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.

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).

Protein-anchoring Strategy for Delivering Acetylcholinesterase to the Neuromuscular Junction

Mikako Ito[1], Yumi Suzuki[1], Takashi Okada[2], Takayasu Fukudome[3], Toshiro Yoshimura[4], Akio Masuda[1], Shin'ichi Takeda[2], Eric Krejci[5] and Kinji Ohno[1]

[1]*Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan;* [2]*Department of Molecular Therapy, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan;* [3]*Division of Clinical Research, Nagasaki Kawatana Medical Center, Nagasaki, Japan;* [4]*Department of Occupational Therapy, Nagasaki University School of Health Sciences, Nagasaki, Japan;* [5]*Université Paris Descartes, CNRS, UMR8194 , Paris, France*

Acetylcholinesterase (AChE) at the neuromuscular junction (NMJ) is anchored to the synaptic basal lamina *via* a triple helical collagen Q (ColQ). Congenital defects of ColQ cause endplate AChE deficiency and myasthenic syndrome. A single intravenous administration of adeno-associated virus serotype 8 (AAV8)-*COLQ* to *Colq*$^{-/-}$ mice recovered motor functions, synaptic transmission, as well as the morphology of the NMJ. ColQ-tailed AChE was specifically anchored to NMJ and its amount was restored to 89% of the wild type. We next characterized the molecular basis of this efficient recovery. We first confirmed that ColQ-tailed AChE can be specifically targeted to NMJ by an *in vitro* overlay assay in *Colq*$^{-/-}$ mice muscle sections. We then injected AAV1-*COLQ-IRES-EGFP* into the left tibialis anterior and detected AChE in noninjected limbs. Furthermore, the *in vivo* injection of recombinant ColQ-tailed AChE protein complex into the gluteus maximus muscle of *Colq*$^{-/-}$ mice led to accumulation of AChE in noninjected forelimbs. We demonstrated for the first time *in vivo* that the ColQ protein contains a tissue-targeting signal that is sufficient for anchoring itself to the NMJ. We propose that the protein-anchoring strategy is potentially applicable to a broad spectrum of diseases affecting extracellular matrix molecules.

Received 28 September 2011; accepted 31 January 2012; advance online publication 28 February 2012. doi:10.1038/mt.2012.34

INTRODUCTION

Acetylcholine (ACh) released from the nerve terminal is rapidly hydrolyzed by acetylcholinesterase (AChE) at the vertebrate neuromuscular junction (NMJ) to terminate cholinergic transmission. Three tetramers of catalytic AChE subunits are linked by a triple helical collagen Q (ColQ) to constitute a ColQ-tailed AChE.[1] The ColQ-tailed AChE is assembled in the endoplasmic reticulum and the Golgi apparatus.[2,3] ColQ carries three domains: (i) an N-terminal proline-rich attachment domain that organizes the catalytic AChE subunits into a tetramer, (ii) a collagenic domain that forms a triple helix, and (iii) a C-terminal domain enriched in charged residues and cysteines. ColQ-tailed AChE is organized in a secretory pathway, excreted, and anchored into the synaptic basal lamina using two domains of ColQ (Figure 1). First, the collagen domain harbors two heparan sulfate proteoglycan (HSPG)-binding domains[4] that bind to heparan sulfate proteoglycan such as perlecan in the synaptic basal lamina.[5] Second, the C-terminal domain of ColQ binds to MuSK, a muscle-specific receptor tyrosine kinase, on the postsynaptic membrane.[6] Human congenital defects of ColQ cause endplate AChE deficiency, in which the neuromuscular transmission is compromised.[7–9] Endplate AChE deficiency is an autosomal recessive disorder, which manifests as generalized muscle weakness, fatigue, amyotrophy, scoliosis, and minor facial abnormalities. Thirty-nine mutations of *COLQ* are currently registered in the Human Gene Mutation Database at http://www.hgmd.cf.ac.uk/. Ephedrine is effective for myasthenic symptoms to some extent,[10,11] though the underlying mechanisms of ephedrine efficacy remain elusive. We have developed a mouse model deficient in ColQ by deletion of the PRAD domain.[12] This strain recapitulates the phenotype of congenital myasthenic syndromes with AChE deficiency.

Gene therapy of endplate AChE deficiency is a complex issue both in humans and mice because ColQ is encoded by alternative promoters with a specific expression in subsynaptic nuclei of slow- and fast-twitch muscles.[13] The levels of AChE at the NMJ are supposed to be precisely controlled by the expression of ColQ and AChE,[14] as well as by a post-translational mechanism.[3] To treat endplate AChE deficiency in *Colq*-deficient mice, we delivered *COLQ* using adeno-associated virus (AAV) serotype 8, which has a tropism for muscles.[15] We used human *COLQ* instead of mouse *Colq* to foresee if the recombinant human *COLQ* is applicable to clinical practice in the future. Efficient rescue of AChE at the NMJ of AAV8-*COLQ*-injected mice prompted us to search for the molecular basis of these unexpected effects. We found that ColQ carries tissue-targeting signals that are necessary and sufficient to cluster AChE at the NMJ. This is the first report of a long-distance delivery of a large extracellular matrix complex over 50 nm in length and weighing over one million kDa in skeletal muscle. The findings of

Correspondence: *Kinji 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*

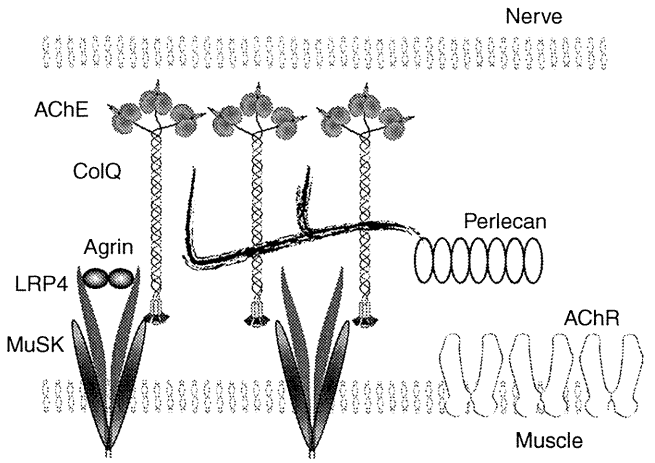

Figure 1 Schematic of anchoring of collagen Q (ColQ) to neuromuscular junction (NMJ). Twelve catalytic subunits of acetylcholinesterase (AChE) are attached to ColQ to form ColQ-tailed AChE. Two heparan sulfate proteoglycan-binding domains of ColQ are bound to perlecan. C-terminal domain of ColQ is bound to muscle-specific kinase (MuSK). Nerve-derived agrin binds to an LRP4–MuSK complex and induces rapsyn-mediated clustering of acetylcholine receptors (AChR) by phosphorylating AChR.

the present study open a new therapeutic avenue for treating many inherited defects of extracellular matrix proteins.

RESULTS

Intravenous administration of AAV8-*COLQ* normalizes motor functions of *Colq*$^{-/-}$ mice

We explored the recovery of the muscular phenotype of *Colq*$^{-/-}$ mice by viral delivery of a functional ColQ molecule. Therefore, we constructed a recombinant AAV serotype 8 carrying human *COLQ* cDNA. AAV serotype 8 (AAV8) is efficiently delivered to skeletal muscle after systemic injection.[16] We intravenously administered 1×10^{11}–2×10^{12} viral genome (vg) copies of AAV8-*COLQ* into 4-week-old *Colq*$^{-/-}$ mice. These mice exhibit muscle weakness, myasthenia, tremor, kyphosis, involuntary vocalization, and a slower growth rate than their wild-type littermates.[12] However, a single injection of 2×10^{12} vg, gradually improved their motor function to reach the level of that of wild type (Figure 2a). Furthermore, there were no signs of fatigue 6 weeks after the therapeutical injection (Figure 2b). Voluntary exercise in the treated mice also increased gradually but did not reach the level of wild type even at 5 weeks after injection (Figure 2c). The improved motor activities of treated mice are also demonstrated in **Supplementary Video 1**. Pairs of treated mice gave birth to *Colq*$^{-/-}$ pups and reared them to maturity. In longitudinal studies of three treated mice, all survived 18–20 months. Motor functions of the treated mice were declined at 48 weeks after injection but to the similar levels as those of wild type (Figure 2a,c). These observations clearly indicate the long-term therapeutic potential of a single viral injection of AAV.

AAV8-*COLQ* normalizes the neuromuscular synaptic transmission

To estimate recovery of neuromuscular transmission, we performed electrophysiological studies (Table 1). Treatment with

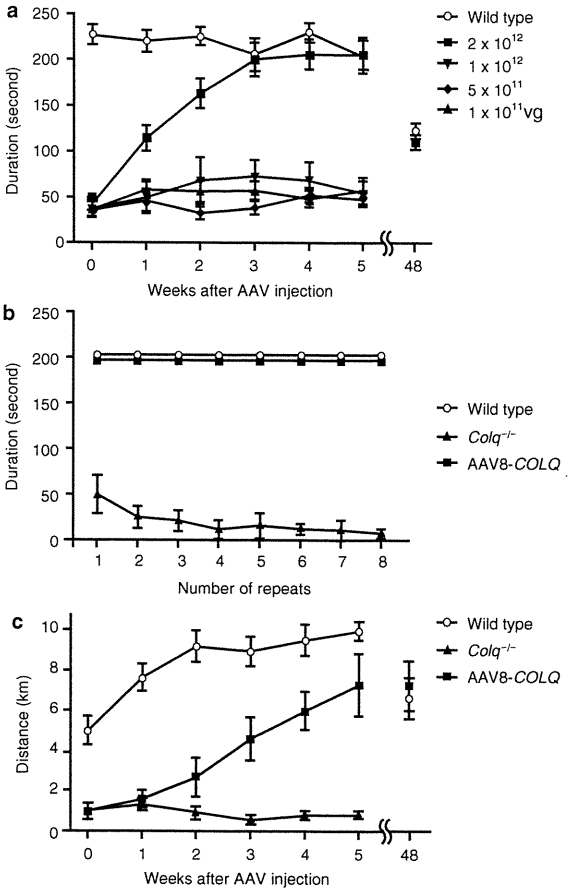

Figure 2 Exploration of motor function after intravenous injection of AAV8-*ColQ* to the tail vein of *Colq*$^{-/-}$ mice. (**a**) Motor function on the rotarod. The rotation was linearly accelerated from 0 to 40 r.p.m. in 240 seconds. Five groups of six mice were studied. Each group consisted of 4-week old mice and was either injected or not (control group) with increasing numbers of viral particles. Three weeks after their AAV8-*COLQ* injection, only the group of mice treated with 2×10^{12} vg remained on the rod as long as the wild-type littermates. Importantly, there was a progressive motor function recovery during the first 3 weeks after injection of *Colq*$^{-/-}$ mice. Symbols indicate mean and SE of six mice for each experiment. Mean and SE of the durations on the rotarod of two treated mice at 48 weeks after treatment is indicated along with that of the four age-matched wild-type mice. (**b**) Fatigue test using the rotarod was performed on three groups of a total of 18 mice. The rotation speed was fixed at 10 r.p.m. and the mice were immediately placed back on the rod each time they fell. Mice injected with 2×10^{12} vg exhibited no fatigue at 6 weeks after injection, whereas untreated *Colq*$^{-/-}$ mice fell increasingly more rapidly off the rod. (**c**) Voluntary movements were quantified by a counter-equipped running wheel. Plots show mean and SE of the number of rotations over 24 hours in each group of six mice (wild type, *Colq*$^{-/-}$, and AAV8-*COLQ*). Only the group of mice treated with 2×10^{12} vg increased the number of rotations every week but they did not reach the level of wild-type mice at 5 weeks after injection. Mean and SE of the number of rotations of two treated mice at 48 weeks after treatment is indicated along with that of the four age-matched wild-type mice. AAV8, adeno-associated virus serotype 8; ColQ, collagen Q.

AAV8-*COLQ* reduced decrements of the compound muscle action potentials in response to repetitive nerve stimulation at 2 Hz, reduced the amplitudes of miniature endplate potentials (MEPPs), shortened the miniature endplate potential decay time constants (Figure 3), and acquired responses to neostigmine. Endplate potential quantal content, which was decreased in

Table 1 Repetitive nerve stimulation and microelectrode studies

	Wild type	Wild type with neostigmine	*Colq*$^{-/-}$	*Colq*$^{-/-}$ with neostigmine	Treated *Colq*$^{-/-}$	Treated *Colq*$^{-/-}$ with neostigmine
Repetitive nerve stimulation[a]	0.92 ± 0.01* (2)	n.a.	0.58 ± 0.05 (3)	n.a.	0.76 ± 0.02* (3)	n.a.
EPP quantal content[b]	39.8 ± 2.3** (18)	n.a.	28.2 ± 1.8 (19)	n.a.	24.1 ± 1.6 (18)	n.a.
MEPP amplitude (mV)[c]	0.77 ± 0.04** (31)	1.52 ± 0.12 (18)	1.52 ± 0.11 (19)	1.52 ± 0.07 (10)	0.68 ± 0.02** (25)	0.98 ± 0.05** (24)
EPP amplitude (mV)[d]	30.6	n.a.	42.9	n.a.	16.4	n.a.
MEPP decay time (ms)[c]	1.77 ± 0.06** (31)	2.27 ± 0.08** (18)	3.07 ± 0.12 (19)	2.99 ± 0.09 (10)	2.45 ± 0.08** (25)	3.66 ± 0.09** (24)

Abbreviations: AChR, acetylcholine receptors; ColQ, collagen Q; EPP, endplate potential; MEPP, miniature endplate potential; n.a., not applicable.
Values represent mean ± SE. T = 29 ± 0.5 °C for EPP and MEPP recordings. Numbers in parenthesis indicate the number of recordings for repetitive nerve stimulation and the number of EPs from one or two mice for the other assays.
[a]Repetitive nerve stimulations were performed at 2 Hz, and the relative areas of compound muscle action potential (CMAP) of the fourth to the first stimulations are indicated. [b]Quantal content of EPP at 0.5 Hz stimulation corrected for resting membrane potential of –80 mV, nonlinear summation, and non-Poisson release. As the quantal contents of EPP are higher than 10, corrected values are indicated according to Cull-Candy *et al.*[45c] Normalized for resting membrane potential of –80 mV and a mean muscle fiber diameter of 55 µm. The actual fiber diameters were 45 ± 3.6 µm (mean ± SD, n = 31) for wild-type mice, 43 ± 3.0 µm (n = 19) for *Colq*$^{-/-}$ mice, and 46 ± 4.2 µm (n = 25) for the treated *Colq*$^{-/-}$ mice. [d]Estimated EPP amplitude is the product of the EPP quantal content and the MEPP amplitude. As AChR was partly blocked with curare for EPP recordings and not for MEPP recordings, we could not directly measure EPP amplitudes. Predicted low EPP amplitudes in treated mice suggest that the improvement of motor function was likely due to amelioration of depolarization block and/or of endplate myopathy.
*P < 0.05 and **P < 0.001 compared to *Colq*$^{-/-}$ mice by Student's t-test.

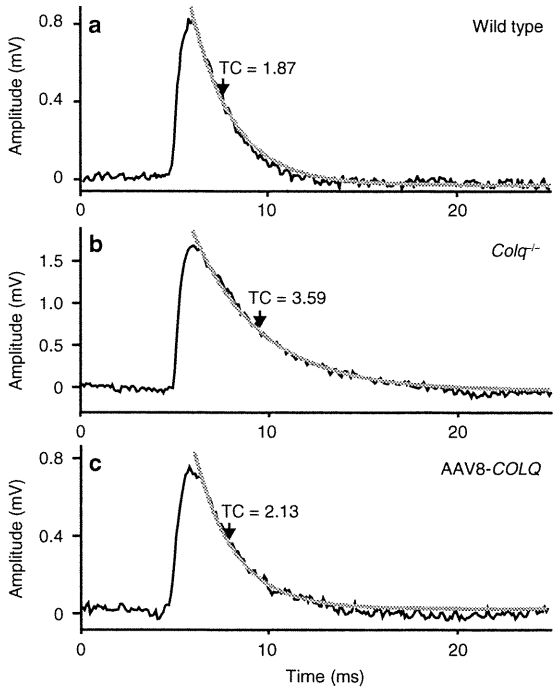

Figure 3 Representative miniature endplate potential (MEPP) recordings of diaphragm muscles of (a) wild type, (b) *Colq*$^{-/-}$, and (c) AAV8-*COLQ*-treated mice. (**b**) *Colq*$^{-/-}$ mice have higher MEPP amplitude and a longer decay time constant (TC) than (**a**) wild-type mice. AAV8-*COLQ* treatment shortened the decay TC and lowered the MEPP amplitude. Gray lines represent fitted exponential decay curves. AAV8, adeno-associated virus serotype 8; ColQ, collagen Q.

Colq$^{-/-}$ mice, was further decreased by the treatment, in contrast to our expectation.

Human ColQ-tailed AChE is anchored to the mouse NMJ *in vivo*

To further evaluate that the rescue was due to restitution of AChE at the NMJ, we used histological methods to visualize ColQ and AChE on muscle sections (Figure 4a–c). ColQ and AChE were colocalized to acetylcholine receptors (AChR) at the NMJ, confirming that ColQ-tailed AChE was specifically clustered to the target tissue. Although, we failed to observe improvement of motor functions with 1 × 10^{12} vg or less (Figure 2a), we still detected ColQ and AChE at NMJs with smaller amounts (data not shown). This suggests that a certain amount of viral genomes is required to exhibit improvement of motor deficits.

The ultrastructural morphology of treated mice also improved compared with age-matched *Colq*$^{-/-}$ mice (Figure 4d–f). The NMJ ultrastructures were variable from one to another in wild type, *Colq*$^{-/-}$, and treated mice, and we quantified the electron micrograph pictures (**Supplementary Table S1**). Quantitative analysis of presynaptic ultrastructures demonstrated that, in soleus slow-twitch muscle, Schwann cell invagination was mitigated, which increased the nerve terminal length, but the nerve terminal area remained essentially the same. Postsynaptic area and postsynaptic membrane length were also increased in soleus muscle of treated mice. In the extensor digitorum longus fast-twitch muscle, however, significant improvement was observed only in the ratio of enwrapped nerve terminal. Thus, the morphological improvements were more prominent in the soleus rather than in extensor digitorum longus muscles.

AAV8-*COLQ* restores the amount of ColQ-tailed AChE in the muscle to 89.3% of wild type

To estimate the efficiency of intravenous administration of AAV8-*COLQ*, we quantified the amount of the transduced *COLQ* mRNA, as well as ColQ-tailed AChE, in the muscle. We estimated the amount of *COLQ* mRNA in hindlimbs by a TaqMan probe, and found that the treated mice expressed the transduced *COLQ* at 92.5 ± 47.8% (mean ± SE, n = 4) of wild type. ColQ-tailed AChE from hindlimbs of the treated mice was fractionated by sucrose density-gradient ultracentrifugation. Sedimentation analysis revealed that AAV8-*COLQ* muscles have similar peaks of ColQ-tailed AChE species as those of wild type (Figure 5a–c). We also quantified the amount of globular AChE and ColQ-tailed AChE in gastrocnemius muscles of treated mice (Figure 5d). As previously reported, the amount of globular AChE was slightly lower in *Colq*$^{-/-}$ mice,[12] and this was normalized by treatment

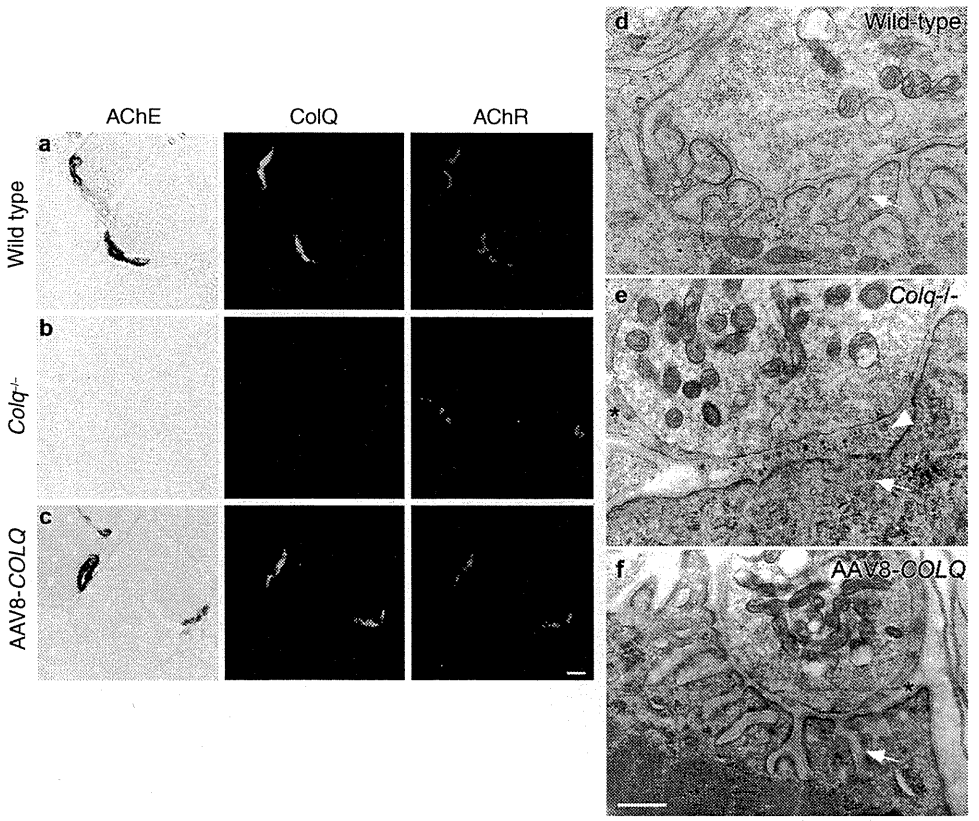

Figure 4 Histologies and ultrastructures of the neuromuscular junctions (NMJs). Localization of acetylcholinesterase (AChE) activity, collagen Q (ColQ), and acetylcholine receptors (AChR) in quadriceps muscles of (**a**) wild type, (**b**) *Colq*$^{-/-}$, and (**c**) AAV8-*COLQ* mice. Mice treated with 2 × 10^{12} vg of intravenous AAV8-*COLQ* express ColQ-tailed AChE at NMJ. AChE is stained for its activity. ColQ and AChR are detected by the polyclonal anti-ColQ antibody and α-bungarotoxin, respectively. Bar = 10 μm (**a–c**). Representative stainings of six mice in each group are indicated. Ultrastructures of soleus muscle NMJ (**d–f**). (**e**) *Colq*$^{-/-}$ mice show simplified synaptic clefts (arrow) and widening of the synaptic space (arrow head), whereas the NMJ ultrastructure of AAV8-*COLQ* mice (**f**) is indistinguishable from that of wild type (**d**). AAV8-*COLQ* mice still have small nerve terminals and invaginated Schwann cells (*). Bar = 1 μm (**d–f**). Representative ultrastructures of 27–41 electron micrograph (EM) pictures (see **Supplementary Table S1**) are indicated. AAV8, adeno-associated virus serotype 8.

with AAV8-*COLQ*. Treatment with AAV8-*COLQ* also restored ColQ-tailed AChE to 89.3 ± 9.6% (mean ± SD, $n = 4$) of wild type at 6 weeks after treatment. We also quantified ColQ-tailed AChE at 48 weeks after treatment and found that the amount was still 81.8 ± 21.6% (mean ± SD, $n = 2$) of the age-matched wild-type mice ($n = 3$). Although soleus slow-twitch muscle exhibited prominent improvement with the ultrastructural analysis, the available amount of soleus muscle was too small for the biochemical assay.

We also examined whether ColQ-tailed AChE was produced in the liver because AAV8 efficiently transduces hepatocytes.[15] AAV8-*COLQ* increased the *COLQ* mRNA level in the liver from 3.4 ± 0.34% (mean ± SE of five wild-type mice) to 61.3 ± 12.6% (mean ± SE of five treated mice) compared to those in the muscle of wild-type mice ($n = 5$). The *Ache* mRNA levels in the liver of wild type, *Colq*$^{-/-}$, and treated mice, however, were estimated to be <0.5% of that in wild-type muscle. The *Ache* mRNA levels in the liver were too low to be accurately quantified by real-time reverse transcription-PCR. Sedimentation profiles revealed no peaks of ColQ-tailed AChE in the liver of either wild type, *Colq*$^{-/-}$, or treated mice (**Supplementary Figure S1a–c**). This was probably due to lack of *Ache* expression. Globular AChE species observed in the sedimentation analysis was likely to represent AChE on the erythrocyte cell membrane.[17] These data suggest that AAV8-*COLQ* did not induce expression of ColQ-tailed AChE in the liver.

Local intramuscular injection of AAV8-*COLQ* expresses ColQ-tailed AChE at NMJs of noninjected limbs

Prominent improvements that we observed in AAV8-*COLQ*-treated mice raised a possibility that ColQ-tailed AChE moved from the transduced muscle cells to other muscle cells. We thus tested this possibility in the following experiments.

First, we have previously reported that the human recombinant ColQ-tailed AChE can be anchored to the synaptic basal lamina of the frog NMJ.[18] We tested this anchoring using mouse NMJs. We purified ColQ-tailed AChE expressed in HEK293 cells and incubated this with a section of skeletal muscle from *Colq*$^{-/-}$ mice. As expected, ColQ and AChE were detected at the mouse NMJ (**Supplementary Figure S2**), which supports the notion that ColQ-tailed AChE can be moved and anchored to the target *in vitro*.

Next, we tested whether ColQ-tailed AChE moved from the transduced muscles to the nontransduced muscles. We injected AAV8-*COLQ* to the left anterior tibial muscle. As expected, AChE and ColQ were rescued at the NMJs of the injected muscle. In addition, AChE and ColQ were also detected at all the examined

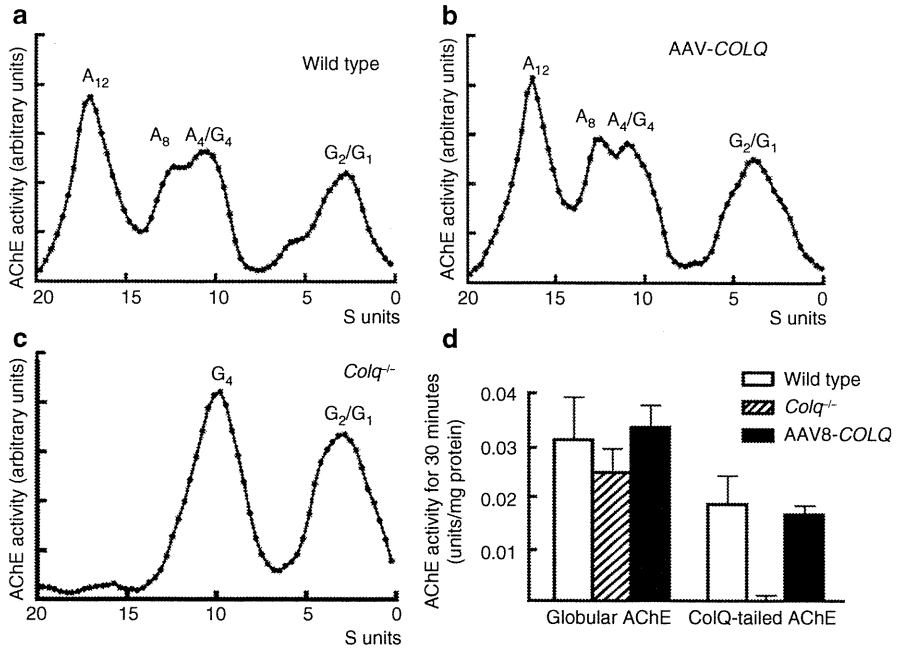

Figure 5 Quantification and biochemical analysis of acetylcholinesterase (AChE) recovery in muscles. Intravenous injection of 2×10^{12} vg of AAV8-*COLQ* into (**b**) *Colq*$^{-/-}$ mice gives rise to a sedimentation profile that is identical to that of (**a**) wild type, whereas (**c**) *Colq*$^{-/-}$ mice carry no collagen Q (ColQ)-tailed AChE. A_4, A_8, and A_{12} species carry 4, 8, and 12 AChE catalytic subunits attached to a triple helical ColQ. G_1, G_2, and G_4 species carry 1, 2, and 4 AChE catalytic subunits but without ColQ. A representative profile of three experiments is indicated. (**d**) Quantification of globular and ColQ-tailed AChE species (mean and SD, $n = 4$). The activity of ColQ-tailed AChE in the skeletal muscle of AAV8-*COLQ* mice is restored to 89 ± 10% of that of wild type. AAV8, adeno-associated virus serotype 8.

NMJs in noninjected muscles (data not shown). These results, however, could not exclude the possibility that AAV8-*COLQ* had been delivered to noninjected muscles in the form of a virus.

Local intramuscular injection of AAV1-*COLQ-IRES-EGFP* expresses ColQ-tailed AChE at NMJs of noninjected limbs

To reduce systemic delivery of AAV8 and to identify infected cells, we packed *COLQ* cDNA into the AAV serotype 1 (AAV1) that is known to transduce the injected muscle fibers locally.[19] In addition, we fused *COLQ* and internal ribosome entry site (*IRES*)-*EGFP* to express green fluorescent protein (GFP) in transduced cells synthesizing ColQ. We injected 2×10^{11} vg of AAV1-*COLQ-IRES-EGFP* into the left anterior tibial muscle of *Colq*$^{-/-}$ mice, while blocking the blood flow with a tourniquet for 20 minutes to restrict the distribution of the virus. The transduction efficiencies of AAV1-*COLQ-IRES-EGFP* were as follows: left anterior tibial muscle, 1.70 ± 0.29 viral copies per nucleus; right gastrocnemius muscle, 0.00100 ± 0.00079 copies; and bilateral brachial muscles, 0.00126 ± 0.00058 copies (mean ± SD, $n = 3$). Although only a fraction of the injected AAV1-*COLQ-IRES-EGFP* moved to noninjected limbs, we observed colocalization of ColQ and AChE at all the examined NMJs of right gastrocnemius, right tibialis anterior, both triceps, and both biceps in four examined mice (Figure 6a). We analyzed a total of 200–400 NMJs per muscle. In contrast, expression of intracellular enhanced GFP (EGFP) was not observed in noninjected limb muscles. We also quantified ColQ-tailed AChE in the noninjected bilateral forelimbs and right hindlimb, and found that the amounts were 21.5 ± 10.2% and 28.4 ± 10.0% (mean ± SD, $n = 4$), respectively, of those of wild type (Figure 6b).

ColQ-tailed AChE protein reaches and binds to remote NMJs

The presence of ColQ in noninjected muscles strongly suggests that the ColQ-tailed AChE is assembled intracellularly in one muscle and has moved to noninjected muscles, where it is anchored to the NMJs. To directly test this possibility, the gluteus maximus muscles of 5-week-old *Colq*$^{-/-}$ mice ($n = 4$) were injected daily with 2 μg of recombinant human ColQ-tailed AChE for 7 days. Histological analysis revealed the presence of ColQ and AChE in all of the examined NMJs from triceps muscles (Figure 7a). Quantitative analysis of ColQ signal intensities at the NMJs of noninjected triceps demonstrated that the ColQ-positive areas normalized for the AChR-positive area per NMJ became indistinguishable from that of wild type (Figure 7b). Furthermore, the ColQ signal intensity normalized for the AChR area per NMJs reached ~41.6% of that of wild type (Figure 7c). The *Colq*$^{-/-}$ mice could not hang on the wire at all, but the protein-injected mice acquired the ability to hang on the wire for two or more minutes from the fourth day of injection.

DISCUSSION

Effective and persistent gene therapy of ColQ with a single intravenous injection of AAV8-*COLQ*

We present an efficient and persistent recovery of AChE at the NMJ after a single intravenous administration of AAV8-*COLQ* in a *Colq*$^{-/-}$ mouse model of congenital myasthenic syndrome

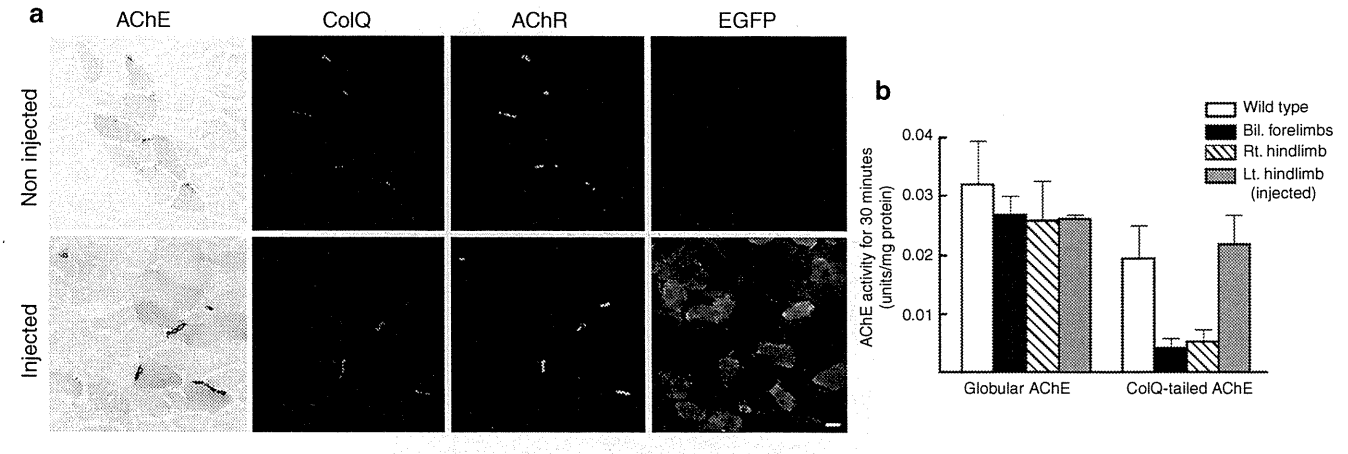

Figure 6 Intramuscular injection of 2×10^{11} vg of AAV1-*COLQ-IRES-EGFP* into the left anterior tibial muscle of *Colq*$^{-/-}$ mice. (**a**) Acetylcholinesterase (AChE) activity and collagen Q (ColQ) are colocalized to the acetylcholine receptors (AChR) in the injected muscle, as well as in the noninjected triceps muscle, although the signal intensities are not as high as those of the injected muscle. In contrast to ColQ, an intracellular molecule, enhanced green fluorescent protein (EGFP), is expressed only in the injected muscle, but not in the noninjected muscle. Bar = 10 μm. (**b**) Quantification of globular and ColQ-tailed AChE species of skeletal muscles (mean and SD, $n = 4$). In the injected left hindlimb, the activity of ColQ-tailed AChE is similar to that of wild type. In the noninjected both forelimbs and right hindlimb, the activities are 21.5 and 28.4% of wild type, respectively. AAV1, adeno-associated virus serotype 1.

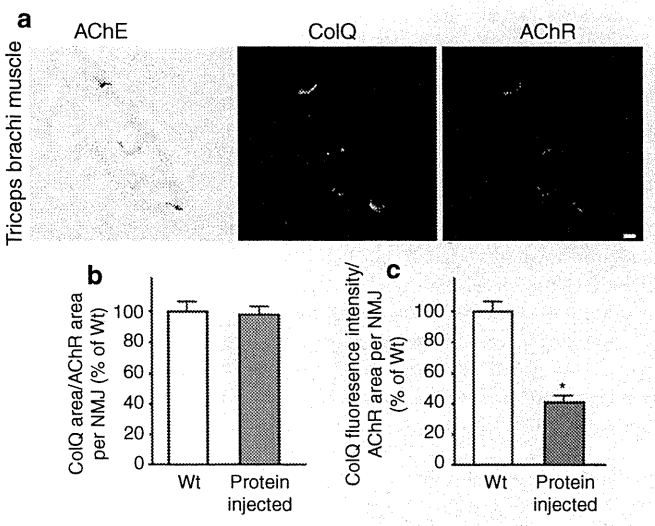

Figure 7 Injection of purified recombinant human collagen Q (ColQ)-tailed acetylcholinesterase (AChE). (**a**) Daily injection of 0.2 μg human recombinant ColQ-tailed AChE into the gluteus maximus muscles of *Colq*$^{-/-}$ mice rescues AChE activity and ColQ in the noninjected triceps where they are colocalized to acetylcholine receptors (AChR). Bar = 10 μm. (**b**) The size of ColQ-positive area is normalized for the size of AChR-positive area at the neuromuscular junctions (NMJs) of noninjected triceps. (**c**) Signal intensities of ColQ at the NMJs of noninjected triceps. Mean and SE are indicated. WT, wild-type mice, number of NMJs = 43; Protein injected, mice injected with ColQ-tailed AChE, number of NMJs = 42. *$P < 0.001$. Signal intensities are normalized to that of *Colq*$^{-/-}$ mice. Quantitative analyses were performed with the BZ-9000 microscope and the Dynamic Cell Count software BZ-H1C (Keyence).

with deficit in AChE. We observed ColQ-tailed AChE at all of the NMJs examined and the amount of the anchored AChE reached 89.3% of the wild-type level. The improved motor functions lasted at least 48 weeks after treatment and the treated mice survived 18–20 months, which is in contrast to at most 6-month lifespan of *Colq*$^{-/-}$ mice.[12] Although >99.5% of the vector genome stays episomal in mouse muscles even at 57 weeks after injection,[20,21] expression of the transgene in skeletal muscle lasts for 1.0–1.5 years without a decline in immunocompetent mice.[22,23] Our studies also underscore the long-lasting expression of the transgene delivered by AAV.

Rat *Colq*[24] and human *COLQ*[7] have two distinct promoters and generate ColQ1 and ColQ1a transcripts, which respectively include exon 1 and exon 1a and encode distinct signal peptides. A nerve-derived factor, calcitonin gene-related peptide, controls the expression of ColQ1a at the NMJs of fast-twitch muscles. However, in slow-twitch muscles, expression of ColQ1 occurs throughout the muscle fibers and is controlled by Ca^{2+}/calmodulin-dependent protein kinase II and myocyte enhancer factor 2.[24,25] As our viral construct was driven by the cytomegalovirus promoter, spatial and temporal regulation of ColQ expression should have been lost. In addition, our construct expressed ColQ1 and not ColQ1a, which was expected to be physiological for slow-twitch muscle but not for fast-twitch muscles. The prominent ultrastructural improvement in slow-twitch muscles rather than fast-twitch muscles may be partly because AAV8-*COLQ* encodes ColQ1 and not ColQ1a. This also suggests that the N-termini of ColQ1 and 1a have different functions. The pattern of ColQ expression resulting from our strategy was not physiological in three ways: (i) lack of subsynaptic nuclei-specific expression of ColQ, (ii) a ubiquitous cytomegalovirus promoter, and (iii) the exclusive expression of ColQ1. Despite these features, the motor and the synaptic functions are improved; AChE is locally accumulated at the NMJ in our treated mice. This suggests that the precise genetic control of the expression of ColQ is not the key factor for clustering of AChE and tissue-targeting signals of ColQ are sufficient to functionally restore AChE at the NMJ.

The protein-anchoring therapy

Although ColQ-tailed AChE in the serum of either wild type, *Colq*$^{-/-}$, or treated mice was less than a detection threshold in the

sedimentation analysis (**Supplementary Figure S1d–f**), anchoring of ColQ-tailed AChE to remote NMJs was supported by two lines of evidence: local intramuscular injections of AAV1-*COLQ-IRES-EGFP* (Figure 6) and of the purified recombinant ColQ-tailed AChE protein complex (Figure 7). In either case, AChE at the NMJ of the noninjected muscle originates from ColQ-tailed AChE arising from another source; not from local secretion and retention of ColQ-tailed AChE synthesized by subsynaptic nuclei followed by assembly and maturation in the postsynaptic area as in wild-type mice. The overlay of recombinant ColQ-tailed AChE *in vitro* either on normal muscle tissue sections of frog[5,18] or of *Colq*$^{-/-}$ mouse (**Supplementary Figure S2**) demonstrates that ColQ harbors a signal that targets AChE to the NMJ. The dual interactions of ColQ with MuSK[6] and perlecan,[5] which are both required in the overlay experiment,[18] are likely to restrict ColQ to the NMJ. In endplate AChE deficiency, point mutations that affect the binding of ColQ to MuSK prevent the accumulation of AChE.[18] Although no mutation has been reported at the heparan sulfate proteoglycan-binding domains of ColQ in endplate AChE deficiency, the reduction of perlecan mimicking Schwartz–Jampel syndrome reduces the level of AChE and ColQ at the NMJ.[26,27] All of these observations suggest that the combination of MuSK and perlecan determines the number of ColQ-tailed AChE anchored at the NMJ. This notion was previously termed as "molecular parking lots" by Rotundo and colleagues.[28]

ColQ-tailed AChE is a nanostructure made of a rigid collagen of 50-nm length and three AChE tetramers. ColQ-tailed AChE is apparently able to move from one muscle to another as demonstrated by clustering in the triceps muscle after protein injection into the gluteus. A similar approach is inherently employed by nature, as exemplified by fibronectin that is ubiquitously present in extracellular matrices and is largely derived from liver.[29] Injection of a protein complex is reported with laminin-111.[30] An intramuscularly or intraperitoneally injected laminin-111 is distributed to the basal lamina of skeletal and cardiac muscles in an mdx-mouse model of Duchenne muscular dystrophy. In contrast to our strategy, laminin-111 is not expressed or accumulated in normal or dystrophic adult muscles. Their studies exploit an ectopic deposition of laminin-111 to induce expression of α_7-integrin that stabilizes the sarcolemma of dystrophic muscle fibers.

The ColQ must be synthesized in cells that produce a splice variant T of AChE for obtaining the correct assembly of the complex.[2] A single muscle fiber harbors hundreds of nuclei that are functionally compartmentalized, and a molecule expressed in a single nucleus goes through the muscle fiber only for a short distance.[31,32] Thus, the multinucleation of muscle fibers is unlikely to have contributed to the restoration of function in the *Colq*$^{-/-}$ mice of our study. Similar specific clustering of a muscle-generated protein to the NMJ has been reported with laminin β2.[33] When laminin β2 is expressed throughout muscle fibers by the MCK promoter in transgenic mice, it is clustered at the NMJ.

The inability to achieve efficient and specific delivery of a transgene to the target tissue often prevents the application of gene therapy to model animals and patients.[34] Here, we propose the protein-anchoring strategy that provides a new therapeutic approach for congenital defects of extracellular matrix proteins.[35] The potential candidate molecules of the protein-anchoring therapy include laminin α2 causing laminin-α2-deficient congenital muscular dystrophy,[36] perlecan causing Schwartz–Jampel syndrome,[26,37] and collagen VI causing Ullrich syndrome.[38] It should be emphasized that this strategy can be potentially used for a huge number of diseases caused by mutations of genes encoding proteins of the extracellular matrices in general.

MATERIALS AND METHODS

Preparation of AAV carrying COLQ. Human *COLQ* cDNA[7] was cloned into pAAV-MCS (AAV Helper-Free system; Stratagene, Santa Clara, CA) that carries the cytomegalovirus promoter to obtain a pAAV-*COLQ*. We also inserted *IRES-EGFP* to make pAAV-*COLQ-IRES-EGFP*. To make AAV8-*COLQ*, HEK293 cells were cotransfected with the following plasmids: the proviral vector plasmid pAAV-*COLQ*, the AAV8 chimeric helper plasmid pRC8, and the adenoviral helper plasmid pHelper (Stratagene) using calcium phosphate coprecipitation method.[39] To make AAV1-*COLQ-IRES-EGFP*, we transfected HEK293 cells with pAAV-*COLQ-IRES-EGFP*, the AAV1 chimeric helper plasmid pRep2Cap1, and pHelper. The AAV particles were concentrated by CsCl gradient ultracentrifugation for 3 hours[40] and further purified with the quick dual ion-exchange procedures.[41] The viral titer was estimated by quantitative PCR in real-time using MX3000p (Stratagene).[42]

Administration of AAV carrying COLQ to Colq$^{-/-}$ mice. All animal studies were approved by the Animal Care and Use Committee of the Nagoya University Graduate School of Medicine. For intravenous administration, 1×10^{11}–2×10^{12} vg of AAV8-*COLQ* were injected into the tail vein of 4-week-old *Colq*$^{-/-}$ mice.[12] For intramuscular administration, 2×10^{11} vg of AAV1-*COLQ-IRES-EGFP* were injected into the left anterior tibial muscle of 4-week-old *Colq*$^{-/-}$ mice. The left proximal thigh was tightly ligated with a tourniquet for 20 minutes during intramuscular injection to prevent vascular delivery of viral particles throughout the body.

Motor activity tests. Muscle weakness and fatigability were measured with a rotarod apparatus (Ugo, Basile, Italy). Mice were first trained three times to be accommodated to the task. Mice were consecutively examined three times and were allowed to take a rest for 1 hour between individual tasks.

Running-wheel activity was used to quantify voluntary exercises. Each mouse was placed in a standard cage equipped with a counter-equipped running wheel (diameter, 14.7 cm, width, 5.2 cm; Ohara Medical, Tokyo, Japan). The running distances were recorded using the counter every 24 hours.

Histology. We raised a polyclonal ColQ antibody by injecting a synthetic peptide of SAALPSLDQKKRGGHKAC, corresponding to codons 34–51 in human ColQ, into rabbits. We confirmed that the raised antibody recognized ColQ by western blotting (**Supplementary Figure S3**) and that no signal was present in a section of *Colq*$^{-/-}$ mice (Figure 4b and **Supplementary Figure S2**). Mice were sacrificed at 6 weeks after treatment. Skeletal muscles of mice were frozen in the liquid nitrogen-cooled isopentane and sectioned at 8-μm thick with a Leica CW3050-4 cryostat at −20 °C. Muscle sections were blocked with 5% horse serum in phosphate-buffered saline for 20 minutes and incubated with the primary antibody (1:100) for 2 hours. Sections were then incubated with a secondary antibody (1:100) for 1 hour, along with Alexa-594-conjugated α-bungarotoxin (2.5 μg/ml) (Sigma, St Louis, MO) for visualizing AChR. Anti-rabbit and anti-mouse secondary antibodies were both FITC-labeled (Vector Lab, Burlingame, CA). For AAV1-*COLQ-IRES-EGFP*, we detected ColQ using anti-rabbit secondary antibody labeled with rhodamine (1:40; Santa Cruz, Santa Cruz, CA) and localized AChR by Alexa-647-conjugated α-bungarotoxin (2.5 μg/ml; Sigma). Signals of ColQ, AChE, AChR, and EGFP were examined with BX60 (Olympus, Tokyo, Japan) or BZ-9000 (Keyence, Osaka, Japan).

Mouse AChE activity was detected by the histochemical method at 6 weeks after treatment. Muscle sections were incubated for 20 minutes at 37 °C in the reaction mixture containing 1.73 mmol/l acetylthiocholine iodide, 38 mmol/l sodium acetate, 51 mmol/l acetic acid, 6 mmol/l sodium

citrate, 4.7 mmol/l copper sulfate, 0.5 mmol/l potassium ferricyanide, and 5 × 10^{-5} mol/l ethopropazine (Sigma), which is an inhibitor of butyrylcholinesterase.

Sedimentation biochemical analyses. Mice were sacrificed at 6 weeks after treatment. Sedimentation analysis was performed as previously described.[7] Proteins were extracted from the muscle and liver in a detergent buffer [10 mmol/l HEPES (pH 7.2), 1% CHAPS, 10 mmol/l EDTA, 2 mmol/l benzamidine, leupeptin (20 μg/ml) and pepstatin (10 μg/ml)] containing 0.8 mol/l NaCl. The eluate was applied on a 5–20% sucrose density gradient, which was made in the detergent buffer containing 0.8 mol/l NaCl, along with β-galactosidase (16S) and alkaline phosphatase (6.1S) as internal sedimentation standards. Centrifugation was performed in a Beckman SW41Ti rotor at 4°C for 21 hours at 38,000 r.p.m. AChE activity was assayed by the colorimetric method of Ellman in the presence of 5 × 10^{-5} mol/l ethopropazine.

For biochemical analysis, skeletal muscle was shattered by the Cool Mill (Toyobo, Osaka, Japan) in liquid nitrogen. We extracted globular forms of AChE into the NaCl-free detergent buffer, and ColQ-tailed AChE into detergent buffer containing 0.8 mol/l NaCl as previously described.[12] AChE activity was assayed using AChE-Specific Assay kit (Dojindo, Kumamoto, Japan) or the Ellman method and normalized by Torpedo AChE activity (Sigma).

Microelectrode studies. Phrenic nerve-diaphragm preparations were obtained from three wild type, three *Colq*$^{-/-}$, and three AAV8-*COLQ*-treated mice at 8 weeks of age, which corresponds to 4 weeks after treatment. We stimulated the sciatic nerve at 2 Hz and recorded compound muscle action potentials of gastrocnemius muscles using a needle electrode under deep anesthesia. For technical reasons, we could not analyze the limb muscles that we used in the other assays. After mice were sacrificed, miniature endplate potentials and evoked EPPs were recorded as described elsewhere.[43] Neostigmine methylsufate (Elkins-Sinn, Cherry Hill, NJ) was used at a concentration of 10^{-6} g/ml in the bath to block cholinesterases. We employed the AxoGraph × 1.1.6 (AxoGraph Scientific, Sydney, Australia) for data analysis.

Electron microscopy. For electron microscopy, extensor digitorum longus and soleus muscles were fixed in ice-cold 3% glutaraldehyde buffered with 0.1 mol/l cacodylate buffer (pH 7.3) at 4 weeks after treatment. The endplate-rich region of the muscle was refixed in 2% OsO_4 in cacodylate buffer, dehydrated, and embedded in Epon812.

All thin sections were cut transversely, stained with lead citrate, and photographed in a JEM 1,200 EX electron microscopy. Morphometric analysis of the motor endplate was performed following the procedure of Engel and Santa,[44] and included the following: (i) presynaptic membrane length, in μm; (ii) nerve terminal area, in μm^2; (iii) number of synaptic vesicles per unit area, in numbers per μm^2; (iv) length of processes of Schwann cells on presynaptic membrane, in μm; (v) percentage of totally enwrapped nerve terminal by processes of Schwann cells; (vi) postsynaptic area of folds and clefts associated with a given nerve terminal, in μm^2; (vii) postsynaptic membrane length associated with a given nerve terminal, in μm; (viii) postsynaptic membrane length per unit postsynaptic area (postsynaptic membrane density), derived by dividing the value of (vii) by that of (vi), in μm per μm^2; (ix) postsynaptic to presynaptic membrane ratio. Endplates were localized and analyzed by established methods, and peroxidase-labeled α-bungarotoxin was used for the ultrastructural localization of AChR.[45] The images were quantified using the NIH Image 1.62 software (National Institutes of Health).

Expression and purification of recombinant ColQ. The plasmids that previously introduced human *COLQ* and human *ACHE* cDNAs into a pTargeT (Promega, Madison, WI)[7] were cotransfected into HEK293 cells. Proteins were extracted from the cells in Tris–HCl buffer [50 mmol/l Tris–HCl (pH 7.0), 0.5% Triton X-100, 0.2 mmol/l EDTA, leupeptin (2 μg/ml), and pepstatin (1 μg/ml)] containing 1 mol/l NaCl. The extracts were loaded onto HiTrap Heparin HP columns (GE Healthcare, Buckinghamshire, UK). The concentration of purified recombinant ColQ-tailed AChE was equivalent to ~4 μg/ml Torpedo AChE. We injected 50 μl of the purified ColQ-tailed AChE in phosphate-buffered saline daily into the gluteus maximus muscles of 5-week-old *Colq*$^{-/-}$ mice for a week. Mice were given a single intraperitoneal injection of 300 mg/kg cyclophosphamide monohydrate (10 mg/ml in saline) at 24 hours after the first ColQ-tailed AChE injection to suppress immunoreaction against the recombinant human protein.[46] After 7 days, mice were sacrificed and brachial muscles were stained for ColQ molecule and AChE activity as described above.

Real-time PCR/reverse transcription-PCR. For expression analysis, total RNAs from skeletal muscle and liver cells were extracted using the RNeasy Mini kit (Qiagen, Hilden, Germany) with *DNaseI* and proteinase K treatment according to the manufacturer's instructions. First-stranded cDNA was synthesized using the ReverTra Ace reverse transcriptase (Toyobo). Expressions of human *COLQ*, mouse *Colq*, and mouse *Ache* were analyzed using the TaqMan (Applied Biosystems, Foster city, CA) probes and primers in LightCycler 480 (Roche, Mannheim, Germany). We also quantified 18S rRNA for normalization.

To quantify the transduction efficiency, total DNA was extracted from skeletal muscle and liver using the QIAamp DNA Mini Kit (Qiagen). The amount of viral genome was quantified by real-time PCR using a TaqMan probe targeting to human *COLQ*, as well as to mouse *Tert* encoding telomerase reverse transcriptase to normalize for the cell numbers.

SUPPLEMENTARY MATERIAL

Figure S1. Sedimentation analyses of AChE in the liver and serum.
Figure S2. Binding of human ColQ-tailed AChE proteins to the NMJ in muscle section of *Colq*$^{-/-}$ mice.
Figure S3. Western blot of a newly raised rabbit polyclonal anti-ColQ antibody (1:1,000).
Table S1. Morphometric analysis of endplate ultrastructures.
Video 1. First part: Two *Colq*$^{-/-}$ mice treated with an intravenous administration of 2 × 10^{12} vg of AAV8-*COLQ* (right cage) move around actively.

ACKNOWLEDGMENTS

We thank James M. Wilson for providing the chimeric helper plasmid pRC8 (identical to p5E18-VD2/8) and pRep2Cap1 (identical to p5E18RXCI). This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, and the Ministry of Health, Labor, and Welfare of Japan, as well as by Grant from ANR maladies rares. The authors declared no conflict of interest.

REFERENCES

1. Krejci, E, Thomine, S, Boschetti, N, Legay, C, Sketelj, J and Massoulié, J (1997). The mammalian gene of acetylcholinesterase-associated collagen. *J Biol Chem* **272**: 22840–22847.
2. Rotundo, RL (1984). Asymmetric acetylcholinesterase is assembled in the Golgi apparatus. *Proc Natl Acad Sci USA* **81**: 479–483.
3. Ruiz, CA and Rotundo, RL (2009). Limiting role of protein disulfide isomerase in the expression of collagen-tailed acetylcholinesterase forms in muscle. *J Biol Chem* **284**: 31753–31763.
4. Deprez, P, Inestrosa, NC and Krejci, E (2003). Two different heparin-binding domains in the triple-helical domain of ColQ, the collagen tail subunit of synaptic acetylcholinesterase. *J Biol Chem* **278**: 23233–23242.
5. Peng, HB, Xie, H, Rossi, SG and Rotundo, RL (1999). Acetylcholinesterase clustering at the neuromuscular junction involves perlecan and dystroglycan. *J Cell Biol* **145**: 911–921.
6. Cartaud, A, Strochlic, L, Guerra, M, Blanchard, B, Lambergeon, M, Krejci, E *et al.* (2004). MuSK is required for anchoring acetylcholinesterase at the neuromuscular junction. *J Cell Biol* **165**: 505–515.
7. Ohno, K, Brengman, J, Tsujino, A and Engel, AG (1998). Human endplate acetylcholinesterase deficiency caused by mutations in the collagen-like tail subunit (ColQ) of the asymmetric enzyme. *Proc Natl Acad Sci USA* **95**: 9654–9659.
8. Donger, C, Krejci, E, Serradell, AP, Eymard, B, Bon, S, Nicole, S *et al.* (1998). Mutation in the human acetylcholinesterase-associated collagen gene, COLQ, is responsible

for congenital myasthenic syndrome with end-plate acetylcholinesterase deficiency (Type Ic). *Am J Hum Genet* **63**: 967–975.
9. Ohno, K, Engel, AG, Brengman, JM, Shen, XM, Heidenreich, F, Vincent, A *et al.* (2000). The spectrum of mutations causing end-plate acetylcholinesterase deficiency. *Ann Neurol* **47**: 162–170.
10. Bestue-Cardiel, M, Sáenz de Cabezón-Alvarez, A, Capablo-Liesa, JL, López-Pisón, J, Peña-Segura, JL, Martin-Martinez, J *et al.* (2005). Congenital endplate acetylcholinesterase deficiency responsive to ephedrine. *Neurology* **65**: 144–146.
11. Mihaylova, V, Müller, JS, Vilchez, JJ, Salih, MA, Kabiraj, MM, D'Amico, A *et al.* (2008). Clinical and molecular genetic findings in COLQ-mutant congenital myasthenic syndromes. *Brain* **131**(Pt 3): 747–759.
12. Feng, G, Krejci, E, Molgo, J, Cunningham, JM, Massoulié, J and Sanes, JR (1999). Genetic analysis of collagen Q: roles in acetylcholinesterase and butyrylcholinesterase assembly and in synaptic structure and function. *J Cell Biol* **144**: 1349–1360.
13. Lee, HH, Choi, RC, Ting, AK, Siow, NL, Jiang, JX, Massoulié, J *et al.* (2004). Transcriptional regulation of acetylcholinesterase-associated collagen ColQ: differential expression in fast and slow twitch muscle fibers is driven by distinct promoters. *J Biol Chem* **279**: 27098–27107.
14. Ruiz, CA and Rotundo, RL (2009). Dissociation of transcription, translation, and assembly of collagen-tailed acetylcholinesterase in skeletal muscle. *J Biol Chem* **284**: 21488–21495.
15. Inagaki, K, Fuess, S, Storm, TA, Gibson, GA, Mctiernan, CF, Kay, MA *et al.* (2006). Robust systemic transduction with AAV9 vectors in mice: efficient global cardiac gene transfer superior to that of AAV8. *Mol Ther* **14**: 45–53.
16. Nakai, H, Fuess, S, Storm, TA, Muramatsu, S, Nara, Y and Kay, MA (2005). Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice. *J Virol* **79**: 214–224.
17. Bouma, SR, Drislane, FW and Huestis, WH (1977). Selective extraction of membrane-bound proteins by phospholipid vesicles. *J Biol Chem* **252**: 6759–6763.
18. Kimbell, LM, Ohno, K, Engel, AG and Rotundo, RL (2004). C-terminal and heparin-binding domains of collagenic tail subunit are both essential for anchoring acetylcholinesterase at the synapse. *J Biol Chem* **279**: 10997–11005.
19. Wang, Z, Zhu, T, Qiao, C, Zhou, L, Wang, B, Zhang, J *et al.* (2005). Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. *Nat Biotechnol* **23**: 321–328.
20. Schnepp, BC, Clark, KR, Klemanski, DL, Pacak, CA and Johnson, PR (2003). Genetic fate of recombinant adeno-associated virus vector genomes in muscle. *J Virol* **77**: 3495–3504.
21. Kay, MA (2007). AAV vectors and tumorigenicity. *Nat Biotechnol* **25**: 1111–1113.
22. Xiao, X, Li, J and Samulski, RJ (1996). Efficient long-term gene transfer into muscle tissue of immunocompetent mice by adeno-associated virus vector. *J Virol* **70**: 8098–8108.
23. Rivière, C, Danos, O and Douar, AM (2006). Long-term expression and repeated administration of AAV type 1, 2 and 5 vectors in skeletal muscle of immunocompetent adult mice. *Gene Ther* **13**: 1300–1308.
24. Krejci, E, Legay, C, Thomine, S, Sketelj, J and Massoulié, J (1999). Differences in expression of acetylcholinesterase and collagen Q control the distribution and oligomerization of the collagen-tailed forms in fast and slow muscles. *J Neurosci* **19**: 10672–10679.
25. Lau, FT, Choi, RC, Xie, HQ, Leung, KW, Chen, VP, Zhu, JT *et al.* (2008). Myocyte enhancer factor 2 mediates acetylcholine-induced expression of acetylcholinesterase-associated collagen ColQ in cultured myotubes. *Mol Cell Neurosci* **39**: 429–438.
26. Arikawa-Hirasawa, E, Rossi, SG, Rotundo, RL and Yamada, Y (2002). Absence of acetylcholinesterase at the neuromuscular junctions of perlecan-null mice. *Nat Neurosci* **5**: 119–123.
27. Stum, M, Girard, E, Bangratz, M, Bernard, V, Herbin, M, Vignaud, A *et al.* (2008). Evidence of a dosage effect and a physiological endplate acetylcholinesterase deficiency in the first mouse models mimicking Schwartz-Jampel syndrome neuromyotonia. *Hum Mol Genet* **17**: 3166–3179.
28. Rotundo, RL, Rossi, SG and Anglister, L (1997). Transplantation of quail collagen-tailed acetylcholinesterase molecules onto the frog neuromuscular synapse. *J Cell Biol* **136**: 367–374.
29. Moretti, FA, Chauhan, AK, Iaconcig, A, Porro, F, Baralle, FE and Muro, AF (2007). A major fraction of fibronectin present in the extracellular matrix of tissues is plasma-derived. *J Biol Chem* **282**: 28057–28062.
30. Rooney, JE, Gurpur, PB and Burkin, DJ (2009). Laminin-111 protein therapy prevents muscle disease in the mdx mouse model for Duchenne muscular dystrophy. *Proc Natl Acad Sci USA* **106**: 7991–7996.
31. Hall, ZW and Ralston, E (1989). Nuclear domains in muscle cells. *Cell* **59**: 771–772.
32. Rossi, SG, Vazquez, AE and Rotundo, RL (2000). Local control of acetylcholinesterase gene expression in multinucleated skeletal muscle fibers: individual nuclei respond to signals from the overlying plasma membrane. *J Neurosci* **20**: 919–928.
33. Miner, JH, Go, G, Cunningham, J, Patton, BL and Jarad, G (2006). Transgenic isolation of skeletal muscle and kidney defects in laminin beta2 mutant mice: implications for Pierson syndrome. *Development* **133**: 967–975.
34. Somia, N and Verma, IM (2000). Gene therapy: trials and tribulations. *Nat Rev Genet* **1**: 91–99.
35. Mueller, C and Flotte, TR (2008). Clinical gene therapy using recombinant adeno-associated virus vectors. *Gene Ther* **15**: 858–863.
36. Helbling-Leclerc, A, Zhang, X, Topaloglu, H, Cruaud, C, Tesson, F, Weissenbach, J *et al.* (1995). Mutations in the laminin alpha 2-chain gene (LAMA2) cause merosin-deficient congenital muscular dystrophy. *Nat Genet* **11**: 216–218.
37. Nicole, S, Davoine, CS, Topaloglu, H, Cattolico, L, Barral, D, Beighton, P *et al.* (2000). Perlecan, the major proteoglycan of basement membranes, is altered in patients with Schwartz-Jampel syndrome (chondrodystrophic myotonia). *Nat Genet* **26**: 480–483.
38. Kawahara, G, Okada, M, Morone, N, Ibarra, CA, Nonaka, I, Noguchi, S *et al.* (2007). Reduced cell anchorage may cause sarcolemma-specific collagen VI deficiency in Ullrich disease. *Neurology* **69**: 1043–1049.
39. Okada, T, Nomoto, T, Yoshioka, T, Nonaka-Sarukawa, M, Ito, T, Ogura, T *et al.* (2005). Large-scale production of recombinant viruses by use of a large culture vessel with active gassing. *Hum Gene Ther* **16**: 1212–1218.
40. Okada, T, Shimazaki, K, Nomoto, T, Matsushita, T, Mizukami, H, Urabe, M *et al.* (2002). Adeno-associated viral vector-mediated gene therapy of ischemia-induced neuronal death. *Meth Enzymol* **346**: 378–393.
41. Okada, T, Nonaka-Sarukawa, M, Uchibori, R, Kinoshita, K, Hayashita-Kinoh, H, Nitahara-Kasahara, Y *et al.* (2009). Scalable purification of adeno-associated virus serotype 1 (AAV1) and AAV8 vectors, using dual ion-exchange adsorptive membranes. *Hum Gene Ther* **20**: 1013–1021.
42. Rohr, UP, Wulf, MA, Stahn, S, Steidl, U, Haas, R and Kronenwett, R (2002). Fast and reliable titration of recombinant adeno-associated virus type-2 using quantitative real-time PCR. *J Virol Methods* **106**: 81–88.
43. Engel, AG, Nagel, A, Walls, TJ, Harper, CM and Waisburg, HA (1993). Congenital myasthenic syndromes: I. Deficiency and short open-time of the acetylcholine receptor. *Muscle Nerve* **16**: 1284–1292.
44. Engel, AG and Santa, T (1971). Histometric analysis of the ultrastructure of the neuromuscular junction in myasthenia gravis and in the myasthenic syndrome. *Ann N Y Acad Sci* **183**: 46–63.
45. Engel, AG, Lindstrom, JM, Lambert, EH and Lennon, VA (1977). Ultrastructural localization of the acetylcholine receptor in myasthenia gravis and in its experimental autoimmune model. *Neurology* **27**: 307–315.
46. Otterness, IG and Chang, YH (1976). Comparative study of cyclophosphamide, 6-mercaptopurine, azathiopurine and methotrexate. Relative effects on the humoral and the cellular immune response in the mouse. *Clin Exp Immunol* **26**: 346–354.

Journal of Human Genetics (2012) 00, 1–7

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

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], Koo 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 advance online publication, 19 April 2012; doi:10.1038/jhg.2012.37

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

[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

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.

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.

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

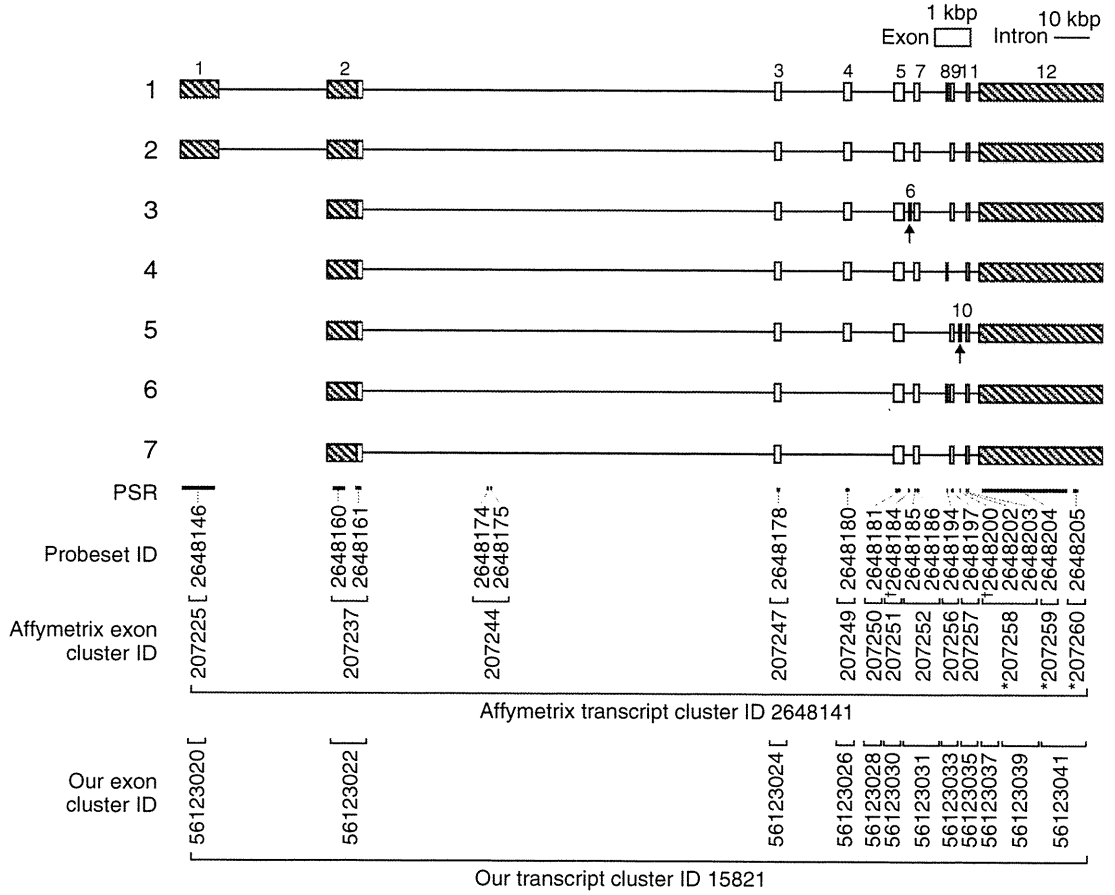

Figure 1 Comparison of Affymetrix annotations and our annotations. Seven alternative transcripts of *MBNL1* according to the NCBI Build 36.3 are drawn to individual exonic and intronic scales indicated at the top. The PSR bars represent the 'core' probe sets by Affymetrix. Each probe set is comprised 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.

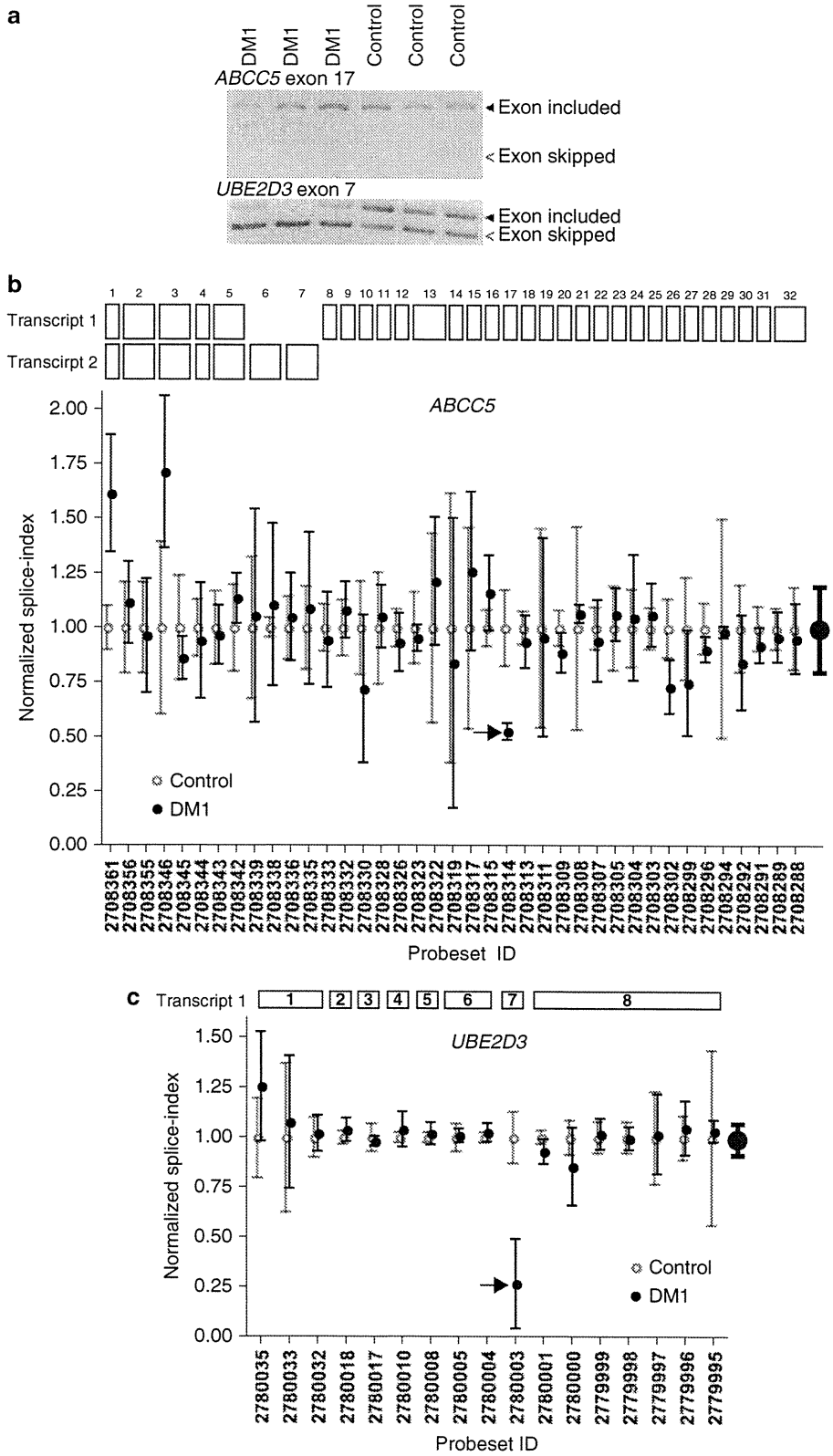

Figure 2 Falsely predicted skipping of *ABCC5* exon 17, and correctly predicted skipping of *UBE2D3* exon 7. (**a**) RT-PCR analysis of three DM1 muscles and three control muscles. (**b**) Splice index (an exonic signals divided by a sum of all the exonic signals throughout a gene) of each probe set 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).

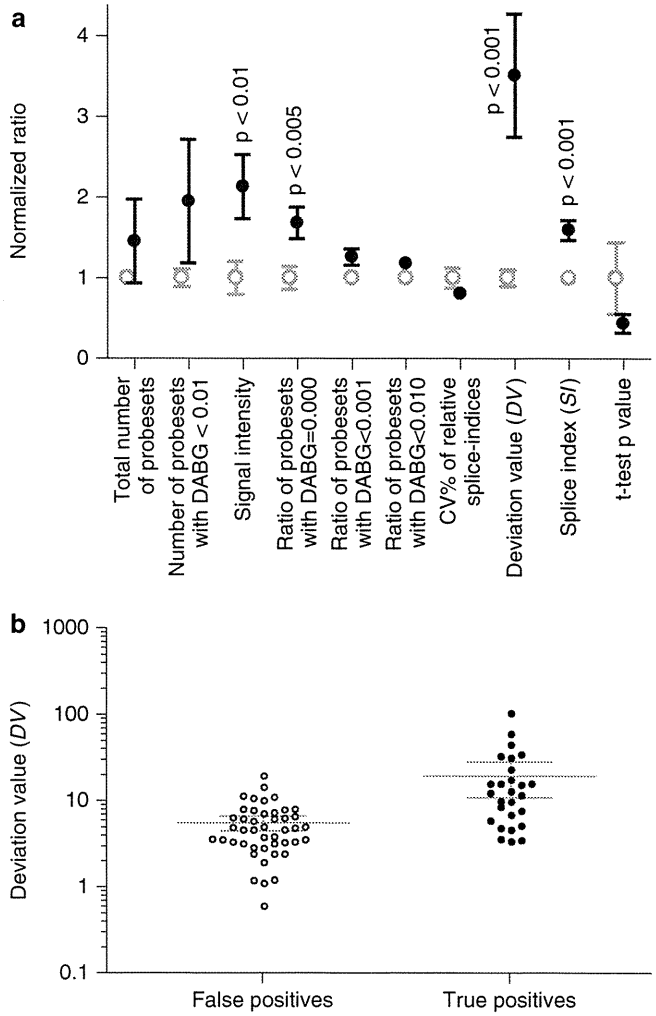

Figure 4 (**a**) Parameters that differentiate the true and false positives. Values are normalized to those of false positives, and the mean and s.e. are indicated. The true and false positives are indicated by solid and gray symbols, respectively. For *SI*, an inverse of the *SI* value is taken when 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

with dependable DABG *P*-values. Inclusion of these probe sets increases the standard deviation and decreases the *DV* of the aberrantly spliced exon.

Potential roles of novel aberrant splicing events in DM1

In this study, we identified 27 DM1-specific aberrant splicing, in which 25 have not been published yet. Among the 25 exons, aberrant splicing events of two exons were 'uniquely' observed in DM1: one is inclusion of the LIM domain binding 3 (*LDB3*) exon 4 and the other is inclusion of titin (*TTN*) exon 45. Interestingly, both encode structural proteins of muscle fiber.

LDB3, also known as Cypher/ZASP (Z-band alternatively spliced PDZ-motif protein), contains a PDZ domain at the N-terminus and one or three LIM domains at the C-terminus. LDB3 is localized to the Z-line and interacts with α-actinin 2 through its PDZ-domain and with protein kinase C via its C-terminal LIM domains.[28] LDB3 is likely to have an essential role in supporting Z-line structure and muscle function during contraction.[29] *LDB3* has several isoforms. As inclusion of exon 4 is preferentially observed in the fetal heart,[30] the aberrant inclusion of exon 4 in the skeletal muscles in DM1 would lead to dysfunction or morphological abnormalities of muscle fiber. Recently, phosphoglucomutase 1 (PGM1), an enzyme involved in glycolysis and gluconeogenesis, has been known to bind to the domain encoded by exon 4 of *LDB3*. *LDB3* mutations in exon 4 reduce the binding to PGM1 and develop dilated cardiomyopathy.[31] On the other hand, the increased binding of PGM1 and LDB3 through aberrant inclusion of exon 4 might be involved in the pathogenesis of muscle atrophy, weakness and histological abnormalities in DM1.

TTN encodes the largest protein in mammals and the third most abundant protein in muscle.[32] An N-terminal Z-disc region and a C-terminal M-line region bind to the Z-line and M-line of the sarcomere, respectively, so that a single molecule extends half the length of a sarcomere. Titin is critically important for myofibril elasticity and structural integrity. Its elasticity lies specifically in the I-band region and contains two elements in series with different properties: the tandem immunoglobulin (Ig) and PEVK domains.[33] Different *TTN* isoforms contribute to differences in elasticity of different muscle types.[34] As exon 45 is located at the tandem Ig domains, aberrant inclusion of exon 45 in DM1 might lead to defective myofibril assembly and function.

ACKNOWLEDGEMENTS

This work was supported by Grants-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology as well as the Ministry of Health, Labor and Welfare of Japan.

1 Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456,** 470–476 (2008).
2 Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40,** 1413–1415 (2008).
3 Smith, C. W. & Valcarcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25,** 381–388 (2000).
4 Caceres, J. F. & Kornblihtt, A. R. Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* **18,** 186–193 (2002).
5 O'Rourke, J. R. & Swanson, M. S. Mechanisms of RNA-mediated disease. *J. Biol. Chem.* **284,** 7419–7423 (2009).
6 Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136,** 777–793 (2009).
7 Harper, P. S. & Monckton, D. G. in *Myology* 3rd edn (ed. Engel, A. G.) Vol. 2, 1039–1076 (McGraw-Hill, New York, 2004).
8 Aslanidis, C., Jansen, G., Amemiya, C., Shutler, G., Mahadevan, M., Tsilfidis, C. *et al.* Cloning of the essential myotonic dystrophy region and mapping of the putative defect. *Nature* **355,** 548–551 (1992).
9 Brook, J. D., McCurrach, M. E., Harley, H. G., Buckler, A. J., Church, D., Aburatani, H. *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).
10 Buxton, J., Shelbourne, P., Davies, J., Jones, C., Van Tongeren, T., Aslanidis, C. *et al.* Detection of an unstable fragment of DNA specific to individuals with myotonic dystrophy. *Nature* **355,** 547–548 (1992).
11 Liquori, C. L., Ricker, K., Moseley, M. L., Jacobsen, J. F., Kress, W., Naylor, S. L. *et al.* Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* **293,** 864–867 (2001).
12 Gharehbaghi-Schnell, E. B., Finsterer, J., Korschineck, I., Mamoli, B. & Binder, B. R. Genotype-phenotype correlation in myotonic dystrophy. *Clin. Genet.* **53,** 20–26 (1998).
13 Consortium T. I. M. D. New nomenclature and DNA testing guidelines for myotonic dystrophy type 1 (DM1). *Neurology* **54,** 1218–1221 (2000).
14 Saito, T., Amakusa, Y., Kimura, T., Yahara, O., Aizawa, H., Ikeda, Y. *et al.* Myotonic dystrophy type 2 in Japan: ancestral origin distinct from Caucasian families. *Neurogenetics* **9,** 61–63 (2008).
15 Osborne, R. J., Lin, X., Welle, S., Sobczak, K., O'Rourke, J. R., Swanson, M. S. *et al.* Transcriptional and post-transcriptional impact of toxic RNA in myotonic dystrophy. *Hum. Mol. Genet.* **18,** 1471–1481 (2009).
16 Wang, G. S., Kearney, D. L., De Biasi, M., Taffet, G. & Cooper, T. A. Elevation of RNA-binding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *J. Clin. Invest.* **117,** 2802–2811 (2007).
17 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).
18 Nezu, Y., Kino, Y., Sasagawa, N., Nishino, I. & Ishiura, S. Expression of MBNL and CELF mRNA transcripts in muscles with myotonic dystrophy. *Neuromuscul. Disord.* **17,** 306–312 (2007).
19 Thornton, C. A., Johnson, K. & Moxley, 3rd R. T. Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. *Ann. Neurol.* **35,** 104–107 (1994).
20 Srinivasan, K., Shiue, L., Hayes, J. D., Centers, R., Fitzwater, S., Loewen, R. *et al.* Detection and measurement of alternative splicing using splicing-sensitive microarrays. *Methods* **37,** 345–359 (2005).
21 Lin, X., Miller, J. W., Mankodi, A., Kanadia, R. N., Yuan, Y., Moxley, R. T. *et al.* Failure of MBNL1-dependent post-natal splicing transitions in myotonic dystrophy. *Hum. Mol. Genet.* **15,** 2087–2097 (2006).
22 Orengo, J. P., Ward, A. J. & Cooper, T. A. Alternative splicing dysregulation secondary to skeletal muscle regeneration. *Ann. Neurol.* **69,** 681–690 (2011).
23 Fugier, C., Klein, A. F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A. *et al.* Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nat. Med.* **17,** 720–725 (2011).
24 Sharma, A., Markey, M., Torres-Munoz, K., Varia, S., Kadakia, M., Bubulya, A. *et al.* Son maintains accurate splicing for a subset of human pre-mRNAs. *J. Cell Sci.* **124,** 4286–4298 (2011).
25 Chang, J. G., Yang, D. M., Chang, W. H., Chow, L. P., Chan, W. L., Lin, H. H. *et al.* Small molecule amiloride modulates oncogenic RNA alternative splicing to devitalize human cancer cells. *PloS One* **6,** e18643 (2011).
26 Dutertre, M., Sanchez, G., De Cian, M. C., Barbier, J., Dardenne, E., Gratadou, L. *et al.* Cotranscriptional exon skipping in the genotoxic stress response. *Nat. Struct. Mol. Biol.* **17,** 1358–1366 (2010).
27 Kalsotra, A., Xiao, X., Ward, A. J., Castle, J. C., Johnson, J. M., Burge, C. B. *et al.* A postnatal switch of CELF and MBNL proteins reprograms alternative splicing in the developing heart. *Proc. Natl Acad. Sci. USA* **105,** 20333–20338 (2008).
28 Zhou, Q., Ruiz-Lozano, P., Martone, M. E. & Chen, J. Cypher, a striated muscle-restricted PDZ and LIM domain-containing protein, binds to alpha-actinin-2 and protein kinase C. *J. Biol. Chem.* **274,** 19807–19813 (1999).
29 Zhou, Q., Chu, P. H., Huang, C., Cheng, C. F., Martone, M. E., Knoll, G. *et al.* Ablation of Cypher, a PDZ-LIM domain Z-line protein, causes a severe form of congenital myopathy. *J. Cell. Biol.* **155,** 605–612 (2001).
30 Huang, C., Zhou, Q., Liang, P., Hollander, M. S., Sheikh, F., Li, X. *et al.* Characterization and *in vivo* functional analysis of splice variants of cypher. *J. Biol. Chem.* **278,** 7360–7365 (2003).
31 Arimura, T., Inagaki, N., Hayashi, T., Shichi, D., Sato, A., Hinohara, K. *et al.* Impaired binding of ZASP/Cypher with phosphoglucomutase 1 is associated with dilated cardiomyopathy. *Cardiovasc. Res.* **83,** 80–88 (2009).
32 Gregorio, C. C., Granzier, H., Sorimachi, H. & Labeit, S. Muscle assembly: a titanic achievement? *Curr. Opin. Cell Biol.* **11,** 18–25 (1999).
33 Horowits, R. The physiological role of titin in striated muscle. *Rev. Physiol. Biochem. Pharmacol.* **138,** 57–96 (1999).
34 Freiburg, A., Trombitas, K., Hell, W., Cazorla, O., Fougerousse, F., Centner, T. *et al.* Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. *Circ. Res.* **86,** 1114–1121 (2000).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)

Epilepsia, 52(7):e66–e69, 2011
doi: 10.1111/j.1528-1167.2011.03139.x

BRIEF COMMUNICATION

Refractory neonatal epilepsy with a de novo duplication of chromosome 2q24.2q24.3

***Akihisa Okumura, †Toshiyuki Yamamoto, †Keiko Shimojima, ‡Yoshinobu Honda, *Shinpei Abe, *Mitsuru Ikeno, and *Toshiaki Shimizu**

***Department of Pediatrics, Juntendo University Faculty of Medicine, Tokyo, Japan; †Tokyo Women's Medical University Institute for Integrated Medical Sciences, Tokyo, Japan; and ‡Department of Premature and Neonatal Medicine, Iwaki Kyoritsu General Hospital, Iwaki, Japan**

Summary

There are only two reports on epileptic patients associated with microduplication of 2q. We found a de novo duplication of chromosome 2q24.2q24.3 in another infant with neonatal epilepsy. The patient had refractory focal seizures since the third day of life. Her seizures were refractory against phenobarbital and levetiracetam, but were controlled by valproate. Array comparative genomic hybridization revealed a 5.3-Mb duplication of 2q24.2q24.3, where at least 22 genes including a cluster of voltage-gated sodium channel genes (*SCN1A*, *SCN2A*, *SCN3A*, *SCN7A*, and *SCN9A*) and one noncoding RNA are located.

KEY WORDS: Neonatal epilepsy, Voltage-gated sodium channels, 2q duplication, Array comparative genomic hybridization.

Chromosomal deletions of 2q21-q31 have been known to be closely related to seizures and epilepsies, and the band 2q24 is the smallest commonly deleted segment in these patients (Davidsson et al., 2008). Several authors including us have reported that deletions of the voltage-gated sodium channel (SCN) cluster at 2q24 are associated with several epilepsy syndromes (Takatsuki et al., 2010). The deletions of the locus, including *SCN1A*, were found in patients with Dravet syndrome (Marini et al., 2009; Suls et al., 2010). The chromosomal deletions of 2q including *SCN1A* and *SCN2A* were reported in some patients with severe epilepsy of infantile onset, developmental delay, and dysmorphic features (Pereira et al., 2004; Langer et al., 2006; Pereira et al., 2006). A small deletion between the locus of *SCN2A* and *SCN3A* was seen in a patient with infantile seizures, mental retardation, and behavioral and psychiatric abnormalities (Bartnik et al., 2010). Compared to that, the duplications of this region have been identified in only two families (Heron et al., 2010; Raymond et al., 2011). Recently, we identified a chromosomal duplication of 2q24.2q24.3, in which SCN gene cluster is located, in an infant with refractory neonatal epilepsy and severe developmental delay. We report on this patient and discuss the genotype–phenotype correlation.

Accepted April 28, 2011; Early View publication June 22, 2011.
Address correspondence to Akihisa Okumura, Department of Pediatrics, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan. E-mail: okumura@juntendo.ac.jp

Patient Report

A female infant was born spontaneously at 40 weeks of gestation following an uncomplicated pregnancy. She was the first product of unrelated healthy parents. Her birth weight was 2,796 g and head circumference was 33 cm. No dysmorphic features were recognized.

Her mother had noticed mild convulsive movement and staring lasting for 10–20 s since the third day of life. The frequency of paroxysmal events gradually increased and she was admitted to the neonatal intensive care unit of Iwaki Kyoritsu Hospital at 9 days of age. Physical examinations were unremarkable other than for mild hypotonia. Head magnetic resonance imaging (MRI) and blood examination including metabolic screening of amino acids and organic acid analyses revealed no abnormalities. Interictal electroencephalography (EEG) showed markedly abnormal background activities with spiky transients. She was diagnosed as having neonatal seizures and treated with phenobarbital and midazolam. Although her seizures were refractory against these antiepileptic drugs, seizures were transiently controlled by 20 mg/kg of oral phenobarbital after 23 days of age.

Her seizures recurred at 2 months of age and she was referred to Juntendo University Hospital. She had 5–20 seizures every day. Although the dose of phenobarbital was increased to reach the serum level of 61.0 μg/ml, her seizures could not be stopped. Next, levetiracetam was added up to the dose of 40 mg/kg. However, the frequency of seizures did not decrease.