Fu, Y., et al. (2011). "AG-dependent 3'-splice sites are predisposed to aberrant splicing due to a mutation at the first nucleotide of an exon." *Nucleic acids research* 39,4396-4404.

Gao, K., et al. (2008). "Human branch point consensus sequence is yUnAy." *Nucleic acids research* 36,2257-2267.

Goren, A., et al. (2006). "Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers." *Molecular cell* 22,769-781.

Gorlov, I. P., et al. (2003). "Missense mutations in hMLH1 and hMSH2 are associated with exonic splicing enhancers." *American journal of human genetics* 73,1157-1161.

Gurvich, O. L., et al. (2008). "DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy." *Annals of neurology* 63,81-89.

Hammond, S. M.&Wood, M. J. (2011). "Genetic therapies for RNA mis-splicing diseases." *Trends in genetics : TIG* 27,196-205.

Hentze, M. W.&Kulozik, A. E. (1999). "A perfect message: RNA surveillance and nonsense-mediated decay." *Cell* 96,307-310.

Imam, J. S., et al. (2010). "Frame-disrupting mutations elicit pre-mRNA accumulation independently of frame disruption." *Nucleic acids research* 38,1559-1574.

Kapustin, Y., et al. (2011). "Cryptic splice sites and split genes." *Nucleic acids research* 39,5837-5844.

Karni, R., et al. (2007). "The gene encoding the splicing factor SF2/ASF is a proto-oncogene." *Nature structural & molecular biology* 14,185-193.

Keren, H., et al. (2010). "Alternative splicing and evolution: diversification, exon definition and function." *Nature reviews. Genetics* 11,345-355.

Kimbell, L. M., et al. (2004). "C-terminal and heparin-binding domains of collagenic tail subunit are both essential for anchoring acetylcholinesterase at the synapse." *The Journal of biological chemistry* 279,10997-11005.

Kolasinska-Zwierz, P., et al. (2009). "Differential chromatin marking of introns and expressed exons by H3K36me3." *Nat Genet* 41,376-381.

Krawczak, M., et al. (2007). "Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing." *Hum Mutat* 28,150-158.

Lander, E. S., et al. (2001). "Initial sequencing and analysis of the human genome." *Nature* 409,860-921.

Langford, C. J.&Gallwitz, D. (1983). "Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts." *Cell* 33,519-527.

Lerner, M. R., et al. (1980). "Are snRNPs involved in splicing?" *Nature* 283,220-224.

Licatalosi, D. D.&Darnell, R. B. (2006). "Splicing regulation in neurologic disease." *Neuron* 52,93-101.

Lim, K. H., et al. (2011). "Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes." *Proceedings of the National Academy of Sciences of the United States of America* 108,11093-11098.

Liu, H. X., et al. (2001). "A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes." *Nat Genet* 27,55-58.

Lytle, J. R.&Steitz, J. A. (2004). "Premature termination codons do not affect the rate of splicing of neighboring introns." *RNA* 10,657-668.

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

Manabe, T., et al. (2003). "Induced HMGA1a expression causes aberrant splicing of Presenilin-2 pre-mRNA in sporadic Alzheimer's disease." *Cell death and differentiation* 10,698-708.

Masuda, A., et al. (2008). "hnRNP H enhances skipping of a nonfunctional exon P3A in CHRNA1 and a mutation disrupting its binding causes congenital myasthenic syndrome." *Human molecular genetics* 17,4022-4035.

Mendell, J. T., et al. (2002). "Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts." *Science* 298,419-422.

Metherall, J. E., et al. (1986). "Beta zero thalassemia caused by a base substitution that creates an alternative splice acceptor site in an intron." *EMBO J* 5,2551-2557.

Muhlemann, O., et al. (2001). "Precursor RNAs harboring nonsense codons accumulate near the site of transcription." *Molecular cell* 8,33-43.

Nalla, V. K.&Rogan, P. K. (2005). "Automated splicing mutation analysis by information theory." *Hum Mutat* 25,334-342.

O'Rourke, J. R.&Swanson, M. S. (2009). "Mechanisms of RNA-mediated disease." *The Journal of biological chemistry* 284,7419-7423.

Ohe, K.&Mayeda, A. (2010). "HMGA1a trapping of U1 snRNP at an authentic 5' splice site induces aberrant exon skipping in sporadic Alzheimer's disease." *Molecular and cellular biology* 30,2220-2228.

Ohno K, M. A. (2011). "RNA pathologies in neurological disorders." *Neurochemical Mechanisms in Disease. Ed. by Blass JP. Advances in Neurobiology 1. Series Ed. by Lajtha A. Springer, New York*399-415.

Ohno, K., et al. (2003). "A frameshifting mutation in CHRNE unmasks skipping of the preceding exon." *Human molecular genetics* 12,3055-3066.

Ohno, K., et al. (2005). "Spectrum of splicing errors caused by CHRNE mutations affecting introns and intron/exon boundaries." *Journal of medical genetics* 42,e53.

Pastuszak, A. W., et al. (2011). "An SF1 affinity model to identify branch point sequences in human introns." *Nucleic acids research* 39,2344-2356.

Piva, F., et al. (2009). "SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans." *Bioinformatics* 25,1211-1213.

Pomponio, R. J., et al. (1997). "Profound biotinidase deficiency caused by a point mutation that creates a downstream cryptic 3' splice acceptor site within an exon of the human biotinidase gene." *Human molecular genetics* 6,739-745.

Pros, E., et al. (2009). "Antisense therapeutics for neurofibromatosis type 1 caused by deep intronic mutations." *Hum Mutat* 30,454-462.

Roca, X., et al. (2012). "Widespread recognition of 5' splice sites by noncanonical base-pairing to U1 snRNA involving bulged nucleotides." *Genes & development* 26,1098-1109.

Roca, X.&Krainer, A. R. (2009). "Recognition of atypical 5' splice sites by shifted base-pairing to U1 snRNA." *Nature structural & molecular biology* 16,176-182.

Roca, X., et al. (2005). "Determinants of the inherent strength of human 5' splice sites." *RNA* 11,683-698.

Rogan, P. K., et al. (1998). "Information analysis of human splice site mutations." *Hum Mutat* 12,153-171.

Sahashi, K., et al. (2007). "In vitro and in silico analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5' splice sites." *Nucleic acids research* 35,5995-6003.

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

Salzberg, S. L. (1997). "A method for identifying splice sites and translational start sites in eukaryotic mRNA." *Comput Appl Biosci* 13,365-376.

Sasaki-Haraguchi, N., et al. (2012). "Mechanistic insights into human pre-mRNA splicing of human ultra-short introns: Potential unusual mechanism identifies G-rich introns." *Biochem Biophys Res Commun.*

Schmitz, J.&Brosius, J. (2011). "Exonization of transposed elements: A challenge and opportunity for evolution." *Biochimie* 93,1928-1934.

Schwartz, S., et al. (2009). "Chromatin organization marks exon-intron structure." *Nature structural & molecular biology* 16,990-995.

Shapiro, M. B.&Senapathy, P. (1987). "RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression." *Nucleic acids research* 15,7155-7174.

Sheth, N., et al. (2006). "Comprehensive splice-site analysis using comparative genomics." *Nucleic acids research* 34,3955-3967.

Smith, C. W., et al. (1989). "Scanning from an independently specified branch point defines the 3' splice site of mammalian introns." *Nature* 342,243-247.

Smith, D. J., et al. (2008). ""Nought may endure but mutability": spliceosome dynamics and the regulation of splicing." *Molecular cell* 30,657-666.

Soares, L. M., et al. (2006). "Intron removal requires proofreading of U2AF/3' splice site recognition by DEK." *Science* 312,1961-1965.

Spritz, R. A., et al. (1981). "Base substitution in an intervening sequence of a beta+-thalassemic human globin gene." *Proceedings of the National Academy of Sciences of the United States of America* 78,2455-2459.

Staden, R. (1984). "Computer methods to locate signals in nucleic acid sequences." *Nucleic acids research* 12,505-519.

Sterne-Weiler, T., et al. (2011). "Loss of exon identity is a common mechanism of human inherited disease." *Genome research* 21,1563-1571.

Taggart, A. J., et al. (2012). "Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo." *Nature structural & molecular biology.*

Tanackovic, G.&Kramer, A. (2005). "Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing in vivo." *Mol Biol Cell* 16,1366-1377.

Tavanez, J. P., et al. (2012). "hnRNP A1 proofreads 3' splice site recognition by U2AF." *Molecular cell* 45,314-329.

Tilgner, H., et al. (2009). "Nucleosome positioning as a determinant of exon recognition." *Nature structural & molecular biology* 16,996-1001.

Tsai, K. N.&Wang, D. (2012). "Identification of activated cryptic 5' splice sites using structure profiles and odds measure." *Nucleic acids research* 40,e73.

Vogel, J., et al. (1997). "Precise branch point mapping and quantification of splicing intermediates." *Nucleic acids research* 25,2030-2031.

Vorechovsky, I. (2010). "Transposable elements in disease-associated cryptic exons." *Human genetics* 127,135-154.

Wachtel, C., et al. (2004). "Stop codon-mediated suppression of splicing is a novel nuclear scanning mechanism not affected by elements of protein synthesis and NMD." *RNA* 10,1740-1750.

Wahl, M. C., et al. (2009). "The spliceosome: design principles of a dynamic RNP machine." *Cell* 136,701-718.

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

Wang, J., et al. (2002). "Nonsense-associated altered splicing: a frame-dependent response distinct from nonsense-mediated decay." *Molecular cell* 10,951-957.

Wang, Z., et al. (2004). "Systematic identification and analysis of exonic splicing silencers." *Cell* 119,831-845.

Westaway, D.&Williamson, R. (1981). "An intron nucleotide sequence variant in a cloned beta +-thalassaemia globin gene." *Nucleic acids research* 9,1777-1788.

Wieringa, B., et al. (1984). "A minimal intron length but no specific internal sequence is required for splicing the large rabbit beta-globin intron." *Cell* 37,915-925.

Wood, M., et al. (2007). "Modulating the expression of disease genes with RNA-based therapy." *PLoS genetics* 3,e109.

Wood, M. J., et al. (2010). "RNA-targeted splice-correction therapy for neuromuscular disease." *Brain : a journal of neurology* 133,957-972.

Wu, S., et al. (1999). "Functional recognition of the 3' splice site AG by the splicing factor U2AF35." *Nature* 402,832-835.

Yeo, G.&Burge, C. B. (2004). "Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals." *J Comput Biol* 11,377-394.

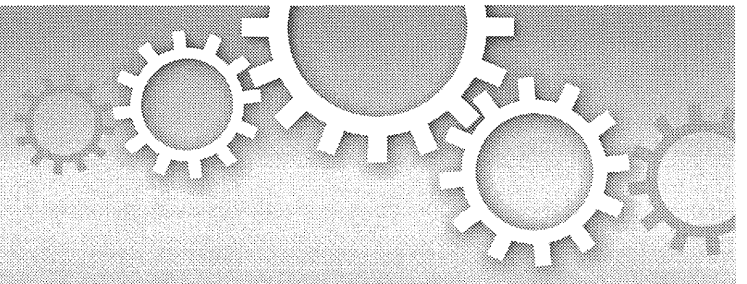
Zhang, M. Q. (1998). "Statistical features of human exons and their flanking regions." *Human molecular genetics* 7,919-932.

Zhang, X. H.&Chasin, L. A. (2004). "Computational definition of sequence motifs governing constitutive exon splicing." *Genes & development* 18,1241-1250.

Zhang, X. H., et al. (2005). "Exon inclusion is dependent on predictable exonic splicing enhancers." *Molecular and cellular biology* 25,7323-7332.

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

SCIENTIFIC REPORTS



OPEN

SUBJECT AREAS:
CELLULAR NEUROSCIENCE
MOTOR SYSTEM
GENE REGULATION
TRANSCRIPTOME

Received
25 August 2011

Accepted
8 December 2011

Published
4 January 2012

Correspondence and requests for materials should be addressed to K.O. (ohnok@med.nagoya-u.ac.jp)

* These authors contributed equally to this work.

CUGBP1 and MBNL1 preferentially bind to 3′ UTRs and facilitate mRNA decay

Akio Masuda[1], Henriette Skovgaard Andersen[2]*, Thomas Koed Doktor[2]*, Takaaki Okamoto[1], Mikako Ito[1], Brage Storstein Andresen[2] & Kinji Ohno[1]

[1]Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan, [2]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark.

CUGBP1 and MBNL1 are developmentally regulated RNA-binding proteins that are causally associated with myotonic dystrophy type 1. We globally determined the *in vivo* RNA-binding sites of CUGBP1 and MBNL1. Interestingly, CUGBP1 and MBNL1 are both preferentially bound to 3′ UTRs. Analysis of CUGBP1- and MBNL1-bound 3′ UTRs demonstrated that both factors mediate accelerated mRNA decay and temporal profiles of expression arrays supported this. Role of CUGBP1 on accelerated mRNA decay has been previously reported, but the similar function of MBNL1 has not been reported to date. It is well established that CUGBP1 and MBNL1 regulate alternative splicing. Screening by exon array and validation by RT-PCR revealed position dependence of CUGBP1- and MBNL1-binding sites on the resulting alternative splicing pattern. This study suggests that regulation of CUGBP1 and MBNL1 is essential for accurate control of destabilization of a broad spectrum of mRNAs as well as of alternative splicing events.

In recent years, new technologies, such as microarray analysis and high throughput sequencing, have dramatically changed our knowledge on gene expression and revealed that extensive regulation takes place during posttranscriptional RNA processing[1]. Numerous RNA processing elements and regulatory RNA-binding proteins play together in a finely tuned interplay to ensure that different mRNAs are made from the primary transcript from a gene and are present in the right cell at the right time and in the correct amounts. Such complex regulation is of course vulnerable and a rapidly increasing number of human diseases are now known to be caused by misregulated RNA processing[2]. An intriguing example where this kind of disease mechanism is in operation is myotonic dystrophy type 1 (DM1), where aberrant regulation of two RNA-binding proteins, CUG-binding protein 1 (CUGBP1) and muscleblind-like 1 (MBNL1) co-operationally cause some of the disease symptoms. DM1 is the most common form of myotonic dystrophy (DM), and is caused by an expansion of CTG-repeats in the 3′ untranslated region (UTR) of the DM protein kinase gene (*DMPK*) on chromosome 19[3–5]. DM1 is a multisystemic disorder and the clinical features include myotonia, muscle degeneration, heart failure, ocular cataracts, impaired glucose tolerance, and mental retardation[6,7]. A dominant negative effect of the *DMPK* mutant allele through RNA gain-of-function has been proposed as the molecular disease mechanism. Many studies support a mechanism where toxic *DMPK* RNA with expanded CUG repeats binds to and sequesters proteins that are important for RNA metabolism including transcription, RNA transport, alternative splicing, translation, and yet unknown processes[6]. The expanded CUG repeats in the *DMPK* mRNA bind to and sequestrate MBNL1 in discrete nuclear foci, which results in depletion of functional MBNL1[8,9]. By a yet unknown mechanism the expanded CUG repeats also activate protein kinase C (PKC), which phosphorylates and stabilizes CUGBP1[10]. Thus, the expanded CUG repeats contribute to DM1 pathogenesis by causing loss of MBNL1 and gain of CUGBP1 activity[11].

Both CUGBP1 and MBNL1 regulate postnatal transitions in alternative splicing patterns during striated muscle development[9,12,13]. Representative targets of CUGBP1 splicing regulation, which are misregulated in DM1 striated muscles, include genes for cardiac troponin T (*TNNT2*)[14,15], insulin receptor (*INSR*)[16], and chloride channel 1 (*CLCN1*)[15,17]. MBNL1 contains four CCCH-type zinc fingers that recognize a YGCY motif that is indeed observed in the CUG repeat (CUGCUG)[18–21]. Mice deficient in *Mbnl1* show aberrant splicing of *Clcn1*, *Tnnt2*, and *Tnnt3*, but not *Insr*[22]. Very recently, MBNL1 was shown to regulate *BIN1* alternative splicing, and dysregulation of *BIN1* splicing in DM1 muscles was suggested to be part of the disease pathology resulting in muscle weakness[23]. Besides an important role in splicing regulation, CUGBP1 mediates mRNA decay of short-lived transcripts by interaction with GU-rich elements in the 3′ UTR[24–27]. In addition, CUGBP1 increases the translation of *CDKN1A*[28] and *Mef2a*[29]. In contrast to the multiple functionalities in posttranscriptional gene regulation of CUGBP1, MBNL1 has so far been exclusively recognized as a splicing regulatory *trans*-factor.

High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation (HITS-CLIP)[30] is a new method that enables global mapping of targets for specific RNA-binding proteins in living cells, thereby shedding light on their role in regulation of RNA processing of known and unknown targets.

In the present study, we performed HITS-CLIP analysis for CUGBP1 and MBNL1 on the mouse myoblast cell line C2C12 to extensively characterize their RNA-binding sites and functional roles in RNA processing. We identified position-dependence of CUGBP1/MBNL1-binding sites in regulating exon inclusion or skipping. Interestingly, we discovered that both CUGBP1 and MBNL1 preferentially bind to the 3′ UTR and destabilize target mRNAs. This points to a new important role of MBNL1 and suggests that binding to the 3′ UTRs and destabilization of mRNA are likely to be a fundamental function shared by CUGBP1 and MBNL1.

Results

Genome-wide CUGBP1/MBNL1-RNA interaction maps. In order to determine global CUGBP1/MBNL1-binding sites *in vivo*, we performed HITS-CLIP experiments using the mouse myoblast cell line, C2C12.

In C2C12 cells, CUGBP1 is constantly expressed throughout myoblast differentiation, whereas expression of MBNL1 is low in undifferentiated cells and gradually increases during differentiation (Supplementary Fig. S1), as previously described[8]. We thus performed HITS-CLIP analysis of CUGBP1 and MBNL1 using undifferentiated and differentiated C2C12 cells, respectively. We also performed CLIP of MBNL1 using undifferentiated cells in three independent experiments, but this yielded an insufficient amount of RNA-protein complexes and failed to yield cDNA libraries suitable for high-throughput sequencing. In the HITS-CLIP analysis of CUGBP1, our first experiment yielded 34,733,815 CLIP tags of 32 nt, of which 29,545,067 (85.06%) were mapped to the mm9 genome allowing at most 2 mismatches and placing reads mapping to multiple locations to a single random site. A second CLIP experiment yielded 10,079,185 CLIP tags of 36 nt, of which 8,516,256 (84.49%) were mapped. In the first MBNL1 CLIP experiment, we obtained 13,218,685 CLIP tags, of which 11,044,152 (83.55%) were mapped, while the second CLIP experiment yielded 13,474,600 CLIP tags with 11,455,886 (85.02%) tags mapped to the mm9 genome. For the analysis of binding motif and binding region annotation, we selected only reads that were aligned uniquely in the genome and removed all potential PCR duplicates by collapsing reads with an identical 5′ start into a single read. This resulted in 177,013 and 130,828 CLIP tags from the two CUGBP1 CLIP experiments, while the two MBNL1 experiments yielded 59,156 and 583,841 CLIP tags respectively.

In an effort to confirm the specificity of our CLIP experiments, we performed CLIP analysis of polypyrimidine tract-binding protein (PTB), a multifunctional RNA-binding protein, using undifferentiated mouse C2C12 cells. We identified 12,841,778 CLIP tags of which 11,184,829 (87.10%) were mapped to the mouse mm9 genome. Removal of non-uniquely aligned reads and PCR duplicates yielded 307,995 unambiguous CLIP tags.

Consensus motifs. To determine RNA-binding motifs associated with CUGBP1/MBNL1 *in vivo*, we used the motif-finding algorithm, Multiple EM for Motif Elicitation (MEME)[31]. We used SeqMonk to identify likely binding regions, and identified 1,841 CUGBP1-binding regions and 302 MBNL1-binding regions. Comparison of SeqMonk's maximum depth scores between samples indicates that binding regions in each replicated experiment are highly overlapping, while PTB binding regions did not overlap with those of the other four CLIP experiments (Supplementary Fig. S2). The lower number of identified MBNL1 regions supported by two independent experiments (Supplementary Fig. S2b) was likely due to the large difference in the number of CLIP tags in the two MBNL1 experiments. The regions demonstrate enrichment of GU-rich motifs for CUGBP1 and YGCY-containing motifs for MBNL1 (Fig. 1).

Our *in vivo* binding motifs are in accordance with previously suggested binding motifs for CUGBP1[25,32] and MBNL1[21,33–35]. We identified 1,824 PTB binding regions in the mouse genome and detected a CU-rich motif, which is essentially identical to the motif for PTB recently identified by HITS-CLIP analysis of a human cell line[36].

We also analyzed the CUGBP1 and MBNL1 motifs enriched in regions containing reads with multiple potential mapping locations (Supplementary Fig. S3), and compared them with the motifs with unique mapping (Fig. 1). Following removal of potential PCR duplicates, we observed 699,382 tags that were non-uniquely aligned in the 1st CUGBP1 CLIP experiment, 219,128 tags in the 2nd CUGBP1 CLIP experiment, 105,432 and 216,882 tags in the two MBNL1 CLIP experiments respectively and finally 851,324 tags in the PTB CLIP experiment. We observed that enriched motifs in these regions (Supplementary Fig. S3) are very similar to the CUGBP1 and MBNL1 motifs enriched in the binding regions containing uniquely aligned reads (Fig. 1), suggesting that these regions share the same properties as the uniquely aligned regions and that they may contain functional binding sites.

HITS-CLIP analysis of splicing targets. We next studied the effects of CUGBP1/MBNL1 binding on alternative splicing. CUGBP1 tags are clustered in intronic regions flanking alternative rather than

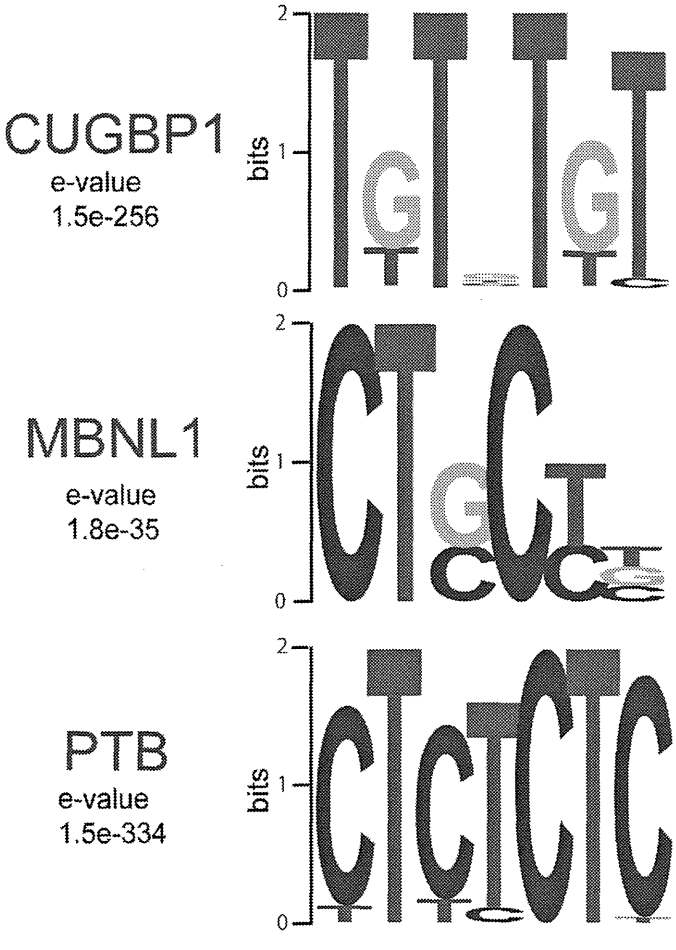


Figure 1 | Binding motif analysis. WebLogos of consensus binding motifs of CUGBP1, MBNL1, and PTB generated by the MEME motif analysis tool. The likelihood of finding the indicated motif by chance is indicated as an E-value.

constitutive exons (Fig. 2a). MBNL1 tags are similarly clustered in intronic regions flanking alternative exons, and are also enriched in alternative and constitutive exons. In order to investigate if and how CUGBP1/MBNL1 binding around splice sites regulate alternative splicing, we knocked down these factors by siRNA in undifferentiated C2C12 cells (Supplementary Fig. S4a). We analyzed alterations of splicing globally using the Affymetrix Mouse Exon 1.0 ST Array (GEO accession number, GSE29990) and identified 8 CUGBP1-responsive and 24 MBNL1-responsive exons (Supplementary Table 1, Figs. S5 and S6abc). We also analyzed 29 CUGBP1-tagged and 51 MBNL1-tagged exons/introns known to be alternatively spliced according to the ENSEMBL version *e!*61, and identified 16 CUGBP1-responsive and 21 MBNL1-responsive exons by RT-PCR (Supplementary Figs. S5 and S6abc). We made the compiled dataset C, which is comprised of the 24 CUGBP1-regulated exons (15 skipped and 9 included), as well as the compiled dataset M consisting of the 45 MBNL1-regulated exons (25 skipped and 20 included). The datasets include 1 and 9 previously identified target exons of CUGBP1 and MBNL1, respectively (Supplementary Fig. S5). In addition, 9 exons are shared between datasets C and M. Mbnl1 siRNA sufficiently suppressed MBNL1 expression up to day 3 after differentiation (Supplementary Fig. S4b), and we observed that as many as 44 of the 45 MBNL1-regulated exons in dataset M respond similarly to MBNL1 knockdown in both differentiated and undifferentiated cells (Supplementary Figs. S4 and S5).

We also made dataset M2 that includes 26 additional MBNL1-dependent cassette exons (15 skipped and 11 included) that were previously identified in skeletal muscle of MBNL1 knockout mice

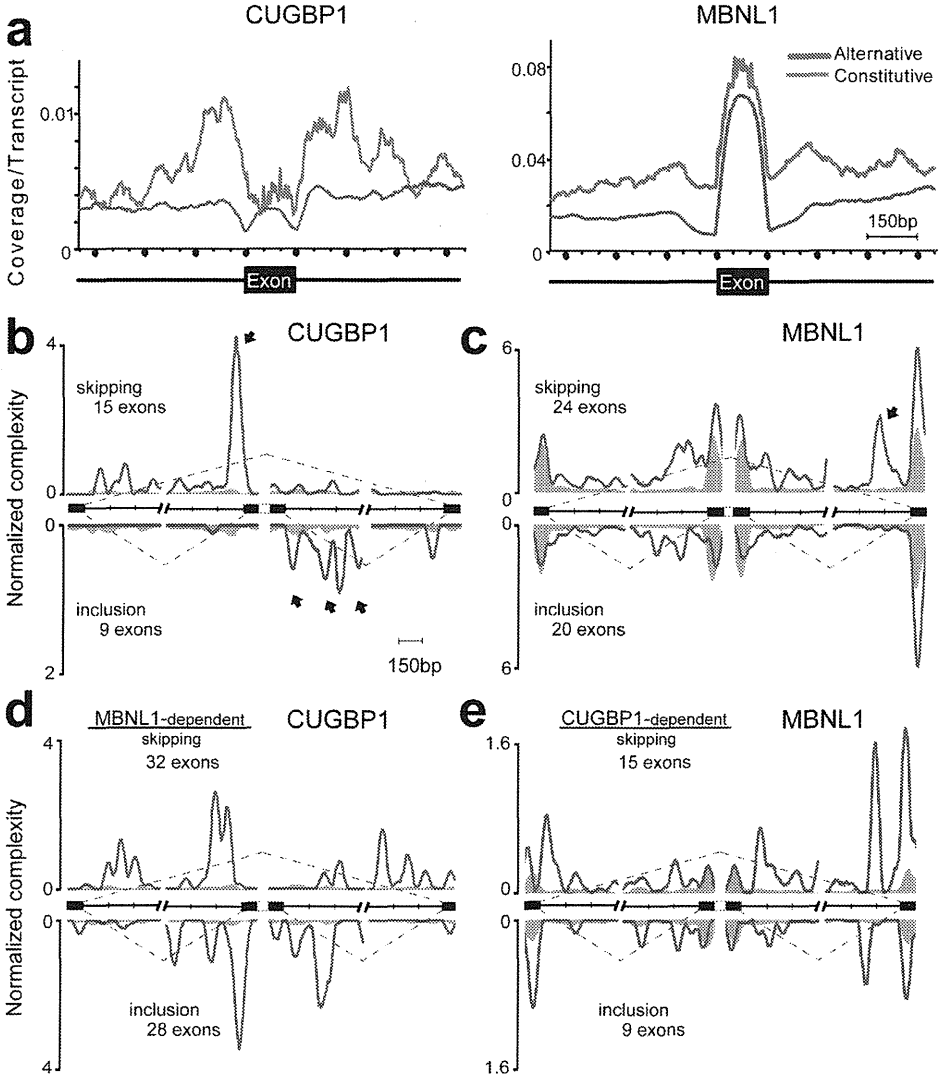


Figure 2 | Mapping of CLIP-tags on exon-intron structures. (a) Distributions of CLIP-tags on constitutively or alternatively spliced exons and the flanking intronic regions. The abscissa indicates an intron-exon-intron structure. The sizes of all the exons are normalized to 150 nucleotides. Numbers of exonic CLIP-tags are also normalized accordingly. Intronic CLIP-tags within 500 nucleotides upstream or downstream of exons are indicated. The number of CLIP-tags is normalized for the number of transcripts belonging to either category of constitutive and alternative exons. (b) Normalized complexity map of CUGBP1 at CUGBP1-dependent splice sites. Twenty-four CUGBP1-regulated splicing events in dataset C in undifferentiated C2C12 cells are compiled. (c) Normalized complexity map of MBNL1 at MBNL1-dependent splice sites. Forty-four MBNL1-regulated splicing events in differentiated C2C12 cells in dataset M are compiled. (d) Normalized complexity map of CUGBP1 at MBNL1-dependent splice sites. Sixty MBNL1-regulated splicing events in undifferentiated C2C12 cells in datasets M and M2 are compiled. (e) Normalized complexity map of MBNL1 at CUGBP1-dependent splice sites. Twenty-four CUGBP1-regulated splicing events in undifferentiated C2C12 cells in dataset C are compiled. Shaded areas represent an average of 100 sets of normalized complexity of 50 (b, c, and d) and 15 (e) randomly selected constitutive exons. Arrows indicate representative peaks that are explained in Results. Graphs represents results of the 2nd CLIP experiments for both CUGBP1 and MBNL1. Resutls of the 1st CLIP experiments are shown in Supplmentary Fig. S7.

(Supplementary Table S2)[33], and found that 18 exons are similarly regulated by *Mbnl1* knockdown in undifferentiated C2C12 cells (Supplementary Fig. S6d and Table S2).

We combined datasets C and M into a single composite pre-mRNA and made integrated RNA maps from our HITS-CLIP reads mapped to the corresponding genomic regions as previously described for Nova[30] and PTB[36]. This showed that CUGBP1 binding to upstream intronic regions facilitates exon skipping, whereas CUGBP1 binding to downstream intronic regions promotes exon inclusion (closed arrows in Fig. 2b and Supplementary Fig. S7a). Results of the 2nd experiments are shown in Fig. 2 and those of the 1st experiments are in Supplementary Fig. S7. In contrast, although the binding sites of MBNL1 are more diffusely distributed and less abundant in regions flanking splice sites (Fig. 2c), MBNL1 binding close to the 3′ end of the downstream intron induces exon skipping (closed arrow in Fig. 2c and Supplementary Fig. S7b). The presence of a similar peak in dataset M2 (closed arrow in Supplementary Fig. S7c) further supports this observation.

We next analyzed the interaction between CUGBP1 and MBNL1 in splicing regulation. We made an RNA map of CUGBP1-binding sites in MBNL1-regulated exons from datasets M and M2 (Fig. 2d and Supplementary Fig. S7e), as well as an RNA map of MBNL1-binding sites in CUGBP1-regulated exons from dataset C (Fig. 2e and Supplementary Fig. S7f). Both RNA maps demonstrate the presence of CUGBP1 clusters in MBNL1-responsive exons and vice versa, which suggests that CUGBP1 and MBNL1 are likely to regulate alternative splicing of some of the same exons.

MBNL1 and CUGBP1 both preferentially bind to the 3′ UTR. MBNL1 has so far solely been categorized as an exon/intron-binding splicing regulatory protein[6], but to our surprise we found that the majority (55%) of MBNL1-binding regions are located in 3′ UTRs (Fig. 3a). The same pattern with preferential binding (53%) in 3′ UTRs is observed for CUGBP1, while only 2% of PTB binding regions are located in 3′ UTRs (Fig. 3a). Similarly, when HITS-CLIP tags are mapped to the size-normalized positions of all the genes in the mouse genome, CUGBP1 and MBNL1 CLIP tags, but not PTB CLIP-tags, are enriched close to the 3′ ends of genes (Fig. 3b). Additionally, 610 3′ UTRs, which constitutes 28.7% of the CUGBP1-tagged 3′ UTRs and 17.4% of the MBNL1-tagged 3′ UTRs, are shared

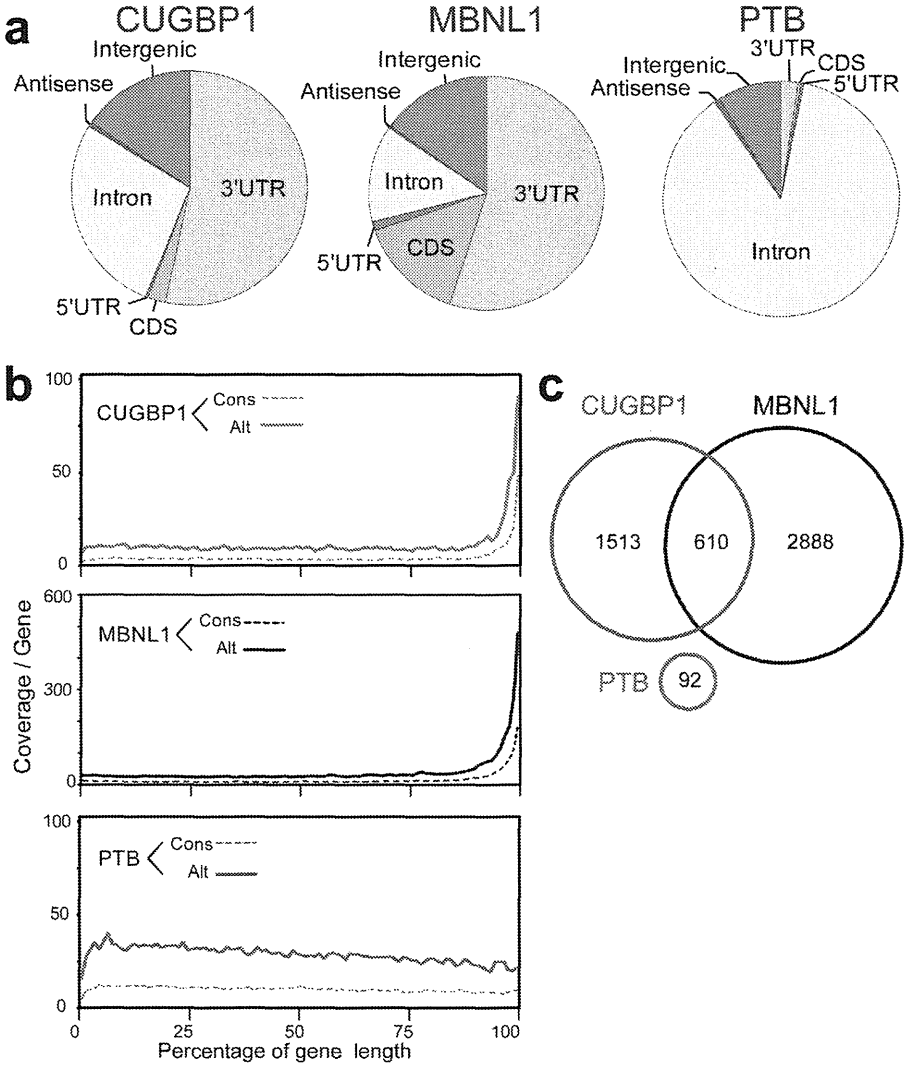


Figure 3 | Enrichment of CUGBP1 and MBNL1 CLIP-tags in the 3′ UTR. (a) Distributions of CUGBP1, MBNL1, and PTB binding regions. Binding regions are mapped to CDS (coding sequence), 5′ and 3′ UTRs, introns, intergenic regions (incl. tRNA and rRNA genes), or antisense within genes according to the UCSC knownGene annotation of the NCBI Build 37.1 mouse genome (mm9). Pie-charts show ratios of binding regions mapped to the indicated regions. (b) Distributions of CUGBP1, MBNL1, and PTB CLIP-tags mapped to the relative positions of all the mouse genes. The relative positions of the genes are shown in percentages of the gene length in abscissa. The broken lines represent 15,638 genes with constitutive transcriptional start and end sites (Cons), and the solid lines represent 7,477 genes with alternative transcriptional start or end site (Alt). (c) Venn diagram of the numbers of genes with CUGBP1-, MBNL1-, and PTB-binding regions within the 3′ UTR. Binding regions were identified using the SeqMonk software.

between these two proteins (Fig. 3c). All these data document that both CUGBP1 and MBNL1 preferentially bind to 3′ UTRs, indicating that this is a key function of both proteins in RNA processing. This suggests that the functional repertoire of MBNL1 should be expanded and that MBNL1, from being primarily regarded as regulator of alternative splicing, should also be considered as an important regulator of 3′ UTR-mediated processes, such as mRNA stability/degradation.

MBNL1 destabilize mRNAs. To analyze the function of CUGBP1/MBNL1 binding to 3′ UTRs, we made luciferase reporter constructs harboring CUGBP1/MBNL1-binding sites in the 3′ UTR. Since no CLIP tags were observed in the 3′ UTR of *Gapdh* (Supplementary Fig. S8), we made a *luciferase-Gapdh* 3′ UTR expression vector, and then inserted 12 repeats of GT and 7 repeats of CTG immediately after the stop codon of *luciferase* to introduce a CUGBP1-binding site (GU rich motif) and an MBNL1-binding site (YGCY motif), respectively (Fig. 4a). We also inserted 12 AC repeats as a control. Due to the high expression level of CUGBP1 in C2C12 cells we used HEK293 cells for transient transfection of these reporter constructs along with CUGBP1/MBNL1 expression vectors. For the constructs with *Gapdh* 3′ UTR alone or with AC repeats inserted, overexpression of CUGBP1 or MBNL1 had no effect on luciferase activity (Fig. 4b). For the GT repeat construct, overexpression of CUGBP1 decreased the luciferase activity, but MBNL1 had no effect. For the CTG repeat construct overexpression of MBNL1 dramatically decreased the luciferase activity, and also overexpression of CUGBP1 significantly reduced luciferase activity (Fig. 4b). In order to shed light on the mechanism underlying the observed decrease in luciferase activity we investigated the decay of *luciferase* mRNA. The SV40 promoter of the luciferase reporter constructs was replaced with a tet-repressible promoter, and HEK293 Tet-off cells were transiently transfected with these constructs. Doxycycline was added to the medium to stop transcription of the tet-responsive promoter, and the temporal profiles of *luciferase* and *GAPDH* mRNA levels were measured. Overexpression of MBNL1 together with the CTG repeat reporter construct resulted in highly increased decay of *luciferase* mRNA and CUGBP1 overexpression together with the GT repeat reporter construct also increased mRNA decay. Overexpression of either protein together with the *Gapdh* 3′ UTR control construct did not alter mRNA decay (Fig. 4c). These data demonstrate that binding of CUGBP1 and MBNL1 to the 3′ UTR promotes mRNA decay. To examine whether CUGBP1 and MBNL1 regulate decay of endogenous mRNAs, we next analyzed mRNA stability in actinomycin D treated C2C12 cells by expression arrays following siRNA knock down of CUGBP1 or MBNL1 (GEO accession number, GSE27583). To identify genes with reliable half-life estimates, we restricted our analysis to 195 transcripts using three conditions: (i) half-life between 2.5–5 hrs; (ii) correlation coefficient of fitting to an exponential decay greater than 0.9; and (iii) RMA-normalized signal values more than 100 at all time points. The median half-life of all the transcripts matching these criteria in the control is 3.56 hrs, whereas those from CUGBP1- and MBNL1-knocked down cells are significantly prolonged to 3.91 hrs and 3.73 hrs, respectively (Fig. 5a). We chose four additional representative mRNAs with a cluster of either CUGBP1- or MBNL1-tags in the 3′UTR, and confirmed by real time PCR that knockdown of either CUGBP1 or MBNL1 results in approximately two-fold increase in mRNA half-life of these target mRNAs (Fig. 5b). The half-lives of 100 out of 195 transcripts are prolonged both by knockdown of CUGBP1 and MBNL1, suggesting overlapping activity in the regulation of mRNA decay by CUGBP1 and MBNL1. We next analyzed the relationship between change in mRNA half-life and coverage of HITS-CLIP tags in the 3′ UTRs. We found that genes displaying prolongation of half-lives in response to CUGBP1 knockdown harbors more CUGBP1-tags in their 3′ UTRs, compared to those displaying shortening of half-lives (Fig. 5c). Similarly, genes that display prolongation of their half-lives in response to MBNL1 knockdown have more MBNL1-tags in their 3′ UTRs (Fig. 5c).

Gene Ontology analysis of CUGBP1/MBNL1-bound 3′ UTRs revealed that the terms 'cytoskeletal protein binding, 'transcription factor binding' and 'RNA binding' are significantly overrepresented for CUGBP1- and MBNL1-bound genes (Table 1).

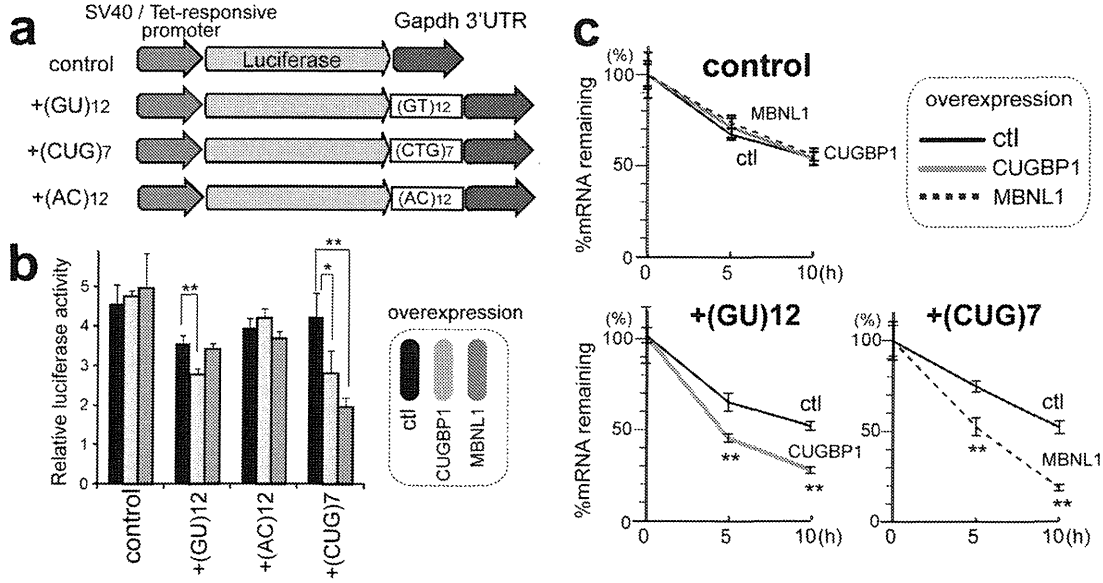


Figure 4 | Decay of *luciferase* mRNA by overexpression of CUGBP1/MBNL1. (a) Schemes of luciferase reporter plasmids harboring *Gapdh* 3′ UTR. Each construct was made carrying either SV40 or tet-responsive promoter. (b) Luciferase activity after overexpression of CUGBP1/MBNL1. HEK293 cells were transfected with the indicated SV40-driven luciferase reporter constructs. Luciferase activity is normalized for the transfection efficiency using co-transfection of pRL/SV40. (c) Decay of luciferase mRNA after overexpression of CUGBP1/MBNL1. HEK293 Tet-off cells were transfected with the indicated tet-responsive promoter-driven luciferase reporter constructs. Doxycycline was added to the medium to stop transcription at time 0. Temporal profiles of luciferase mRNA decay were quantified by real time RT-PCR and are normalized for *Gapdh* mRNA levels. All experiments were triplicated, and the mean and s.d. are indicated (* $p < 0.05$; ** $p < 0.01$).

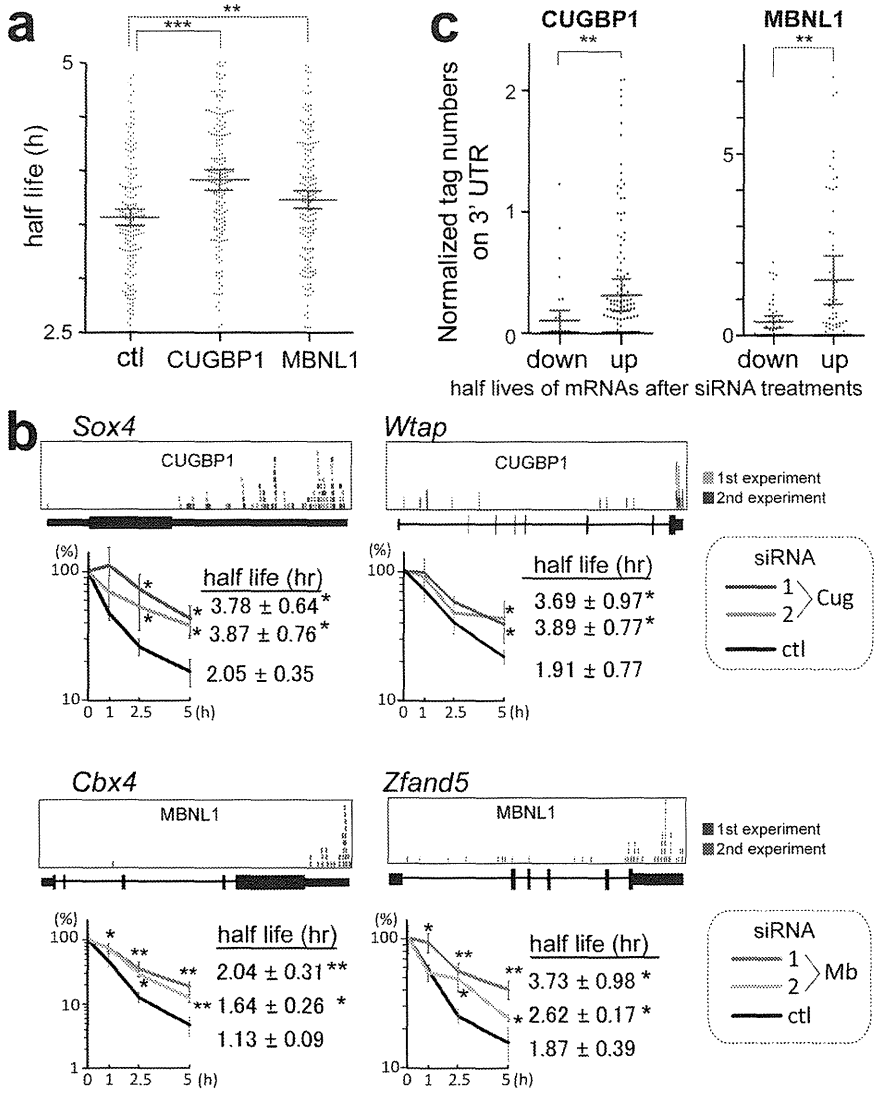


Figure 5 | Global analysis of mRNA decay by expression array of C2C12 cells treated with CUGBP1/MBNL1 siRNA. (a) Half-lives of mRNAs in C2C12 cells with the indicated siRNAs. Red lines represent means and 95% confidence intervals. ** $p < 0.01$ and *** $p < 0.001$. (b) Real-time RT-PCR analysis of the stability of four representative endogenous mRNAs, which were detected by expression arrays. CLIP-tag distributions are shown above each gene structure. C2C12 cells were treated with either control (ctl), CUGBP1 (Cug), or MBNL1 (Mb) siRNA. Actinomycin D was added to the medium to stop transcription at time 0. Temporal profiles of decay of the indicated genes were analyzed by real-time RT-PCR and are normalized for *Gapdh* mRNA levels. All experiments were triplicated, and the mean and s.d. are indicated (* $p < 0.05$ and ** $p < 0.01$). (c) Tag counts in the 3′ UTR of each gene are plotted in two categories of prolonging (up) and shortening (down) of half-lives after MBNL1 and CUGBP1 siRNAs. Red lines represent means and 95% confidence intervals. ** $p < 0.01$. Tag counts were normalized by the gene expression level at 0 h of cells treated with control siRNA.

Table 1 | The five most frequent Gene Ontology terms of mRNAs that are bound by CUGBP1 and MBNL1 to the 3′ UTR

CLIP data	GO ID	Term	P Value
CUGBP1	GO:0008092	cytoskeletal protein binding	1.58E-06
	GO:0003723	RNA binding	1.40E-04
	GO:0008134	transcription factor binding	9.65E-04
	GO:0051082	unfolded protein binding	0.003184
	GO:0019904	protein domain specific binding	0.006603
MBNL1	GO:0008092	cytoskeletal protein binding	7.31E-20
	GO:0008134	transcription factor binding	2.20E-08
	GO:0003723	RNA binding	0.001893
	GO:0019899	enzyme binding	0.002046
	GO:0032553	ribonucleotide binding	0.004210

We utilized the mRNAs that have more than 8-fold coverage of CLIP tags in their 3′UTR for the analysis by DAVID[53,54].

PITX2 is a homeobox transcription factor that regulates left-right asymmetric morphogenesis[37,38] and it is also deeply implicated in myogenesis during mouse embryonic development[39–41]. We found that the decay of *Pitx2* mRNA is prolonged by knocking down MBNL1, but not CUGBP1 in undifferentiated C2C12 cells (Fig. 6b and c). This is consistent with the fact that *Pitx2* harbors a much higher number of MBNL1-CLIP tags than that of CUGBP1-CLIP tags in the 3′ UTR (Fig. 6a). We also observed that down regulation of both CUGBP1 and MBNL1 decreases the decay of *Myod1* and *Mbnl2* mRNA, but not that of *Gapdh* mRNA (Supplementary Fig. S8). Similarly, down regulation of CUGBP1 decreases the decay of other myogenic transcription factors such as *Myog* and *Mef2a* mRNAs, and also of *Cugbp2* (Supplementary Fig. S9). Furthermore, knockdown of CUGBP1 and MBNL1 prolongs decay of *Mbnl1* and *Cugbp1* mRNAs, respectively, suggesting a mechanism for cross-regulation of expression of MBNL1, CUGBP1, and their family proteins (Supplementary Fig. S8).

To analyze more directly the role of MBNL1 binding to the 3′ UTR in regulation of mRNA decay, we examined the mRNA stability of firefly luciferase fused with the 3′ UTR of *Pitx2* (Fig. 6a). There are 11 YGCY motifs in the 3′ UTR of *Pitx2*, and 4 of the 11 motifs have MBNL1-CLIP tags. We introduced artificial mutations in these 4 motifs to prevent binding of MBNL1 (Fig. 6a). Consistent with the proposed role for MBNL1 in mRNA decay, we observe that disruption of the MBNL1-binding motifs in the *Pitx2*-3′ UTR abolished responsiveness to MBNL1 knockdown (Fig. 6d). Furthermore, immunoblots demonstrated that MBNL1-knockdown enhanced expression of endogenous PITX2 in C2C12 cells (Fig. 6e). These data suggest that MBNL1 promotes decay of *Pitx2* mRNA and thereby represses expression of the PITX2 protein.

Taken together, all of our data are consistent with a model where CUGBP1 and MBNL1 facilitate mRNA decay through binding to the 3′ UTR of target genes.

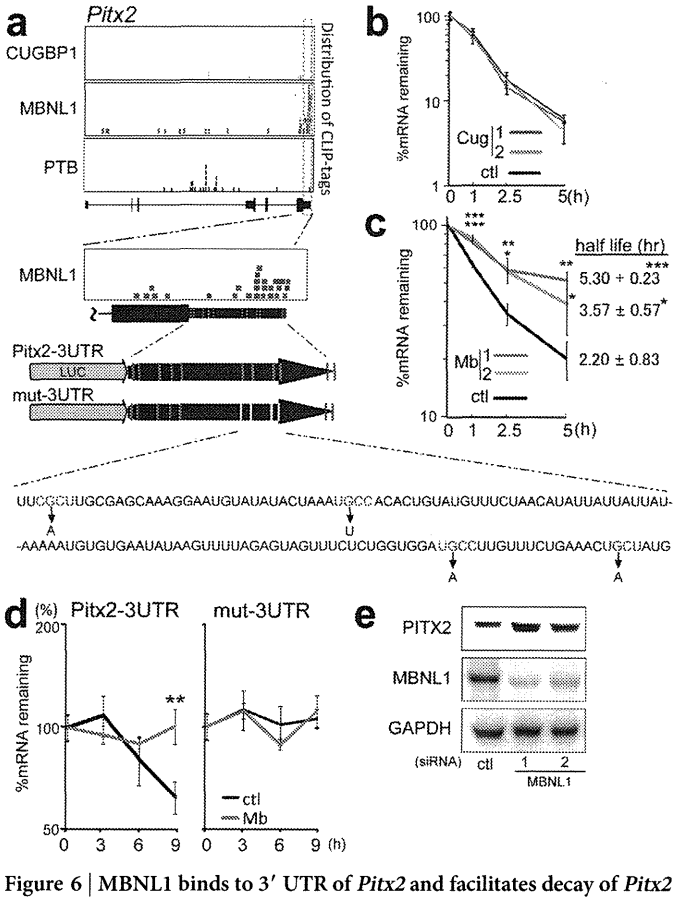


Figure 6 | MBNL1 binds to 3′ UTR of *Pitx2* and facilitates decay of *Pitx2* mRNA. (a) Distributions of MBNL1 CLIP-tags in the *Pitx2*-gene structure, and schemes of luciferase reporter plasmids harboring wild-type (*Pitx2*-3UTR) and mutated (mut-3UTR) 3′ UTRs of *Pitx2*. Red bars indicate locations of native YGCY motifs and yellow bars indicate those of mutated YGCY motifs. Individual mutations disrupting YGCY motifs are shown at the bottom. We introduced U at the 2nd YGCY motif to prevent formation of AU and AC repeats, which constitute a potential binding site of Hu proteins and hnRNP L, respectively. (b and c) Decay of endogenous *Pitx2* mRNA by siRNA-mediated knockdown of CUGBP1 (b) and MBNL1 (c). C2C12 cells were treated with either control (ctl), CUGBP1 (Cug) or MBNL1 (Mb) siRNA. Actinomycin D was added to the medium to stop transcription at time 0. Temporal profiles of decay of the indicated genes were analyzed by real-time RT-PCR and were normalized for *Gapdh* mRNA levels. All experiments were triplicated, and the mean and s.d. are indicated (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$). (d) Decay of luciferase mRNA in the Pitx2-3UTR and mut-3UTR constructs after either control (ctl) or MBNL1 (Mb) siRNA in TetON-C2C12 cells. Culture medium was replaced by medium without doxycycline to stop transcription at time 0. Temporal profiles of luciferase mRNA decay are quantified by real-time RT-PCR and are normalized for *Gapdh* mRNA levels. All experiments were triplicated, and the mean and s.d. are indicated (** $p < 0.01$). (e) Immunoblot of PITX2 in undifferentiated C2C12 cells at 48 hrs after transfection of MBNL1 siRNA.

Discussion

CUGBP1 and MBNL1 are developmentally regulated RNA-binding proteins that are causally associated with myotonic dystrophy type 1. In this study, we show that both CUGBP1 and MBNL1 preferentially bind to 3′ UTRs and destabilize the bound mRNAs. In particular, we show that CUGBP1 and MBNL1 destabilize myogenic differentiation factors and RNA-binding proteins. In addition, our results confirm and significantly expand the current knowledge of the splicing-regulatory effects of CUGBP1 and MBNL1. Taken together, the data from the present study indicates that CUGBP1 and MBNL1 are closely related and cross regulate alternative splicing and mRNA decay.

MBNL1 binding to 3′ UTRs has not been previously reported. We show for the first time that MBNL1 binds to 3′ UTRs and promotes mRNA decay in both artificial constructs and in endogenous genes. We also demonstrate by expression arrays that both CUGBP1 and MBNL1 facilitate mRNA decay by binding to 3′ UTRs. The present study demonstrates global *in vivo* interactions between CUGBP1 and 3′ UTRs and reveals that CUGBP1 also preferentially binds to 3′ UTR rather than exons/introns. We provide *in vivo* evidence that CUGBP1 facilitates mRNA decay of a broad spectrum of genes in addition to the previously reported genes[25–27,42–44].

Interestingly, we find that MBNL1 promotes decay of *Cugbp1* mRNA and that CUGBP1 promotes decay of *Mbnl1* mRNA, and that this is associated with corresponding changes at the protein level during differentiation of C2C12 cells (Supplementary Fig. S4b). This may suggest that expression of CUGBP1 and MBNL1 are mutually regulated in myogenic differentiation. Kuyumcu-Martinez and colleagues report that expanded CUG repeats of *DMPK* through an unknown mechanism leads to phosphorylation and thereby to stabilization of CUGBP1 in DM1 myoblasts[10]. Our studies additionally suggest that loss of MBNL1 in DM1 could lead to decreased decay of *CUGBP1* mRNA and hence to further increase of CUGBP1 activity. Although CUGBP1 is not upregulated in adult MBNL1-knockout mice, this mechanism could lead to increased misregulation of splicing and decay of the mRNAs of target genes in embryonic development that culminates in the DM1 phenotype.

We find that binding sites for CUGBP1 and MBNL1 are enriched around alternative cassette exons (Fig. 2a). The binding sites for CUGBP1 are prominent in adjacent intronic regions flanking alternative exons. Our functional analysis reveals that binding of CUGBP1 to the upstream intron facilitates exon skipping, whereas

binding to the downstream intron enhances exon inclusion (Fig. 2b). Interestingly, similar regulation of alternative splicing has been observed for NOVA, FOX2 and PTB[30,45,46], indicating the presence of a common underlying mechanism shared by these proteins.

In contrast to CUGBP1, MBNL1 tags are also enriched in coding exons. Until now, splicing *cis*-elements of MBNL1 have been mapped exclusively to introns, and no exonic *cis*-element has been reported to our knowledge[20,23,34,47,48]. Although MBNL1 preferentially binds to exons, MBNL1 binding to introns is enriched at alternative rather than constitutive splice sites (Fig. 2a. This enrichment is diffusely distributed throughout regions harboring 500 nt upstream or downstream of alternative exons, in contrast to the prominent intronic peaks observed for CUGBP1 tags. This could suggest that MBNL1 needs to bind simultaneously to the target exon and adjacent introns to regulate splicing. Functional analysis of MBNL1 reveals that binding of MBNL1 close to the 3′ end of the downstream intron facilitates exon skipping, whereas no characteristic binding pattern is observed for exons included in response to MBNL1 (Fig. 2c). PTB has also been reported to regulate alternative splicing through binding close to the 3′ end of the downstream intron[36]. In contrast to MBNL1, however, binding of PTB to this region promotes exon inclusion. We similarly find binding of PTB to this region in our HITS-CLIP data in MBNL1-regulated exons (Supplementary Fig. S7d). Interestingly, the MBNL1-binding motif is enriched in PTB-regulated exons[46]. MBNL1 may thus compete for binding with other splicing factors like PTB and regulate alternative splicing events.

Post-transcriptional gene expression regulation is crucial to achieve precise developmental and tissue-specific control of cellular processes. Our studies reveal that CUGBP1 and MBNL1 preferentially bind to the 3′ UTRs of mRNAs encoding RNA-binding proteins and transcription factors, which can regulate cell development. During development of murine skeletal muscles, the nuclear level of MBNL1 increases, while that of CUGBP1 decreases[9,12]. Genes with mRNAs that can be bound both by CUGBP1 and MBNL1 are likely to be down-regulated by CUGBP1 in undifferentiated cells. If these genes need to be tightly down-regulated also in differentiated cells, MBNL1 can substitute for CUGBP1 in order to achieve continued destabilization of the target mRNA. We conclude that finely-tuned expression of CUGBP1 and MBNL1 may be important regulators of myogenic differentiation through precise regulation of both alternative splicing and mRNA stability.

Methods

Antibodies. Antibodies to CUGBP1 (3B1), MHC (H300), myogenin (M225) and PTB (N20) were purchased from Santa Cruz Biotechnology. Anti-GAPDH pAb was purchased from Sigma. Anti-PITX2 pAb was purchased from Abcam. Anti-MBNL1 rabbit serum (A2764) was a kind gift of Dr. Charles A. Thornton at University of Rochester. The specificity of antibodies against CUG BP1 and MBNL1 is supported by the data in previous reports 2,3 and also by our siRNA experiments (Supplementary Fig. S1).

Cell culture. Detailed methods are included in the Supplementary Information.

HITS-CLIP. C2C12 cells were UV-irradiated at 400 mJ and CLIP was performed as previously described[49]. High-throughput 36-bp single-end and 40-bp single-end sequencing was performed using an Illumina Genome Analyzer II. All HITS-CLIP data were registered in ArrayExpress with an accession number E-MTAB-414 and in ENA with an accession number ERP000789. Detailed information is provided in the Supplementary Information.

Bioinformatics analysis. Illumina reads were first prepared by removing the 4-bp tag and filtering sequences composed primarily of Illumina adapter. The resulting reads were mapped to the mouse genome (NCBI Build 37.1/mm9) with default parameters using the BWA[50] mapping software. To extract consensus motifs from the mapped reads, we considered only uniquely aligned reads and first removed duplicate reads to avoid potential PCR-mediated deviations in addition to bias from very highly expressed transcripts. We then extended the reads to 110 nt, the expected mean of the CLIP fragments and used the SeqMonk software (www.bioinformatics.bbsrc.ac.uk/projects/seqmonk) to identify binding regions by using the program's built-in peak detection algorithm. Peaks were scored using both a reads per peak scoring scheme and a maximum depth scoring scheme (effectively the height of the peak) in order to filter out peaks. For the identification of CUGBP1- and MBNL1-binding regions, we used PTB as a negative control and removed peaks present in the PTB dataset as well. We then selected CUGBP1 peaks that were present in the two independent CUGBP1 CLIP experiments and MBNL1 peaks that were similarly corroborated by the two MBNL1 experiments. PTB binding regions were identified by removing peaks that were present in either of the four CUGBP1 and MBNL1 experiments. Finally, we restricted the set of binding regions to only those spanning 70–150 bp since this was the fragment length used in the CLIP experiments. We analyzed each dataset using a motif analysis tool, MEME[31], using a background Markov model based on the entire mouse genome.

We analyzed the mapped Illumina reads and binding regions and mapped them to UCSC knownGene annotations[51] of the mouse genome (NCBI Build 37.1/mm9) by writing and running Perl and Excel VBA programs, as well as by running BEDTools utilities[52]. Normalized complexity maps of CUGBP1/MBNL1/PTB-RNA interactions were generated as previously described[30]. For the control, normalized complexity map was similarly generated by analyzing 100 sets of 15 to 50 constitutive exons that were randomly selected from 118,969 constitutive exons in mm9. To identify enriched Gene Ontology terms, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID 6.7)[53,54].

Construction of plasmids. To construct luciferase reporter vectors with the 3′ UTR of *Gapdh* and *Pitx2*, 3′ UTRs of these genes were amplified by PCR. Amplified DNA was ligated into the *Xba*I and *Bam*HI sites of the pGL3-promoter vector (Promega) to substitute for the 3′ UTR of the firefly luciferase gene. DNA fragments harboring GT and CTG repeats were amplified by self-priming PCR using primers terminating in a *Xba*I site, and ligated into the *Xba*I site to make the pGL3P-*Gapdh*-3′ UTR.

To construct tet-responsive luciferase constructs, the tet-responsive promoter region was excised from pTRE-Tight vector (Clontech) with *Xho*I-*Hin*dIII site and cloned into the *Xho*I-*Hin*dIII site of the pGL3-promoter vector with the 3′ UTR of *Gapdh* and *Pitx2*. To introduce mutations in 3′ UTR of *Pitx2* in the luciferase construct, we used the QuikChange site-directed mutagenesis kit (Stratagene).

To construct expression vectors for MBNL1 and CUGBP1, the human MBNL1 cDNA and human CUGBP1 cDNA (Open Biosystems) were subcloned into the mammalian bidirectional expression vector pBI-CMV2 (Clontech), which should constitutively expresses the insert and AcGFP1.

RNA interference and transfection. The siRNA duplexes against CUGBP1 and MBNL1 were synthesized by Sigma. The sense sequences of the siRNAs were as follows: Cugbp1-1, 5′-GCUUUGGUUUUGUAAGUUA-3′; Cugbp1-2, 5′-GGCUUAAAGUGCAGCUCAA-3′; Mbnl1-1, 5′-CACUGGAAGUAUGUAGAGA-3′; and Mbnl1-2, 5′-GCACAAUGAUUGAUACCAA-3′. We purchased the AllStar Negative Control siRNA (1027281) from Qiagen. C2C12 cells were seeded on 24-well plates, and transfected with siRNA using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Tet-off advanced HEK293 cells were seeded on 96-well plates, and were transfected with luciferase reporter gene constructs using FuGENE 6 (Roche) according to the manufacturer's instructions. At 48 hrs after transfection, cells were either harvested for RNA extraction or processed for isolation of total proteins or nuclear extracts.

RT-PCR for splicing analysis. Total RNA was extracted using Trizol (Invitrogen) according to the manufacturer's instructions. cDNA was synthesized using an oligo-dT primer and ReverTra Ace (Toyobo), and PCR amplifications were performed using GoTaq (Promega) for 30–35 cycles. Sequences of the primers used for PCR are listed in the Supplementary Table S3. The intensities of PCR-amplified spliced products were quantified with the ImageJ 1.42q software (NIH). We then calculated a percentage of exon inclusion (% inclusion) as the ratio of the intensity of the upper band divided by the sum of intensities of all the bands.

Real-time RT-PCR for RNA stability analysis. Total RNA was extracted using RNeasy mini kit (Qiagen) or CellAmp Direct RNA Prep Kit (Takara) according to the manufacturer's instructions. cDNA was synthesized as described above and real-time PCR was performed using the Mx3005P QPCR System (Stratagene) and the SYBR Premix Ex Taq II (Takara). Sequences of the primers used for PCR are listed in Supplementary Table S4.

Microarray analysis. Total RNA was extracted using the RNeasy mini kit according to the manufacturer's instructions. We synthesized and labeled cDNA fragments from 100 ng of total RNA using the GeneChip WT cDNA Synthesis Kit (Ambion). The labeled cDNAs were hybrized to the Affymetrix Mouse Exon 1.0 ST Arrays for splicing analysis or the Affymetrix Mouse Gene 1.0 ST Arrays for analyzing temporal profiles of expression of CUGBP1/MBNL1-targeted genes following the manufacturer's protocols. The robust multichip analysis (RMA) algorithm was used to normalize the array signals across chips with the Affymetrix Expression Console software 1.1.2. All microarray data were uploaded to the Gene Expression Omnibus database (accession numbers, GSE29990 for exon arrays and GSE27583 for expression arrays).

Western blotting. For preparation of total cell lysates, cells were lysed in buffer A (10 mM HEPES pH 7.8, 10 mM KCl, 0.1 mM EDTA, 1 mM DTT, 2 μg/ml Aprotinin, 0.5 mM PMSF, 0.1% NP-40) and incubate on ice for 20 min. After sonication, samples were centrifuged (15,000 rpm, 5 min) and the supernatants were stored at −80°C for further experiments. For preparation of nuclear cell lysates, cells were suspended in 400 μl of buffer A. Nuclei were pelleted, and the cytoplasmic

proteins were carefully removed. The nuclei were then resuspended in buffer C (50 mM HEPES pH 7.8, 420 mM KCl, 0.1 mM EDTA, 5 mM $MgCl_2$, 2% Glycerol, 1 mM DTT, 2 μg/ml Aprotinin, and 0.5 mM PMSF). After vortexing and stirring for 20 min at 4°C, the samples were centrifuged, and the supernatants were stored at −80°C. Samples were analyzed on a 10% SDS polyacrylamide gel, and the proteins were transferred to Immobilon polyvinylidene difluoride membranes (Millipore). Membranes were blocked with 1% BSA in Tris-buffered saline containing 0.05% Tween20 (TBST) for 1 hr, incubated for 1 hr with primary antibodies in TBST, washed three times with TBST, and incubated for 1 hr with horseradish peroxidase-conjugated anti-mouse or -rabbit immunoglobulin (GE) diluted 1 : 5,000 in TBST. After three washes in TBST, the blot was developed with the enhanced chemiluminescence system (GE) according to the manufacturer's instructions.

Luciferase assay. HEK293 cells seeded on a 96 well plate were transfected with 10 ng of pGL3P-*Gapdh*-3′ UTR with or without GT and CTG repeats, 5 ng of pRL/SV40 (Promega), and 40 ng of pBI-CMV2-based CUGBP1 or MBNL1 expression vector using FuGENE 6. At 48 hrs after the transfection, the luciferase activity was measured using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions.

1. Licatalosi, D. D. & Darnell, R. B. RNA processing and its regulation: global insights into biological networks. *Nat Rev Genet* **11**, 75–87 (2010).
2. Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* **8**, 749–61 (2007).
3. Brook, J. D. *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* **68**, 799–808 (1992).
4. Day, J. W. & Ranum, L. P. RNA pathogenesis of the myotonic dystrophies. *Neuromuscul Disord* **15**, 5–16 (2005).
5. Larkin, K. & Fardaei, M. Myotonic dystrophy—a multigene disorder. *Brain Res Bull* **56**, 389–95 (2001).
6. Lee, J. E. & Cooper, T. A. Pathogenic mechanisms of myotonic dystrophy. *Biochem Soc Trans* **37**, 1281–6 (2009).
7. Turner, C. & Hilton-Jones, D. The myotonic dystrophies: diagnosis and management. *J Neurol Neurosurg Psychiatry* **81**, 358–67 (2010).
8. Miller, J. W. *et al.* Recruitment of human muscleblind proteins to (CUG)(n) expansions associated with myotonic dystrophy. *EMBO J* **19**, 4439–48 (2000).
9. Lin, X. *et al.* Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum Mol Genet* **15**, 2087–97 (2006).
10. Kuyumcu-Martinez, N. M., Wang, G. S. & Cooper, T. A. Increased steady-state levels of CUGBP1 in myotonic dystrophy 1 are due to PKC-mediated hyperphosphorylation. *Mol Cell* **28**, 68–78 (2007).
11. Iwahashi, C. K. *et al.* Protein composition of the intranuclear inclusions of FXTAS. *Brain* **129**, 256–71 (2006).
12. Kalsotra, A. *et al.* A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc Natl Acad Sci U S A* **105**, 20333–8 (2008).
13. Bland, C. S. *et al.* Global regulation of alternative splicing during myogenic differentiation. *Nucleic Acids Res* (2010).
14. Philips, A. V., Timchenko, L. T. & Cooper, T. A. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**, 737–41 (1998).
15. Ho, T. H., Bundman, D., Armstrong, D. L. & Cooper, T. A. Transgenic mice expressing CUG-BP1 reproduce splicing mis-regulation observed in myotonic dystrophy. *Hum Mol Genet* **14**, 1539–47 (2005).
16. Savkur, R. S., Philips, A. V. & Cooper, T. A. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nat Genet* **29**, 40–7 (2001).
17. Charlet, B. N. *et al.* Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol Cell* **10**, 45–53 (2002).
18. Begemann, G. *et al.* muscleblind, a gene required for photoreceptor differentiation in Drosophila, encodes novel nuclear Cys3His-type zinc-finger-containing proteins. *Development* **124**, 4321–31 (1997).
19. Teplova, M. & Patel, D. J. Structural insights into RNA recognition by the alternative-splicing regulator muscleblind-like MBNL1. *Nat Struct Mol Biol* **15**, 1343–51 (2008).
20. Ho, T. H. *et al.* Muscleblind proteins regulate alternative splicing. *EMBO J* **23**, 3103–12 (2004).
21. Cass, D. *et al.* The four Zn fingers of MBNL1 provide a flexible platform for recognition of its RNA binding elements. *BMC Mol Biol* **12**, 20 (2011).
22. Kanadia, R. N. *et al.* A muscleblind knockout model for myotonic dystrophy. *Science* **302**, 1978–80 (2003).
23. Fugier, C. *et al.* Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nature Medicine* **17**, 720–5 (2011).
24. Moraes, K. C., Wilusz, C. J. & Wilusz, J. CUG-BP binds to RNA substrates and recruits PARN deadenylase. *Rna* **12**, 1084–91 (2006).
25. Vlasova, I. A. *et al.* Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1. *Mol Cell* **29**, 263–70 (2008).
26. Lee, J. E., Lee, J. Y., Wilusz, J., Tian, B. & Wilusz, C. J. Systematic analysis of cis-elements in unstable mRNAs demonstrates that CUGBP1 is a key regulator of mRNA decay in muscle cells. *PLoS One* **5**, e11201 (2010).
27. Rattenbacher, B. *et al.* Analysis of CUGBP1 Targets Identifies GU-Repeat Sequences That Mediate Rapid mRNA Decay. *Mol Cell Biol* **30**, 3970–80 (2010).
28. Timchenko, N. A., Iakova, P., Cai, Z. J., Smith, J. R. & Timchenko, L. T. Molecular basis for impaired muscle differentiation in myotonic dystrophy. *Mol Cell Biol* **21**, 6927–38 (2001).
29. Timchenko, N. A. *et al.* Overexpression of CUG triplet repeat-binding protein, CUGBP1, in mice inhibits myogenesis. *J Biol Chem* **279**, 13129–39 (2004).
30. Licatalosi, D. D. *et al.* HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature* **456**, 464–9 (2008).
31. Bailey, T. L. & Elkan, C. The value of prior knowledge in discovering motifs with MEME. *Proc Int Conf Intell Syst Mol Biol* **3**, 21–9 (1995).
32. Marquis, J. *et al.* CUG-BP1/CELF1 requires UGU-rich sequences for high-affinity binding. *Biochem J* **400**, 291–301 (2006).
33. Du, H. *et al.* Aberrant alternative splicing and extracellular matrix gene expression in mouse models of myotonic dystrophy. *Nat Struct Mol Biol* **17**, 187–93 (2010).
34. Goers, E. S., Purcell, J., Voelker, R. B., Gates, D. P. & Berglund, J. A. MBNL1 binds GC motifs embedded in pyrimidines to regulate alternative splicing. *Nucleic Acids Res* (2010).
35. Kino, Y. *et al.* Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats. *Hum Mol Genet* **13**, 495–507 (2004).
36. Xue, Y. *et al.* Genome-wide analysis of PTB-RNA interactions reveals a strategy used by the general splicing repressor to modulate exon inclusion or skipping. *Mol Cell* **36**, 996–1006 (2009).
37. Hamada, H., Meno, C., Watanabe, D. & Saijoh, Y. Establishment of vertebrate left-right asymmetry. *Nat Rev Genet* **3**, 103–13 (2002).
38. Yashiro, K., Shiratori, H. & Hamada, H. Haemodynamics determined by a genetic programme govern asymmetric development of the aortic arch. *Nature* **450**, 285–8 (2007).
39. Dong, F. *et al.* Pitx2 promotes development of splanchnic mesoderm-derived branchiomeric muscle. *Development* **133**, 4891–9 (2006).
40. Shih, H. P., Gross, M. K. & Kioussi, C. Cranial muscle defects of Pitx2 mutants result from specification defects in the first branchial arch. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5907–12 (2007).
41. Gherzi, R. *et al.* Akt2-mediated phosphorylation of Pitx2 controls Ccnd1 mRNA decay during muscle cell differentiation. *Cell Death and Differentiation* **17**, 975–83 (2010).
42. Chen, H. H., Xu, J., Safarpour, F. & Stewart, A. F. LMO4 mRNA stability is regulated by extracellular ATP in F11 cells. *Biochem Biophys Res Commun* **357**, 56–61 (2007).
43. Zhang, L., Lee, J. E., Wilusz, J. & Wilusz, C. J. The RNA-binding protein CUGBP1 regulates stability of tumor necrosis factor mRNA in muscle cells: implications for myotonic dystrophy. *J Biol Chem* **283**, 22457–63 (2008).
44. Horb, L. D. & Horb, M. E. BrunoL1 regulates endoderm proliferation through translational enhancement of cyclin A2 mRNA. *Dev Biol* (2010).
45. Yeo, G. W. *et al.* An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat Struct Mol Biol* **16**, 130–7 (2009).
46. Llorian, M. *et al.* Position-dependent alternative splicing activity revealed by global profiling of alternative splicing events regulated by PTB. *Nat Struct Mol Biol* **17**, 1114–23 (2010).
47. Hino, S. *et al.* Molecular mechanisms responsible for aberrant splicing of SERCA1 in myotonic dystrophy type 1. *Hum Mol Genet* **16**, 2834–43 (2007).
48. Sen, S. *et al.* Muscleblind-like 1 (Mbnl1) promotes insulin receptor exon 11 inclusion via binding to a downstream evolutionarily conserved intronic enhancer. *J Biol Chem* **285**, 25426–37 (2010).
49. Ule, J., Jensen, K., Mele, A. & Darnell, R. B. CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* **37**, 376–86 (2005).
50. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–95 (2010).
51. Rhead, B. *et al.* The UCSC Genome Browser database: update 2010. *Nucleic Acids Res* **38**, D613–9 (2010).
52. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–2 (2010).
53. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
54. Dennis, G., Jr. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).

Acknowledgements

This work was supported by a JST-DASTI joint grant entitled "Strategic Japanese-Danish Cooperative Program on Molecular Medical Research", by Grants-in-Aid from the MEXT and MHLW of Japan, and by a grant from the Danish Medical Research Council (FSS Grant no. 271-07-342).

Author contributions

A.M., H.S.A, T.O., and M.I. performed the experiments. A.M., T.K.D., B.S.A., and K.O. analyzed the data. A.M., T.K.D., B.S.A. and K.O. prepared the manuscript. All authors reviewed the manuscript.

Additional information

Accession codes: All HITS-CLIP data were registered in ArrayExpress with an accession number E-MTAB-414 and in ENA with an accession number ERP000789.

All microarray data were uploaded to the Gene Expression Omnibus database with accession numbers, GSE29990 for exon arrays and GSE27583 for expression arrays.

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/

How to cite this article: Masuda, A. *et al.* CUGBP1 and MBNL1 preferentially bind to 3′ UTRs and facilitate mRNA decay. *Sci. Rep.* **2**, 209; DOI:10.1038/srep00209 (2012).

Journal of Human Genetics (2012) 57, 368–374
© 2012 The Japan Society of Human Genetics All rights reserved 1434-5161/12 $32.00
www.nature.com/jhg

ORIGINAL ARTICLE

Four parameters increase the sensitivity and specificity of the exon array analysis and disclose 25 novel aberrantly spliced exons in myotonic dystrophy

This article has been corrected since advance Online Publication, and a corrigendum is also printed in this issue

Yoshihiro Yamashita[1,6], Tohru Matsuura[1,2,6], Jun Shinmi[1], Yoshinobu Amakusa[1], Akio Masuda[1], Mikako Ito[1], Masanobu Kinoshita[3], Hirokazu Furuya[4], Koji Abe[2], Tohru Ibi[5], Ko Sahashi[5] and Kinji Ohno[1]

Myotonic dystrophy type 1 (DM1) is an RNA gain-of-function disorder in which abnormally expanded CTG repeats of *DMPK* sequestrate a splicing *trans*-factor MBNL1 and upregulate another splicing *trans*-factor CUGBP1. To identify a diverse array of aberrantly spliced genes, we performed the exon array analysis of DM1 muscles. We analyzed 72 exons by RT-PCR and found that 27 were aberrantly spliced, whereas 45 were not. Among these, 25 were novel and especially splicing aberrations of *LDB3* exon 4 and *TTN* exon 45 were unique to DM1. Retrospective analysis revealed that four parameters efficiently detect aberrantly spliced exons: (i) the signal intensity is high; (ii) the ratio of probe sets with reliable signal intensities (that is, detection above background *P*-value = 0.000) is high within a gene; (iii) the splice index (*SI*) is high; and (iv) *SI* is deviated from *SIs* of the other exons that can be estimated by calculating the deviation value (*DV*). Application of the four parameters gave rise to a sensitivity of 77.8% and a specificity of 95.6% in our data set. We propose that calculation of *DV*, which is unique to our analysis, is of particular importance in analyzing the exon array data.
Journal of Human Genetics (2012) **57**, 368–374; doi:10.1038/jhg.2012.37; published online 19 April 2012

Keywords: CUGBP1; Exon array; MBNL1; myotonic dystrophy

INTRODUCTION

Alternative splicing regulates developmental stage-specific and tissue-specific gene expressions and markedly expands the proteome diversity with a limited number of genes. High-throughput sequencing of total mRNAs expressed in cells has revealed that 98% or more of multiexon genes are alternatively spliced,[1] with an average of seven alternative splicing per multiexon gene.[2] Alternative splicing is achieved by exonic/intronic splicing enhancers/silencers (ESE, ISE, ESS, ISS) in combination with spatial and temporal expression of *trans*-acting splicing factors, such as serine/arginine-rich (SR) proteins and heterogeneous nuclear ribonucleoproteins.[3,4] Aberrations of alternative splicing are mediated by either mutations disrupting splicing *cis*-elements or dysregulation of splicing *trans*-factors.[5,6]

Myotonic dystrophy is an autosomal dominant multisystem disorder affecting the skeletal muscles, eye, heart, endocrine system and central nervous system. The clinical symptoms include muscle weakness and wasting, myotonia, cataract, insulin resistance, hypogonadism, cardiac conduction defects, frontal balding and intellectual disabilities.[7] 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)[8–10] 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).[11] In DM1, normal individuals have 5–30 repeats; mildly affected individuals have 50–80 repeats; and severely affected individuals have 2000 or more repeats of CTG.[12,13] In DM2, the size of expanded repeats is extremely variable, ranging from 75 to 11 000 repeats, with a mean of 5000 CCTG repeats.[11,14]

In DM1 and DM2, expanded CTG or CCTG repeats in the non-coding regions sequestrate a splicing *trans*-factor muscleblind encoded by *MBNL1* to intranuclear RNA foci harboring mutant RNA.[15] In addition, in DM1 cells, another splicing *trans*-factor CUG-binding protein encoded by *CUGBP1* is hyperphosphorylated by protein kinase C and is stabilized.[16–18] Dysregulation of the two splicing *trans*-factors then causes aberrant splicing of their target genes. A total

[1]Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan; [2]Department of Neurology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama, Japan; [3]Department of Frontier Health Sciences, Graduate School of Human Health Sciences, Tokyo Metropolitan University, Tokyo, Japan; [4]Neuro-Muscular Center, National Oomuta Hospital, Fukuoka, Japan and [5]Department of Neurology, Aichi Medical University School of Medicine, Aichi, Japan
[6]These authors contributed equally to this work.
Correspondence: Professor K Ohno, Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, 65 Tsurumai, Showa-ku, Nagoya 466-8550, Japan.
E-mail: ohnok@med.nagoya-u.ac.jp
Received 24 November 2011; revised 10 February 2012; accepted 23 March 2012; published online 19 April 2012

Table 1 A total of 28 aberrantly spliced exons and introns identified to date in skeletal and cardiac muscles in myotonic dystrophy

Gene[a]	*Affected exon/intron*[b]
ATP2A1 (*SERCA1*)[21,31]	Exon 22
ATP2A2 (*SERCA2*)[31]	Intron 19
BIN1[32]	Exon 11
CAPN3[21]	Exon 16
CLCN1[33]	Intron 2
CLCN1[34,35]	Exons 6b/7a
DMD[36]	Exon 71
DMD[36]	Exon 78
DTNA[37]	Exons 11A and 12
FHOD1 (*FHOS*)[21]	Exon 11a
FN1[38]	Exon 33
GFPT1 (*GFAT1*)[21]	Exon 10
INSR[39]	Exon 11
KCNAB1[40]	Exons 2b/2c
LDB3 (*ZASP*)[21]	Exon 11 (189-nt exon 7)
MBNL1[21]	Exon 7 (54-nt exon 6)
MBNL2[21]	Exon 7 (54 nt, no exonic annotation)
MEF2C[41]	Exons 4 and 5
MTMR1[42]	Exons 2.1 and 2.2
MYOM1[43]	Exon 17a
MYH14[44]	Exon6
NRAP[21]	Exon 12
PDLIM3 (*ALP*)[21]	Exons 5a/5b
RYR1[31]	Exon 70
TNNT2[45]	Exon 5
TNNT3[46]	Fetal exon
TTN[21]	Exons Zr4 and Zr5 (138-nt exon 11 and 138-nt exon 12)
TTN[21]	Exon Mex5 (303-nt exon 315)

[a]Gene symbols by the HUGO Gene Nomenclature Committee. Symbols in parentheses represent alternative symbols that are used in the reference.
[b]Exon numbers in parentheses represent annotations of the NCBI Build 36.3.

of 28 exons/introns of 22 genes have been identified to date in the skeletal and cardiac muscles in myotonic dystrophy (Table 1).

The Affymetrix GeneChip Human Exon 1.0 ST array contains ~1.4 million probe sets comprised of ~5.4 million probes. The exon array is designed to measure the expression level of each exon and to enable quantitative analysis of alternative splicing. In designing the exon array, probe selection region(s) (PSR) are placed within each exon throughout the genome. Each PSR has a unique probe set ID and carries four probes. A group of PSRs placed on a single exon is given a unique exon cluster ID. In most instances, each exon cluster ID, which represents an individual exon, carries a single PSR. A group of exon clusters spanning a single gene has a unique transcript cluster ID. The exon array thus carries 30–40 probes along the entire length of each gene. Data analysis of the exon array, however, is more complicated than that of the expression array as in the HuEx1.0 ST exon array, (i) each probe set is comprised of only four probes, (ii) each probe does not have a corresponding mismatched probe and (iii) each probe cannot be optimally designed due to a short span of the PSR.

In an effort to elucidate a diverse array of alternatively spliced genes in myotonic dystrophy, we performed the exon array analysis with skeletal muscles of three DM1 patients and three normal controls. In the course of the analysis, we tested alternative splicing of 72 exons by RT-PCR and found that 27 were alternatively spliced, whereas 45 were not. We sought for parameters that best discriminate true and false positives, and found that four parameters discriminate the true and false positives with a sensitivity of 77.8% and a specificity of 95.6%.

MATERIALS AND METHODS

Patient samples

Skeletal muscles were previously biopsied for diagnostic purposes. Clinical features of the patients are summarized in Supplementary Table S1. Two control muscles were biopsied muscle specimens that showed no pathological abnormalities. One control muscle RNA was the Human Skeletal Muscle PCR-Ready cDNA from Life Technologies (Carlsbad, CA, USA). All experiments were performed under the IRB approvals of the Nagoya University Graduate School of Medicine and the Aichi Medical University. The samples were used for the current studies after appropriate informed consents were given. High-molecular weight DNA was extracted by the conventional proteinase K and phenol chloroform method. We determined the CTG repeat numbers at the 3′ UTR of the *DMPK* gene by Southern blotting and found that skeletal muscles of patients 1, 2 and white blood cells of patient 3 carried 3430, 4500 and 1500 CTG repeats, respectively. Our analysis underscored a notion that skeletal muscles have larger numbers of repeats compared to leukocytes.[19]

RNA preparation and array hybridization

Total RNA was extracted by the RNeasy Mini Kit (Qiagen, Hilden, Germany). We confirmed that the RNA integrity numbers were all above 7.0. Hybridization and signal acquisition of the HuEx1.0 ST exon array (Affymetrix, Santa Clara, CA, USA) were performed according to the manufacturer's instructions. The signal intensities were normalized by the RMA method using the Expression Console 1.1 (Affymetrix).

Exclusion of genes with undependable signals using four criteria

Before we analyzed our exon array data, we excluded genes and probe sets with undependable signals using the following criteria. First, the gene must be comprised of four or more exons. Second, the smaller detection above background (DABG) *P*-value in either controls or DM1 muscles is $\leqslant 0.01$ for a probe set to be analyzed. DABG is a detection metric generated by comparing perfectly matched probes to a distribution of background probes. Affymetrix expression arrays used a mismatched probe to measure the background signal for a specific probe, whereas Affymetrix exon arrays use shared background probes to estimate the background signals. As exon skipping results in low signals that give rise to high DABG *P*-values, we did not discard probe sets with unreliable signals in either controls or DM1, but not in both. Third, three or more dependable probe sets with DABG *P*-value $\leqslant 0.01$ should be included in a gene to be analyzed, and such probe sets should comprise 15% or more of all the probe sets on the gene. Fourth, the average signal intensities of either the controls or DM1 should be no less than 150. Among the 336 293 exonic probe sets in our data set, 103 543 probe sets met these criteria.

Unique exon cluster IDs and unique transcript cluster IDs

In order to provide our unique exon cluster IDs and unique transcript cluster IDs, we analyzed annotations of the NCBI human gene database build 36.3 by writing and running Perl programs on the PrimePower HPC2500/Solaris 9 supercomputer (Fujitsu Ltd, Tokyo, Japan). We analyzed the exon array signals on Microsoft Excel by making VBA programs. Partitioning of parameters to distinguish true and false positives was performed by the JMP statistical software Ver. 8.0.1 (SAS Institute, Cary, NC, USA) with its default settings.

RESULTS

We provided our unique transcript cluster IDs and exon cluster IDs for the exon array based on the NCBI RefSeq database

We analyzed muscle specimens of three DM1 and three controls using the HuEx1.0 ST exon array. In the course of the analysis, we noticed that the exon array annotations provided by the manufacturer are based on comprehensive collation of several different gene databases and do not match to any single annotation database. We thus exploited the NCBI RefSeq annotation and provided our unique exon cluster IDs (Supplementary data). We also provided our unique transcript cluster IDs because in the manufacturer's annotations some exon clusters either upstream or downstream of the RefSeq-defined

gene region are given the same transcript cluster ID as intragenic exon clusters. An example of the *MBNL1* annotations is shown in Figure 1. Our data set was comprised of 336 293 probe sets that were grouped into 218 622 exonic clusters on 27 208 transcript clusters. For 1766 probe sets, we assigned duplicated exonic and transcript clusters, as two genes shared the same exonic regions. Our data set thus utilized 23.5% (336 293/1 432 144) of probe sets placed on the array.

The Affymetrix HuEx-1.0-st-v2 annotations release 32 carries 284 805 'core' probe sets. The 'core' probe sets represent the RefSeq transcripts and the full-length GenBank mRNAs. The 'core' probe sets are grouped into 192 554 exonic clusters on 19 231 transcript clusters. When the Affymetrix 'core' annotations are compared with the NCBI Build 36.3 database that we utilized, 17 372 (6.1%) of the 284 805 'core' probe sets are on non-exonic regions according to NCBI. Conversely, 69 278 (6.0%) of the 1 147 338 'non-core' probe sets are on exonic regions according to NCBI.

Deviation values (*DVs*) of splice indices (*SIs*) provide essential information to distinguish true and false positives

According to the manufacturer's suggestions, we first calculated the normalized intensity (*NI*) of each exon cluster by dividing the signal intensity of a given exon cluster by a sum of all the exonic signals throughout the gene. We next calculated the *SI* by dividing NI_{DM1} by $NI_{control}$.[20] We also calculated *t*-test *P*-values of *SIs* between three controls and three patients.

$$NI\ (exoncluster_i) = signal\ intensity\ of\ exon\ cluster_i / expression\ level\ of\ a\ gene$$

$$SI\ (exoncluster_i) = NI\ (exon\ cluster_i)_{DM1} / NI(exon\ cluster_i)_{control}$$

Validation by RT-PCR of ~20 exons, however, revealed that only about a quarter of candidate exons were aberrantly spliced, whereas three quarters were not. In the course of analysis, we noticed that, in most cases, the *SIs* of the truly positive gene were all close to 1.0 throughout the gene, whereas those of the falsely positive gene were variable from probe set to probe set (Figure 2). In order to quantify how much the normalized *SI* of a particular exon is deviated from those of the other exons, we calculated the mean and standard deviation (*s.d.*) of *SIs* of the other exons. We then calculated the *DV* of the *SI* of an exon cluster of our interest.

$$DV\ (exon\ cluster_i) = [SI\ (exon\ cluster_i) - mean_{SI}]/SD_{SI}$$

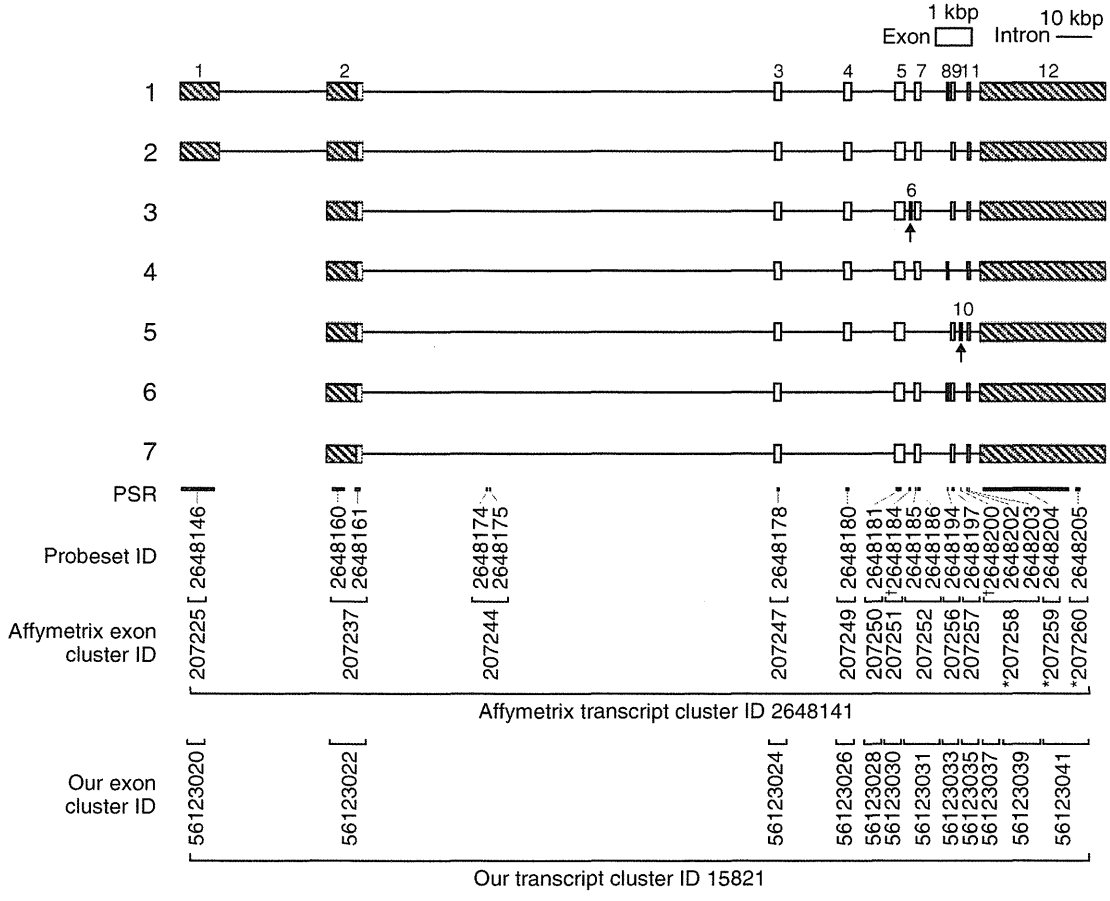


Figure 1 Comparison of Affymetrix annotations and our annotations. Seven alternative transcripts of *MBNL1* according to the NCBI Build 36.3 are drawn to individual exonic and intronic scales indicated at the top. The PSR bars represent the 'core' probe sets by Affymetrix. Each probe set is comprised of four probes (not shown). Each exon cluster corresponds to a single exon and carries one or more probe sets. A transcript cluster is comprised of exon clusters on the gene. Note that three 3′ exon clusters (asterisks) by Affymetrix are discordant to the genomic structure by NCBI: the Affymetrix exon cluster ID 207258 corresponds to exons 10 and 11 by NCBI, and we provided two different exon cluster IDs of 56123037 and 56123039. The exon cluster IDs 207259 and 207260 correspond to exon 12 by NCBI, and provided a single exon cluster ID of 56123041. Our unique exon cluster IDs and transcript cluster IDs throughout the genome are in the Supplementary data. Arrows indicate aberrant exons 6 and 10 identified in the current studies, and daggers indicate the corresponding probe sets. Shaded areas represent non-coding regions.

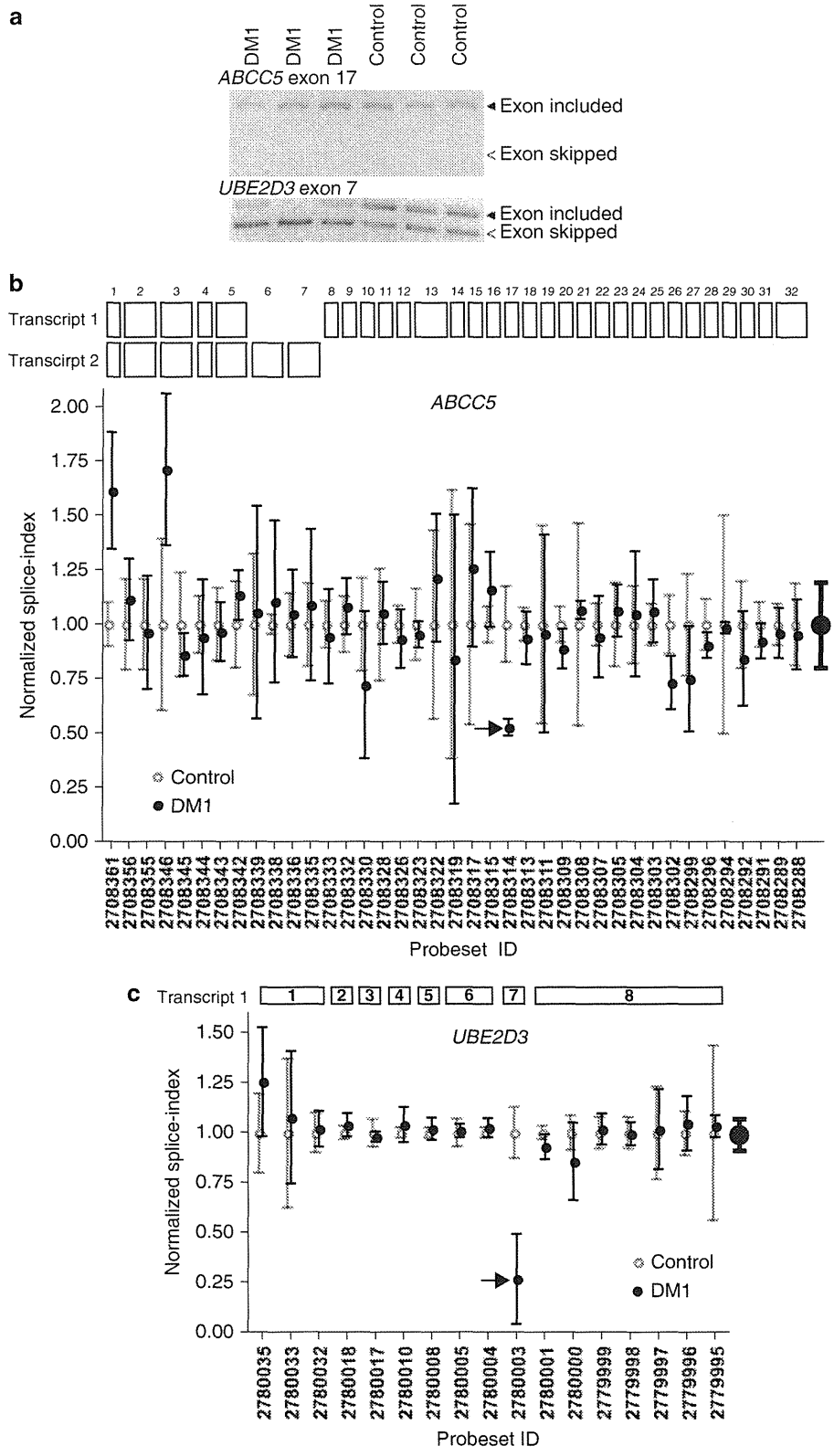


Figure 2 Falsely predicted skipping of *ABCC5* exon 17, and correctly predicted skipping of *UBE2D3* exon 7. (**a**) RT-PCR analysis of three DM1 muscles and three control muscles. (**b**) Splice index (an exonic signals divided by a sum of all the exonic signals throughout a gene) of each probe set on *ABCC5* in DM1 is normalized for that in controls. The normalized splice indices in DM1 are variable from probe set to probe set, giving rise to the mean and s.d. of 1.00 ± 0.20 (bold symbols). An arrow points to exon 17. The deviation value of probe set ID 2708314 on exon 17 is -3.7 s.d. (**c**) Normalized splice indices of *UBE2D3* in DM1 are less variable compared with those of *ABCC5*, which gives rise to the mean and s.d. of 1.00 ± 0.08 (bold symbols). An arrow points to exon 7. The deviation value of probe set ID 2780003 on exon 7 is -9.7 s.d.

We indeed found that *SIs* of the truly positive genes were all close to 1.0 and alternative exons tend to give rise to high *DV* values as explained in Figure 4a and the relevant statements below.

We analyzed 72 exons and identified 27 aberrant exons in DM1

In order to seek for aberrantly spliced exons/introns in DM1, we arbitrarily set three thresholds of $DV > 3.0$, $SI > 1.5$ and *t*-test *P*-value < 0.1. The three criteria were satisfied in 256 exons. Among these, we arbitrarily chose 72 exons. As we started our analysis without knowing which parameters were efficiently able to predict true positives, there were no strict objective criteria how we chose these exons. We, however, looked into the following features when we chose candidate exons: (i) a conspicuous value in one or more of the three parameters; (ii) alternative spliced exons annotated in the NCBI database; or (iii) a possible pathogenic gene that can be causally associated with DM1. We also avoided previously known aberrant splicing except for *LDB3* exon 7 and *MBNL1* exon 6, as the two exons were of special interest to us. RT-PCR analysis of the 72 exons revealed that 27 exons were indeed aberrantly spliced (Supplementary Figure S1; Table 2), whereas 45 exons were not (Supplementary Table S2), which gave rise to a positive predictive value of $27/72 = 37.5\%$ (Figure 3b). Most aberrant fragments were observed in normal and disease controls to variable extents, and we defined 'aberrant' splicing when the ratios of aberrant fragments in DM1 were more than those in any normal controls. In 11 of the 27 exons, intensities of aberrant fragments in DM1 exceeded those of normal controls but not all of disease controls, indicating that the splicing aberrations are not specific to DM1. In the remaining 16 'specific' exons (asterisks in Supplementary Figure S1), the ratios of aberrant fragments in DM1 were more than those in any disease controls. Especially, splicing aberrations in *LDB3* exon 4 and *TTN* exon 45 were almost exclusively observed in DM1, and were 'unique' to DM1. To summarize, among the 27 aberrantly spliced exons that we identified in the current studies, 25 were novel, 16 were 'specific' to DM1 and 2 were 'unique' to DM1 (*LDB3* exon 4 and *TTN* exon 45).

Table 2 A total of 27 aberrantly spliced exons in DM1 identified in the current studies

Gene	Exon (size in bp)	Probe set ID	Deviation value (DV)	Splice index (SI)	t-test P-value
AKAP13	16 (62)[b]	3606402	3.8	2.4[e]	0.0132
ATP5G2	1 (273)[b]	3456337	108.2	2.6	0.0065
FGD6	2 (2425)[c]	3466416	16.5	2.0	0.0144
ILF3	18 (1352)[b]	3820705	16.1	2.5	0.0048
LDB3	4 (368)[c]	3255989	36.0	6.9	0.0889
LDB3	7 (189)[a,c]	3256033	34.4	6.6	0.0000
MAP4K4	17 (231)[b]	2496832	16.5	3.1	0.0063
MBNL1	6 (54)[a, b]	2648184	16.5	3.1	0.0063
MBNL1	10 (64)[b]	2648200	5.4	4.0[e]	0.0041
MBNL2	8 (95)[b]	3497646	10.2	2.7	0.0008
MSI2	14 (73)	3728314	24.3	2.5	0.0550
MXRA7	4 (81)[b]	3771753	13.6	3.5	0.0464
MYBPC1	23 (54)[b]	3428645	46.9	4.8	0.0060
MYBPC1	31 (59)[b]	3428661	10.4	1.8	0.0808
NCOR2	10 (225)[c]	3476468	7.2	1.7	0.0163
NDUFV3	3 (1095)[b]	3922938	6.2	3.4[e]	0.0034
NEB	116 (105)[c]	2581073	32.7	2.8	0.0004
NEDD4L	13 (132)[c]	3790056	3.5	3.0[e]	0.0267
NEDD4L	14 (120)[c]	3790058	3.6	3.2[e]	0.0184
NEXN	2 (42)[c]	2343241	12.9	1.9	0.0169
NFIX	7 (123)[c]	3822162	18.3	2.3	0.0087
NR4A1	4 (1642)[b]	3415256	12.1	2.9[e]	0.0012
PPHLN1	7 (57)[d]	3412039	8.1	3.0[e]	0.0001
SOS1	21 (45)[c]	2549106	5.0	2.0[e]	0.0334
TBC1D15	7 (51)[c]	3422345	4.9	3.8[e]	0.0149
TTN	45 (375)[b]	2589787	62.7	4.9	0.0484
UBE2D3	11 (50)[b]	2780003	8.9	3.3[e]	0.0087

Exon numbers are according to the NCBI Build 36.3.
[a]Aberrant splicing in DM1 has been previously reported.
[b]Exons are known to be alternatively spliced according to (i) NCBI and ENSEMBL release 50.
[c](ii) ENSEMBL release 50.
[d]or (iii) dbEST alone.
[e]Signals in DM1 are weaker than those in controls, and the *SI* values are inversed.

In an effort to understand the rarity of 'specific' and 'unique' aberrant splicing, we examined the disease specificity of four previously reported aberrant splicing in DM1.[21] Limited availability of biopsied muscles hindered us from analyzing all the 28 previously reported splicing aberrations shown in Table 1. We found that aberrant splicing of *PDLIM3* exon 5 was 'specific'; that of *CAPN3* exon 16 was 'unique'; and those of *GFPT1* exon 10 and *NRAP* exon 12 were observed in normal and disease controls (Supplementary Figure S2). Thus, some of the previously reported splicing aberrations in DM1 (Table 1) are likely to represent muscle degeneration and/or regeneration.[22]

Four parameters increased the sensitivity of exon array analysis

Using the 72 analyzed exons, we next asked which parameters were able to discriminate the true and false positives. We analyzed 10 parameters and found that *DV* were most discriminative and *SI* followed (Figure 4). Additionally, we observed significant differences in average signal intensities and in ratios of probe sets with DABG *P*-value = 0.000. A DABG *P*-value is attached to a signal intensity of each probe set and represents reliability of the signal intensity. The *t*-test *P*-values were lower in true positives, but without statistical significance.

Comparison of the true (Table 2) and false (Supplementary Table S2) positives using the recursive partitioning functionality of the JMP 8.0.1 statistical software indicated four thresholds. The partitioning functionality of JMP seeks for the best splitting point of the best factor, X_i, among a group of factors that best discriminate the response Y. The 10 factors indicated Figure 4a were analyzed to discriminate the true and false positives. First, the signal intensities of either controls or DM1 should be more than 270. Second, the ratio of probe sets with DABG *P*-value = 0.000 either in controls or DM1 should be more than 0.05. Third, *DV* should be more than 10.0 or *SI* should be more than 2.6. Application of the four thresholds excluded 6 out of 27 true positives and 43 out of 45 false positives, and gave rise to a sensitivity of $21/27 = 77.8\%$ and a specificity of $43/45 = 95.6\%$ (Figure 3b). If we exclude the threshold for the *DV* and include the threshold for *t*-test *P*-value of less than 0.05, which are commonly used in the analysis of exon arrays, the sensitivity becomes as low as $15/27 = 55.6\%$, whereas the specificity rather becomes $45/45 = 100\%$. Thus, the inclusion of *DVs* in the analysis increases a chance of identifying aberrantly spliced exons by 22.2%, although a chance of detecting false positives is rather increased by 4.4%.

Although the significance of four parameters is demonstrated in our data set (Figure 4a), the thresholds should be unique to our data set and different thresholds need to be applied to different data sets. To prove this, we analyzed four human exon arrays of GSE21795,[23] GSE28672,[24] GSE24581[25] and GSE21840[26] in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). Each data set was comprised of a pair of three to five samples, and aberrant and alternative splicing events of a total of 23 exons were validated by RT-PCR in the original papers. Although *SIs* and *DVs* of the 23

a

256 exons met the arbitrary criteria
Fold-change > 1.5
t-test p-value < 0.10
Deivation value > 3.0
72 exons were validated by RT-PCR

Criteria	Aberrant	Normal	Total
Positive	27	45	72
Negative	n.a.	n.a.	n.a.

Positive predictive value = 37.5% (27/72)

b

Best discriminative criteria indicated by JMP
Signal intensities > 270
Ratio of probesets with reliable DABG p-value > 0.05
Splice index (*SI*) > 2.6
Deviation value (*DV*) > 10.0

Criteria	Aberrant	Normal	Total
Positive	21	2	23
Negative	6	43	49
Total	27	45	72

Sensitivity = 77.8% (21/27)
Specificity = 95.6% (43/45)

Figure 3 (**a**) Arbitrary criteria to search for aberrant splicing (left panel) and their results (right panel). (**b**) Four discriminative criteria indicated by JMP-IN (left panel) and their results (right panel).

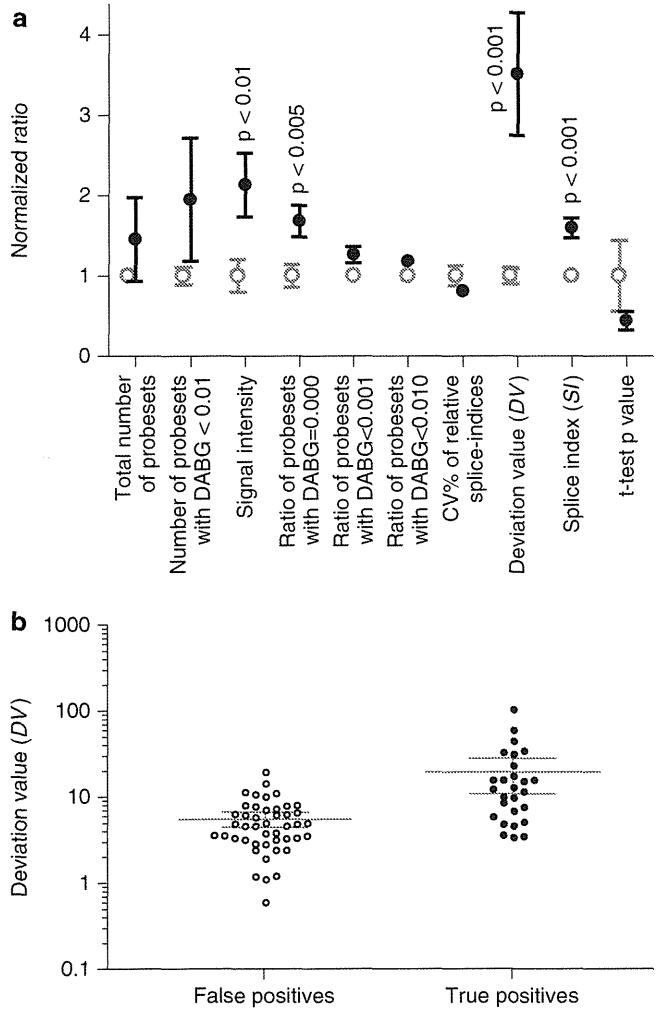


Figure 4 (**a**) Parameters that differentiate the true and false positives. Values are normalized to those of false positives, and the mean and s.e. are indicated. The true and false positives are indicated by solid and gray symbols, respectively. For *SI*, an inverse of the *SI* value is taken when the signal is decreased in DM1. Four parameters exhibit statistical significance with the Student's *t*-test. (**b**) Deviation values of 27 true and 45 false positives are plotted on a logarithmic scale. Gray lines indicate means and 95% confidence intervals.

exons were as high as 2.58 ± 1.05 and 8.3 ± 12.8 (mean and s.d.), respectively, 9 exons did not meet the criteria indicated in Figure 3b, which gave rise to a sensitivity of 14/23 = 60.9%. As no false positive results were documented in these papers, we could not calculate the specificity, and we could not estimate if calculation of four parameters indeed increases the sensitivity and specificity of the exon array analysis for these data sets.

DISCUSSION

A total of 25 novel aberrantly spliced exons in DM1

We identified 25 novel aberrantly spliced exons in DM1. Among these, aberrant splicing events of *LDB3* exon 4 and *TTN* exon 45 are 'unique' to DM1. Aberrations of the other 23 exons are observed in other muscle diseases with variable degrees. MBNL1 normally translocates from cytoplasm to nucleus in the postnatal period to induce adult-type splicing, and lack of muscleblind due to sequestration to RNA foci in myotonic dystrophy recapitulates fetal splicing patterns.[21,27] Downregulation of MBNL1 and upregulation of CUGBP1 is likely to occur in rejuvenating muscle fibers, and is likely to result in altered splicing patterns that we observe in disease controls. Not all aberrantly spliced exons in DM1, however, are observed in disease controls. Pathological significance of aberrant splicing in disease controls thus remains to be elucidated.

Exon array analysis

In expression arrays, fold-changes and *t*-test *P*-values have been successfully employed to detect altered gene expressions. On the other hand, these parameters are not sufficient to detect aberrant splicing in the exon array data. We thus sought for additional parameters and found that four parameters are informative to discriminate true and false positives: (i) the *DV*, (ii) the normalized splice index, (iii) the signal intensity and (iv) the ratio of probe sets with DABG *P*-value = 0.000. Application of these four parameters has enabled us to achieve a sensitivity of 77.8% and a specificity of 95.6%. On the other hand, the *t*-test *P*-values are not significantly lower in true positives. This represents that the threshold of *t*-test *P*-value $\leqslant 0.10$ is likely to be sufficient to exclude a large amount of false positives and that further stringent *P*-values would not help discriminate true and false positives.

In addition, our unique annotations of exon cluster IDs and transcript cluster IDs also make the *DV*s more dependable. This is because probe sets on rare transcripts or probe sets outside of the NCBI-defined gene region sometimes give rise to falsely strong signals