

© 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/> The published version is available via <https://doi.org/10.1016/j.gene.2018.07.030>

## Title page

### Use of modified U1 small nuclear RNA for rescue from exon 7 skipping caused by 5'-splice site mutation of human cathepsin A gene

Naoshi Yamazaki <sup>a, b, \*</sup>, Keisuke Kanazawa <sup>a</sup>, Maria Kimura <sup>a</sup>, Hironobu Ike <sup>b</sup>, Makiko Shinomiya <sup>a</sup>, Shouko Tanaka <sup>a</sup>, Yasuo Shinohara <sup>a, b, c</sup>, Noriaki Minakawa <sup>a, b</sup>, Kohji Itoh <sup>a, b</sup>, and Yoshiharu Takiguchi <sup>a, b</sup>

a Faculty of Pharmaceutical Sciences, Tokushima University, Shomachi 1-78, Tokushima 770-8505, Japan

b Graduate School of Pharmaceutical Sciences, Tokushima University, Shomachi 1-78, Tokushima 770-8505, Japan

c Institute of Advanced Medical Sciences, Tokushima University, Kuramoto 3-18, Tokushima 770-8503, Japan

\* Corresponding Author:

Naoshi Yamazaki, Ph. D. and Associate Professor

Address: Shomachi 1-78, Tokushima 770-8505, Japan

E-mail: [nyamazaki@tokushima-u.ac.jp](mailto:nyamazaki@tokushima-u.ac.jp)

Tel: +81-88-633-9516

Declarations of interest: none

## Abstract

Cathepsin A (CTSA) is a multifunctional lysosomal enzyme, and its hereditary defect causes an autosomal recessive disorder called galactosialidosis. In a certain number of galactosialidosis patients, a base substitution from adenine to guanine is observed at the +3 position of the 7th intron (IVS7 +3a>g) of the *CTSA* gene. With this mutation, a splicing error occurs; and consequently mRNA lacking the 7th exon is produced. This skipping of exon 7 causes a frame shift of the transcripts, resulting in a non-functional CTSA protein and hence galactosialidosis. This mutation seems to make the interaction between the 5'-splice site of intron 7 of pre-mRNA and U1 small nuclear RNA (U1 snRNA) much weaker. In the present study, to produce properly spliced mRNA from the *CTSA* gene harboring this IVS7 +3a>g mutation, we examined the possible usefulness of modified U1 snRNA that could interact with the mutated 5'-splice site. Toward this goal, we first prepared a model system using a mutant *CTSA mini gene* plasmid for delivery into HeLa cells. Then, we examined the effectiveness of modified U1 snRNA on the formation of properly spliced mRNA from this mutant *CTSA mini gene*. As a result, we succeeded in obtaining improved formation of properly spliced CTSA mRNA. Our results suggest the usefulness of modified U1 snRNA for rescue from exon 7 skipping caused by the IVS7 +3a>g mutation of the *CTSA* gene.

### **Key words**

exon skipping, galactosialidosis, RNA splicing, splice defect

### **Abbreviations**

bp, base pair(s); cDNA, complementary DNA; CTSA, cathepsin A; EGFP, enhanced green fluorescent protein; kbp, kilobase pairs; kDa, kilodaltons; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; pre-mRNA, mRNA precursor; rAAV, recombinant adeno-associated virus; RT-PCR, reverse transcription-PCR; U1 snRNA, U1 small nuclear RNA; SD, standard deviation; SDS, sodium dodecyl sulfate; ss, splice site; ss-cDNA, single-stranded complementary DNA

## **1. Introduction**

In the various processes involved in eukaryotic gene expression, the splicing of nuclear mRNA precursor (pre-mRNA) is one of the critical steps of RNA processing. During this splicing process, introns are precisely excised at the exon/intron boundaries, and adjacent exons are joined together to form the mature mRNA. This splicing reaction is initiated by the association of U1 small nuclear RNA (U1 snRNA) with the 5'-splice site (ss) of pre-mRNA [Krämer et al., 1984; Zhuang and Weiner, 1986; Roca et al., 2013]. In general, the third to eleventh bases of U1 snRNA (5'-ACUUACCUG) bind to the 5'-ss of pre-mRNA by RNA-RNA base pairing. The consensus sequence of the 5'-ss was reported to be 5'-MAG/GURAGU, with "/" indicating the boundary between exon and intron,

and "M" and "R" representing A or C and A or G, respectively [Mount, 1982]. In many cases, however, the nucleotide sequences of the 5'-ss do not completely match with the above consensus sequence. Even though not completely conserved, such a 5'-ss can be recognized by U1 snRNA, and the splicing reaction is initiated. However, if the degree of conservation of the 5'-ss is too low, the association of U1 snRNA with pre-mRNA does not properly occur, and a splice defect can thus be induced [Faustino and Cooper, 2003; Wang and Cooper, 2007; Havens et al., 2013].

Cathepsin A (EC 3.4.16.5), hereafter abbreviated as CTSA, is a multifunctional lysosomal enzyme showing the catalytic activities of deamidase, esterase, and carboxypeptidase. In addition, this enzyme is essential for the stabilization of  $\beta$ -galactosidase and the activation of neuraminidase. Therefore, a genetic defect of CTSA causes a significant decrease in the activities of  $\beta$ -galactosidase and neuraminidase, resulting in an autosomal recessive lysosomal storage disorder called galactosialidosis [Shimmoto et al., 1993; Hiraiwa, 1999; Ketterer et al., 2017]. About 80 patients with galactosialidosis have been reported globally. Based on the age of onset and severity of their symptoms, patients are classified as early infantile type, late infantile type, or juvenile/adult type. More than 60% of these patients have been reported in Japan, and most of them are of the juvenile/adult type [Hossain et al., 2016].

The gene for human CTSA is located on 20q13.1 [Wiegant et al., 1991]. It consists of 15 exons (one 5'-non-coding exon and 14 coding

exons) spanning approx. 7.5 kilobase pairs (kbp) and encodes a 480 amino acid-inactive precursor protein [Shimmoto et al., 1996]. In the human *CTSA* gene, more than 10 missense mutations have been reported to cause galactosialidosis [OMIM®, 2009]. In addition, an adenine to guanine base substitution at the +3 position of the 7th intron (NG\_008291.1:g.7363A>G) is frequently found in juvenile/adult-type Japanese galactosialidosis patients (homozygous for this mutation or heterozygous with the other allele being a missense mutation) [Shimmoto et al., 1990, 1993]. In this manuscript, this mutation is referred to as IVS7 +3a>g. With such a mutation, mature mRNA lacking the 7th exon (SpDEX7) is formed. Skipping of the 7th exon causes a frame shift in the transcripts, and thus functional mature CTSA protein is not produced. As a result, the IVS7 +3a>g mutation of the *CTSA* gene causes galactosialidosis. As treatment for galactosialidosis, transplantation of cells overexpressing human CTSA [Zhou et al., 1995; Hahn et al., 1998] and enzyme replacement therapy [Itoh et al., 2016] are currently being investigated. In addition, the injection of a recombinant adeno-associated virus (rAAV) vector expressing human CTSA under the control of a liver-specific promoter was reported to be effective for galactosialidosis treatment of CTSA-deficient mouse [Hu et al., 2012]; and a low-molecular-weight chemical chaperone was reported to increase the endogenous  $\beta$ -galactosidase activity in CTSA-deficient cells [Hossain et al., 2016]. However, for the development of new therapies for human galactosialidosis, a novel

approach for treatment of this disorder is necessary.

In the IVS7 +3a>g mutation of the *CTSA* gene, it is predicted that U1 snRNA cannot interact with the 5'-ss of the 7th intron of the pre-mRNA; and, hence, skipping of the 7th exon occurs. To negate the splice defect caused by the 5'-ss mutation, the application of a U1 snRNA of which its 5'-end was modified to enable association with the mutated 5'-ss was reported [Havens et al., 2013]. In the present study, we constructed a model experimental system using a *CTSA mini gene* harboring the IVS7 +3a>g mutation and examined possible modified U1 snRNA-mediated rescue from the splice defect of the human *CTSA* gene caused by this mutation.

## **2. Materials and methods**

### **2.1. Construction of expression plasmids encoding U1 snRNAs**

All recombinant DNA experiments were performed according to the guidelines of Tokushima University.

A genomic DNA fragment encoding the human *U1 snRNA* gene (accession number V00591), from its own promoter to the transcription termination signal, i.e., nucleotide sequence from -393 to +199 (nucleotide sequence numbered taking the 5'-end of mature U1 snRNA as +1) [Murphy et al., 1982], was prepared from human genomic DNA (Clontech, Palo Alto, CA, USA) by use of the polymerase chain reaction (PCR). The obtained DNA fragment was subcloned into pUC19. In this paper, the expression plasmid encoding wild-type U1 snRNA is referred to as the wild-type U1 plasmid. The expression plasmids of modified U1 snRNA were prepared by PCR using

mutated oligonucleotide primers and the wild-type U1 plasmid as a template. Details of the plasmid construction are summarized in **Supplemental methods**.

## **2.2. Construction of *CTSA mini gene* plasmids**

Genomic DNA fragments encoding the human *CTSA gene* (accession number NG\_008291.1) were isolated by PCR using human genomic DNA (Clontech) as a template. Two regions, from exon 6 to exon 8, and from exon 3 to exon 11, were amplified with the primer pairs of primer 6 (sense) and primer 8 (antisense), and primer 3 (sense) and primer 11 (antisense), respectively. The obtained DNA fragment was subcloned into pcDNA3.1/Hygro(-) mammalian expression plasmid (Invitrogen, Carlsbad, CA, USA), in which DNA fragment encoding CTSA was sandwiched between those encoding enhanced green fluorescent protein (EGFP, Clontech) and c-Myc tag. These mini gene plasmids enabling transcription of the pre-mRNA of parts of the *CTSA gene* are referred to as wild-type *CTSA mini gene* plasmids ("exons 6-8" or "exons 3-11"). Details of the plasmid construction are summarized in **Supplemental methods** and **Supplemental results Fig.S1A**.

For construction of the *CTSA mini gene* plasmid harboring the IVS7 +3a>g mutation, the overlap extension PCR method was carried out with mutated oligonucleotide primers [Ho et al., 1989] using a wild-type *CTSA mini gene* plasmid as a template. The plasmids including the mutation were referred to as mutant *CTSA mini gene* plasmids ("exons 6-8" or "exons

3-11"). The detailed nucleotide sequence of *CTSA mini gene* plasmid (exons 3-11) is shown in **Supplemental results Fig.S1B and C**.

### **2.3. Cell culture and transfection**

HeLa cells were cultured in Minimum Essential Medium (Sigma-Aldrich, St. Louis, MO, USA) containing 10% fetal bovine serum (MP Biomedicals, Santa Ana, CA, USA) and 1% MEM Non-essential Amino Acids (Sigma-Aldrich). COS7 cells were cultured in Dulbecco's Modified Eagle's Medium (Sigma-Aldrich) containing 10% fetal bovine serum. Eighteen to 24 hrs before transfection,  $3.5 \times 10^5$  cells were suspended in 2 mL of culture medium and plated in each well of a 6-well plate.

For RNA analysis (HeLa cells), basically, a 1- $\mu$ g aliquot of the *CTSA mini gene* plasmid and a 2- $\mu$ g aliquot of U1 plasmid were mixed with 3  $\mu$ L of X-treme GENE HP (Roche Diagnostics, Indianapolis, IN, USA) in 100  $\mu$ L/well Opti-MEM (Invitrogen) and added to each culture well. For protein analysis (COS7 cells), a 2- $\mu$ g aliquot of the *CTSA mini gene* plasmid, a 1- $\mu$ g aliquot of U1 plasmid, and a 0.3- $\mu$ g aliquot of EGFP expression plasmid (only EGFP-coding region was inserted into pcDNA3.1/Hygro(-) plasmid) were mixed with 5  $\mu$ L of Lipofectamine 2000 (Invitrogen) in 500  $\mu$ L/well Opti-MEM and added to each well.

After incubation for 24 hrs, total RNA was prepared from HeLa cells by use of an RNeasy Plus Kit (QIAGEN, Hilden, Germany) or ISOGEN II (Nippon Gene, Tokyo, Japan) according to the methods recommended by the suppliers. For protein analysis, COS7 cells were dissolved in RIPA



buffer (Nacalai Tesque, Kyoto, Japan) supplemented with 0.1% sodium dodecyl sulfate (SDS), and their lysates were prepared.

#### **2.4. Reverse transcription-PCR (RT-PCR)**

To obtain single-stranded complementary DNA (ss-cDNA), total RNA from HeLa cells was reversely transcribed with oligo(dT) primer. For specific amplification of the transcripts derived from the *CTSA mini gene* plasmid, ss-cDNA was amplified by 2 rounds of PCR. Briefly, ss-cDNA corresponding to the region sandwiched between the T7 promoter and BGH polyadenylation signal region of the pcDNA3.1/Hygro(-) plasmid was first amplified by using the primer pair of T7 primer (sense) and BGH pA primer (antisense). Then, the reaction mixture was diluted with distilled water. To obtain the complementary DNA (cDNA) fragment corresponding to the region from exon 6 to exon 8 of the human *CTSA gene*, we amplified a diluted reaction mixture by using the primer pair of primer 6 and primer 8. The reaction mixtures of second PCR were subjected to polyacrylamide gel electrophoresis (PAGE) with DNA size marker (pUC19 digested with *MspI*). The gel was stained with ethidium bromide and visualized by use of an LAS500 (Fujifilm, Tokyo, Japan). The band intensity was quantified with Multi Gauge 3.2 (Fujifilm), and the degree (%) of exon 7 inclusion was expressed as the mean  $\pm$  standard deviation (SD) from at least 3 different transfection experiments.

As a check for the transcripts in normal human cells, samples of poly(A)<sup>+</sup> RNA of human liver (Clontech) and total RNA of HeLa cells

were reversely transcribed with oligo(dT) primer, and the cDNA fragment corresponding to the region from exon 6 to exon 8 of the human *CTSA* gene was amplified by using the primer pair of primer 6 and primer 8. Details are summarized in **Supplemental methods**.

## **2.5. Western blot analysis**

SDS-PAGE and Western blot analysis were carried out as described previously [Matsuo et al., 2010], using lysates of COS7 cells transfected with plasmids. The protein concentration of the lysate was measured with a BCA protein assay kit (Thermo Scientific, Rockford, IL, USA) using bovine serum albumin as the standard protein. A 20- $\mu$ g protein sample was mixed with an equal volume of 2 $\times$ SDS-sample buffer and then loaded onto a 10% polyacrylamide gel with WIDE-VIEW Pre-stained Protein Size Marker III (Wako Pure Chemical Industries, Osaka, Japan). Proteins in the gel were transferred onto a nitrocellulose membrane (GE Healthcare, Buckinghamshire, England), after which the membrane was incubated in Tris-buffered saline supplemented with 0.05% Tween-20, 3% nonfat dry milk, and anti-EGFP antibody (MBL, Nagoya, Japan) or anti-Myc tag antibody (Cell Signaling Technology, Danvers, MA, USA). To detect anti-EGFP or anti-Myc tag antibodies on the membrane, anti-rabbit IgG (Cell Signaling Technology) or anti-mouse IgG (Sigma-Aldrich) conjugated with horse radish peroxidase, respectively, were used. The immunoreactive signals were visualized by use of the ECL detection system (GE Healthcare) and LAS500.

## 2.6. Calculation of the splice site score

Splice site score of the 5'-ss of intron 7 of the *CTSA* gene was calculated using splice site prediction programs; Human Splicing Finder [Desmet et al., 2009] (<http://www.umd.be/HSF3/>), MaxEntScan [Yeo and Burge, 2004] (<http://www.umd.be/HSF3/>), NNSPLICE 0.9 [Reese et al., 1997] ([http://www.fruitfly.org/seq\\_tools/splice.html](http://www.fruitfly.org/seq_tools/splice.html)).

## 3. Results

### 3.1. Construction of model system mimicking the IVS7 +3a>g mutation of the *CTSA* gene

The nucleotide sequence of the 5'-end of wild-type U1 snRNA and those of the 5'-ss of intron 7 of *CTSA* pre-mRNAs are depicted in **Fig. 1A**. As mentioned, in general, the third to eleventh bases of the wild-type U1 snRNA bind to the 5'-ss of pre-mRNA. The 5'-ss of the intron 7 of pre-mRNA transcribed from the wild-type *CTSA* gene has 2 nucleotides non-complementary with the 5'-end of U1 snRNA. Nevertheless, it seems to be properly recognized by wild-type U1 snRNA. This interpretation was supported by our experimental results: when we examined whether the transcript encoded by the wild-type *CTSA* gene was precisely spliced in human healthy tissue (liver) and HeLa cells, almost all transcripts were found to be correctly spliced in a manner containing exon 7 ("exons 6-8") as clearly demonstrated in **Fig. 1B**. Although a faint band indicating the skipping of exon 7 was observed in the sample of HeLa cells ("exon

6 and exon 8"), its level was far from that of the ordinary transcript. We do not have any information as to how it was formed. On the contrary, in the case of the IVS7 +3a>g mutation, the 5'-ss of intron 7 of pre-mRNA has 3 nucleotides non-complementary with the 5'-end of U1 snRNA. Thus, this mutation is thought to cause a splice defect at this exon/intron boundary, forming mature mRNA lacking exon 7 [Shimmoto et al., 1990, 1993].

Next, we established a model system mimicking the *CTSA* gene harboring the IVS7 +3a>g mutation. For this purpose, we chose to use HeLa cells, because this cell line is often used for splicing analysis with mini genes and produces the correctly spliced mRNA from the endogenous *CTSA* gene (**Fig. 1B**). As for the mini gene, we first prepared small mini gene plasmids carrying just exon 6 to exon 8 of the wild-type *CTSA* gene and one having the mutation (see **Fig. 1C**). As shown in **Fig. 1D**, when HeLa cells were transfected with mutant *CTSA* mini gene plasmid, we observed no signal corresponding to the RNA containing exon 7 ("exons 6-8"). When HeLa cells were transfected with wild-type *CTSA* mini gene plasmid, we observed a weak signal corresponding to the RNA containing exon 7. These results indicate that introduction of mini gene plasmids encoding the nucleotide sequence from exon 6 to exon 8 was not suitable to mimic splicing properties of wild-type and mutant *CTSA* gene.

Thus, we next prepared much longer mini gene plasmids carrying exon 3 to exon 11 of the wild-type *CTSA* gene and its mutant. When HeLa cells were transfected with these plasmids and examined for the expressed

RNA species, this system was found to be much better; i.e., cells transfected with the mini gene plasmid carrying the mutation only showed the signal corresponding to the RNA lacking exon 7 ("exon 6 and exon 8"), and those with the wild-type mini gene plasmid showed a strong signal corresponding to the RNA containing exon 7. Although the ratio of exon 7 inclusion in cells introduced with the wild-type mini gene plasmid containing exon 3 to exon 11 varied in different transfection experiments, it was always significantly higher than that obtained in the case of the wild-type mini gene plasmid containing exon 6 to exon 8. This result was not surprising, because the use of mini genes was reported not always to reflect the splicing pattern of the endogenous gene, probably due to lack of the necessary genomic sequence context [Baralle et al., 2006; Sánchez-Alcudia et al., 2011]. This finding suggests that splicing of intron 7 of CTSA was influenced not only by the proximal but also the distal region of the gene. Based on these observations, we concluded that the second mini gene plasmid, i.e., the one carrying exon 3 to exon 11 of the IVS7 +3a>g mutation, would be more suitable for evaluation of the effects of modified U1 snRNAs.

It should be noted that we often observed an extra signal of amplified DNA band (designated as "X" in **Fig. 1D**) between two DNA bands of "exons 6-8" and "exon 6 and exon 8". We considered the possibility that there was an RT-PCR product having an insertion of "gt" nucleotides between exon 6 and exon 8 (this product is designated as "exon 6-gt-exon 8") and that this band "X" might contain heteroduplex DNA comprising

the "exon 6 and exon 8" and "exon 6-gt-exon 8". Detailed characterization of band "X" is summarized in **Supplemental results, Fig. S2A, B, C, and D.**

### **3.2. Screening of modified U1 snRNA effective for proper splicing of pre-mRNA encoded by the human CTSA gene having the IVS7 +3a>g mutation**

*In silico* analysis using splice site prediction programs suggested that the splicing score is slightly decreased by IVS7 +3a>g mutation (from 89.26 to 88.10 according to Human Splicing Finder; from 9.99 to 6.34 according to MaxEntScan; and from 0.95 to 0.57 according to NNSPLICE 0.9). Although the IVS7 +3a>g mutation shows only 1 base substitution, there is a possibility that the wild-type U1 snRNA becomes incapable of associating with the pre-mRNA having the IVS7 +3a>g mutation, and thus exon 7 is skipped. Thus, to enable the formation of RNA/RNA hybrids between U1 snRNA and 5'-ss of intron 7, we first expected that the 6th uracil base of U1 snRNA, which should form a base pair with mutated guanine base at the 5'-ss of intron 7, should be reversely mutated to a cytosine base (hereafter, this modified U1 snRNA is referred to as 6th U>C; **Fig. 2A, upper panel**). To examine the effectiveness of this modified U1 snRNA (6th U>C) for rescue from the splice defect, we prepared the expression plasmid of this U1 snRNA. Then, HeLa cells were transfected with plasmids encoding exon 3 to exon 11 of mutant *CTSA mini gene* and those encoding modified U1 snRNA (6th U>C). After harvesting of the cells, expressed RNA was analyzed by RT-PCR. However, this

modification was not effective for rescue from the splice defect (**Fig. 2A, lower panel**).

Based on the above result, we thought that a much stronger interaction between U1 snRNA and 5'-ss would be required for rescue from the splice defect caused by the mutation. Thus, we designed a modified U1 snRNA, referred to as modified-1, having a nucleotide sequence completely complementary to that of the mutated 5'-ss (**Fig. 2B, upper panel**). As a result, this modified-1 showed a moderate rescue (**Fig. 2B, lower panel**). Next, we sought the mostly effective modified U1 snRNA by serial substitution of the nucleotide. The 5'-nucleotide of the modified-1, i.e., U base, was serially substituted with another nucleotide (**Fig. 2B, upper panel**). When we examined the effectiveness of these modified U1 snRNAs (modified-1, -2, -3, and -4) in rescuing cells from the splice defect, modified-4 was found to be the most effective (**Fig. 2B, lower panel**). Then, by fixing the first nucleotide as "A", the second nucleotide was serially substituted (modified-4, -5, -6, and -7). Likewise, after optimization of the second nucleotide, the third nucleotide was also serially substituted (modified-7, -8, -9, and -10). From the band intensity, exon 7 inclusion was evaluated as shown in **Fig. 2C**. By performing these nucleotide substitution studies, we found that all modified U1 snRNAs except for 6th U>C increased the exon 7 inclusion; and modified-7 U1 snRNA was shown to have the relatively highest activity in rescuing the splice defect caused by the IVS7 +3a>g mutation. The dose dependency of the modified-7 U1 expression plasmid

was determined as shown in **Fig. 2D**. Experimentally, it was difficult to distinguish wild-type U1 snRNA (endogenous) and modified-7 U1 snRNA, because their nucleotide sequences were completely the same except for only 3 bases at their 5'-end; and also the expression level of endogenous U1 snRNA would be much higher than that of modified-7 U1 snRNA. Therefore, at this stage, we did not check the expression level of modified-7 U1 snRNA and did not estimate the ratio of wild-type U1 snRNA and modified-7 U1 snRNA and that of modified-7 U1 snRNA and pre-mRNA transcribed from the *CTSA mini gene* plasmid. However, the amount of properly spliced mRNA was shown to increase in accordance with the amount of modified-7 U1 expression plasmid transfected into the cells (**Fig. 2D**). Detailed characterizations, including the effect on exon 7 splicing of the wild-type *CTSA* gene, of the modified-7 U1 expression plasmid are shown in **Supplemental results, Fig. S3A, B, C, and D**.

### **3.3. Modified U1 snRNA was also effective for the synthesis of the corresponding protein**

We further examined whether the rescue effect of modified-7 U1 snRNA could be also observed at the level of protein expression. The *CTSA mini gene* plasmid that we used was designed to be applicable for studies on protein synthesis; i.e., the 5' and 3'-ends of the nucleotide sequence of mini gene were flanked with nucleotide sequences encoding EGFP and c-Myc tag, respectively (see **Fig. 1C**). If the nucleotide sequences corresponding to exons 3-11 of *CTSA gene* were properly spliced,



the transcript formed would code for a 63-kilodaltons (kDa) protein composed of EGFP, the part of CTSA encoded by exon 3 to exon 11, and c-Myc tag (for detailed amino acid sequences, see **Supplemental results, Fig. S1D**). As exon 7 contains 92 nucleotides (**Supplemental results, Fig. S1C**), skipping of exon 7 causes a frame shift and the stop codon arises by another frame in exon 8. In this case, a 45-kDa protein without the Myc tag was produced.

To examine whether the modified-7 U1 snRNA would be effective for formation of this ideal protein, we used COS7 cells because of their high protein expression from pcDNA3.1 plasmid. COS7 cells were transfected with *CTSA mini gene* plasmid and U1 plasmid, and then cell lysates were prepared and subjected to immunodetection. As shown in **Fig. 3**, when COS7 cells were transfected with the wild-type *CTSA mini gene* plasmid and wild-type U1 plasmid, a 63-kDa protein (shown by arrow) was detected by both anti-EGFP and anti-Myc antibodies. This protein was not detected when mutant *CTSA mini gene* plasmid and wild-type U1 plasmid were used. On the other hand, when the cells were transfected with mutant *CTSA mini gene* plasmid and modified-7 U1 plasmid, the 63-kDa protein was detected by both antibodies. The observed differences in the expression levels of EGFP-CTSA-Myc protein among these samples did not reflect the differences in the transfection efficiency of COS7 cells, because the expression levels of EGFP (27-kDa), co-transfected to estimate the transfection efficiency, were almost the same among the samples (**Fig. 3**, 27-kDa EGFP). Based on these results, we concluded

that the modified-7 U1 snRNA designed in the present study was not only effective for proper splicing of the mRNA coded by the *CTSA* gene having the IVS7 +3a>g mutation, but also for the synthesis of the target protein.

The smaller protein observed in both EGFP and Myc panels for cells transfected with the mini gene plasmid (shown by broken arrows) might have been translated from a splice variant. Although we could not identify the corresponding transcript, a possible explanation about this protein is given in **Supplemental results, Fig. S2D**.

#### 4. Discussion

For the correct pre-mRNA splicing, the 5'-ss, the 3'-ss including the polypyrimidine tract, the branch point located upstream of the 3'-ss, and intronic and/or exonic splice regulatory elements are required. Generally, the base pairing between U1 snRNA and the 5'-ss of pre-mRNA is necessary for the initiation of the splicing reaction, although U1-independent pre-mRNA splicing has been reported [Lund and Kjems, 2002; Fukumura et al., 2009]. For initiation of the splicing reaction, the 5'-end of U1 snRNA, 11 bases as maximum, interacts with the 5'-ss of pre-mRNA by making base pairs [Roca et al., 2013]. Because the nucleotides at positions +7 and +8 of the 5'-ss are not consistently conserved in human genes, the 9 bases from -3 to +6 of the 5'-ss (consensus sequence of 5'-MAG/GURAGU) are generally recognized by complementary bases from the third to eleventh position of U1 snRNA. However, if the degree of complementation is decreased by mutation at the 5'-ss, U1 snRNA

may not properly anneal with the pre-mRNA, thus causing a splice defect.

In the case of the human *CTSA gene*, the adenine to guanine base substitution at the +3 position of intron 7 (IVS7 +3a>g) causes skipping of exon 7 in its mRNA. Both nucleotides adenine and guanine at the +3 position match the consensus nucleotide sequence of 5'-ss; and, hence, the nucleotide substitution from adenine to guanine seemed not to influence the splicing reaction. However, this substitution caused a splice defect (**Fig. 1D**). Similarly, a splice defect due to nucleotide substitution of adenine for guanine at the +3 position has been reported to occur in several genes [Ohno et al., 1999; Madsen et al., 2006; Le Guédard-Méreuze et al., 2009; Sánchez-Alcudia et al., 2011]; and the relationship between the nucleotide sequence of this region and degree of splice defect may be summarized as follows [Ohno et al., 1999]: In the 5'-ss having the guanine at the +3 position, but showing proper splicing, there tends to be a consensus sequence at +4 to +6. Inversely, in the 5'-ss having the adenine at the +3 position, and showing the splice defect, there tends to be a non-consensus sequence at +4 to +6. In the case of the 5'-ss of intron 7 of the human *CTSA gene*, nucleotides from +4 to +6 (or +4 to +8) also showed a poor association with the wild-type U1 snRNA, as was depicted in **Fig. 1A**. There is no evidence, but we think that the long intron caused by the skipping of exon 7 does not have any physiological role and that it is degraded as an ordinary intron.

To rescue the splice defect due to mutation at the 5'-ss, trials using modified U1 snRNA have been reported for some diseases. Therefore,

we examined the effect of modified U1 snRNA on the splicing of pre-mRNA encoded by the *CTSA* gene having the IVS7 +3a>g mutation. All modified U1 snRNAs, except for the 6th U>C, were effective in proper splicing of pre-mRNA encoded by the gene having the mutation (**Fig. 2C**). At this stage, the reasons causing the difference in the proper splicing effects among the modified U1 snRNAs remain uncertain. Because exon 7 inclusion was not completely achieved in the cells transfected with a mini gene even if it was a wild-type one (**Fig. 1D**), the estimated effect of modified U1 snRNAs on correct splicing might be low.

For *in vivo* application for rescue from a splice defect, modified U1 snRNA constructed in a plasmid or rAAV vector has been examined by using mice having a 5'-ss mutation of the coagulation *F7* gene [Balestra et al., 2014] and the *dopa decarboxylase* gene [Lee et al., 2016]. As a result, although exon inclusions by modified U1 snRNA were observed, an off-target effect by modified U1 snRNA was suggested. To improve specificity and reduce the potential off-target effects, exon-specific U1 snRNA (ExSpeU1), targeting non-conserved intronic sequences downstream of the 5'-ss, was developed earlier [Fernandez Alanis et al., 2012]. Intraperitoneal injection of rAAV9 containing an ExSpeU1 expression unit was shown to increase the amount of mRNA including exon 7 of the *survival motor neuron 2* gene in brain, heart, kidney, liver, and skeletal muscle of transgenic mice [Dal Mas et al., 2015]. As galactosialidosis is a condition that affects many areas of the body including the brain, systemic rAAV-mediated gene delivery for modified

U1 snRNA or ExSpeU1 could be effective for treatment of galactosialidosis caused by the *CTSA* gene harboring the IVS7 +3a>g mutation.

In cells derived from galactosialidosis patients who were homozygous for the IVS7 +3a>g *CTSA* gene allele, CTSA enzyme activity was not detected [Shimmoto et al., 1993]. In some lysosomal storage diseases, it was suggested that 2-10% of normal enzyme activity was required to rescue cells from the metabolic defects [Hu et al., 2012]. In our present study, although we did not examine the recovery of the enzyme activity, the expression of appropriate protein synthesized from the mutant mini gene plasmid was apparently observed when the modified U1 snRNA was used. To investigate whether modified U1 snRNA could be a potential medicinal tool for the treatment of galactosialidosis caused by the *CTSA* gene harboring the IVS7 +3a>g mutation, further analysis using patient-derived cells and animal models are necessary.

#### **Acknowledgements**

This work was supported by JSPS KAKENHI Grant Number JP16K08235. N.Y. thanks the research program for development of intelligent TOKUSHIMA-artificial Exosome (iTEX) from Tokushima University. The authors would like to thank Ms. Asaka Kaneda, Mr. Akira Ama, Mr. Yuji Fujiwara, Ms. Yuki Asakura, Ms. Akari Saito, Ms. Rina Takahashi, Ms. Kanae Koide, and Mr. Seitaro Takaishi (Faculty of Pharmaceutical Sciences, Tokushima University) for their support in conducting the experiments.

### Disclosure statement

The authors have no conflict of interest to declare for this manuscript.

### References

- Baralle M, Skoko N, Knezevich A, De Conti L, Motti D, Bhuvanagiri M, Baralle D, Buratti E, Baralle FE. 2006. NF1 mRNA biogenesis: effect of the genomic milieu in splicing regulation of the NF1 exon 37 region. *FEBS Lett* 580:4449-4456.
- Balestra D, Faella A, Margaritis P, Cavallari N, Pagani F, Bernardi F, Arruda VR, Pinotti M. 2014. An engineered U1 small nuclear RNA rescues splicing-defective coagulation F7 gene expression in mice. *J Thromb Haemost* 12:177-185.
- Dal Mas A, Rogalska ME, Bussani E, Pagani F. 2015. Improvement of SMN2 pre-mRNA processing mediated by exon-specific U1 small nuclear RNA. *Am J Hum Genet* 96:93-103.
- Desmet FO, Hamroun D, Lalande M, Collod-Bérout G, Claustres M, Bérout C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37:e67.
- Faustino NA, Cooper TA. 2003. Pre-mRNA splicing and human disease. *Genes Dev* 17:419-437.
- Fernandez Alanis E, Pinotti M, Dal Mas A, Balestra D, Cavallari N, Rogalska ME, Bernardi F, Pagani F. 2012. An exon-specific U1 small

- nuclear RNA (snRNA) strategy to correct splicing defects. *Hum Mol Genet* 21:2389-2398.
- Fukumura K, Taniguchi I, Sakamoto H, Ohno M, Inoue K. 2009. U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. *Nucleic Acids Res* 37:1907-1914.
- Hahn CN, del Pilar Martin M, Zhou XY, Mann LW, d'Azzo A. 1998. Correction of murine galactosialidosis by bone marrow-derived macrophages overexpressing human protective protein/cathepsin A under control of the colony-stimulating factor-1 receptor promoter. *Proc Natl Acad Sci U S A* 95:14880-14885.
- Havens MA, Duelli DM, Hastings ML. 2013. Targeting RNA splicing for disease therapy. *WIREs RNA* 4:247-266.
- Hiraiwa M. 1999. Cathepsin A/protective protein: an unusual lysosomal multifunctional protein. *Cell Mol Life Sci* 56:894-907.
- Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77:51-59.
- Hossain MA, Higaki K, Shinpo M, Nanba E, Suzuki Y, Ozono K, Sakai N. 2016. Chemical chaperone treatment for galactosialidosis: Effect of NOEV on  $\beta$ -galactosidase activities in fibroblasts. *Brain Dev* 38:175-180.
- Hu H, Gomero E, Bonten E, Gray JT, Allay J, Wu Y, Wu J, Calabrese C, Nienhuis A, d'Azzo A. 2012. Preclinical dose-finding study with a liver-tropic, recombinant AAV-2/8 vector in the mouse model of

- galactosialidosis. *Mol Ther* 20:267-274.
- Itoh K, Kobayashi I, Nishioka S, Sezutsu H, Machii H, Tamura T. 2016. Recent progress in development of transgenic silkworms overexpressing recombinant human proteins with therapeutic potential in silk glands. *Drug Discov Ther* 10:34-39.
- Ketterer S, Gomez-Auli A, Hillebrand LE, Petrera A, Ketscher A, Reinheckel T. 2017. Inherited diseases caused by mutations in cathepsin protease genes. *FEBS J* 284:1437-1454.
- Krämer A, Keller W, Appel B, Lührmann R. 1984. The 5' terminus of the RNA moiety of U1 small nuclear ribonucleo protein particles is required for the splicing of messenger RNA precursors. *Cell* 38:299-307.
- Le Guédard-Méreuze S, Vaché C, Molinari N, Vaudaine J, Claustres M, Roux AF, Tuffery-Giraud S. 2009. Sequence contexts that determine the pathogenicity of base substitutions at position +3 of donor splice-sites. *Hum Mutat* 30:1329-1339.
- Lee NC, Lee YM, Chen PW, Byrne BJ, Hwu WL. 2016. Mutation-adapted U1 snRNA corrects a splicing error of the dopa decarboxylase gene. *Hum Mol Genet* 25:5142-5147.
- Lund M, Kjems J. 2002. Defining a 5' splice site by functional selection in the presence and absence of U1 snRNA 5' end. *RNA* 8:166-179.
- Madsen PP, Kibaek M, Roca X, Sachidanandam R, Krainer AR, Christensen E, Steiner RD, Gibson KM, Corydon TJ, Knudsen I, Wanders RJ, Ruiten JP, Gregersen N, Andresen BS. 2006. Short/branched-chain acyl-CoA dehydrogenase deficiency due to an IVS3+3A>G mutation that causes exon



skipping. Hum Genet 118:680-690.

Matsuo T, Yamamoto A, Yamamoto T, Otsuki K, Yamazaki N, Kataoka M, Terada H, Shinohara Y. 2010. Replacement of C305 in heart/muscle-type isozyme of human carnitine palmitoyltransferase I with aspartic acid and other amino acids. Biochem Genet 48:193-201.

Mount SM. 1982. A catalogue of splice junction sequences. Nucleic Acids Res 10:459-472.

Murphy JT, Burgess RR, Dahlberg JE, Lund E. 1982. Transcription of a gene for human U1 small nuclear RNA. Cell 29:265-274.

Ohno K, Brengman JM, Felice KJ, Cornblath DR, Engel AG. 1999. Congenital end-plate acetylcholinesterase deficiency caused by a nonsense mutation and an A-->G splice-donor-site mutation at position +3 of the collagenlike-tail-subunit gene (COLQ): how does G at position +3 result in aberrant splicing? Am J Hum Genet 65:635-644.

OMIM®, Online Mendelian Inheritance in Man. Johns Hopkins University, Baltimore, MD. MIM Number: 613111: 11/20/2009. World Wide Web URL: <https://omim.org/>

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. J Comput Biol 4:311-323.

Roca X, Krainer AR, Eperon IC. 2013. Pick one, but be quick: 5' splice sites and the problems of too many choices. Genes Dev 27:129-144.

Sánchez-Alcudia R, Pérez B, Pérez-Cerdá C, Ugarte M, Desviat LR. 2011. Overexpression of adapted U1snRNA in patients' cells to correct a 5' splice site mutation in propionic acidemia. Mol Genet Metab

102:134-138.

Shimmoto M, Takano T, Fukuhara Y, Oshima A, Sakuraba H, Suzuki Y. 1990.

Japanese-type adult galactosialidosis. A unique and common splice junction mutation causing exon skipping in the protective protein/carboxypeptidase gene. *Proc Japan Acad* 66B:217-222.

Shimmoto M, Fukuhara Y, Itoh K, Oshima A, Sakuraba H, Suzuki Y. 1993.

Protective protein gene mutations in galactosialidosis. *J Clin Invest* 91:2393-2398.

Shimmoto M, Nakahori Y, Matsushita I, Shinka T, Kuroki Y, Itoh K, Sakuraba H.

1996. A human protective protein gene partially overlaps the gene encoding phospholipid transfer protein on the complementary strand of DNA. *Biochem Biophys Res Commun* 220:802-806.

Wang GS, Cooper TA. 2007. Splicing in disease: disruption of the splicing

code and the decoding machinery. *Nat Rev Genet* 8:749-761.

Wiegant J, Galjart NJ, Raap AK, d'Azzo A. 1991. The gene encoding human

protective protein (PPGB) is on chromosome 20. *Genomics* 10:345-349.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs

with applications to RNA splicing signals. *J Comput Biol* 11:377-394.

Zhuang Y, Weiner AM. 1986. A compensatory base change in U1 snRNA

suppresses a 5' splice site mutation. *Cell* 46:827-835.

Zhou XY, Morreau H, Rottier R, Davis D, Bonten E, Gillemans N, Wenger

D, Grosveld FG, Doherty P, Suzuki K, Grosveld GC, d'Azzo A. 1995. Mouse

model for the lysosomal disorder galactosialidosis and correction of

the phenotype with overexpressing erythroid precursor cells. *Genes*

### Figure legends

**Fig. 1. The nucleotide sequences of U1 snRNA and 5'-ss of the 7th intron of pre-mRNA encoding CTSA and RT-PCR analysis of the CTSA transcript**

(A) Manner of interaction between the nucleotide sequences of U1 snRNA and 5'-ss of the 7th intron of pre-mRNA encoding CTSA. The nucleotide position corresponding to the IVS7 +3a>g mutation is indicated by the asterisk. The 5'-end of U1 snRNA shown with lowercase letters fails to base pair with those in the pre-mRNA of CTSA having the mutation.

(B) RT-PCR of the CTSA transcript in human normal cells. RNA samples of human liver and HeLa cells were reversely transcribed; then, the nucleotide sequence corresponding to the regions of exon 6 to exon 8 was amplified by PCR with primer 6 and primer 8. The DNA band strongly stained by ethidium bromide contained the nucleotide sequence composed of "exons 6-8" (350 base pairs [bp]). The position of the expected migration of the DNA band containing the nucleotide sequence composed of "exon 6 and exon 8" (258 bp) is also shown.

(C) Structures of the mini gene plasmids encoding the regions of "exon 6 to exon 8" or "exon 3 to exon 11" of human *CTSA* gene, harboring the IVS7 +3a>g mutation (shown by asterisks). The plasmids of these regions of the wild-type *CTSA* gene were also prepared and used for the control experiment. Primers (T7 primer, BGH pA primer, primer 6, and

primer 8) used for specific RT-PCR of transcripts from the mini gene plasmids are also shown.

(D) RT-PCR of the CTSA transcript from the mini genes. HeLa cells were transfected with one of these 4 mini gene plasmids. The RNA samples from cells were reverse transcribed, and the cDNAs thus prepared were amplified by using T7 primer and BGH pA primer; and then DNA fragments containing the region between exon 6 and exon 8 were re-amplified with primer 6 and primer 8 (note that the PCR was designed not to amplify the transcript encoded by the endogenous *CTSA* gene). Upper and lower bands represent DNA fragments containing "exons 6-8 (350 bp)" and "exon 6 and exon 8 (258 bp)", respectively. As for band "X", see **Supplemental results, Fig. S2A, B, C, and D.**

**Fig. 2. Effects of modified U1 snRNAs on rescue from the splice defect at 5'-ss of the *CTSA* gene having the IVS7 +3a>g mutation**

Panels A and B show the results of the 6th U>C (A) and serial mutation analysis (B), respectively. In the upper panel, the nucleotide sequences of the 5'-end of the U1 snRNAs used in this experiment and 5'-ss of IVS7 +3a>g CTSA pre-mRNA are shown. The IVS7 +3a>g mutation is indicated with an asterisk. In the 5'-end of U1 snRNA, nucleotides shown by uppercase or lowercase letters indicate those complementary or non-complementary, respectively, to the nucleotide sequence of the IVS7 +3a>g allele. In the lower panel, the results of the analysis of the RNA species are shown. HeLa cells were co-transfected with the mini

gene plasmid (exons 3-11, mutant) and U1 plasmid, and then RNA samples were analyzed by RT-PCR. In the case of products obtained by using modified-7 U1 snRNA, the DNA fragment containing the upper band ("exons 6-8") was purified from the electrophoretic gel, and the exon boundaries (exon 6/exon 7 and exon 7/exon 8) were confirmed by direct sequencing (data not shown).

(C) Histograms indicate the percentage of exon 7 inclusion upon densitometric analysis of the bands. The band intensities of "exons 6-8", "X", and "exon 6 and exon 8" were calculated from the electrophoretic gel; and total band intensity was obtained as sum of those of 3 bands. The percentage of the band intensity of "exons 6-8" is expressed as the mean  $\pm$  SD from 4 independent experiments except for control-1 and -2 (3 independent experiments). As for the control-1 and -2, the RNA samples were prepared from cells transfected with only mini gene plasmid (exons 3-11, wild-type) and both mini gene plasmid (exons 3-11, wild-type) and wild-type U1 plasmid, respectively. The electrophoregrams of control-1 and -2 are shown in **Fig. 1D** (control-1) and **Supplemental results, Fig. S3D** (control-1 and -2).

(D) Dose dependency of modified-7 U1 expression plasmid for splicing was examined. The mini gene plasmid (exons 3-11, mutant) and U1 plasmid (wild-type and/or modified-7) were introduced into HeLa cells at the indicated amounts ( $\mu$ g). The percentage of exon 7 inclusion is indicated below each lane, expressed as the mean  $\pm$  SD from 3 independent experiments.

**Fig. 3. Immunodetection of EFGP-CTSA-Myc fusion protein expressed in COS7 cells**

COS7 cells were transfected with i) pcDNA3.1/Hygro(-) plasmid (empty vector) or *CTSA mini gene* plasmid, ii) U1 plasmid, and iii) EGFP expression plasmid (for estimation of transfection efficiencies). After harvesting, cell lysates were subjected to SDS-PAGE and subsequent Western blotting using anti-EGFP and anti-Myc antibodies. Two independently prepared samples were subjected to the analyses. Upper and middle photos show EFGP-CTSA-Myc fusion protein (approx. 63-kDa) detected by anti-EGFP antibody and by anti-Myc antibody, respectively (shown by arrows). Bottom photo represents the expression level of the 27-kDa EGFP (not fused with another peptide) as a transfection control. As for the smaller protein shown by the broken arrows, see **Supplemental results, Fig. S2D**.

### **Highlights**

1. Model system with a mini gene reproduced exon 7 skipping by IVS7 +3a>g of CTSA gene.
2. One base substitution of U1 snRNA for +3a>g was not effective for proper splicing.
3. Increased complementarity of U1 snRNA to splice site produced properly spliced mRNA.

Fig. 1

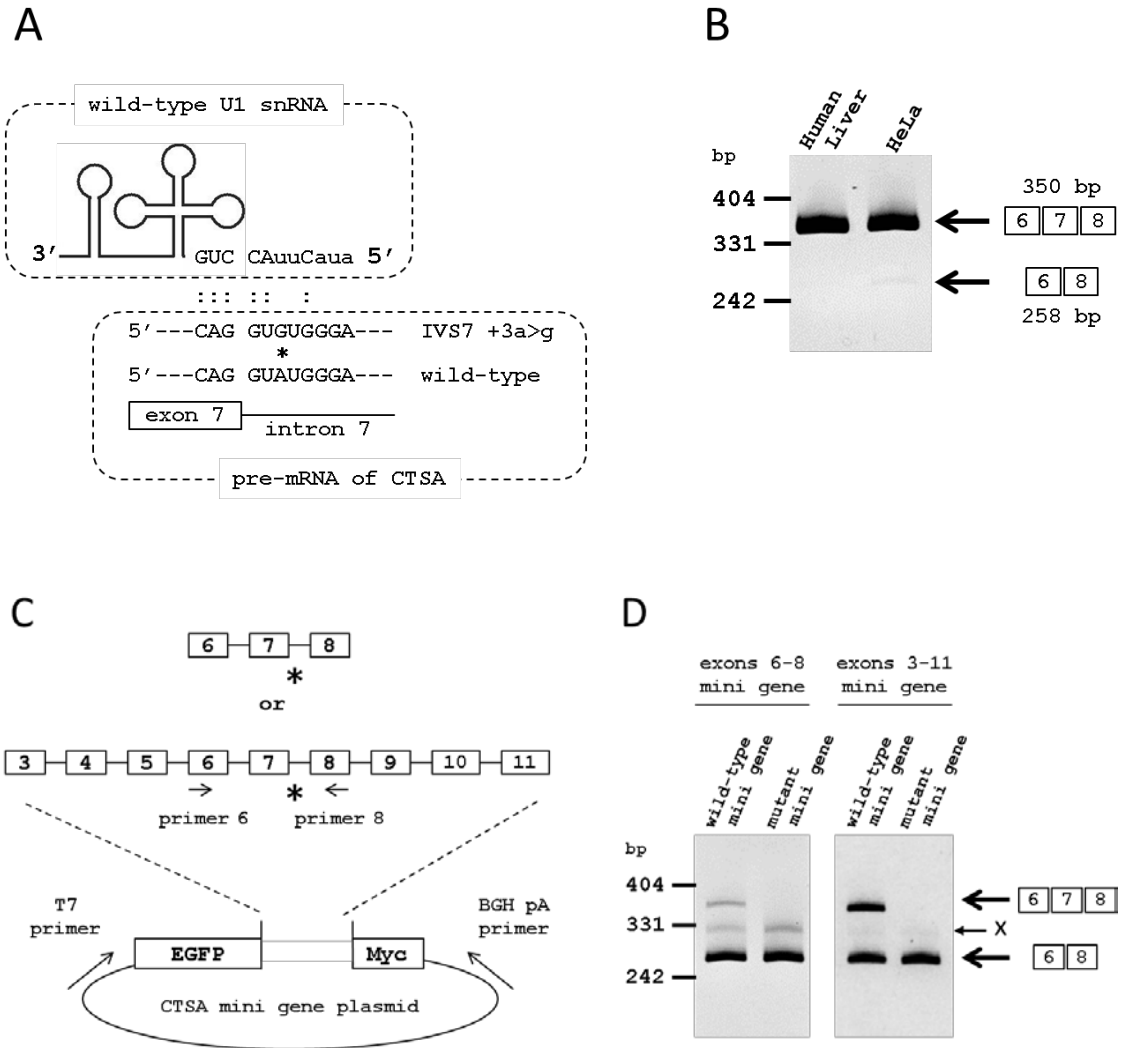




Fig. 2

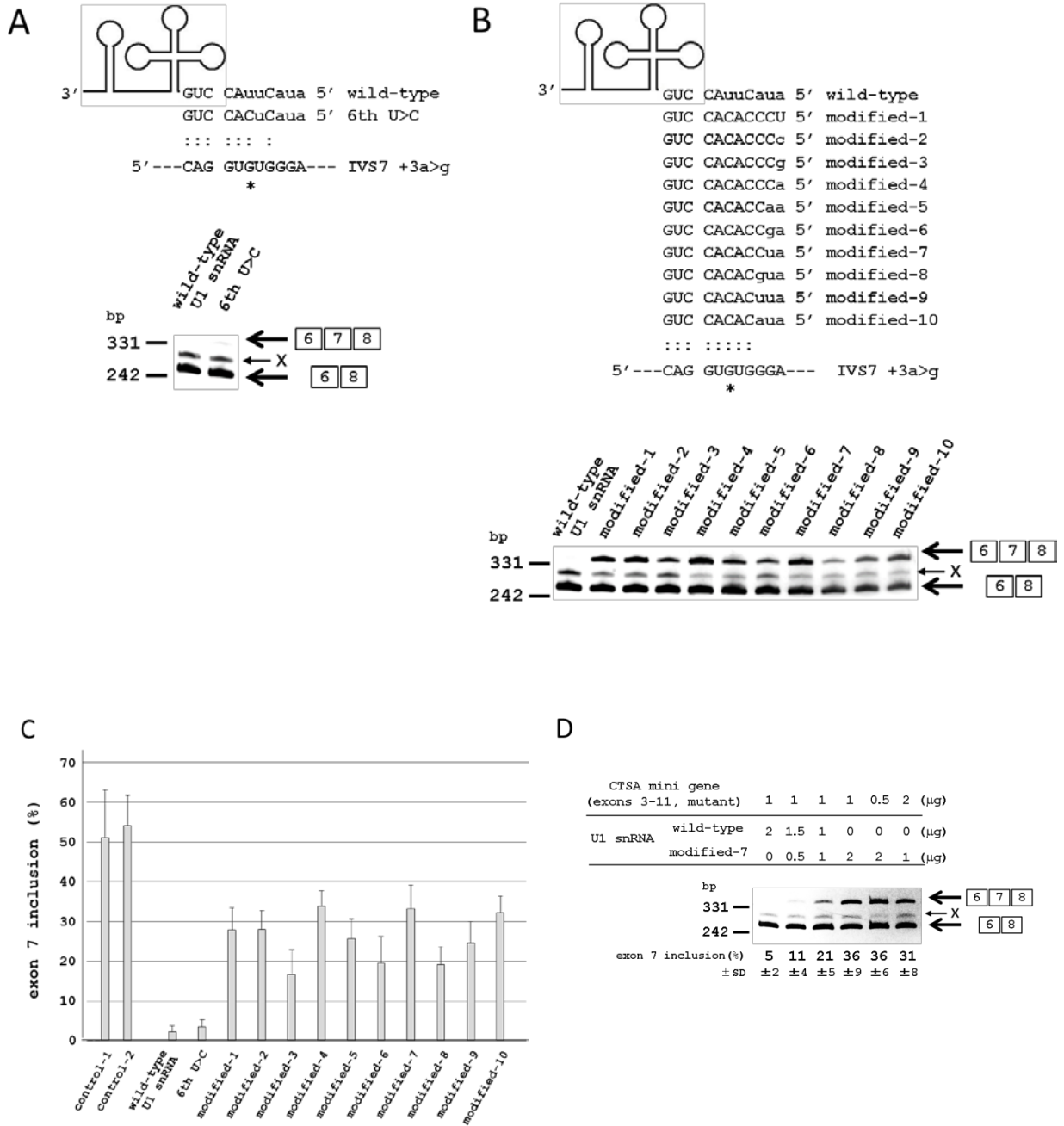
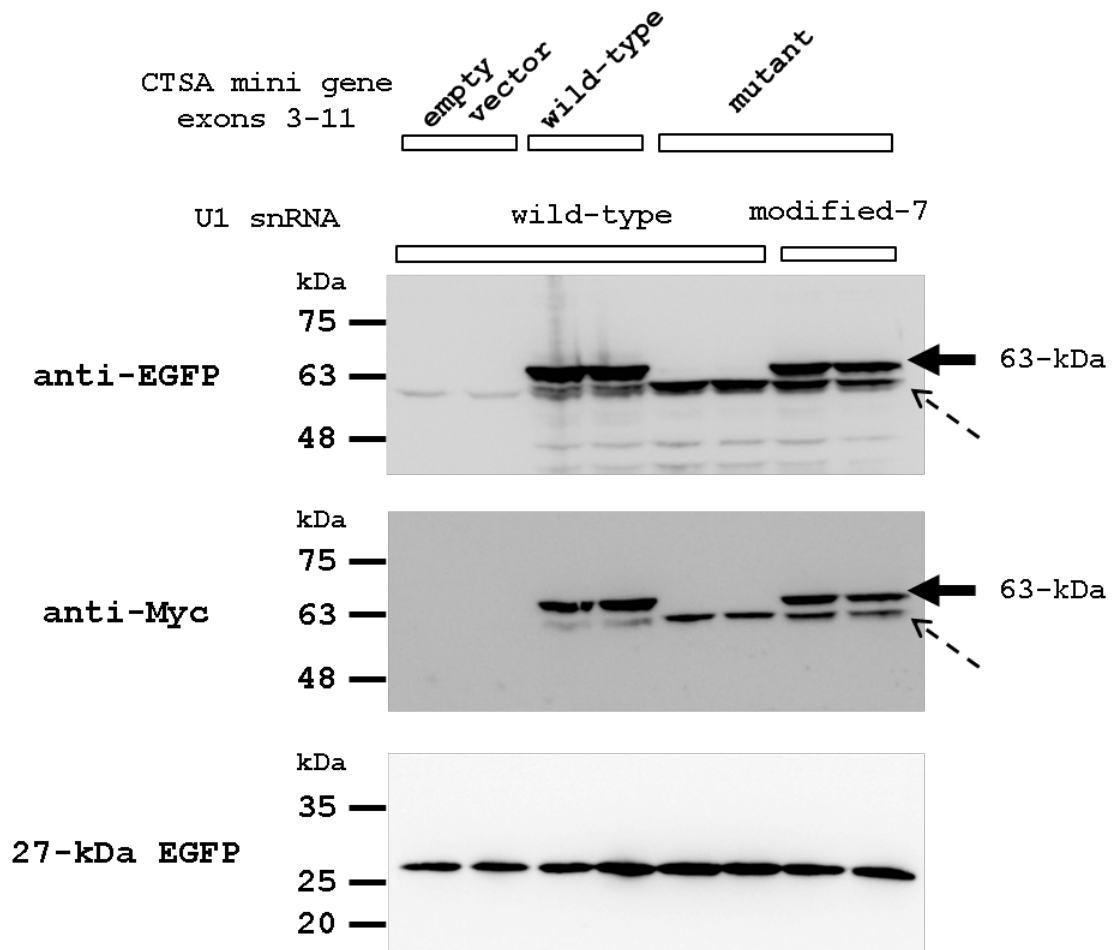


Fig. 3



## Supplemental methods

### 1. Construction of expression plasmids of wild-type and modified U1 snRNAs

To construct the expression plasmid of human U1 snRNA, we amplified human genomic DNA by PCR using PrimeSTAR HS DNA Polymerase (TaKaRa, Otsu, Japan). For amplification from its own promoter to transcription termination signal (from -393 to +199), the primer pair of 5'-aaaggatccTAAGGACCAGCTTCTTTGGGAGAGAACAGAC (sense) and 5'-aaagtcgactTAGCGTACAGTCTACTTTTGAAACTCCAG (antisense) was used (nucleotides artificially added for restriction sites [*Bam*HI and *Sal*II] and for efficient digestion are shown by the lowercase letters). The obtained DNA fragment was digested with *Bam*HI and *Sal*II, and subcloned into these restriction sites of pUC19. This plasmid is referred to as wild-type U1 plasmid.

By PCR using mutated oligonucleotide primers and the wild-type U1 plasmid, the coding regions for modified U1 snRNAs were prepared. The obtained PCR product was digested with *Bgl*III (position -7) and *Sal*I (artificially added to the 3'-end of *U1 snRNA gene*). To construct the expression plasmid of modified U1 snRNA, the corresponding region of the wild-type U1 plasmid was substituted with the thus prepared mutated DNA fragment.

### 2. Construction of wild-type *CTSA mini gene* plasmids

To construct the mini gene plasmids of human *CTSA* from exon 6

to exon 8 and from exon 3 to exon 11, we amplified human genomic DNA by PCR using PrimeSTAR HS DNA Polymerase. For amplification from exon 6 to exon 8, 5'-gcactcgaGTCGCCAGAGCAATTTTGAGGCCCTTCAAG (primer 6, sense) and 5'-gcagtcgacATTGGTCACGCATTCCAGGTCTTTGTTGTC (primer 8, antisense) were used; and for that from exon 3 to exon 11, 5'-gcatgggtactctcgcgagcgtgGTTTGTGGAGTCCAGAAGGATCCC (primer 3, sense) and 5'-gcatgggtacttagtcgacgTTGCACATGTCCCATTGTGGCAGCTG (primer 11, antisense) were employed. Nucleotides artificially added for restriction sites (*Xho*I and *Sal*I) and for efficient digestion are shown by lowercase letters. The obtained DNA fragment was digested with *Xho*I and *Sal*I and subcloned into pcDNA3.1/Hygro(-), in which DNA fragment encoding CTSA was sandwiched between those fragments encoding EGFP and c-Myc tag. These mini gene plasmids are referred to as wild-type CTSA mini gene plasmids ("exons 6-8" or "exons 3-11").

### 3. RT-PCR

A 1- $\mu$ g aliquot of total RNA from HeLa cells was reversely transcribed with oligo(dT) primer. ss-cDNA corresponding to 50 ng of total RNA was first amplified by 35 cycles of PCR with T7 primer and BGH pA primer in a 25- $\mu$ L reaction mixture using *Ex Taq* Hot Start Version (TaKaRa). The reaction conditions for 1 cycle of PCR were 94 °C for 0.5 min, 56 °C for 0.5 min, and 72 °C for 1.5 min. Approx. 1.7-1.8 kbp products were obtained by this PCR. Then, the reaction mixture was diluted 5,000-10,000 times with distilled water, and a 1- $\mu$ L aliquot of

diluted mixture was amplified by 20 cycles of PCR with primer 6 and primer 8 in a 25- $\mu$ L reaction mixture. The reaction conditions for 1 cycle of PCR were 94 °C for 0.5 min, 56 °C for 0.5 min, and 72 °C for 0.5 min. The reaction mixtures were then subjected to PAGE.

Transcript from cells without transfection of mini gene plasmids was analyzed by 1 round of PCR with primer 6 and primer 8. Briefly, samples of poly(A)<sup>+</sup> RNA of human liver and total RNA of HeLa cells were reversely transcribed with oligo(dT) primer. Samples of ss-cDNA corresponding to 5 ng (liver) and 50 ng (HeLa cells) of RNA were amplified by 35 cycles of PCR with primer 6 and primer 8 in a 25- $\mu$ L reaction mixture using *Ex Taq* Hot Start Version. The reaction conditions for 1 cycle of PCR were 94 °C for 0.5 min, 56 °C for 0.5 min, and 72 °C for 0.5 min.

### Supplemental results

#### **Fig. S1. Structures of the *CTSA mini gene* plasmid and nucleotide sequence**

(A) The overall structure of *CTSA mini gene* plasmids ("exons 6-8" and "exons 3-11") is depicted. Asterisk indicates the IVS7 +3a>g mutation.

(B) The nucleotide sequences (upper line) and amino acid sequences (lower line) of EGFP and c-Myc tag coding regions are shown. The underlined nucleotides indicate artificially added nucleotides to create a restriction site(s) and a termination codon. The amino acids are shown by the one-letter abbreviation code, and underlined amino acids represent those encoded by artificially added nucleotides.

(C) The nucleotide sequence (upper line) and amino acid sequence (lower line) of the *CTSA* gene from exon 3 to exon 11 are shown. Nucleotides in the exon and intron are shown by uppercase and lowercase letters, respectively. The asterisk at the exon 7/intron 7 boundary indicates the IVS7 +3a>g mutation. In both lines, underlined nucleotides and amino acids indicate those as described in "B".

(D) Amino acid sequence of the EGFP-CTSA-Myc fusion protein produced by transfection with the *CTSA mini gene* plasmid (exons 3-11). The amino acids coded by additional nucleotides containing restriction sites are underlined as described in "B". When pre-mRNAs transcribed from the *CTSA mini gene* plasmids (exons 3-11) were properly spliced, an approx. 63-kDa protein (total of 557 amino acids) was produced.

**Fig. S2. Characterization of the band "X" obtained by RT-PCR**

(A) The nucleotide sequence of the cDNA corresponding to exon 6 to exon 8 of *CTSA* mRNA is shown. Boxes above nucleotides show exons (exon 6, 7, and 8). Arrows indicate oligonucleotide primers (primer 6, primer 8, primer F, and primer R) used in this experiment. Primer F and primer R were used in "C" and "D".

(B) RT-PCR products were obtained with T7 primer and BGH pA primer (1st PCR) and primer 6 and primer 8 (2nd PCR). The reaction mixture of RT-PCR was loaded onto a polyacrylamide gel and stained with ethidium bromide (lane 1). DNA in band "X" was carefully extracted from the electrophoretic gel. In addition, DNAs in band "exons 6-8 (350 bp)"

and "exon 6 and exon 8 (258 bp)" were also purified. After purification, DNAs from "exons 6-8", from "X", and from "exon 6 and exon 8" were loaded (lane 2, 3, and 4, respectively).

(C) The DNA molecules contained in band "X" (lane 3 in "B") was directly sequenced using primer F. The sequencing peak showed the nucleotide sequences following exon 6 of the PCR products in band "X". As a result, band "X" appeared to contain mainly 2 DNA species, i) a DNA fragment of exon 6 directly connected to exon 8, and ii) a DNA fragment having an insertion of "gt" nucleotides between exon 6 and exon 8 (hereafter, this latter DNA is designated as "exon 6-gt-exon 8"). DNAs contained in band of "exons 6-8" (lane 2 in "B") and "exon 6 and exon 8" (lane 4 in "B") were also sequenced. DNA in band of "exons 6-8" was confirmed to have the expected nucleotide sequence. Whereas, in the case of the "exon 6 and exon 8" band, DNA in which exon 6 was directly connected to exon 8, was shown to be the main component. In addition, although we could not identify the nucleotide sequence, there was other DNA in which exon 6 was not directly connected to exon 8 (data not shown).

(D) To further characterize the DNA species in bands observed in "B", we re-amplified gel-purified DNAs by using the primer pair of primer F and primer R. After PCR, the reaction mixtures were loaded onto a polyacrylamide gel. When the RT-PCR mixture (lane 1 in "B") was used as a template, 3 bands were observed. From their electrophoretic mobility, it seemed that DNAs in bands "a (262 bp)", "b", and "c (170 bp)" had arisen from "exons 6-8", "X", and "exon 6 and exon 8",

respectively. When purified DNA from lane 2 in "B" ("exons 6-8") was used as a template, only band "a" was observed. On the contrary, when purified DNA from lane 4 in "B" ("exon 6 and exon 8") was used as a template, 2 bands, band "b" and band "c", were observed. A similar band pattern was also observed when DNA from lane 3 in "B" ("X") was used as a template.

These observations suggest that the DNA of band "X" was a heteroduplex comprising the "exon 6 and exon 8" and "exon 6-gt-exon 8". Moreover, if a splice variant including "exon 6-gt-exon 8" existed, a 59-kDa protein (total of 527 amino acids) with EGFP and Myc tag would be synthesized from a mini gene plasmid containing exons 3-11. Accordingly, the smaller protein shown by the broken arrows in **Fig. 3** (observed in both EGFP and Myc panels) might have been translated from this splice variant.

### **Fig. S3. Detailed analysis of modified-7 U1 expression plasmid**

The experiments in this section were conducted to ascertain that the splicing correction was achieved by "U1 snRNA molecule" expressed from modified U1 plasmid (A, B, and C). In addition, the effect on exon 7 inclusion of wild-type *CTSA* gene by modified U1 snRNA was examined (D).

(A) The structures of the expression plasmid of U1 snRNA lacking its promoter ( $\Delta$ Pro and RevPro) are depicted. For construction of these plasmids, the promoter region of the expression plasmids of modified-7 U1 snRNA was digested with *Bam*HI and *Bgl*III. After elimination of the



promoter region, self-ligation was done to make the promoter-less plasmid ( $\Delta$ Pro). To make a plasmid having a promoter region with its direction reversed (RevPro), we inserted the digested promoter region in the reverse direction.

(B) Mutations of loop1 and Sm domain of U1 snRNA are shown. Loop1 and the Sm domain of U1 snRNA are necessary for U1 snRNA function [Surowy et al., 1989; Hamm et al., 1990; Hamm and Mattaj, 1990; Gunderson et al., 1998] and substitution of cytosine to adenine at 33 (C33A) was reported to not have any influence on its function [Surowy et al., 1989]. By use of the overlap extension PCR method using mutated oligonucleotide primers, modified-7 U1 expression plasmid was further modified to create expression plasmids of  $\Delta$ Loop1,  $\Delta$ Sm, and C33A. Lowercase letters mean that they were substituted from the wild-type U1 snRNA sequences.

(C) The results of the analysis of RNA species are shown. Transfection with the plasmids and RT-PCR experiments were performed as described in **Fig. 2A and 2B**. Transfection with either  $\Delta$ Pro or RevPro of modified-7 U1 expression plasmids was not effective in rescue from the splice defect.  $\Delta$ Loop1 and  $\Delta$ Sm mutants were also not effective, but C33A of the modified-7 U1 expression plasmid showed activity similar to that of the parental modified-7. These results showed that the modified-7 U1 expression plasmid was transcribed in HeLa and COS7 cells and that U1 snRNA synthesized in the cells acted as a splicing factor.

(D) The effect of modified-7 U1 snRNA on exon 7 inclusion of the wild-type *CTSA* gene was examined. In the left panel, HeLa cells were

transfected with a 2- $\mu$ g aliquot of U1 plasmid without the mini gene plasmid. In the right panel, HeLa cells were transfected with a 2- $\mu$ g aliquot of U1 plasmid and a 1- $\mu$ g aliquot of the *CTSA mini gene* plasmid (exons 3-11, wild-type). In both panels, the lanes represented as "non" mean that RNA samples were prepared from cells without transfection with the U1 plasmid. RT-PCR experiments were performed as described in **Fig. 1D** (without mini gene plasmid) and **Fig. 2A and 2B** (with mini gene plasmid). The percentages of exon 7 inclusion are indicated below each lane, expressed as the mean  $\pm$  SD from 3 independent experiments. As a result, modified-7 U1 snRNA was shown to have no effect on the splicing of intron 7 of the wild-type *CTSA gene*.

## References

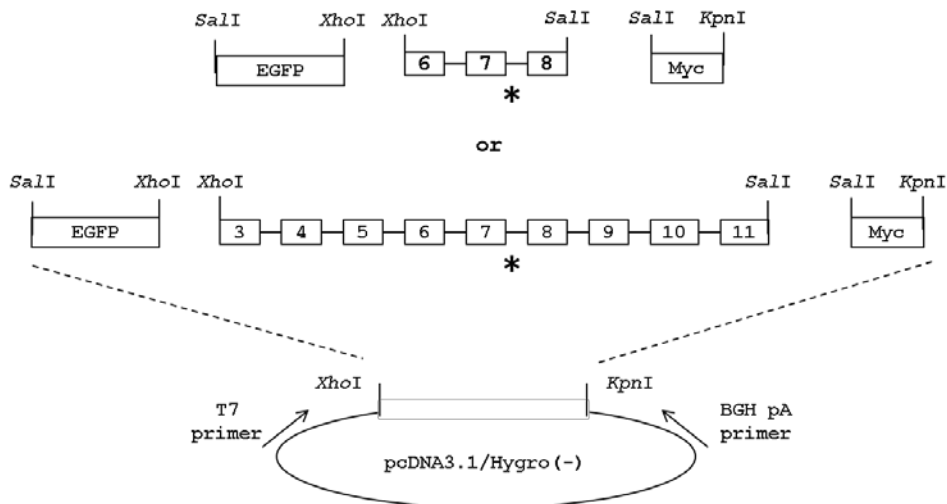
### (Supplemental methods, Supplemental results)

- Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. 1998. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell* 1:255-264.
- Hamm J, Dathan NA, Scherly D, Mattaj IW. 1990. Multiple domains of U1 snRNA, including U1 specific protein binding sites, are required for splicing. *EMBO J* 9:1237-1244.
- Hamm J, Mattaj IW. 1990. Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* 63:109-118.
- Surowy CS, van Santen VL, Scheib-Wixted SM, Spritz RA. 1989. Direct, sequence-specific binding of the human U1-70K ribonucleoprotein

antigen protein to loop I of U1 small nuclear RNA. Mol Cell Biol  
9:4179-4186.

Supplemental results Fig. S1.

A



**B** EGFP coding region (*Sali*-*XhoI*)

```

GTCGACATGGTGAGCAAGGGCGAGGAGCTGTTCCACGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTC
M V S K G E E L F T G V V P I L V E L D G D V N G H K F
AGCGTGTCCGGCGAGGGCGAGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGCCCTGG
S V S G E G E G D A T Y G K L T L K F I C T T G K L P V P W
CCCACCTCGTGACCACCCTGACCTACGGCGTGCAGTGCCTTCAGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCC
P T L V T T L T Y G V Q C F S R Y P D H M K Q H D F F K S A
ATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
M P E G Y V Q E R T I F F K D D G N Y K T R A E V K F E G D
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGC
T L V N R I E L K G I D F K E D G N I L G H K L E Y N Y N S
CACAAAGCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAG
H N V Y I M A D K Q K N G I K V N F K I R H N I E D G S V Q
CTCGCCGACCACTACAGCAGAACACCCCATCGGGCAGGCCCGCTGCTGCTGCCGACAACCACTACCTGAGCACCCAGTCCGCCCTG
L A D H Y Q Q N T P I G D G P V L L P D N H Y L S T Q S A L
AGCAAAGACCCCAACGAGAAGCGGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTAC
S K D P N E K R D H M V L L E F V T A A G I T L G M D E L Y
AAGTCCGGACTCAGATCTCGAG
K S G L R S R
  
```

**c-Myc tag coding region (*Sali*-*KpnI*)**

```

GTCGACGAACAAAACTCATCTCAGAAGAGGATCTGTAAGCTTGGTACC
V D E Q K L I S E E D L stop
  
```

Supplemental results Fig. S1.

**C** human CTSA gene exon 3-exon 11 (XhoI-SalI)

```

ex3  CTCGAGCGTGGTTTGTGGAGTCCCAGAAGGATCCCAGAACAGCCCTGTGGTGTCTTTGGTCAATGGGGTCCCGGCTGCAGCTCACTAG
     R A W F V E S Q K D P E N S P V V L W L N G G P G C S S L D
     ATGGGCTCCTCACAGAGCATGGCCCTTCCTGgtgagtggacagcaggggaaagcacagtccccaaagtaaaggctgggaaagcagag
     G L L T E H G P F L
     agaggggctttgtgattttocaaaaagtcttcttcttaagcctcgggattttctctatgceattggccttgctgcctgtacctctc
     agaggttttaccacatgtaaaagtgcctctcatggtggcccttccctccaccagctggcccttgggacttactcagccatctcttcc

ex4  tcctcagGTCCAGCCAGATGGTGTCCACCTGGAGTACAACCCTATTCTTGGAAATCTGGtatagtggagctgtgggtgtgtctgggcac
     V Q P D G V T L E Y N P Y S W N L

ex5  ttggatgggggtggcatttagctaattttccccctctagATTGCCAATGTTTATACCTGGAGTCCCAGCTGGGGTGGGTTCCTCC
     I A N V L Y L E S P A G V G F S
     TACTCCGATGACAAGTTTTATGCAACTAATGACACTGAGgtgagctctgtgctgcccatctgccccaggctccccggtcctccatccat
     Y S D D K F Y A T N D T E
     ggcagatgctggggagagagagcatggccttggagcttactttctgctgtcatatgctattgtggtatggtggccagctcctctct
     tgcacaaaataggggggttaagtcatatctctgaagtcttccagttcaataccaaaagctcctgattcctctgtcttggctctgc

ex6  catccccagTCCGCCAGAGCAATTTGAGGCCCTTCAAGATTTCTTCGGCTCTTTCCGGAGTACAAGAACAACAACTTTTCCTGACC
     V A Q S N F E A L Q D F F R L F P E Y K N N K L F L T
     GGGAGAGCTATGTGGCATCTACATCCCCACCTGGCCGTGTGGTCATGCAGGATCCCAGCATGAACCTTCAGgtcagggtagctgc
     G E S Y A G I Y I P T L A V L V M Q D P S M N L Q
     aggagggaaaggaggtagcttgaggtgtggccttacagttagcaaggtcagactgactggtcaatgtccccaaccaaccagctcctctg
     cataacctccccaccggctgcccagtagctgtctgtgtgtgctccccacaggaactccactgtaggctgccagctctgtgttttg
     cattgagtggtctctcatacccaccgctctacataaaacccctgggtttcagcttccggatgctgaaggtctgggtgttgggtgaggtgag

ex7  gaggtctctctttttgcccctccatgagctgagcaccctgggtgtttcacagGGGCTGGTGTGGCAATGGACTCCTCCTCATGAG
     G L A V G N G L S S Y E
     CAGAATGACAACCTCCCTGGTCTACTTTGCCTACTACCATGGCCTTCTGGGGAACAGgttgggtagggcagtgggcaatctctgggt
     Q N D N S L V Y F A Y Y H G L L G N R *
     gaggcaggtcacatgatctcaggtctgtctccaggeacatgatagacagccagcccaggtttagagcagcaggtgggtgtgcagaa
     agaacctagtgctgaagcccaagaaatagagatcagttttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt
     aggctggagtgctgagtgccacgactcttagctcactgcaacctccaccctccaggttcaagcaattctcctcagactcccaggcagct
     gggactacagcgcgtaccaccaacacctgcctaattttggatattttagtagagatggggttttgcgtgtggtggccagctggacttga
     actcctggcctcaaatgatctcccaagtgtgggttacaggeatgagctacagcagccagcagaaaatcagattttaaaggaactaa
     ggagcctgtctgtatgaccctggggcaactctttgggagctagcttagatgacatacagaccacacctgtgatacttccagctctc
     ctagggtcaggggttggtgggaggtgggggtatgctcgcctcctcctccttctctcttcccaaccagcagctgtagccttttcaactc

ex8  catctcctacagGCTTGTCTTCTCTCCAGCCCACTGCTGCTCTCAAAACAAGTGAAGTCTCTATGACAACAAGAGCTGGAATGCGT
     L W S S L Q T H C C S Q N K C N F Y D N K D L E C V
     GACCAATgtgaggttctgccatcactttgcatgagctctccatcccaatcctgagaacaggaccacattcctccttggggactcctt
     T N
     gttctccgtccctcattgtcttctgacaccttaagcaatttagtgtctcctgggacatggtctaaagtgtattccccaccaccctc
     ctatggtggcaaatcccataacctccgtagggagaaaggactcttccatcttaacaaccctgcaactggcagggcactgtgtgaac
     tgtgtctatgtagattatgaagctcaagatccccctcaaatgagatcactaccacatttccagataaggaactgagggctat
     gaatttgaatatcactgtctgaagtcaaggtgtggagggcagtgatgctaggatttggaaccaggtctgggaagattcctgccactgag
     cctcagtttccctacctgtgaaaaaagggtgagaggtgaaaatccaagagcagtaaatctgggacaacttggtgggtctgtggaga

ex9  ggctggggtctgtaaacatggtgcttgggtatatttgcagCTTCAGGAAGTGGCCCGCATGTTGGCAACTCTGGCCCTCAACATCT
     L Q E V A R I V G N S G L N I Y
     ACAATCTCTATGCCCCGTGTGCTGGAGGGTGGCCAGCCATTTAGgtaggtgtgtgtgtgggtccccctggagccaaccagcccaatc
     N L Y A P C A G G V P S H F R
     ggaggtccacaccatccccccacctcacatttgcagGTATGAGAAGGACACTGTTGTGTCAGGATTTGGGCAACATCTTCACTCGCC

ex10  Y E K D T V V V Q D L G N I F T R L
     TGCCACTCAAGCGGATGTGGCATCAGgtgtgagggcgtgggttctctcctgtgagggtggggcagggggaggggaggggaagcagag
     P L K R M W H Q
     gccctgaccctgtctgtgcttccagGCACTGCTGCGCTCAGGGGATAAAGTGGCAGTGGACCCGCCCTGCACCAACACAACAGCTGC

ex11  A L L R S G D K V R M D P P C T N T T A A
     TTCCACTACCTCAACAACCCGTACGTGCGGAGGCCCTCAACATCCGGAGCAGCTGCCACAATGGGACATGTGCAACGTCGAC
     S T Y L N N P Y V R K A L N I P E Q L P Q W D M C N V D
  
```

Supplemental results Fig. S1.

D

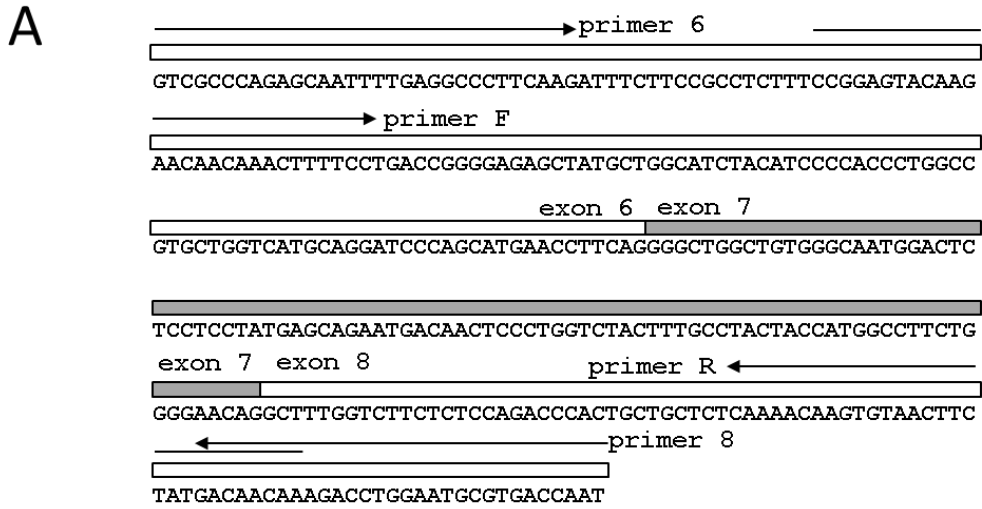
```
MVSKGEELFT GVPVILVELD GDVNGHKFSV SGEGEDATY GKLTLLKFICT TGKLPVPWPT 60
LVTTLTYGVQ CFSRYPDHMK QHDFFKSAMP EGYVQERTIF EKDDGNYKTR AEVKFEEDTL 120
VNRIELKGID FKEDGNILGH KLEYNYNSHN VYIMADKQKN GIKVNEFKIRH NIEDGSVQLA 180
DHYQQNTPIG DGPVLLPDNH YLSTQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYKS 240
GLRSRAWFVE SQKDPENSPV VLWLNCGPGC SSLDGLLTHEH GPFLVQPDGV TLEYNEYSWN 300
LIANVLYLES PAGVGFSYSD DKFYATNDTE VAQSNFEALQ DFFRLFPEYK NNKLELTGES 360
YAGIYIPTLA VLVMQDPSMN LQGLAVGNGL SSYEQNDNSL VYFAYYHGLL GNRLWSSLQT 420
HCCSQNKCNF YDNKDLECVT NLQEVARIVG NSGLNIYNLY APCAGGVPSH FRYEKDTPVV 480
QDLGNIFTRL PLKRMWHQAL LRSQDKVRMD PPCTNTTAAS TYLNNPYVRK ALNIPEQLPQ 540
WDMCNVDEQK LISEEDL 557
```

1-247 (28-kDa) : EGFP + linker

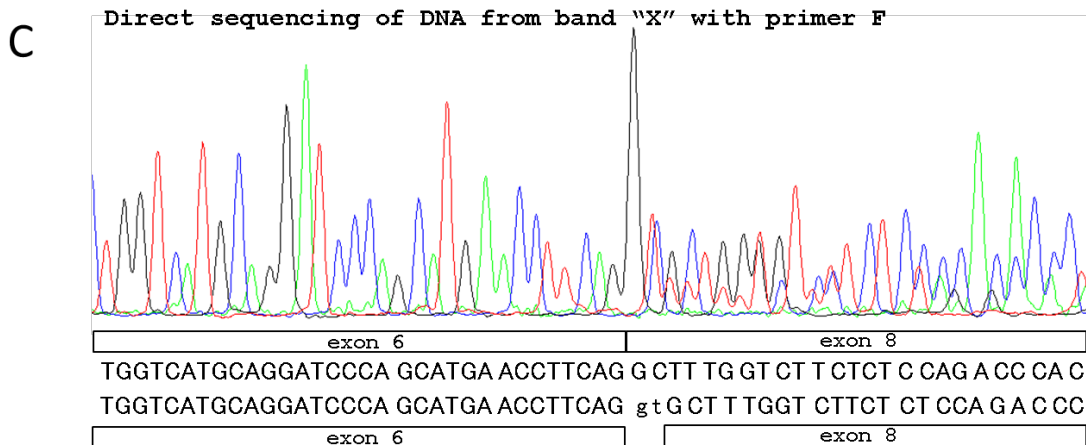
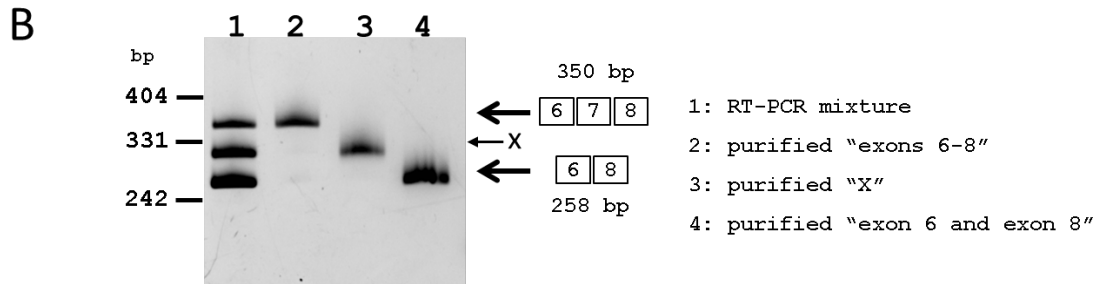
248-545 (33-kDa) : CTSA exon 3-exon 11  
(amino acid 66-363, according to UniProt P10619-1)

546-557 (1.4-kDa) : linker + c-Myc

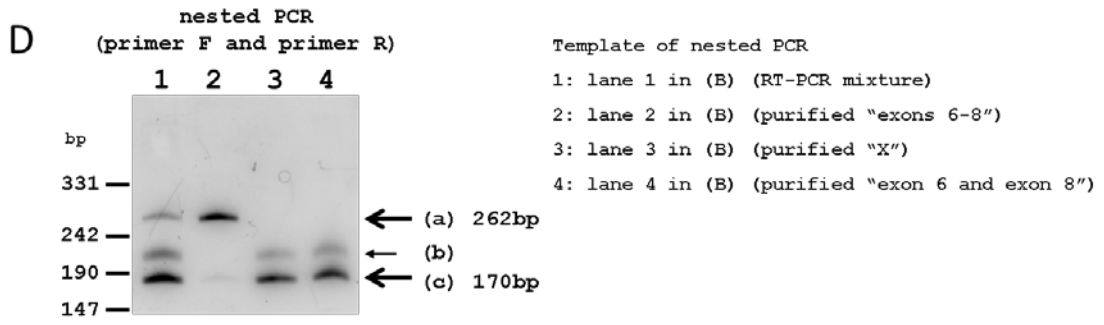
Supplemental results Fig. S2.



primer 6 and primer 8



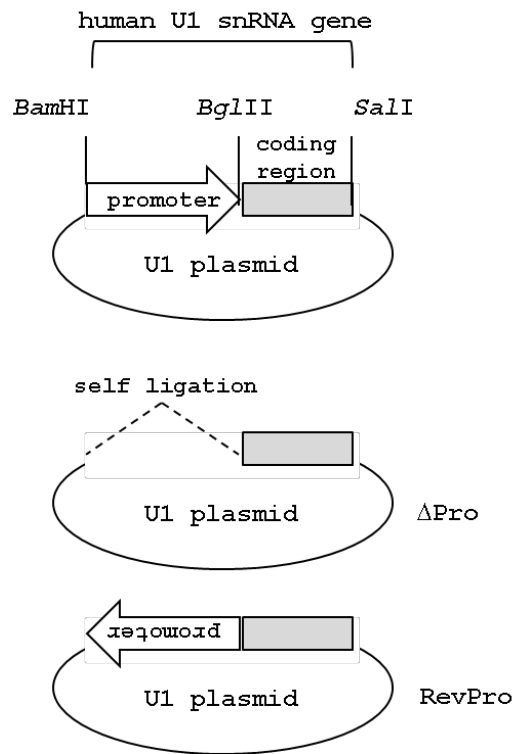
Supplemental results Fig. S2.



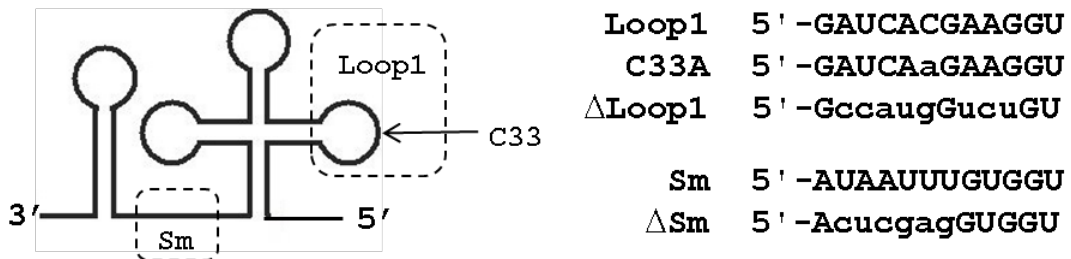


Supplemental results Fig. S3.

A



B



Supplemental results Fig. S3.

