sequences nor re-grow on the SD/−His/−Leu/+3-AT medium. On the other hand, the reporter strain containing the bait DNA in a reverse orientation (R4-3-2-1, Fig. 1*A*) gave no positive clone.

To refine the binding regions of three representative clones (pACT2 clones 2, 8, and 11), we further carried out the yeast one-hybrid assay using four yeast reporter strains containing the bait DNA of R1 (−364 to −214), R2 (−231 to −114), R3 (−131 to +51), and R4 (+29 to +158) DNA fragment (Fig. 1*A*), respectively. Both pACT2 clone 2 and pACT2 clone 8 revealed strong binding to the R2 region in the promoter of the *HD* gene (Fig. 1*B*). In addition, they also showed weak binding activity to the R1 region, indicating that they bind to an overlap region (−231 to −214) between R1 and R2 or that there are several distinct binding sites within R1 and R2 regions. On the other hand, pACT2 clone 11 was shown to associate with the R3 region (data not shown). As both pACT2 clone 2 and pACT2 clone 8 revealed specific binding to the promoter region of the *HD* gene, we characterized these two clones in detail.

Isolation of the Transcripts Encoding HDBP1 and HDBP2—Based upon the sequence analysis, the 1.3-kb cDNA of pACT2 clone 2 appears to be a part of transcript; therefore, we screened further a human testis cDNA library using the 1.3-kb cDNA as a probe to obtain a full-length transcript. Two independent cDNA clones, whose inserts are ~1.5 and 1.8 kb long, respectively, were identified. DNA sequence analysis revealed that the 1.8-kb cDNA encompassed 1,776 nucleotides (nt) with a single ORF encoding a 46-kDa protein of 387 amino acids (GenBank™ accession no. AB052777). As 1.8 kb of the cDNA is equivalent to the size of the transcript represented by Northern blotting (see Fig. 2*A*), the 1.8-kb cDNA might contain the entire sequence of the transcript. We thus designated this gene product as HDBP1 (Huntington's disease gene regulatory region-binding protein 1).

DNA sequencing of the 1.5-kb cDNA (pACT2 clone 8) revealed that the termination codon was included within the sequence, but no authentic initiation codon for methionine was found. Therefore, 5′-RACE using a human testis cDNA library was carried out to obtain the most upstream of the 5′ sequence of the transcript including the initiation codon and 5′-untranslated region (UTR). Several clones comprising the initiation codon and the 5′-UTR were identified (data not shown). The composite DNA sequence of pACT2 clone 8 transcript encompasses 1,874 nt with a single ORF and is predicted to encode a 62-kDa protein consisting of 513 amino acids (GenBank™ accession no. AB073627). Northern blot analysis demonstrated that the transcript for the pACT2 clone 8 was ~5.0 kb long (Fig. 2*B*), indicating that this composite cDNA sequence was still missing a portion of 3′-UTR. Despite of this fact, pACT2 clones 8 obviously contain the entire coding region, and the gene product was designated as HDBP2.

Tissue Distribution of mRNAs Encoding HDBP1 and HDBP2—The expression level and distribution of the transcripts encoding HDBP1 and HDBP2 were analyzed by Northern blot analysis with mRNAs extracted from various human tissues. Both transcripts are ubiquitously expressed, but with significantly less expression in brain and ileum (Fig. 2, *A* and *B*).

Data Base Searches and Deduced Amino Acid Sequences of HDBP1 and HDBP2—The genes coding for HDBP1 and HDBP2 have been named and registered as *SLC2A4* and *ZNF395*, respectively (HUGO Gene Nomenclature Committee). Data base search of the deduced amino acid sequences for HDBP1 and HDBP2 was performed using a BLASTP program. The middle region of HDBP1 amino acid sequence (387 aa), corresponding to 68–293 aa, was found to be an exact match of GEFdb (GLUT4 enhancer factor DNA-binding domain) (19) (Fig. 3*A*). The 236-amino acid sequence for GEFdb filed is apparent that it is deduced from a partial mRNA sequence and lacks an initiation codon and also most part of the N-terminal sequences as well as corresponding amino acid sequences to the C-terminal region of HDBP1. Thus, GEFdb may be a potential splice variant of HDBP1. It is also revealed that HDBP2 is the same protein that has been known as papillomavirus binding factor PBF, which play a role in papillomavirus genes' transcription and in E2-mediated repression (20). Homology search in conjunction with the multiple alignments of the protein sequences showed that the overall homology of the amino acid sequences between HDBP1/GEFdb and HDBP2/PBF was significantly high with a 36% identity (Fig. 3*A*, DNASIS-Mac 3.5). Moreover, mouse glucocorticoid-induced gene 1 (GIG1) (GenBank™ accession no. AF292939) was found to be highly homologous to both HDBP1and HDBP2 (Fig. 3*A*). HDBP1, HDBP2, and GIG1 proteins share the similar domain structures: a single zinc finger domain (C_2H_2 type), several putative nuclear localization signals (NLSs) and nuclear export signals (NESs), serine-rich region (SRR), proline-rich region (PRR), and three highly conserved regions (CR1, CR2, and CR3), implying functional similarity to one another. The predicted CR1, CR2, and CR3 peptide sequences among the proteins are 54–70, 68–86, and 84–86% identities, respectively (Fig. 3*B*). Taken together, these data imply that HDBP1 and HDBP2 belong to a novel protein family.

The C-terminal Conserved Region (CR3) of HDBP1 and HDBP2 Interacts with DNA—To confirm the DNA-protein interaction observed in the yeast one-hybrid screening, we examined whether HDBP1 and HDBP2 proteins bind to the promoter region of the *HD* gene *in vitro*. GST fusion HDBP1 and HDBP2 proteins prepared from *E. coli* were subjected to EMSA using 4 *HD* promoter DNA fragments (R1, R2, R3, and R4; Fig. 1*A*) as probes. Equal amounts of GST-fused HDBP1 and HDBP2 peptides were used in this binding experiment (Fig. 5*A*, *arrowhead*). By the addition of the varied amount of purified GST-HDBP1, the intensity of signals representing DNA-protein complexes was increased in a protein-concentration-dependent manner with fixed amount of ^{32}P-labeled R2 and R3 probes, but not with R1 or R4 probes (Fig. 4*A*). To prove the specific bindings, a competition assay was performed. Both GST-HDBP1 bindings to the R2 and R3 probes were preferentially competed by the addition of unlabeled R2 but not significantly by R3 (Fig. 4B). The binding of GST-HDBP1 to R3 probe implies as a result of an overlap of 19 bp on each probe or some sequence similarity between R2 and R3 (detailed below). EMSA and the competition analysis also revealed the specific binding of GST-HDBP2 to the R2 probe, but not to the R1, R3, or R4 (Fig. 4, *C* and *D*). These results are consistent with the result of the one-hybrid screening (Fig. 1*B*). However, weak binding activities to the R1 region in one-hybrid experiment (Fig. 1*B*) were not confirmed by the *in vitro* binding experiments. It could be because of a weak and/or unstable interaction of HDBP1/HDBP2 with R1. In fact, there is an overlap of 18 bp on R1 and R2, and one of the triplicate DNA binding sequences resides in the overlapping region (detailed in the following section).

To define HDBP1/HDBP2 DNA-binding domains, which mediate the interaction with R2 DNA fragment, we prepared three truncated peptides of HDBP1 fused with GST, GST-HDBP1-N (1–105 aa), GST-HDBP1-Zn (106–321 aa), and GST-HDBP1-C (322–387 aa), and three truncated peptides of HDBP2 fused with GST, GST-HDBP2-N (1–168 aa), GST-HDBP1-Zn (169–437 aa), and GST-HDBP1-C (438–513 aa) for the EMSA. DNA-protein complexes were formed with only GST-HDBP1-C but not GST-HDBP1-N or GST-HDBP1-Zn

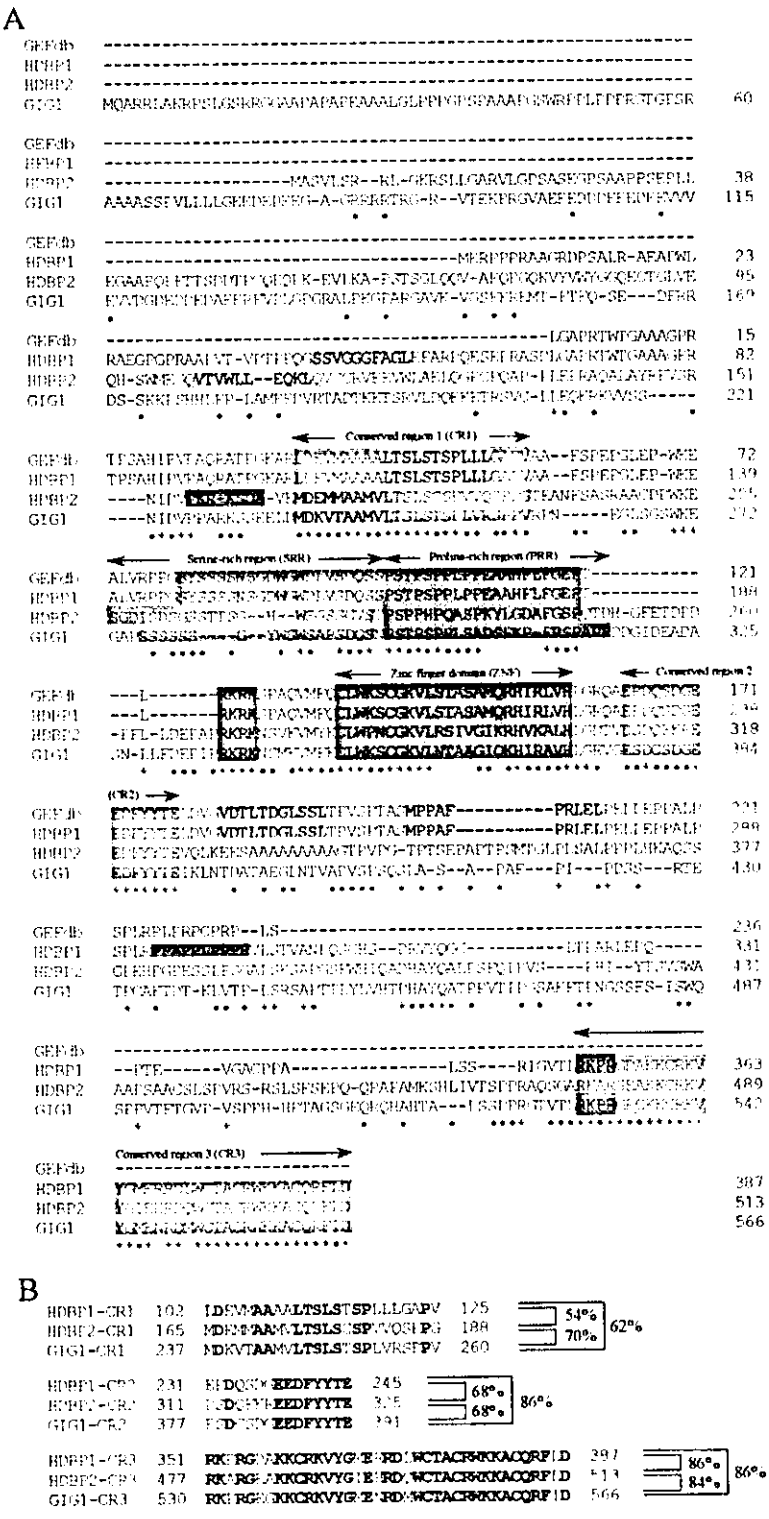

FIG. 3. **Comparison of the deduced amino acid sequences of GEFdb, HDBP1, HDBP2, and GIG1.** *A*, amino acid sequence alignment was carried out using multiple alignment system of GENETYX-MAC. Amino acid residues matched in three or four proteins are shown with *asterisk* at the *bottom* of the alignment, and those matched in two proteins are shown with *dot* at the *bottom* of the alignment. *Dashes* represent gaps in the sequence alignment. The sequences are *numbered* with the predicted initiation methionine codon of HDBP1, HDBP2, or GIG1 designated 1. Most N-terminal residue of GEFdb was designated as 1, as it was deduced from the partial cDNA sequence. Putative NES motifs, putative NLS motifs, a serine-rich region (*SRR*), a proline-rich region (*PRR*), a single zinc-finger domain (*ZNF*), and three CR are represented by the *gray*, *red*, *orange*, *purple*, *green*, and *blue shadings*, respectively. *B*, multiple amino acid sequence alignment of CR1, CR2, and CR3 among HDBP1/GEFdb, HDBP2, and GIG1. Percentage of the identities in CRs between HDBP1 and HDBP2, HDBP2, and GIG1, and HDBP1 and GIG1 are shown on the *right* of the sequence alignment. Amino acid residues that matched among all three proteins are shown in *bold letters*.

(Fig. 5*B*). The EMSA using a series of truncated HDBP2 peptides also represented the DNA-protein complexes formation with the C-terminal peptide, GST-HDBP2-C (438–513 aa), but not others (Fig. 5*C*). For further analysis of the DNA-binding domains, we prepared the GST fusion peptides consisting of the conserved C-terminal amino acid sequences (CR3; 37 aa) of HDBP1 (351–387 aa) and HDBP2 (477–513 aa). Both CR3 peptides were found to retain binding activities to the R2 probe (Fig. 5*D*). Moreover, these complex formations were specifically competed out by the addition of unlabeled R2 (Fig. 5*E*). These results demonstrate that highly conserved CR3 region in the C terminus of HDBP1 and HDRF2 is the DNA-binding domain and directly interacts with the R2 region of the *HD* promoter. As a protein sequence for CR3 matches to none of the publicly available motif/domains listed so far, CR3 might be a novel DNA-binding domain.

HDBP1 and HDBP2 Bind to the Unique DNA Sequence—To determine the DNA sequence that interacts with HDBP1 and

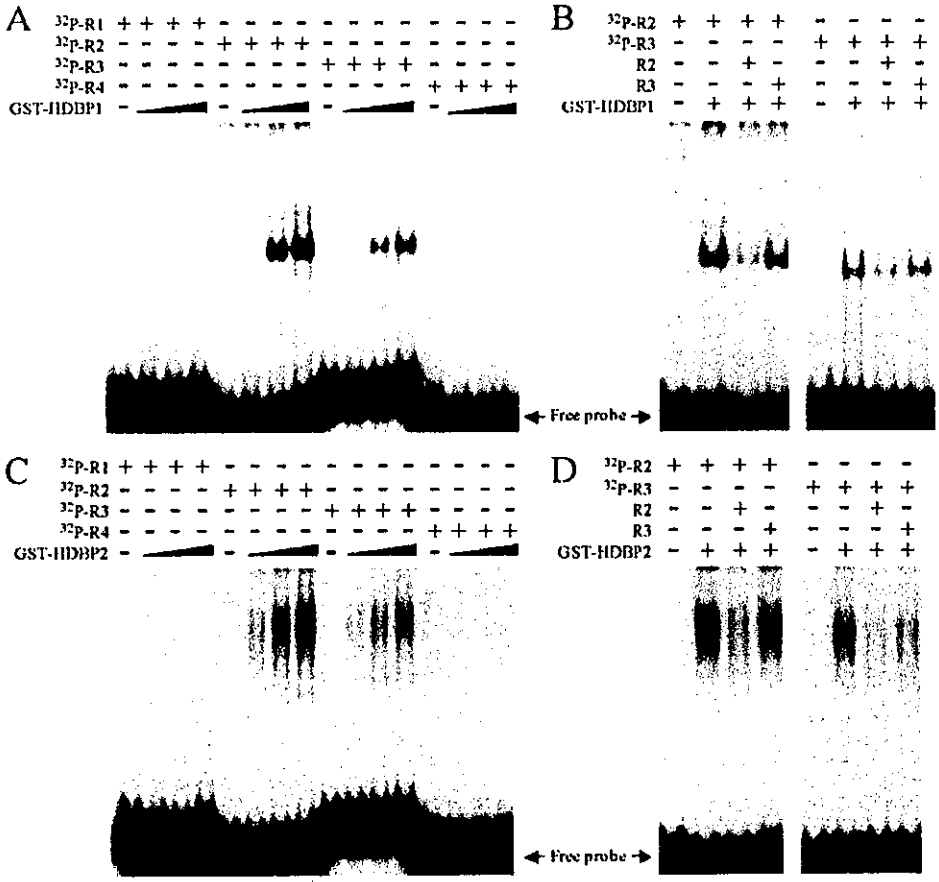

FIG. 4. **EMSA analysis of the promoter region of the *HD* gene.** EMSA was performed using four MluI-EcoRI fragments, R1, R2, R3, and R4, which were labeled with [α-^{32}P]dATP at their 3'-EcoRI end. *A*, three different amounts (20, 100, and 200 ng) of partially purified GST-HDBP1 (see Fig. 5*A*) were incubated with each ^{32}P-labeled DNA for 20 min at room temperature. *B*, for the competition assay, the reaction mixture containing 200 ng of GST-HDBP1 was incubated with 50-fold molar excess of unlabeled R2 or R3 for 5 min at room temperature before the addition of ^{32}P-labeled R2 or R3. *C*, three different amounts (20, 100, and 200 ng) of partially purified GST-HDBP2 (see Fig. 5*A*) were incubated with each ^{32}P-labeled DNA for 20 min at room temperature. *D*, for the competition assay, the reaction mixture containing 200 ng of GST-HDBP2 was incubated with 50-fold molar excess of unlabeled R2 or R3 for 5 min at room temperature prior to the addition of ^{32}P-labeled R2 or R3. DNA-protein complexes were separated on a 4% polyacrylamide gel by electrophoresis. The positions of the free probe are shown.

HDBP2, DNase I footprint analysis was achieved using a 253-bp DNA fragment (R1-2), corresponding to nucleotides −365 to −113 of the *HD* promoter sequence, in the presence of GST-HDBP1-C and GST-HDBP2-C. The footprint patterns exhibited the strongly protected sequences at three regions (nucleotides −221 to −215, −201 to −195, and −181 to −175) (Fig. 6) and a weakly protected sequence (nucleotide −142 to −136; data not shown) by the addition of either GST-HDBP1-C or GST-HDBP2-C. All three strongly protected sequences are the same as the consensus sequence (C/G)GCCGGCG(C/T). One copy of the consensus sequence resides in each of the duplicated 20-bp direct repeats, and the third one resides at proximal to the repeat. The 20-bp direct repeat sequence has been postulated as an enhancer element for HD transcription; thus, the 7-bp consensus sequence in triplicate at 13-nucleotide intervals is a potential *cis*-regulatory core element.

The weakly protected sequence (nucleotide −142 to −136) is GCCGGGA. In addition, another similar sequence as GCCGGGC is in R3 (nucleotide −18 to −12). This single copy of 7-bp sequences in R3 might be the cause of the weak interaction with HDBP1 and HDBP2 *in vitro* experiments as described above.

To confirm further the binding between the 7-bp consensus sequence and HDBP1/HDBP2 molecules is specific, we generated radioactive 86-bp DNA fragments carrying with a wild type sequence of the 7-bp consensus sequences (wt-7bp) and a mutated 7-bp consensus sequence of GCAAGCG in all of three consensus sequences (m3-7bp) respectively, and performed EMSA with purified GST-HDBP1-C and GST-HDBP2-C. The mutation introduced into three consensus sequences abolished DNA-HDBP1/HDBP2 complexes formation (Fig. 7*A*). The bindings of GST-HDBP1 and GST-HDBP2 to the wt-7bp probe were competed by the addition of excess amount of the unlabeled wt-7bp, but not by the unlabeled m3-7bp (Fig. 7*B*), indicating that the 7-bp consensus sequence was the specific site recognized by HDBP1 and HDBP2.

According to the binding analysis of PBF, PBF recognized the sequence CCGG in papillomavirus genomes (20). However, it is apparent to our *in vitro* binding assay that HDBP2 does not bind to the sequence CCGG in R1 (nucleotides −345 to −342 and −220 to −217) (Fig. 4*C*).

7-bp Consensus Sequence Is Essential for the HD Promoter Activity in a Neuronal Cell Line—The study of the *HD* promoter has demonstrated that the 20-bp direct repeat in the promoter is required for enhanced transcription (18). In this report, we identified a novel *cis*-regulatory core element with the 7-bp consensus sequence repeat in triplicate including the 20-bp direct repeat (Fig. 6). To know whether the 7-bp consensus sequence repeat in triplicate is required and acts as a *cis*-regulatory element for the *HD* promoter function *in vivo*, we carried out luciferase reporter gene assays in a non-neuronal (HeLa; human epitheloid carcinoma) and a neuronal (IMR32;

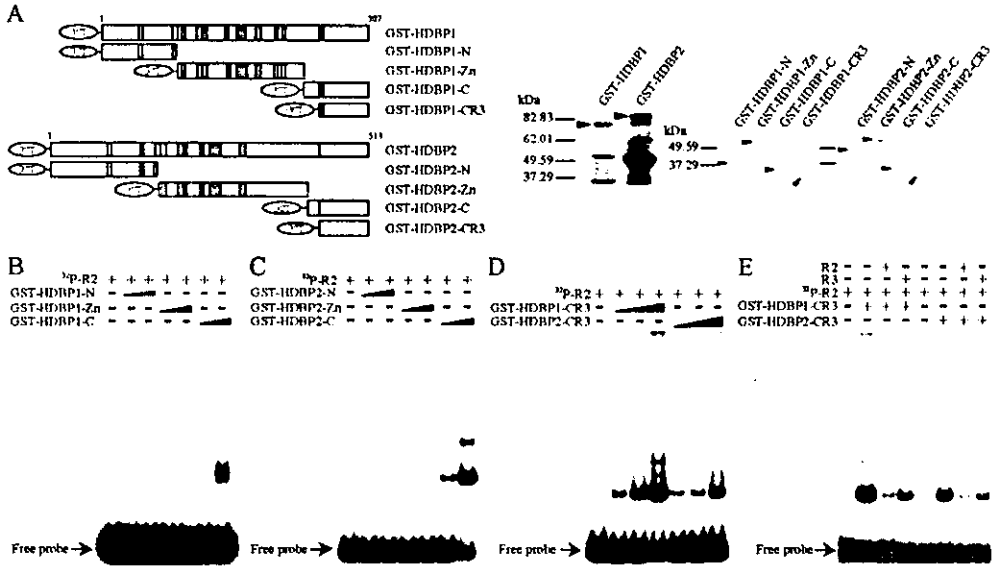

FIG. 5. **Identification of a DNA-binding domain in HDBP1 and HDBP2.** *A*, schematic representation of the GST-fused full-length and truncated forms of HDBP1 and HDBP2 proteins: GST-HDBP1 (1–387 aa), GST-HDBP1-N (1–105 aa), GST-HDBP1-Zn (106–321 aa), GST-HDBP1-C (322–387 aa), GST-HDBP1-CR3 (351–387 aa), GST-HDBP2 (1–513 aa), GST-HDBP2-N (1–168 aa), GST-HDBP2-Zn (169–437 aa), GST-HDBP2-C (438–513 aa), and GST-HDBP2-CR3 (477–513 aa). The equal amount of GST-HDBP1 and GST-HDBP2 (200 ng each) or GST-HDBP1 (2)-N, -Zn, -C, and -CR3 (50 ng each), which were confirmed by SDS-PAGE (Ready Gels, 5–20% (Bio-Rad)), were used in *in vitro* binding assay. *B–E*, EMSA analysis of the *HD* promoter using various fragments of GST fusion HDBP proteins; *B*, 10 and 50 ng of GST-HDBP1-N, GST-HDBP1-Zn, or GST-HDBP1-C; *C*, 10 and 50 ng of GST-HDBP2-N, GST-HDBP2-Zn, or GST-HDBP2-C; *D*, 10, 20, and 50 ng of GST-HDBP1-CR3 or GST-HDBP2-CR3. The reaction solution containing each GST fusion protein was incubated with ^{32}P-labeled R2 for 20 min at room temperature. *E*, for the competition analysis, 50 ng of GST-HDBP1-CR3 or GST-HDBP2-CR3 was incubated with a 50-fold molar excess of unlabeled R2 or R3 for 5 min at room temperature prior to the addition of ^{32}P-labeled R2. DNA-protein complexes were separated on a 4% polyacrylamide gel by electrophoresis. The positions of the free probe are indicated.

human neuroblastoma) cell lines. The reporter constructs with the wild-type and mutated 7-bp consensus sequences as described above were generated (pGV-E2-wt-7bp/Luc, pGV-E2-m3-7bp/Luc; Fig. 8A) and transfected to the cells. Each of the promoter activities in HeLa and IMR32 cells was represented as a relative luciferase activity normalized to the control sea pansy luciferase activity (Figs. 8, *B* and *C*). The *HD* promoter activity in HeLa cells was not significantly affected by the mutations of the 7-bp consensus sequence (Fig. 8*B*); however, the promoter activity in IMR32 cells relied on the 7-bp consensus sequence (Fig. 8*C*). Thus, the *HD* promoter tends to comply with the 7-bp consensus sequence in a neuronal cell line but not in a non-neuronal cell line. These results imply that the 7-bp consensus sequence act as an essential *cis*-regulatory element in neuronal cells for the transcription steered by the *HD* promoter.

Subcellular Localization of Ectopically Expressed HDBP1/HDBP2 in HeLa and IMR32 Cells—To investigate the subcellular localization of HDBP1 and HDBP2, we generated N-terminal fusion of HDBP1 and HDBP2 to GFP (HDBP1-GFP and HDBP2-GFP) (Fig. 9*A*). The transient expression in HeLa cells exhibits that both proteins localize in the cytoplasm (Fig. 9*B*, *control (−)*), and a nuclear localization of the GFP fusion proteins is observed in a small number of cells. On the other hand, an immunohistochemical analysis of the endogenous HDBP1 and HDBP2 is in preparation, and is waiting for antibodies against the proteins.

Several of the putative NES and NLS motifs are found in both HDBP1 and HDBP2; we examined whether these motifs are genuine. HeLa cells expressing HDBP1-GFP and HDBP2-GFP were treated with LMB, a specific inhibitor of the NES receptor Crm1 (21, 22). LMB treatment (Fig. 9*B*, *+LMB*) resulted in nuclear localizations of HDBP1-GFP and HDBP2-GFP, implying that these proteins could be imported into the nucleus by their own NLS function and retained within the nucleus by inhibiting the Crm1-dependent nuclear exports.

We next examined the subcellular localization of HDBP1-GFP and HDBP2-GFP in IMR32 cells. When HDBP1-GFP and HDBP2-GFP were expressed transiently, both proteins localized in the cytoplasm (Fig. 9*C*, *control (−)*) and in the nuclei (Fig. 9*C*, *LMB*) under non-treated and treated conditions with LMB, respectively. Consequently, distributions of HDBP1 and HDBP2 in IMR32 cells were identical to those in HeLa cells. These results demonstrate that both HDBP1 and HDBP2 possess functional NES as well as NLS, and shuttle between cytoplasm and nuclei through a NLS/NES-dependent pathway in either neuronal/non-neuronal cells.

Identification of Functional NES—The identification of a genuine NES among multiple NES motifs deduced in HDBP1 and HDBP2 was achieved by site-directed NES mutagenesis. HDBP1 contains four putative NES motifs. The N-terminal peptide (1–198 aa) carrying first and second putative NESs and the C-terminal peptide (199–387 aa) carrying third and fourth putative NESs were expressed in HeLa and IMR32 cells as GFP fusion proteins. The N-terminal peptide localized only in the nuclei (Fig. 10*A* and data not shown), suggesting that both first and second NESs are non-functional. This result parallels the function of NLS in N-terminal region of HDBP1. In fact, HDBP1 protein sequence search identified a candidate NLS at 187–193 aa. To confirm whether the third and/or fourth NES of HDBP1 are functional, we mutated both Leu-257 and Leu-260 to alanine residues obtaining the third NES mutant (HDBP1-L257A/L260A). We also mutated both Phe-273 and Leu-276 to alanine residues generating the fourth NES mutant (HDBP1-F273A/L276A). The ectopic expression of each C-terminal GFP fusion mutant protein in HeLa and IMR32 cells revealed that HDBP1-L257A/L260A localized in the nuclei, whereas HDBP1-F273A/L276A was in the cytoplasm (Fig. 10 (*A* and *C*) and data

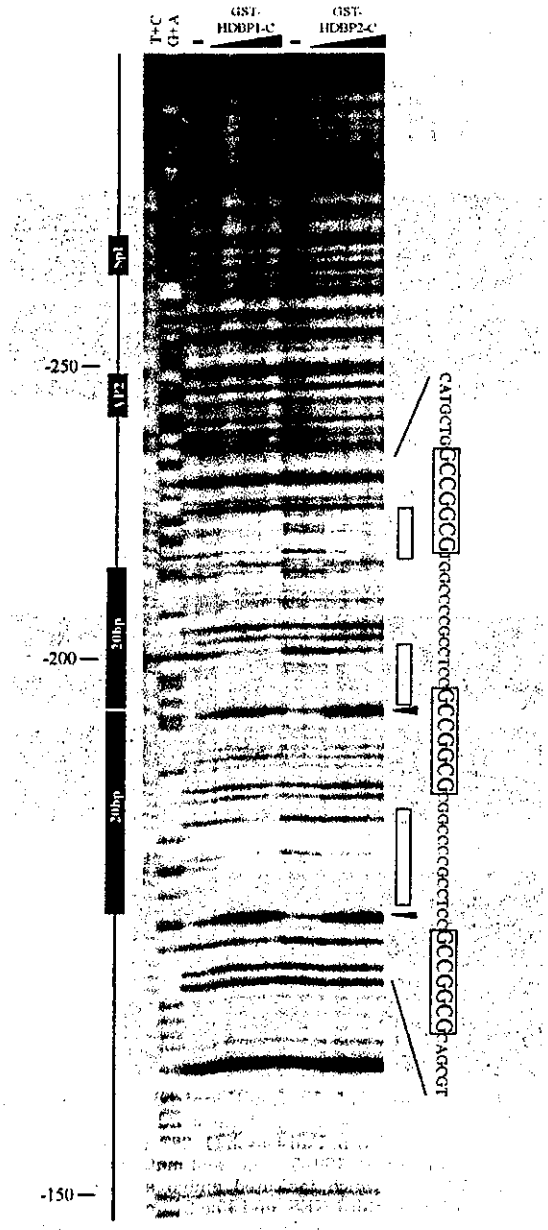

FIG. 6. **DNase I footprinting analysis of the *HD* promoter region between nucleotides −365 and −113.** The 253-bp MluI-XhoRI fragment labeled at its 3′-XhoI end was used for the determination of the sequence protected by HDBP1 and HDBP2. The labeled fragment was incubated with 10, 50, 200, and 400 ng of GST-HDBP1-C or GST-HDBP2-C for 20 min at room temperature. The samples were digested with an appropriate concentration of DNase I and were separated on an 8% urea/8% acrylamide gel by electrophoresis. The regions protected by GST-HDBP1-C or GST-HDBP2-C were *highlighted* as *white boxes* on the *right*, and DNA sequences protected are *boxed*. The *lanes* shown as *minus* (−) are controls without GST fusion protein. The *arrowheads* shown on the *right* are indicative of DNase I-hypersensitive sites. The lanes of A+G and T+C show the products of A+G and T+C chemical reaction as markers, respectively. *Numbers* on the *left* indicate the nucleotide positions relative to the translation start site. Sp1 and AP2 binding sites and 20-bp direct repeats are also indicated.

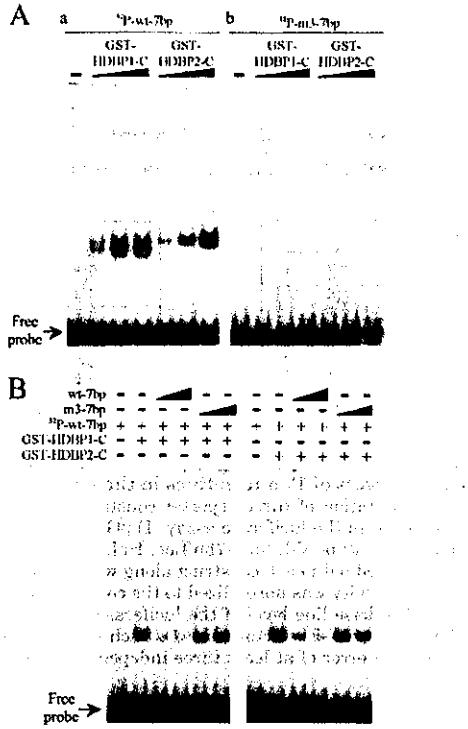

FIG. 7. **Specific binding between wild-type triplicated 7-bp consensus sequences and HDBP1 or HDBP2.** EMSA was performed using two 86-bp *Mlu*I-XhoRI fragments, wt-7bp and m3-7bp, corresponding to nucleotides −231 to −146 of the *HD* promoter sequence, which were labeled with [α-^{32}P]dCTP. *A*, different amounts (5, 25, and 50 ng) of GST-HDBP1-C and GST-HDBP2-C (see Fig. 5*A*) were incubated with ^{32}P-labeled wt-7bp (*a*) or m3-7bp (*b*) for 20 min at room temperature. *B*, for the competition assay, the reaction mixture containing 5 ng of GST-HDBP1-C and GST-HDBP2-C was incubated with 20- and 50-fold molar excess of unlabeled wt-7bp or m3-7bp for 5 min at room temperature before the addition of ^{32}P-labeled wt-7bp. DNA-protein complexes were separated on a 4% polyacrylamide gel by electrophoresis. The positions of the free probe are indicated.

not shown). These results indicate that only a third NES (253–263 aa) of HDBP1 is functional indeed. In addition, the C-terminal peptide showed the nuclear localization after LMB treatment (data not shown), implying that the second NLS (351–354 aa) of HDBP1 is also functional candidate.

As HDBP2 carries two putative NES motifs, we generated the N- and C-terminal HDBP2 peptides (1–168 and 169–513 aa, respectively) of GFP fusion protein-expressing constructs to determine the functional NES (Fig. 10*B*). When each terminal peptide was expressed in HeLa and IMR32 cells, both the N- and C- terminal peptides showed the nuclear localization (Fig. 10*B* and data not shown). Taking this result together with a fact that the second NES (165–174 aa) is split into the N- and C-terminal halves of HDBP2, the first intact NES is not functional and the second one may be a functional candidate. To confirm this, two NES mutants of HDBP2, HDBP2-L109A/L113A and HDBP2-M169A/M172A, were generated and expressed in HeLa and IMR32 cells. HDBP2-M169A/M172A localized in the nuclei (Fig. 10, *B* and *C*), whereas HDBP2-L109A/L113A was in the cytoplasm (Fig. 10*B* and data not shown). These results indicate that the second NES (165–174 aa) is functional. In addition, both putative NLSs (156–162 and 267–273 aa) of HDBP2 seem to be functional, based upon the fact of the nuclear localization of N- and C-terminal halves of HDBP2 protein.

DISCUSSION

HD is an autosomal dominant neurodegenerative disorder caused by an expansion of the CAG trinucleotide repeat within exon 1 of the *HD* gene. The molecular pathogenesis of HD is not yet fully understood. Recently, the conditional HD mouse has been generated by Yamamoto *et al.* (15). Their results brought

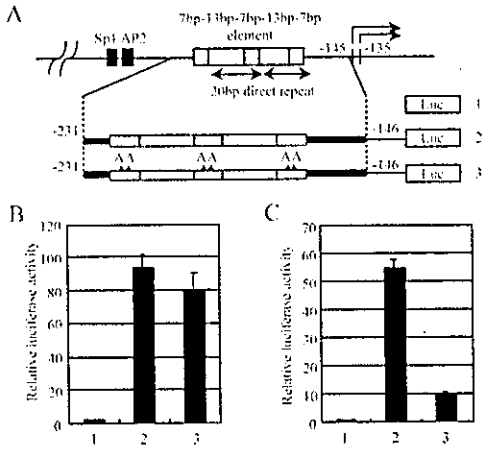

FIG. 8. **Functional analysis of 7-bp consensus sequences in triplicate at intervals of 13 nucleotides in the *HD* promoter.** *A*, a schematic representation of three reporter constructs (see "Materials and Methods") used in the luciferase assay: 1) pGV-E2/Luc (vector), 2) pGV-E2-wt-7bp/Luc, 3) pGV-E2-m3-7bp/Luc. HeLa (*B*) and IMR32 (*C*) cells were transfected with each construct along with the pRL-TK. The firefly luciferase activity was normalized to the control sea pansy luciferase activity. The base-line level of the luciferase activity was determined in the empty pGV-E2 vector (*lane 1*). Each value represents the mean and standard error of at least three independent experiments.

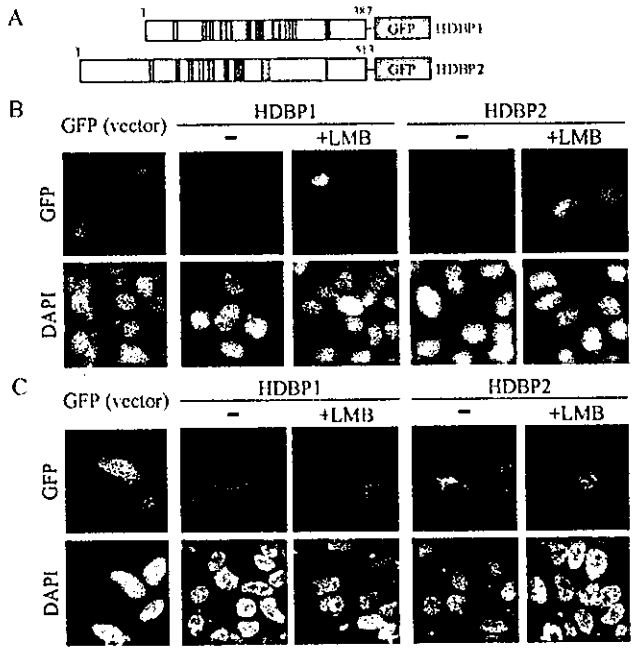

FIG. 9. **Subcellular localization of HDBP1 and HDBP2 expressed transiently in HeLa and IMR32 cells.** *A*, schematic representation of transiently expressed HDBP1-GFP and HDBP2-GFP fusion proteins. *B*, HeLa cells were transfected with GFP alone (GFP vector), HDBP1-GFP, and HDBP2-GFP constructs. *C*, IMR32 cells were transfected with GFP alone (GFP vector), HDBP1-GFP and HDBP2-GFP constructs. After 24 h, transfected cells were treated with (+LMB) or without (−) LMB (10 ng/ml) for 1 h. The localization of HDBP1 and HDBP2 were observed by GFP fluorescence (*GFP*), and the nuclei were visualized by 4′,6-diamidino-2-phenylindole, dihydrochloride staining (*DAPI*). Note that LMB treatment resulted in nuclear localization of HDBP1 and HDBP2. All images were obtained at a magnification of ×200 (*B*) and ×630 (*C*).

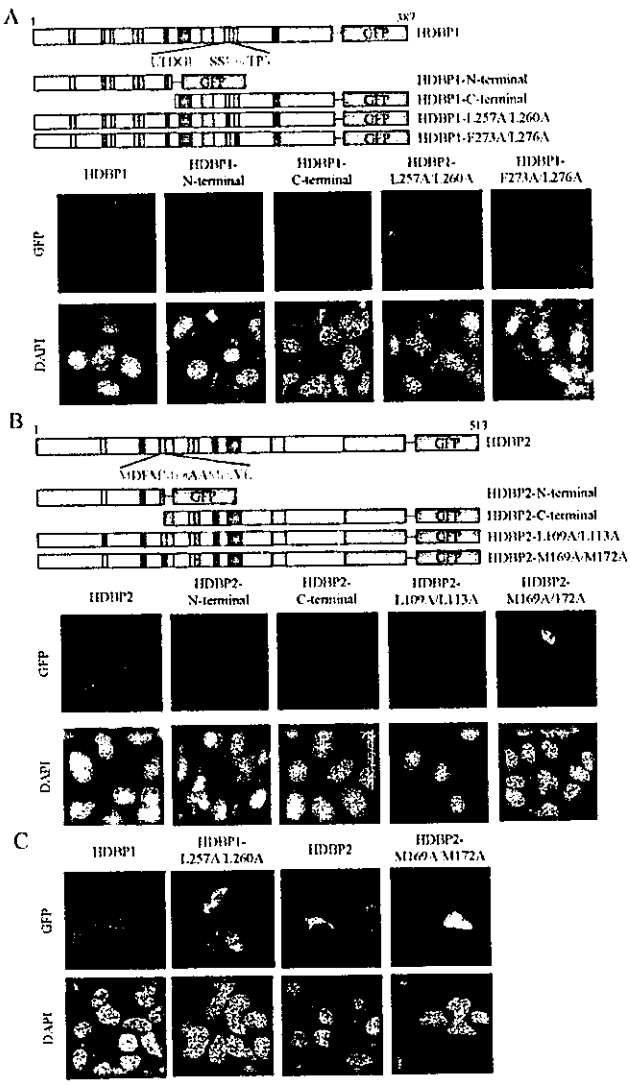

FIG. 10. **Identification of the functional NES in HDBP1 and HDBP2.** *A*, schematic representation of transiently expressed HDBP1-GFP fusion proteins is shown. Full-length (1–387 aa), N-terminal half (1–198 aa), C-terminal half (199–387 aa), and mutants (L257A/L260A and F273A/L276A), in which indicated amino acid residues (Leu-257and Leu-260 for the third NES, and Phe-273 and Leu-276 for the fourth NES) were substituted for alanine, were expressed in HeLa cells. *B*, schematic representation of transiently expressed HDBP2-GFP fusion proteins is shown. Full-length (1–513 aa), N-terminal half (1–168 aa), C-terminal half (169–513 aa), and mutants (L109A/L113A and M169A/M172A), in which indicated amino acid residues (Leu-109and Leu-113 for the first NES, and Met-169 and Met-172 for the second NES) were substituted for alanine, were expressed in HeLa cells. *C*, wild-type full-length (1–387 aa) and mutant (L257A/L260A) of HDBP1-GFP fusion proteins and wild-type full-length (1–513 aa) and mutant (M169A/M172A) of HDBP2-GFP fusion proteins were transiently expressed in IMR32 cells. The localization of fusion proteins was observed by GFP fluorescence (*GFP*) and the nuclei were visualized by 4′,6-diamidino-2-phenylindole, dihydrochloride staining (*DAPI*). All images were obtained at a magnification of ×200 (*A* and *B*) and ×630 (*C*).

a new aspect of the HD pathogenesis by means of perpetual expression of the mutated *HD* gene that might be a major causative for HD, and imply that HD is a reversible neurodegenerative disorder. Besides the unknown function of HD protein, molecular base of the transcriptional regulation of the *HD* gene in neurons has not been delineated as yet.

In this study, we identified two independent ubiquitously expressed genes, *SLC2A4* and *ZNF395*, which encode proteins termed HDBP1 and HDBP2, respectively. Analysis of amino acid sequences of HDBP1 and HDBP2 revealed that both proteins are highly homologous and shared several motifs as well as the structural features including NLS, NES, a single zinc

finger motif of C_2H_2 type, serine-rich region, proline-rich region, and the highly conserved C-terminal region (Fig. 3). The biochemical analyses of HDBP1 and HDBP2 demonstrated that the highly conserved C-terminal region (CR3) of each protein contained the novel DNA-binding domain, which mediated the interaction to the region containing the unique 20-bp direct repeat, which is a potential enhancer for the *HD* promoter (18) (Fig. 4). DNase I footprint analysis indicated that both novel DNA-binding domains for HDBP1 and HDBP2 recognized the heptanucleotide (GCCGGCG: 7-bp consensus sequence) in triplicate, which is located within and proximal to 20-bp direct repeats (Fig. 6). In addition, the specific binding between the 7-bp consensus sequence and HDBP1 (HDBP2) was confirmed by EMSA (Fig. 7). These results imply that HDBP1 and HDBP2 are first candidate *trans*-acting factors coupled with the 7-bp consensus sequence, which is in triplicate at intervals of 13 bp in the *HD* promoter and is also a novel *cis*-regulatory element.

The mutation of 7-bp consensus sequence, GCCGGCG to GCAAGCG, not only obstructed HDBP1 (HDBP2) binding to the sequence but also abolished the *HD* promoter activity in a neuronal cell line, IMR32 (Fig. 8*C*). Under the present condition of the *HD* promoter assay, IMR32 cells in non-differentiated phase may tend to less dose stringency of the 7-bp consensus sequences. However, the mutation of 7-bp consensus sequence did not significantly affect the *HD* promoter function in a non-neuronal cell line, HeLa (Fig. 8*B*). It is noteworthy that the *HD* promoter relies on the 7-bp consensus sequence in a neuronal cell line but not in a non-neuronal cell line. These results imply that the 7-bp consensus sequence acts as an essential *cis*-regulatory element in neuronal cells for the transcription steered by *HD* promoter. Thus, these elements (HDBP1, HDBP2, and 7-bp consensus sequence) provide new clues to the understanding of the transcriptional regulation of the *HD* gene.

Moreover, a set of the triplicate DNA-binding 7-bp consensus sequence at 13-bp intervals may contribute to form a rigid HDBP1 (HDBP2)-DNA complex through a positive cooperative manner, and a complex with single copy of the 7-bp consensus sequence as well as similar (GCCGGGA/GCCGGGC) sequence is vulnerable to dissociation. In this study, we have no results yet to show that either single HDBP1 (HDBP2) molecule or polymeric HDBP1 and/or HDBP2 molecules bind to the triplicate 7-bp consensus sequences.

Data base searches found that a part of HDBP1 peptide sequence was an exact match of GEFdb playing role in a transcriptional regulation of the human GLUT4 gene (19). However, GEFdb that has been cloned and sequenced so far seems to be a part of the full length of the peptide, lacking an initiation methionine (19). Our data indicate that GEFdb could be a potential splice variant omitting the conserved C-terminal region of HDBP1, through which HDBP1 bound to the *HD* promoter (Fig. 3). Conversely, HDBP1 might represent the full-length polypeptide encoded by *SLC2A4* gene. Nevertheless, it has been shown that GEFdb binds to DNA sequence ACCGG within the *GLUT4* (19). This implicates that each protein (HDBP1, GEFdb) has a distinct DNA-binding domain, which recognizes a unique DNA sequence and is implicated in the transcriptional regulation of the different genes (such as *HD* and *GLUT4*), possibly depending on different signaling pathways or tissue specificities. On the other hand, amino acid sequence of HDBP2 is an exact match of that of PBF (20). Recent study has reported that PBF recognizes the sequence CCGG in the 3′ half of the E2 binding site of papillomavirus genomes and has played a role in transcription of papillomavirus genes and in E2-mediated repression (20). However, a significant correlation between the transcriptional regulation of the *HD* gene and that of papillomavirus genes is unclear. In addition, both HDBP1 and HDBP2 proteins share the similar domain structures (Fig. 3), suggesting that HDBP1/GEFdb and HDBP2/PBF are members of the novel protein family. Data base searches also revealed that mouse GIG1 shared the several motifs and domain features, including a conserved C-terminal region, with those found in HDBP1 and HDBP2 (Fig. 3). Although human homologue of GIG1 has not been identified yet, the function of GIG1 may be analogous to that of either HDBP1 or HDBP2. So far, it is unknown whether GIG1 interacts with the promoter region for the *HD* gene.

The 7-bp consensus sequence, GCCGGCG, in triplicate at intervals of 13 nucleotides is unique to the *HD* promoter gene, but a single copy of the 7-bp sequence is also found in other human genes. For example, the H^+/K^+-ATPase β subunit gene (25), thrombopoietin (*TPO*) gene (26), $p27^{Kip1}$ gene (27), and Werner's syndrome gene (*WRN*) (28) carry a single copy of the 7-bp sequence within their promoter regions. However, neither *cis*-regulatory elements nor *trans*-acting factors among these genes have been identified as yet. Our results using the reporter gene assay demonstrated that the 7-bp consensus sequence(s) was sufficiently effective in a neuronal cell line to promote the transcription of heterologous gene (Fig. 8, *B* and *C*). In addition, Sp1-like sites, which lie within 13-nucleotide intervals, are also associated with the enhanced transcription (18). Thus, it is conceivable that both HDBP1 and HDBP2 are novel *trans*-acting factors, and the triplicate 7-bp consensus sequence at 13-bp intervals is a novel *cis*-regulatory element in the transcriptional regulation of *HD* gene.

To understand the transcriptional regulation of the *HD* gene via the interactions of HDBP1 and/or HDBP2 to the *HD* promoter, characterization of the subcellular distribution of these proteins is thought to be the first step. Unexpectedly, the ectopically expressed GFP-tagged HDBP1 and HDBP2 in HeLa and IMR32 cells localized in cytoplasm (Fig. 9, *B* and *C*). However, LMB treatment induced rapid accumulation of GFP-HDBP1 and -HDBP2 in the nuclei (Fig. 9, *B* and *C*). Moreover, the mutations in NES also led to the nuclear localization of HDBP1 and HDBP2 (Fig. 10). These findings clearly show that both NESs in HDBP1 and HDBP2 are functional in either neuronal/non-neuronal cells.

The nuclear transport of protein factors is regulated by various stimuli such as phosphorylation and/or dephosphorylation (29), stress response (24), or protein processing (30). For instance, transcription factor has to be re-localized from the cytoplasm into the nuclei to modulate the expression of target genes (31, 32). The nuclear localization of Aft1 transcription factor responds to iron status and leads to iron-regulated expression of the target genes, and NES associates with nuclear export of Aft1 (32). Also, class II histone deacetylase 5 localizes in the nuclei in the dephosphorylated state and associates with MEF2 transcription factor to block expression MEF2 target genes (29). In the phosphorylated state, histone deacetylase 5 is excluded from the nuclei through the NES-dependent pathway. In fact, we found a putative 14-3-3 protein binding site within the primary sequence of HDBP1 and HDBP2. As 14-3-3 protein plays critical roles in signal transduction pathways that control mitogen-activated protein kinase activation and gene expression (33), it may regulate the nucleocytoplasmic transport of HDBP1 and HDBP2. Taken together, it is conceivable that HDBP1 and HDBP2 may shuttle between nuclei and cytoplasm through a NES-dependent pathway that is regulated by the specific signal transduction mediated by factors such as 14-3-3 protein. In particular, factor(s) that was specifically expressed in neuronal cells might influence the subcellular distribution of

HDBP1 and HDBP2 in neuronal cell-specific manners, and HDBP1 and HDBP2 might interact with other neuronal cell-specific factors regulating the *HD* promoter activity in IMR32 cells.

Recently, functional differences in the *HD* promoter between neuronal and non-neuronal cell lines have been reported (18). Furthermore, the *HD* gene expression seems to be regulated by Sp1 as well, because the Sp1 element is conserved in the *HD* promoter and 20-bp direct repeats also contain Sp1-like element (18). Moreover, the 20-bp direct repeats in the human *HD* gene containing the 7-bp recognition sequence show polymorphisms in its copy number ranging from one to three copies (34), suggesting that the bindings of HDBP1 and HDBP2 to the *HD* promoter may also be influenced by this polymorphism.

Future studies on HDBP1 and HDBP2 may provide new clues to the understanding of the neuron-specific transcriptional regulation of the *HD* gene and will pave the way toward the development of a novel therapy for HD.

Acknowledgments—We thank Shinji Hadano for helpful discussion and owe Alexander E. MacKenzie (University of Ottawa, Ottawa, Ontario, Canada) a debt for the critical reading of the manuscript.

REFERENCES

1. Marin, J. B., and Gusella, J. F. (1986) *N. Engl. J. Med.* **315**, 1267–1276
2. The Huntington's Disease Collaborative Research Group (1993) *Cell* **72**, 971–983
3. DiFiglia, M., Sapp, E., Chase, K. O., Davies, S. W., Bates, G. P., Vonsattel, J. P., and Aronin, N. (1997) *Science* **277**, 1990–1993
4. Davies, S. W., Turmaine, M., Cozens, B. A., DiFiglia, M., Sharp, A. H., Ross, C. A., Scherzinger, E., Wanker, E. E., Mangiarini, L., and Bates, G. P. (1997) *Cell* **90**, 537–548
5. Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G. P., Davies, S. W., Lehrach, H., and Wanker, E. E. (1997) *Cell* **90**, 549–558
6. Dunah, A. W., Jeong, H., Griffin, A., Kim, Y.-M., Standaert, D. G., Hersch, S. M., Mouradian, M. M., Young, A. B., Tanese, N., and Krainc, D. (2002) *Science* **296**, 2238–2243
7. Li, S.-H., Cheng, A. L., Zhou, H., Lam, S., Rao, M., Li, H., and Li, X.-J. (2002) *Mol. Cell. Biol.* **22**, 1277–1287
8. Cha, J.-H. J. (2000) *Trends Neurosci.* **23**, 387–392
9. Dawson, T. M., and Ginty, D. D. (2002) *Nat. Med.* **8**, 450–451
10. Nucifora, F. C., Jr., Sasaki, M., Peters, M. F., Huang, H., Cooper, J. K., Yamada, M., Takahashi, H., Tsuji, S., Troncoso, J., Dawson, V. L., Dawson, T. M., and Ross, C. A. (2001) *Science* **291**, 2423–2428
11. Shimohata, T., Nakajima, T., Yamada, M., Uchida, C., Onodera, O., Naruse, S., Kimura, T., Koide, R., Nozaki, K., Sano, Y., Ishiguro, H., Sakoe, K., Ooshima, T., Sato, A., Ikeuchi, T., Oyake, M., Sato, T., Aoyagi, Y., Hozumi, I., Nagatsu, T., Takiyama, Y., Nishizawa, M., Goto, J., Kanazawa, I., Davidson, I., Tanese, N., Takahashi, H., and Tsuji, S. (2000) *Nat. Genet.* **26**, 29–36
12. Steffan, J. S., Kazantsev, A., Spasic-Boskovic, O., Greenwald, M., Zhu, Y.-Z., Gohler, H., Wanker, E. E., Bates, G. P., Housman, D. E., and Thompson, L. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6763–6768
13. Zuccato, C., Ciammola, A., Rigamonti, D., Leavitt, B. R., Goffredo, D., Conti, L., MacDonald, M. E., Friedlander, R. M., Silani, V., Heyden, M. R., Timmusk, T., Sipione, S., and Cattaneo, E. (2001) *Science* **293**, 493–498
14. Leavitt, B. R., Guttman, J. A., Hodgson, J. G., Kimel, G. H., Singaraja, R., Vogl, A. W., and Hayden, M. R. (2001) *Am. J. Hum. Genet.* **68**, 313–324
15. Yamamoto, A., Lucas, J. J., and Hen, R. (2000) *Cell* **101**, 57–66
16. Holzmann, C., Schmidt, T., Thiel, G., Epplen, J. T., and Riess, O. (2001) *Mol. Brain Res.* **92**, 85–97
17. Lin, B., Nasir, J., Kalchman, M. A., McDonald, H., Zeisler, J., Goldberg, Y. P., and Hayden, M. R. (1995) *Genomics* **25**, 707–715
18. Coles, R., Caswell, R., and Rubinsztein, D. C. (1998) *Hum. Mol. Genet.* **7**, 791–800
19. Oshel, K. M., Knight, J. B., Cao, K. T., Thai, M. T., and Olson, A. L. (2000) *J. Biol. Chem.* **275**, 23666–23673
20. Boeckle, S., Pfister, H., and Steger, G. (2002) *Virology* **293**, 103–117
21. Nishi, K., Yoshida, M., Fujiwara, D., Nishikawa, M., Horinouchi, S., and Beppu, T. (1994) *J. Biol. Chem.* **269**, 6320–6324
22. Kudo, N., Taoka, H., Toda, T., Yoshida, M., and Horinouchi, S. (1999) *J. Biol. Chem.* **274**, 15151–15158
23. Arai, N. R., Akiyama, R., Niimi, N. H., Nakatsubo, H., and Inoue, T. (1999) *Genes Genet. Syst.* **74**, 159–167
24. Hoshino, H., Kobayashi, A., Yoshida, M., Kudo, N., Oyake, T., Motohashi, H., Hayashi, N., Yamamoto, M., and Igarashi, K. (2000) *J. Biol. Chem.* **275**, 15370–15376
25. Yoshida, T., Sato, R., Mahmood, S., Kawasaki, S., Futai, M., and Maeda, M. (1997) *FEBS Lett.* **414**, 333–337
26. Kamura, T., Handa, H., Hamasaki, N., and Kitajima, S. (1997) *J. Biol. Chem.* **272**, 11361–11368
27. Ito, E., Iwahashi, Y., Yanagisawa, Y., Suzuki, Y., Sugano, S., Yuasa, Y., and Maruyama, K. (1999) *Gene (Amst.)* **228**, 93–100
28. Yamabe, Y., Shimamoto, A., Goto, M., Yokota, J., Sugawara, M., and Furuichi, Y. (1998) *Mol. Cell. Biol.* **18**, 6191–6200
29. McKinsey, T. A., Zhang, C. L., and Olson, E. N. (2001) *Mol. Cell. Biol.* **21**, 6312–6321
30. Ura, S., Masuyama, N., Graves, J. D., and Gotoh, Y. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10148–10153
31. Brunet, A., Kanai, F., Stehn, J., Xu, J., Sarbassova, D., Frangioni, J. V., Dalal, S. N., DeCaprio, J. A., Greenberg, M. E., and Yaffe, M. B. (2002) *J. Cell Biol.* **156**, 817–828
32. Yamaguchi-Iwai, Y., Ueta, R., Fukunaka, A., and Sasaki, R. (2002) *J. Biol. Chem.* **277**, 18914–18918
33. Yaffe, M. B. (2002) *FEBS Lett.* **513**, 53–57
34. Coles, R., Leggo, J., and Rubinsztein, D. C. (1997) *J. Med. Genet.* **34**, 371–374
35. Altschul, S. F., Mockten, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402

Human Molecular Genetics, 2004, Vol. 13, No. 6 *589–600*
DOI: 10.1093/hmg/ddh064
Advance Access published on January 20, 2004

Inhibition of myogenesis in transgenic mice expressing the human DMPK 3′-UTR

Christopher J. Storbeck[1,2], Suzana Drmanic[1], Kate Daniel[3], James D. Waring[1], Frank R. Jirik[4], David J. Parry[5], Nazim Ahmed[1], Luc A. Sabourin[3,5], Joh-E Ikeda[6] and Robert G. Korneluk[1,2,*]

[1]Solange Gauthier–Karsh Molecular Genetics Laboratory, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Rd, Ottawa, Ontario, Canada K1H 8L1, [2]Department of Biochemistry, Microbiology and Immunology, University of Ottawa, 451 Smyth Rd, Ottawa, Ontario, Canada K1H 8M5, [3]Ottawa Hospital Research Institute Neuroscience Program, University of Ottawa, 451 Smyth Rd, Ottawa, Ontario, Canada K1H 8M5, [4]Department of Biochemistry and Molecular Biology, University of Calgary, 2500 University Drive NW, Calgary, AB, Canada, T2N 1N4, [5]Department of Cellular and Molecular Medicine, University of Ottawa, 451 Smyth Rd, Ottawa, Ontario, Canada K1H 8M5 and [6]NeuroGenes Project, International Cooperative Research Project/Japan Science and Technology Corporation, Children's Hospital of Eastern Ontario Research Institute, 401 Smyth Rd, Ottawa, Ontario, Canada K1H 8L1

Received September 24, 2003; Revised January 4, 2004; Accepted January 12, 2004

Myotonic dystrophy (DM1) is a multisystemic disorder caused by a CTG repeat expansion within the 3′-UTR of the *DMPK* gene. DM1 is characterized by delayed muscle development, muscle weakness and wasting, cardiac conduction abnormalities, cognitive defects and cataracts. Recent studies have demonstrated that the disease mechanism involves a dominant gain-of-function conferred upon mutant transcripts by expanded repeats. However, further attempts to model aspects of DM muscle pathology in cultured myoblasts suggest that 3′-UTR sequences flanking the CTG repeat tract are also required for full expression of the disease phenotype. Here, we report that overexpression of the *DMPK* 3′-UTR including either wild-type (11) or expanded (91) CTG repeats results in aberrant and delayed muscle development in fetal transgenic mice. In addition, transgenic animals with both expanded and wild-type CTG repeats display muscle atrophy at 3 months of age. Primary myoblast cultures from both 11 and 91 repeat mice display reduced fusion potential, but a greater reduction is observed in the 91 repeat cultures. Taken together, these data indicate that overexpression of the *DMPK* 3′-UTR interferes with normal muscle development in mice and that this is exacerbated by inclusion of a mutant repeat. This suggests that the delayed muscle development in DM1 involves an interplay between the expanded CTG repeat and adjacent 3′-UTR sequences.

INTRODUCTION

Myotonic dystrophy type I (DM1) is the most prevalent inherited neuromuscular disorder in adults, affecting approximately 1/8000. It is a multisystemic disorder clinically divided into an adult and a more severe congenital form. Features of adult DM1 include myotonia and progressive muscular weakness and wasting, cardiac conduction defects, testicular atrophy, cataracts and premature balding (1). Congenital DM infants are hypotonic, have excessive numbers of muscle satellite cells and experience difficulties suckling due to poorly developed facial muscles (1). In addition, they are often mentally retarded.

The causative mutation for DM1 is a CTG trinucleotide repeat expansion in the 3′-untranslated region (UTR) of the myotonic dystrophy protein kinase (*DMPK*) gene (2–4). This repeat is polymorphic in normal individuals with alleles ranging from five to 37 in length. Repeats exceeding a threshold of approximately 50–80 result in disease, with increasing severity and an earlier age-of-onset as the repeats lengthen.

*To whom correspondence should be addressed. Tel: +1 6137383281; Fax: +1 6137384833; E-mail: bob@mgcheo.med.uottawa.ca

Repeats in adult patients range from roughly 100 to 1000 in length while in congenital DM they range from 1000 to 3000 repeats (2). Recently, the mutation responsible for a second form of DM (DM2) was identified as a CCTG tetranucleotide repeat within the first intron of the *ZNF9* gene on chromosome 3. Expansions of the CCUG repeat in the most severe DM2 cases can be as large as 44 kb (5).

Although it is not entirely clear how CTG repeat expansions result in DM pathology, recent evidence has shown that an RNA-mediated dominant gain-of-function conferred upon mutant transcripts plays a major role. Overexpression of the *DMPK* 3′-UTR including an expanded CUG repeat (6) or repeat tracts alone (7) inhibits differentiation of the mouse myoblast cell line C2C12, which has been used to model the impaired muscle development seen in congenital DM1 cases. Mutant DMPK mRNA has been shown to be selectively retained in aberrant foci within the nuclei of patient cells, strongly suggesting that perturbed transcript processing underlies the dominant gain-of-function (8). As a potential mediator of this effect, a protein was characterized which binds to CUG repeat oligonucleotides *in vitro* and displays increasing affinity as the repeat sequence lengthens (9). This protein was identified as the human homolog (MBNL) of *Drosophila muscleblind*, which is important for muscle and eye differentiation in flies (10). Furthermore, MBNL and related isoforms were found to co-localize with mutant transcript foci in both DM1 and DM2 patient cell nuclei (11–13). Finally, insertion of 250 CUG repeats into the 3′-UTR of a foreign gene (mouse skeletal muscle actin) was sufficient to elicit specific features of DM muscle pathology in transgenic mice, notably myotonia (14).

Other experiments suggest that sequences within the *DMPK* 3′-UTR which flank the repeats may be important to the full expression of the disease phenotype. We found that over-expression of a wild-type 3′-UTR inhibits C2C12 differentiation and mapped this activity to the region 5′ of the repeat tract (15). In agreement, in a separate study the inhibition of differentiation by a mutant 3′-UTR also required the same region 5′ of the repeat (16). Also, transgenic mice over-expressing a wild-type *DMPK* gene displayed certain features of DM1, including type I fibre atrophy, increased numbers of central nuclei and reduced fusion potential of cultured myoblasts (17). To test this hypothesis, we overexpressed the *DMPK* 3′-UTR with either a wild-type (11) or mutant repeat sequence (91) in transgenic mice and examined the effects on muscle differentiation *in vivo*. We found that both strains of mice displayed defects in muscle differentiation at an early stage of development, including developmental abnormalities of face, jaw, back and neck musculature. In addition, young adult mice from both strains show type I and/or type II muscle fiber atrophy. Primary myoblasts obtained from young adult mice showed reduced fusion potential compared with cultures derived from control mice, but interestingly, this effect was more pronounced for 91 repeat mice compared with 11 repeat mice. This is the first demonstration of inhibition of myogenesis *in vivo* by overexpression of a wild-type *DMPK* 3′-UTR. This phenotype is exacerbated by the presence of a CUG repeat expansion and therefore suggests that cooperativity between the *DMPK* 3′-UTR sequences and the CUG repeat may be involved in the full expression of DM pathology in skeletal muscle.

RESULTS

Generation and characterization of transgenic mice overexpressing the human *DMPK* 3′-UTR

In vitro studies have demonstrated that overexpression of a normal *DMPK* 3′-UTR (15) or of mutant CTG repeats (6,7) inhibits the differentiation of C2C12 myoblasts. As these cells represent a faithful model of *in vivo* muscle differentiation, we investigated the possibility that overexpression of a normal (11 CTG repeats) or a mutant (91 CTG repeats) *DMPK* 3′-UTR in the muscle of transgenic mice could reproduce the observations seen in C2C12 myoblasts. Wild-type and mutant *DMPK* 3′ UTR sequences were fused downstream of the reporter gene *GFP* to facilitate detection of the gene product (Fig. 1). To direct expression of the transgene in muscle, we utilized a region derived from 5′-UTR of *DMPK* that was previously shown to have promoter activity in combination with a region from the first intron that had skeletal muscle specific enhancer activity (18); these sequences were fused upstream of *GFP*. The activities of these elements in the non-muscle tissues in which *DMPK* is normally expressed have not been investigated. Transgenic mice were generated and founders were identified by Southern analysis using a *GFP* probe (Fig. 1B).

To confirm that the *DMPK* regulatory elements employed conferred the appropriate expression patterns *in vivo* and to assess expression levels between different transgenic lines, northern blot analysis was performed using a human specific *DMPK* 3′-UTR probe (Fig. 2A). Total RNA was isolated from hind-limb muscle, heart and brain from five lines of transgenic mice containing 11 CUG repeats (CUG11; Fig. 2A and B and Table 1) and three lines with 91 CUG repeats (CUG91; Fig. 2B, Table 1 and data not shown) at 3–6 months of age. Transgene expression in all lines was highest in heart, followed by hind-limb muscle, while low amounts were detected in brain in some lines. This general expression pattern is consistent with that of the human *DMPK* gene (2,3). Transgene expression was generally higher in the CUG11 lines, with line CUG11–1913 being highest of the two CUG11 lines used in subsequent experiments (CUG11–1913, CUG11–1923: Fig. 2B and Table 1). Using quantitative real-time PCR analysis, we found that, of the three CUG91 founder lines generated, expression was comparable to that of the CUG11 lines in only one line (2043-1: Fig. 2B and Table 1). This was not unexpected as expanded CTG repeats have been shown to significantly decrease the expression of a reporter gene *in cis* in C2C12 (relative to wild-type repeats); these mutant constructs exerted a negative effect on differentiation despite their lower expression levels (6). Similar effects of repeat expansions on DMPK are also observed in patient myoblasts (19). Despite this difference in expression level, we used two CUG91 lines (2043-1 and 2038-3) in key experiments and noted similar results in both lines (Figs 5, 6 and 8 and data not shown). We also analyzed the expression of GFP in tissue extracts by western blot. Tongue, hind-limb muscle, heart and brain (data not shown) were analyzed from CUG11 line 1923 and a wild-type control littermate (Fig. 2C). An extract from C2C12 myoblasts transfected with a GFP expression construct was included as a positive control. GFP was clearly detected in the tongue and the heart and faintly in the hind-limb muscle, in agreement with the RNA analysis.

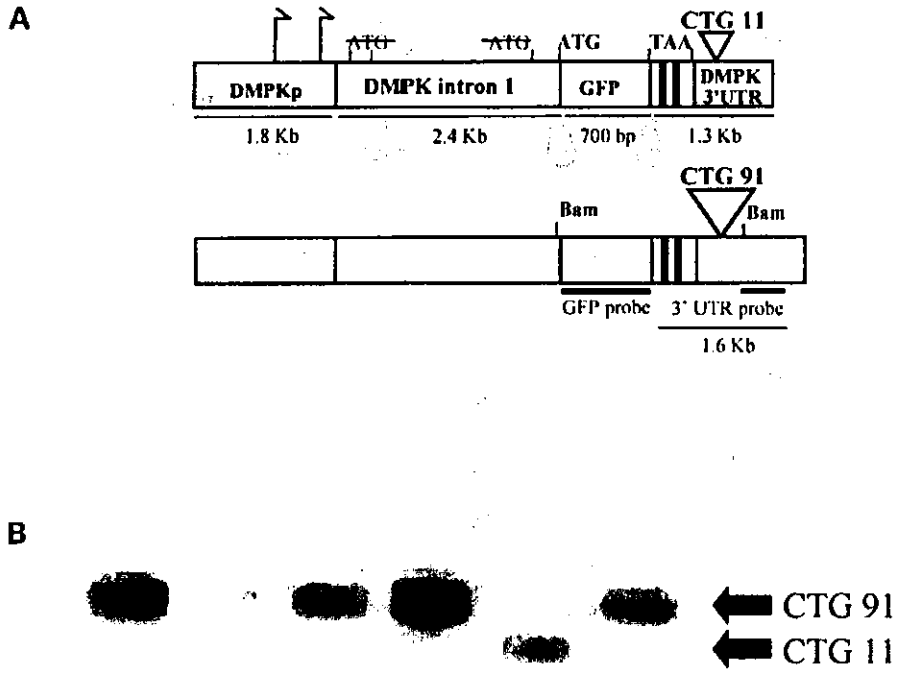

Figure 1. DMPK CUG11 and CUG91 transgenes and identification of transgenic mice by Southern blot analysis. (**A**) The transgene used in the construction of CUG11 and CUG91 transgenic mice is shown. The *Bam*HI sites used in Southern blot analysis are indicated and the location of the probe used for screening Southern blots for transgene positive animals. (**B**) Southern blots were probed with the GFP cDNA. Five 91 CUG repeat transgenic mice and one 11 CUG repeat transgenic animal are shown, highlighting the mobility difference between the two transgenes.

In situ hybridization analysis of transgene expression

To determine the precise expression pattern of the transgene and identify potentially affected tissues, *in situ* hybridization was performed on transgenic mouse embryos aged 20 days post-coitum (pc). Antisense probes for *GFP* and *myogenin* were used to detect the transgene mRNA and to identify developing muscle, respectively (Fig. 3A–F). GFP and myogenin were detected in the tongue, snout and to a lesser extent in muscles of the neck and throat (Fig. 3A and B). GFP was also expressed in several regions of the brain including the frontal lobe, midbrain, hindbrain and choroid plexus and ependymal cells of the subventricular layer (arrowheads, Fig. 3A and B). In the torso area, expression of both genes was found in intercostal muscles, muscles of the back and the diaphragm (Fig. 3C and D). GFP was also expressed in the heart and was evident in the smooth muscle lining of pulmonary and amniotic blood vessels (arrows and arrowheads, Fig. 3C and D). Finally, in the gut area, only GFP was observed in smooth muscle lining the large intestines whereas both genes were expressed in skeletal muscle of the lower limbs (arrows and arrowheads, Fig. 3E and F). These data are consistent with both *in situ* hybridization data for the endogenous mouse *DMPK* gene (20) and the known expression profile in humans. Therefore, in addition to achieving expression in skeletal muscle, the regulatory elements utilized conferred accurate expression of the transgene in heart, smooth muscle and brain. However, no overt pathology was observed in transgenic embryos at this (late) developmental stage.

DMPK 3′-UTR overexpression disrupts early muscle development

We examined transgenic mice for effects of *DMPK* 3′-UTR overexpression at day 12 pc, a time that coincides with the onset of primary myogenesis (21). Embryos from CUG11 line 1923 were identified as transgenic by *in situ* hybridization with an antisense *GFP* probe (Fig. 4A, embryo 3) and serial sections from these and wild-type littermates were stained using a probe for *myogenin* as a marker for developing muscle (22). Two wild-type and one transgenic embryo are shown (Fig. 4). Myogenin staining was clearly observed in the somites in all embryos (small arrows, Fig. 4B) and in developing muscles of the neck region in wild-type embryos (small arrowhead, embryos 1 and 2, Fig. 4B). In contrast, there was a striking lack of staining in this region of the neck in the CUG11 embryo (small arrowhead, embryo 3, Fig. 4B). GFP-positive tissue can be seen in this region (extending dorsal and ventral from the large arrowhead, embryo 3, Fig. 4A) but expresses little myogenin (large arrow and arrowhead, embryo 3, Fig. 4B). This suggests that pre-muscle tissue is present but has not fused to form myotubes. GFP-positive tissue can also be seen in a more ventral position (large arrow, embryo 3, Fig. 4A), in the area of developing jaw muscles. Several corresponding myogenin-positive structures are present in wild-type embryos (for example, black arrowheads in embryos 1 and 2, Fig. 4B), but only one can be seen in the transgenic embryo (large arrow, embryo 3, Fig. 4B). These data indicate that overexpression of the *DMPK* 3′-UTR disrupts primary myogenesis in the developing mouse embryo, particularly in future facial and neck areas.

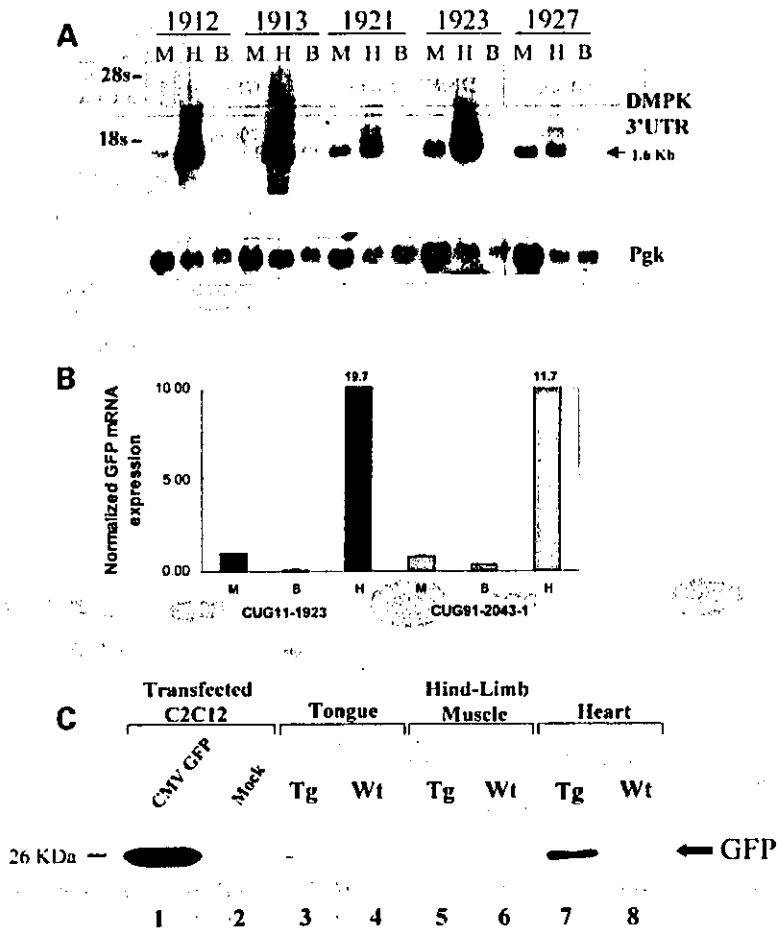

Figure 2. Expression analysis of mRNA and protein from various tissues of several lines of CUG11 transgenic mice. (A) Northern blot analysis was performed on total RNA from skeletal muscle, heart and brain of five CUG11 transgenic mouse lines (1912, 1913, 1921, 1923 and 1927). The northern blot was probed with the 3′ end of the DMPK 3′ UTR. To control for RNA loading, the blot was probed with the Pgk cDNA. (B) Quantitative real-time PCR expression analysis of CUG11 and CUG91 transgenic mice. Total RNA from heart, muscle and brain of two CUG11 and two CUG91 transgenic mouse lines were analyzed. Tabulated transgene expression levels were normalized to GAPDH expression levels. Shown are relative expression profiles for CUG11 line 1923 and CUG91 line 2043-1. CUG91 line 2043-1 expressed less transgene mRNA than CUG11 line 1923 in heart and muscle but more in brain. Experiment was performed three times in triplicate, twice with one set of primers and once with a second set of primers yielding near identical results. Shown is a representative experiment. (C) Western blot analysis using a polyclonal GFP antibody was performed on tissue lysates from CUG11 transgenic mouse and wild-type littermates. Lanes 1 and 2 are protein lysates from C2C12 cells transfected with a GFP expression vector or mock transfected. Protein lysates from tongue, skeletal muscle and heart from transgenic (lanes 3, 5 and 7) and wild-type littermates (lanes 4, 6 and 8) were analyzed. Expression of CUG11 mRNA and protein is highest in cardiac tissue and is present but at lower levels in tongue and skeletal muscle.

Table 1. Relative transgene expression in muscle, brain and heart of CUG11 and CUG91 lines assessed by quantitative RT–PCR

Line	Muscle	Brain	Heart
CUG 11–1913	1.85 ± 0.33	13.58 ± 0.25	43.16 ± 4.44
CUG 11–1923	1 ± 0.01	0.09 ± 0.01	19.66 ± 0.70
CUG 91–2038-3	0.02 ± 0.00	0.01 ± 0.00	0.51 ± 0.06
CUG 91–2043-1	0.81 ± 0.09	0.37 ± 0.02	11.66 ± 0.96

DMPK 3′-UTR overexpression delays secondary muscle development

The above data suggests that overexpression of a wild-type *DMPK* 3′-UTR can delay normal muscle development *in vivo*, as previously speculated (15,17). To determine whether overexpression of a mutant *DMPK* 3′-UTR resulted in any additional pathology, we performed immunohistological analysis by staining CUG11, CUG91 and wild-type animals for myogenin at day 15 pc, a time which marks the boundary between primary and secondary myogenesis. In back, neck and tongue musculature, a lack of fully developed myotube structures was observed in transgenic embryos from CUG11 (line 1913; Fig. 5B, D and F) compared with a wild-type littermate (Fig. 5A, C and E). This phenotype was also observed in the other CUG11 line (1923) as well as both CUG91 lines examined (2038-3 and 2043-1; data not shown). Transgenic embryos had either disorganized, incompletely formed myotubes containing only one or a few myonuclei

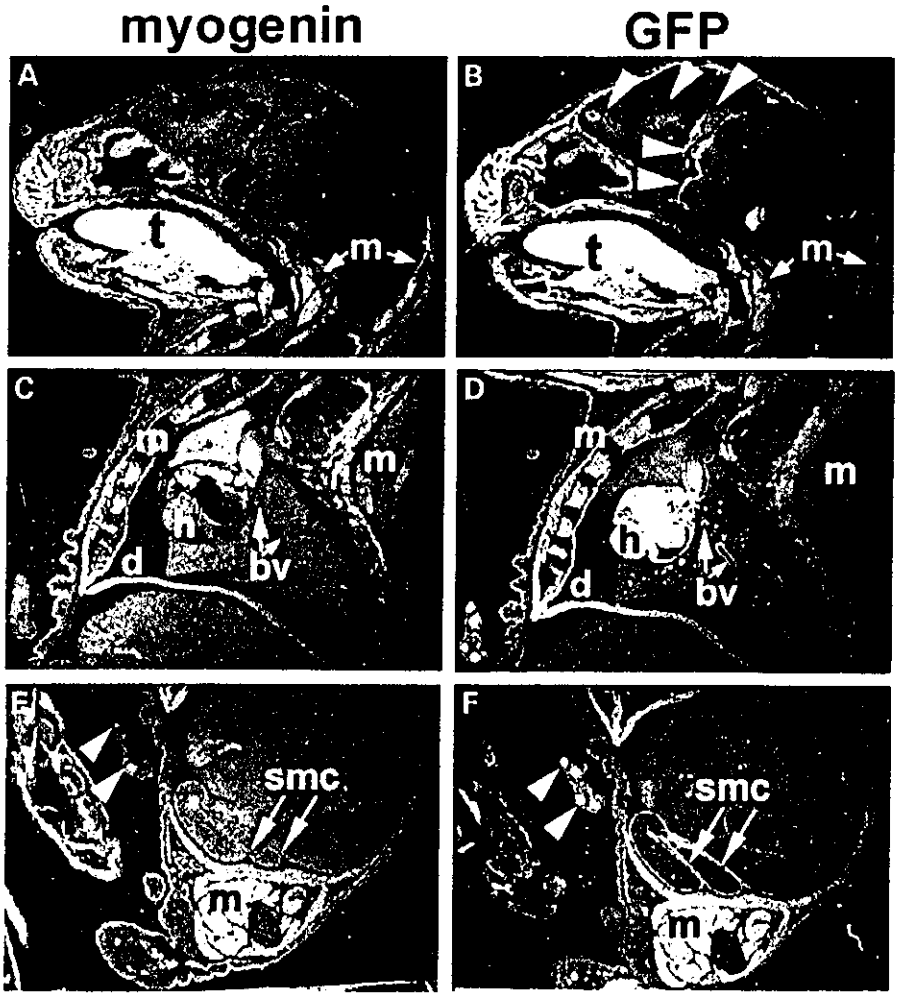

Figure 3. Expression of the CUG11 transgene (GFP) in E20 embryos. *In situ* hybridization was performed on transgenic (line 1923) E20 embryo sections using antisense probes for myogenin (**A, C, E**) and GFP (**B, D, F**). (A, B) Expression of myogenin and the GFP is clearly observed in tongue (T), muscles of the head and neck area (M), cell linings of the ventricles and within regions of the forebrain, midbrain and hindbrain (arrowheads) (B). Myogenin and GFP are present in intercostal muscles (m), muscles of the back and neck (m) and in the diaphragm (d) (C, D). Expression of GFP, but not myogenin was observed in smooth muscle cells lining the blood vessels (bv) and cardiac cells of the heart (h). Expression of both myogenin and GFP is observed in the muscles of the lower limb (m) (E, F). Only GFP is expressed in the smooth muscle cell (smc) lining of the large intestines and amniotic blood vessels (arrows, arrowheads) (F). GFP is expressed in skeletal muscle, heart, smooth muscle and brain at E20. This pattern of expression indicates that regulatory elements required for tissue specific expression of DMPK were included in the transgene regulatory elements.

or many mononucleated myogenic cells, instead of large multinucleated myotubes (arrowheads, Fig. 5D and F). Immunostaining of transgenic CUG91 (2038-3) E15 embryos for total sarcomeric myosin protein revealed, like myogenin staining, poorly developed myotube structures (Fig. 6). Myotubes in the developing back and neck musculature of the transgenic animals were very thin and underdeveloped (Fig. 6B, D), while no myotubes were evident in the jaw musculature of the transgenic embryo (compare Fig. 6E and F). In addition, it was clearly evident that sarcomeric myosin protein expression was considerably lower in the developing muscle of the transgenic embryo compared with the wild-type, indicating reduced myogenic differentiation (Fig. 6A–F). Again, this phenotype was observed in all other lines analyzed (1913, 1923 and 2043-1).

A differentiation defect seen in primary myoblasts from transgenic animals is exacerbated by mutant CTG repeats

In order to quantitate the difference in myotube development, primary myoblast cell cultures were established from one CUG11 line (1923), one CUG91 line (2043-1) and from wild-type mice and differentiated *in vitro* by growth factor withdrawal. Myotubes were poorly developed in both CUG11 (not shown) and CUG91 cultures compared with wild-type (Fig. 7A and B). Fusion indices were determined (see Materials and Methods) and a clear reduction was seen in both transgenic cultures compared with wild-type (Fig. 7C). Interestingly, CUG91 cultures exhibited an approximately 50% lower fusion index than CUG11 cultures. These results are in agreement with

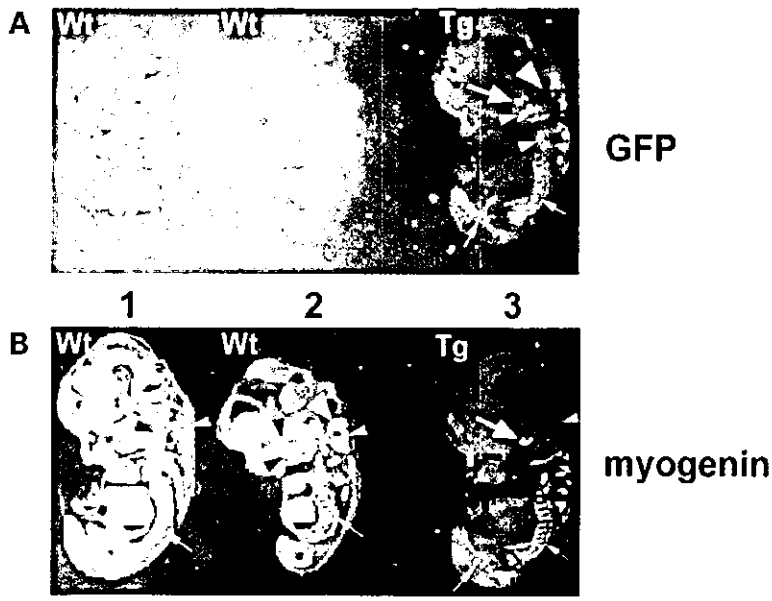

Figure 4. Expression of the CUG11 mRNA during early mouse embryogenesis. (A) *In situ* hybridization was performed on embryo (12 days pc) sections using an antisense GFP probe. The presence of GFP transcripts indicates embryo 3 is transgenic. GFP is expressed in skeletal muscle precursor cells of the somite (small arrows) and in muscles of the future neck, face and jaw (small arrowheads). (B) Primary myogenesis was detected *in vivo* using an antisense probe for myogenin, which is expressed in developing skeletal muscle. GFP is expressed in a similar pattern to myogenin confirming *in vivo* muscle specificity of the *DMPK* regulatory elements employed. The white arrowhead in embryo 3 indicates neck musculature that does not express myogenin transcripts (compare small white arrowhead in embryo 2 and 3). Small white arrows mark myogenin expression in somites. The black arrowheads in embryos 1 and 2 indicate expression of myogenin in facial muscles of wild-type embryos while the white arrow in embryo 3 indicates poor myogenin staining of facial muscle in the transgenic embryo. Transgenic embryos display a muscle development phenotype during primary myogenesis, particularly in developing neck and facial muscle.

both the poor development of muscle seen *in vivo* in these mice and with a study demonstrating inhibition of myoblast fusion as a result of 3′-UTR expression in C2C12 myoblasts (16).

Muscle fibre atrophy in young transgenic mice

Finally, we closely examined muscle from mice at various times after birth. A histological analysis of muscle at 6 months revealed no overt changes in CUG11 and CUG91 mice, consistent with the overtly normal appearance of day 20 embryos. As type I muscle fiber atrophy is a hallmark feature of DM, we determined the mean fiber area in soleus muscle from younger transgenic and wild-type animals. Cross sections were stained for type I MHC to distinguish type I and type II fibers and the cross sectional areas were plotted (Fig. 8A and B). In 1 to 3-month-old CUG11 mice, significant type I fiber atrophy was observed in line 1913 ($P<0.05$, Student's *t*-test) compared with wild-type littermates, whereas line 1923 showed significant type I and II fiber atrophy (Fig. 8B). The CUG91 mice analyzed also displayed significant type I and II fiber atrophy (Fig. 8B). We are uncertain whether the unexpected type II atrophy is reflective of a more pronounced defect than that normally seen in humans, of aberrant expression of the transgene at the fiber level, or of some difference between mouse and human physiology. In older mice (6 months and 1 year of age) muscle hypertrophy, but never atrophy, was sometimes observed. Taken together, our data demonstrate that the effects of *DMPK* 3′-UTR overexpression are pronounced early in development but diminish with age.

DISCUSSION

Considerable evidence now supports the hypothesis that the dominant nature of the DM mutation is exerted at the RNA level. In particular, expression of 250 CTG repeats placed within the 3′-UTR of an α-actin transgene causes myotonia and other histological features of DM in mouse skeletal muscle (14). However, a number of observations suggest that expression of expanded repeats alone does not fully account for disease. Weakness and wasting of skeletal muscle was not observed in the α-actin/mutant repeat mice, suggesting that the myopathy observed was incomplete. In studies with C2C12 myoblasts, we and others have found that expression of the *DMPK* 3′-UTR inhibits differentiation (6,15,16). While these studies have disagreed on the importance of a mutant repeat, they have agreed that the region 5′ of the repeat within the 3′-UTR is necessary. Interestingly, transcripts with repeats alone (i.e. removed from their normal 3′-UTR context) aggregate in nuclear foci, yet no longer inhibit differentiation, demonstrating for the first time that the phenomena are separable (16). Finally, it is noteworthy that DM2 does not have a severe congenital form despite the fact that MBNL (and related isoforms) are likewise sequestered by mutant *ZNF9* transcripts. While this might simply be due to intrinsic properties of CCUG versus CUG repeat sequences, it is equally possible that the sequence context for expanded repeats plays a role. Together, these studies suggest an interplay between repeat expansions and some normal function of the 3′-UTR, perhaps related to development or differentiation.

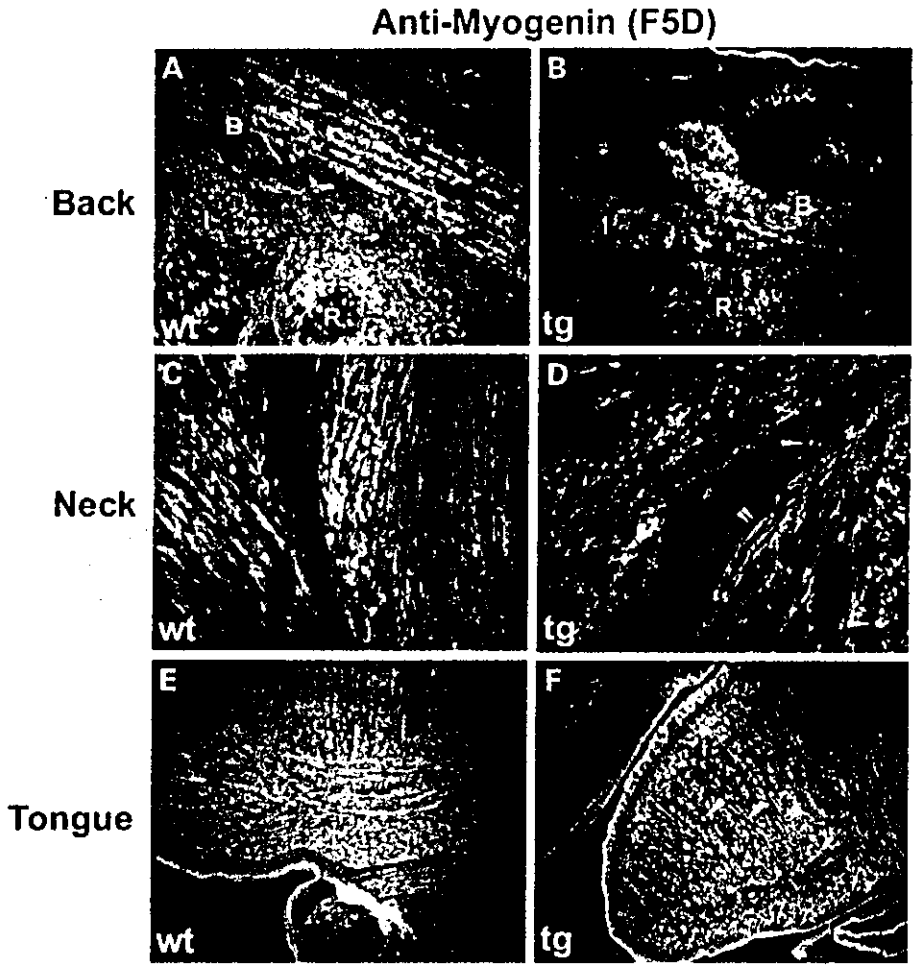

Figure 5. Inhibition of muscle differentiation in E15 embryos from CUG11 and CUG91 mice. Wild-type and transgenic embryo littermates from two lines of CUG11 (1913, 1923) and CUG91 (2038-3, 2043-1) were sectioned and stained with anti-myogenin antibody F5D. Shown are muscles of the back (**A**, **B**), neck (**C**, **D**) and tongue (**E**, **F**) of CUG11 (1913) mouse and a wild-type control. An identical phenotype was observed in embryos from both (2038-3, 2043-1) CUG91 lines and CUG11 line 1923. Formation of myotubes *in vivo* is inhibited by expression of the DMPK 3′-UTR with 11 or 91 CUG repeats in developing muscle.

We chose to study the effects of overexpression of the human DMPK 3′-UTR using an endogenous promoter/enhancer system (18), in an attempt to most closely mimic the normal timing and expression patterns for *DMPK*. It was hoped that this approach would avoid possible artefacts due to very high transgene levels and eventually extend the tissues available for study. As noted, transgene expression levels in the heart, skeletal muscle and brain generally paralleled those of the endogenous gene. Transgene expression was also seen in ventricular epithelium and various smooth muscle compartments, also sites of DMPK expression (20,23). Thus, our construct contains the regulatory elements necessary for faithful expression in several tissues and therefore represents an excellent system for the extension of our studies into the heart, brain and smooth muscle.

Skeletal muscle in our mice appeared to be normal at later stages in development but we observed obvious defects at the time at which primary myogenesis occurs (E12–E15) (21). Muscle groups in the neck and face that normally strongly express myogenin were conspicuously unstained, while the somites were normal. At this time, we cannot be certain whether poorly differentiated precursor tissue is present at the correct position or whether muscle groups are ectopically localized, but favor the former explanation. In any case, this observation is particularly intriguing as these muscle groups display the most consistent involvement in DM. At the onset of secondary myogenesis, staining for myogenin and sarcomeric actin again indicated that back, neck and tongue muscle remain poorly developed in both CUG11 and CUG91 transgenic animals, indicating that the delay in myogenesis continues through this stage.

Although the myogenic phenotype observed in CUG11 and CUG91 transgenic embryos was similar, differentiated myoblasts cultured from CUG91 mice had the lowest fusion index compared with CUG11 and wild-type primary cultures. Clearly, the presence of a CUG expansion further impaired the fusion of primary myoblasts, indicating cooperation between the CUG repeat and other 3′-UTR sequences, as previously

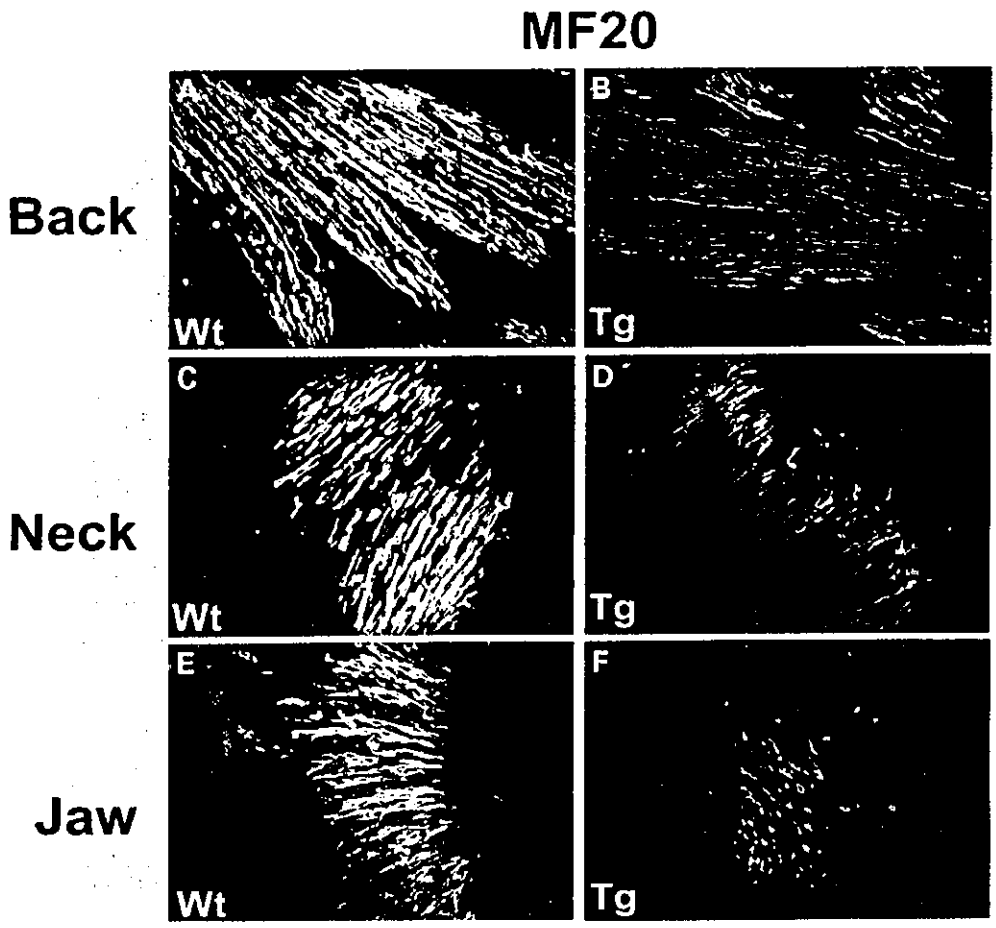

Figure 6. Impairment of muscle differentiation and reduction of total myosin levels in E15 embryos from CUG11 and CUG91 mice. Wild-type and transgenic embryos from two lines of CUG11 (1913, 1923) and CUG91 (2038-3, 2043-1) and wild-type littermates were sectioned and stained with an antibody against total sarcomeric myosin (MF20). Shown are muscles of the back (**A, B**), neck (**C, D**) and jaw (**E, F**) of CUG91 (line 2038-3) mouse and a wild-type control. An identical phenotype was observed with embryos from CUG 91 line 2043-1 and CUG11 lines 1913 and 1923. Myotube formation and expression of total sarcomeric myosin is inhibited *in vivo* as a result of expression of the DMPK 3′ UTR with 11 or 91 CUG repeats in developing muscle.

suggested (16). This exacerbation of the myogenic phenotype occurs despite the typically lower expression of the CUG91 transgene (Table 1).

It should be noted that, although very high levels of GFP have been reported to cause cellular pathology in some cases (24), GFP expression was consistently much higher in the heart of our mice than in other tissues and yet we observed no developmental defects here (data not shown). More importantly, as mentioned, the myoblast fusion deficit was more severe in CUG91 lines despite the fact that they typically expressed less transgene than the CUG11 lines. These results strongly argue against a non-specific toxicity in transgenic mouse skeletal muscle.

While the effects of the transgene were primarily developmental, post-natal defects in muscle composition were also seen. Type I and sometimes type II fiber atrophy was seen in the soleus muscle of 3-month-old animals from both CUG11 and CUG91 transgenic mice, generally consistent with our previous study of *DMPK* overexpression (17). *DMPK* 3′-UTR expression in adult muscle may cause mild muscle degeneration or, alternatively, muscle maturation in younger transgenic mice may continue to be delayed as a result of the delay during development. Obviously, we favor the latter explanation as the fiber atrophy does not persist after 6 months. Type II fiber atrophy was not expected because type I fiber atrophy is predominately observed in DM (1). The type II fiber atrophy may be due to differences between mouse and human physiology or may be the result of ectopic transgene overexpression in type II fibers.

It is not clear at this point how the *DMPK* 3′-UTR inhibits muscle development, but we suspect that it may normally have a negative role during differentiation and that overexpression increases this function to a deleterious state. A role for the 3′-UTR's of structural genes such as tropomyosin during muscle differentiation has been elucidated (25), but in this case these sequences actually promote differentiation by suppressing proliferation. In any case, according to our proposal, mutant repeats may also enhance the negative impact of the DMPK 3′-UTR in a specific fashion, which could only be observed in primary myoblast cultures in this study. One possibility is that

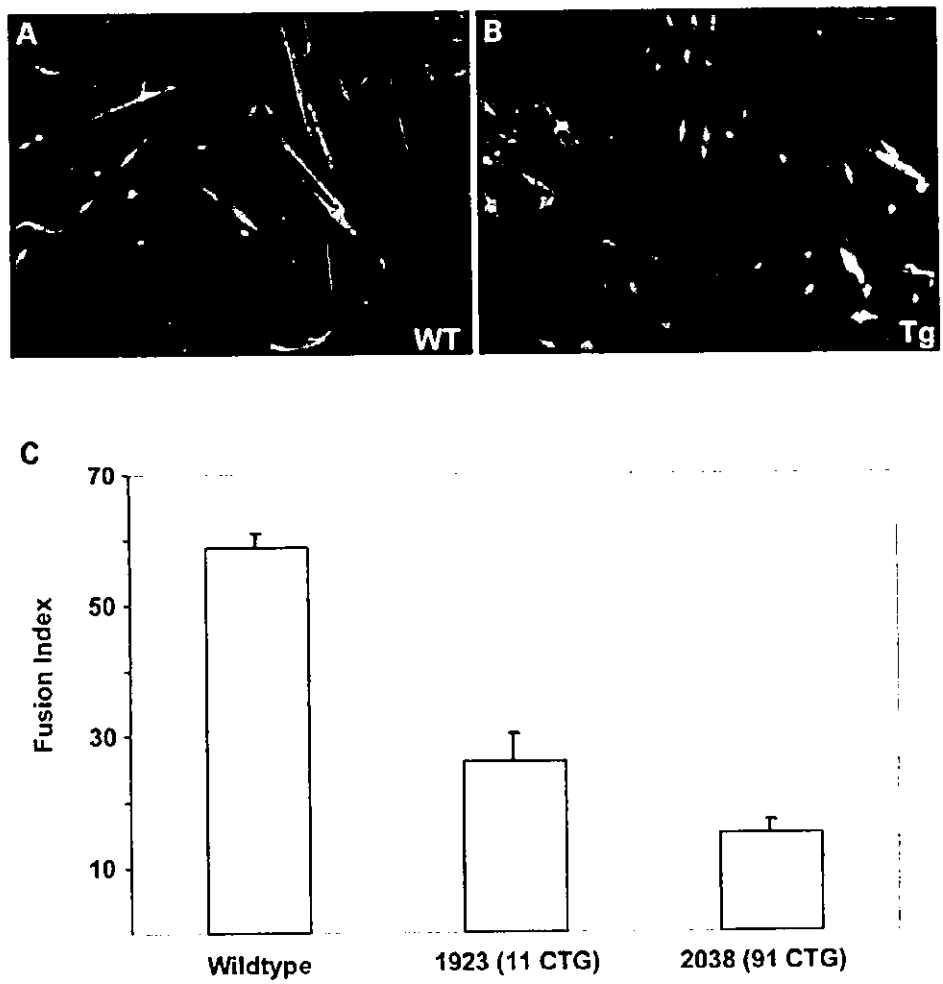

Figure 7. Inhibition of muscle differentiation in myoblasts obtained from CUG11 and CUG91 mice. (A, B) Primary myoblasts were obtained from CUG11 and 91 mice and wild-type littermates, induced to differentiate for 5 days then stained for total sarcomeric myosin (MF20) and DAPI. Shown is a typical field of (A) wild-type and (B) CUG91 myotube cultures. (C) Degree of *in vitro* differentiation (fusion index) was calculated (see Materials and Methods). Myoblasts cultured from both CUG11 and CUG91 transgenic mice exhibited reduced differentiation potential compared with wild-type myoblasts, although the effect was more pronounced in CUG91 myoblasts. Experiment was performed twice from separate myoblast isolations, shown is a representative experiment.

sites for RNA-binding proteins exist in the portion of the 3′-UTR upstream of the repeat and that both overexpression and repeat expansion increase factor binding, such that they become depleted.

Congenital myotonic dystrophy presents with a distinct set of symptoms, notably a profound delay in muscle maturation, while other pathological features seen in adult muscle (e.g. sarcoplasmic masses, centronucleation) are not observed. Infants that survive this phase often have improved status and go on to develop symptoms of the adult disease later (1). Therefore, very large repeats have additional effects that exert themselves at an early developmental phase and may not be simply extensions of the adult disease. As mentioned, MBNL sequestration probably also plays a role in DM skeletal muscle pathology. Therefore, factors that bind upstream of the repeat sequence must operate in addition to the role proposed for MBNL. It is possible that these unknown factors interact with MBNL. Further studies that delineate the possible roles of MBNL during human muscle development and in adult muscle will shed further light in this regard. Finally, recent studies have shown that myotonia in DM is a consequence of mis-splicing of the ClC-1 muscle chloride channel pre-mRNA, ultimately resulting in loss of the protein from the muscle membrane (26,27). Myotonia is a symptom of adults and thus the associated splicing defect may not be directly related to the congenital developmental defect. As mentioned, the α-actin/mutant repeat mice did not exhibit muscle weakness and wasting (26). Taken together, these observations raise the possibility that repeat expansion may cause various symptoms of DM through multiple mechanisms.

In conclusion, we demonstrate for the first time *in vivo* that overexpression of the human *DMPK* 3′ UTR utilizing *DMPK* regulatory elements results in myogenic defects. Aspects of the myogenesis phenotype are exacerbated by inclusion of a CUG repeat expansion within the 3′-UTR. Therefore, the molecular basis of the myogenic defect in DM may involve an interplay

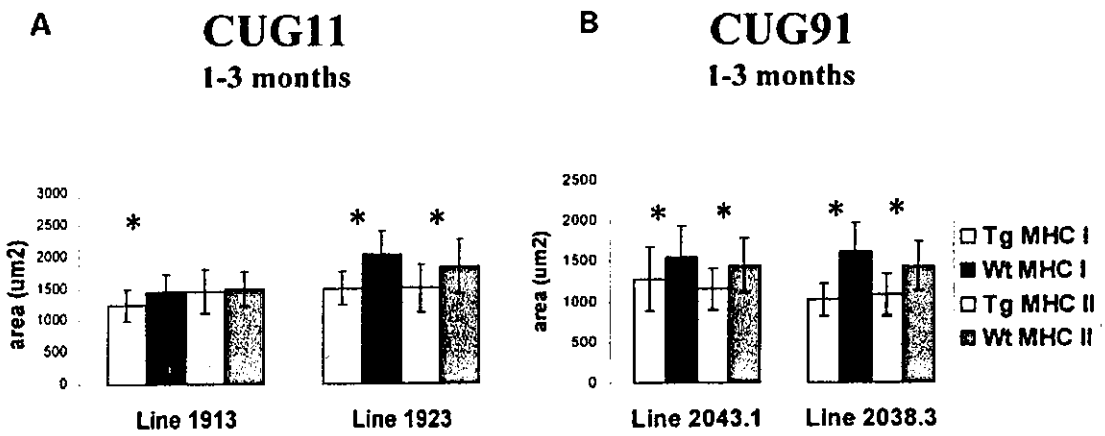

Figure 8. Determination of type I and II fibre sizes from cross-sectioned soleus muscle of CUG11 and 91 transgenic mice. Soleus muscle was cross-sectioned through the middle third of the muscle. This was confirmed by α-bungarotoxin staining to detect neuromuscular junctions (not shown). Mean type I and II fibre areas for two lines of CUG11 (**A**) and CUG91 (**B**) transgenic mice are shown. Both CUG11 lines exhibit significant type I atrophy in the transgenic sample but only line 1923 had type II atrophy. Both CUG91 lines analyzed showed significant type I and II atrophy in the transgenic sample. Significance is indicated with asterisks ($P<0.05$, type 1 fibres, $n=100$, type II fibres, $n=50$).

between the CUG repeat expansion and adjacent 3′-UTR sequences. Elucidation of these molecular interactions will be critical to further our understanding of this disease.

MATERIALS AND METHODS

Transgene construction and generation of transgenic mice

A 1.8 kb *Eco*RV/*Sma*I fragment (GenBank L08835-nt 291–2139) with promoter activity from the *DMPK* 5′-UTR (19) was ligated into pBluescript (SK-). A 2.1 kb fragment from the *DMPK* first intron (nt 2362–4470) with muscle-specific enhancer activity was cloned into a *Not*I site downstream of the DMPK promoter. To omit a first intron ATG codon and not exclude the end of the first intron, a 261 bp fragment (nt 4474–4735) was ligated into an *Spe*I site downstream of the enhancer. A fragment encoding GFP was excised from pEGFP-N1 (Clontech) using *Bam*HI and *Not*I (blunted with Klenow) and cloned into *Bam*HI/*Eco*RI (blunt) sites downstream of the DMPK promoter/enhancer. *DMPK* 3′-UTR fragments were PCR amplified from normal (11 CTG repeats) or patient DNA (91 CTG repeats) cloned with *Bam*HI (blunt) and *Hin*dIII (nt 12129–13408) and ligated into *Eco*RV/*Hin*dIII sites downstream of GFP. The final 3′-UTR fragment (nt 13408–13747) was PCR-amplified and cloned into the *Hin*dIII site to give the transgene constructs. Both transgenes including CTG repeats were verified by sequencing and restriction digests. Transgenes were excised from pBluescript and 2 ng was microinjected into oocytes obtained from the mating of CBA/C57BL6 F_1 mice. Identified founder transgenic mice were crossed onto a C57BL6 background.

Nucleic acid isolation and analysis

Mouse litters were weaned at 3 weeks of age at which time pups were ear-tagged and tail-clipped. DNA was isolated from mouse tails using the DNeasy DNA isolation kit (Qiagen) according to the manufacturer's instructions. DNA was digested with *Bam*HI and electrophoresed on a 1% agarose gel. Southern blotting was performed as described (2). RNA was isolated from animal tissues using Trizol reagent (Gibco-BRL) according to manufacturer's instructions. Northern blotting was performed as described (18).

Real-time quantitative RT–PCR

GFP mRNA (isolated from skeletal muscle, brain and heart) was measured using real-time quantitative RT–PCR as per the Taqman method (6,28). The *GFP* forward and reverse primers (forward 5′ CATGGTCCTGCTGGAGTTC, reverse 5′ TTACTTGTACAGCTCGTCCATGC) and fluorescent probe (probe 5′ [6∼FAM]CCGCCGCCGGGATCACTCTC[Tamra∼Q]),were designed to amplify a 68 bp region of *GFP*. Total RNA was isolated as described above and further purified using Rneasy mini-spin columns combined with DNase I treatment (Qiagen). For each sample, 5 ng of total RNA were used for the analysis. The RNA was reverse-transcribed and PCR amplified using the Taqman EZ RT–PCR kit (Applied Biosystems) on an ABI Prism 7700 Sequence Detector (PE Applied Biosystems). Results were quantified using the cycle threshold (Ct) method and normalized to GAPDH mRNA levels, determined using PE-ABI supplied mouse primers and probe (JOE or VIC-labeled). *GFP* primers and probes were used at concentrations of 600 and 200 nM, respectively. The RT reaction was carried out at 50°C for 2 min, 60°C for 30 min, and 95°C for 5 min and was followed by 45 cycles of PCR at 94°C for 20 s and 60°C for 1 min.

Protein lysates and western blotting

Protein lysates from tissues were obtained by homogenization on ice with a polytron in 10 mM Tris pH 8.0, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM PMSF. Samples were spun for 5 min at

500 rpm in a microfuge; supernatants were adjusted to 2% SDS and boiled for 20 min. Lysates from cultured cells were prepared in RIPA buffer (150 mM NaCl, 1% NP40, 0.5% deoxycholate, 0.1% SDS, 50 mM Tris–HCl pH 7.5, 100 μg/ml PMSF, 2 μg/ml aprotinin, 1 μg/ml pepstatin A, 2 μg/ml antipain). Protein concentrations were determined using a Pierce micro BCA kit according to the manufacturer's instructions. A 30–50 μg sample of lysate was mixed with 3 × loading buffer (187.5 mM Tris–HCl pH 6.8, 6% SDS, 30% glycerol, 0.03% bromphenol blue) and boiled for 5 min. Samples were electrophoresed on a 10% SDS–polyacrylamide gel and transferred to PVDF membrane by semi–dry electroblotting. The membrane was incubated with a polyclonal GFP antibody (Clontech; 1 : 1000 dilution), washed in PBST and visualized by chemiluminescence (ECL kit, Amersham) according to the manufacturer's instructions.

^{33}P *in situ* hybridization

Riboprobes encoding *GFP* and *myogenin* were generated by run-off transcription using a MAXIscript kit (Ambion) supplemented with ^{33}P-rUTP according to the manufacturer's instructions. A 396 bp fragment of the *GFP* cDNA (nt 126–522) was PCR-amplified and cloned into pCR 2.1 (linearized with *Bam*HI) and the most 3′ 449 bp (*Eco*RI fragment) of the myogenin cDNA also cloned into pCR2.1 (linearized with *Bam*HI) were used as templates. Embryos were obtained from pregnant C57 Bl6 mice. The day the vaginal plug appeared was designated E0. Pregnant mice were euthanized by CO_2 and embryos removed at various times between E9.5 and E20. Subsequent procedures were performed as described (22).

Immunostaining of cryosectioned embryos and muscle

Embryos were flash frozen in dry-ice cooled isopentane and sectioned on a cryostat. 12 μm cryosections were dipped in H_2O for 1 min and then fixed in 4% PFA in PBS (8 mM Na_2HPO_4, 2.2 mM NaH_2PO_4, 137 mM NaCl, 2.7 mM KCl) for 3 min at room temperature. All antibody incubations and washes were performed in PBS. Slides were washed three times for 3 min with shaking. Specimens were blocked in 5% serum in PBS at room temperature for 15 min. Sections were incubated with primary antibody solution containing 0.3% Triton X-100 overnight at 4°C, washed three times for 3 min with shaking and incubated with secondary antibody solution (1 : 200 dilution) containing 0.3% Triton X-100. Slides were washed three times for 3 min with shaking and mounted in antifade solution (Dako).

Primary myoblast culture and fusion assay

Primary myoblast cultures were isolated essentially as described (29). Myoblasts were grown to a density of 1×10^4 cells in 24-well collagen coated dishes. Cells were then induced to differentiate in DMEM medium containing 5% FBS for 5 days. Cells were fixed in 4% PFA for 10 min and stained for MHC-I with MF20 hybridoma surpernatant (1 : 5 dilution). Hoechst dye (1 μl of 1 mg/ml) was used to visualize nuclei. Fusion index was calculated by tabulating the percentage of total nuclei located within myotube structures. Ten fields were counted, approximately 400 nuclei per field. The values shown represent the average fusion index of the 10 fields. Two separate primary myoblast isolations were performed on one CUG11 and CUG91 mouse and corresponding littermate controls, shown is a representative experiment.

Muscle sectioning and fiber size determination

Mouse hindlimb muscle was hemisectioned such that the soleus was bisected at the midline. Muscles were submersed in liquid nitrogen cooled isopentane for approximately 30 s, sectioned and adhered to precleaned Superfrost-Plus slides (VWR). Sections were stained for neuromuscular junctions with α-bungarotoxin–CY3 (1 : 500; Molecular Probes) as described to ensure sections were from the middle third of the soleus muscle. Sections were stained with monoclonal antibody BAD5 (1 : 5 dilution) specific for type I myosin heavy chain (MHC-I) as described above except that the secondary antibody was conjugated to horseradish peroxidase. The peroxidase reaction was visualized with 3, 3′-diaminobenzidine tetrahydrochloride (DAB) substrate. Images of type I (DAB positive) and type II (DAB negative) fibres from transgenic and wild-type mice were traced manually and cross section areas determined using the ScionImage program (Scion corporation). Cross sectional areas were determined for one transgenic and wild-type littermate for two lines of CUG11 (1913, 1923) and two lines of CUG91 (2038-3, 2043-1) mice. For type II fibers, 50 cross section areas were included in the calculations and 100 for type I fibers.

ACKNOWLEDGEMENTS

Thanks to Catherine Neville and Jin-Ying Xuan for sequencing. We are grateful to Martine St Jean and Aegera Therapeutics for technical assistance. C.J.S. was supported by the Aurthur Minden pre-doctoral fellowship from the Muscular Dystrophy Association of Canada (MDAC). This work was supported by grants from the MDA (USA), the Canadian Institutes of Health Research (CIHR) and the Canadian Genetic Diseases Network (CGDN). R.G.K. is a CIHR Senior Investigator and a Howard Hughes Medical Institute (HHMI) International Research Scholar.

REFERENCES

1. Harper, P.S. (1989) *Myotonic Dystrophy*, 2nd edn. W.B. Saunders, London.
2. 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.
3. 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*, **69**, 799–808.
4. 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.J. *et al.* (1992) An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science*, **255**, 1256–1258.

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. Amack, J.D., Paguio, A.P. and Mahadevan, M.S. (1999) Cis and trans effects of the myotonic dystrophy (DM) mutation in a cell culture model. *Hum. Mol. Genet.*, **8**, 1975–1984.
7. Bhagwati, S., Shafiq, S.A. and Xu, W. (1999) $(CTG)_n$ repeats markedly inhibit differentiation of the C2C12 myoblast cell line: implications for congenital myotonic dystrophy. *Biochim. Biophys. Acta.*, **1453**, 221–229.
8. 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.
9. 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 [in Process Citation]. *EMBO J.*, **19**, 4439–4448.
10. 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.
11. 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.
12. Fardaei, M., Larkin, K., Brook, J.D. and Hamshere, M.G. (2001) *In vivo* co-localisation of MBNL protein with DMPK expanded-repeat transcripts. *Nucl. Acids Res.*, **29**, 2766–2771.
13. Fardaei, M., Rogers, M.T., Thorpe, H.M., Larkin, K., Hamshere, M.G., Harper, P.S. and Brook, J.D. (2002) Three proteins, MBNL, MBLL and MBXL, co-localize *in vivo* with nuclear foci of expanded-repeat transcripts in DM1 and DM2 cells. *Hum. Mol. Genet.*, **11**, 805–814.
14. 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.
15. Sabourin, L.A., Tamai, K., Narang, M.A. and Korneluk, R.G. (1997) Overexpression of 3′-untranslated region of the myotonic dystrophy kinase cDNA inhibits myoblast differentiation *in vitro*. *J. Biol. Chem.*, **272**, 29626–29635.
16. Amack, J.D. and Mahadevan, M.S. (2001) The myotonic dystrophy expanded CUG repeat tract is necessary but not sufficient to disrupt C2C12 myoblast differentiation. *Hum. Mol. Genet.*, **10**, 1879–1887.
17. Narang, M.A., Waring, J.D., Sabourin, L.A., Rajcan-Separovic, E., Parry, D., Jirik, F. and Korneluk, R.G. (2000) Skeletal myopathy in mice over-expressing the human myotonic dystrophy kinase (DMPK) gene. *Gene Funct. Dis.*, **3**, 1–15.
18. Storbeck, C.J., Sabourin, L.A., Waring, J.D. and Korneluk, R.G. (1998) Definition of regulatory sequence elements in the promoter region and the first intron of the myotonic dystrophy protein kinase gene. *J. Biol. Chem.*, **273**, 9139–9147.
19. Furling, D., Lam le, T., Agbulut, O., Butler-Browne, G.S. and Morris, G.E. (2003) Changes in myotonic dystrophy protein kinase levels and muscle development in congenital myotonic dystrophy. *Am. J. Pathol.*, **162**, 1001–1009.
20. 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.
21. Ashby, P.R., Pincon-Raymond, M. and Harris, A.J. (1993) Regulation of myogenesis in paralyzed muscles in the mouse mutants peroneal muscular atrophy and muscular dysgenesis. *Dev. Biol.*, **156**, 529–536.
22. Sassoon, D., Lyons, G., Wright, W.E., Lin, V., Lassar, A., Weintraub, H. and Buckingham, M. (1989) Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis. *Nature*, **341**, 303–307.
23. Whiting, E.J., Waring, J.D., Tamai, K., Somerville, M.J., Hincke, M., Staines, W.A., Ikeda, J.E. and Korneluk, R.G. (1995) Characterization of myotonic dystrophy kinase (DMK) protein in human and rodent muscle and central nervous tissue. *Hum. Mol. Genet.*, **4**, 1063–1072.
24. Liu, H.S., Jan, M.S., Chou, C.K., Chen, P.H. and Ke, N.J. (1999) Is green fluorescent protein toxic to the living cells? *Biochem. Biophys. Res. Commun.*, **260**, 712–717.
25. Rastinejad, F., Conboy, M.J., Rando, T.A. and Blau, H.M. (1993) Tumor suppression by RNA from the 3′ untranslated region of alpha-tropomyosin. *Cell*, **75**, 1107–1117.
26. Mankodi, A., Takahashi, M.P., Jiang, H., Beck, C.L., Bowers, W.J., Moxley, R.T., Cannon, S.C. and Thornton, C.A. (2002) Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. *Mol. Cell*, **10**, 35–44.
27. Charlet, B.N., Savkur, R.S., Singh, G., Philips, A.V., Grice, E.A. and Cooper, T.A. (2002) Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol. Cell*, **10**, 45–53.
28. Heid, C.A., Stevens, J., Livak, K.J. and Williams, P.M. (1996) Real time quantitative PCR. *Genome Res.*, **6**, 986–994.
29. Sabourin, L.A., Girgis-Gabardo, A., Seale, P., Asakura, A. and Rudnicki, M.A. (1999) Reduced differentiation potential of primary $MyoD^{-/-}$ myogenic cells derived from adult skeletal muscle. *J. Cell Biol.*, **144**, 631–643.