Nuclear localization of MBNL1: splicing-mediated autoregulation and repression of repeat-derived aberrant proteins

Yoshihiro Kino1,2,3,4, Chika Washizu3, Masaru Kurosawa1,2,3, Yoko Oma5, Nobutaka Hattori1, Shoichi Ishiura5 and Nobuyuki Nukina1,2,3,*

1Department of Neuroscience for Neurodegenerative Disorders, Juntendo University Graduate School of Medicine, Tokyo 113-0033, Japan, 2CREST (Core Research for Evolutionary Science and Technology), JST, Saitama 332-0012, Japan, 3Laboratory for Structural Neuropathology, Brain Science Institute, RIKEN, Saitama 351-0198, Japan, 4Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, Tokyo 204-8588, Japan and 5Department of Life Sciences, Graduate School of Arts and Sciences, the University of Tokyo, Tokyo 153-8902, Japan

Received September 1, 2014; Revised September 1, 2014; Accepted September 22, 2014

In some neurological diseases caused by repeat expansions such as myotonic dystrophy, the RNA-binding protein muscleblind-like 1 (MBNL1) accumulates in intranuclear inclusions containing mutant repeat RNA. The interaction between MBNL1 and mutant RNA in the nucleus is a key event leading to loss of MBNL function, yet the details of this effect have been elusive. Here, we investigated the mechanism and significance of MBNL1 nuclear localization. We found that MBNL1 contains two classes of nuclear localization signal (NLS), a classical bipartite NLS and a novel conformational NLS. Alternative splicing of exon 7 acts as a switch between these NLS types and couples MBNL1 activity and intracellular localization. Depending on its nuclear localization, MBNL1 promoted nuclear accumulation of mutant RNA containing a CUG or CAG repeat, some of which produced proteins containing homopolymeric tracts such as polyglutamine. Furthermore, MBNL1 repressed the expression of these homopolymeric proteins including those presumably produced through repeat-associated non-ATG (RAN) translation. These results suggest that nuclear retention of expanded RNA reflects a novel role of MBNL proteins in repressing aberrant protein expression and may provide pathological and therapeutic implications for a wide range of repeat expansion diseases associated with nuclear RNA retention and/or RAN translation.

INTRODUCTION

Expansions of repetitive sequences in non-coding regions cause neurological diseases such as myotonic dystrophy (DM) types 1 and 2 (DM1 and DM2), and spinocerebellar ataxia type 8 (SCA8) (1). The number of non-coding repeat expansion diseases is still growing, as exemplified by recent findings of the mutations causing spinocerebellar ataxia type 36 and chromosome 9p-linked amyotrophic lateral sclerosis/frontotemporal dementia (C9-ALS/FTD) (2–4). One common pathological feature of these diseases is the presence of nuclear RNA inclusions called ribonuclear inclusions or RNA foci. Notably, muscleblind-like 1 (MBNL1) is a common component of ribonuclear inclusions in DM1, DM2, SCA8 and Huntington’s disease-like 2 (HDL2) (5–8). MBNL1 binds directly to CUG and CCUG-repeat RNA (9), which are expressed in these diseases. Recruitment of MBNL1 into ribonuclear inclusions is thought to cause functional depletion of MBNL1 (10,11).

MBNL proteins are evolutionarily conserved RNA-binding proteins with variable numbers of C3H-type zinc finger motifs. In vertebrates, three MBNL orthologs, MBNL1, MBNL2 and MBNL3, have been identified (12). MBNL1 and MBNL2 are expressed in a wide variety of adult tissues including brain, heart and skeletal muscle, whereas MBNL3 is expressed predominantly in the placenta (12). MBNL1 is a splicing factor that directly regulates alternative splicing of numbers of transcripts. Importantly, splicing misregulation of MBNL-regulated genes has been found in DM1 and DM2 tissues (11). Indeed,
In addition, MBNL1 and its orthologs can modulate the toxicity (21). Thus, gene products from an expanded allele are much repeats and accumulates in CAG-repeat RNA foci (9,26,27). MBNL1 binds to CAG repeats in addition to CUG and CCUG more complex than previously recognized. Interestingly, effects of CAG-repeat RNA have been suggested in polyQ diseases, not only CTG/CCTG repeat diseases. Proteins have become potential common players in many repeat expansion diseases, which include Huntington’s disease (HD). An expansion of CAG repeats encoding a polyglutamine (polyQ) tract leads to production of an aggregation-prone protein with various toxic effects (18). Although coding CAG-repeat diseases and non-coding CTG repeat diseases have been attributed to protein and RNA gain-of-function mechanisms, respectively, recent studies have revealed overlapping features between these diseases. CAG-repeat-containing transcripts are expressed from the disease allele and produce polyQ-containing proteins in DM1, SCAS and HDL2 (19–21), whereas toxic effects of CAG-repeat RNA have been suggested in polyQ diseases (22–25). Furthermore, a recent study reports that expanded-repeat RNA can be translated even without an ATG codon (21). In this process, RAN translation, a single repeat tract can produce homopolymeric proteins in multiple frames (21). Thus, gene products from an expanded allele are much more complex than previously recognized. Interestingly, MBNL1 binds to CAG repeats in addition to CUG and CCUG repeats and accumulates in CAG-repeat RNA foci (9,26,27). In addition, MBNL1 and its orthologs can modulate the toxicity of the CAG repeat in animal models (22,25). Thus, MBNL1 proteins have become potential common players in many repeat expansion diseases, not only CTG/CCTG repeat diseases.

A key issue related to both the abnormal regulation and function of MBNL1 in repeat expansion diseases is the alternative splicing of exon 7 in this protein. The inclusion of this exon is elevated in DM1, DM2 and SCA8 patients and model mice (8,11,28). Moreover, the same alteration is observed in Mbnl1Δex3/Δex3 knockout mice, suggesting an autoregulation of exon 7 by MBNL1 (11,29). Furthermore, isoforms containing this exon exhibit nuclear localization (11,30,31), although the mechanism of nuclear localization of MBNL1 is still elusive. Another key issue related to MBNL1 in the nucleus is the precise effect of the RNA foci formation. As demonstrated in DM pathogenesis, nuclear interaction between MBNL proteins and mutant RNA leads to loss of MBNL function. However, it is still unclear whether the colocalization of MBNL1 and repeat RNA in the inclusions reflects some active response of MBNL1 or a passive sequestration of MBNL1. Indeed, recent studies have shown that depletion of MBNL1 reduces ribonuclear inclusions (32–34), supporting the notion that MBNL1 is involved in the process of foci formation.

In this study, we investigated the mechanism, regulation and significance of MBNL1 nuclear localization. We found that nuclear localization is a major determinant of MBNL1 function. We then focused on the regulation of nuclear localization of MBNL1 and identified novel conserved motifs that cooperatively mediate nuclear localization of MBNL1. Alternative splicing of exon 7 acts as a regulatory switch that couples MBNL1 nuclear activity and nuclear localization. Finally, we identified the significance of the interaction between MBNL1 protein and repeat RNA in the nucleus. MBNL1 promotes the nuclear retention of repeat-containing transcripts, which results in repression of aberrant protein expression from the expanded repeats, pointing to a novel aspect of RNA foci formation as well as a potential strategy for preventing protein toxicity in repeat expansion diseases.

**RESULTS**

**Intracellular localization of MBNL isoforms**

Mammalian MBNL proteins show complex splicing variations (Fig. 1A and Supplementary Material, Fig. S1A). We first tried to identify the determinants of MBNL1 properties in relation to splice variations. In this study, subcellular localization of protein or RNA was evaluated by automated fluorescence analysis of individual transfected cells (35). The nucleocytoplasmic index (NCI) ranges from 0 to 1 with higher values indicating stronger cytoplasmic localization in a cell. An averaged NCI value from a cell population was used as a representative value of the transfected construct. Using this system, intracellular localization of EGFP-fused MBNL1 isoforms was analyzed in comparison with the control, EGFP. In COS-7 cells, only MBNL142 showed nuclear localization (Fig. 1B and C). Thus, exon 7, which is included only in this isoform, must be involved in nuclear localization (Fig. 1A). The other isoforms showed more cytoplasm localization with MBNL140s showing the most cytoplasmic localization (Fig. 1B and C). Protein expression of these isoforms was verified by western blot (Supplementary Material, Fig. S1C). We also confirmed that C-terminally myc-tagged MBNL1 isoforms showed subcellular localization similar to that of N-terminally EGFP-tagged ones (Supplementary Material, Fig. S5C).

Next, we compared splicing regulatory activity of these isoforms using an Actn1 minigene (36,37). This minigene contains two mutually exclusive alternative exons, NM (non-muscle) and SM (smooth muscle) (Fig. 1D). In the basal splicing pattern, NM inclusion was predominant. When the amount of MBNL1 was increased, NM inclusion was gradually reduced, whereas SM inclusion was transiently increased and then decreased at maximum MBNL1 dose (Fig. 1D). Among MBNL1 isoforms, MBNL142 had the strongest effect on Actn1 splicing (Fig. 1E), consistent with its strong nuclear localization. Isoforms 41, 40, 41 and 40s exhibited Actn1 splicing regulation weaker than that of MBNL142 but stronger than isoforms 36 and 35 (Fig. 1E and Supplementary Material, Fig. S1D). Both MBNL136 and MBNL135 lacked exon 5 (Fig. 1A), suggesting that the linker region between ZnfF1/2 and ZnfF3/4 encoded by this exon determines the splicing regulatory activity of MBNL1. An N-terminal region of MBNL1 with four ZnF motifs (MBNL1-N) alone showed much weaker activity compared with MBNL142 (Fig. 1F). This reduction can be explained by its weaker nuclear localization (Fig. 1C); the activity of MBNL1-N splicing regulation was fully recovered by the addition of an NLS, whereas it was further reduced by the addition of a nuclear export signal (NES) (Fig. 1C and F). The regulatory activity of MBNL1-N was weakened when exon 5 was deleted (Fig. 1F, MBNL1-NΔex5). However, the presence of exon 5 is not an absolute requirement for splicing regulation, as
NLS-fused NΔex5 showed splicing regulation similar to that of MBNL1Δ2 or NLS-MBNL1-N (Fig. 1F). In conclusion, the splicing regulatory activity of MBNL1 can be attributed to the N-terminus region containing four ZnF motifs and a linker region, although the activity is also determined by the degree of nuclear localization that is mediated by the C-terminus including exon 7. Similarly, MBNL2 has an alternative exon corresponding to MBNL1 exon 7, which is associated with nuclear localization and stronger splicing regulation (Supplementary Material, Fig. S2A–C).
Identification of a nuclear localization signal in the C-terminus of MBNL1

As seen earlier, nuclear localization is a major determinant for the functions of MBNL1. We sought to identify the mechanism of its nuclear localization using EGFP-fused MBNL1 mutants. In both COS-7 and Neuro2a (N2a) cells, a C-terminal region of MBNL1 was essential for nuclear localization, as the deletion of the C-terminus disrupted nuclear localization (1–248, Fig. 2A). In addition, the C-terminus of MBNL1 (C42), but not of the other isoforms (C40, C41, C40s and C41s), was sufficient for inducing nuclear localization (Fig. 2A). Deletion analysis of this region revealed that nuclear localization activity could be narrowed to the regions 276–308, which comprises exon 7 and

![Figure 2](https://academic.oup.com/hmg/article-abstract/24/3/740/2900971/246137402390971)
a part of exon 8 (Fig. 2B and Supplementary Material, Fig. S3A). Consistently, a peptide covering 270–308 was sufficient for driving nuclear localization (Fig. 2B and Supplementary Material, Fig. S3A). Point mutation analysis of this region in the context of full-length MBNL1 demonstrated the importance of four basic amino acids, K278, R279, K299 and R300 (Fig. 2C and Supplementary Material, Fig. S3B). Single point mutation of either R279A or K299A was sufficient for disrupting nuclear localization. In the case of K278A and R300A, single mutation showed little or no effect, whereas the simultaneous mutation of K278A and R300A completely disrupted nuclear localization (K278A/ R300A, Fig. 2C and Supplementary Material, Fig. S3B). We also examined the effect of NLS mutations on the splicing activity. Both R279A and K299A mutants showed weakened splicing regulation of MBNL1, even though the protein expression of these mutants was equivalent to that of wild type (WT) (Supplementary Material, Fig. S3C and D).

All MBNL1 isoforms analyzed in this study, except for MBNL142s, contain a highly conserved motif in exon 8, KRPSE (Fig. 2D, Supplementary Material, Fig. S4A, and pink boxes in Fig. 2B). A corresponding motif in the fruit fly is involved in nuclear localization (30). However, this motif was not sufficient for mammalian MBNL1 nuclear localization, as demonstrated by the non-nuclear localization of C40 and 288–388, which contain this motif (Fig. 2B and Supplementary Material, Fig. S3A). The identified residues essential for nuclear localization were partly similar to a classical bipartite NLS (bNLS), which is typically summarized as KRX(10-12)KR and acts through the importin α/β pathway. Co-expression of red fluorescent protein (RFP)-fused Bimax2 peptide, a specific inhibitor of the importin α/β pathway (38), prevented nuclear localization of C42 (Fig. 2E). Therefore, the NLS in the C-terminus of MBNL1 appears to be a bipartite NLS, even though the spacing of two KR motifs (19 amino acids) is larger than a typical bNLS (10–12 amino acids). In conclusion, both exon 7 and the conserved motif of exon 8 provide KR motifs that constitute a bNLS only when exon 7 is included in the alternative splicing.

The second NLS activity of MBNL1 involves multiple conserved motifs

During the above-mentioned analysis, we noticed a cell-type-dependent difference in nuclear localization of MBNL1. In N2a cells, MBNL140 showed significant nuclear localization despite the absence of exon 7 (Fig. 2A). Interestingly, Bimax2 disrupted nuclear localization of MBNL142 in COS-7 but only partially in N2a cells (Fig. 2E). Furthermore, nuclear localization of MBNL140 in N2a was not affected by Bimax2 (Fig. 2E). These results suggested that MBNL1 contains another NLS independent of the importin α/β pathway. We performed deletion analysis of MBNL140 (identical to 1–388Δex7 in Fig. 3A, a designation based on MBNL142 to maintain the amino acid numbering). In N2a cells, nuclear localization of MBNL140 required at least parts of both N- and C-terminal regions (Fig. 3A). Comparison of 1-308Δex7 and 1–298Δex7 demonstrated the necessity of the region around the conserved motif in exon 8 (pink box, Fig. 3A). In addition, deletion of an N-terminal region containing ZnF1/2 resulted in the loss of nuclear localization, as shown by 71–388Δex7 and 71–308Δex7 (Fig. 3A). Thus, the region around the conserved motif in exon 8 and the N-terminal region were essential for nuclear localization. We then tested point mutations of the conserved motif on exon 8 in the context of MBNL140. Similar to the case of MBNL142, mutation of K299A and R300A disrupted the nuclear localization of MBNL140 (Fig. 3B). However, other residues in the conserved motif were also essential for nuclear localization as revealed by the moderate to severe effects of P301D, A302D, L303A and E304A (Fig. 3B). This stood in contrast to the case of MBNL142, in which mutations of these residues did not affect nuclear localization (Fig. 2C).

Next, the involvement of the N-terminal region was examined. ZnF1/2 is a very highly conserved motif through evolution. We noticed that flanking regions of ZnF1/2 are also highly conserved and contain motifs, RD/KWL and KxQL/NGR (Fig. 3C and Supplementary Material, Fig. S4A). These motifs are not found in the flanking regions of ZnF3/4 of MBNL1 or other C3H ZnF proteins. Deletion of the first 13 amino acids from the N-terminus disrupted nuclear localization (Δ13, Fig. 3C). This could be partly explained by the deletion of the RD/ KWL motif as shown by its mutation, mut(9–14) (Fig. 3C). Deletion of the KxQL/NGR motif, Δ(76–85), moderately weakened nuclear localization of MBNL1 (Fig. 3C). Interestingly, ZnF1/2, but not ZnF3/4, was found to contribute to nuclear localization (ΔZnF1/2 and ΔZnF3/4, Fig. 3C). Therefore, the N-terminal region of MBNL1 that comprises ZnF1/2 and its flanking motifs and the C-terminal conserved motif in exon 8 are essential for a second NLS activity of MBNL1. In contrast to the bNLS that comprises a linear motif, the second NLS was mediated by multiple and discrete regions of MBNL1. This is consistent with the recent classification of a conformational NLS (conNLS) (39).

We then analyzed whether the conNLS can act in the context of MBNL142 in N2a cells. A mutation of the KR motif in exon 7, which disrupts the bNLS, weakened but still preserved nuclear localization [1–388(K278A/R279A), Fig. 3D], suggesting that conNLS activity is present in MBNL142. However, deletion of a C-terminal region including the conserved motif of exon 8 also weakened but still preserved nuclear localization (Fig. 3D, compare 1–308, 1–298 and EGFP). This result was unexpected as this deletion was predicted to disrupt both bNLS and conNLS. Further deletion of exon 7 resulted in the loss of nuclear localization (1–269, Fig. 3D). Thus, exon 7 is involved in nuclear localization even in the absence of exon 8. As exon 7 itself was not sufficient for inducing nuclear localization (239–269, Fig. 2B), this effect of exon 7 requires the presence of the N-terminal region. We noticed that the sequence around the KR motif in exon 7 (KRPSE) is similar to that of the conserved motif in exon 8, KRPSE (Fig. 2D). When the KR motif of exon 7 was mutated in the context of 1–298, the mutant, 1–298 (K278A/R279A), exhibited loss of nuclear localization similar to 1–269 (Fig. 3D). Therefore, it appears that exon 7 contains another copy of the conserved motif and constitutes a conNLS together with the N-terminal region. In conclusion, MBNL1 has two classes of NLS (Fig. 3E). One is the bNLS consisting of two KR motifs embedded in conserved motifs of exons 7 and 8. The other is the conNLS, which involves the N-terminal region and either one of the conserved motifs in exon 7 or exon 8. Thus, MBNL142 possesses three possible NLS combinations, bNLS, conNLS using exon 7 and conNLS using exon 8. Most other isoforms contain only one conNLS using exon 8. MBNL140 lacks any known NLS.
Splicing-mediated coupling of MBNL activity and subcellular localization

As demonstrated earlier, exon 7 is an important determinant of the localization and function of MBNL1. In both N2a cells and adult mouse striatum, we detected Mbnl1 isoforms with and without exon 7 (Supplementary Material, Fig. S1A). Endogenous Mbnl1 was localized in the nucleus and cytoplasm (Supplementary Material, Fig. S5A). To analyze splicing regulation of this exon, we analyzed a minigene covering exons from 6 to 8 of mouse Mbnl1. Overexpression of the N-terminal regions of all three Mbnl paralogs strongly reduced the inclusion of exon 7 (Fig. 4A). In contrast to the case of Actn1, the N-terminus of Mbnl proteins did not require an NLS for strong repression of Mbnl1 exon 7, suggesting the high sensitivity of this exon to Mbnl proteins. Co-expression of the Mbnl1 minigene and a DMPK construct with an interrupted CTG480, but not CTG18, increased the inclusion of exon 7 (Fig. 4B). We also observed a small but significant increase in exon 7 inclusion when the
CAG480 construct was overexpressed (Fig. 4B). RNAi-mediated knockdown of endogenous Mbnl1 induced a moderate increase in the inclusion of exon 7, whereas knockdown of Mbnl2 alone exhibited little effect (Fig. 4C and Supplementary Material, Fig. S2D). Importantly, simultaneous knockdown of both Mbnl1 and Mbnl2 increased exon 7 inclusion more strongly than that of Mbnl1 alone (miMbnl1/2, Fig. 4C and Supplementary Material, Fig. S2D). These results demonstrated that alternative splicing of Mbnl1 exon 7 is regulated by MBNL proteins in a dose-dependent manner. We confirmed intracellular association between EGFP-MBNL1 and endogenous Mbnl1 transcripts by ribonucleoprotein immunoprecipitation analysis using a cell line stably expressing EGFP-MBNL1N (Supplementary Material, Fig. S4B). Thus, MBNL1 associates with its transcript and autoregulates alternative splicing.

As the Mbnl1 minigene contains a region corresponding to the biNLS identified earlier, we expected that the localization of EGFP fused upstream of the minigene might be changed according to the splicing pattern of exon 7. However, the EGFP fluorescence of this construct was not strong enough to be detected clearly. We made another shortened minigene (Mbnl1#2, Fig. 4D), which still contained the biNLS region and showed stronger fluorescence than the original minigene. A N2a-based cell line stably expressing this minigene showed EGFP fluorescence in both nucleus and cytoplasm (Supplementary Material, Fig. S4C). Patterns of minigene-derived protein or RNA were altered when MBNL1 was overexpressed or knocked down (Fig. 4D). The patterns of the transcript and the protein expressed from the minigene were altered depending on the dose of MBNL1 (Fig. 4D). Notably, cytoplasmic localization of EGFP...
was enhanced when MBNL1 was transfected, whereas nuclear localization was enhanced when endogenous Mbnl1 and Mbnl2 were knocked down (Fig. 4E and Supplementary Material, Fig. S4C). Thus, alternative splicing of Mbnl1 exon 7 directly reflects the activity of MBNL proteins and alters nucleocytoplasmic localization of MBNL1. This mechanism can be regarded as a dynamic negative feedback where an increase in nuclear MBNL activity leads to a decrease in the production of nuclear isoform containing exon 7.

MBNL1 promotes nuclear accumulation of RNA with expanded CUG and CAG repeats

Next, we analyzed the effect of MBNL1 on the localization of transcripts containing expanded CUG or CAG repeats. It has been known that CUG repeat-containing RNA is retained in the nucleus and co-aggregates with MBNL1 (10). As reported previously (32), knockdown of MBNL proteins greatly reduced RNA foci in DM1 fibroblasts (Supplementary Material, Fig. S11B and C), indicating an essential role of MBNL proteins in RNA foci formation. We then used constructs that contain an interrupted repeat similar to the ones widely used in DM1 studies (Supplementary Material, Fig. S6A). CUG-repeat RNA in the context of a DMPK fragment (DMPK3’-CTG480) was visualized by fluorescence in situ hybridization (FISH). Unexpectedly, many cells showed RNA localization in the cytoplasm (Fig. 5A). In contrast, co-transfection of MBNL1.42 with the repeat constructs greatly enhanced nuclear accumulation of repeat RNA mainly as foci (arrows, Fig. 5B and Supplementary Material, Fig. S7A and B). We quantified the relative nucleocytoplasmic distribution of repeat RNA based on the FISH signal. Nuclear localization of CUG repeat was enhanced by MBNL1.42 compared with EGFP (Fig. 5C). Mutation analysis revealed that the enhancement of nuclear RNA localization was dependent on the composition of zinc finger motifs (comparison of MBNL1-N versus 71–248), linker region (MBNL1.40 versus MBNL1.35 and 71–388 versus 174–388) and nuclear localization (NLS-MBNL1-N versus NES-MBNL1-N) (Fig. 5C and Supplementary Material, Fig. S8A, and see also Fig. 2A for the localization of MBNL1 mutants). We also tested non-coding repeats in a DMPK-independent context, in which CTG or CAG repeat is downstream of a CMV promoter and no ATG codon is located upstream of repeat tract (Fig. 5D). EGFP-MBNL1.42 did not colocalize with them (Supplementary Material, Fig. S9A). Moreover, both overexpression and knockdown of MBNL1 did not affect the localization of CAA-repeat RNA (Supplementary Material, Fig. S9B and C).

Thus, the effect of MBNL1 is dependent on repeat sequences.

MBNL1 suppresses the expression of aberrant proteins containing an expanded CUG or CAG repeat

A recent study reports protein expression from expanded repeats in the absence of an initiation codon, in which the repeats tract could be translated in all three frames (21). We examined whether any proteins are expressed from the non-coding repeat constructs used in Figure 5D (CTG240 and CAG240). CAG240 produced a high-molecular-weight protein detected by antibody against V5-tag, which is located downstream of the repeat tract (Fig. 6A). We also examined different lengths of CAG and CTG repeats. Interestingly, CAG constructs expressed V5-positive proteins in a length-dependent manner with peak expression at 360 repeats. In contrast, CTG constructs did not express proteins detected by anti-V5 (Fig. 6A). Owing to 5-base interruptions for every 20 CAG or CTG repeats, the interrupted CAG and CTG repeat tract encodes (Q20-S22-A20)$_n$ and (C20-A20-L21)$_n$ polypeptides, respectively, if translated (Supplementary Material, Fig. S6A). The large protein expressed from the CAG construct was detected by 1C2 antibody, indicating the presence of polyQ tracts. In the absence of an apparent start codon upstream of the repeat tract, the reading frame of the repeat is not clear. We examined three different reading frames of the V5 tag downstream of the repeat tract (Fig. 6B); reading frame 1 was used in the above-mentioned experiments. CAG-repeat constructs with different reading frames produced high-molecular-weight proteins detected by anti-V5 and 1C2 antibodies (Fig. 6B). These constructs showed some difference in the pattern of lower-molecular-weight proteins (corresponding to 40–70 kDa) (Fig. 6B). In contrast, little or no protein was expressed from the CTG constructs with different reading frames of the V5 tag (Fig. 6B). Next, we tested the effect of MBNL1 on the expression of the protein from the CAG construct. When the CAG360 construct was co-expressed with MBNL1.42 or MBNL2.42, the anti-V5-reactive product was decreased compared with control (Fig. 6C). MBNL1.42 enhanced nuclear localization of CAG360 mRNA (Supplementary Material, Fig. S8B). Simultaneous knockdown of both Mbnl1 and Mbnl2 increased the expression of CAG360 (Fig. 6D). Therefore, nuclear retention of repeat-containing RNA induced by MBNL1 can result in the reduction of aberrant proteins expressed from the repeat RNA.

MBNL1 suppresses the expression of polyglutamine-containing proteins

As shown earlier, MBNL1 can alter the localization of seemingly non-coding CAG repeats and reduce the expression of
polyQ-containing proteins. We next asked whether MBNL1 can alter the expression from apparently coding CAG repeats. For this purpose, the effect of MBNL1 on the localization of huntingtin exon 1 RNA was tested. Similar to the results in Figure 5, MBNL1 enhanced the nuclear localization of CAG-containing mRNA of huntingtin, as seen by FISH analysis.
When protein expression was examined, co-expression of EGFP-MBNL1 greatly reduced mutant huntingtin protein fused with RFP or V5 tag detected mainly at the gel top (Fig. 7C and Supplementary Material, Fig. S6B). Reduction in protein expression was much milder for huntingtin with CAG18 (Fig. 7C). Knockdown of both Mbnl1 and Mbnl2 resulted in the accumulation of mutant huntingtin (Fig. 7D). Depletion of Mbnl1/2 also led to cytoplasmic localization of mutant huntingtin mRNA and increased the cytotoxicity (Fig. 7E and F). RNAi-resistant mutant of MBNL142, MBNL1(res)42, but not WT MBNL142 counteracted the effect of Mbnl1/2 knockdown (Supplementary Material, Fig. S10). The expression of RFP-fused CAG78 encoding Q78 was reduced by the expression of MBNL1 (Supplementary Material, Fig. S6C). Furthermore, MBNL1 reduced the expression of EGFP-fused CAG78 in different reading frames, which expressed polyglutamine, polyalanine or polyserine (Supplementary Material, Fig. S6D). Thus, the repressive effect of MBNL1 was independent of the context of huntingtin and was not restricted to the reading frame of polyQ. In conclusion, MBNL1 has a repressive effect on the expression of CAG-repeat-derived proteins.

In primary cultured mouse cortical neurons, we observed RNA foci and RNA localization in the nucleus and cytoplasm when DMPK3′-CTG480 or mutant huntingtin exon 1 was transfected with EGFP, whereas co-expression of MBNL142 enhanced the localization of repeat RNA in the nuclear foci (Fig. 8A and Supplementary Material, Fig. S11A). We also analyzed the striatum of HD model mice, the R6/2 strain. Transcripts
from Huntingtin transgene were detected by FISH in the transgenic (TG) mice but not in the WT control (Fig. 8B). Mutant RNA were localized mostly in the cytoplasm, and we hardly observed RNA foci in the nucleus (Fig. 8B). Mbnl proteins were distributed in both nucleus and cytoplasm, and the staining patterns were similar between WT and TG mice, except for slight reduction of Mbnl2 staining in TG mice reflecting reduction in its mRNA level (Fig. 8C, Supplementary Material, Fig. S2G).

Alternative splicing of Mbnl1 exon 7 was not altered in TG mice (Fig. 8D), suggesting that the function of Mbnl proteins was preserved in TG mice. In cells derived from striatal neurons of HD knock-in mice (StHdh 111/111), we did not detect specific FISH signals above background, as we observed diffuse staining in StHdh 7/7 that may reflect CAG-containing transcripts other than huntingtin (Fig. 8E, upper panels). However, overexpression of Mbnl1 localized RNA foci in StHdh 111/111 cells (Fig. 8E, lower panels). These results suggest that the expression of expanded-repeat transcripts is not sufficient for RNA foci formation in some cell-types. However, overexpression of MBNL proteins could induce RNA foci formation and/or nuclear RNA retention in such cells.

**DISCUSSION**

**Interplay of conserved motifs determines nuclear localization of MBNL1**

Though previous studies indicated that exon 7 of MBNL1 is involved in nuclear localization (11,31,40), the mechanism has been uncharacterized. Here, we clarified the role of exon 7 in nuclear localization. Two clusters of KR residues in the C-terminus region were essential for the function of the NLS.
Figure 8. Localization of expanded RNA and MBNL proteins in cultured neurons and disease model cells. (A) Subcellular localization of mRNA of DMPK3′-CTG480 in primary cultured cortical neurons. The localization of repeat RNA was detected by FISH using Cy3-labeled CAG8. MAP2 serves as a marker of neurons. (B) FISH analysis of huntingtin transgene in the striatum of R6/2 mice. Mutant huntingtin mRNA was detected using DIG-labeled antisense RNA of human huntingtin. (C) Immunohistochemical analysis of Mbnl proteins in the striatum of R6/2 mice. (D) Splicing analysis of Mbnl1 exon 7 in the striatum of R6/2 mice. The inclusion of exon 7 was not different between WT and TG mice (n = 4). (E) FISH analysis of StHdh cells transfected with EGFP-MBNL142. RNA foci were detected in StHdh 111/111 cells expressing EGFP-MBNL1 (arrows), but not in untransfected cells. Scale bars indicate 10 μm.
The activity of this NLS was inhibited by a specific inhibitor of the importin alpha/beta pathway (Bimax2), clearly demonstrating that the two KR clusters constitute a biNLS. We also found the second NLS that is dependent on the cell type. This latter NLS involves multiple conserved regions of the protein (Fig. 3E). None of the conserved motifs, namely KRPAL, ZnF1/2, RD/KWL and KxQL/NGR, was sufficient for the NLS activity by itself. Thus, the second NLS is consistent with a conNLS (39). It is currently unclear which import pathway mediates the activity of the conNLS. As ZnF1/2 is involved in RNA-binding activity (9), the status of RNA binding of MBNL1 might be related to the activity of the conNLS. We also found that MBNL1 contains two KRP(A)LE motifs in exons 7 and 8 and that each of these motifs can act as a part of the conNLS. Thus, MBNL142 can utilize two combinations of the conNLS in addition to the biNLS. Therefore, the inclusion of exon 7 acts as a switch for multiplying the pathways of nuclear import, ensuring efficient nuclear localization. Finally, the residues in the KRP(A)LE motifs are differently required in the context of biNLS and conNLS (Figs 2C and 3B), providing a unique example of motif usage in two different import pathways.

Alternative splicing of exon 7 couples nuclear activity and intracellular localization of MBNL1

It has been suggested that MBNL1 autoregulates exon 7 splicing (11,29). Mouse studies also suggested cross-regulation among MBNL proteins (11,41). Here, we show that alternative splicing of exon 7 directly couples activity of MBNL proteins and nuclear localization of MBNL1 at the cellular level. Overexpression of any Mbnl paralogs repressed exon 7 inclusion, whereas simultaneous knockdown of Mbnl1 and Mbnl2 was required for efficient skipping of exon 7 (Fig. 4A and C). MBNL2 has an alternative exon corresponding to exon 7 of MBNL1, which is involved in the nuclear localization as well as splicing activity of MBNL2 (Supplementary Material, Fig. S2B and C) and is misregulated in DM1 (11). These explain why knockdown of either Mbnl1 or Mbnl2 showed weak effects on the splicing regulation and the repression of repeat-derived proteins (Figs 4C, 6D and 7D), because a reduction in one MBNL protein would be compensated by the increased nuclear isoforms of the other MBNL protein. Importantly, we showed that the NLS activity of MBNL1 closely reflects its splicing activity in a dynamic manner using a cell line expressing a fluorescent minigene, demonstrating that depletion of Mbnl proteins induced nuclear accumulation of the Mbnl1 minigene (Fig. 4D and E). Moreover, co-expression of repeat RNA and MBNL1 per se did not induce translocation of cytoplasmic MBNL1 into the nucleus (Fig. 5E), suggesting that an extensive nuclear accumulation of MBNL1 observed in disease conditions may require enhanced production of nuclear isoforms. Altogether, we established an activity-dependent nuclear localization of MBNL1, in which MBNL isoforms containing exon 7 are predominantly produced when the nuclear amount or activity of MBNL proteins is low, whereas isoforms without exon 7 are predominantly produced after the amount of nuclear MBNL proteins reaches a certain amount. This mechanism would keep the amount of nuclear MBNL protein to a certain level in an autoregulatory feedback loop (Fig. 9A) and suggests the status of exon 7 splicing as an indicator of MBNL activity. Moreover, the increase of nuclear isoforms in disease conditions would accelerate the interaction between MBNL1 and CUG-repeat RNA in the nucleus (Fig. 9B).

Repression of aberrant protein expression by MBNL1

Previous reports suggest that knockdown of MBNL proteins reduces RNA foci formation (32–34). We have shown that MBNL1 promotes the accumulation of expanded-repeat-containing RNA in the nucleus. As a consequence of nuclear RNA retention, MBNL1 represses the expression of mutant protein from transcripts containing an expanded CAG repeat (Figs 6 and 7). The effect of MBNL1 was observed for both coding and non-coding repeat tracts. To date, mislocalization or accumulation of MBNL1 in the nuclear foci has been thought to cause a loss of MBNL1 function leading to misregulation of alternative splicing (Fig. 9B). We propose that such accumulation of MBNL1 has another aspect, the repression of cytoplasmic transport of mutant RNA and its subsequent translation (Fig. 9B). This process might be regarded as an RNA quality control mechanism that limits the production of aberrant proteins from non-physiological transcripts with certain structural properties. Thus, MBNL1 is not simply a victim of expanded repeats but can be a guardian against protein toxicity. As MBNL proteins have cytoplasmic roles (42,43), we do not rule out the contribution of cytoplasmic MBNL1 in repressing repeat-derived proteins.

We observed protein expression from non-coding repeats with interruptions, which have not been examined previously (21). We observed the expression of polyQ-containing proteins from interrupted CAG repeats, which were repressed by MBNL proteins (Fig. 6C). Importantly, similar results were obtained for mutant huntingtin with a pure CAG repeat (Fig. 7C), suggesting that some important properties of pure repeats are retained in interrupted repeats. We hardly observed protein expression from interrupted CTG repeats even when Mbnl proteins were depleted (data not shown). There might be an orientation-dependent preference in the expression of these repeats. Consistently, RAN translation is dependent on the context surrounding the repeat (21). Alternatively, the homopolymeric proteins may have different susceptibility to degradation.

RNA-binding proteins as potential modifiers of repeat expansion diseases

Our results suggest MBNL proteins as potential modifiers of polyQ diseases. Interestingly, MBNL1-containing RNA foci were recently detected in the fibroblasts of human HD and spinocerebellar ataxia type 3 (24,44). However, RNA foci were hardly detected in HD model mice and cells derived from HD knock-in mice, in our hands. There might be cell-type-dependent differences in the interaction between MBNL proteins and repeat RNA. Mbnl2 is a major MBNL protein in the mouse brain (Supplementary Material, Fig. S2F). We observed a slight reduction in the expression of Mbnl2 (Fig. 8C and Supplementary Material, Fig. S2G). However, alternative splicing of Mbnl1 exon 7 was unchanged, suggesting that the activity of Mbnl proteins was not markedly altered in HD mice. In contrast, MBNL proteins are found in RNA foci in DM1 and SCA8 that is accompanied by splicing misregulation of Mbnl1 exon 7 (8,11,28). These
results are consistent with the notion that repeat-MBNL interaction induces RNA foci formation that have two consequences, reduced MBNL activities and suppression of repeat-derived proteins (Fig. 9B). In HD mice and StHdh cells, this interaction was somehow inefficient, possibly because (i) expanded CAG repeat binds to MBNL proteins less strongly than CUG repeat, (ii) the length of CAG repeat was relatively short, (iii) the level of endogenous MBNL proteins or mutant RNA was not sufficient and/or (iv) some co-factor or modifications of Mbnl proteins were lacked in the brain of HD mice and/or StHdh cells. However, in all cell types we have tested, overexpression of MBNL142 induced RNA foci formation, suggesting that a certain level of nuclear MBNL proteins can overcome the limitations mentioned earlier. Our results would warrant further investigations of HD mouse models based on the overexpression of MBNL proteins.

MBNL proteins might provide a basis for the development of therapies to reduce aberrant protein expression associated with repeat expansion. For example, compounds that mimic MBNL1 and induce nuclear retention of mutant transcripts would prevent toxic effects of homopolymeric proteins (45). Another therapeutic implication is that MBNL-mediated
nuclear RNA retention would facilitate selective degradation of mutant RNA, as RNA in the nucleus is susceptible to antisense oligonucleotide-mediated degradation (46).

Recently, dipeptide repeat proteins were found in C9-ALS/FTD that is caused by a hexanucleotide repeat expansion (47–52). In addition, polyglycine-containing protein was found in fragile X-associated tremor/ataxia syndrome (FXTAS) (50). These repetitive proteins are generated from non-coding tracts through RAN translation. The pronounced expression of expanded-repeat-derived proteins in HD, C9-ALS/FTD and FXTAS indicates that nuclear retention of expanded RNA might be inefficient in the cells of these diseases. Remarkably, the distributions of RNA foci and dipeptide repeat inclusions were largely segregated in the brain regions of C9-ALS/FTD (52), consistent with our model of reciprocal relationship between RNA foci formation and RAN translation (Fig. 9B) and its dependence on cell types. Owing to RAN translation, all transcribed repeat tracts may potentially have both RNA and protein toxicities (21). The relative contribution of these toxicities in each disease would be an important future subject. Our results suggest the importance of RNA-binding proteins as key factors involved in both RNA- and protein-mediated pathogenesis of repeat expansion diseases.

**MATERIALS AND METHODS**

**cDNA clones and constructs**

cDNA of human MBNL1, MBNL2 and MBNL3 were described previously (37). To facilitate mutagenesis, a silent mutation of L281L (c843t), which eliminates a XhoI site, was introduced into MBNL1 cDNA and the mutant was designated Mbnl1/L281L. For making Mbnl1 deletion, L281L was used as a PCR template together with a BamHI-added forward primer and an XhoI-added reverse primer. The amplified fragments were digested with BamHI and Xhol and then inserted in the BglII-SalI restriction sites of pEGFP-C1 (Clontech). N-terminus regions of Mbnl1, Mbnl2 and Mbnl3 were described previously (37). Constructs encoding an Actnl minigene and RFP-Bimax2 were described previously (35,37). A DMPK3′-N3 construct with CTG18, interrupted CTG480, interrupted CTG900 or interrupted CAG480 contained two terminal exons as well as the intervening intron of human DMPK gene, which were excised from RFP-DMPK3′ constructs (37) and inserted into pCMV-N3, a tag-less vector made by deleting the EGFP fragment from pEGFP-N3 (Clontech). The Mbnl1 minigene fragment covering exon 6 to exon 8 was amplified from mouse genomic DNA by PCR using a primer set of BamHI-Mbnl1-ex6-Fw and SalI-Mbnl1-ex8-Rv, in which restriction sites for BamHI and SalI were added, respectively. These fragments were cleaved by BglII and the SalI and then subcloned into the BglII-SalI restriction sites of pEGFP-C1. Non-coding CAG or CTG repeat constructs were made by inserting interrupted repeat fragments into the XhoI site of pcDNA3.1 V5-His A, B or C (Invitrogen). Constructs for microRNA-based RNAi were described previously (37). miMbnl1-236 and miMbnl2-675 were used for knockdown of murine endogenous Mbnl1 and Mbnl2, respectively. For simultaneous knockdown of Mbnl1 and Mbnl2, miMbnl1-146 was utilized (37).

**Cell culture and transfection**

COS-7 and N2a cells were maintained in Dulbecco’s modified Eagle’s medium and 10% fetal bovine serum. For minigene assays, cells were typically cultured in 12-well plates and transfected with 0.5 μg of plasmids for protein expression (or cognate empty vector) and 10 or 20 ng of plasmids for the expression of a minigene using Lipofectamine 2000 (Invitrogen) according to the manufacturer’s instructions. When multiple plasmids were transfected for protein expression, the total amount of plasmids was adjusted to 0.5 μg. For microscopic studies, cells were plated in 8-well chamber slides and transfected with 120 ng of plasmids in total. For subcellular localization studies, cells were plated in a 24-well or 96-well plate. The cell line expressing EGFP-MBNL1(46) was described previously (37). The following cell lines were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research (GM04647:DM1000 and GM04795:control). SHdh cells were obtained from Dr. Macdonald and CHDI through Coriell Institute. These cell lines were maintained as recommended by the supplier. We used FuGENE HD (Promega) diluted in OPTI-MEM (Invitrogen) for transfection of SHdh cells.

**Cellular splicing assay**

Cellular splicing assay was performed as previously described (37). For Mbnl1 splicing, minigene fragments were amplified by PCR using a fluorescein isothiocyanate (FITC)-labeled forward primer for a 3′ region of EGFP sequence, FITC-GFP-Fw, and a gene-specific reverse primer, Mbnl1-splicing-Rv2. PCR products were resolved by 2.0–2.5% agarose gel electrophoresis. The fluorescence of PCR products was captured and visualized by LAS-1000 (FUJIFILM) or ImageQuant LAS 4000 (GE Healthcare Japan). Intensity of band signals was analyzed using Multigauge software (FUJIFILM).

**Quantitative PCR**

Gene-specific primers were designed using Primer Express software (Applied BioSystems) and are listed in Supplementary Material, Table S1. cDNA and gene-specific primer sets were mixed with Power SYBR Green PCR Master Mix (Applied Biosystems) or Fast Start Universal SYBR Green Master (Roche). Real-time amplification and quantification were performed using ABI7700 (Applied BioSystems) or LightCycler 480 II (Roche) following the manufacturer’s protocol.

**Fluorescence in situ hybridization**

Fluorescence in situ hybridization was performed essentially as previously described with minor modifications (6,53). Cells plated in the 8-well chamber slides or 96-well plates were transfected using Lipofectamine 2000 (Invitrogen). Twenty-four hours post-transfection, cells were fixed with 4% paraformaldehyde–PBS for 0.5–1 h. Cells were treated for 5 min with 2% acetone–PBS pre-chilled at ~30°C for permeabilization, followed by 10 min of pre-hybridization using 40% formamid–2 × SSC solution. When necessary, DNase or RNase was treated before pre-hybridization. Hybridization mixture...
contains 40% formamide, 2× SSC, vanadyl-nucleoside complex, BSA, yeast tRNA and 1 mg/ml (CTG)8 or (CAG)8 probe conjugated with Alexa-647 or -546. We also used Cy3-conjugated 2-O-methyl-modified RNA (CUG)8 or (CAG)8 probes at the same concentration. Hybridization was performed at 37°C for 2–3 h. After hybridization, cells were washed with 40% formamide–2× SSC and PBS. Cells were treated with Hoechst33342 (300 mg/ml) in PBS for 40 min to stain the nuclei followed by PBS wash three times. VECTASHIELD Mounting Medium (Vector Laboratories, Inc.) was used for mounting. Fluorescent images were obtained by sequential scan using a Leica TCS SP2 MP confocal microscope. 40× and 63× objectives were used. Obtained images were processed using Adobe Photoshop 5.5.

Subcellular localization analysis

Cells were plated on a 24-well or 96-well plate. Typically at 24 h after transfection, cells were fixed by 4% PFA–PBS containing Hoechst33342 for 0.5–1 h at room temperature followed by washing with PBS three times. In the case of repeat RNA localization analysis, FISH was performed using oligonucleotides labeled with Alexa-647, Alexa-546 or Cy3 (Supplementary Material, Table S1). Quantitative localization analysis was performed using ArrayScanVTI High Content Screening Reader (Cellomics) as previously described (35). An NCI was defined as (cytoplasmic intensity per pixel)/(cytoplasmic intensity per pixel + nuclear intensity per pixel). In one experiment, at least 150 cells were analyzed for each construct and the average NCI value of the cell population was calculated.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS

We thank Dr M. MacDonald and CHDI for StHdh cells and Ms Tomoko Yoda and Ms Itsuko Yamamoto for animal maintenance and the staff members of Research Resource Center at RIKEN Brain Science Institute for DNA sequencing and cell sorting. Y.K. was supported by the RIKEN Special Postdoctoral Researchers Program.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the Grant-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan to N.N. (22110004, 22240037, 24659436 and 25253066) and to Y.K. (19790620, 21790854, 23791007 and 25461299) and by a Grant-in-Aid for the Research on Measures for Ataxic Diseases from the Ministry of Health, Welfare and Labor to N.N.

REFERENCES


CAG repeat transcript at JPH3 locus mediates expanded polyglutamine protein toxicity in Huntington’s disease-like 2 mice. Neuron, 70, 427–440.


