of the *Drosophila* muscleblind protein, which is involved in the terminal differentiation of photoreceptor and muscle cells in the fly[13]. All three MBNL proteins can colocalize with RNA inclusions of expanded CUG/CCUG repeats in both DM1 and DM2 cells[14]. MBNL1 binds directly to both CUG and CCUG repeat RNA in a length-dependent manner *in vitro*[15]. Therefore, these proteins are considered to be sequestered by the expanded RNA through direct interactions, and their cellular functions can be disrupted in both types of DM. It is important to note that cellular studies have demonstrated that MBNL proteins can directly regulate the alternative splicing of the *cTNT* and *IR* genes, which are misregulated in DM1 patients[16,17]. These results strongly support the hypothesis that loss of function of MBNL proteins leads to the misregulation of splicing in DM.

CELF proteins are multi-functional proteins that play regulatory roles in translation, RNA editing, mRNA stability, as well as splicing[18]. CUG-BP regulates the alternative splicing of *cTNT* exon 5, *IR* exon 11, and *CLCN1* intron 2[7,8]. In DM1 patients, the expression of CUG-BP protein is elevated because of protein stabilization induced by PKC-mediated phosphorylation[7,19]. CUG-BP acts antagonistically against MBNL proteins in the splicing regulation of *cTNT* and *IR*[16,17] but their activities are independent, suggesting that altered CELF activities, in addition to the loss of MBNL function, can induce aberrant splicing in DM1 (see **Fig. 1**). However, the extent to which these proteins can account for splicing abnormalities and the pathogenesis of DM remains unclear.

Thus, it is important to characterize the roles of MBNL and CELF proteins in the regulation of *Clcn1* splicing to understand the mechanism of myotonia in DM. Although increased exon 7A inclusion is the most frequent abnormality of *CLCN1*/*Clcn1* splicing in DM[9], the mechanism of its regulation is still unclear.

We established a *Clcn1* minigene assay system and identified multiple *cis*- and *trans*-acting factors that regulate the alternative splicing of *Clcn1* exon 7A. The essential role of MBNL proteins in the normal splicing pattern of *Clcn1* was verified. Our results also highlight some CELF proteins as antagonistic regulators against MBNL proteins.

2. Materials and methods

MBNL1 and MBNL2 were amplified by PCR from a human skeletal muscle cDNA library (BD Marathon-Ready human cDNA ; Clontech). MBNL3 was amplified from a human liver cDNA library. CELF proteins were amplified from cDNA libraries of either brain or skeletal muscle of human origin[20].

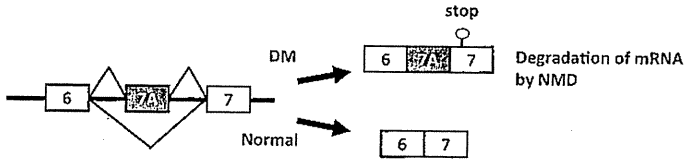

Fig. 2 Splicing regulation of *Clcn1* by MBNL and CELF proteins
Structure of chloride channel minigenes. Mouse *Clcn1* minigene was subcloned between the BglII and SalI sites of pEGFP-C1. Black boxes represent exons of the minigenes. Arrows indicate the position of primers used in the splicing assays.

Cells transfected with plasmids for the expression of a protein and a minigene were harvested 48 h post-transfection. Typically, cells were cultured in 12-well plates and transfected with 0.5 μg plasmids for protein expression (or cognate empty vector) and 0.01 μg plasmids for the expression of a minigene. Total RNA was extracted and purified using either the acidic guanidine phenol chloroform method or RNeasy Mini kit (Qiagen) including DNase treatment. Typically, 1.0 μg total RNA was reverse-transcribed using the ThermoScript RT-PCR System (Invitrogen) or Revertra Ace-α-(Toyobo) with a 1 : 1 mixture of oligo dT and random hexamer as primers. Minigene fragments were amplified by PCR using a fluorescein isothiocyanate (FITC)-labeled forward primer for the 3' region of the EGFP sequence (FITC-GFP-Fw) and a gene-specific reverse primer (Clcn1-Rv for *Clcn1* or CLCN1-Rv for *CLCN1*). PCR products were resolved by 2.0-2.5% agarose gel electrophoresis. By sampling at multiple cycles, the cycle numbers of PCR were adjusted such that the amplification was within the logarithmic phase. The fluorescence of PCR products was captured and visualized by LAS1000 or LAS3000 (Fujifilm). The intensity of band signals was quantified using Multigauge software (Fujifilm). The ratio of exon 7A inclusion in *Clcn1* and *CLCN1* was calculated as (7A inclusion)/(7A inclusion + 7A skipping) × 100.

3. Results and discussion

To examine whether the MBNL and CELF family proteins can regulate the splicing of *Clcn1*, we created a minigene covering exons 6 to 7 of the mouse *Clcn1* gene (**Fig. 2**). It is important to note that because the inclusion of exon 7A does not produce a premature termination codon in the context of our *Clcn1* minigene, the spliced products containing

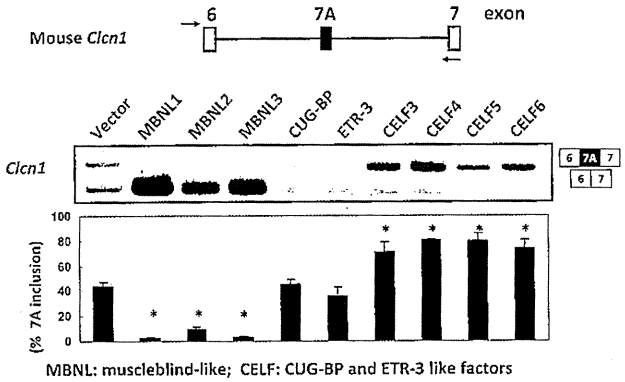

Fig. 3 Splicing regulation of MBNL and CELF proteins.

Representative results of cellular splicing assays using the *Clcn1* minigene in COS-7 cells. The upper bands correspond to a splice product containing exon 7A, whereas lower bands correspond to a splice product lacking exon 7A. Bar chart shows quantified results of exon 7A inclusion (mean ± SD, $n=3$). Statistical significance was analyzed by analysis of variance and Dunnett's multiple comparison. All MBNL proteins and CELF proteins except for CUG-BP and ETR-3 showed significant differences ($^{*}p<0.0001$) compared to the empty vector.

exon 7A are not substrates of nonsense-mediated mRNA decay (NMD). Thus, the minigene would provide more faithful splicing patterns compared to the endogenous *Clcn1*. We utilized non-muscle cell lines to minimize the effect of muscle-dependent backgrounds and focus on the direct effects of transgenes. When the *Clcn1* minigene was transfected into COS-7 cells, 45% of the spliced products contained exon 7A (**Fig. 3**). Next we expressed myc-tagged MBNL or CELF proteins with the *Clcn1* minigene and examined the patterns of *Clcn1* splicing. The expressions of MBNL and CELF proteins were confirmed by Western blotting using an anti-myc antibody (data not shown). All three MBNL proteins strongly repressed exon 7A inclusion (**Fig. 1B**). In contrast, CELF3, CELF4, CELF5, and CELF6 proteins significantly promoted the inclusion of 7A. Remarkably, CUG-BP (CELF1) and ETR-3 (CELF2) did not alter the ratio of exon 7A inclusion. These two proteins increased the unspliced product and reduced the spliced products with or without exon 7A (data not shown).

CLCN1/Clcn1 splicing is a key event in DM. Although the misregulation of splicing has been well established as a characteristic abnormality of DM, few misregulated genes have a clear causal relationship to symptoms of DM. *Clcn1* misregulation can account for myotonia in DM model mice[21]. As demonstrated recently, the skipping of exon 7A induced by antisense oligonucleotide reversed the myotonic phenotype of DM model mice[22], making *CLCN1* splicing a promising target for therapeutic approaches. Understanding *Clcn1/CLCN1* splicing would aid in the design of rational strategies for correcting CLCN1 expression to perhaps prevent myotonia.

Here, we have demonstrated that the splicing regulation of *Clcn1* exon 7A by MBNL1 was observed in COS-7 as well as HeLa, and Neuro2A cell lines (**Fig. 3** and ref. 20). Thus, the regulation of exon 7A can be determined directly by the expression level of MBNL proteins. The inclusion of exon 7A was repressed by the overexpression of MBNL proteins but increased by their knockdown[20]. These results are consistent with the model that MBNL proteins directly regulate *CLCN1/Clcn1* and that the loss of MBNL function leads to *CLCN1/Clcn1* misregulation in DM.

In contrast to MBNL proteins, CELF3/4/5/6 promoted increased inclusion of exon 7A of mouse *Clcn1* (**Fig. 3**). Among these CELF proteins, CELF4 is expressed in a wide variety of tissues, including muscle[18,23]. Although mice deficient in *Celf4* have been reported to manifest a complex seizure phenotype[24], the physiological function of CELF4 is largely unclear. Although an elevation of CUG-BP and ETR-3 proteins was observed in DM1 patients, the other CELF proteins have not been well characterized. The expression level, intracellular localization, and activity of CELF4 (and CELF3/5/6) should be investigated in the context of DM. Although *Clcn1* is enriched in muscle, it is expressed in other tissues (including the brain) even at a low level. Because some CELF proteins are enriched in the brain[23], they might play a role in keeping *Clcn1* expression at a low level in tissues other than muscle through a splicing-mediated regulation of expression.

In order to understand the regulatory mechanism controlling splice site selection, it should be clarified how these RNA-binding proteins activate splicing of one substrate and repress splicing of another. Whether antagonistic regulation by MBNL and CELF is linked or not? We hope future work will help the way to treat DM.

Acknowledgments

This work was supported by research grants from the Ministry of Health, Labor and Welfare ; and the Ministry of Education, Culture, Sports, Science, and Technology of Japan. We would like to thank Dr. Yoshihide Hayashizaki for FANTOM3 clones and Dr. Gen

Matsumoto for the R-miR vector and Dr. Chika Washizu for discussion. We are also grateful to the staff of the Research Resource Center of RIKEN Brain Science Institute for DNA sequencing. Y. K. was supported by a JSPS Research Fellowship for Young Scientists and the RIKEN Special Postdoctoral Researchers Program.

References

1) Harper, P. S. *Myotonic Dystrophy*, 3rd edn. WB Saunders, London, 2001.
2) Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K, Stanton VP, Thirion JP, Hudson T, et al. 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. 1992 ; 68 : 799-808.
3) Fu YH, Friedman DL, Richards S, Pearlman JA, Gibbs RA, Pizzuti A, Ashizawa T, Perryman MB, Scarlato G, Fenwick RG, et al. Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy. Science. 1993 ; 260 : 235-238.
4) Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K, et al. Myotonic dystrophy mutation : an unstable CTG repeat in the 3' untranslated region of the gene. Science. 1992 ; 255 : 1253-1255.
5) Mankodi A, Logigian E, Callahan L, McClain C, White R, Henderson D, Krym M, Thornton CA. Myotonic dystrophy in transgenic mice expressing an expanded CUG repeat. Science. 2000 ; 289 : 1769-1773.
6) Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP. Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science. 2001 ; 293 : 864-867.
7) Savkur RS, Philips AV, Cooper TA. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. Nat Genet. 2001 ; 29 : 40-47.
8) Charlet-B N, Savkur RS, Singh G, Philips AV, Grice EA, Cooper TA. Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. Mol Cell. 2002 ; 10 : 45-53.
9) Mankodi A, Takahashi MP, Jiang H, Beck CL, Bowers WJ, Moxley RT, Cannon SC, Thornton CA. Expanded CUG repeats trigger aberrant splicing of ClC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. Mol Cell. 2002 ; 10 : 35-44.
10) Kanadia RN, Johnstone KA, Mankodi A, Lungu C, Thornton CA, Esson D, Timmers AM, Hauswirth WW, Swanson MS. A muscleblind knockout model for myotonic dystrophy. Science. 2003 ; 302 : 1978-1980.
11) Kimura T, Nakamori M, Lueck JD, Pouliquin P, Aoike F, Fujimura H, Dirksen RT, Takahashi MP, Dulhunty AF, Sakoda S. Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca2+-ATPase in myotonic dystrophy type 1. Hum Mol Genet. 2005 ; 14 : 2189-2200.
12) Ranum LP, Cooper TA. RNA-mediated neuromuscular disorders. Annu Rev Neurosci. 2006 ; 29 : 259-277.
13) Begemann G, Paricio N, Artero R, Kiss I, Perez-Alonso M, Mlodzik M. muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. Development. 1997 ; 124 : 4321-4331.
14) Fardaei M, Rogers MT, Thorpe HM, Larkin K, Hamshere MG, Harper PS, Brook JD. 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. 2002 ; 11 : 805-814.

15) Kino Y, Mori D, Oma Y, Takeshita Y, Sasagawa N, Ishiura S. Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. Hum Mol Genet. 2004 ; 13 : 495-507.
16) Ho TH, Charlet-B N, Poulos MG, Singh G, Swanson MS, Cooper TA. Muscleblind proteins regulate alternative splicing. EMBO J. 2004 ; 23 : 3103-3112.
17) Dansithong W, Paul S, Comai L, Reddy S. MBNL1 is the primary determinant of focus formation and aberrant insulin receptor splicing in DM1. J Biol Chem. 2005 ; 280 : 5773-5780.
18) Ladd AN, Charlet N, Cooper TA. The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol Cell Biol. 2001 ; 21 : 1285-1296.
19) Kuyumcu-Martinez NM, Wang GS, Cooper TA. Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. Mol Cell. 2007 ; 28 : 68-78.
20) Kino Y, Washizu C, Oma Y, Onishi H, Nezu Y, Sasagawa N, Nukina N, Ishiura S. MBNL and CELF proteins regulate alternative splicing of the skeletal muscle Chloride channel *CLCN1*. Nucleic Acids Research. 2009 ; 37 : 6477-6490.
21) Lueck JD, Lungu C, Mankodi A, Osborne RJ, Welle SL, Dirksen RT, Thornton CA. Chloride channelopathy in myotonic dystrophy resulting from loss of posttranscriptional regulation for CLCN1. Am J Physiol Cell Physiol. 2007 ; 292 : C1245-1247.
22) Wheeler TM, Lueck JD, Swanson MS, Dirksen RT, Thornton CA. Correction of ClC-1 splicing eliminates chloride channelopathy and myotonia in mouse models of myotonic dystrophy. J Clin Invest. 2007 ; 117 : 3952-3957.
23) Ladd AN, Nguyen NH, Malhotra K, Cooper TA. CELF6, a member of the CELF family of RNA-binding proteins, regulates muscle-specific splicing enhancer-dependent alternative splicing. J Biol Chem. 2004 ; 279 : 17756-17764.
24) Yang Y, Mahaffey CL, Bérubé N, Maddatu TP, Cox GA, Frankel WN. Complex seizure disorder caused by Brunol4 deficiency in mice. PLoS Genet. 2007 ; 3 : e124.

Handbook of Neurochemistry and Molecular Neurobiology 3rd Edition

Volume No. 25 "Neurochemical Mechanisms in Disease"

"RNA Pathologies in Neurological Disorders"

Kinji Ohno, Akio Masuda

Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan

E-mail: ohnok@med.nagoya-u.ac.jp

Abstract

RNA is not a simple intermediate linking DNA and protein. RNA is widely transcribed from a variety of genomic regions, and extensive studies on the functional roles and regulations of non-coding RNAs including antisense RNAs and small RNAs are in progress. In addition, the human genome project revealed that we humans carry as few as ~22,000 genes. Humans exploit tissue-specific and developmental stage-specific alternative splicing to generate a large variety of molecules in specific cells at specific developmental stages. Neurological disorders are also subject to aberrations of the splicing mechanisms. This review focuses mostly on splicing abnormalities due to pathological alterations of splicing *cis*-elements and *trans*-factors. Pathomechanisms associated with disrupted splicing *cis*-elements can be applied to any human diseases, and we did not restrict the descriptions to neurological diseases. On the other hand, we limited the descriptions of dysregulated splicing *trans*-factors to neurological disorders. Neurological diseases covered in this review include congenital myasthenic syndromes, spinal muscular atrophy, myotonic dystrophy, Alzheimer's disease, frontotemporal dementia with parkinsonism linked to

chromosome 17, facioscapulohumeral muscular dystrophy, fragile X-associated tremor/ataxia syndrome, Prader-Willi syndrome, Rett syndrome, spinocerebellar atrophy type 8, and paraneoplastic neurological disorders.

Key words: The RNA world, pre-mRNA splicing, splicing *cis*-elements, splicing *trans*-factors, branch point sequence (BPS), exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE), intronic splicing silencer (ISS), nonsense-mediated mRNA decay (NMD), nonsense-associated skipping of a remote exon (NASRE), congenital myasthenic syndromes, spinal muscular atrophy (SMA), myotonic dystrophy (DM1, DM2), Alzheimer's disease, frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17), facioscapulohumeral muscular dystrophy (FSHD), fragile X-associated tremor/ataxia syndrome (FXTAS), Prader-Willi syndrome, Rett syndrome, spinocerebellar atrophy type 8 (SCA8), and paraneoplastic neurological disorders (PND).

1. Introduction

The central dogma first enunciated by Francis Crick depicts that RNA is an intermediate that links DNA and protein (Crick, 1970). The beginning of life, however, was the RNA world where there was no DNA or proteins (Gilbert, 1986). In the RNA world, RNA was the only carrier of genetic information that DNA currently serves, and the only functional molecule that proteins currently serve. Although the RNA transmits no genetic information to progeny and constitutes a limited number of functional molecules in our human body, the RNA world is still in effect in our body. Humans transcribe more than half of our entire genome including noncoding regions. The transcripts work as antisense RNAs, microRNAs, and snoRNAs. Researchers are now working to disclose the functional significance of these noncoding RNAs.

The human genome project and the subsequent annotation efforts revealed that we humans carry as few as 22,000 genes. Tissue-specific and developmental stage-specific splicing enables to us to generate more than 100,000 molecules from a limited number of genes (Black, 2003; Licatalosi and Darnell, 2006). Small RNA molecules and RNA splicing mechanisms potentially become targets of neurological diseases (Ranum and Cooper, 2006). This review focuses mostly on splicing aberrations associated with neurological disorders.

2. Physiology of splicing mechanisms

In higher eukaryotes, pre-mRNA splicing is mediated by degenerative splicing *cis*-elements comprised of the branch point sequence (BPS), the polypyrimidine tract (PPT), the 5' and 3' splice sites, and exonic/intronic splicing enhancers/ silencers **(Fig. 1)**. Stepwise assembly of the spliceosome starts from recruitment of U1 snRNP to the 5' splice site, SF1 to the BPS, U2AF65 to the PPT, and U2AF35 to the 3' end of an

intron to form a spliceosome complex E (Sperling et al., 2008). SF1, a 75-kDa protein, is a mammalian homolog of yeast BBP (branch point-binding protein). U2AF65 and U2AF35 bring U2 snRNP to the BPS in place of SF1 (Wu et al., 1999; Zorio and Blumenthal, 1999). The BPS establishes base pairing interactions with a stretch of 'GUAGUA' of U2 snRNA (Arning et al., 1996; Abovich and Rosbash, 1997), which then bulges out the branch site nucleotide, usually an adenosine to form a spliceosome complex A (Query et al., 1994). Thereafter, pre-mRNAs are spliced in two sequential transesterification reactions mediated by the spliceosome. In the first step, the 2'-OH moiety of the branch site nucleotide carries out a nucleophilic attack against a phosphate at the 5' splice site, generating a free upstream exon, as well as a lariat carrying the intron and the downstream exon. In the second step, the 3'-OH moiety of the upstream exon attacks the 3' splice site of the lariat leading to intron excision and ligation of the upstream and downstream exons (Query et al., 1996).

In addition to the "classical" spliceosomal mechanisms, splicing is modulated by exonic/intronic splicing enhancers/silencers (ESE, ISE, ESS, ISS). The *trans*-factors for the splicing enhancers/silencers carry repeats of arginine and serine, are accordingly called SR proteins. Tissue-specific and developmental stage-specific expressions of the splicing *trans*-factors enable precise spatial and temporal regulations of the gene expressions. In addition, the splicing *trans*-factors also work on constitutively spliced exons to compensate for highly degenerative "classical" splicing *cis*-elements.

3. Disorders associated with disruption of splicing *cis*-elements

3.1 Aberrations of the 5' splice sites

Mutations disrupting the 5' splice sites have been most frequently reported. U1 snRNA recognizes three nucleotides at the end of an exon and six nucleotides at the beginning of an intron **(Fig. 2)**. The completely matched nucleotides to U1 snRNA are CAG|GTAAGT, where the vertical line represent the exon/intron boundary. The completely matched sequence is observed at 1,597 sites out of the entire 189,249 5' splice sites in the human genome (Sahashi et al., 2007), which is the 10th most common sequence. The completely matched 5' splice site is rather avoided because, in the second stage of splicing, U1 snRNA is substituted for with U5 snRNA. If U1 snRNA is tightly bound to the 5' splice site, it hinders binding of U5 snRNA.

Degeneracy of the 5' splice site and its vulnerability to disease-causing mutations have been extensively studied. Three algorithms have been proposed. First, Shapiro and Senapathy collated nucleotide frequencies at each position of the 5' splice site. They assumed that nucleotide frequencies at each position of the 5' splice site represent the splicing signal intensity. They thus constructed a linear regression model so that the most preferred 5' splice site becomes 1.0 and the most unfavorable 5' splice site becomes 0.0 (Shapiro and Senapathy, 1987). Second, Rogan and Schneider invented the information contents, R*i*. For example, at a specific position, if a single nucleotide is exclusively used, the information content at this position becomes $-\log_2(1/4) = 2$ bits. Similarly, if two nucleotides are equally used, the information content becomes $-\log_2(2/4) = 1$ bit. In R*i*, the similarity to the consensus sequence is represented by the sum of information bits (Rogan and Schneider, 1995; O'Neill et al., 1998). Third, we found that a new parameter, the SD-Score, which represents a common logarithm of the frequency of a specific 5' splice site in the human genome, efficiently predicts the splicing signal intensity (Sahashi et al., 2007). Our algorithm predicts the splicing consequences of mutations with the sensitivity of

97.1% and the specificity of 94.7%. Simulation of all the possible mutations in the human genome using the SD-score algorithm predicts high frequencies of splicing mutations from exon -3 to intron +6 **(Table 1)**. Especially at exon position -3, about one third of mutations are predicted to cause aberrant splicing. Using our algorithm, we predicted and proved that *DYSF* G1842D in Miyoshi myopathy, *ABCD1* R545W in adrenoleucodystrophy, *GLA* Q333X in Fabry disease, and *DMD* Q119X and Q1144X in Duchenne muscular dystrophy are not missense or nonsense mutations but are splicing mutations. Algorithms by us and by others all point to the notion that aberrant splicing caused by mutations at the 5' splice sites are likely to be underestimated.

3.2 Human branch point consensus sequence

In an effort to seek for an algorithm to predict the position of the branch point sequence (BPS) in human, we sequenced 367 clones of lariat RT-PCR products arising from 52 introns of 20 human housekeeping genes and identified that the human consensus BPS is simply yUnAy, where 'y' represents U or C (Gao et al., 2008) **(Fig. 3)**. The consensus BPS was more degenerative than we had expected and we failed to construct a dependable algorithm that predicts the position of the BPS. Sixteen disease-causing mutations and a polymorphism, however, have been reported to date that disrupt a BPS and cause aberrant splicing (Gao et al., 2008). Among these, eight mutates U at position -2, whereas nine affects A at position 0, which also supports the notion that U at -2 and A at 0 are essential nucleotides.

3.3 Ectopic AG dinucleotide abrogates the AG-scanning mechanism

The 3' end of an intron and the 5' end of an exon carry a consensus sequence of CAG|G, where the vertical line represents the intron/exon boundary. The AG dinucleotide is scanned from the branch point and the first AG is recognized as the 3' end of the intron (Chen et al., 2000). In a patient with congenital myasthenic syndrome, we identified duplication of a 16-nt segment comprised of 8 intronic and 8 exonic nucleotides at the intron 10/exon 10 boundary of *CHRNE* encoding the acetylcholine receptor epsilon subunit (Ohno et al., 2005). We found that the upstream AG of the duplicated segment is exclusively used for splicing and that one or two mutations in the upstream BPS had no effect whereas complete deletion of the upstream BPS partially activated the downstream AG. Similar exclusive activation of the upstream AG is reported in *HEXB* (Dlott et al., 1990) and *SLC4A1* (Bianchi et al., 1997). Creation of a cryptic AG dinucleotide close to the 3' end of an intron should be carefully scrutinized in mutation analysis.

3.4 Mutations that disrupt ESE and ESS

Gorlov and colleagues predicted that more than 16-20% of missense mutations are splicing mutations that disrupt an ESE (Gorlov et al., 2003). According to our own experience, their estimates are likely to be too high. Most ESE/ESS-disrupting mutations, however, are likely to be underestimated, because the positions and sequences of ESE/ESS are highly degenerative.

Four web services provide valuable information to locate ESE and ESS. First, the ESE Finder (http://rulai.cshl.org/ESE/) calculates the similarity of a given nucleotide sequence to the consensus sequences of four splicing *trans*-factors, SF2/ASF, SC35, SRp40, and SRp55 (Cartegni et al., 2003; Smith et al., 2006). Second, the RESUCE-ESE web server (http://genes.mit.edu/burgelab/rescue-ese/)

shows the similarity of a given sequence to ESE elements of unidentified splicing *trans*-factors (Fairbrother et al., 2002). The same group also provides the FAS-ESS web service to screen for ESS elements (http://genes.mit.edu/fas-ess/) (Wang et al., 2004). Third, the PESX web server (http://cubweb.biology.columbia.edu/pesx/) indicates an RNA octamer with putative exonic splicing enhancing or silencing activities (Zhang and Chasin, 2004; Zhang et al., 2005). Fourth, the ESRsearch web server (http://ast.bioinfo.tau.ac.il/) shows 285 candidate ESE/ESS sequences (Goren et al., 2006), as well as ESE/ESS elements indicated by the RESCUE-ESE, FAS-ESS, and PESX services.

In patients with congenital myasthenic syndromes, we identified that *CHRNE* E154X and EF157V (Ohno et al., 2003), as well as *COLQ* E415G (Kimbell et al., 2004), disrupt an ESE and cause aberrant splicing. The ESE/ESS servers above indicate disruption of candidate splicing *cis*-elements for all the three mutations, but we frequently obtain false positives and we cannot simply rely on the servers. Analysis of patient mRNA or analysis using a minigene is generally expected.

3.5 Mutations that disrupt ISE and ISS

Identification of mutations disrupting intronic splicing *cis*-elements is more challenging than that of exonic mutations, because introns are longer than exons and splicing mutations can be anywhere in the introns, and because we do not have a dependable algorithm to predict ISE/ISS. The ESRsearch web server described above is able to indicate consensus sequences recognized by a variety of splicing *trans*-factors including intronic ones.

In a patient with congenital myasthenic syndrome, we identified that *CHRNA1* IVS3-8G>A attenuates binding of hnRNP H ~100-fold and causes exclusive inclusion

of the downstream exon P3A (Masuda et al., 2008) **(Fig. 4)**. We also identified that polypyrimidine tract binding protein (PTB) silences recognition of exon P3A and tannic acid facilitates the expression of PTB by activating its promoter region (submitted).

3.6 Spinal muscular atrophy (SMA)

SMA is an autosomal recessive disorder characterized by degeneration of the anterior horn cells of the spinal cord, which causes muscular weakness and atrophy. SMA is caused by loss-of-function mutations including deletion of the *SMN1* gene that encodes the survival of motor neuron 1. Humans carry almost identical *SMN1* and *SMN2* genes both on chromosome 5q13. *SMN2* carries a C-to-T transition at position 6 of exon 7 compared to *SMN1*, which results in loss of an SF2/ASF-dependent ESE activity (Cartegni et al., 2006). In addition, *SMN2* carries an A-to-G transition at position +100 of intron 7, which creates a high-affinity hnRNP A1-binding site and promotes skipping of exon 7 (Kashima et al., 2007). Skipping of exon 7 in *SMN2* can be ameliorated by therapeutic doses of valproic acid (Brichta et al., 2003; Brichta et al., 2006) and of salbutamol (Angelozzi et al., 2008).

4. Skipping of multiple exons caused by a single splicing mutation

4.1 Skipping of multiple contiguous exons

A mutation disrupting a splicing *cis*-element generally affects splicing of a single exon or intron, but sometimes generates aberrant transcripts affecting multiple neighboring exons. Skipping of multiple contiguous exons is accounted for by

ordered removal of introns and consequent clustering of neighboring exons (Schwarze et al., 1999; Takahara et al., 2002).

4.2 Nonsense-associated skipping of a remote exon (NASRE)

A single mutation infrequently causes skipping of a remote exon. In a patient with congenital myasthenic syndrome, we found that a 7-nt deletion in exon 7 of *CHRNE* causes complete skipping of the preceding exon 6. *CHRNE* exon 6 is composed of 101 nucleotides. It carries weak splicing signals and is partially skipped even in normal subjects. The exon 6-skipped transcript, however, is removed by the nonsense-mediated mRNA decay (NMD) mechanism. The 7-nt deletion in exon 7 restores the open reading frame of the exon 6-skipped transcript and renders it immune to NMD. On the other hand, the normally spliced transcript carries a premature stop codon (PTC) after the 7-nt deletion, and is degraded by NMD **(Fig. 5)**. We dubbed this mechanism NASRE, and found that it is in effect in *SLC25A20* (Hsu et al., 2001), *DBT* (Fisher et al., 1993), *BTK* (Haire et al., 1997), and *MLH1* (Clarke et al., 2000).

5. Disorders associated with dysregulation of splicing *trans*-factors

5.1 Myotonic dystrophy

Myotonic dystrophy is an autosomal dominant multisystem disorder affecting skeletal muscles, eye, heart, endocrine system, and central nervous system. The clinical symptoms include variable degrees of muscle weakness and wasting, myotonia, cataract, insulin resistance, hypogonadism, cardiac conduction defects, frontal balding, and intellectual disabilities (Harper and Monckton, 2004). Myotonic

dystrophy is caused by abnormally expanded CTG repeats in the 3' untranslated region of the *DMPK* gene encoding the dystrophia myotonica protein kinase on chromosome 19q13 (myotonic dystrophy type 1, DM1) (Brook et al., 1992) or by abnormally expanded CCTG repeats in intron 1 of the *ZNF9* gene encoding the zinc finger protein 9 on chromosome 3q21 (myotonic dystrophy type 2, DM2) (Liquori et al., 2001). In DM1, normal individuals have 5 to 30 repeats, mildly affected patients have 50 to 80 repeats, and severely affected individuals have 2,000 or more copies of CTG (Gharehbaghi-Schnell et al., 1998). In DM2, the size of expanded repeats is extremely variable, ranging from 75 to 11,000 repeats, with a mean of 5,000 CCTG repeats (Liquori et al., 2001).

In both DM1 and DM2, expanded CTG or CCTG repeats in the noncoding regions sequestrate a splicing *trans*-factor muscleblind encoded by *MBNL1* to intranuclear RNA foci harboring the mutant RNA, and somehow upregulate another splicing *trans*-factor CUG-binding protein encoded by *CUGBP1* (Ranum and Cooper, 2006) **(Fig. 6)**. Dysregulation of the two splicing *trans*-factors then causes aberrant splicing of their target genes. The aberrantly spliced genes identified to date in skeletal and cardiac muscles include *ATP2A1* (*SERCA1*) exon 22, *ATP2A2* (*SERCA2*) intron 19, *CAPN3* exon 16, *CLCN1* intron 2 and exons 6b/7a, *DMD* exons 71 and 78, *DTNA* exons 11A and 12, *FHOD1* (*FHOS*) exon 11a, *GFPT1* (*GFAT1*) exon 10, *INSR* exon 11, *KCNAB1* exons 2b/2c, *LDB3* (*ZASP*) exon 11 (189-nt exon 7 according to RefSeq Build 36.3), *MBNL1* exon 7 (54-nt exon 6 according to RefSeq), *MBNL2* exon 7 (54 nt, no exonic annotation in RefSeq), *MTMR1* exons 2.1 and 2.2, *NRAP* exon 12, *PDLIM3* (*ALP*) exons 5a/5b, *RYR1* exon 70, *TNNT2* exon 5, *TNNT3* fetal exon, *TTN* exons Zr4 and Zr5 (138-nt exon 11 and 138-nt exon 12 according to RefSeq), and *TTN* exon Mex5 (303-nt exon 315 according to RefSeq) (Philips et al.,

1998; Savkur et al., 2001; Kimura et al., 2005; Lin et al., 2006). Lin and colleagues report that alternative transcripts observed in myotonic dystrophy are all fetal isoforms (Lin et al., 2006). Muscleblind normally translocates from cytoplasm to nucleus in the postnatal period to induce adult-type splicings, and lack of muscleblind in nucleus due to sequestration to RNA foci recapitulates fetal splicing patterns.

5.2 Alzheimer's disease (AD) and frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17)

AD is the most common neurodegenerative disease representing dementia. It is characterized by intracellular neurofibrillary tangles (NFTs) and extracellular amyloid plaques. NFTs are composed of aggregates of the hyperphosphorylated tau protein encoded by *MAPT*. The amyloid plaques are composed of amyloid β peptide (Aβ) that originates from enzymatic cleavage of the amyloid precursor protein (*APP*) by β-secretase followed by γ-secretase (LaFerla et al., 2007). The γ-secretase is an enzyme complex composed of presenilin-1 (*PS1*) or presenilin-2 (*PS2*), as well as nicastrin, anterior pharynx defective (APH-1), and presenilin enhancer 2 (PEN-2) (Takasugi et al., 2003). Autosomal dominant forms of AD constitutes ~5% of AD and are caused by mutations in *APP*, *PS1*, or *PS2* (Bertram and Tanzi, 2008).

Although the pathomechanisms underlying sporadic AD remain mostly unknown, *PS2* exon 5 is exclusively skipped in brains of sporadic AD, which is mediated by overexpression of a splicing *trans*-factor, HMGA1a (Sato et al., 1999; Manabe et al., 2003). As hypoxia induces the overexpression of HMGA1a, the upregulation of HMGA1a in sporadic AD may or may not represent an agonal state of AD, in which respiratory insufficiency possibly associated with pneumonia frequently becomes the cause of death.

Mutations in *MAPT* are not observed in AD, but are present in FTDP-17. *MAPT* exon 10 is alternatively spliced in normal brain. N279K, K280del, and L284L mutations on exon 10 provoke aberrant splicing of exon 10 by disrupting or enhancing exonic splicing *cis*-elements, and cause FTDP-17 (D'Souza et al., 1999) **(Fig. 7)**. The splicing *trans*-factors for these *cis*-elements are also identified (Jiang et al., 2004; Kondo et al., 2004).

5.3 Facioscapulohumeral muscular dystrophy (FSHD)

FSHD is the third most common hereditary muscular dystrophy after Duchenne muscular dystrophy and myotonic dystrophy. As its name represents, the disease predominantly affects the face, the scapulae, and the proximal arm muscles. In FSHD, the number of a 3.3-kb repeat in the subtelomeric region of 4q (4q35), designated *D4Z4*, are abnormally reduced (Wijmenga et al., 1992). Loss of *D4Z4* causes upregulation of *FRG1* located upstream of *D4Z4* (Gabellini et al., 2002). FRG1 is a splicing *trans*-factor, and its overexpression causes aberrant splicing of *TNNT3* encoding the troponin T type 3 of fast skeletal muscle and *MTMR1* encoding the myotubularin-related protein 1 (Gabellini et al., 2006). The reported splicing aberrations in FSHD, however, have not been confirmed by us (unpublished data) or by the other groups (personal communications).

5.4 Fragile X-associated tremor/ataxia syndrome (FXTAS)

Fragile X mental retardation syndrome is caused by abnormal expansion of a CGG repeat in the 5' untranslated region of *FMR1*, which culminates in hypermethylation of *FMR1* and silences its expression (Kremer et al., 1991). On the other hand, moderate expansion of the CGG repeat in *FMR1* causes FXTAS, which is

characterized by intention tremor, parkinsonism, cognitive decline, and neuropathy (Hagerman and Hagerman, 2004). In FXTAS, CGG-binding proteins including hnRNP A2 and muscleblind are excessively bound to the expanded CGG repeats of *FMR1* and are depleted from the cellular pool (Iwahashi et al., 2006), which results in the loss their functions in other regulatory processes (Jacquemont et al., 2007).

5.5 Prader-Willi syndrome (PWS)

PWS is an autosomal dominant disorder characterized by obesity, muscular hypotonia and weakness, mental retardation, short stature, hypogonadotropic hypogonadism, and small distal extremities. The proximal long arm of chromosome 15 (15q11-q13) is normally imprinted in order to achieve parent-specific monoallelic gene expressions. Some genes in this region are expressed only from the maternal allele, and some others are only from the paternal allele. Lack of a functional paternal copy of 15q11-13 causes PWS, whereas lack of a functional maternal copy of *UBE3A* in the same region results in Angelman syndrome (Horsthemke and Wagstaff, 2008). PWS is caused by a deletion of the paternal 15q11-q13 or by maternal uniparental disomy 15.

A snoRNA HBII-52 is located in the defective region of PWS. HBII-52 binds to an ESS in exon Vb of *HTR2C* encoding the serotonin receptor 2C, and its disruption in PWS causes aberrant splicing of *HTR2C* and potentially accounts for dysfunctional serotonergic system in PWS (Kishore and Stamm, 2006).

5.6 Rett syndrome

Rett syndrome is a neurodevelopmental disorder in females, which is characterized by loss of speech, stereotypical movements of hands, microcephaly,

seizures, and mental retardation. Rett syndrome is caused by a mutation in *MECP2* encoding the metyl-CpG-binding protein 2 (Amir et al., 1999). MeCP2 binds to a splicing *trans*-factor YB-1 and the abnormal regulation of YB-1 causes aberrant splicing of its target genes (Young et al., 2005).

5.7 Spinocerebellar ataxia type 8 (SCA8)

SCA8 is caused by an abnormal expansion of CTA/CTG repeats in the protein-noncoding *ATXN8OS*, which represents the *ATXN8* opposite strand (Ikeda et al., 2008). Expanded CUG repeats on the *ATXN8OS* transcript potentially bind to and sequestrate CUG-binding proteins, as we observe in myotonic dystrophy (Mutsuddi and Rebay, 2005). In addition, *ATXN8* on the opposite strand of *ATXN8OS* encodes the Kelch-like 1, and the expanded CAG repeats on *ATXN8* give rise to a polyglutamine tract that forms a cytotoxic aggregate in neuronal cells (Moseley et al., 2006). Furthermore, expression of *ATXN8OS* is colocalized with that of *ATXN8* (Chen et al., 2008). *ATXN8OS* thus potentially serves as an antisense RNA for *ATXN8*, and the abnormal CTA/CTG expansion in *ATXN8OS* may dysregulate the expression of *ATXN8* **(Fig. 8)**.

5.8 Paraneoplastic neurological disorders (PND)

In PND, tumors outside of the nervous system excrete humoral factors such as hormones and cytokines, or provoke an immune response against specific molecules expressed in tumors, and cause a wide range of neurological symptoms. In paraneoplastic opsoclonus myoclonus ataxia (POMA), autoantibodies are raised against the Nova family of neuron-specific splicing *trans*-factor (Jensen et al., 2000; Ule et al., 2003; Ule et al., 2006; Licatalosi et al., 2008). In paraneoplastic

encephalomyelitis and sensory neuropathy (PEN/SN or Hu syndrome), autoantibodies recognize the Hu family of RNA-binding protein (Szabo et al., 1991), a human homolog of the Drosophila splicing *trans*-factor Elav (Koushika et al., 2000; Soller and White, 2003). In both disorders, autoantibodies downregulate the splicing *trans*-factors and cause aberrant splicing in neuronal cells.

Acknowledgements

Works from the authors' laboratories have been supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and from the Ministry of Health, Labor, and Welfare of Japan.

References

Abovich N, Rosbash M. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. Cell 89: 403-412

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, et al. 1999. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 23: 185-188

Angelozzi C, Borgo F, Tiziano FD, Martella A, Neri G, et al. 2008. Salbutamol increases SMN mRNA and protein levels in spinal muscular atrophy cells. J Med Genet 45: 29-31

Arning S, Gruter P, Bilbe G, Kramer A. 1996. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. RNA 2: 794-810

Bertram L, Tanzi RE. 2008. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. Nat Rev Neurosci 9: 768-778

Bianchi P, Zanella A, Alloisio N, Barosi G, Bredi E, et al. 1997. A variant of the EPB3 gene of the anti-Lepore type in hereditary spherocytosis. Br J Haematol 98: 283-288

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72: 291-336

Brichta L, Hofmann Y, Hahnen E, Siebzehnrubl FA, Raschke H, et al. 2003. Valproic acid increases the SMN2 protein level: a well-known drug as a potential therapy for spinal muscular atrophy. Hum Mol Genet 12: 2481-2489

Brichta L, Holker I, Haug K, Klockgether T, Wirth B. 2006. In vivo activation of SMN in spinal muscular atrophy carriers and patients treated with valproate. Ann Neurol 59: 970-975