Human Molecular Genetics, 2002, Vol. 11, No. 7 805–814

Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells

Majid Fardaei, Mark T. Rogers[1], Helena M. Thorpe, Kenneth Larkin, Marion G. Hamshere[2], Peter S. Harper[1] and J. David Brook*

Institute of Genetics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK, [1]Institute of Medical Genetics, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN, UK and [2]School of Life and Environmental Sciences, University of Nottingham, Biological Sciences Building, University Park, Nottingham NG7 2RD, UK

Received December 5, 2001; Revised and Accepted January 25, 2002

Myotonic dystrophy is a complex neuromuscular disorder associated with DNA expansion mutations in two different genes. In DM1 a CTG repeat in the 3′-untranslated region of *DMPK* is expanded, whereas in DM2 an intronic CCTG expansion occurs in the gene *ZNF9*. Transcripts containing expanded repeats form foci in the nuclei of DM1 and DM2 cells. Recent work using antibodies has shown that proteins related to *Drosophila* muscleblind co-localize with repeat foci in DM1 and DM2 cells. We show that rather than there being a single human muscleblind gene producing multiple proteins through alternative splicing, there are in fact three different muscleblind genes, *MBNL*, *MBLL* and *MBXL*, which map to chromosomes 3, 13 and X, respectively, and which show extensive alternative splicing. Two of the genes, *MBNL* and *MBLL*, are expressed in many adult tissues whereas *MBXL* is expressed predominantly in the placenta. Green fluorescent protein-tagged versions of MBNL, MBLL and MBXL co-localize with nuclear foci in DM1 and DM2 cells, suggesting that all three proteins may play a role in DM pathophysiology.

INTRODUCTION

Myotonic dystrophy (DM) is the most common form of muscular dystrophy affecting adults (1). It is characterized by myotonia with muscle weakness and wasting but also involves defects in the heart, ocular and endocrine systems. Two different loci for DM have been identified; DM1 on 19q (2–4) and DM2 on 3q (5). The most common form, DM1, which accounts for ~98% of cases, is caused by the expansion of a trinucleotide repeat (CTG) in the 3′-untranslated region of a protein kinase gene, *DMPK* (6–8). DM2 is caused by the expansion of a tetranucleotide sequence (CCTG) in intron 1 of *ZNF9*, a zinc finger protein gene (5). In both cases transcripts containing expanded repeats form foci that are retained within the nuclei of DM cells (5,9–11). This discovery has considerable implications for the molecular mechanism underlying both forms of the disorder as the phenotypic overlap between DM1 and DM2 is so striking that a common molecular mechanism would appear likely (12). One explanation for the common DM phenotype, the dominant RNA gain-of-function mechanism, is that in both DM1 and DM2 cells the foci of transcripts titrate the same nuclear proteins. One protein, CUG-BP, has been identified by virtue of the fact that it binds to expanded CUG repeats *in vitro* (13,14). Although there is a growing body of evidence to suggest an involvement for CUG-BP in DM (15,16), studies have failed to demonstrate a co-localization of CUG-BP with nuclear foci (17,18). However, support for an RNA gain-of-function mechanism via protein sequestration has come from recent work in DM1 (18–20) and DM2 (5,21). For DM1 it has been shown that the zinc finger protein, MBNL, a mammalian homologue of the *Drosophila* muscleblind protein (22), co-localizes with expanded repeat containing transcripts in nuclear foci (18,19). It has also been shown that the anti-MBNL antibody co-localizes with foci in the nuclei of DM2 cells (21). In this latter study using antibodies (21), it was not possible to distinguish between muscleblind proteins arising through alternative splicing of a single gene and muscleblind proteins encoded by different genes.

In this paper we show that as well as MBNL two related proteins, MBLL and MBXL, co-localize with foci of retained transcripts in the nuclei of cells from both DM1 and DM2 patients. The expression patterns of MBNL and MBLL appear similar, with expression in all tissues tested and highest expression in heart and skeletal muscle. MBXL is predominantly expressed in the placenta. The genes encoding these proteins map to different chromosomes but show similar genomic organization and alternative splice forms. This study showing the sequestration of three muscleblind proteins by expanded repeat-containing transcripts provides a further molecular link

*To whom correspondence should be addressed. Tel: +44 115 8493217; Fax: +44 115 9709906; Email: david.brook@nottingham.ac.uk

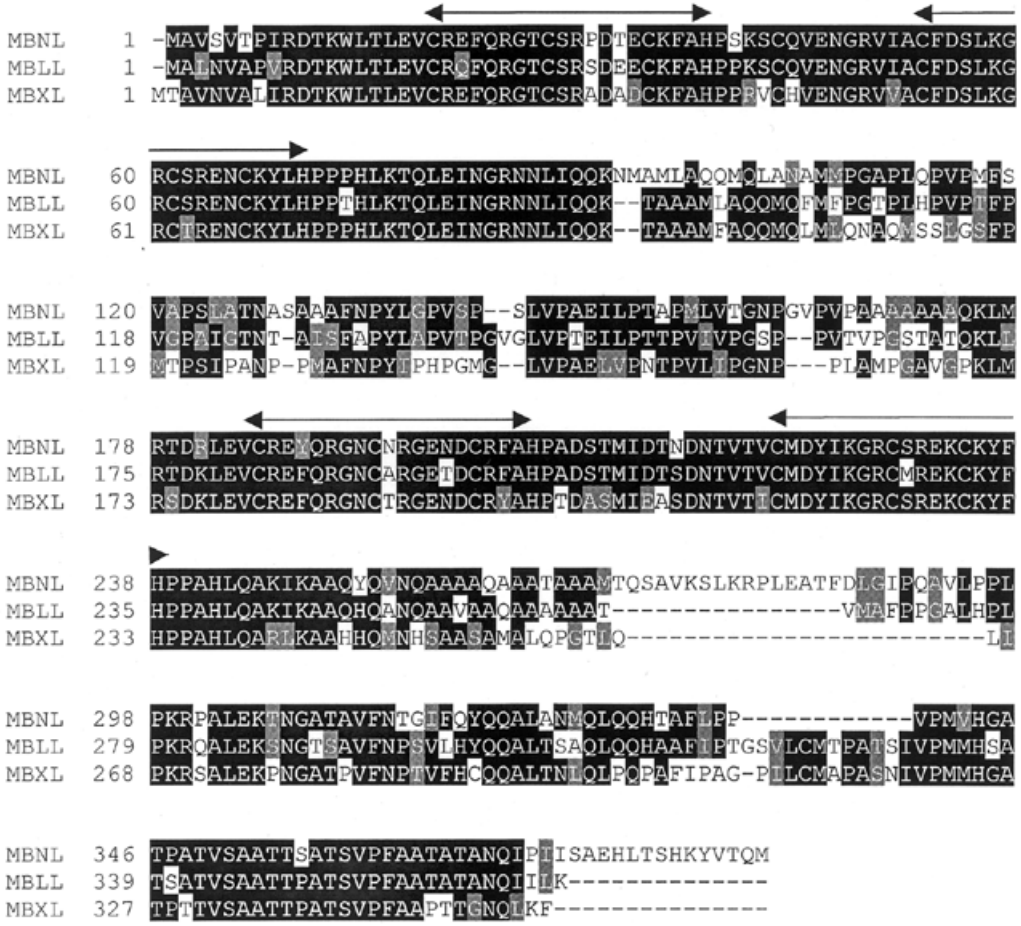

Figure 1. Alignment of the amino acid sequences corresponding to human MBNL, MBLL and MBXL. Sequences were aligned using CLUSTALW and assembled using BOXSHADE (http://www.ch.embnet.org/software/BOX_form.html). The arrows indicate the zinc knuckle domains. Identical amino acids are boxed in black and similar amino acids in grey.

between DM1 and DM2, but also points to increased complexity of sequestered proteins.

RESULTS

Muscleblind proteins are encoded by three different genes

In order to establish whether there are other proteins similar to MBNL, which may be implicated in DM1 and to test their involvement in DM2, we searched the protein database SWall, which contains Swiss-Prot and translated EMBL, with the translated sequence of *MBNL* from clone EXP42/MBNL (19) using the programme Fasta3 (see http://www.ebi.ac.uk). Fourteen strong hits were identified including four human; 095205, Q9NUKO, Q9UF37 and Q9HOZ7 (SWall database numbers), which are virtual proteins containing zinc knuckle domains but which do not correspond to MBNL. The protein sequence 095205 is derived from the translation of NCBI DNA sequence accession no. AF061261, a cDNA that maps to chromosome 13. This protein has been named MBLL, for muscleblind-like protein. The three other hits correspond to the translation of DNA sequence accession no. AK002178 which maps to the X chromosome. We suggest the name MBXL, muscleblind-like-protein-X-linked, for this protein. The locations of *MBLL* and *MBXL* differ from that of *MBNL*, which maps to chromosome 3. 5′- and 3′-rapid amplification of cDNA ends (5′- and 3′-RACE) were employed to extend the cDNA clones encoding MBXL and MBLL, respectively, and the alignment of these proteins to MBNL is shown in Figure 1.

MBLL and MBXL co-localize with foci of expanded-repeat transcripts in DM1

To test whether MBLL and MBXL show similar cellular distributions to MBNL in DM1 cells (18,19), we generated constructs from which green fluorescent protein (GFP)-tagged versions of the two proteins could be produced. Constructs pGFP/MBLL and pGFP/MBXL were individually transfected into control and DM1 cells and the distribution of the fusion protein determined by fluorescence microscopy (Fig. 2). In control cells, GFP/MBLL and GFP/MBXL are distributed throughout the nucleus and cytoplasm with greater concentration in the nucleus (Fig. 2A and E). In DM1 cells GFP/MBLL and GFP/MBXL are again distributed throughout the nucleus and cytoplasm. However, as reported previously for GFP/MBNL (18), both proteins also show multiple nuclear foci that represent areas of increased concentration, which are absent from the nucleus of control cells (Fig. 2B and F). To determine whether these concentrated spots of GFP/MBLL and GFP/MBXL co-localize with the foci of expanded transcripts we

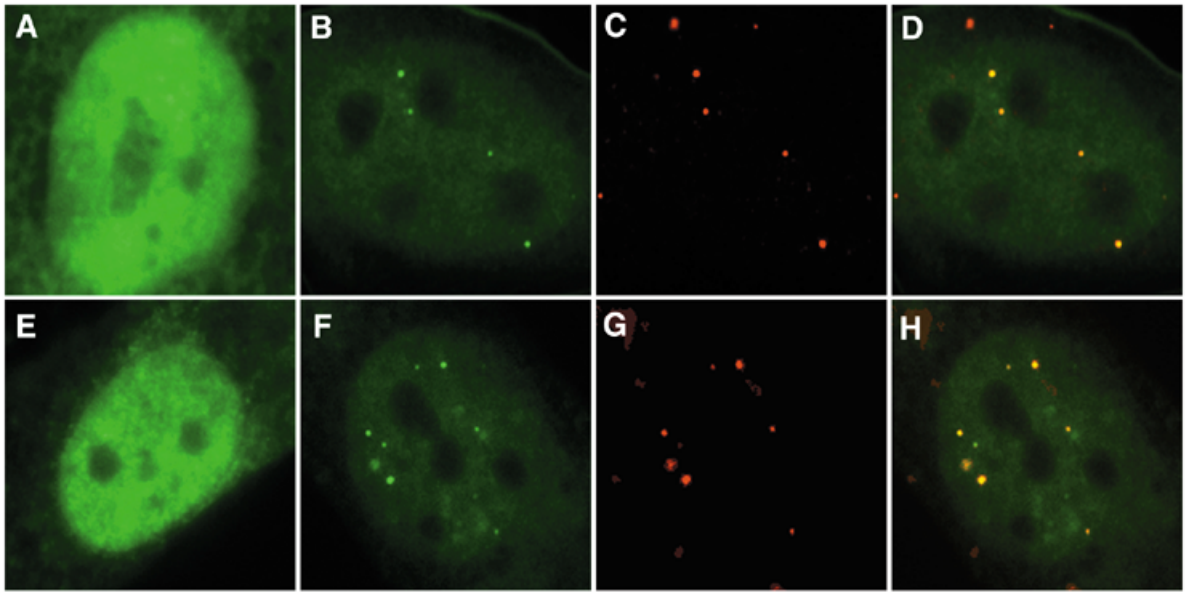

Figure 2. Transient expression of GFP-tagged proteins in control and DM1 fibroblast cells. Cells were transfected with either of two constructs: GFP/MBLL (**A–D**) and GFP/MBXL (**E–H**). Two cell types were used: control fibroblasts (A and E) and DM1 fibroblast cells (B–D and F–H). (A) GFP/MBLL and (E) GFP/MBXL distribution in control cells. (B and F) GFP-tagged protein distribution in DM1 fibroblast cells. (C and G) The location of *DMPK* expanded transcripts using *in situ* hybridization with $(CAG)_{10}$-Cy3 probe. Merged images show that GFP/MBLL (D) and GFP/MBXL (H) co-localize with the foci of expanded *DMPK* transcripts.

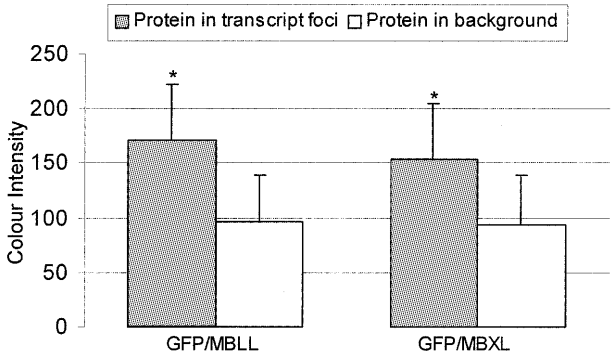

Figure 3. Quantification of GFP/MBLL and GFP/MBXL in DM1 cells. Fluorescence from GFP-tagged proteins was quantified using digital images in which mean intensities were estimated per unit pixel using IP Lab (Scanalytic) software. The levels of GFP/MBLL and GFP/MBXL in the transcript foci of DM1 cells were significantly higher than background. In both cases 62 transcript foci were analysed. $*P < 0.01$ (*t*-test).

measured the intensity of GFP at these sites. Foci of expanded repeat transcripts were visualized by *in situ* hybridization with a $(CAG)_{10}$-Cy3 probe (Fig. 2C and G), and the overlaid images are shown in Figure 2D and H. The intensities of GFP were measured for each protein in 20 cells from two DM1 patients. A significant increase in intensity of GFP, above background, was noted for both GFP/MBLL and GFP/MBXL in regions of the nucleus that contain foci of expanded repeats (Fig. 3). This indicates that in DM1 cells MBLL and MBXL show a similar distribution to that observed for MBNL (18,19) as they co-localize with the foci of expanded CUG-repeat containing transcripts. Occasionally spots were observed that showed either GFP labelling, or Cy3-repeat probe labelling, but not both. This was not common, accounting for <10% of cases. We assume that the absence of Cy3 labelling of some foci is due to incomplete probe penetration and that the Cy3 labelled spots with no GFP are artefacts of the labelling process.

MBNL, MBLL and MBXL co-localize with foci of expanded-repeat transcripts in DM2

It has been reported recently that DM2 is caused by the expansion of a CCTG repeat within the first intron of zinc finger gene *ZNF9* and that transcripts containing this repeat form foci within the nucleus of DM2 cells (5). To examine whether MBNL, MBLL and MBXL co-localize with the foci of CCUG-containing transcripts we performed transfection experiments with GFP-tagged versions of these proteins in DM2 cell lines. We also examined the distribution of two other GFP-tagged proteins, CUG-BP and hnRNP C. GFP-tagged proteins were visualized by fluorescence microscopy and repeat expansion transcripts determined by *in situ* hybridization with a $(CAGG)_{10}$-Cy3 probe. GFP/MBNL, GFP/MBLL and GFP/MBXL are distributed throughout the nucleus and cytoplasm in DM2 cells, but they also show multiple nuclear foci as observed in DM1 cells (Fig. 4A–I). The distribution of GFP/CUG-BP and GFP/hnRNP C in DM2 cells was as observed in DM1 and control cells where proteins were distributed throughout the nucleus and cytoplasm but they did not co-localize with the foci of expanded repeat containing transcripts (Fig. 4J–O). To quantify the co-localization of MBNL, MBLL and MBXL with CCUG repeat-containing foci, we measured the intensity of GFP for each protein in 20 cells from two DM2 patients. Figure 5 shows a significant increase in the intensity of GFP, above the background, for all three proteins in regions of the nucleus that contain foci of expanded repeat transcripts visualized by *in situ* hybridization. The distribution of GFP-tagged CUG-BP and hnRNP C, was the same across the foci as background

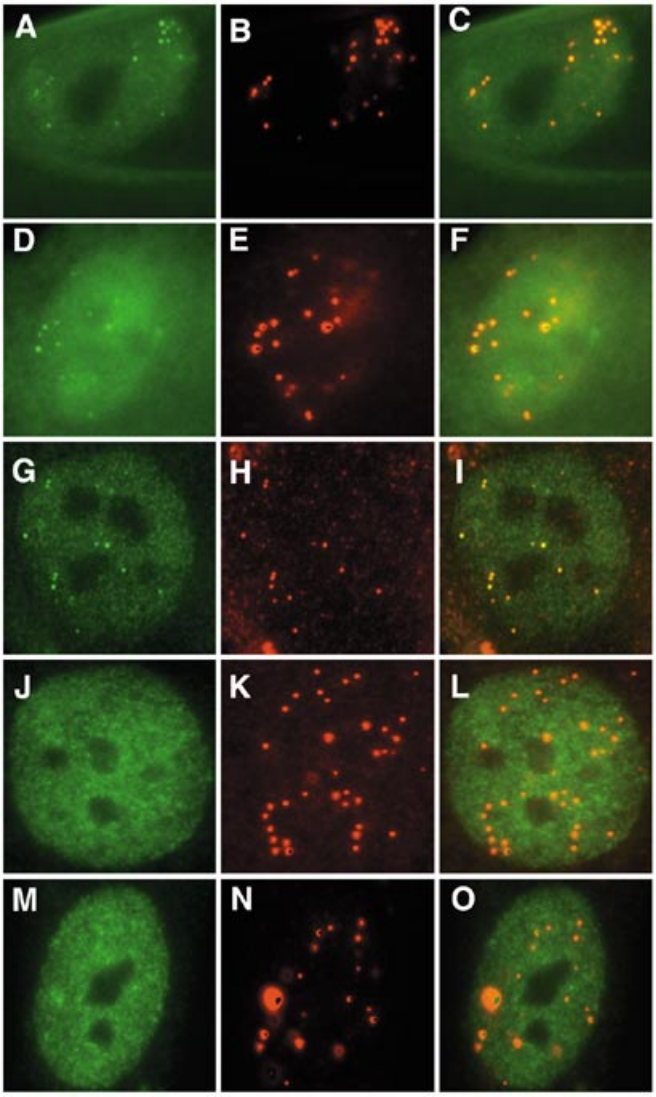

Figure 4. Transient expression of GFP-tagged proteins in DM2 cells. Cells were transfected with five different constructs: GFP/MBNL (**A**–**C**), GFP/MBLL (**D**–**F**), GFP/MBXL (**G**–**I**), GFP/hnRNP C (**J**–**L**) and GFP/CUG-BP (**M**–**O**). (A, D, G, J and M) show the GFP-tagged protein distribution in DM2 fibroblast cells. (B, E, H, K and N) show the location of transcripts containing expanded repeats using *in situ* hybridization with $(CAGG)_{10}$-Cy3 probe. (C, F, I, L and O) are the merged images which show that GFP/hnRNP C (L) and GFP/CUG-BP (O) do not co-localize with foci of expanded repeat transcripts. The merged images of the $(CAGG)_{10}$-Cy3 probe and GFP/MBNL (C), GFP/MBLL (F) and GFP/MBXL (I) show that these proteins co-localize with the foci of transcripts containing expanded repeats.

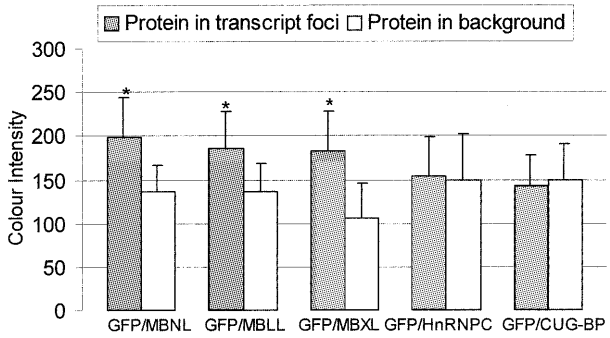

Figure 5. Quantification of GFP-tagged proteins in DM2 cells. Fluorescence from GFP-tagged proteins was quantified as in Figure 3. The number of transcript foci analysed was 92 for GFP/MBNL, 96 for GFP/MBLL and 87 for GFP/MBXL. The levels of these three GFP-tagged proteins in the transcript foci were significantly higher than background in DM2 cells (in each case $*P < 0.01$, *t*-test). There was no significant difference between the levels of protein in the background and at the transcript foci for GFP/hnRNP C (92 foci analysed) or GFP/CUG-BP (101 foci analysed).

(Fig. 5). Thus, as in DM1 cells, MBNL, MBLL and MBXL co-localize with foci of transcripts containing expanded repeats in the nuclei of DM2 cells.

Expression patterns of *MBLL* and *MBXL*

In order to examine the expression of *MBLL* and *MBXL* in various tissues for comparison to that observed for *MBNL* (19), we performed northern blot analysis. Figure 6 shows the tissue distribution of *MBNL*, *MBLL* and *MBXL* expression in heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. *MBNL* identifies two bands of ~6 and 6.5 kb. *MBLL* identifies a transcript of ~5.5 kb, slightly smaller than the predominant transcripts recognized by *MBNL*. Unlike *MBNL*, which shows strongest expression in muscle and heart, *MBLL* shows a similar level of expression across all tissues tested. *MBXL* identifies a transcript of 10 kb and is expressed most strongly in placenta with weak expression in heart, liver, kidney and pancreas.

Genomic organization and alternative splicing of *MBNL*, *MBLL* and *MBXL*

Using expressed sequence tag (EST) database analysis, RT–PCR and comparison to published cDNA (19,21) and genomic sequences we have established the genomic organization of *MBNL*, *MBLL* and *MBXL* and identified more than 10 different splice variants from transcripts of these genes. The genomic organization of *MBNL*, *MBLL* and *MBXL* is similar, and is shown schematically in Figure 7. Exons 1, 2 and 4 contain the highly conserved CCCH zinc finger domains and are present in all forms analysed. Two published forms of *MBNL* lack exon 3 (19), although nine ESTs which cover this region all contain exon 3. Exon 5 is frequently absent from transcripts of *MBNL* and *MBLL*, and we were unable to identify sequence corresponding to this exon in 10 *MBXL* ESTs and in genomic sequence of *MBXL*. Exon 7, the smallest exon, is also often absent from the various forms of muscleblind. The presence/absence of exons 5 and 7 does not affect the reading frame whereas splicing out of either exon 6 or 8, which was observed frequently in *MBLL* (six of 18 ESTs) does affect the reading frame resulting in different stop codons. Altogether we have identified seven forms of *MBNL*, five forms of *MBLL* and three forms of *MBXL*. Two of the splice forms of *MBLL* and two of *MBXL* differ from the seven forms of *MBNL* identified. Those spliceforms that result in frameshifts and therefore, different C-terminal protein sequences are poorly represented in the EST database in comparison to the in-frame spliceforms. The accession numbers of the relevant sequences are shown in the legend to Figure 7. The northern blot data shows a single predominant band for *MBLL* and for *MBXL*, and two major bands for *MBNL*. Most of the alternative splice forms

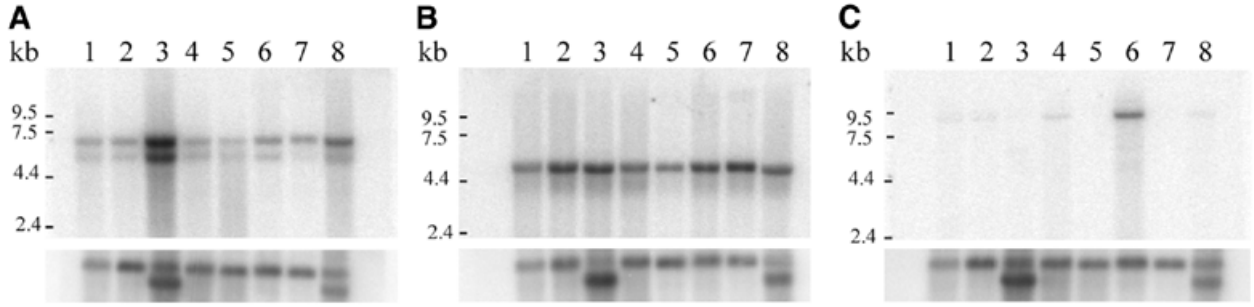

Figure 6. Expression patterns of *MBNL*, *MBLL* and *MBXL*. Northern blots containing ~2 µg of poly(A)$^+$ RNA per lane, hybridized with (**A**) *MBNL*, (**B**) *MBLL* and (**C**) *MBXL*, are shown. Tissues present in each lane are (1) pancreas, (2) kidney, (3) skeletal muscle, (4) liver, (5) lung, (6) placenta, (7) brain and (8) heart. The size markers are shown to the left of each blot and a control probe, human β-actin, is shown in each case to confirm loading.

described here produce size differences of 50–200 bp, which are difficult to resolve on northern blots, and would probably appear as a single broad band. The composition of the two *MBNL* bands cannot be reconciled at this time with the splice patterns observed.

MBNL, MBLL and MBXL and other zinc finger proteins

MBNL, MBLL and MBXL contain CCCH zinc finger motifs. In order to identify other similar proteins we have searched protein sequence databases at the NCBI web site. From the resulting list, sequences were selected for comparison to MBNL, MBLL and MBXL on the basis of the presence of one or more CCCH zinc fingers. Plant, bacterial, yeast and hypothetical protein sequences were excluded from the analysis. The evolutionary relationships of these proteins was examined by sequence alignment using ClustalW (23). Figure 8 shows a neighbour-joining tree (24) of CCCH protein sequences. Several groups of proteins were identified, which differed in the number of zinc fingers present and the spacing of cysteine residues within the zinc fingers. The MBL proteins are of unknown function and contain four CCCH fingers in two pairs with the cysteines spaced $CX_7CX_6CX_3H$—$CX_7CX_4CX_3H$. A second group includes zebrafish and human No-arches, and *Drosophila*, mouse, bovine and human Clipper, which are RNA hairpin cleavage proteins containing five CCCH fingers with cysteines spaced $CX_7CX_5CX_3H$—$CX_7CX_5CX_3H$—$CX_8CX_4CX_3H$—$CX_7CX_5CX_3H$—$CX_7CX_5CX_3H$. Another group includes proteins such as tristetraprolin (TTP) and butyrate response factor (BRF) which are involved in the destabilization of mRNA and contain two CCCH fingers both of which conform to the pattern $CX_8CX_5CX_3H$. Other proteins with two CCCH zinc fingers of the type $CX_8CX_5CX_3H$—$CX_7CX_5CX_3H$ include members of the U2AF 35 kDa subunit family, which are involved in mRNA splicing. The mouse protein Fliz is unusual in that it has three CCCH fingers in which the cysteines are spaced $CX_7CX_5CX_3H$—$CX_7CX_5CX_3H$—$CX_8CX_5CX_3H$. Human and *Drosophila* unkempt proteins have six CCCH fingers with cysteines spaced $CX_7CX_7CX_3H$—$CX_9CX_5CX_3H$—$CX_9CX_6CX_3H$—$CX_6CX_5CX_3H$—$CX_8CX_5CX_3H$—$CX_8CX_5CX_3H$.

DISCUSSION

Three different molecular mechanisms have been suggested to explain the pathophysiology of DM1 including reduction in levels of DMPK (25–27), effects on the expression of neighbouring genes (28–32) and the dominant RNA model (9–11,33). Although each mechanism may contribute to the DM phenotype, recent work has provided compelling data for the RNA dominance mechanism (5,19,20). For example, transgenic mice expressing an untranslated CUG repeat in an unrelated gene develop myotonia and myopathy (20), something that has also been observed in transgenic mice carrying the CTG expansion in its human DM1 context (34), and a protein, MBNL, similar to *Drosophila* mbl (22), has been identified that is recruited to the foci of DMPK transcripts containing expanded repeats in the nuclei of DM1 cells (19). The significance of this model has been further enhanced by the discovery that DM2 is caused by the expansion of a tetranucleotide repeat within an intron of a gene, *ZNF9*, and that transcripts containing this expanded repeat also show nuclear retention (5). Furthermore, it has been shown that muscleblind proteins also co-localize to nuclear foci in DM2 (21). This study (21), which preceded the identification of the DM2 mutation, employed a CAG probe to visualize transcript foci and an antibody to identify muscleblind proteins. Thus it was not possible to establish whether the human muscleblind proteins reported (21) were the product of alternative splicing of transcripts from a single gene, or produced by different genes. In the present study we show that three different human genes, *MBNL*, *MBLL* and *MBXL*, map to different chromosomes and encode muscleblind-related proteins, which co-localize with foci of transcripts in the nuclei of cells from DM1 and DM2 patients. Each of the proteins exists in multiple isoforms due to extensive alternative splicing of the respective transcripts. The precise role of MBNL, MBLL and MBXL in normal and disease processes awaits further investigation.

DM1 and DM2 share many clinical features (35,36) and it is only recently that they have been recognized as distinct genetic entities (37). Both show anticipation (38) and are associated with muscle weakness and myotonia, with cardiac conduction defects and smooth muscle involvement. In both, endocrine features include insulin resistance or frank diabetes mellitus, male hypogonadism and hypothyroidism (36,39), and central nervous system involvement includes cognitive impairment and hypersomnolence. Polychromatic posterior lens cataracts, of a type once thought to be pathognomic for DM1 alone, are also present in both. There are however differences including the distribution of weakness, which is usually more distal in DM1 and proximal in DM2, myotonia, which is generally

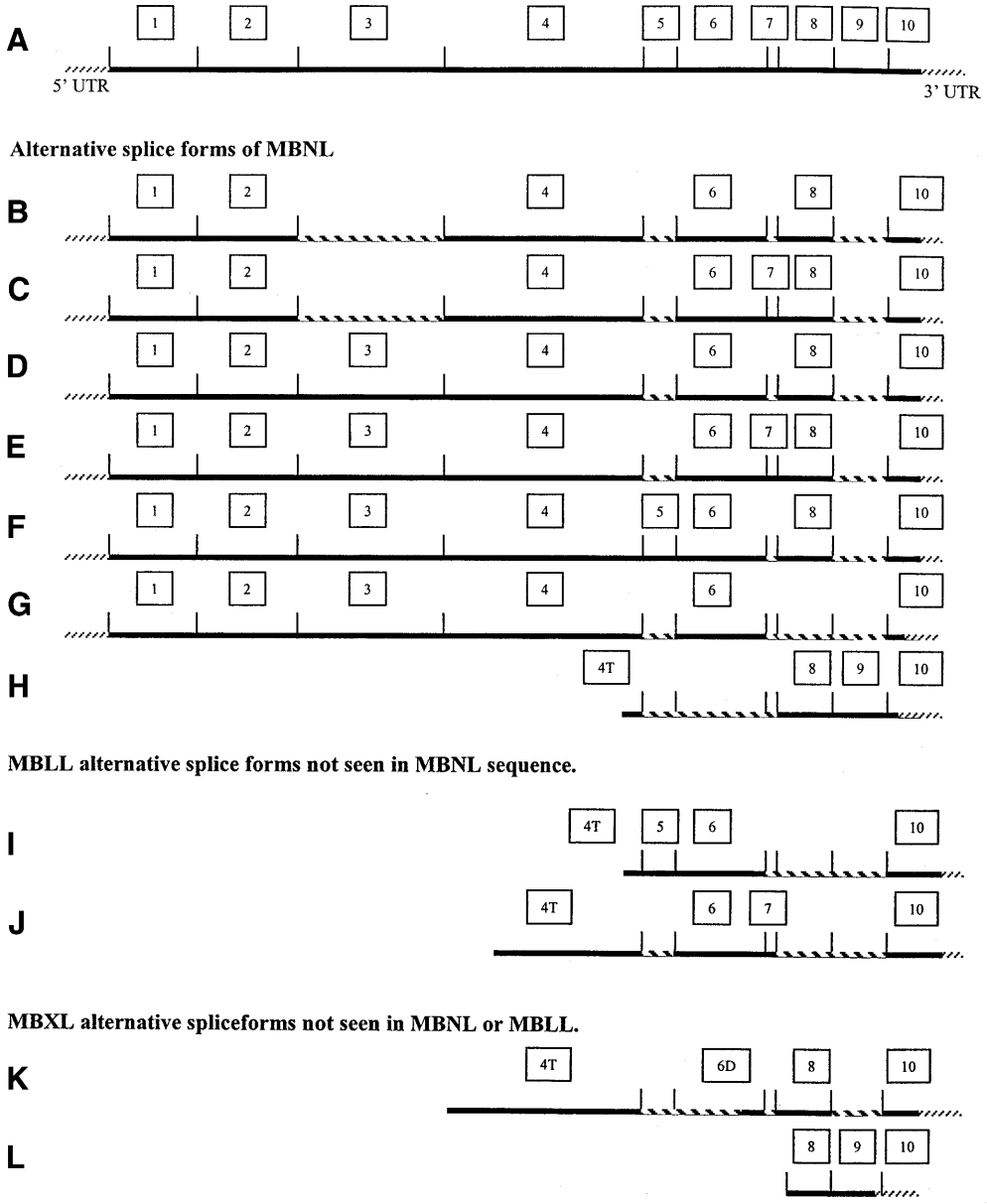

Figure 7. Genomic organization of *MBNL*, *MBLL* and *MBXL*. Schematic representation of the genomic organization of human muscleblind genes showing all possible exons (**A**). This hypothetical 'full length' splice form has not been observed. Alternative splice forms of *MBNL* are illustrated (**B**–**H**). Spliceforms (B) (accession no. AF255334), (C) (AF395876), (D) (NM_021038), (E) (AF401998) and (F) (Y13829) have been reported elsewhere (19,21). Spliceforms (G), which is an RT–PCR product (unpublished data), and (H), sequence from an EST (AA595519), use alternative stop codons in exon 10. Spliceforms (**I**) (AA019370) and (**J**) (BG702549), which also use an alternative stop codon in exon 10, are novel forms identified only in *MBLL* ESTs. Also represented in the *MBLL* EST collection are spliceforms with the same 3′ organization as *MBNL* forms (D) (BF940381), (F) (BF432190) and (G) (AW968217). All forms of *MBXL* surveyed lacked exon 5. Spliceforms (**K**) (BG259690) and (**L**) (BE302133) were both identified in EST databases. (K) uses an alternative splice site in the middle of exon 6 that results in an in-frame deletion of the 5′ end of this exon and (L) uses an alternative stop codon within exon 9, with exon 10 untranslated. A form of *MBXL* (BM069866) that is similar to the 3′ end of *MBNL* form (F) was also identified in the EST database.

more obvious in DM1, and the congenital form, which has not yet been described in DM2. The extent to which the sequestration of muscleblind proteins is responsible for the shared aspects of the DM1 and DM2 phenotypes remains to be established. However, as the titration of these three proteins is common to both forms of DM it represents a good starting point for further investigation. It will also be interesting to establish the extent to which the subtle distinction in phenotype may be explained by differences in the expression patterns of *DMPK* and *ZNF9*, or whether other molecular mechanisms such as reduced expression of these genes following repeat expansion, or effects on neighbouring genes, are responsible.

One of the emerging features of the molecular basis of DM1 is the dysregulation of alternative splicing. Three recent pieces of work point to effects on the distribution of alternative splice forms of cardiac troponin T (*cTNT*) (15), insulin receptor (*IR*) (16) and tau protein (40) within DM1 tissues. The splicing of *cTNT* is affected in the striated muscle of DM patients and in

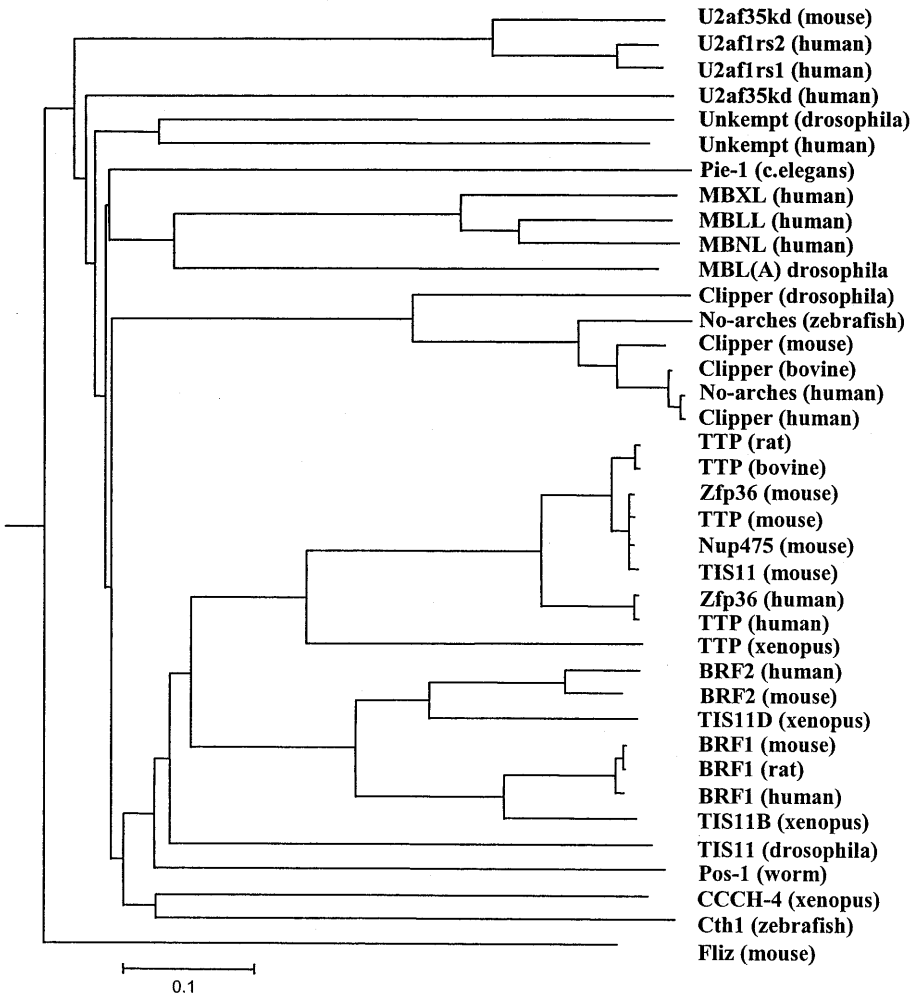

Figure 8. Multiple alignment of CCCH zinc finger proteins. Thirty-eight full-length CCCH zinc finger protein sequences were retrieved from the protein database at the NCBI website and aligned using a multiple sequence alignment method, from which a bootstrapped neighbour joining tree was constructed. Fliz, a nuclear protein of unknown function, with three zinc fingers, was selected as the out-group for the construction of this tree. The length of the branch represents time to common ancestor. The scale is shown with 0.1 equivalent to 10% sequence difference.

normal cells expressing transcripts containing expanded CTG repeats (15). Also in DM1 skeletal muscle the alternative splicing of insulin receptor pre mRNA is altered with the predominant expression of the lower signalling non-muscle isoform. CUG-BP has been implicated in both examples as it has been shown to regulate the alternative splicing of *cTNT* and its overexpression in cultured cells induces a switch to the lower signalling isoform of *IR* via interaction with an intronic element upstream of alternatively spliced exon 11. It will be important to establish whether the isoform shifts reported for DM1 are also found in DM2 and how this relates to the sequestration of muscleblind proteins. In addition, the relationship between CUG-BP and the muscleblind proteins requires further investigation. Clearly the functional inactivation of the muscleblind proteins in knockout mice will be necessary to establish whether the sequestration of these proteins represents a critical event in DM pathophysiology rather than a coincidental epi-phenomenon that is not involved in the disease pathway. The identification of three different muscleblind genes, two of which are expressed in a wide range of tissues, raises the possibility of functional redundancy amongst the muscleblind proteins, something that could compound single-gene knockout analysis in the mouse. Interestingly in *Drosophila*, homozygous mutations that inactivate mbl produce a disruption of muscle Z-band structures (41), which may reflect a similar underlying process to that responsible for the Z-band disruption observed in electron microscopic studies of DM patient muscle (42).

The CCCH family of zinc finger proteins is extensive and diverse, not only with respect to the number and spacing of the fingers but also with respect to the spacing of the cysteine residues within each of the fingers. It will be important to establish whether any other members of the CCCH family of zinc finger proteins co-localize to foci of repeat transcripts and what features of these proteins confer this specificity. It is clear that the sequestration of the muscleblind-related proteins provides a common molecular mechanism underlying both forms of DM. However, future studies to understand the molecular basis of DM will have to allow for the possibility of functional redundancy amongst these proteins in humans and mice.

MATERIALS AND METHODS

Databases and sequence analysis

Database searches were performed through the National Centre for Biotechnology Information (NCBI) and the European Bioinformatics Institute (EBI). Protein database searches identified SWall protein accession nos 095205, Q9NUKO, Q9UF37 and Q9HOZ7. Accession no. 095205 corresponds to NCBI protein accession no. AAC67242, which is derived from the translation of DNA sequence submission no. AF061261. The three other hits, Q9NUKO, Q9UF37 and Q9HOZ7, are identical to NCBI protein accession no. BAA92124, which is derived from the translation of DNA sequence accession no. AK002178. The multiple sequence alignment for MBNL, MBLL and MBXL, was produced with the ClustalW program using a default gap-opening penalty of 10 and a gap extension penalty of 0.05.

The genomic organization of all three genes was established by direct comparison of cDNA sequences with the human genomic sequences using the blastn program (43). The cDNA sequences used were Y13829 (*MBNL*), AF061261 (*MBLL*) and AK002178 (*MBXL*). These sequences for *MBLL* and *MBXL* do not contain the full-length cDNA; this was established by using the predicted full-length protein sequence of MBNL (CAA74155) to analyse the sequence of the human chromosome corresponding to *MBLL* or *MBXL* using the tblastn program. Full-length cDNA sequence was then used to identify alternative splice forms by searching the human EST database using the blastn program. Alternative splice forms were then translated using Translation Machine to establish alternative STOP codon usage. Websites used were www.ensembl.org, www.ncbi.nlm.nih.gov and www.ebi.ac.uk.

GFP fusion constructs

To produce a GFP/MBLL fusion construct, a PCR product was amplified from IMAGE clone 1683133 (HGMP Resource Centre), corresponding to bases 783–1538 of NCBI DNA sequence accession no. AF061261, using primers 5′-TACTTTGGTACCATGGCTTTGAACGTTGCCCCAG-3′ and 5′-ATAGGATCCGCGGCCGCGGCTGCGCAGC-3′ and cloned into pGEM-T Easy (Promega). The construct was sequence verified and the insert excised as a *Kpn*I/*Bam*HI fragment and subcloned into pEGFP-C1 (Clontech) at the same sites. This GFP/MBLL construct was verified by sequencing.

To generate a GFP/MBXL fusion construct a two-stage cloning strategy was employed. First a PCR product corresponding to bases 31–961 of NCBI DNA sequence accession no. AK002178 was amplified from IMAGE clone 3391556 (HGMP Resource Centre) using primers 5′-GGGGGATCCTGGGCGAATTCCATGGAGAGGGC-3′ and 5′-CTGATAGTCGACAACTCTGCTGTTCAG-3′. The resulting PCR product was digested with *Eco*RI and *Sal*I, and subcloned into pEGFP-C1 at the same sites to produce pEGFP-MBXL3′. The 5′ part of *MBXL* was amplified by PCR from HeLa cell cDNA (18) using primers 5′-TTTTGCGAATTCTATGACGGCTGTCAATGTTGCCCTGATTCG-3′ and 5′-GGAGGATGAAAGTACTTGCATTTCTCCCGC-3′. The resulting PCR fragment was cloned into pCR2.1-ToPo (Invitrogen) to give pTOPO-MBXL5′ and sequenced. To produce the complete GFP/MBXL construct, the *Eco*RI/*Sca*I fragment from pEGFP-MBXL3′ was replaced with the *Eco*RI/*Sca*I fragment from pTOPO-MBXL5′ and the final construct was sequenced.

Other GFP constructs used in these experiments, GFP/CUG-BP, GFP/hnRNP C and GFP/MBNL, have been described previously (18). The tagged versions of MBNL and MBXL correspond to isoforms G and E respectively, in Figure 7. The tagged version of MBLL contains alternatively spliced exon 3, but is truncated after exon 4.

Northern blot analysis

Human northern blot membranes (MTN blot, Clontech) containing RNA from eight different tissues were hybridized with three different probes for *MBNL*, *MBLL* and *MBXL*. The hybridization probes were generated by PCR. For *MBNL*, primers 5′-TGGACGCAATAACTTGATTC-3′ and 5′-CCTCAAGTCTGTCTGTTCGC-3′ were used with pGFP/MBNL as template, generating a probe corresponding to bases 757–1061 of NCBI accession no. Y13829. For *MBLL* the primers were 5′-CGGTTACGGTGTTGTCACTTG-3′ and 5′-CTGCAGCAGCAATGCTTGC-3′ with pGFP/MBLL as template. This probe corresponds to bases 1057–1431 of NCBI accession no. AF06126. For *MBXL* primers 5′-AGCAGATGCAGCTTATGCTC-3′ and 5′-CCAACAGCTCCTGGCATTGC-3′ were used with pGFP/MBXL as template, to generate a probe corresponding to bases 184–388 of NCBI accession no. AK002178. Probe labelling with ^{32}P was performed as described previously (6) and filters hybridized for 17 h at 65°C. Filters were washed at room temperature for 30 min in 2× SSC and 30 min at 65°C in 1× SSC. The radioactive signal was detected using a PhosphorImager (Molecular Dynamics) following a 19 h exposure. The membrane was subsequently rehybridized with a β-actin cD-NA probe as an internal control.

Cell culture and transfections, *in situ* hybridization and fluorescent microscopy

The repeat sizes in the DM1 cells are 8.2 and 10.4 kb. Cell culture, transfections, *in situ* hybridization and fluorescent microscopy were performed as described previously (18). In the fluorescence *in situ* hybridization studies a Cy3-conjugated $(CAG)_{10}$ oligonucleotide was used for DM1 cells and a Cy3-conjugated $(CAGG)_{10}$ oligonucleotide was used for DM2 cells.

CCCH protein analysis

A multiple sequence alignment using ClustalW (23) was performed on 38 CCCH zinc finger protein sequences retrieved from the protein database at the NCBI. Pairwise distances among sequences were estimated with a correction for multiple hits. The tree was constructed by the neighbour-joining method (24) applied to the pairwise distances; 1000 bootstrap samples were examined. Spacing of the cysteine residues within the zinc fingers was analysed by visual comparison of sequences in the alignment.

ACKNOWLEDGEMENTS

The authors wish to thank Tom Cooper (University of Texas) for his gift of a plasmid containing CUG-BP, Steve Cross for help with the sequence searches, and the DM patients and their

families for support and for assisting with this research. This work was supported by the Muscular Dystrophy Association, the Muscular Dystrophy Campaign, the Muscular Dystrophy Support Group and the Human Frontiers Science Programme Organisation.

REFERENCES

1. Harper,P.S. (2001) *Myotonic Dystrophy*, 3rd edn. Saunders, London and Philadelphia.
2. Aslanidis,C., Jansen,G., Amemiya,C., Shutler,G., Mahadevan,M., Tsilfidis,C., Chen,C., Alleman,J., Wormskamp,N.G., Vooijs,M. *et al.* (1992) Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature*, **355**, 548–551.
3. Buxton,J., Shelbourne,P., Davies,J., Jones,C., Van Tongeren,T., Aslanidis,C., de Jong,P., Jansen,G., Anvret,M., Riley,B. *et al.* (1992) Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature*, **355**, 547–548.
4. Harley,H.G., Brook,J.D., Rundle,S.A., Crow,S., Reardon,W., Buckler,A.J., Harper,P.S., Housman,D.E. and Shaw,D.J. (1992) Expansion of an unstable DNA region and phenotypic variation in myotonic dystrophy. *Nature*, **355**, 545–546.
5. Liquori,C.L., Ricker,K., Moseley,M.L., Jacobsen,J.F., Kress,W., Naylor,S.L., Day,J.W. and Ranum,L.P. (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science*, **293**, 864–867.
6. Brook,J.D., McCurrach,M.E., Harley,H.G., Buckler,A.J., Church,D., Aburatani,H., Hunter,K., Stanton,V.P., Thirion,J.P., Hudson,T. *et al.* (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3′ end of a transcript encoding a protein kinase family member. *Cell*, **68**, 799–808.
7. Fu,Y.H., Pizzuti,A., Fenwick,R.G.,Jr, King,J., Rajnarayan,S., Dunne,P.W., Dubel,J., Nasser,G.A., Ashizawa,T., de Jong,P. *et al.* (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*, **255**, 1256–1258.
8. Mahadevan,M., Tsilfidis,C., Sabourin,L., Shutler,G., Amemiya,C., Jansen,G., Neville,C., Narang,M., Barcelo,J., O'Hoy,K. *et al.* (1992) Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene. *Science*, **255**, 1253–1255.
9. Taneja,K.L., McCurrach,M., Schalling,M., Housman,D. and Singer,R.H. (1995) Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J. Cell Biol.*, **128**, 995–1002.
10. Davis,B.M., McCurrach,M.E., Taneja,K.L., Singer,R.H. and Housman,D.E. (1997) Expansion of a CUG trinucleotide repeat in the 3′ untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc. Natl Acad. Sci. USA*, **94**, 7388–7393.
11. Hamshere,M.G., Newman,E.E., Alwazzan,M., Athwal,B.S. and Brook,J.D. (1997) Transcriptional abnormality in myotonic dystrophy affects DMPK but not neighboring genes. *Proc. Natl Acad. Sci. USA*, **94**, 7394–7399.
12. Tapscott,S.J. and Thornton,C.A. (2001) Perspectives: reconstructing myotonic dystrophy. *Science*, **293**, 816–817.
13. Timchenko,L.T., Miller,J.W., Timchenko,N.A., DeVore,D.R., Datar,K.V., Lin,L., Roberts,R., Caskey,C.T. and Swanson,M.S. (1996) Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res.*, **24**, 4407–4414.
14. Timchenko,L.T., Timchenko,N.A., Caskey,C.T. and Roberts,R. (1996) Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Hum. Mol. Genet.*, **5**, 115–121.
15. Philips,A.V., Timchenko,L.T. and Cooper,T.A. (1998) Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science*, **280**, 737–741.
16. Savkur,R.S., Philips,A.V. and Cooper,T.A. (2001) Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat. Genet.*, **29**, 40–47.
17. Michalowski,S., Miller,J.W., Urbinati,C.R., Paliouras,M., Swanson,M.S. and Griffith,J. (1999) Visualization of double-stranded RNAs from the myotonic dystrophy protein kinase gene and interactions with CUG-binding protein. *Nucleic Acids Res.*, **27**, 3534–3542.
18. Fardaei,M., Larkin,K., Brook,J.D. and Hamshere,M.G. (2001) *In vivo* co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucleic Acids Res.*, **29**, 2766–2771.
19. Miller,J.W., Urbinati,C.R., Teng-Umnuay,P., Stenberg,M.G., Byrne,B.J., Thornton,C.A. and Swanson,M.S. (2000) Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J.*, **19**, 4439–4448.
20. Mankodi,A., Logigian,E., Callahan,L., McClain,C., White,R., Henderson,D., Krym,M. and Thornton,C.A. (2000) Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. *Science*, **289**, 1769–1773.
21. Mankodi,A., Urbinati,C.R., Yuan,Q.P., Moxley,R.T., Sansone,V., Krym,M., Henderson,D., Schalling,M., Swanson,M.S. and Thornton,C.A. (2001) Muscleblind localizes to nuclear foci of aberrant RNA in myotonic dystrophy types 1 and 2. *Hum. Mol. Genet.*, **10**, 2165–2170.
22. Begemann,G., Paricio,N., Artero,R., Kiss,I., Perez-Alonso,M. and Mlodzik,M. (1997) Muscleblind, a gene required for photoreceptor differentiation in *Drosophila*, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development*, **124**, 4321–4331.
23. Higgins,D.G. and Sharp,P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, **73**, 237–244.
24. Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
25. Reddy,S., Smith,D.B., Rich,M.M., Leferovich,J.M., Reilly,P., Davis,B.M., Tran,K., Rayburn,H., Bronson,R., Cros,D. *et al.* (1996) Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat. Genet.*, **13**, 325–335.
26. Jansen,G., Groenen,P.J., Bachner,D., Jap,P.H., Coerwinkel,M., Oerlemans,F., van den Broek,W., Gohlsch,B., Pette,D., Plomp,J.J. *et al.* (1996) Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat. Genet.*, **13**, 316–324.
27. Berul,C.I., Maguire,C.T., Aronovitz,M.J., Greenwood,J., Miller,C., Gehrmann,J., Housman,D., Mendelsohn,M.E. and Reddy,S. (1999) DMPK dosage alterations result in atrioventricular conduction abnormalities in a mouse myotonic dystrophy model. *J. Clin. Invest.*, **103**, R1–R7.
28. Thornton,C.A., Wymer,J.P., Simmons,Z., McClain,C. and Moxley,R.T.,III (1997) Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat. Genet.*, **16**, 407–409.
29. Klesert,T.R., Otten,A.D., Bird,T.D. and Tapscott,S.J. (1997) Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat. Genet.*, **16**, 402–406.
30. Alwazzan,M., Newman,E., Hamshere,M.G. and Brook,J.D. (1999) Myotonic dystrophy is associated with a reduced level of RNA from the DMWD allele adjacent to the expanded repeat. *Hum. Mol. Genet.*, **8**, 1491–1497.
31. Sarkar,P.S., Appukuttan,B., Han,J., Ito,Y., Ai,C., Tsai,W., Chai,Y., Stout,J.T. and Reddy,S. (2000) Heterozygous loss of Six5 in mice is sufficient to cause ocular cataracts. *Nat. Genet.*, **25**, 110–114.
32. Klesert,T.R., Cho,D.H., Clark,J.I., Maylie,J., Adelman,J., Snider,L., Yuen,E.C., Soriano,P. and Tapscott,S.J. (2000) Mice deficient in Six5 develop cataracts: implications for myotonic dystrophy. *Nat. Genet.*, **25**, 105–109.
33. Wang,J., Pegoraro,E., Menegazzo,E., Gennarelli,M., Hoop,R.C., Angelini,C. and Hoffman,E.P. (1995) Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. *Hum. Mol. Genet.*, **4**, 599–606.
34. Seznec,H., Agbulut,O., Sergeant,N., Savouret,C., Ghestem,A., Tabti,N., Willer,J.C., Ourth,L., Duros,C., Brisson,E. *et al.* (2001) Mice transgenic for the human myotonic dystrophy region with expanded CTG repeats display muscular and brain abnormalities. *Hum. Mol. Genet.*, **10**, 2717–2726.
35. Ricker,K., Koch,M.C., Lehmann-Horn,F., Pongratz,D., Speich,N., Reiners,K., Schneider,C. and Moxley,R.T.,III (1995) Proximal myotonic myopathy. Clinical features of a multisystem disorder similar to myotonic dystrophy. *Arch. Neurol.*, **52**, 25–31.
36. Meola,G. (2000) Clinical and genetic heterogeneity in myotonic dystrophies. *Muscle Nerve*, **23**, 1789–1799.
37. Ricker,K., Koch,M.C., Lehmann-Horn,F., Pongratz,D., Otto,M., Heine,R. and Moxley,R.T.,III (1994) Proximal myotonic myopathy: a new dominant disorder with myotonia, muscle weakness, and cataracts. *Neurology*, **44**, 1448–1452.

38. Schneider,C., Ziegler,A., Ricker,K., Grimm,T., Kress,W., Reimers,C.D., Meinck,H., Reiners,K. and Toyka,K.V. (2000) Proximal myotonic myopathy: evidence for anticipation in families with linkage to chromosome 3q. *Neurology*, **55**, 383–388.
39. Phillips,M.F., Rogers,M.T., Barnetson,R., Braun,C., Harley,H.G., Myring,J., Stevens,D., Wiles,C.M. and Harper,P.S. (1998) PROMM: the expanding phenotype. A family with proximal myopathy, myotonia and deafness. *Neuromusc. Disord.*, **8**, 439–446.
40. Sergeant,N., Sablonniere,B., Schraen-Maschke,S., Ghestem,A., Maurage,C.A., Wattez,A., Vermersch,P. and Delacourte,A. (2001) Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1. *Hum. Mol. Genet.*, **10**, 2143–2155.
41. Artero,R., Prokop,A., Paricio,N., Begemann,G., Pueyo,I., Mlodzik,M., Perez-Alonso,M. and Baylies,M.K. (1998) The muscleblind gene participates in the organization of Z-bands and epidermal attachments of *Drosophila* muscles and is regulated by Dmef2. *Dev. Biol.*, **195**, 131–143.
42. Ludatscher,R.M., Kerner,H., Amikam,S. and Gellei,B. (1978) Myotonia dystrophica with heart involvement: an electron microscopic study of skeletal, cardiac, and smooth muscle. *J. Clin. Pathol.*, **31**, 1057–1064.
43. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.