A previously unidentified *MECP2* open reading frame defines a new protein isoform relevant to Rett syndrome

Gevork N Mnatzakanian[1], Hannes Lohi[1], Iulia Munteanu[1,2], Simon E Alfred[3], Takahiro Yamada[1], Patrick JM MacLeod[4], Julie R Jones[5], Stephen W Scherer[1,2], N Carolyn Schanen[6], Michael J Friez[5], John B Vincent[2,3,8] & Berge A Minassian[1,2,7,8]

Rett syndrome is caused by mutations in the gene *MECP2* in ~80% of affected individuals. We describe a previously unknown MeCP2 isoform. Mutations unique to this isoform and the absence, until now, of identified mutations specific to the previously recognized protein indicate an important role for the newly discovered molecule in the pathogenesis of Rett syndrome.

Rett syndrome (RTT; OMIM 312750) is characterized by onset, in girls, of a gradual slowing of neurodevelopment in the second half of the first year of life that proceeds towards stagnation by age 4 years, followed by regression and loss of acquired fine motor and communication skills. A pseudostationary period follows during which a picture of preserved ambulation, aberrant communication and stereotypic hand wringing approximates early autism. Regression, however, remains insidiously ongoing and ultimately results in profound mental retardation[1].

Up to 80% of individuals with RTT have mutations[2,3] in exons 3 and 4 of the four-exon gene *MECP2* (**Fig. 1a**)[4] encoding the transcriptional repressor MeCP2. In the known transcript of the gene, all four exons are used, the translation start site is in exon 2, and exon 1 and most of exon 2 form the 5′ untranslated region (UTR)[4]. For clarity, we refer to this transcript as *MECP2*A and its encoded protein as MeCP2A. We sought to identify *MECP2* splice variants contributing new coding sequence that might contain mutations in the remaining individuals with RTT. Inspection of the 5′ UTR showed that, whereas exon 2 has a number of in-frame stop codons upstream of the ATG start codon, exon 1 contains an open reading frame across its entire length, including an ATG. Submitting a theoretical construct composed of exons 1, 3 and 4 to the ATGpr program (http://www.hri.co.jp/atgpr/), which predicts the likelihood that an ATG will be an initiation codon based on the significance of its surrounding Kozak nucleotide context, returned a reliability score of 97%, as compared with 64% for *MECP2*A. A search in EST databases identified eight examples of our theorized transcript, which we named *MECP2*B (**Fig. 1b**), as compared with 14 examples of *MECP2*A. *MECP2*B is predicted to encode a new isoform, MeCP2B, with an alternative, longer N terminus determined by exon 1 (see **Supplementary Table 1** online).

To confirm that *MECP2*B is expressed and not merely an artifact of cDNA library preparation, we amplified cDNA by PCR from a variety of tissues using a 5′ primer in exon 1 and a 3′ primer in exon 3 (**Fig. 1a**). We obtained two PCR products corresponding in size and sequence to *MECP2*A and *MECP2*B in all tissues, including fetal and adult brain and different brain subregions (**Fig. 1c**). Results in mouse were similar (**Fig. 1c**). We quantified the expression levels of the two transcripts in adult human brain. Expression of *MECP2*B was ten times higher than that of *MECP2*A (**Fig. 1d**). We studied the subcellular localization of MeCP2B after transfection of 3′ myc-tagged *MECP2*B into COS-7 cells and found it to be principally in the nucleus (**Fig. 1e**).

To determine whether the new coding region is mutated in RTT, we PCR-amplified and sequenced exon 1 and flanking sequences in 19 girls with typical RTT in whom no mutations had been found in the other exons. One affected individual (V1) carries a deletion of 11 bp in exon 1 (**Fig. 2**). The deletion occurs within the predicted exon 1 open reading frame of *MECP2*B and leads to a frame shift that results in a missense amino acid sequence followed by a premature stop codon after amino acid 36. It does not affect the coding sequence of *MECP2*A. We did not find this sequence change in any of 200 control individuals, including individual V1's parents and brother. To search, in the remaining affected individuals, for additional exon 1 deletions not detectable by our PCR reaction, we carried out multiplex ligation-dependent probe amplification (MLPA)[5] in all four exons and detected hemizygous deletion of exon 1 in one individual with RTT (individual V2; see **Supplementary Note** online). An additional affected individual with an MLPA-detected deletion restricted to exon 1 was recently documented in a published abstract, though the effect on *MECP2*B was not appreciated (Boulanger, S. *et al.*, *Am. J. Hum. Genet.* **73**, 572, 2003).

In contrast, no mutation specific to the MeCP2A-defining exon 2 has been identified despite the analysis of several hundred individuals with RTT for mutations in this exon (31 publications; most reviewed in ref. 3). These studies did not include exon 1, as it was considered noncoding.

Exon 1 deletions result in absent or truncated MeCP2B protein. However, they also result in shortening of *MECP2*A's 5′ UTR and could potentially affect its expression. We tested this possibility in individual V1 by carrying out RT-PCR of whole blood, however, and found no diminution of *MECP2*A expression (**Fig. 2c**). In conclusion, mutation data indicate that inactivation of MeCP2B is sufficient to cause RTT, whereas the same cannot so far be said of MeCP2A.

Why was *MECP2*B not detected in previous expression studies? Northern analyses showed three transcripts, of 1.9, 5 and 10.1 kb, with the differences in size resulting from alternative polyadenylation signal usage[4,6–8] (**Fig. 1a**). *MECP2*B differs from *MECP2*A in lacking the 124-nucleotide exon 2. At the 5- and 10.1-kb positions on the gel, the two transcripts would not be separable. In the 1.9-kb range, published northern blots do show a thick or double band that is likely to correspond to the two transcripts. Likewise, conventional western-blot analysis would not allow resolution of the two MeCP2 isoforms (molecular weight difference <0.9 kDa; **Supplementary Table 1** online).

Recent studies in frog (*Xenopus laevis*) have afforded useful insight into the role of MeCP2 in neurodevelopmental transcription regulation. MeCP2 was shown to be a component of the SMRT (silencing mediator of retinoic acid and thyroid receptors) complex that participates in the regulation of genes involved in neuronal dif-

[1]Program in Genetics and Genomic Biology, Research Institute, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [2]Institute of Medical Sciences, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [3]The Centre for Addiction and Mental Health, 250 College Street, Toronto, Ontario M5T 1R8, Canada. [4]Division of Medical Genetics, Victoria General Hospital, One Hospital Way, Victoria, British Columbia V8R 6R5, Canada. [5]Greenwood Genetic Center, One Gregor Mendel Circle, Greenwood, South Carolina 29646, USA. [6]Nemours Research Program, Alfred I. duPont Hospital for Children, 1600 Rockland Road, Wilmington, Delaware 19803, USA. [7]Division of Neurology, Department of Paediatrics, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [8]These authors contributed equally to this work. Correspondence should be addressed to B.A.M. (bminass@sickkids.ca).

Published online 21 March 2004; doi:10.1038/ng1327

ferentiation following developmental stage–specific mediation by Notch-Delta[9]. The frog *Mecp2* transcript targeted for silencing in these experiments is an ortholog of *MECP2*B. In fact, *Mecp2*B appears to be the only form of *Mecp2* in nonmammalian vertebrates (**Supplementary Table 1** online).

The new MeCP2 N terminus is a distinctive 21-amino-acid peptide including polyalanine and polyglycine tracts (MAAAAAAAPS GGGGGGEEERL). A similar N terminus occurs in extracellular signal-regulated kinase-1 (ERK1, or MAPK3; **Supplementary Table 1** online), a key common component of multiple signal transduction pathways. Notably, in neurons, both ERK1 (ref. 10) and MeCP2 (ref. 11) are present in the postsynaptic compartment in addition to the nucleus, and ERK1 translocates between the two compartments to link synaptic activity to transcriptional regulation[10]. It is possible that MeCP2B similarly links synaptic function, in this case neurodevelopmental synaptic-contact guidance, with transcriptional regulation. The only other proteins in which we found consecutive polyalanine and polyglycine tracts are some members[12] of the homeobox family. These, like MeCP2, are developmental transcription regulators.

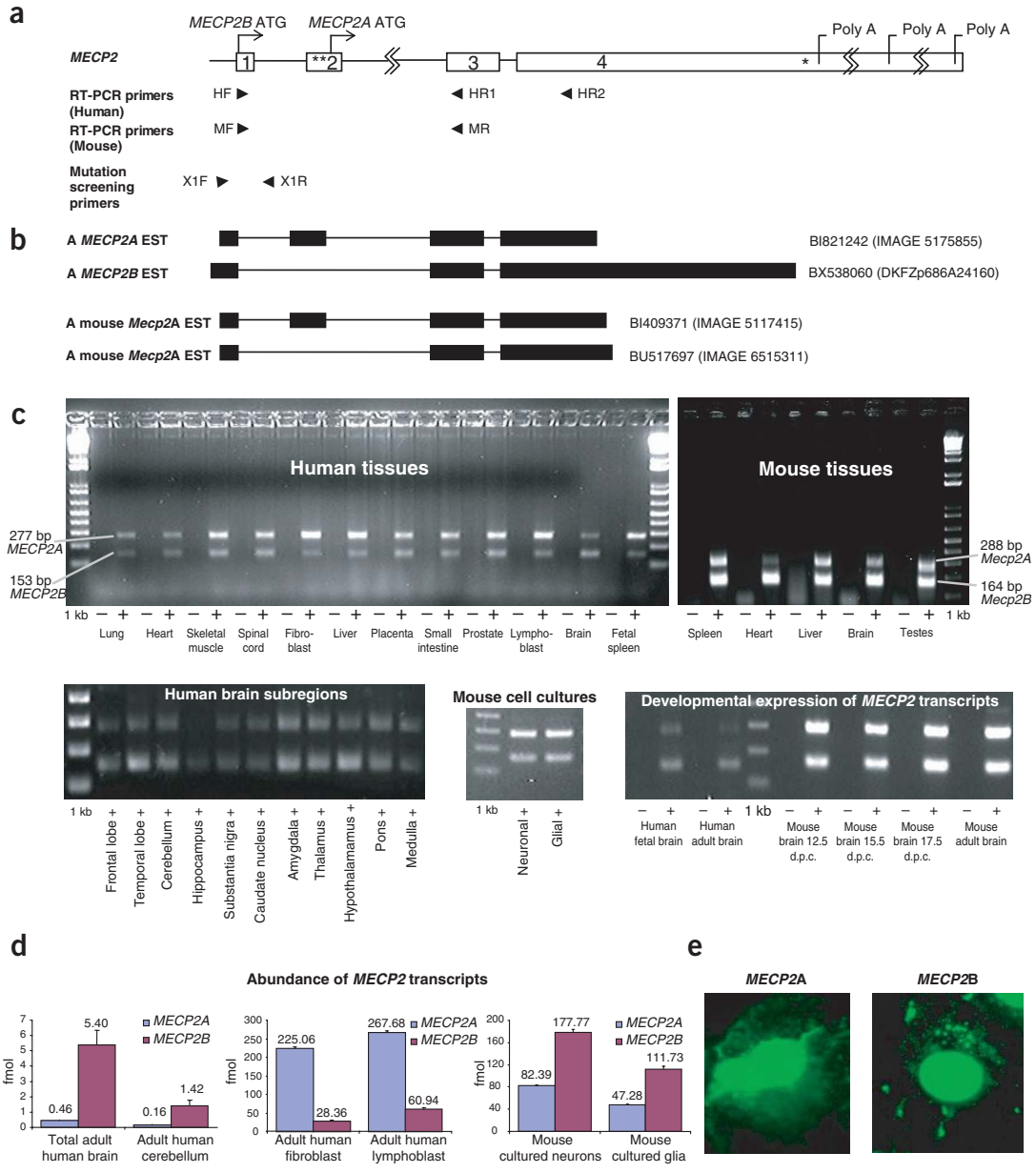

Figure 1 *MECP2* 5′ splice-variant profiles. (**a**) Structure of the *MECP2* gene. Numbered boxes, exons; asterisks, in-frame stop codons. In the traditional *MECP2*A splice variant, the start codon is in exon 2. In *MECP2*B, exon 2 is not present and the start codon is in exon 1. HF, HR1 and MF, MR, human and mouse primer pairs, respectively, used in the RT-PCR experiments illustrated in **c**; HR2, a second human reverse primer, which confirms the results obtained with HR1 (data not shown); X1F, X1R, mutation screening primers (see **Fig. 2**). Primer sequences are available on request. (**b**) Examples of *MECP2* ESTs. (**c**) Results of PCR using HF and HR1, or MF and MR, primers on cDNA from indicated adult tissues (except where noted otherwise) and cell cultures. d.p.c., days post coitum. (**d**) Transcript-specific real-time quantitative PCR (SYBR Green detection) on cDNA from indicated tissues or cell cultures. (**e**) 3′ myc-tagged MeCP2B (and MeCP2A) localize principally in the nucleus and in indeterminate puncti in the cytoplasm. See **Supplementary Methods** online for details.

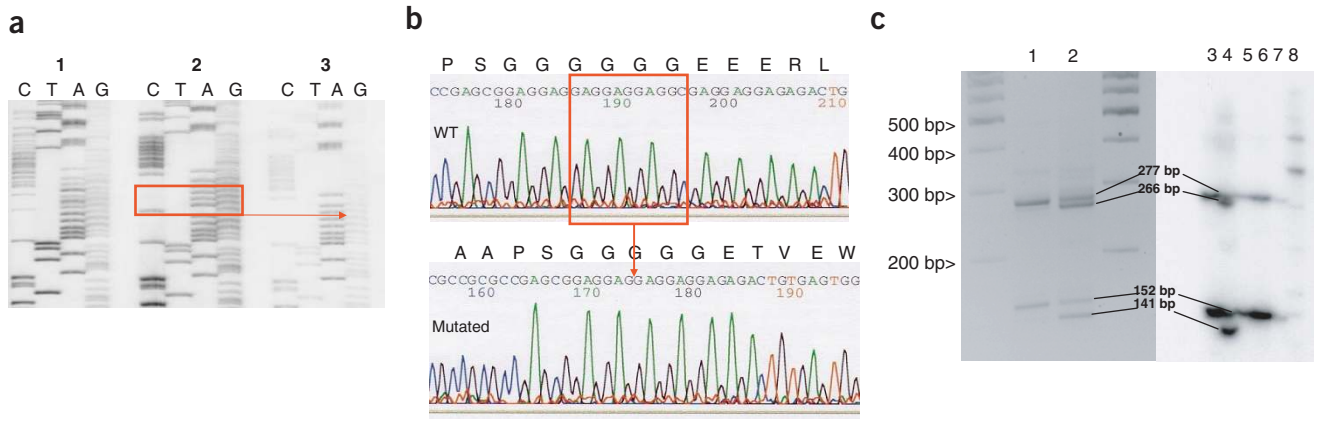

Figure 2 Deletion mutation in individual V1. V1 is presently 27 years of age. She first smiled at 4 weeks, sat at 11 months and walked at 22 months, at which time she had a vocabulary of 6 words. By 30 months, she had lost all her words and all interest in her parents and surroundings. She developed ritualistic behaviors, including continuous hand-wringing, which continue to the present. Intermittent seizures subsided in late adolescence. (**a**) Left, sequence of PCR product from genomic DNA using primers X1F and X1R (**Fig. 1a**). Note mixed sequence. Middle and right, sequences of clones of individual V1's wild-type (WT) and mutant alleles, respectively. Red box, 11 nucleotides deleted in the mutated allele. (**b**) Electropherograms of the same cloned wild-type and deleted alleles. (**c**) PCR on indicated cDNAs using primers HF and HR1 (**Fig. 1a**,**c**). Lanes 1 and 2 (on 2.5% high-resolution agarose gel) are from whole blood from a control and the affected individual, respectively. Lanes 3–8 (on 6% denaturing polyacrylamide) are from control blood (3), affected individual blood (4), control fetal brain (5), control adult brain (6), control testis (7) and control genomic DNA (8). Note that expression of the affected individual's *MECP2*A transcript with the 11-bp exon 1 deletion (band of 266 bp) was not diminished as compared to the nondeleted allele (277 bp). The bands of 141 and 152 bp are the deleted and nondeleted *MECP2*B transcripts, respectively. See **Supplementary Methods** online for details.

Finally, non-inactivating *MECP2* mutations have been associated with phenotypes that overlap RTT, such as mental retardation and autism[13]. The MeCP2 isoform discovered in this study is a candidate for involvement in these disorders.

Accession number. *MECP2*B mRNA, AY541280.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS
This work was funded by the Canadian Institutes of Health Research–Neuromuscular Research Partnership. H.L. is supported by the Sigrid Juselius Foundation, Finland.

COMPETING INTEREST STATEMENT
The authors declare that they have no competing financial interests.

Received 1 December 2003; accepted 10 Februrary 2004

Published online at http://www.nature.com/naturegenetics

1. Hagberg, B. *Ment. Retard. Dev. Disabil. Res. Rev.* **8**, 61–65 (2002).
2. Amir, R.E. *et al. Nat. Genet.* **23**, 185–188 (1999).
3. Miltenberger-Miltenyi, G. & Laccone, F. *Hum. Mutat.* **22**, 107–115 (2003).
4. Reichwald, K. *et al. Mamm. Genome.* **11**, 182–190 (2000).
5. Schouten, J.P. *et al. Nucleic Acids Res.* **30**, e57 (2002).
6. D'Esposito, M. *et al. Mamm. Genome.* **7**, 533–535 (1996).
7. Coy, J.F., Sedlacek, Z., Bachner, D., Delius, H. & Poustka, A. *Hum. Mol. Genet.* **8**, 1253–1262 (1999).
8. Shahbazian, M.D., Antalffy, B., Armstrong, D.L. & Zoghbi, H.Y. *Hum. Mol. Genet.* **11**, 115–124 (2002).
9. Stancheva, I., Collins, A.L., Van den Veyver, I.B., Zoghbi, H. & Meehan, R.R. *Mol. Cell.* **12**, 425–435 (2003).
10. Hardingham, G.E., Arnold, F.J. & Bading, H. *Nat. Neurosci.* **4**, 565–566 (2001).
11. Aber, K.M. *et al. Neuroscience.* **116**, 77–80 (2003).
12. Utsch, B. *et al. Hum. Genet.* **110**, 488–494 (2002).
13. Hammer, S., Dorrani, N., Dragich, J., Kudo, S. & Schanen, C. *Ment. Retard. Dev. Disabil. Res. Rev.* **8**, 94–98 (2002).

Erratum: Touching Base

Nat. Genet. **36, 327 (2004).**

The Executive Director of the IGC expO project (http://www.intgen.org) is Michael Berens.

Erratum: A previously unidentified *MECP2* open reading frame defines a new protein isoform relevant to Rett syndrome

G N Mnatzakanian, H Lohi, I Munteanu, S E Alfred, T Yamada, P J M MacLeod, J R Jones, S W Scherer, N C Schanen, M J Friez, J B Vincent & B A Minassian
Nat. Genet. **36, 339–341 (2004).**

In the bottom half of **Figure 1b**, the second EST is a mouse *Mecp2B* EST (BU517697; IMAGE 6515311).