Journal of Human Genetics (2013) 58, 33–39

www.nature.com/jhg

ORIGINAL ARTICLE

Molecular characterization of an X(p21.2;q28) chromosomal inversion in a Duchenne muscular dystrophy patient with mental retardation reveals a novel long non-coding gene on Xq28

Thi Hoai Thu Tran[1,3,4], Zhujun Zhang[1,3,5], Mariko Yagi[1], Tomoko Lee[1], Hiroyuki Awano[1], Atsushi Nishida[1,6], Takeshi Okinaga[2], Yasuhiro Takeshima[1] and Masafumi Matsuo[1,6]

Duchenne muscular dystrophy (DMD) is the most common inherited muscular disease and is characterized by progressive muscle wasting. DMD is caused by mutations in the *dystrophin* gene on Xp21.2. One-third of DMD cases are complicated by mental retardation, but the pathogenesis of this is unknown. We have identified an intrachromosomal inversion, inv(X)(p21.2;q28) in a DMD patient with mental retardation. We hypothesized that a gene responsible for the mental retardation in this patient would be disrupted by the inversion. We localized the inversion break point by analysis of *dystrophin* complementary DNA (cDNA) and fluorescence *in situ* hybridization. We used 5′ and 3′ rapid amplification of cDNA ends to extend the known transcripts, and reverse transcription-PCR to analyze tissue-specific expression. The patient's *dystrophin* cDNA was separated into two fragments between exons 18 and 19. Exon 19 was dislocated to the long arm of the X-chromosome. We identified a novel 109-bp sequence transcribed upstream of exon 19, and a 576-bp sequence including a poly(A) tract transcribed downstream of exon 18. Combining the two novel sequences, we identified a novel gene, named *KUCG1*, which comprises three exons spanning 50 kb on Xq28. The 685-bp transcript has no open-reading frame, classifying it as a long non-coding RNA. *KUCG1* mRNA was identified in brain. We cloned a novel long non-coding gene from a chromosomal break point. It was supposed that this gene may have a role in causing mental retardation in the index case.
Journal of Human Genetics (2013) **58,** 33–39; doi:10.1038/jhg.2012.131; published online 6 December 2012

Keywords: dystrophin; long non-coding gene; mental retardation

INTRODUCTION

Duchenne muscular dystrophy (DMD) is the most common inherited muscle disease affecting approximately one in 3500 males and is characterized by progressive muscle wasting during childhood. DMD shows muscle dystrophin deficiency because of mutations in the *dystrophin* gene that comprises 79 exons spanning >2500 kb on chromosome Xp21.2.[1] Mutations in the *dystrophin* gene range from single-nucleotide changes to chromosomal abnormalities (http://www.dmd.nl/).[2] Deletions encompassing one or more exons of the *dystrophin* gene are the most common cause of DMD and account for ~60% of mutations.[3] Disastrous mutations such as an out-of-frame deletion or nonsense mutation result in severe DMD.[4] DMD is complicated by mental retardation in one-third of patients.[5] Many studies have been conducted to elucidate the pathogenic mechanism of this complicating mental retardation. There are now several reports describing that mutations at the 3′ end of the *dystrophin* gene are related to complication with mental retardation.[6,7]

In a small portion of DMD patients, gross chromosomal rearrangements have been reported as the cause of dystrophin deficiency. In fact, a huge intrachromosomal deletion showing contiguous gene deletion syndrome was used to clone the *dystrophin* gene.[8] Intrachromosomal inversions have been identified in DMD.[9,10] X-autosome translocations involving the *dystrophin* gene have also been identified in a limited number of DMD patients.[11,12]

Disease-associated chromosomal rearrangements have been frequently used as a starting point in the elucidation of congenital

[1]Department of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan and [2]Department of Developmental Medicine (Pediatrics), Osaka University Graduate School of Medicine, Osaka, Japan
[3]These authors contributed equally to this work.
[4]Current address: Department of Pediatrics, Medical University of Pham Ngoc Thach, Ho Chi Minh City, Vietnam.
[5]Current address: Department of Pathology, Nankai University School of Medicine, Nankai University, Tianjin, People's Republic of China.
[6]Current address: Department of Medical Rehabilitation, Faculty of Rehabilitation, Kobegakuin University, Kobe, Japan.
Correspondence: Professor M Matsuo, Department of Medical Rehabilitation, Faculty of Rehabilitation, Kobegakuin University, 518 Arise, Ikawadani-cho, Nishi-ku, Kobe, Hyogo 6512180, Japan.
E-mail: mmatsuo@reha.kobegakuin.ac.jp or matsuo@kobe-u.ac.jp
Received 18 August 2012; revised 26 September 2012; accepted 10 October 2012; published online 6 December 2012

disorders. Disrupted X-chromosomal genes are even more promising in this respect as they often represent knockouts.[10,13] In one DMD patient with complicating mental retardation, for example, an intrachromosomal inversion led to the identification of a Ras-like GTPase gene that causes mental retardation.[9] In addition, >20 genes have been identified by studying balanced X-chromosome rearrangements.[14]

The genes for X-linked mental retardation are largely unknown.[14–16] In a series of 442 Japanese mutations in the *dystrophin* gene, we have described a karyotype of 46,Y,inv (X)(p21.2;q28) to be the cause of one case of DMD.[2] This case was complicated with moderate mental retardation and it is thought very likely that the inversion disrupts one of the >40 genes responsible for mental retardation at Xq28.[17]

A diverse population of non-protein-coding RNAs has been reported in the human genome.[18,19] Long non-coding RNAs (lncRNAs), defined as greater than 200 nucleotides (nt) in length,[20] have a wide range of functions, including the regulation of transcription, RNA editing and organelle biogenesis.[19,21,22] It has been suggested that a subset of lncRNAs could contribute to neurological disorders when they become dysregulated.[23]

In this study, we characterized an intrachromosomal inversion inv(X)(p21.2;q28). We identified a novel long non-coding gene named *KUCG1* at the break point on Xq28. As this gene was expressed in the brain, we propose that disruption of the *KUCG1* gene may have a role in causing the mental retardation in the index case.

MATERIALS AND METHODS

Patient

The index patient is a 3-year-old Japanese boy. He is the first child of healthy, non-consanguineous, Japanese parents. Family history was unremarkable. When he was born at term, blood sampling was performed because of birth asphyxia. Unexpectedly, his serum creatine kinase level was highly elevated (25 510 $IU l^{-1}$; normal: <270 $IU l^{-1}$). He walked unassisted at the age of 15 months. As the high creatine kinase level persisted, a muscle biopsy was conducted at the age of 2 years to examine dystrophin expression. Dystrophin staining using monoclonal antibodies to three different domains revealed no reactive material in his skeletal muscle, confirming the diagnosis of DMD. He was referred to Kobe University Hospital for a genetic diagnosis (KUCG481). At the age of 3 years, his serum creatine kinase was 21 776 $IU l^{-1}$. His growth parameters were normal but he displayed moderate mental retardation (developmental quotient: 40). Brain magnetic resonance imaging findings were normal. His karyotype has been described in our previous report as 46,Y, inv(X)(p21.2;q28).[2] The inversion was inherited through his mother (data not shown). The protocol for the following study was approved by the ethical committee of Kobe University School of Medicine.

Dystrophin mRNA analysis

RNA was isolated from biopsied skeletal muscle and analyzed by reverse transcription-PCR as described previously.[24,25] The full-length *dystrophin* complementary DNA (cDNA) was amplified as 10 separate fragments.[26] To identify the break point within the *dystrophin* cDNA, fragments encompassing exons 18 and 19 were amplified using different sets of primers. The ends of two separate *dystrophin* cDNAs were confirmed by PCR amplification using newly designed primers; a reverse primer on exon 18 and a forward primer on exon 19, respectively (Table 1).

PCR amplification

PCR amplification was performed in a total volume of 20 μl, containing 2 μl of cDNA, 2 μl of 10× ExTaq buffer (Takara Bio, Inc., Shiga, Japan), 0.5 U of ExTaq polymerase (Takara Bio, Inc.), 500 nM of each primer and 250 μM deoxyribonucleotide triphosphates (Takara Bio, Inc.). Thirty-five cycles of amplification were performed on a Mastercycler Gradient PCR machine (Eppendorf, Hamburg, Germany) using the following conditions: initial denaturation at 94 °C for 5 min, subsequent denaturation at 94 °C for 0.5 min, annealing at 59 °C for 0.5 min and extension at 72 °C for 1 min. The conditions were sometimes slightly modified for optimization. For nested or semi-nested PCR, 2 μl of the first reaction mixture was used as the template for the second amplification. The amplified PCR products were electrophoresed on 2% agarose gels with a low-molecular weight DNA standard (ϕX174-Hae III digest; Takara Bio, Inc.) and stained with ethidium bromide.

Table 1 Primers used in this study

Primer name	*Primer sequence (5′-3′)*
Dystrophin cDNA	
Exon 18r	GCAGAGTCCTGAATTTGCAATC
Exon 19f	CATTCACCATCTGTTCCACCA
5′-RACE	
c24r	CAGCCATCCATTTCTTCAGG
c21r	TTGTCTGTAGCTCTTTCTCT
c20r	ACTGGCAGAATTCGATCCAC
3′-RACE	
c16f	CTGATCTAGAGGTACCGGATCC
c18f	GCAGAGTCCTGAATTTGCAATC
KUCG1 mRNA	
Bf	GGTGAACCCCTCAATGTAAG
Cr	CTCTTGTATTCGCTGCAGTG
Cr2	CAGCAAACTTGTACAGTTGC

Abbreviations: cDNA, complementary DNA; RACE, rapid amplification of cDNA ends.

Fluorescence *in situ* hybridization

Fluorescence *in situ* hybridization was conducted on metaphase spreads from the patients' lymphocytes with digoxigenin-labeled PCR product containing exons 18 or 19 of the *dystrophin* gene in combination with DXZ1 spectrum green probe for the X centromere (Vysis, Inc., Downers Grove, IL, USA). The exon 18 and 19 probes were detected by immunocytochemistry. This assay was carried out commercially by Mitsubishi Chemical Medience Co. (Tokyo, Japan).

5′-Rapid amplification of cDNA ends

5′-Rapid amplification of cDNA ends (RACE) was performed to obtain the 5′-end of the transcript using the 5′-RACE System Version 2 (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions, with primers specific for the *dystrophin* mRNA (Table 1). Total RNA isolated from the patient's skeletal muscle was reverse transcribed using a gene-specific primer (c24r) and SuperScript II, a derivative of Moloney Murine Leukemia Virus Reverse Transcriptase (Invitrogen). PCR amplification was then performed using Taq DNA polymerase (Takara Bio, Inc.), a nested gene-specific primer (c21r), and a deoxyinosine-containing anchor primer provided with the system. A nested amplification using an inner gene-specific primer (c20r) and the anchor primer from the provider was also performed.

3′-Rapid amplification of cDNA ends

3′-RACE was performed to obtain the 3′-end of the transcript using the 3′-RACE System Version 2 (Invitrogen) with primers specific for the *dystrophin* mRNA (Table 1). First-strand cDNA synthesis was initiated at the poly(A) tail of mRNA using the adapter primer from the provider. After first-strand cDNA synthesis, the original mRNA template was destroyed with *RNase* H. Amplification was performed using a gene-specific primer (c16f) and a universal amplification primer from the provider that targets the cDNA complementary to the 3′-end of the mRNA. A nested amplification using an inner gene-specific primer (c18f) and the anchor primer from the provider was also performed.

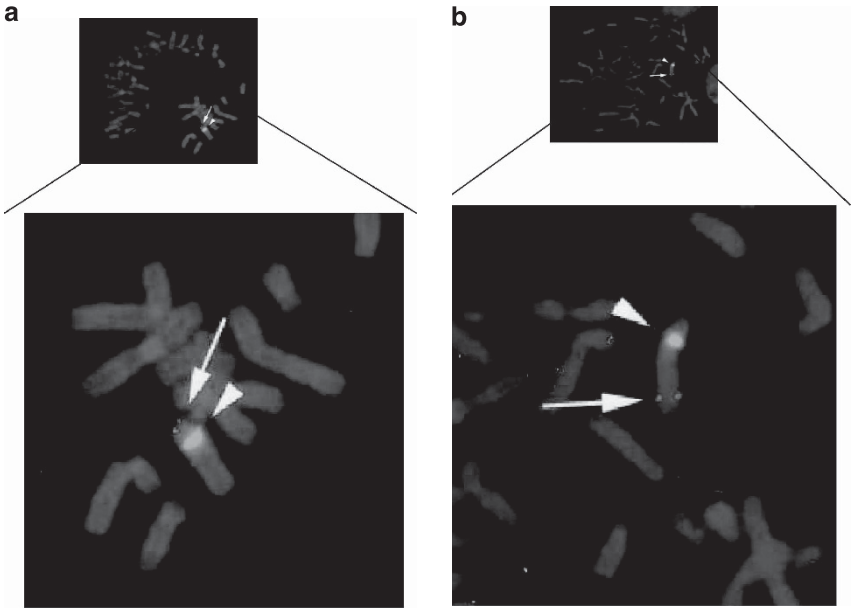

Figure 1 FISH analysis revealing disruption of the *dystrophin* gene. Results of FISH examination are shown with an enlarged panel (below). Centromeric signal is marked by arrowheads. (**a**) Exon 18 probe. Hybridization signals (arrow) are present on the short arm of the X-chromosome. (**b**) Exon 19 probe. Signals (arrow) are present on the long arm of the X-chromosome. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

DNA sequencing

PCR-amplified bands were excised from the gel with a sharp razor blade, pooled and purified using a QIAGEN gel extraction kit (QIAGEN, Inc., Hilden, Germany) according to the manufacturer's instructions. Purified products were sequenced either directly or after subcloning into the pT7 Blue T-vector (Novagen, San Diego, CA, USA). DNA sequencing was performed using a BigDye 1.1 Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA) in a Mastercycler Gradient (Eppendorf). The DNA sequences were determined using an automated DNA sequencer (ABI 310; Applied Biosystems).

mRNA expression of KUCG1

The expression of the *KUCG1* transcript was examined by reverse transcription-PCR. Human total RNA from 21 tissues (adrenal gland, bone marrow, brain, colon, fetal brain, fetal liver, heart, kidney, liver, lung, lymphocytes, placenta, prostate, salivary gland, skeletal muscle, spinal cord, testis, thymus, thyroid gland, trachea and uterus) was obtained from a human total RNA Master Panel II (Clontech Laboratories, Inc., Mountain View, CA, USA). cDNA was synthesized as described previously[27] from 2.5 μg of each total RNA. The *KUCG1* transcript spanning exon 2 to exon 3 was amplified by semi-nested PCR using primers Bf and Cr2, then Bf and Cr1 (Table 1), yielding a 314-bp fragment.

To check the integrity and concentration of the cDNA, the glyceraldehyde-3-phosphate dehydrogenase gene was also reverse transcription-PCR amplified, as described previously.[28]

Database searches and multiple sequence alignments

Homology searching was performed using the National Center for Biotechnology Information BLAST program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The cloned 658-bp sequence was searched using NONCODE v3.0.[29] The core promoter of the *KUCG1* gene was analyzed using Genety X (Ver. 8.2.0) (GENETYX corporation, Tokyo, Japan).

RESULTS

We performed a molecular characterization of an intrachromosomal inversion in a DMD patient, inv(X)(p21.2;q28). We were able to amplify all 79 *dystrophin* exon-encompassing regions from the patient's genome (data not shown), indicating that the overall structure of the gene was intact. We examined the full-length *dystrophin* cDNA as 10 separate fragments. All the cDNA fragments could be obtained by PCR except one that covered exons 17 to 25 (data not shown). This suggested that the *dystrophin* cDNA was separated into two fragments; one from exons 1 to 18 and the other from exons 19 to 79 (data not shown). We used fluorescence *in situ* hybridization to confirm this. As expected, an exon 19 probe hybridized to the long arm of the X-chromosome, while an exon 18 probe hybridized to the short arm (Figure 1). We concluded that the exon 19 dislocation from the short arm to the long arm was the cause of DMD.

We were surprised the distal *dystrophin* cDNA (exons 19 to 78) could be PCR amplified, because this indicated that it formed a new fusion gene after dislocation. We, therefore, examined the full-length transcript using skeletal muscle RNA from the patient (Figure 2). We obtained a 5′-RACE product from exon 20, which contained 109 bp between the adapter and *dystrophin* exon 19 sequence (Figure 2). Homology searching of the identified sequence revealed that, although it did not match any known gene, it was identical to a portion of Xq28 (GenBank ID: NW001842413.1). The first nucleotide of the cloned sequence was 89,813 bp downstream from the melanoma antigen family A, 9 (*MAGEA9*) gene (Figure 3). Examination of the genomic sequence 3′ of the cloned 109-bp sequence revealed a GT dinucleotide, a splice donor consensus sequence (Figure 2). Although an AG dinucleotide-a consensus splice acceptor sequence was not present at the 5′-end, we did identify a TATA-box 5′-(ATATATAA CAATTTA)-3′, GC-box 5′-(TAAGGGCATACCCT)-3′ and CCAAT-box 5′-(CCTAGCCAATAG)-3′ at 168, 266 and 372 bp upstream of the cloned sequence, respectively (Figure 2). Additionally, a cap signal sequence (TCAGCAAC) was present 24 bp upstream. These characteristics indicated that the cloned sequence was the first exon of an unknown gene that is transcribed in the centromere-to-telomere direction. We concluded that, in the patient, the first exon of the

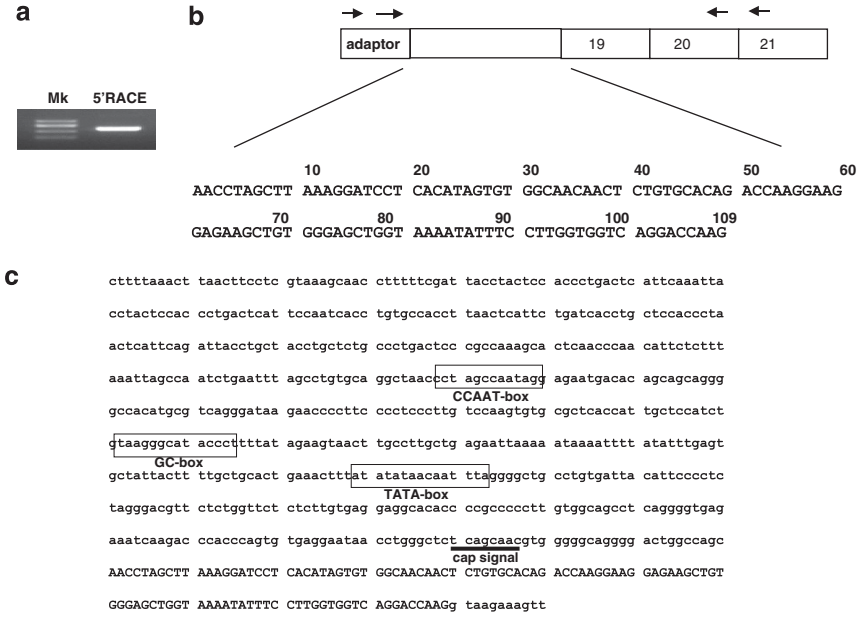

Figure 2 5′-RACE of *dystrophin* transcript. (**a**) Product of 5′-RACE of skeletal muscle RNA from the patient is shown (5′-RACE). Mk refers to ϕX174-*Hae III* molecular weight marker. (**b**) Schematic description of the amplified product. Numbered boxes indicate *dystrophin* exons. The open box indicates the novel 109-bp sequence. Arrows indicate primers used for PCR. (**c**) Part of Xq28 genomic sequence indicating the identified 109 nt (upper case). The boxed regions indicate the TATA-box, GC-box and CCAAT-box at 168, 266 and 372 bp upstream, respectively. A cap signal (thick underline) was identified 24 nt upstream of the 109-bp sequence.

unknown gene spliced to the dislocated part of the *dystrophin* gene, producing a chimeric *dystrophin* transcript.

To identify the rest of the novel gene, we conducted 3′-RACE using a primer in exon 16, and obtained one clear product (Figure 4). Sequencing of the amplified product revealed a 583-bp sequence inserted between *dystrophin* exon 18 and the adapter sequence (Figure 4). Homology searching revealed that this sequence, apart from the last seven 'A' nt, matched two separate regions of Xq28. The first 123 bp that were continuous with the 3′-end of exon 18 completely matched nt 148986563–148986685 and the last 453 bp matched nt 149008147–149008599 (NC 00023.10). The last nucleotide was located 4448 bp upstream of the melanoma antigen family A, 8 (*MAGEA8*) gene (Figure 3). Examination of the genomic sequences flanking the first 123 bp revealed consensus splice donor and acceptor sites at the 3′ and 5′ ends, respectively, indicating that it is an internal exon of an unknown gene. The last 453 bp had an AG dinucleotide immediately upstream but no GT dinucleotide downstream. Instead, a consensus polyadenylation signal (AATAAA) was identified 14 bp upstream of the 3′-end (Figure 4).[30] Considering the stretch of seven 'A's as part of a poly(A) tail, we concluded that the 453 bp sequence was the last exon of the unknown gene. The *dystrophin* promoter would produce a chimeric transcript comprising *dystrophin* exons 1–18 and two novel exons at the 3′-end.

Combining the results of 5′ and 3′-RACE, we had cloned a 685-bp-long transcript, the sequence of which we deposited in GenBank under the accession number JX283354. Homology searching did not reveal any transcript with significant similarity. The transcript had no significant open-reading frame, but because of its mRNA-like structure and length of >200 bp, we concluded that it was a novel lncRNA. We named it *KUCG1*. *KUCG1* spans nearly 50 kb on Xq28 and is located 9.0 kb downstream of *MAGEA9* and 4.4 kb upstream of *MAGEA8* (Figure 3). It has three exons separated by two introns (32 kb and 20 kb long, respectively). The site of recombination of the intrachromosomal inversion inv(X)(p21.2;q28) was intron 1. The inversion caused a head-to-tail fusion of *KUCG1* and *dystrophin* at the recombination sites. We searched for homologous lncRNAs using NONCODE v3.0,[29] but did not identify any significant matches. This indicated that *KUCG1* is a novel lncRNA. It was found that exon 3 of KUCG1 overlaps with the antisense transcript RP5-869M20.2, an lncRNA of unknown function (Figure 3).

We next examined the tissue-specific expression of *KUCG1* in humans. We amplified a fragment comprising exons 2 to 3 by reverse transcription-PCR of total RNA from 21 human tissues. The expected size product was obtained by semi-nested PCR from four tissues (lung, thyroid gland, brain and placenta), whereas no product was obtained from the other 17 tissues (Figure 5). Considering the brain expression of *KUCG1*, we consider that its disruption may be responsible for the moderate mental retardation in the index case.

DISCUSSION

In this report, we describe molecular characterization of an inverted X (p21.2;q28) chromosome in a patient with DMD and mental retardation. The inversion disrupted both the *dystrophin* gene, presumed to be the cause of the DMD, and a novel lncRNA, *KUCG1*, which may be the cause of the mental retardation. This is the third intrachromosomal inversion to be molecularly clarified in DMD,[9,10] but the first to disrupt unknown gene directly.

The *KUCG1* mRNA was detected in 4 out of 21 tissues: lung, thyroid gland, brain and placenta (Figure 5), indicating tissue-specific gene regulation despite the presence of three common consensus sequences in the promoter. The tissue-restricted expression and low expression level (semi-nested PCR was required to detect a product) could explain why this lncRNA has not been previously detected among the thousands of ncRNAs identified by high-throughput sequencing.[31]

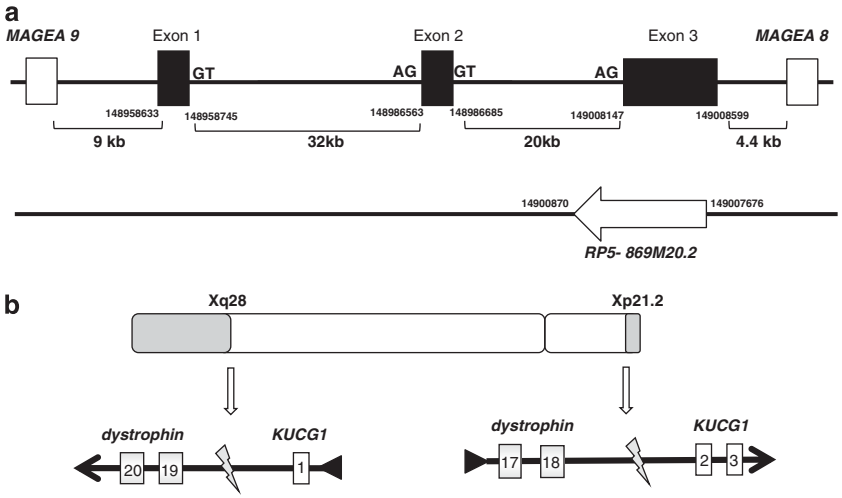

Figure 3 Schematic description of the gene and X-chromosome. (**a**) Schematic description of the *KUCG1* gene. The *KUCG1* gene that spans nearly 50 kb on Xq28 is transcribed in a centromere-to-telomere direction, and comprises three exons (black boxes) of 109, 123 and 453 bp, respectively. Numbers below the exons indicate the chromosomal nucleotide position according to GenBank NC00023.10. Introns 1 and 2 span 32 kb and 20 kb, respectively. *KUCG1* is located between *MAGEA9* and *MAGEA8* (open boxes). Another non-coding gene, *RP5-869M20.2* (ENSG00000230899.1) has been mapped to this region (nt 149007636–149009870) but is transcribed in the antisense direction (horizontal arrow). (**b**) Schematic description of the translocated X-chromosome schema of inv(X)(p21.2;q28) is described. At Xq28 intron 1 of the *KUCG 1* gene directly joined to intron 18 of the *dystrophin* gene. In contrast, intron 18 of the *dystrophin* gene joined to intron 1 of the *KUCG1* gene. Open and shaded boxes are normal and translocated parts of X-chromosome, respectively. Horizontal arrows and triangles indicate the direction and the promoter region of fused genes, respectively. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

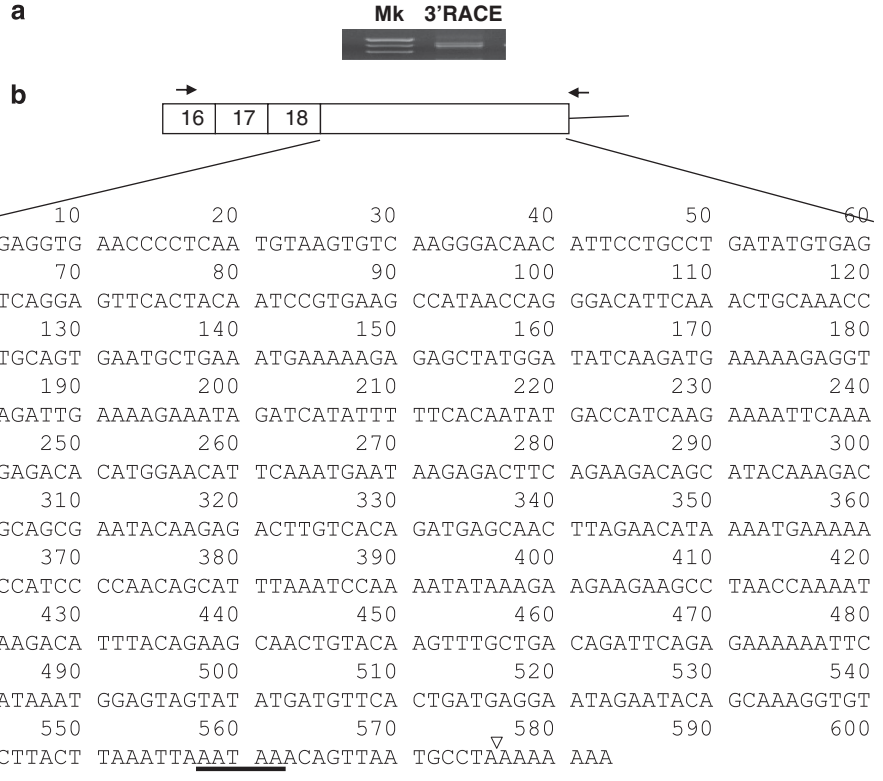

Figure 4 3′-RACE of *dystrophin* transcript. (**a**) Product of 3′-RACE of skeletal muscle RNA from the patient is shown (3′-RACE). Mk refers toϕX174-*Hae III* molecular weight marker. (**b**) Schematic description and sequence of the 3′-RACE product. *Dystrophin* exons are indicated as numbered open boxes. The product contained a 583-nt sequence (open box) downstream of *dystrophin* exon 18 (numbered box). The 583-nt sequence contains a polyadenylation signal (thick underline) followed by a short poly(A) tail (open triangle). The first 123 nt and the last 453 nt of the sequence (separated at the filled triangle) matched two separate regions on Xq28.

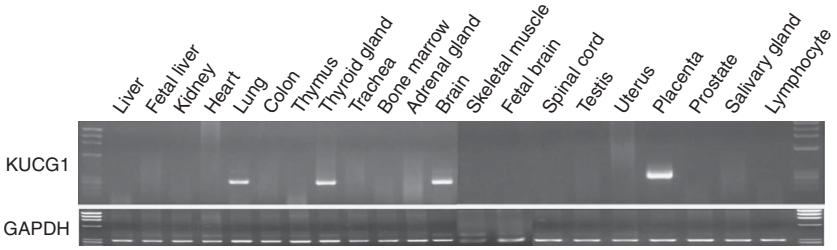

Figure 5 Tissue-specific expression of *KUCG1* mRNA. Products of reverse transcription-PCR amplification of *KUCG1* mRNA are shown. Reverse transcription-PCR amplification of 21 human tissues revealed a product in lung, thyroid gland, brain and placenta. The correct identity of the product was validated by sequencing. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) mRNA levels were used as a reference.

What is the function of the *KUCG1* gene? As it undergoes splicing, is >200 nt long, and contains features such as a poly(A) signal/tail, *KUCG1* can be considered an mRNA-like ncRNA.[32,33] lncRNAs have been shown to have key roles in imprinting control, immune responses and human disease;[20] for instance, an ncRNA cloned from a chromosomal inversion was recently demonstrated to cause autosomal dominant hypertension and brachydactyly (OMIM 112410).[34] In the central nervous system, the increasing variety of ncRNAs shown to be expressed has suggested a strong connection between ncRNAs and the complexity of the system.[35] Hundreds of lncRNAs have been shown to localize to specific neuroanatomical regions, cell types or subcellular compartments within the brain[36] and a subset of lncRNAs is likely to contribute to neurological disorders.[23] For instance, the levels of the linc-MD1 lncRNA are strongly reduced in DMD,[37] indicating a role for this lncRNA in the disease pathology of DMD.

The mechanism of action of lncRNAs is thought to involve direct binding to target sites on proteins and RNAs.[33,37] It is interesting that exon 3 of *KUCG1* overlaps with the antisense transcript *RP5-869M20.2*, an lncRNA of unknown function. It is possible that transcripts from *KUCG1* and *RP5-869M20.2* form a double-stranded RNA that has a particular physiological role.

As *KUCG1* is expressed in the brain, we suspect that its disruption is responsible for the moderate mental retardation in the index case. Although >40 genes responsible for X-linked mental retardation have been annotated to Xq28,[17] the gene(s) responsible for many cases of X-linked mental retardation remain unidentified.[14] To test whether *KUCG1* is responsible for other cases of X-linked mental retardation, we sequenced *KUCG1* in ten Japanese families with X-linked mental retardation for which no responsible gene mutation has been identified. No mutations were identified (data not shown). Although we have not provided direct evidence linking mental retardation to mutation of *KUCG1*, further studies of its function, and mutation analysis in other X-linked mental retardation families, is warranted.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to thank Ms K Yoshikawa for her secretarial help. This work was supported by a Health and Labor Sciences Research Grant for Research on Psychiatric and Neurological Diseases and Mental Health, a Grant-in-Aid for Scientific Research grant (B) from the Japan Society for the Promotion of Science (MEXT/JSPS KAKENHI, grant number 24390267), and a research grant for Nervous and Mental Disorders from the Ministry of Health, Labor, and Welfare, Japan.

1 Ahn, A. H. & Kunkel, L. M. The structural and functional diversity of dystrophin. *Nat. Genet.* **3,** 283–291 (1993).
2 Takeshima, Y., Yagi, M., Okizuka, Y., Awano, H., Zhang, Z., Yamauchi, Y. *et al.* Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. *J. Hum. Genet.* **55,** 379–388 (2010).
3 Koenig, M., Hoffman, E. P., Bertelson, C. J., Monaco, A. P., Feener, C. & Kunkel, L. M. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell* **50,** 509–517 (1987).
4 Monaco, A. P., Bertelson, C. J., Liechti-Gallati, S., Moser, H. & Kunkel, L. M. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics* **2,** 90–95 (1988).
5 Daoud, F., Candelario-Martinez, A., Billard, J. M., Avital, A., Khelfaoui, M., Rozenvald, Y. *et al.* Role of mental retardation-associated dystrophin-gene product Dp71 in excitatory synapse organization, synaptic plasticity and behavioral functions. *PLoS One* **4,** e6574 (2009).
6 Taylor, P. J., Betts, G. A., Maroulis, S., Gilissen, C., Pedersen, R. L., Mowat, D. R. *et al.* Dystrophin gene mutation location and the risk of cognitive impairment in Duchenne muscular dystrophy. *PLoS One* **5,** e8803 (2010).
7 Pane, M., Lombardo, M. E., Alfieri, P., D'Amico, A., Bianco, F., Vasco, G. *et al.* Attention deficit hyperactivity disorder and cognitive function in Duchenne muscular dystrophy: phenotype–genotype correlation. *J. Pediatr.* **161,** 705–709 (2012).
8 Kunkel, L, Hejtmancik, JF, Caskey, CT, Speer, A, Monaco, AP, Middlesworth, W *et al.* Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. *Nature* **322,** 73–77 (1986).
9 Saito-Ohara, F., Fukuda, Y., Ito, M., Agarwala, K. L., Hayashi, M., Matsuo, M. *et al.* The Xq22 inversion break point interrupted a novel ras-like GTPase gene in a patient with Duchenne muscular dystrophy and profound mental retardation. *Am. J. Hum. Genet.* **71,** 637–645 (2002).
10 Bhat, S. S., Schmidt, K. R., Ladd, S., Kim, K. C., Schwartz, C. E., Simensen, R. J. *et al.* Disruption of DMD and deletion of ACSL4 causing developmental delay, hypotonia, and multiple congenital anomalies. *Cytogenet. Genome Res.* **112,** 170–175 (2006).
11 van Bakel, I., Holt, S., Craig, I. & Boyd, Y. Sequence analysis of the breakpoint regions of an X;5 translocation in a female with Duchenne muscular dystrophy. *Am. J. Hum. Genet.* **57,** 329–336 (1995).
12 Kalz-Fuller, B., Sleegers, E., Schwanitz, G. & Schubert, R. Characterisation, phenotypic manifestations and X-inactivation pattern in 14 patients with X-autosome translocations. *Clin. Genet.* **55,** 362–366 (1999).
13 Wheway, J. M., Yau, S. C., Nihalani, V., Ellis, D., Irving, M., Splitt, M. *et al.* A complex deletion-inversion-deletion event results in a chimeric IL1RAPL1-dystrophin transcript and a contiguous gene deletion syndrome. *J. Med. Genet.* **40,** 127–131 (2003).
14 Ropers, H. H. & Hamel, B. C. X-linked mental retardation. *Nat. Rev. Genet.* **6,** 46–57 (2005).
15 Raymond, F. L. X linked mental retardation: a clinical guide. *J. Med. Genet.* **43,** 193–200 (2006).
16 Tarpey, P. S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C. *et al.* A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat. Genet.* **41,** 535–543 (2009).
17 Kolb-Kokocinski, A., Mehrle, A., Bechtel, S., Simpson, J. C., Kioschis, P., Wiemann, S. *et al.* The systematic functional characterisation of Xq28 genes prioritises candidate disease genes. *BMC Genomics* **7,** 29 (2006).
18 Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mattick, J. S. The eukaryotic genome as an RNA machine. *Science* **319,** 1787–1789 (2008).
19 Derrien, T., Guigo, R. & Johnson, R. The long non-coding RNAs: a new (P)layer in the 'Dark Matter'. *Front. Genet.* **2,** 107 (2012).
20 Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10,** 155–159 (2009).
21 Mercer, T. R., Qureshi, I. A., Gokhan, S., Dinger, M. E., Li, G., Mattick, J. S. *et al.* Long noncoding RNAs in neuronal-glial fate specification and oligodendrocyte lineage maturation. *BMC Neurosci.* **11,** 14 (2010).
22 Taft, R. J., Pang, K. C., Mercer, T. R., Dinger, M. & Mattick, J. S. Non-coding RNAs: regulators of disease. *J. Pathol.* **220,** 126–139 (2010).

23 Niland, C. N., Merry, C. R. & Khalil, A. M. Emerging roles for long non-coding RNAs in cancer and neurological disorders. *Front. Genet.* **3,** 25 (2012).
24 Shiga, N., Takeshima, Y., Sakamoto, H., Inoue, K., Yokota, Y., Yokoyama, M. *et al.* Disruption of the splicing enhancer sequence within exon 27 of the dystrophin gene by a nonsense mutation induces partial skipping of the exon and is responsible for Becker muscular dystrophy. *J. Clin. Invest.* **100,** 2204–2210 (1997).
25 Kubokawa, I., Takeshima, Y., Ota, M., Enomoto, M., Okizuka, Y., Mori, T. *et al.* Molecular characterization of the 5'-UTR of retinal dystrophin reveals a cryptic intron that regulates translational activity. *Mol. Vis.* **16,** 2590–2597 (2010).
26 Roberts, R. G., Barby, T. F., Manners, E., Bobrow, M. & Bentley, D. R. Direct detection of dystrophin gene rearrangements by analysis of dystrophin mRNA in peripheral blood lymphocytes. *Am. J. Hum. Genet.* **49,** 298–310 (1991).
27 Matsuo, M., Masumura, T., Nishio, H., Nakajima, T., Kitoh, Y., Takumi, T. *et al.* Exon skipping during splicing of dystrophin mRNA precursor due to an intraexon deletion in the dystrophin gene of Duchenne muscular dystrophy Kobe. *J. Clin. Invest.* **87,** 2127–2131 (1991).
28 Tran, V. K., Zhang, Z., Yagi, M., Nishiyama, A., Habara, Y., Takeshima, Y. *et al.* A novel cryptic exon identified in the 3' region of intron 2 of the human dystrophin gene. *J. Hum. Genet.* **50,** 425–433 (2005).
29 Bu, D., Yu, K., Sun, S., Xie, C., Skogerbo, G., Miao, R. *et al.* NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res.* **40,** D210–D215 (2011).
30 Proudfoot, N. J. Ending the message: poly(A) signals then and now. *Genes Dev.* **25,** 1770–1782 (2012).
31 Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* **25,** 1915–1927 (2011).
32 Erdmann, V. A., Szymanski, M., Hochberg, A., de Groot, N. & Barciszewski, J. Collection of mRNA-like non-coding RNAs. *Nucleic Acids Res.* **27,** 192–195 (1999).
33 Nagano, T. & Fraser, P. No-nonsense functions for long noncoding RNAs. *Cell* **145,** 178–181 (2011).
34 Bähring, S., Kann, M., Neuenfeld, Y., Gong, M., Chitayat, D., Toka, H. *et al.* Inversion region for hypertension and brachydactyly on chromosome 12p features multiple splicing and noncoding RNA. *Hypertension* **51,** 426–431 (2008).
35 Cao, X., Yeo, G., Muotri, A. R., Kuwabara, T. & Gage, F. H. Noncoding RNAs in the mammalian central nervous system. *Annu. Rev. Neurosci.* **29,** 77–103 (2006).
36 Mercer, T. R., Dinger, M. E., Sunkin, S. M., Mehler, M. F. & Mattick, J. S. Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. USA* **105,** 716–721 (2008).
37 Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M. *et al.* A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147,** 358–369 (2011).