, **13**, 629–636.
49. Nagarajan,R.P., Hogart,A.R., Gwye,Y., Martin,M.R. and LaSalle,J.M. (2006) Reduced MeCP2 expression is frequent in autism frontal cortex and correlates with aberrant MECP2 promoter methylation. *Epigenetics*, **1**, e1–e11.
50. 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.
51. Friez,M.J., Jones,J.R., Clarkson,K., Lubs,H., Abuelo,D., Bier,J.A., Pai,S., Simensen,R., Williams,C., Giampietro,P.F. *et al.* (2006) Recurrent infections, hypotonia, and mental retardation caused by duplication of MECP2 and adjacent region in Xq28. *Pediatrics*, **118**, e1687–e1695.
52. Smrt,R.D., Eaves-Egenes,J., Barkho,B.Z., Santistevan,N.J., Zhao,C., Aimone,J.B., Gage,F.H. and Zhao,X. (2007) Mecp2 deficiency leads to delayed maturation and altered gene expression in hippocampal neurons. *Neurobiol. Dis.*, **27**, 77–89.
53. Pelka,G.J., Watson,C.M., Radziewic,T., Hayward,M., Lahooti,H., Christodoulou,J. and Tam,P.P. (2006) Mecp2 deficiency is associated with learning and cognitive deficits and altered gene activity in the hippocampal region of mice. *Brain*, **129**, 887–898.
54. Chang,Q., Khare,G., Dani,V., Nelson,S. and Jaenisch,R. (2006) The disease progression of Mecp2 mutant mice is affected by the level of BDNF expression. *Neuron*, **49**, 341–348.
55. Chen,W.G., Chang,Q., Lin,Y., Meissner,A., West,A.E., Griffith,E.C., Jaenisch,R. and Greenberg,M.E. (2003) Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science*, **302**, 885–889.
56. Stancheva,I., Collins,A.L., Van den Veyver,I.B., Zoghbi,H. and Meehan,R.R. (2003) A mutant form of MeCP2 protein associated with human Rett syndrome cannot be displaced from methylated DNA by notch in Xenopus embryos. *Mol. Cell*, **12**, 425–435.
57. Kishi,N. and Macklis,J.D. (2004) MECP2 is progressively expressed in post-migratory neurons and is involved in neuronal maturation rather than cell fate decisions. *Mol. Cell. Neurosci.*, **27**, 306–321.
58. Mnatzakanian,G.N., Lohi,H., Munteanu,I., Alfred,S.E., Yamada,T., MacLeod,P.J., Jones,J.R., Scherer,S.W., Schanen,N.C., Friez,M.J. *et al.* (2004) A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat. Genet.*, **36**, 339–341.
59. 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.
60. 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.
61. 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.
62. Agarwal,N., Hardt,T., Brero,A., Nowak,D., Rothbauer,U., Becker,A., Leonhardt,H. and Cardoso,M.C. (2007) MeCP2 interacts with HP1 and modulates its heterochromatin association during myogenic differentiation. *Nucleic Acids Res.*, **35**, 5402–5408.
63. Kimura,H. and Shiota,K. (2003) Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J. Biol. Chem.*, **278**, 4806–4812.
64. Kudo,S. (1998) Methyl-CpG-binding protein MeCP2 represses Sp1-activated transcription of the human leukosialin gene when the promoter is methylated. *Mol. Cell. Biol.*, **18**, 5492–5499.
65. 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.
66. Buschdorf,J.P. and Stratling,W.H. (2004) A WW domain binding region in methyl-CpG-binding protein MeCP2: impact on Rett syndrome. *J. Mol. Med.*, **82**, 135–143.
67. Sudol,M., Chen,H.I., Bougeret,C., Einbond,A. and Bork,P. (1995) Characterization of a novel protein-binding module–the WW domain. *FEBS Lett.*, **369**, 67–71.
68. Pires,J.R., Parthier,C., Aido-Machado,R., Wiedemann,U., Otte,L., Bohm,G., Rudolph,R. and Oschkinat,H. (2005) Structural basis for APPTPPPLPP peptide recognition by the FBP11WW1 domain. *J. Mol. Biol.*, **348**, 399–408.
69. Shahbazian,M.D., Antalffy,B., Armstrong,D.L. and Zoghbi,H.Y. (2002) Insight into Rett syndrome: MeCP2 levels display tissue- and cell-specific differences and correlate with neuronal maturation. *Hum. Mol. Genet.*, **11**, 115–124.
70. 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.
71. 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.
72. Kishi,N. and Macklis,J.D. (2005) Dissecting MECP2 function in the central nervous system. *J. Child Neurol.*, **20**, 753–759.
73. 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.
74. Liu,J. and Francke,U. (2006) Identification of cis-regulatory elements for MECP2 expression. *Hum. Mol. Genet.*, **15**, 1769–1782.
75. Carninci,P., Sandelin,A., Lenhard,B., Katayama,S., Shimokawa,K., Ponjavic,J., Semple,C.A., Taylor,M.S., Engstrom,P.G., Frith,M.C. *et al.* (2006) Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.*, **38**, 626–635.
76. Sandelin,A., Carninci,P., Lenhard,B., Ponjavic,J., Hayashizaki,Y. and Hume,D.A. (2007) Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.*, **8**, 424–436.
77. Koch,C.M., Andrews,R.M., Flicek,P., Dillon,S.C., Karaoz,U., Clelland,G.K., Wilcox,S., Beare,D.M., Fowler,J.C., Couttet,P. *et al.* (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res.*, **17**, 691–707.
78. Decker,C.J. and Parker,R. (1995) Diversity of cytoplasmic functions for the 3' untranslated region of eukaryotic transcripts. *Curr. Opin. Cell. Biol.*, **7**, 386–392.
79. Margulies,E.H., Cooper,G.M., Asimenos,G., Thomas,D.J., Dewey,C.N., Siepel,A., Birney,E., Keefe,D., Schwartz,A.S., Hou,M. *et al.* (2007) Analyses of deep mammalian sequence alignments and constraint predictions for 1% of the human genome. *Genome Res.*, **17**, 760–774.
80. Brown,T.A. (2007) *Genomes 3*, 3rd edn. Garland Science, New York.
81. Chen,C.Y. and Shyu,A.B. (1995) AU-rich elements: characterization and importance in mRNA degradation. *Trends Biochem. Sci.*, **20**, 465–470.
82. Klein,M.E., Lioy,D.T., Ma,L., Impey,S., Mandel,G. and Goodman,R.H. (2007) Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA. *Nat. Neurosci.*, **10**, 1513–1514.
83. Denoeud,F., Kapranov,P., Ucla,C., Frankish,A., Castelo,R., Drenkow,J., Lagarde,J., Alioto,T., Manzano,C., Chrast,J. *et al.* (2007) Prominent use of distal 5' transcription start sites and discovery of a large number of additional exons in ENCODE regions. *Genome Res.*, **17**, 746–759.
84. Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, **420**, 563–573.
85. Mercer,T.R., Dinger,M.E., Sunkin,S.M., Mehler,M.F. and Mattick,J.S. (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA*, **105**, 716–721.
86. Perez,D.S., Hoage,T.R., Pritchett,J.R., Ducharme-Smith,A.L., Halling,M.L., Ganapathiraju,S.C., Streng,P.S. and Smith,D.I. (2008) Long, abundantly expressed non-coding transcripts are altered in cancer. *Hum. Mol. Genet.*, **17**, 642–655.
87. Ravasi,T., Suzuki,H., Pang,K.C., Katayama,S., Furuno,M., Okunishi,R., Fukuda,S., Ru,K., Frith,M.C., Gongora,M.M. *et al.* (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.*, **16**, 11–19.
88. Sasaki,Y.T., Sano,M., Ideue,T., Kin,T., Asai,K. and Hirose,T. (2007) Identification and characterization of human non-coding RNAs with tissue-specific expression. *Biochem. Biophys. Res. Commun.*, **357**, 991–996.
89. Kim,T.H., Abdullaev,Z.K., Smith,A.D., Ching,K.A., Loukinov,D.I., Green,R.D., Zhang,M.Q., Lobanenkov,V.V. and Ren,B. (2007) Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell*, **128**, 1231–1245.
90. Martinowich,K., Hattori,D., Wu,H., Fouse,S., He,F., Hu,Y., Fan,G. and Sun,Y.E. (2003) DNA methylation-related chromatin remodeling in activity-dependent BDNF gene regulation. *Science*, **302**, 890–893.

91. Ballestar,E., Ropero,S., Alaminos,M., Armstrong,J., Setien,F., Agrelo,R., Fraga,M.F., Herranz,M., Avila,S., Pineda,M. *et al.* (2005) The impact of MECP2 mutations in the expression patterns of Rett syndrome patients. *Hum. Genet.*, **116**, 91–104.
92. Samaco,R.C., Hogart,A. and LaSalle,J.M. (2005) Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. *Hum. Mol. Genet.*, **14**, 483–492.
93. Makedonski,K., Abuhatzira,L., Kaufman,Y., Razin,A. and Shemer,R. (2005) MeCP2 deficiency in Rett syndrome causes epigenetic aberrations at the PWS/AS imprinting center that affects UBE3A expression. *Hum. Mol. Genet.*, **14**, 1049–1058.
94. Nuber,U.A., Kriaucionis,S., Roloff,T.C., Guy,J., Selfridge,J., Steinhoff,C., Schulz,R., Lipkowitz,B., Ropers,H.H., Holmes,M.C. *et al.* (2005) Up-regulation of glucocorticoid-regulated genes in a mouse model of Rett syndrome. *Hum. Mol. Genet.*, **14**, 2247–2256.
95. Peddada,S., Yasui,D.H. and LaSalle,J.M. (2006) Inhibitors of differentiation (ID1, ID2, ID3 and ID4) genes are neuronal targets of MeCP2 that are elevated in Rett syndrome. *Hum. Mol. Genet.*, **15**, 2003–2014.
96. McGill,B.E., Bundle,S.F., Yaylaoglu,M.B., Carson,J.P., Thaller,C. and Zoghbi,H.Y. (2006) Enhanced anxiety and stress-induced corticosterone release are associated with increased Crh expression in a mouse model of Rett syndrome. *Proc. Natl Acad. Sci. USA*, **103**, 18267–18272.
97. Itoh,M., Ide,S., Takashima,S., Kudo,S., Nomura,Y., Segawa,M., Kubota,T., Mori,H., Tanaka,S., Horie,H. *et al.* (2007) Methyl CpG-binding protein 2 (a mutation of which causes Rett syndrome) directly regulates insulin-like growth factor binding protein 3 in mouse and human brains. *J. Neuropathol. Exp. Neurol.*, **66**, 117–123.
98. Deng,V., Matagne,V., Banine,F., Frerking,M., Ohliger,P., Budden,S., Pevsner,J., Dissen,G.A., Sherman,L.S. and Ojeda,S.R. (2007) FXYD1 is an MeCP2 target gene overexpressed in the brains of Rett syndrome patients and Mecp2-null mice. *Hum. Mol. Genet.*, **16**, 640–650.
99. Jordan,C., Li,H.H., Kwan,H.C. and Francke,U. (2007) Cerebellar gene expression profiles of mouse models for Rett syndrome reveal novel MeCP2 targets. *BMC Med. Genet.*, **8**, 36.
100. Jordan,C. and Francke,U. (2006) Ube3a expression is not altered in Mecp2 mutant mice. *Hum. Mol. Genet.*, **15**, 2210–2215.