



HOKKAIDO UNIVERSITY

Title	A Splicing Reporter Tuned to Non-AG Acceptor Sites Reveals that Luteolin Enhances the Recognition of Non-canonical Acceptor Sites
Author(s)	Chiba, Masanori; Ariga, Hiroyoshi; Maita, Hiroshi
Citation	Chemical biology & drug design, 87(2), 275-282 https://doi.org/10.1111/cbdd.12656
Issue Date	2016-02
Doc URL	https://hdl.handle.net/2115/64435
Rights	This is the peer reviewed version of the following article: Chiba, M., Ariga, H. and Maita, H. (2016), A Splicing Reporter Tuned to Non-AG Acceptor Sites Reveals that Luteolin Enhances the Recognition of Non-canonical Acceptor Sites. Chemical Biology & Drug Design, 87: 275-282., which has been published in final form at http://dx.doi.org/10.1111/cbdd.12656 . This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.
Type	journal article
File Information	manuscript.pdf



A Splicing Reporter Tuned to Non-AG Acceptor Sites Reveals that Luteolin Enhances the Recognition of Noncanonical Acceptor Sites

Relaxing splicing fidelity

Keywords

pre-mRNA splicing, 3' splice site, reporter assay, fidelity, luteolin

Masanori Chiba^a, Hiriyoshi Ariga^b, and Hiroshi Maita^{b,1}

^a Graduate School of Life Science,

^b Faculty of pharmaceutical Sciences,

Hokkaido University, Sapporo, Japan 060-0812

¹Corresponding author:

Kita12 Nishi6, Kita-ku, Sapporo, Hokkaido, Japan, 060-0812

Phone: 011-706-3921

Fax: 011-706-4988

E-mail: maimai@pharm.hokudai.ac.jp

Abstract

Removal of an intron requires precise recognition of the splice donor and acceptor sites located at the 5' and 3' termini of introns. Although the roles of these sequences differ, mutations in both sites easily block normal splicing and produce an aberrant mRNA. For example, many splice-site mutations occur in patients with inherited diseases. Several approaches have been evaluated to restore expression of a functional protein; however, because of the strict requirement for an AG dinucleotide at the 3' terminus of a U2-type intron, no method is available to correct splicing at a mutated sequence. To identify compounds that allow splicing at the non-AG acceptor site, in the present study we constructed a reporter gene with a modified polypyrimidine (Py) tract. However, the modified Py tract mediated splicing at adjacent noncanonical acceptor sites, including the original mutated site. Further, we show that certain flavones such as luteolin and apigenin enhanced aberrant splicing at the noncanonical acceptor site of the reporter gene. These results suggest that the reporter gene and luteolin may be useful for further screening to identify molecules that correct aberrant splicing caused by a disease-associated splice acceptor site mutation.

Introduction

Pre-mRNA splicing must remove introns with high fidelity because incorrect splicing can alter an open reading frame, resulting in the synthesis of an aberrant protein. Most genes contain multiple introns and encode pre-mRNAs that are alternatively spliced. To ensure high fidelity and plasticity, intron definition is an important step, and different strategies are employed for recognition of the splice donor and acceptor sites located at the 5' and 3' termini of an intron, respectively. The donor splice site is determined according to its complementarity with U1 and U6 small nuclear RNAs (1), and the acceptor site is recognized and bound by the heterodimeric U2AF35/U2AF65 complex (2) at the beginning of a splicing reaction.

Introns are classified to be U2- and U12-dependent according to the splice-site sequence and are spliced by the major and the minor spliceosomes, respectively. The canonical U2-dependent acceptor splice site comprises an AG dinucleotide at its 3' terminus and a polypyrimidine (Py) tract precedes the AG sequence. U2AF65 and U2AF35 bind the Py tract and AG dinucleotide, respectively (3-6). The U2AF35/65 heterodimer stabilizes the U2 snRNP bound to the branch point sequence to allow further assembly of the spliceosome. The functions of U2AF65 and U2AF35 partially overlap, and a longer Py tract that results in stronger binding of U2AF65 confers independence from U2AF35 of a splice acceptor site, whereas an intron with a short, low-affinity Py tract requires binding of U2AF35 to the AG dinucleotide (7, 8). Further, the *DEK* proto-oncogene product and hnRNP A1 enhance the fidelity of recognition of the canonical AG by impeding U2AF35 loading at the non-AG site (9, 10), ensuring the rejection of a substrate with a mutated

acceptor site. Approximately 18% of splicing-related mutations found in inherited diseases involve mutations of the AG dinucleotide. (11). These mutations are expected to convert the acceptor site to a noncanonical site, which is skipped in favor of novel cryptic acceptor sites around the original sites (12), resulting in frame-shifts or deletions of mRNA sequences.

In contrast, a recent genome-wide study revealed that noncanonical acceptor sites such as TG dinucleotides are used in particular genes (13), suggesting the presence of a mechanism to bypass the highly accurate selection for the AG dinucleotide. Reduction of hnRNPA1 expression contributes to increased splicing at the noncanonical TG dinucleotide in human $G\alpha_s$ gene (14). If a compound relaxes the fidelity of acceptor site recognition by mimicking the selection mechanism, splicing at the original mutated site would be restored. Because the protein-coding sequence is normal even in the presence of an acceptor site mutation, such a compound may provide a general strategy to cure diseases caused by acceptor site mutations. Drug discovery research that is focused on splicing has identified several novel classes of splicing inhibitors and modulators of alternative splicing (15). However, these compounds induce or inhibit splicing at the canonical acceptor site. To the best of our knowledge, no compound has been analyzed for the ability to restore splicing at the noncanonical acceptor site. Therefore, to discover such molecules, we aimed to develop a reporter that detects changes in splicing efficiency at noncanonical acceptor sites.

Here, we report that modifying the Py tract improves the splicing efficiency at the mutated noncanonical acceptor site. The comparison of a variety of mutations of the AG

dinucleotide revealed that splicing at the noncanonical site is sensitive to the substituted nucleotide. Furthermore, we identified that the flavone, luteolin enhances the use of a noncanonical acceptor site in a reporter gene. This reporter gene and luteolin may help elucidate the mechanism of selection of the noncanonical acceptor site, resulting in the discovery of compounds that relax the fidelity of splicing at pathologically mutated acceptor sites.

Methods and Materials

Plasmids

To construct a model minigene, a genomic DNA fragment of the human fibronectin gene (*FNI*) comprising exons 17–19 (NG_012196.1; 32872–34831) was amplified using the polymerase chain reaction (PCR), and the amplicon was inserted into the pEGFP-C1 vector to fuse the sequence encoding green fluorescent protein (GFP) and that of the *FNI* exon 17. Using the GFP-*FNI* ex17–19 recombinant gene as a template, a fragment including the GFP sequence to introns 18–19 was further amplified using PCR and ligated to the sequence of Emerald Luciferase (Toyobo) present in the modified pcDNA3 vector that contains the I-SceI site upstream of the CMV promoter. To introduce various mutations into the acceptor site, mutated fragments were prepared using PCR, and each fragment was exchanged using HindIII and EcoRI sites. The reporter gene sequences were verified using Sanger sequencing. Sequences of the two introns contain several substitutions and deletions of nucleotides compared with the reference sequence (supplementary file).

Reporter assay

The 293T cell line was cultured in DMEM (Nissui) supplemented with 10% calf serum in an atmosphere containing 5% CO₂. For transfection, cells were plated in 60-mm or 100-mm dishes. The reporter plasmid DNAs were transfected using Jet-PEI according to the manufacturer's protocol. Cells were collected and plated in a multiwell plate 6 h after transfection. Further, compounds were added to the transfected cells 24 h after transfection. After incubation for another 24 h, cells were harvested and lysed in SL buffer (16). Whole-cell extracts were transferred to a black-wall plate to measure GFP fluorescence using Glomax (Promega). After measuring fluorescence, cell extracts were transferred to a white-wall plate and mixed with Steady-Glo (Promega) for detecting luciferase activity using Glomax. After subtraction of average counts of empty wells (background), relative luciferase activity was calculated as a ratio of the luciferase luminescence to GFP fluorescence.

Reverse transcription (RT)-PCR

RNAiso (TaKaRa) and ReverTra Ace RT master mix (Toyobo) were used for RNA extraction and reverse transcription, respectively. ExTaq was used for labeled PCR, and 18–24 PCR cycles were used to adjust for differences in transfection efficiencies. To detect spliced mRNAs, fragments were amplified using the sense primer A (ex17-up: 5'-CAACAAGAAACCACTGGCAC-3') and the antisense primer (pFD-as-IRD1: 5'-GCCCAGCTTCACAGTAGTG-3'-IRD800), and the amplicons were electrophoresed through 2% agarose gel. To detect and quantify labeled PCR products, agarose gels were scanned using an Odyssey imager (Aloka). Applied volumes of PCR products were adjusted for the total intensity of each lane to appear even in the figures. To obtain data

with higher resolution, PCR products generated using the sense primer B (ex2-sense: 5'-CAACAAGAAACCACTGGCAC-3') and pFD-as-IRD1 were electrophoresed on a denaturing polyacrylamide gel using a dNA Sequencer LONG READIR 4200 (Aloka) according to the manufacturer's instruction. After gel electrophoresis, larger original scanned images were cropped using Image analysis (Aloka) and saved as TIFF files to transfer the files to another PC.

Sequencing of PCR amplicons

A GsJr bench-top system (Roche) was used to sequence the RT-PCR amplicons of the spliced mRNAs generated using the reporter gene. For this purpose, fusion primers were designed as per the manufacturer's instructions. Sequences of the fusion primer are divided into four regions: adaptor sequence, key sequence, MID-tag, and annealing region. Adaptor and key sequences were added to purify PCR amplicon using the capture beads and added to indicate the start position of the target gene sequence, respectively. Each MID-tag had a unique sequence and each was added to the sequence data different from samples that were generated by a single sequencing run. To read the target sequence from both direction, sense and antisense fusion primers were designed, and sequences of the annealing region for these primers were 5'-GCC ATC AGT AGA AGG TAG CAG-3' and 5'-CAG TAG TGT CGA AGA ACT CC-3', respectively. RT-PCR was performed using a protocol similar to that described above, except for that 30 cycles were used to detect fragments present in low quantities. Preparation of emPCR, purification of PCR products, and analysis using the GsJr were performed according to the manufacturer's protocol. Obtained data were analyzed using Amplicon Variant Analyzer (Roche) and the cryptic splice sites were manually curated.

Results and Discussion

Modification of the polypyrimidine tract activates several noncanonical acceptor sites near the original site

We attempted to develop a reporter gene that contains the non-AG acceptor site present in an intron predicted to be U2-dependent. For this purpose, we chose intron 18–19 of *FNI*, because this intron is constitutively spliced, and the length of the fragment is suitable for cloning using a plasmid vector (17). The reporter genes contained intron 17–18 and intron 18–19 of *FNI* (Fig. 1A). The splice acceptor site of the second intron (intron 18–19) was mutated, and the first intron (intron 17–18) was unchanged. Because splicing is coupled with other mRNA processing steps, and inhibition of splicing causes defects in other processes (18), we did not mutate the sequence of intron 17–18 of the minigene; therefore, its transcript was always spliced at least once. We modified the Py tract preceding the mutated AG to potentially increase the binding affinity of the mutated transcripts to U2AF65.

If both introns are spliced normally, GFP and luciferase are expressed as a fusion protein, and GFP fluorescence and luciferase luminescence are detected. If a mutation of the AG dinucleotide impedes splicing, intron 18–19 will be retained or a cryptic downstream acceptor site occurring in the luciferase ORF will be used. In either case, GFP-fusion reporter protein will lack the active luciferase domain by a frame-shift or large deletion mutation, resulting in GFP fluorescence alone (Fig. 1B). Thus, the comparison of luciferase luminescence with GFP fluorescence indicates the splicing efficiency at the

mutated acceptor site. We found that the relative luciferase activity of the wild-type reporter pFD-Luc-AG was approximately the same as that of the positive control reporter pFD-Luc-cDNA that did not contain introns, indicating that the reporter construct was normally spliced when there was no mutation (Fig. 2A). In contrast, the pFD-Luc-AT reporter that contained a G to A mutation did not generate detectable luciferase activity; however, the pFD-Luc-Py-AT reporter with the modified Py tract generated a detectable signal, although it was much weaker compared with that of the wild-type reporter.

To confirm that the signal was reflecting altered splicing efficiency, the spliced mRNA was analyzed using RT-PCR. Shorter bands, which represented splicing at the downstream cryptic acceptor site, were generated by mutant reporters (Fig. 2B). The position of the cryptic site is indicated as #8 in the sequence shown in Figure 2D. In addition to the shorter band, a longer band was generated by pFD-Luc-Py-AT that was similar to the size generated by the wild-type reporter. These findings suggest that the increased number of pyrimidine residues in the Py tract allowed splicing at the mutated acceptor site. However, it was unclear if the band corresponded to a transcript spliced at the original site. Therefore, we analyzed RT-PCR products using a sequencing gel that resolved DNA molecules with higher resolution.

Consequently, the larger band detected using pFD-Luc-Py-AT was resolved into seven bands, whereas a single band was detected using the wild-type reporters (Fig. 2C).

Sequence analysis revealed that band #2 was generated by splicing at the original site and that bands #1 and #3–8 were each generated by splicing at seven different adjacent sites, respectively (Fig. 2D). Among these cryptic sites, splicing products labeled #1, #2, and

#4 were in-frame, and these three isoforms were likely responsible for the luciferase activity. This explains the lack of strict correlation between the relative luciferase activities to the ratios of the shorter to longer bands detected in the agarose gels. Because the intron sequence of our reporter does not match that of the consensus sequence of U12-dependent introns, we assume that splicing events detected here at the noncanonical acceptor site were catalyzed by the major spliceosome. The requirement for an extended Py tract that may enhance the binding of transcripts to U2AF65 is consistent with this assumption, although further studies are required to elucidate the mechanism.

Sequence-specific effects on the recognition of noncanonical acceptor sites

To determine if these unusual splicing events were caused by other mutation of the AG dinucleotide, we tested other mutant reporters. The luciferase activities generated by these mutant reporters were generally comparable with that of the original mutant reporter (Fig. 3B and C). In particular, the activities of all-1G mutants were weaker compared with those of the-2A mutants, although there were no large differences among the band intensities in the agarose gels. Further, sequence analysis revealed a significant difference in the splicing efficiency at the original position (Fig. 3D). The-1G > T mutation that we first tested was the only mutation among the-1G mutations that allowed splicing at the original position. Lack of splicing at the AA or AC sites suggests the presence of another fidelity mechanism, which cannot be bypassed by modification of Py-tract in mammalian cells. All-2A mutations allowed splicing at the original position. TG is the most frequently used sequence in the U2-dependent noncanonical acceptor sites in the human genome (13), which may correlate to the higher levels of luciferase activity generated by the TG mutants.

Identification of luteolin as a splicing enhancer at noncanonical acceptor sites

Because our reporter detected splicing at the non-AG acceptor site, we tested the effects on the splicing of 11 compounds used in Chinese traditional medicine. Because the effects of these compounds on cell viability and reporter activity varied, we calculated the difference in induction relative to the luciferase activity generated by the pFD-Luc-cDNA (Fig. 4A) and found that luteolin (marked as #8) increased the relative luciferase activity approximately two-fold, although luteolin increased the basal level of luciferase activity of all reporters through an unknown mechanism.

To assess whether luteolin affected splicing efficiency, we treated cells with luteolin at concentrations ranging from 5 μ M to 50 μ M and analyzed the sizes of the RT-PCR products using agarose gel electrophoresis (Fig. 4B). Luteolin increased the ratio of the larger to the shorter band in a concentration-dependent manner, suggesting that luteolin enhances splicing at the proximal noncanonical acceptor site (Fig. 4C). Because luteolin is a flavone, we tested the flavones apigenin, genistein, and naringenin and found that only apigenin had a similar effect on splicing as that of luteolin (Fig. 4D and E). A recent study has identified hnRNPA2 to be a target of apigenin and shows that apigenin affects alternative splicing (19). Furthermore, the study reported that luteolin, but not naringenin or genistein, binds hnRNPA2; this is notable because the specificity of hnRNPA2 binding is consistent with the present results.

Luteolin does not change the splice-site selection at the noncanonical sites

Because the mutations shown above generated different splicing patterns (Fig. 3C), we tested the effects of luteolin on the splicing of transcripts with other mutations. Luteolin

uniformly increased the levels of the longer isoforms of all mutant reporters, (Fig. 5A and B) and did not generate isoforms that differed from those identified in the absence of luteolin (Fig. 5C). Thus, luteolin may enhance the use of the noncanonical acceptor site by changing the balance of splice-site selection between proximal noncanonical sites (#1–7, Fig. 2D) and distal cryptic AG sites (#8, Fig. 2D). Alternatively, luteolin may decrease rejection rate of a noncanonical acceptor site via relaxing fidelity in AG recognition by U2AF35.

Pre-mRNA processing steps are closely interrelated (20) and that coupling is important for precise recognition of the splice site (21). Because our method combined transient transfection and gene expression driven by the CMV promoter, an excess level of transcripts may uncouple transcription and splicing. The extended Py tract may contribute to restoration of splicing of the mutated reporter transcript in the present model system. Thus, it remains unclear whether a naturally occurring mutation in a U2-dependent intron causes similar use of noncanonical splice sites, whereas the fidelity of acceptor site recognition in U12-dependent introns is lower and activation of adjacent noncanonical acceptor sites occurs in an endogenous gene in a mutated U12-dependent intron (22-24). And also to verify the applicability of our reporter system, it would be interesting to test the effect of luteolin on an endogenous mutated acceptor site. Further investigations to determine the mechanism employed by luteolin to enhance unusual splicing may facilitate the development of strategies to prevent pathologically aberrant splicing by a mutation into the acceptor AG dinucleotide.

Acknowledgements

We thank Rie Nakaido, Yumi Komatsu, and the staff of the Center for Research and Education on Drug Discovery, Hokkaido University, for their assistance. This study was supported by JSPS KAKENHI Grant Number 25460057 (to HM), by a Grant-in-Aid (2013) from the Cooperative Research Project of the Institute of Natural Medicine, University of Toyama, and by a grant from the Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Conflict of Interest

The authors declare no competing financial interest.

Figure Legends

Figure 1 (A) Structure of reporters. RT-PCR primers are indicated. (B) Schematic representation of the reporter assay.

Figure 2 (A) Ratio of relative luciferase activity to that of the wild-type reporter pFD-Luc-AG. Numbers on the bars indicate the average activity of each reporter (ND, no signal detected). Error bars represent the standard deviation (SD) ($n = 3$). (B) Agarose gel electrophoresis (2%) of RT-PCR products in. (C) Analysis of RT-PCR products electrophoresed using a sequencing gel. Numbers on the right side of the panel correspond to the numbers of the sequences indicated below. (D) Junction sequence between the second and the third exons. Nucleotides indicated with a capital letter are substitutions from (g) or (a) nucleotides shown below. Original AG and distal cryptic AG

splice sites and their polypyrimidine tracts (dotted line) are enclosed in boxes.

Figure 3 (A) Mutated sequences. (B) Ratio of relative luciferase activity to that of the wild-type AG reporter. Numbers above the bars represent the average ratio, and error bars represent the SD ($n = 3$). $**P < 0.01$ (ANOVA and Tukey-Kramer test). (C) Detection of RT-PCR products using agarose gel electrophoresis. (D) Detection of RT-PCR products using a sequencing gel. Marker and arrowheads are same as those in Figure 2C.

Figure 4 (A) Differences in induction of relative luciferase activity by the compounds as follows: #1, alkannin; #2, atropine sulfate; #3, bisdemethoxycurcumin; #4, cinobufagin; #5, cinobufotalin; #6, dehydrocostuslactone; #7, dimethylesculetin, #8, luteolin; #9, saikosaponin d; #10, shikonin; and #11, timosaponin A-III. Error bars shown indicate the standard error of the mean SE ($n = 3$). In case of alkannin, two independent experiments were performed and error bars indicate the SD values. (B) Agarose gel electrophoresis analysis of luteolin's effect on splicing. (C) Quantification of changes in splicing showing the ratio of intensities of the larger and smaller bands. (D) Structures of tested flavones (E) Quantification of changes in flavone-mediated splicing. Bars indicate the differences between flavone-treated compare with DMSO-treated samples. The SD values are shown as error bars ($n = 3$). $*P < 0.05$, $**P < 0.01$ (ANOVA and Tukey-Kramer test).

Figure 5 (A) RT-PCR analysis of luteolin's effect on splicing transcripts with different mutations. (B) Quantification of the changes in splicing change detected using RT-PCR.

N.S., no significant difference detected using ANOVA and the Tukey–Kramer test. (C)

High-resolution analysis of RT-PCR products using a sequencing gel. Markers and arrowheads are same as those shown in Figure 2C.

Supplementary file

Results of BLAST search using the FN1-derived sequence in a reporter gene. Detected changes are indicated by a gray box. Arrows indicate the start and end positions of each exons.

References

- [1] Roca, X., Krainer, A.R., Eperon, I.C. (2013) Pick one, but be quick: 5' splice sites and the problems of too many choices. *Genes Dev* 27: 129-144.
- [2] Reed, R. (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr Opin Cell Biol* 12: 340-345.
- [3] Wu, S., Romfo, C.M., Nilsen, T.W., Green, M.R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402: 832-835.
- [4] Zamore, P.D., Patton, J.G., Green, M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 355: 609-614.
- [5] Merendino, L., Guth, S., Bilbao, D., Martinez, C., Valcarcel, J. (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* 402: 838-841.
- [6] Zorio, D.A., Blumenthal, T. (1999) Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* 402: 835-838.
- [7] Reed, R. (1989) The organization of 3' splice-site sequences in mammalian introns. *Genes Dev* 3: 2113-2123.
- [8] Guth, S., Martinez, C., Gaur, R.K., Valcarcel, J. (1999) Evidence for substrate-specific requirement of the splicing factor U2AF(35) and for its function after polypyrimidine tract recognition by U2AF(65). *Mol Cell Biol* 19: 8263-8271.
- [9] Soares, L.M., Zanier, K., Mackereth, C., Sattler, M., Valcarcel, J. (2006) Intron removal requires proofreading of U2AF/3' splice site recognition by DEK.

- Science 312: 1961-1965.
- [10] Tavanez, J.P., Madl, T., Kooshapur, H., Sattler, M., Valcarcel, J. (2012) hnRNP A1 proofreads 3' splice site recognition by U2AF. *Mol Cell* 45: 314-329.
 - [11] Wang, J., Zhang, J., Li, K., Zhao, W., Cui, Q. (2012) SpliceDisease database: linking RNA splicing and disease. *Nucleic Acids Res England* 40: D1055-1059.
 - [12] Aebi, M., Hornig, H., Padgett, R.A., Reiser, J., Weissmann, C. (1986) Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* 47: 555-565.
 - [13] Parada, G.E., Munita, R., Cerda, C.A., Gysling, K. (2014) A comprehensive survey of non-canonical splice sites in the human transcriptome. *Nucleic Acids Res* 42: 10564-10578.
 - [14] Pollard, A.J., Krainer, A.R., Robson, S.C., Europe-Finner, G.N. (2002) Alternative splicing of the adenylyl cyclase stimulatory G-protein G alpha(s) is regulated by SF2/ASF and heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) and involves the use of an unusual TG 3'-splice Site. *J Biol Chem* 277: 15241-51.
 - [15] Bonnal S., Vigevani L., Valcarcel J. (2012) The spliceosome as a target of novel antitumour drugs. *Nat Rev Drug Discov*;11: 847-59.
 - [16] Maita, H., Tomita, K., Ariga, H. (2014) A split luciferase-based reporter for detection of a cellular macromolecular complex. *Anal Biochem* 452: 1-9.
 - [17] Pandya-Jones, A., Black, D.L. (2009) Co-transcriptional splicing of constitutive and alternative exons. *RNA* 15: 1896-1908.
 - [18] Sanford, J.R., Caceres, J.F. (2004) Pre-mRNA splicing: life at the centre of the central dogma. *J Cell Sci* 117: 6261-6263.
 - [19] Arango, D., Morohashi, K., Yilmaz, A., Kuramochi, K., Parihar, A., Brahimaj, B., Grotewold, E., Doseff, A.I. (2013) Molecular basis for the action of a dietary flavonoid revealed by the comprehensive identification of apigenin human targets.
 - [20] Neugebauer, K.M. (2002) On the importance of being co-transcriptional. *J Cell Sci* 115: 3865-3871.
 - [21] Pandya-Jones, A. (2011) Pre-mRNA splicing during transcription in the mammalian system. *Wiley Interdiscip Rev RNA* 2: 700-717.
 - [22] Levine, A., Durbin, R. (2001) A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res* 29: 4006-4013.
 - [23] Hastings, M.L., Resta, N., Traum, D., Stella, A., Guanti, G., Krainer, A.R. (2005) An LKB1 AT-AC intron mutation causes Peutz-Jeghers syndrome via splicing at noncanonical cryptic splice sites. *Nat Struct Mol Biol* 12: 54-59.
 - [24] Dietrich, R.C., Fuller, J.D., Padgett, R.A. (2005) A mutational analysis of

U12-dependent splice site dinucleotides. RNA 11: 1430-1440.

Figure 1

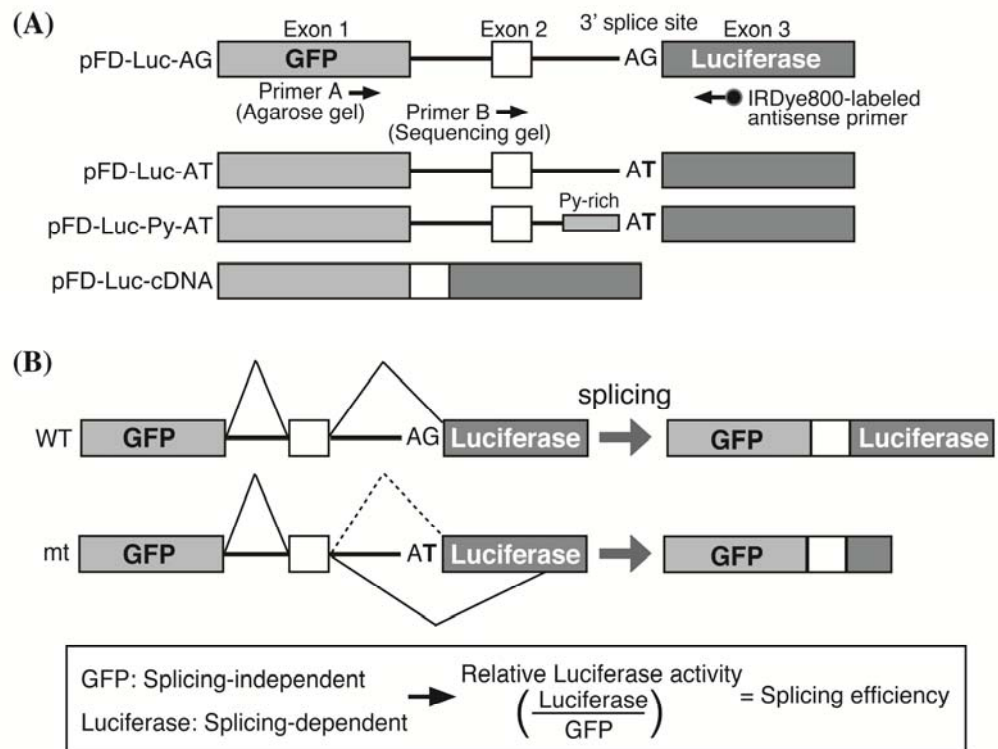


Figure 2

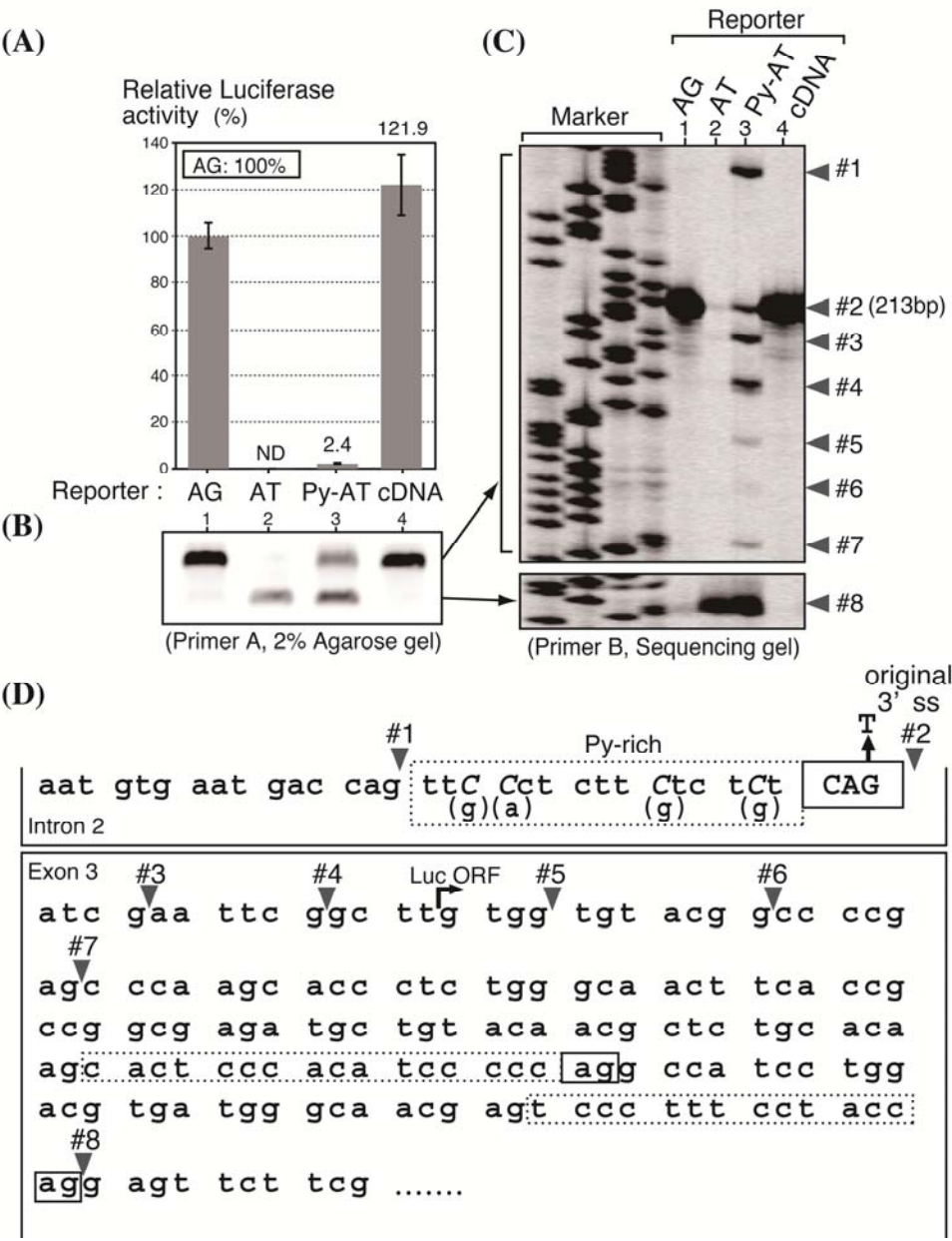


Figure 3

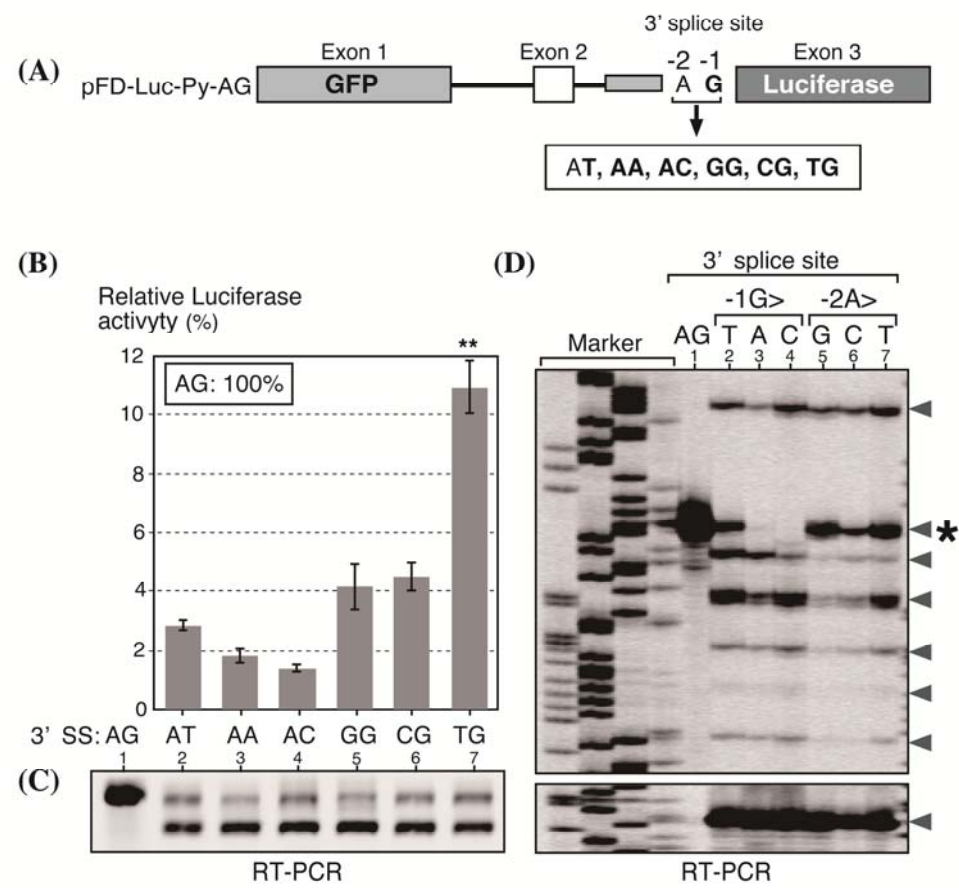


Figure 4

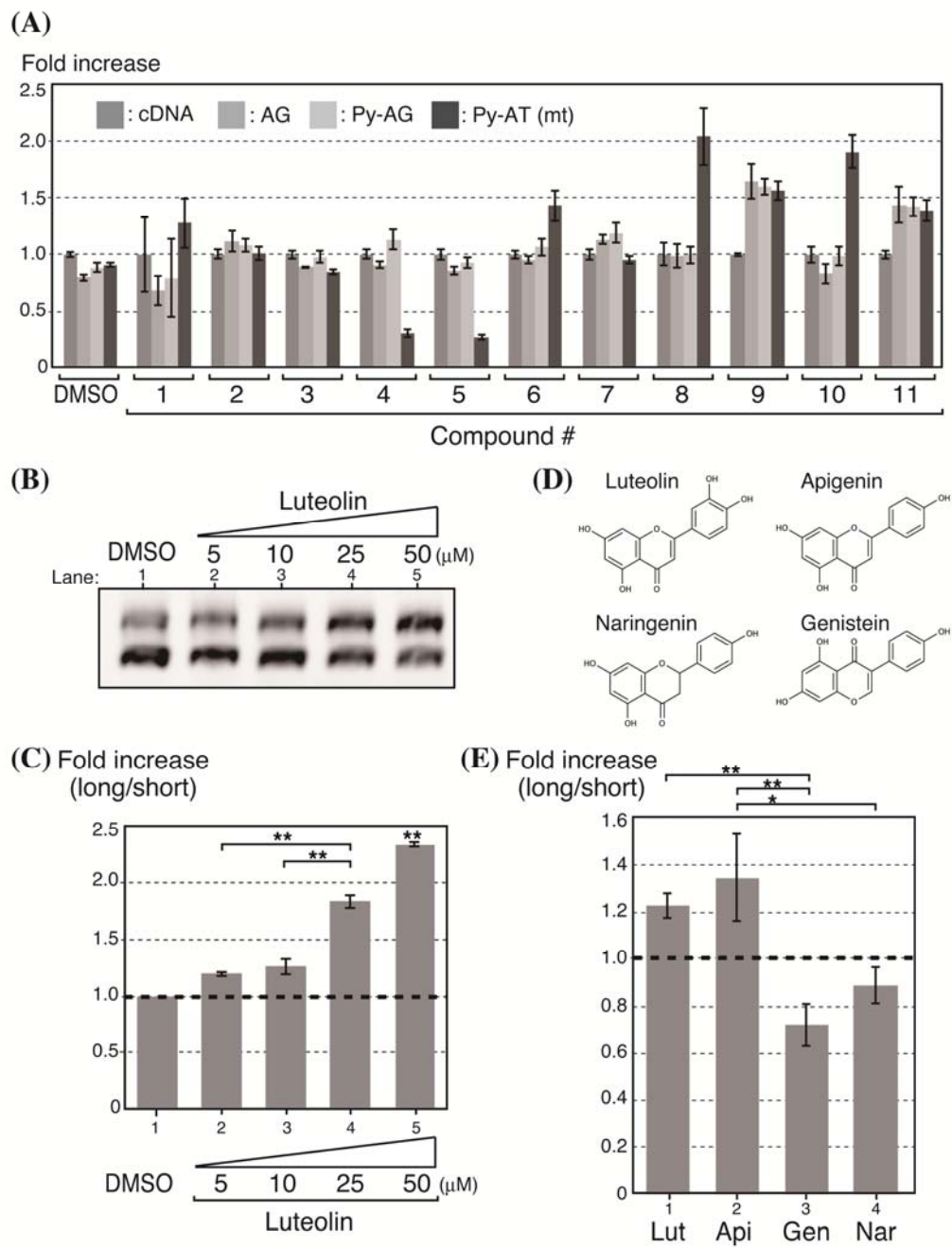
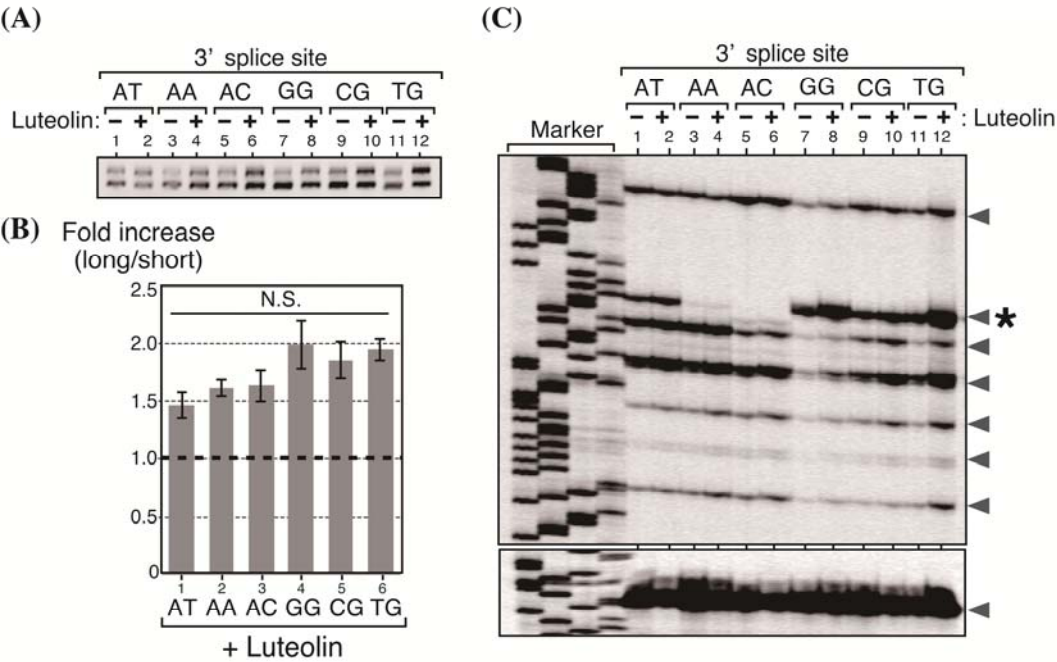


Figure 5



Supplementary file 1

Results of BLAST search using the FN1-derived sequence in a reporter gene. Changes are indicated by a gray box. Allows indicate the start and end positions of each exons.

Query 1 **exon17**→ CGCCTGATGCCCCCTCCTGACACGACTGTGGACCAAGTTGATGACACCTCAATTGTTGTTC 60
 |||||||||||||||||||
 Sbjct 32872 CGCCTGATGCCCCCTCCTGACCCGACTGTGGACCAAGTTGATGACACCTCAATTGTTGTTC 32931
 Query 61 ←**exon17** GCTGGAGCAGACCCCAGGCTCCCATCACAGGTCAGCTAAGCGTCCCCCTCTTTGGCTGCT 120
 |||||||||||||||||||
 Sbjct 32932 GCTGGAGCAGACCCCAGGCTCCCATCACAGGTCAGCTAAGCGTCCCCCTCTTTGGCTGCT 32991
 Query 121 ATGTTAATCTTAATGACATCAGCAGGGAGGGCGCAGATTCTGACTGCGGACCTGCATATC 180
 |||||||||||||||||||
 Sbjct 32992 ATGTTAATCTTAATGACATCAGCAGGGAGGGCGCAGATTCTGACTGCGGACCTGCATATC 33051
 Query 181 ACTTTAAATCTCCAATATAATTTATGGGAGAGGGGTTtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 240
 |||||||||||||||||||
 Sbjct 33052 ACTTTAAATCTCCAATATAATTTATGGGAGAGGGGTTGTGTGTGTGTGTGTGTGTGTGTGT 33111
 Query 241 g--GCggggg--gggggAGTTATTTCTATGGCACATTTCCTTGAACCATTTCACCAA 297
 |||||||
 Sbjct 33112 GTGGCGGGGTGGGGAGTTATTTCTATGGCACATTTCCTTGAACCATTTCACCAA 33171
 Query 298 CTCCCTTATACATACACACCACAACATACACACAACCTGTAAAGCCAGCTCATTGGCTTA 357
 |||||||
 Sbjct 33172 CTCCCTTATACATACACACCACAACATACACACAACCTGTAAAGCCAGCTCATTGGCTTA 33231
 Query 358 TTAAAGCAAGTGTTCCCAGGGTTGAAGAGGTGTAATTCCTGAAAACGTTGCTCTAAGAT 417
 |||||||
 Sbjct 33232 TTAAAGCAAGTGTTCCCAGGGTTGAAGAGGTGTAATTCCTGAAAACGTTGCTCTAAGAT 33291
 Query 418 TTATCCTTAAGGAGAAAGCTGAGCTGTCGTCTTAGCTCATTAGGTGATTCAACTGCCTCA 477
 |||||||
 Sbjct 33292 TTATCCTTAAGGAGAAAGCTGAGCTGTCGTCTTAGCTCATTAGGTGATTCAACTGCCTCA 33351
 Query 478 TCACTGAAGTTCCAAAAGACACACACAGTGCTAGACAACCTCTGCTTAGGCTGGTTCATT 537
 |||||||
 Sbjct 33352 TCACTGAAGTTCCAAAAGACACACACAGTGCTAGACAACCTCTGCTTAGGCTGGTTCATT 33411
 Query 538 AATTGCTTCCCTCGTCTGGAGCTCAAAGAGGAAAAATCAGCTTAACATGAATATTTTCAC 597
 |||||||
 Sbjct 33412 AATTGCTTCCCTCGTCTGGAGCTCAAAGAGGAAAAATCAGCTTAACATGAATATTTTCAC 33471
 Query 598 CTAATGGCATCTCTAATTGACATTTATTAAGGATGTCAGGTCTTCAAGGATGACATTTAT 657
 |||||||
 Sbjct 33472 CTAATGGCATCTCTAATTGACATTTATTAAGGATGTCAGGTCTTCAAGGATGACATTTAT 33531

Query	658	TATTA AAAAGGTTTCGCATGACTGTTCTTATTTTATCTTCGTGCTGAATAGTCATTATTAG	717
Sbjct	33532	TATTA AAAAGGTTTCGCATGACTGTTCTTATTTTATCTTCGTGCTGAATAGTCATTATTAG	33591
Query	718	AAGAAGTGGCAATATTCAAAGAGTCAAAAAGTATCACTGGCTCTTCACTAATCAAGCAAG	777
Sbjct	33592	AAGAAGTGGCAATATTCAAAGAGTCAAAAAGTATCACTGGCTCTTCACTAATCAAGCAAG	33651
Query	778	ATGCTAAGGGATATTAGAAAAGGAGGATTTATGGTGTTCCTAAGCCTCTGTCTCAAAGA	837
Sbjct	33652	ATGCTAAGGGATATTAGAAAAGGAGGATTTATGGTGTTCCTAAGCCTCTGTCTCAAAGA	33711
Query	838	AAACAGTGCATCTTACTTTTGCTCATGAATCTGCAGGGTACAGAATAGTCTATTTCGCCAT	897
Sbjct	33712	AAACAGTGCATCTTACTTTTGCTCATGAATCTGCAGGGTACAGAATAGTCTATTTCGCCAT	33771
Query	898	CAGTAGAAGGTAGCAGCACAGAACTCAACCTTCCTGAAACTGCAAACTCCGTCACCCCTCA	957
Sbjct	33772	CAGTAGAAGGTAGCAGCACAGAACTCAACCTTCCTGAAACTGCAAACTCCGTCACCCCTCA	33831
Query	958	GTGACTTGCAACCTGGTGTTCAGTATAACATCACTATCTATGCTGTGGAAGAAAATCAAG	1017
Sbjct	33832	GTGACTTGCAACCTGGTGTTCAGTATAACATCACTATCTATGCTGTGGAAGAAAATCAAG	33891
Query	1018	AAAGTACACCTGTTGTCAATTCAACAAGAAACCACTGGCACCCTCGCTCAGGTAACTTT	1077
Sbjct	33892	AAAGTACACCTGTTGTCAATTCAACAAGAAACCACTGGCACCCTCGCTCAGGTAACTTT	33951
Query	1078	TTAAGAAGACACTTCCTATGTTATCTTAAACAGGATTGTTCCCTGAAGGAGGGTTGTTTGT	1137
Sbjct	33952	TTAAGAAGACACTTCCTATGTTATCTTAAACAGGATTGTTCCCTGAAGGAGGGTTGTTTGT	34011
Query	1138	CTCTGTCAACAGTCCTCATTCAAGAAATCTTATATATTAGTTTTCCCTAAACTTCTG	1197
Sbjct	34012	CTCTGTCAACAGTCCTCATTCAAGAAATCTTATATATTAGTTTTCCCTAAACTTCTG	34071
Query	1198	ATATTTAGCTGAAATGTCATAAGTAACCTATCAAAGCTGGCTACTGGCCTTTCTGATTAA	1257
Sbjct	34072	ATATTTAGCTGAAATGTCATAAGTAACCTATCAAAGCTGGCTACTGGCCTTTCTGATTAA	34131
Query	1258	AAACTGACACCATAACATCCATCTACAAATTTCCCTAGAAAGCTTAAGGGTCATTTTCAT	1317
Sbjct	34132	AAACTGACACCATAACATCCATCTACAAATTTCCCTAGAAAGCTTAAGGGTCATTTTCAT	34191

Query	1318	TTTGTATTCTTAAGTATTATAAAATGATTCAGTAAAACAAAACTGTTACATTATTTCT	1377
Sbjct	34192	TTTGTATTCTTAAGTATTATAAAATGATTCAGTAAAACAAAACTGTTACATTATTTCT	34251
Query	1378	GCAGTTATTTCAAAGGCTTTTCTTaaaaaaTTTGTAGTATCTTTCTTATAACCCCTCTC	1437
Sbjct	34252	GCAGTTATTTCAAAGGCTTTTCTTAAAAAATTTTGTAGTATCTTTCTTATAACCCCTCTC	34311
Query	1438	CCCCCACCACCATATGACACTTCATATGCTGGTAACTCACTTATTACCCATTTTAGATAA	1497
Sbjct	34312	CCCCCACCACCATATGACACTTCATATGCTGGTAACTCACTTATTACCCATTTTAGATAA	34371
Query	1498	AAGGTTCAAATGTCATAATTTAAGCACTATGACTGGACCATGAAAATGTGATCTGATTAA	1557
Sbjct	34372	AAGGTTCAAATGTCATAATTTAAGCACTATGACTGGACCATGAAAATGTGATCTGATTAA	34431
Query	1558	GAGACAAAACCTTAGCAAAACTCCCGAATGAGAGGCTCAATGAACTGCCTAACATATCCAA	1617
Sbjct	34432	GAGACAAAACCTTAGCAAAACTCTCGAATGAGAGGCTCAATGAACTGCCTAACATATCCAA	34491
Query	1618	GAAACTGGCTTCTTAAAAGTAATCTTCAGCAGAGAATGTGAATGACCAGTT	1668
Sbjct	34492	GAAACTGGCTTCTTAAAAGTAATCTTCAGCAGAGAATGTGAATGACCAGTT	34542

exon19 →