Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-CpG-binding protein 2

Ruthie E. Amir[1], Ignatia B. Van den Veyver[2,3], Mimi Wan[5], Charles Q. Tran[3], Uta Francke[5,6] & Huda Y. Zoghbi[1,2,4]

Rett syndrome[1] (RTT, MIM 312750) is a progressive neurodevelopmental disorder and one of the most common causes of mental retardation in females, with an incidence of 1 in 10,000–15,000 (ref. 2). Patients with classic RTT appear to develop normally until 6–18 months of age, then gradually lose speech and purposeful hand use, and develop microcephaly, seizures, autism, ataxia, intermittent hyperventilation and stereotypic hand movements[3]. After initial regression, the condition stabilizes and patients usually survive into adulthood. As RTT occurs almost exclusively in females, it has been proposed that RTT is caused by an X-linked dominant mutation with lethality in hemizygous males[3–8]. Previous exclusion mapping studies using RTT families mapped the locus to Xq28 (refs 6,7,9–11). Using a systematic gene screening approach, we have identified mutations in the gene (*MECP2*) encoding X-linked methyl-CpG-binding protein 2 (MeCP2) as the cause of some cases of RTT. MeCP2 selectively binds CpG dinucleotides in the mammalian genome and mediates transcriptional repression through interaction with histone deacetylase and the corepressor SIN3A (refs 12,13). In 5 of 21 sporadic patients, we found 3 *de novo* missense mutations in the region encoding the highly conserved methyl-binding domain (MBD) as well as a *de novo* frameshift and a *de novo* nonsense mutation, both of which disrupt the transcription repression domain (TRD). In two affected half-sisters of a RTT family, we found segregation of an additional missense mutation not detected in their obligate carrier mother. This suggests that the mother is a germline mosaic for this mutation. Our study reports the first disease-causing mutations in RTT and points to abnormal epigenetic regulation as the mechanism underlying the pathogenesis of RTT.

We carried out systematic mutational analysis of genes located in Xq28 in RTT patients. We chose a number of candidate genes from this region based on their known function and expression patterns, but recently excluded these genes[14,15]. We then analysed the gene encoding methyl-CpG binding protein 2 (*MECP2*), which maps to Xq28 between *L1CAM* and the *RCP*/*GCP* loci and undergoes X inactivation[16]. MeCP2 is an abundant chromosome-binding protein that selectively binds 5-methyl cytosine residues in symmetrically positioned CpG dinucleotides in mammalian genomes[17]. These residues are preferentially located in the promoter regions of genes that are subject to transcriptional silencing after DNA methylation. MeCP2 contains two functional domains, an 85 amino acid methyl-CpG binding domain (MBD), essential for its binding to 5-methylcytosine[18], and a 104 amino acid transcriptional repression domain (TRD) that interacts with histone deacetylase and the transcriptional corepressor SIN3A. Interactions between this transcription repressor complex and chromatin-bound MeCP2 leads to deacetylation of core histones, which in turn leads to transcriptional repression[12,13]. This complex also can inhibit transcription from a promoter at a distance[19].

Using published genomic sequence of *MECP2*, we designed primers complementary to intronic sequences for PCR amplification of all *MECP2* coding exons, including the splice junctions. We screened genomic DNA from 21 sporadic and 8 familial RTT patients by conformation-sensitive gel electrophoresis (CSGE) to look for heteroduplexes and by direct sequencing. We confirmed that all sporadic patients screened in this analysis had classic RTT. The familial cases included five pairs of full sisters (unpublished cases), two pairs of half-sisters and a pair of second half-cousins[6]. Among the sporadic patients we identified three missense mutations, one frameshift mutation and a nonsense mutation (Table 1 and Fig. 1). The R133C mutation in patient 39 replaces the basic amino acid arginine with cysteine. The F155S and the T158M mutations in patients 24 and 6, respectively, substitute hydrophobic amino acids with polar amino acids. These changes may disrupt the structure of

Table 1 • *MECP2* mutations in RTT

Patient	Nucleotide[a]	Protein[a]	Parents
sporadic-39	471C→T	R133C	*de novo*
sporadic-24	538T→C	F155S	*de novo*
sporadic-6	547C→T	T158M	*de novo*
sporadic-22	837C→T	nonsense	*de novo*
sporadic-29	694insT	frameshift[b]	not present in the mother[c]
familial: C2[d], C3[d]	390C→T	R106W	not present in the mother[c]
Benign variants			
familial: F3[e], F4[e]	656C→T	none	present in sibs and father
sporadic-10	1307C→T	none	not present in the mother[c]

[a]Nucleotide and amino acid numbering according to GenBank. [b]Stop codon after 27 out-of-frame amino acids. [c]Father is unavailable. [d]Two affected half-sisters. [e]Two affected full sisters.

Departments of [1]Pediatrics, [2]Molecular and Human Genetics and [3]Obstetrics and Gynecology and [4]Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030, USA. [5]Department of Genetics and [6]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA. Correspondence should be addressed to H.Y.Z. (e-mail: hzoghbi@bcm.tmc.edu).

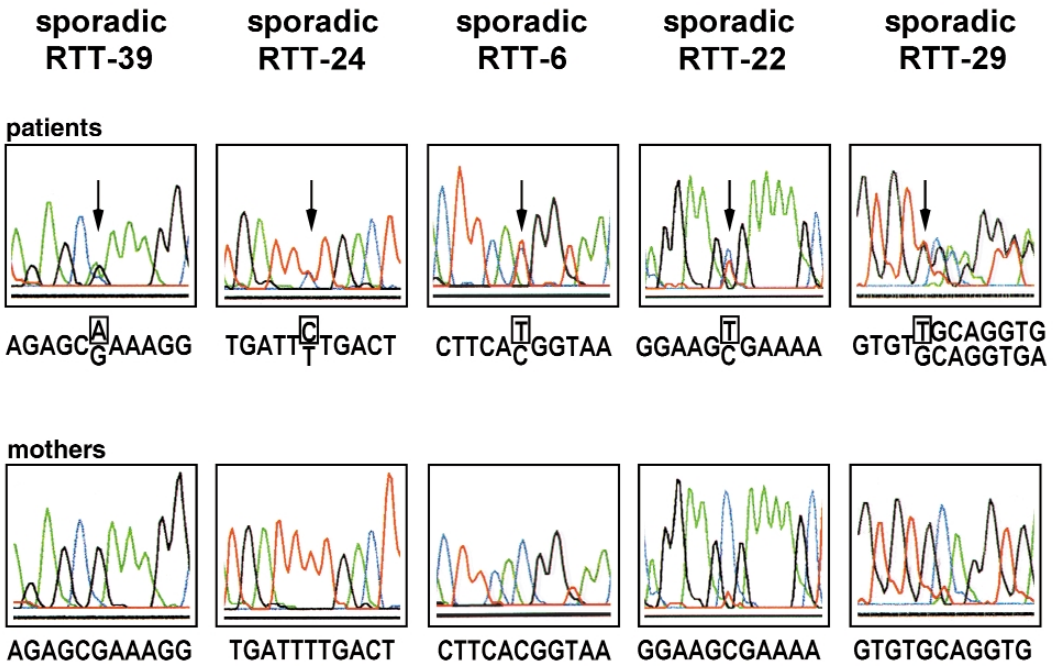

Fig. 1 *MECP2* mutations in sporadic RTT patients. Portions of the electropherograms illustrating 5 mutations found in sporadic patients 6, 22, 24, 29 and 39. Top, mutated sequences in patients; bottom, normal sequence from each patient's mother. The boxed nucleotides and arrows indicate mutated nucleotides for each patient in panels 39 (A), 24 (C), 6 (T) and 22 (T), and the inserted nucleotide (T) in panel 29. The two sequences under the electropherogram of patient 29 represent the superimposed sequences caused by the frameshift. All sequences are in the sense orientation except for that of patient 39.

the methyl-binding domain, thereby interfering with its function. The nonsense mutation in patient 22 is a C→T (bp 837) substitution, which converts a CGA to a TGA (R255X) that predicts truncation of the MeCP2 protein at residue 255 of 486. In patient 29, an insertion (694insT) at codon 208 shifts the reading frame and introduces a stop codon after 27 amino acids. In these last two cases, the predicted truncated proteins lack an intact transcription repression domain. We analysed DNA samples from both parents for all patients except 29 (frameshift mutation), whose father's DNA was not available. None of the parents' samples showed any abnormalities by CSGE or sequence analysis, demonstrating that these are *de novo* mutations (Fig. 1). As we analysed DNA from only the mother of patient 29, we cannot exclude mosaicism in the father. We also identified a missense mutation, 390C→T, changing a conserved amino acid (R106W) in the MBD of the protein in a family with two affected half-sisters who have the same mother (family 1 in ref. 6; Fig. 2). Because the half-sisters carry the same mutation, we conclude that their mother must be an obligate carrier. This obligate carrier female is normal, and previous studies have shown that she has a random X-inactivation pattern in her peripheral blood leukocytes, in contrast to the several carrier females who have skewed X-inactivation patterns[5,7,11]. Neither sequence nor heteroduplex analyses detected the mutation in her genomic DNA. These findings suggest that germline mosaicism is likely to be the mechanism by which she transmitted the disease to both daughters, but we cannot exclude the possibility that she has low-level somatic mosaicism in other tissues. All four of the missense mutations change amino acids in the methyl-binding domain that are conserved in human, mouse, chicken and *Xenopus laevis* (Fig. 3). We detected none of these mutations in 96 non-RTT chromosomes. We did identify two silent single-nucleotide polymorphisms (SNPs): a 656C→T substitution that occurred in two affected sisters and was inherited from the normal father, and a 1307C→T substitution in a sporadic patient whose mother's DNA does not have the polymorphism and whose father's DNA is not available. These SNPs were not detected in the 96 non-Rett chromosomes; the presence of the 656C→T SNP in the normal father, together with the finding that these nucleotide substitutions do not alter the respective codons, suggests that they are benign.

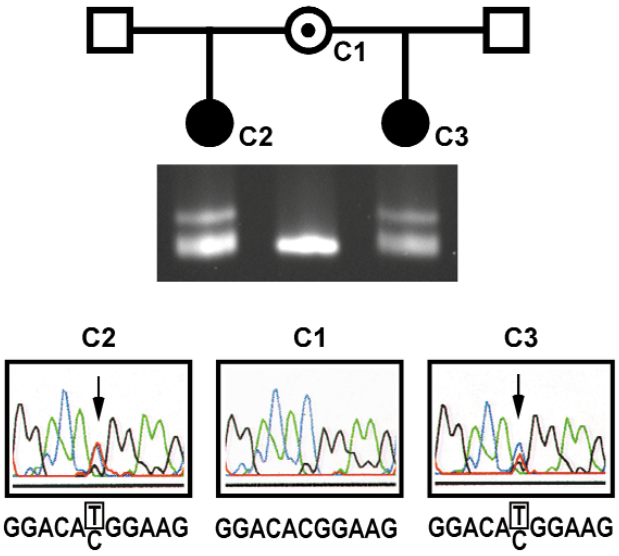

Fig. 2 Mutations in the family of affected half-sisters. The pedigree is shown on top. The gel picture in the middle presents the result of the heteroduplex analysis: no heteroduplex was found in the mother (C1), but both affected daughters (C2, C3) have a slower-migrating band representing a heteroduplex. The electropherograms of tested individuals are below their respective pedigree symbols. The affected half-sisters share the same missense mutation (390C→T), which alters a conserved amino acid (R106W), whereas their mother, who is their common parent, has a cytosine at this position.

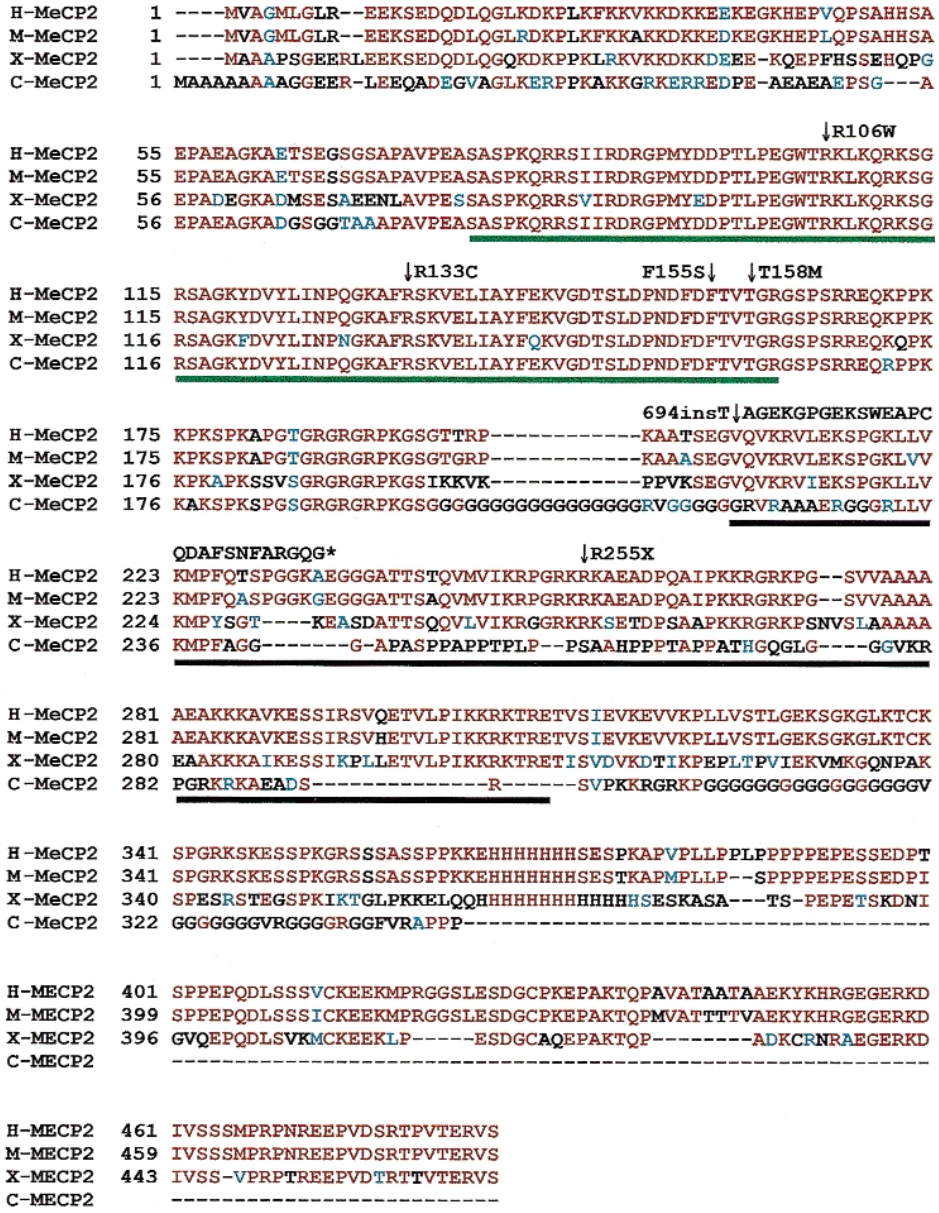

Fig. 3 Alignment of MeCP2 sequences from different species with the positions of the mutations in RTT. Red, identical amino acids between species; blue, similar amino acids; green, conserved methyl-cytosine-binding domain; black, the transcription repression domain. Arrows show the precise positions of the mutations. The 694insT mutation leads to 27 out-of-frame amino acids and a stop codon (*). The protein sequence alignment allows comparison of human (H-MeCP2), mouse (M-MeCP2), chicken (C-MeCP2) and *X. laevis* (X-MeCP2) proteins.

Given that all mutations we identified are *de novo* in sporadic cases, one mutation segregates in familial RTT, all missense mutations change conserved amino acids in the MBD and both truncating mutations disrupt the TRD of MeCP2, we conclude that mutations in *MECP2* are the cause of RTT in these individuals. The nature of these mutations makes it likely that they lead to either partial or complete loss of function of MeCP2. The random pattern of X inactivation in most RTT patients, according to *PGK1*, *HPRT1* and *AR* methylation assays[5,20], ensures expression of the normal allele in some cells. The normal allele probably enables survival of affected females but does not protect them from major neurodevelopmental abnormalities.

MECP2 is essential once cellular differentiation begins. Targeted deletion of this gene in embryonic stem (ES) cells did not have notable effects on ES survival and proliferation, but chimaeric embryos with a high level of contribution from mutant ES cells failed to gastrulate and died between embryonic days 8.5 and 12 (ref. 21). This is consistent with the X-linked, dominant, male-lethal phenotype of RTT.

To our knowledge, RTT is the first human disease to be caused by mutations in a gene encoding a *trans*-acting factor that has a role in the epigenetic regulation of gene expression. Why is the Rett phenotype limited for the most part to the nervous system? MeCP2 is widely expressed, and is abundant in the brain; alternative polyadenylation in the 3′ UTR results in a variety of transcripts, some of which are differentially expressed in human brain[16,22]. The longest transcript (10.1 kb) is most highly expressed in fetal brain, whereas the 5-kb transcript is enriched in adult brain[22]. It is conceivable that loss of function of this protein in some cells, especially differentiated and postmitotic neurons, would lead to overexpression of some genes that may be detrimental during nervous system maturation. We have found mutations in only 5 of 21 sporadic patients and 1 familial patient, but we have screened only for mutations in the coding region. The high degree of conservation across species of several regions in the 3′ UTR suggests that these sequences are under evolutionary selection and that they are important for post-transcriptional regulation of *MECP2* (ref. 22). It is plausible that mutations in the 3′ UTR of *MECP2* may be the underlying cause of RTT in

patients whose mutations were not detected here. Another possibility is that some cases of RTT might be caused by autosomal mutations in proteins related to MeCP2. For example, MeCP2 belongs to a family of MBD-containing proteins that may mediate transcriptional regulation[23]. The genomic structure and mapping data of four additional members of this protein family have recently been described[24]; mutations in any of these proteins or their interactors may cause RTT or related phenotypes such as autism and non-syndromic mental retardation.

The discovery of *MeCP2* as a RTT gene will enable the development of a test for early diagnosis and prenatal detection, and the finding that epigenetic regulation has a role in the pathogenesis of RTT may provide opportunities for therapy. Although it is not clear at this point what the pathogenic mechanism is, it is possible that partial loss of function of MeCP2 would decrease transcriptional repression of some genes. The relatively normal development during the first 6–18 months of life may allow for presymptomatic therapeutic intervention, especially if newborn screening programs can identify affected females.

Note added in proof: The missense mutation causing the R106W substitution in the familial case has been identified in an additional sporadic RTT patient.

Methods

CSGE analysis. We prepared total genomic DNA from peripheral blood leukocytes or from lymphoblastoid cell lines using standard protocols[5]. We designed the following primer pairs using the available genomic sequence of the *MECP2* locus and amplified the coding exons and portions of the 3′ UTR: exon 1 forward, 5′–GTTATGTCTTTAGTCTTTGG–3′, and reverse, 5′–TGTGTTTATCTTCAAAATGT–3′; exon 2 forward, 5′–CCTGCCTCTGCTCACTTGTT–3′, and reverse, 5′–GGGGTCATCATACATGGGTC–3′; forward, 5′–AGCCCGTGCAGCCATCAGCC–3′, and reverse, 5′–GTTCCCCCCGACCCCACCCT–3′; exon 3 forward, 5′–TTTGTCAGAGCGTTGTCACC–3′, and reverse, 5′–CTTCCCAGGACTTTTCTCCA–3′; forward, 5′–AACCACCTAAGAAGCCCAAA–3′, and reverse, 5′–CTGCACAGATCGGATAGAAGAC–3′; forward, 5′–GGCAGGAAGCGAAAAGCTGAG–3′, and reverse, 5′–TGAGTGGTGGTGATGGTGGTGG–3′; forward, 5′–TGGTGAAGCCCCTGCTGGT–3′, and reverse, 5′–CTCCCTCCCCTCGGTGTTTG–3′; forward, 5′–GGAGAAGATGCCCAGAGGAG–3′, and reverse, 5′–CGGTAAGAAAAACATCCCCAA–3′. We performed PCR amplification in a final volume (25–50 µl) with 1×PCR buffer (50 mM KCL, 10 mM Tris HCL, 1.5 mM $MgCl_2$, 0.1% w/v gelatin), dNTPs (0.25 mM), *Taq* polymerase (0.625 U; Cetus) and primers (1 µm each). PCR conditions were: initial denaturation at 95 °C for 5 min followed by 35 cycles of denaturation at 95 °C, annealing at (T_m) and extension at 72 °C for 1 min each. The T_m was 58–62 °C for exon 2 and exon 3 and 50 °C for exon 1. The amplified products were denatured at 95 °C for 5 min, allowed to reanneal at 68 °C for 60 min, and electrophoresed at 450–500 V for 16 h on conformation-sensitive polyacrylamide gels to resolve heteroduplexes according to the manufacturer's specifications[25] (Bio-Rad).

Sequence analysis. We purified PCR products using a Qiagen PCR purification kit and sequenced amplimers directly using the ABI PRISM dye terminator cycle sequencing ready reaction kit (Perkin-Elmer). An ABI 377 DNA sequencer (Applied Biosystems) performed automated sequencing. We used GCG software, Wisconsin package version 10.0-unix, to analyse sequences.

GenBank accession number. *MECP2* locus, AF030876; *MECP2*, X99686.

Acknowledgements

We thank the Rett families for their participation and for motivating us as we sought the cause of this disease; the Blue Bird Circle Rett Center, D.G. Glaze and R.J. Schultz for following patients and facilitating sample collection; E.J.R. Dahle for previous research contributions; and A.L. Beaudet for discussions and critical review of the manuscript. This research was supported by the Howard Hughes Medical Institute and NIH grants HD24234 (H.Y.Z. and U.F.) and MRRC HD 24064, the International Rett Syndrome Association (R.A.), the Society for Gynecologic Investigation (I.V.) and the L.M. Chandler Research Fund (M.W.).

Received 26 August; accepted 1 September 1999.

1. Rett, A. *Uber ein zerebral-atrophisches Syndrome bei Hyperammonemie* (Bruder Hollinek, Vienna, 1966).
2. Hagberg, B. Rett's syndrome: prevalence and impact on progressive severe mental retardation in girls. *Acta Paediatr. Scand.* **74**, 405–408 (1985).
3. Hagberg, B., Aicardi, J., Dias, K. & Ramos, O. A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: report of 35 cases. *Ann. Neurol.* **14**, 471–479 (1983).
4. Zoghbi, H. Genetic aspects of Rett syndrome. *J. Child Neurol.* **3**, S76–78 (1988).
5. Zoghbi, H.Y., Percy, A.K., Schultz, R.J. & Fill, C. Patterns of X chromosome inactivation in the Rett syndrome. *Brain Dev.* **12**, 131–135 (1990).
6. Ellison, K.A. *et al.* Examination of X chromosome markers in Rett syndrome: exclusion mapping with a novel variation on multilocus linkage analysis. *Am. J. Hum. Genet.* **50**, 278–287 (1992).
7. Schanen, N.C. *et al.* A new Rett syndrome family consistent with X-linked inheritance expands the X chromosome exclusion map. *Am. J. Hum. Genet.* **61**, 634–641 (1997).
8. Schanen, C. & Francke, U. A severely affected male born into a Rett syndrome kindred supports X-linked inheritance and allows extension of the exclusion map. *Am. J. Hum. Genet.* **63**, 267–269 (1998).
9. Archidiacono, N. *et al.* Rett syndrome: exclusion mapping following the hypothesis of germinal mosaicism for new X-linked mutations. *Hum. Genet.* **86**, 604–606 (1991).
10. Curtis, A.R. *et al.* X chromosome linkage studies in familial Rett syndrome. *Hum. Genet.* **90**, 551–555 (1993).
11. Sirianni, N., Naidu, S., Pereira, J., Pillotto, R.F. & Hoffman, E.P. Rett syndrome: confirmation of X-linked dominant inheritance, and localization of the gene to Xq28. *Am. J. Hum. Genet.* **63**, 1552–1558 (1998).
12. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).
13. Jones, P.L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.* **19**, 187–191 (1998).
14. Amir, R., Roth Dahle, E., Toniolo, D. & Zoghbi, H.Y. Candidate gene analysis in Rett syndrome and the identification of twenty-one SNPs in Xq. *Am. J. Med. Genet.* (in press).
15. Wan, M. & Francke, U. Evaluation of two X chromosomal candidate genes for Rett syndrome: glutamate dehydrogenase-2 (*GLUD2*) and rab GDP-dissociation inhibitor (*GDI1*). *Am. J. Med. Genet.* **78**, 169–172 (1998).
16. D'Esposito, M. *et al.* Isolation, physical mapping, and northern analysis of the X-linked human gene encoding methyl CpG-binding protein, MECP2. *Mamm. Genome* **7**, 533–535 (1996).
17. Lewis, J.D. *et al.* Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* **69**, 905–914 (1992).
18. Nan, X., Meehan, R.R. & Bird, A. Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. *Nucleic Acids Res.* **21**, 4886–4892 (1993).
19. Nan, X., Campoy, F.J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471–481 (1997).
20. Allen, R.C., Zoghbi, H.Y., Moseley, A.B., Rosenblatt, H.M. & Belmont, J.W. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am. J. Hum. Genet.* **51**, 1229–1239 (1992).
21. Tate, P., Skarnes, W. & Bird, A. The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nature Genet.* **12**, 205–208 (1996).
22. Coy, J.F., Sedlacek, Z., Bachner, D., Delius, H. & Poustka, A. 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 (1999).
23. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18**, 6538–6547 (1998).
24. Hendrich, B. *et al.* Genomic structure and chromosomal mapping of the murine and human mbd1, mbd2, mbd3, and mbd4 genes. *Mamm. Genome* **10**, 906–912 (1999).
25. Ganguly, A., Rock, M.J. & Prockop, D.J. Conformation-sensitive gel electrophoresis for rapid detection of single-base differences in double-stranded PCR products and DNA fragments: evidence for solvent-induced bends in DNA heteroduplexes. *Proc. Natl Acad. Sci. USA* **90**, 10325–10329 (1993); erratum: **91**, 5217 (1994).