npg

ORIGINAL ARTICLE

Contemporary retrotransposition of a novel non-coding gene induces exon-skipping in dystrophin mRNA

Hiroyuki Awano, Rusdy Ghazali Malueka, Mariko Yagi, Yo Okizuka, Yasuhiro Takeshima and Masafumi Matsuo

Non-autonomous retrotransposon-mediated mobilizations of the Alu family are known pathogenic mechanisms of human disease. Here, we report a pathogenic, contemporary, non-autonomous retrotransmobilization of part of a novel non-coding gene into the *dystrophin* gene. In a Japanese Duchenne muscular dystrophy patient, a 330-bp-long *de novo* insertion was identified in exon 67 of dystrophin. The insertion induced exon 67-skipping in the dystrophin mRNA, creating a premature stop codon. The sequence of the insertion had certain characteristics of retrotransposons: an antisense polyadenylation signal accompanied by a poly(T) sequence and a target site duplication. The insertion site matched the consensus recognition sequence for the L1 endonuclease, indicating a retrotransposon-mediated event, although the inserted sequence did not match any known retrotransposons. The origin of the inserted sequence was mapped to a gene-poor region of chromosome 11. The inserted fragment was expressed in multiple human tissue RNAs, indicating that it is a novel transcript. The full length of the transcript was cloned and showed no meaningful protein coding ability.

INTRODUCTION

Mobile DNA elements are discrete DNA sequences that have the remarkable ability to transport or duplicate themselves to other regions of the genome. Mobile elements can be divided into two different classes based on how they duplicate themselves within the genome.[1,2] DNA transposons mobilize through a DNA intermediate, typically using a so-called 'cut-and-paste' mechanism. Retrotransposons mobilize through RNA and proceed via a 'copy-and-paste' mechanism. In this process, an RNA copy is first generated from the original retrotransposon and is subsequently reverse transcribed back into DNA using the enzyme reverse transcriptase. cDNA is then inserted into a new location in the genome, sometimes disrupting host gene function.[3–5]

Retrotransposons can be further subdivided into those elements that are autonomous, meaning that they encode their own replication machinery (for example, long interspersed nuclear element 1: L1 or LINE1),[6] and those elements that are non-autonomous, such as the Alu family.[2] Non-autonomous retrotransposons borrow the enzymatic machinery required for their propagation from L1 elements. L1 endonuclease-dependent retrotransposition has been reported to cause many human genetic diseases.[7]

Processed pseudogenes are another retrotransposable element resulting from the random integration of reverse-transcribed mature RNA molecules into genomes. They are characterized by a lack of introns, the presence of a poly(A) tail and the presence of flanking direct repeats.[8] This gene retrotransposition may arise as a by-product

of L1 retrotransposition.[9] Recently, an ancient retrotranspositional insertion of a transcript from chromosome 6 (chromosome 6 open-reading frame 68) has been shown to disrupt the *SLC25A13* gene.[10] Contemporary retrotransposition of a gene transcript has never been shown to cause a genetic disease.

Mutations in the *dystrophin* gene, the largest human gene spreading over 2500 kb on the short arm of the X chromosome, cause Duchenne muscular dystrophy (DMD), the most common inherited muscular disease affecting one in every 3500 male subjects. Although deletions removing one or more exon of the gene have been reported as the most common mutation, more than 200 mutations have been identified in the *dystrophin* gene (http://www.dmd.nl/). To date, retrotranspositional insertions into this gene have been reported in four cases. In our previous Japanese study, an L1 insertion was identified in one Japanese DMD patient.[11] All three other insertions were derived from L1 retrotransposons.[12–14]

Here, we identify a contemporary retrotranspositional insertion of a novel non-coding gene from chromosome 11 into exon 67 of the *dystrophin* gene in a Japanese DMD patient. This is a novel retrotransposon-mediated transmobilization that causes human disease.

MATERIALS AND METHODS

Case

The proband (KUCG732) was a 4-year-old Japanese boy. He had no family history of neuromuscular disease. At 3 years old, his serum creatine kinase level

786

was found to be high $(14\,780\,IU\,l^{-1})$ on blood chemical examination. At 4 years old, a muscle biopsy was performed and immunohistochemical examination, using three dystrophin monoclonal antibodies that recognize three different epitopes, found an absence of dystrophin staining, and the DMD diagnosis was confirmed. Informed consent was obtained from his parents for molecular analysis and the study was approved by the ethics committees of Kobe University School of Medicine (approval no. 28 in 1998).

Methods

Mutation analysis. The patient's genomic DNA was extracted from peripheral blood. Each of the 79 exons of the *dystrophin* gene was polymerase chain reaction (PCR) amplified as described previously.[15] The region encompassing exon 67 was amplified using the forward primer DMD-67U (5′-GAAGTA ACCCCACTACTGTGGAA-3′) and the reverse primer DMD-67L (5′-AAACGA AGCTCTGTGGGTTT-3′).

The dystrophin mRNA expressed in the skeletal muscle was examined by reverse transcription PCR (RT-PCR) as described previously.[16] Briefly, total RNA was isolated from thinly sliced (6-μm) sections of frozen muscle using Isogen (Nippon Gene, Toyama, Japan). After synthesizing cDNA with reverse transcriptase (Invitrogen Corp., Carlsbad, CA, USA), a fragment extending from exons 64 to 68 was amplified using a forward primer corresponding to a segment of exon 64 (c64f, 5′-CTCCGAAGACTGCAGAAGGC-3′) and a reverse primer complementary to a segment of exon 68 (5D, 5′-TTTCTGCAGCCA CTCT-3′) as described previously.[17]

The PCR-amplified products were electrophoresed on agarose gels. Purified PCR products were subjected to sequencing either directly or after subcloning into the pT7 blue T vector (Novagen, Madison, WI, USA).

Transcript analysis. A fragment covering the inserted sequence was amplified by RT-PCR from human tissue RNAs (Human Total RNA Panel; Clontech, Mountain View, CA, USA). First-strand cDNA synthesis was carried out with 3 μg of RNA using SuperScript II reverse transcriptase (Invitrogen Corp.). To amplify the fragment covering the inserted sequence, the forward primer awa11q3Lf (5′-GCCTCTGGATCAGGAAGAGC-3′) and the reverse primer awa11q3Rr (5′-TTTTTGAAATTTGAAGCATTTTTCC-3′) were used. Thirty-five PCR cycles were performed in a mixture as described before,[17] using the following conditions: initial denaturation at 94 °C for 4 min, subsequent denaturation at 94 °C for 1 min, annealing at 60 °C for 1 min and extension at 72 °C for 1 min. The final extension reaction was carried out at 72 °C for 1 min. An aliquot of amplified DNA was electrophoresed on a 3% agarose gel and stained with ethidium bromide along with a low-molecular-weight DNA standard (φ174X–HaeIII digest; Takara Bio, Shiga, Japan). In addition, a fragment of the glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) gene was also amplified using two primers: the forward primer GAPDH-F106 (5′-CCCTTCATTGACCTCAAC-3′) and the reverse primer GAPDH-R407 (5′-TTCACACCCATGACGAAC-3′), as described before. These PCR products were verified by sequencing.

Cloning of the novel transcript. To obtain the 5′ end of the novel transcript, 5′-rapid amplification of cDNA ends (5′-RACE) was performed using the 5′-RACE System (version 2.0; Invitrogen Corp.). Single-stranded cDNA was synthesized from brain RNA (Clontech) using a gene-specific primer, GSP1 (5′-TGAAATTTGAAGCATTTTTCCAA-3′). A homopolymeric T-tail was added to the 3′ end of the cDNA using terminal deoxynucleotidyl transferase and dCTP. The dC-tailed cDNA was amplified with the gene-specific primer, GSP2 (5′-GGCTGTGAATAATAGCATTCT-3′), and the cassette primer, Abridged Anchor (5′-GGCCACGCGTCGACTAGTACGGGGGGGGGGGG-3′). The resulting product was re-amplified in a second round of PCR using the primers nestedGSP (5′-CCACCAAACTGTTAAACTCA-3′) and AUAP (5′-GGC CACGCGTCGACTAGTAC-3′). PCR products were separated by agarose gel electrophoresis and subjected to subcloning and sequencing.

DNA sequencing. DNA sequencing was performed using the BigDye 2.0 or 3.1 Terminator Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA). PCR products for sequencing were either gel-purified and/or cloned into the pT7 blue T vector (Novagen) using the TOPO TA Cloning kit (Invitrogen Corp.). The primers used for sequencing PCR products were identical to the primers used for amplification of the corresponding targets. Sequencing of PCR fragments cloned into the pT7 blue T vector was performed using the forward primer PT7-F (5′-CTATAGGGAAAGCTTGCATGC-3′) and reverse primer PT7-R (5′-GTTTTCCCAGTCACGACGTTG-3′). Sequencing was performed on an ABI 310 capillary sequencer (Applied Biosystems).

Database searches and multiple sequence alignments. Homology searches were conducted using the Basic Local Alignment Search Tool (http://blast. ncbi.nlm.nih.gov/) at the nucleotide, transcript or protein level in GenBank (http://www.ncbi.nlm.nih.gov/genbank/), Refseq_rna (http://www.ncbi.nlm.nih. gov/RefSeq/), dbEST (http://www.ncbi.nlm.nih.gov/projects/dbEST/), Swissprot (http://www.expasy.org/sprot/) and Refseq_protein (http://www.ncbi.nlm.nih. gov/RefSeq/). Micro-RNA analysis was performed using miRBase (http:// mirbase.org/). DNA sequences encompassing the inserted sequence were analyzed for repetitive elements using the RepeatMasker web server (www. repeatmasker.org) with Repbase (http://www.girinst.org/repbase/) database. The core promoter was analyzed using Genetyx, version 8.2.0 (Genetyx Corp., Osaka, Japan).

RESULTS

To identify the responsible mutation in the *dystrophin* gene in the index case, all 79 exons of the *dystrophin* gene were PCR amplified using primers in the flanking introns. All exons except exon 67 could be amplified at their normal lengths. The amplified region encompassing exon 67 was obtained as a fragment larger than the same region amplified from the patient's father or mother (Figure 1a). Sequencing of the amplified product (Figure 1b) revealed that the 5′ portion of the sequence was identical to the normal sequence until the eighth nucleotide of exon 67 (c.9657C), but this was followed by approximately 330 bp of unknown sequence. The unknown sequence was followed by the 3′ portion of exon 67, beginning at c.9655T, and the amplified portion of intron 67. This indicated that there was an approximately 330-bp insertion mutation in exon 67. As the patient's mother displayed only one normal-sized amplicon for this region, as did his father (Figure 1a), she was deemed a non-carrier for this mutation. Therefore, we concluded that the insertion event occurred *de novo* in the patient.

The impact of the large insertion on splicing was examined by RT-PCR amplification of dystrophin mRNA obtained from skeletal muscle. When the region extending from exons 64 to 68 was amplified, the amplified product was shorter than expected (Figure 2). Sequencing of this product showed that the 3′ end of exon 66 joined directly to the 5′ end of exon 68, indicating complete exon 67-skipping. As the result of a frame shift, a premature stop codon was created at the second codon in exon 68. We concluded that the insertion caused exon 67-skipping, which led to the DMD phenotype.

To identify the insertion, we examined the inserted sequence and discovered two important characteristics: (1) TTC trinucleotides from c.9655 to 9657 were present at both ends, indicating that TTC was the target site for duplication (Figure 1b); and (2) the inserted sequence had an approximately 115-bp stretch of T (Figure 1b). These hallmarks indicated that the inserted fragment was a retrotransposed element. In addition, the remaining 212 bp of the inserted sequence had the reverse sequence of the polyadenylation signal (TTTATT) at the 26th nucleotide from the end (Figure 1b). However, a homology search for the 212-bp unknown sequence revealed no homology in any retrotransposon or transcript sequence database. Instead, we found a single genomic sequence on the long arm of chromosome 11 (11q22) that the complementary sequence of the 212-bp insertion matched perfectly (9041614–9041825, GenBank accession no., NT_033899.8) (Figure 3). As expected, a poly(A) stretch complementary to the poly(T) was not present in this genomic region. These results indicated that the inserted fragment was a reverse-transcribed product from a transcript with a poly(A) tail.
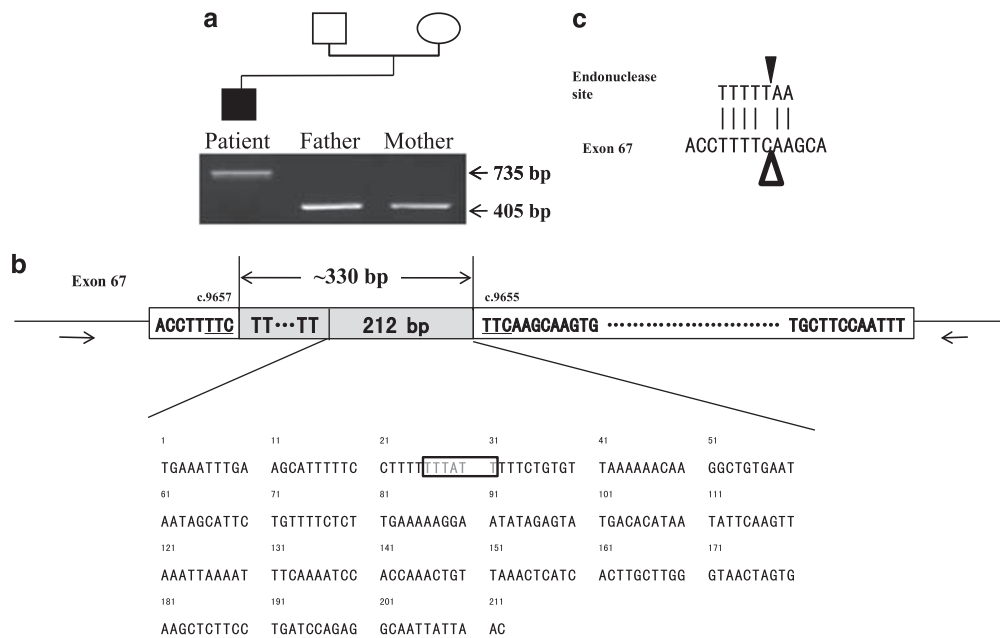
Figure 1 Analysis of the exon 67-encompassing region of the *dystrophin* gene. (a) Polymerase chain reaction (PCR) amplification of the exon 67-encompassing region. From the index patient, one clear product was obtained, but its size (735 bp) was larger than that amplified from his father or mother (405 bp). The pedigree of the family is also shown. (b) Schematic description of the exon 67-encompassing region of the patient. Open bars indicate the separated exon 67. The horizontal lines indicate introns. Horizontal arrows indicate the location and directions of primers. The shaded bar indicates a ~330-bp inserted sequence. The 5′ portion of exon 67 ended at the eighth nucleotide (c.9657), and the 3′ portion started at the sixth nucleotide of exon 67 (c.9655). The underlined TTC appears twice, at the end of the 5′ portion of exon 67 and at the beginning of the 3′ portion of exon 67. The inserted sequence is divided into two parts: the first has a poly(T) stretch of approximately 115 bp and the second has a unique 212-bp sequence that is described below. The candidate polyadenylation signal (TTTATT) is boxed. (c) Sequences around the insertion site. The consensus sequence for the L1 endonuclease site is shown on the upper line. The wild-type sequence of exon 67 is described on the lower line. Vertical lines indicate nucleotide matches. The filled triangle indicates the cleavage site. The open triangle indicates the site of the insertion.
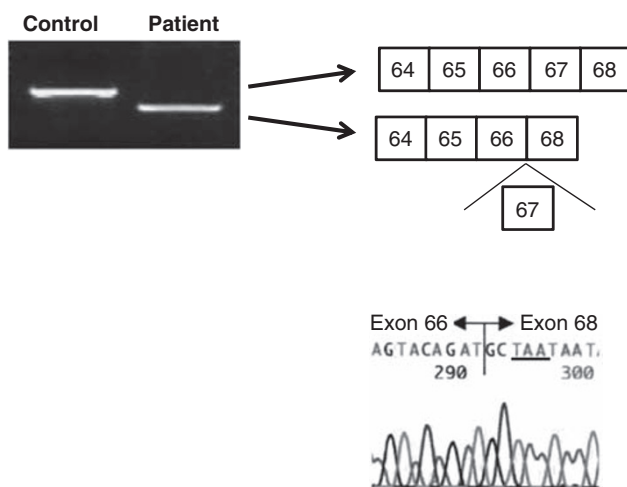

Figure 2 Reverse transcription-polymerase chain reaction (RT-PCR) amplification of dystrophin exons 64–68. RT-PCR products encompassing dystrophin exons 64–68 are shown (left). The size of the product obtained from the patient was smaller than the expected 582 bp (control). On the right-hand side is a schematic representation of the exon organization of the amplified fragment. Exon 67 has been completely removed from the dystrophin cDNA in the patient. The lower panel shows part of the sequence at the junction of exons 66 and 68. The 3′ terminal sequence of exon 66 (GAT) is directly joined to the 5′ end of exon 68 (GCT). The underlined sequence represents a stop codon (TAA).

When the sequence around the inserted site in exon 67 was examined, TTTTCAA, which is highly similar to the consensus sequence for the L1 endonuclease cleavage site (TTTTT/AA; '/' denotes the cleavage site), was found in the wild-type exon 67 sequence (Figure 1c). These sequences differed by only one nucleotide, with the fifth T replaced with the other pyrimidine nucleotide, C (underlined). Remarkably, the insertion was present at exactly the endonuclease cleavage site (Figure 1c). This indicated that an L1 endonuclease cut at TTTTC/AA creating the TTC target duplication. From the characteristics of the inserted sequence and the insert site sequence, we concluded that the insertion event was an L1-mediated retrotransposition.

Our findings strongly suggested that the source region on chromosome 11 is actively transcribed and thus can be reverse transcribed. As a database search failed to disclose the presence of this transcript in the human transcriptome, we assumed that the transcribed sequence has gone undetected because of a high tissue or developmental specificity. To observe expression of the source region, RT-PCR amplification of a fragment of the inserted segment was conducted using 10 human tissue RNAs (Figure 4). Remarkably, the expected product (206 bp) was observed clearly in the brain, thyroid, placenta, skeletal muscle and testis, and faintly in the heart, lung and kidney. The validity of the PCR products was confirmed by sequencing. No product was observed from the liver and bone marrow. Accordingly, no product was obtained from any of the 10 examined tissues (data not shown) when the PCR was conducted without the reverse transcription step. This indicates that the inserted sequence was present as a transcript in these tissues.
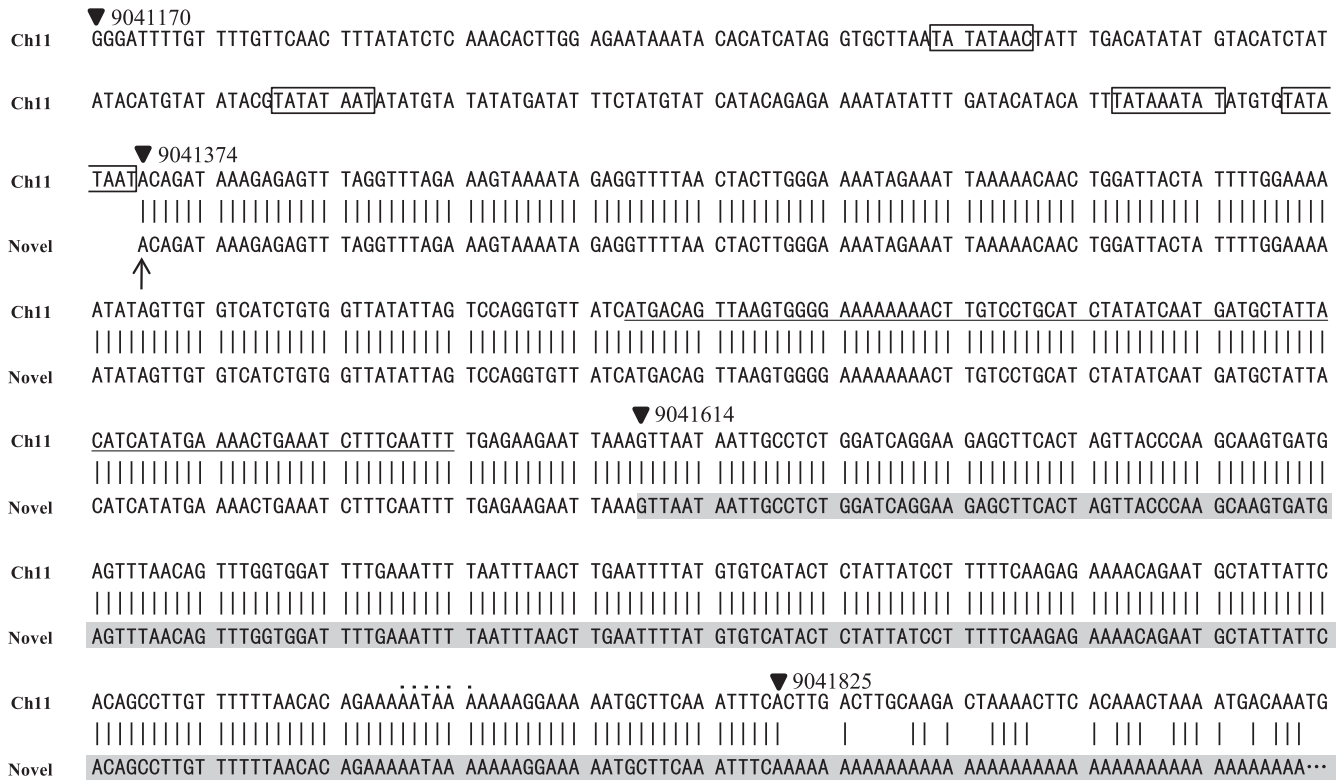
```
          ▼ 9041170
Ch11   GGGATTTTGT TTTGTTCAAC TTTATATCTC AAACACTTGG AGAATAAATA CACATCATAG GTGCTTAATA TATAACTATT TGACATATAT GTACATCTAT

Ch11   ATACATGTAT ATACGTATAT AATATATGTA TATATGATAT TTCTATGTAT CATCACAGAGA AAATATATTT GATACATACA TTTATAAATA TATGTGTATA

          ___   ▼ 9041374
Ch11   TAATACAGAT AAAGAGAGTT TAGGTTTAGA AAGTAAAATA GAGGTTTTAA CTACTTGGGA AAATAGAAAT TAAAAACAAC TGGATTACTA TTTTGGAAAA
          |||||  |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| ||||||||||
Novel     ACAGAT AAAGAGAGTT TAGGTTTAGA AAGTAAAATA GAGGTTTTAA CTACTTGGGA AAATAGAAAT TAAAAACAAC TGGATTACTA TTTTGGAAAA
          ↑
Ch11   ATATAGTTGT GTCATCTGTG GTTATATTAG TCCAGGTGTT ATCATGACAG TTAAGTGGGG AAAAAAAACT TGTCCTGCAT CTATATCAAT GATGCTATTA
       |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| ||||||||||
Novel  ATATAGTTGT GTCATCTGTG GTTATATTAG TCCAGGTGTT ATCATGACAG TTAAGTGGGG AAAAAAAACT TGTCCTGCAT CTATATCAAT GATGCTATTA

                                                     ▼ 9041614
Ch11   CATCATATGA AAACTGAAAT CTTTCAATTT TGAGAAGAAT TAAAGTTAAT AATTGCCTCT GGATCAGGAA GAGCTTCACT AGTTACCCAA GCAAGTGATG
       |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| ||||||||||
Novel  CATCATATGA AAACTGAAAT CTTTCAATTT TGAGAAGAAT TAAAGTTAAT AATTGCCTCT GGATCAGGAA GAGCTTCACT AGTTACCCAA GCAAGTGATG

Ch11   AGTTTAACAG TTTGGTGGAT TTTGAAATTT TAATTTAACT TGAATTTTAT GTGTCATACT CTATTATCCT TTTTCAAGAG AAAACAGAAT GCTATTATTC
       |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| ||||||||||
Novel  AGTTTAACAG TTTGGTGGAT TTTGAAATTT TAATTTAACT TGAATTTTAT GTGTCATACT CTATTATCCT TTTTCAAGAG AAAACAGAAT GCTATTATTC

                                    . . . . . .                     ▼ 9041825
Ch11   ACAGCCTTGT TTTTTAACAC AGAAAAATAA AAAAAGGAAA AATGCTTCAA ATTTCACTTG ACTTGCAAGA CTAAAACTTC ACAAACTAAA ATGACAAATG
       |||||||||| |||||||||| |||||||||| |||||||||| |||||||||| ||||||   |      |      || |    ||||     | ||| ||| | | |||
Novel  ACAGCCTTGT TTTTTAACAC AGAAAAATAA AAAAAGGAAA AATGCTTCAA ATTTCAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAA···
```

Figure 3 The origin of the inserted sequence on chromosome 11. The normal sequence of chromosome 11 (5′–3′) is shown on the top line (ch11). The bottom line shows the complement of the inserted sequence (shaded), and the upstream sequence obtained by 5′-rapid amplification of cDNA ends (5′-RACE) (novel). The vertical arrow indicates the 5′ end of the 5′-RACE product. Vertical lines indicate nucleotide matches. Boxes indicate TATA boxes. The polyadenylation signal (AATAAA) is marked by dots over the nucleotides. The longest open-reading frame, encoding 29 amino acids (MTVKWGKKTCPASISMMLLHHMKTEIFQF), is underlined. Vertical arrowheads indicate nucleotide numbers on chromosome 11 (GenBank accession no., NT_033899.8).
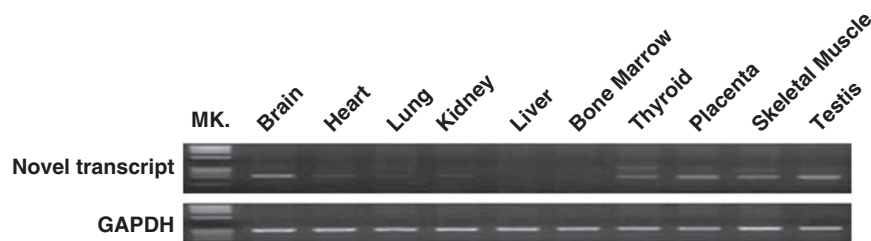

Figure 4 Expression of the inserted sequence. Reverse transcription-polymerase chain reaction (RT-PCR) products of a fragment of the inserted sequence from 10 human tissues (upper panel). A product of the expected size (206 bp) was observed clearly in the brain, thyroid, placenta, skeletal muscle and testis, but only faintly in the heart, lung and kidney. The validity of the amplified products was confirmed by sequencing (data not shown). No visible product was observed in the liver or bone marrow. The lower panel shows glyceraldehyde 3-phosphate dehydrogenase (GAPDH) RT-PCR products from the same samples. MK, DNA size marker.

As the inserted sequence corresponded to the 3′ end of the unknown transcript, we cloned the 5′ end of the transcript by 5′-RACE from the brain mRNA. This generated a single product with an additional 240 bp at the 5′ end of the inserted sequence (Figure 3). In the genome, this additional sequence was contiguous with the inserted fragment. Therefore, we concluded that the entire 452-bp region was expressed in the human brain. Examination of the genomic sequence upstream of the transcribed sequence revealed four TATA boxes (Figure 3). These results indicated that this region contains an intronless gene structure.

When the novel transcript was examined for its protein coding ability, the longest open-reading frame encoded 29 amino acids (MTVKWGKKTCPASISMMLLHHMKTEIFQF) (Figure 3). Homology searches for this peptide revealed no significantly homologous proteins and no significant domains. A strong consensus sequence for the translation initiation site[18] was not present in this frame. Taken together, these findings indicated no appreciable protein coding ability. The possibility of the transcript being a micro-RNA was examined by screening it against the miRBase database, but the results were negative. Therefore, this transcript is currently considered

a novel non-coding RNA transcribed from an apparently silent genomic region.

DISCUSSION

We identified an approximately 330-bp insertion at the ninth nucleotide of exon 67 of the *dystrophin* gene (Figure 1). Even though the enlarged exon 67 maintained its wild-type splicing consensus sequences at either end, the full sequence of exon 67 was skipped during splicing (Figure 2). As a result, the dystrophin mRNA would contain a premature stop codon within exon 68. We concluded that this insertion mutation causes DMD by inducing a secondary splicing error. The exon 67-skipping is likely owing to the enlarged exon size (approximately 480 bp) that escapes proper recognition by the splicing machinery, as has been reported previously.[11]

The identified insertion sequence had the hallmarks of a retrotransposon: an approximately 115-bp T nucleotide stretch that would be complementary to the poly(A) tract of the mRNA and 3-bp (TTC) target site duplications (Figure 1b). In addition, sequences at the insertion site within exon 67 were well matched to the consensus cleavage site for the L1 endonuclease (TTTTCAA) (Figure 1c). However, the inserted sequence did not encode any meaningful protein, including reverse transcriptase (Figure 3). We assume that the novel transcript was retrotransposed using autonomous L1 retrotranscriptase and endonuclease.[9] It has been shown that protein-coding mRNAs are occasionally reverse transcribed and integrated into genomic DNA, possibly as a by-product of L1 retrotransposition.[19] L1-encoded proteins bind to a processed cytoplasmic mRNA instead of L1 RNA. The abundance of cellular mRNAs and their 3′ poly(A) tails are thus thought to be the critical factors allowing mRNAs to take advantage of L1-encoded proteins for retrotransposition.[3]

It has been previously reported that the L1 retrotransposon machinery retrotransposed a partial *ATM* gene sequence from chromosomes 11 to 7, although no full-length L1 has been identified around the *ATM* gene.[20] Considering that the *ATM* gene is 2614 kb centromeric to the novel transcript, it is likely that the same L1 that retrotransposed the *ATM* gene also retrotransposed the novel non-coding gene into dystrophin.

A total of 118 disease events attributable to retrotransposons of L1s, Alus and SVAs have been reported to date, comprising 0.27% of all human mutations identified.[4] In the *dystrophin* gene, four retrotransposons have been identified to cause DMD, the largest being a 1400-bp L1 insertion.[12] Previously, one L1 insertion was identified in our Japanese patient.[11] This report increases the number of retrotransposon-related insertions to two out of the 442 identified mutations in Japan,[21] and we calculated the rate of retrotransposon-related insertion to be 0.47% of the mutations identified in Japanese dystrophinopathy. This higher incidence may be owing to a detection bias for mutations in the *dystrophin* gene on the X chromosome, which are more easily detected than mutations in autosomal genes.[5]

One non-autonomous retrotransposon insertion causing human disease has been reported in the *SLC25A13* gene, resulting in citrin deficiency.[10] The 2667-bp sequence from a gene on chromosome 6 (chromosome 6 open-reading frame 68) was found inserted into intron 16 of the *SLC25A13* gene. This insertion has a repetitive sequence (17 nt) derived from *SLC25A13* at both ends of the insert. Even though it was inserted within an intron, this insertion created a novel exon that included a stop codon and a poly(A) addition signal. The insertion was identified not only in the Japanese, but also in other East Asian populations such as the Chinese and Koreans. Therefore, this is most likely an ancient retrotransposition that occurred before the Japanese and Chinese became separated. In contrast, our insertion

has two novel characteristics: (1) the insertion occurred in the patient, indicating contemporary non-autonomous retrotranspositional activity and (2) the inserted sequence was a transcript from a region where no gene has been mapped.

As the novel transcript was expressed in the brain (Figure 4), it may be involved in brain function. Recently, it has been reported that normally quiescent 'jumping genes' can be activated in neural progenitor cells.[22] The novel transcript may be one of these quiescent genes, although its expression is probably under the control of a TATA box (Figure 3). Further studies are required to elucidate the physiological role of this novel non-coding gene.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1   Deininger, P. L. & Batzer, M. A. Mammalian retroelements. *Genome Res.* **12,** 1455–1465 (2002).

2   Goodier, J. L. & Kazazian, Jr. H. H. Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* **135,** 23–35 (2008).

3   Chen, J. M., Stenson, P. D., Cooper, D. N. & Ferec, C. A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. *Hum. Genet.* **117,** 411–427 (2005).

4   Callinan, P. A. & Batzer, M. A. Transposable elements and human disease. *Genome Dyn.* **1,** 104–115 (2006).

5   Belancio, V. P., Deininger, P. L. & Roy-Engel, A M. LINE dancing in the human genome: transposable elements and disease. *Genome Med.* **1,** 97 (2009).

6   Ostertag, E. M. & Kazazian, Jr. H. H. Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* **35,** 501–538 (2001).

7   Chen, J. M., Ferec, C. & Cooper, D. N. LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease: mutation detection bias and multiple mechanisms of target gene disruption. *J. Biomed. Biotechnol.* **2006,** 56182 (2006).

8   Drouin, G. Processed pseudogenes are more abundant in human and mouse X chromosomes than in autosomes. *Mol. Biol. Evol.* **23,** 1652–1655 (2006).

9   Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24,** 363–367 (2000).

10 Tabata, A., Sheng, J. S., Ushikai, M., Song, Y. Z., Gao, H. Z., Lu, Y. B. *et al.* Identification of 13 novel mutations including a retrotransposal insertion in SLC25A13 gene and frequency of 30 mutations found in patients with citrin deficiency. *J. Hum. Genet.* **53,** 534–545 (2008).

11 Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Ishikawa, Y., Minami, R. *et al.* Insertion of a 5′ truncated L1 element into the 3′ end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.* **91,** 1862–1867 (1993).

12 Holmes, S. E., Dombroski, B. A., Krebs, C. M., Boehm, C. D. & Kazazian, Jr. H. H. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat. Genet.* **7,** 143–148 (1994).

13 Yoshida, K., Nakamura, A., Yazaki, M., Ikeda, S. & Takeda, S. Insertional mutation by transposable element, L1, in the DMD gene results in X-linked dilated cardiomyopathy. *Hum. Mol. Genet.* **7,** 1129–1132 (1998).

14 Musova, Z., Hedvicakova, P., Mohrmann, M., Tesarova, M., Krepelova, A., Zeman, J. *et al.* A novel insertion of a rearranged L1 element in exon 44 of the dystrophin gene: further evidence for possible bias in retroposon integration. *Biochem. Biophys. Res. Commun.* **347,** 145–149 (2006).

15 Tran, V. K., Takeshima, Y., Zhang, Z., Yagi, M., Nishiyama, A., Habara, Y. *et al.* Splicing analysis disclosed a determinant single nucleotide for exon skipping caused by a novel intra-exonic four-nucleotide deletion in the dystrophin gene. *J. Med. Genet.* **43,** 924–930 (2006).

16 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).

17 Tran, V. K., Takeshima, Y., Zhang, Z., Habara, Y., Haginoya, K., Nishiyama, A. *et al.* A nonsense mutation-created intraexonic splice site is active in the lymphocytes, but not in the skeletal muscle of a DMD patient. *Hum. Genet.* **120,** 737–742 (2007).

18 Kozak, M. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15,** 8125–8148 (1987).

19 Ding, W., Lin, L., Chen, B. & Dai, J. L1 elements, processed pseudogenes and retrogenes in mammalian genomes. *IUBMB Life* **58,** 677–685 (2006).

20 Ejima, Y. & Yang, L. Trans mobilization of genomic DNA as a mechanism for retro-transposon-mediated exon shuffling. *Hum. Mol. Genet.* **12,** 1321–1328 (2003).

21 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).

22 Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460,** 1127–1131 (2009).