Keiko Tadokoro · Mayu Yamazaki-Inoue
Maki Tachibana · Mina Fujishiro · Kazuaki Nagao
Masashi Toyoda · Miwako Ozaki · Masami Ono
Nobuhiro Miki · Toshiyuki Miyashita · Masao Yamada

Frequent occurrence of protein isoforms with or without a single amino acid residue by subtle alternative splicing: the case of Gln in *DRPLA* affects subcellular localization of the products

Received: 12 May 2005 / Accepted: 17 May 2005 / Published online: 10 August 2005
© The Japan Society of Human Genetics and Springer-Verlag 2005

Abstract Protein isoforms with or without a single amino acid residue make a subtle difference. It has been documented on a few genes that alternative splicing generated such isoforms; however, the fact has attracted little attention. We became aware of a subtle sequence difference in *DRPLA*, a polyglutamine disease gene for dentatorubral pallidoluysian atrophy. Some reported cDNA sequences lacked 3 nucleotides (nt) (CAG), which were positioned apart from the expandable and polymorphic CAG repeats and also coded for glutamine. We experimentally confirmed that the difference was indeed generated by alternative splicing utilizing two acceptors separated by 3 nt. In *DRPLA*, the expression ratio of two mRNA isoforms was almost constant among tissues, with the CAG-included form being major. The glutamine-included protein isoform was more predominantly localized in the nucleus. Database searching revealed that alternative splice

acceptors, as well as donors, are frequently situated very close to each other. We experimentally confirmed two mRNA isoforms of 3 nt difference in more than 200 cases by RT-PCR and found interesting features associated with this phenomena. Inclusion of 3 nt tends to result in single amino acid inclusion despite the phase of translational frame. The expression ratio sometimes varied extensively among tissues.

Keywords Subtle alternative splicing · Splice acceptors · Polyglutamine diseases · *DRPLA* · Subcellular localization

K. Tadokoro · M. Yamazaki-Inoue · M. Tachibana
M. Fujishiro · K. Nagao · M. Toyoda · T. Miyashita
M. Yamada (✉)
National Research Institute for Child Health and Development,
2-10-1 Ohkura, Setagaya-ku, Tokyo 157-8535, Japan
E-mail: myamada@nch.go.jp
Tel.: +81-3-34160181
Fax: +81-3-54947035

M. Tachibana
Department of Pediatrics, Tokyo Medical University,
Tokyo, Japan

M. Fujishiro
Laboratory of Nucleic Acid Science, Nihon University,
Fujisawa, Japan

M. Ozaki
Laboratory for Memory and Learning, Brain Science Institute,
Riken, Wako, Japan

M. Ono · N. Miki
Institute of Clinical Endocrinology, Tokyo Women's
Medical University, Tokyo, Japan

Introduction

Alternative splicing of pre-mRNA is an important regulatory mechanism for quantitative control of gene expression and also for generating diverse protein species (Lopez 1998; Modrek and Lee 2002; Black 2003). It is highlighted, as the human-genome-sequencing project has revealed a limited number of genes (20,000–25,000) in the genome, which were far less than expected for functional complexity (International Human Genome Sequencing Consortium 2004). Recent surveys in databases revealed that as many as 40–60% of human genes are associated with alternative splicing (Black 2003). Three fundamental types of alternative splicing are noted: one is a choice in use of the entire potential exonic sequence (cassette exon or skipped exon) and the other two are selection of the site for a donor or acceptor in the potential continuous exonic sequence. Other types, mutually exclusive exon and intron retention, may be explained by the combination of the fundamental types. The choice for the first exon in alternative splicing is sometimes associated with differential regulation in expression, as regulatory elements are usually situated near the transcriptional start site. Alternative splicing

occurring in the coding region gives rise to generation of protein isoforms. Recent surveys also revealed that 70–88% of alternative splices cause a change in protein products (Black 2003). As one event of alternative splicing in a coding region usually generates a pair of protein isoforms, a multiple combination of alternative splicing events in a gene may generate an enormous number of protein isoforms, sometimes reaching to dozens or even hundreds.

Among diversified protein isoforms, a pair of isoforms with or without a single amino acid residue is the case with a minimum difference. Although it has been documented that such isoforms are generated by alternative splicing on a handful of genes (Manrow and Berger 1993; Condorelli et al. 1994; Vogan et al. 1996; Oberkoffler et al. 1997; Lin et al. 2000) (Table 1), the phenomenon seems not to be justly recognized. The difference in amino acid as well as nucleotide sequences is so subtle that it may be disregarded or explained by experimental and/or human errors. We became aware of such a subtle difference in a previously reported cDNA sequence after comparison with sequences that other laboratories reported.

Dentatorubral pallidoluysian atrophy (DRPLA) is an autosomal dominant neurodegenerative disorder characterized by selective neuron loss in the cerebellar and pallidal outflow pathways (Kanazawa 1998). Patients show a combination of ataxia and extrapyramidal signs (chorea and athetosis) to varying degrees. We detected expansion of CAG repeats on chromosome 12p13 in the patients (Nagafuchi et al. 1994a) and then determined the entire cDNA sequence of the gene (Nagafuchi et al. 1994b). Several other neurodegenerative disorders, including Huntington's disease have been shown to be caused by expansion of CAG repeats in the coding region of respective genes (Cummings and Zoghbi 2000). As the CAG unit is situated in the translational frame and encodes glutamine, these disorders are collectively called polyglutamine diseases. It has often been observed through these diseases that the age of onset becomes younger with more severity in successive generations. This phenomenon, genetic anticipation, is now explained by the following two facts. The severity is correlated to the number of repeat iterations, and the extended repeats are further expanded when transmitted to the next generation. DRPLA is typical in genetic anticipation when the extended allele is transmitted paternally (Nagafuchi et al. 1994a; Cummings and Zoghbi 2000). As repeat expansion is a novel type of mutation associated with a subset of disorders and shows unique features, it attracts broad attention. The molecular mechanism underlying cell death has been demonstrated as induction of apoptosis at the final step, and several processes have been proposed for the route where the cohesive force of the polyglutamine tract is central (Okamura-Oho et al. 2003; Michalik and Van Broeckhoven 2003; Forman et al. 2004). In contrast, it is still uncertain how and why specific subsets of neurons are degenerated in respective polyglutamine diseases. To

address this issue, we have been studying the normal functions of the *DRPLA* gene and its product.

In this report, we demonstrate an alternative splice in the *DRPLA* gene, which originated from a subtle sequence difference. We characterized the DRPLA protein isoforms with or without a single glutamine residue and found that such a subtle difference affected the protein nature. We extensively searched for similar phenomena in the literature and databases, experimentally examined cases, and revealed widespread occurrence of protein isoforms with or without a single amino acid residue in the human proteome due to alternative splicing. We propose the term of “subtle alternative splicing” for alternative splicing to generate a subtle difference, such as 3 nt, in mRNA and a single amino acid residue in protein.

Materials and methods

RT-PCR

RNA preparations from human, mouse, and rat tissues were purchased from several companies, including BD Biosciences Clontec, and Ambion and RNA from cultured cells and rodent tissue was prepared with a standard method, as described previously (Tadokoro et al. 1993; Miyashita et al. 1997). Fine, dissected, rat brain was performed with the aid of markers and ascertained by other means (Ozaki et al. 2004). One or two microgram of total RNA was subjected to direct cDNA synthesis using Superscript II RNaseH minus reverse transcriptase with the AP primer (an oligo-dT primer with an attached sequence) under the conditions recommended for the 3'-RACE system (Invitrogen, UK). An aliquot of synthesized cDNA was subjected to PCR amplification in a reaction containing 200 μ M of each deoxynucleotide triphosphate, 0.5 μ M of each primer, and 1 U of rTaq polymerase (Takara-bio). The PCR was carried out with a thermocycler in a cycling condition with an initial denaturation of 4 min at 94°C followed by 25–35 cycles of denaturation at 94°C for 1 min, annealing at T_m –10°C for 1 min, and extension at 72°C for 1 min, with a final extension at 72°C for 4 min (Tadokoro et al. 1993). The amounts of synthesized DNA from respective tissues were primarily adjusted to generate an equal level of *GAPDH* products, but the amounts were sometimes readjusted to facilitate the detection of two products with a 3 nt difference. Amplified PCR products were separated by electrophoresis through a 10% polyacrylamide or 3% SeaKem LE agarose gel (FMC BioProducts) and then detected by silver staining (Bio-Rad, Mississauga, ON, Canada) or ethidium bromide staining. Images were captured through a CCD camera, and the integrated optical density of detected bands was measured by the ImagePro-Plus and GelPro image analysis software (Media Cybernetics, Silver Spring, USA). In a mass screening, the ratio of two products with a 3 nt difference was represented in terms

Table 1 Representative genes with subtle alternative splicing utilizing two acceptor sites separated by 3 nucleotides (nt)

Genes ^a	Accession numbers		Genome	Exon	References
	cDNA with 3 nt	cDNA without 3 nt			
DRPLA	D31840	NM_001940 (U23851)	NT_000012 (U47924)	4–5	This study
GHRHR	–	NM_000823(L01406)	NT_000007(AC005155)	1–2	This study
BAIAP2	BC014020	NM_017450 (AB015019)	NT_000017(AC115099)	9–10	This study
PTMA	M14630	NM_002823(BC022433)	NT_000002(AC073476)	2–3	Manrow et al. (1993)
IGF1R	NM_000875 (X04434)	–	NC_0000156 (AY332722)	13–14	Condorelli et al. (1994)
PAX3	NM_181460 (AY251280)	BC008826	NT_000002 (AC010980)	2–3	Vogan et al. (1996)
PAX7	NM_002584 (X96743)	–	NT_000001 (AL021528)	2–3	Vogan et al. (1996)
LEP	NM_000230 (U43653)	D49487	NT_000007 (AC018635)	2–3	Oberkofler et al. (1997)
Dnmt1	AF162282	NM_010066 (X14805)	NT_039472 (CAAA01059910)	4–5	Lin et al. (2000)
CAST	NM_001750 (D16217)	NM_173061 (U58996)	NT_000005 (AC008906)	27–28	This study
MAN2B1	NM_000528 (U60266)	U05572	NT_000017 (AC010422)	7–8	This study
PSEN2	NM_000447 (L43964)	NM_012486 (U34349)	NT_000001 (AL391628)	10–11	This study
LAP1B	AK001780	NM_015602 (AK021613)	NT_000001 (AL353708)	2–3	ENSG00000143337
NOXO1	AB097667	NM_144603 (BC015917)	NT_037887	2–3	ENSG00000162042
CCL20	NM_004591 (D86955)	U64197	NT_005403	1–2	ENSG00000115009
SGNE1	BC005349	NM_003020 (Y00757)	NT_010194 (AJ290438)	3–4	ENSG00000166922
TGFA	NM_003236 (X70340)	BT006833	NT_022184	2–3	ENSG00000163235
Nucleotide sequences ^b			Amino acid changes	Expression ratio in tissues ^c (longer/shorter)	
Splice donor site			Splice acceptor site		
DRPLA	TGAGgtgggaa	gagttcctctttctaacg	CAGGAACTC	Gln/none	8:2–9:1
GHRHR	GACCGtgagta	atcgtgtcactgtttccag	CAGGTATTG	Gln/none	2:8
BAIAP2	ACAAGtaagg	taacgtcctgtccccag	CAGCCGAGA	ThrAla/Thr	1:9–0:10
PTMA	TGCTgtgagtg	atggcgtttttctgag	GAGAAATGAG	Glu/none	0:10–1:9
IGF1R	AAAAGtaaggc	tttcctcctttgtgcag	CAGGATATG	ThrGly/Arg	7:3–8:2
PAX3	CAAGgtgagcg	gccgcgtttctttaaag	CAGGTGACA	Gln/none	0:10–10:0
PAX7	CAGAgtgagtg	tccacctccacctgaag	CAGGTGGCG	Gln/none	0:10–9:1
LEP	CACGgtaagg	tcccttctcctgcatag	CAGTCAGTC	Gln/none	1:9–10:0
Dnmt1	CTTgtgaagg	cactttctgttttaag	CAGTTGAAA	SerVal/Phe	4:6–6:4
CAST	CTCGgtaagca	cagcattatttcttcag	CAGAGTGAC	Gln/none	9:1–10:0
MAN2B1	GCAAGtgagtg	tccgtcctcccccag	CAGGCAGAA	Gln/none	0:10–3:7
PSEN2	ATGGgtgagta	ttcttctcgacacccag	AAGAAGACT	GluGlu/Glu	7:3
LAP1B	CCAGgtaagaa	gttttcttctctatag	CAGTGAGTG	AlaVal/Val	0:10–10:0
NOXO1	CAAGgtgagtg	gccctgttcccgcctcag	AAGAGCCCTC	Lys/none	0:10–9:1
CCL20	GAAGgtaagtg	tcacttttttttttag	CAGCAAGCA	AlaAla/Ala	4:6–9:1
SGNE1	ACAgtgtaacag	aaaccttggcgtttgag	CAGATGATG	AlaAsn/Asn	4:6–8:2
TGFA	AGTgtgagtg	tgcattcttctcccag	CAGACCCGC	AlaAsn/Asn	5:5–9:1

^aAll the genes listed in this table are human, except for mouse *Dnmt1*, as the structure of the corresponding intron–exon boundary in humans differs from that in the mouse. The phenomena for *PAX3* and *PAX7* were originally identified in mice, but the same phenomena were confirmed to occur in humans in this study

^bExon and intron for a longer isoform are indicated with upper and lower cases, respectively. Three nucleotides (nt) excluded in the shorter isoform by alternative splicing (optional 3 nt) are indicated in *bold*. The nucleotides encoding the indicated amino acid residues in the longer form are *underlined*

^cThe expression ratio of the longer to shorter isoform among tissues with a considerable expression level is indicated. When the ratio varied among tissues, the range is given. The figures are rounded out to the nearest whole number to bring the total to ten

37°C. Transient transfection was performed with the polycationic liposome method (LipofectAMINE PLUS Reagent, Invitrogen) according to the manufacturer's instructions. To assess the localization of EGFP-DRPLA fusion protein, cells were cultured in poly-d-lysine-coated glass-bottom plates (MatTek) and incubated in a medium containing 50 μ M Z-VAD-FMK (Peptide Institute Inc.) after transfection (Miyashita et al. 1997). The cells were observed under a confocal microscope (FLUOVIEW, Olympus), and fluorescent-positive cells were counted in terms of the subcellular localization at the indicated period. The expression levels were ascertained by Western blotting, as described previously (Miyashita et al. 1997).

Database search and analyses

Previous reports on alternative splicing making a subtle difference were searched in Entrez Pub Med at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>. It was generally difficult to find because no common terms have been given for such cases. Nevertheless, we have detected a handful of reports although some others may remain undetected. For a subtle difference in amino acid sequences, we surveyed the ExPASy database (Expert Protein Analysis System) and found a file, humpvar.txt, at <http://www.expasy.org/cgi-bin/lists?humpvar.txt>, in which information on sequence variants is summarized from the Swiss-Prot human protein sequence entries. Among the list, we selected the cases of "one amino acid missing" and other subtle differences and then ascertained whether it could occur by subtle alternative splicing based on the standard genome and cDNA sequences at the National Center for Biotechnology Information (NCBI). A source data set of humans in the AltSplice database at <http://www.ebi.ac.uk/asd/altsplice/> was downloaded and analyzed with database software, Kiri (KanriKogaku Kenkyusho Ltd., Japan). Although we used the prerelease Version 2, the release Version 1 was available after May 21, 2004. Throughout this study, we used genomic sequences to facilitate further analyses although RNA sequences may be more appropriate for splicing.

Results

Subtle alternative splicing in DRPLA

After detection of CAG repeat expansion in DRPLA pedigrees, we determined the entire cDNA sequence of the gene (accession number, D31840, Nagafuchi et al. 1994b). Since then, other laboratories have also reported cDNA sequences for the human *DRPLA* gene and orthologues (U23851, D38529 and others; Onodera et al. 1995; Love et al. 1995; Margolis et al. 1996). It is reasonable to find a difference in the number of iterations of the CAG repeats, which start at the 1,700th position of D31840, as it is polymorphic and expandable. Besides the repeats, the U23851 sequence lacked 3 nt occurring in other sequences (CAG at the 518th position of D31840), which resulted in the absence of the single glutamine residue at 94 in the DRPLA protein (also known as atrophin-1). After the entry of the sequence in the NCBI nucleotide database in 1995, we gradually considered the possibility of alternative splicing, as the inconsistent 3 nt was situated at the boundary of exon 4 and 5. To confirm this, we made a minigene system that contained genomic DNA covering a portion of exon 4, intron 4, and exon 5 under the control of SP6 promoter (Fig. 1). When in vitro transcribed products were subjected to react with nuclear extracts, two products with a 3 nt difference in size were generated. Sequencing verified that the products had the authentic cDNA sequences with or without the CAG nucleotides. This result clearly shows that alternative splicing takes place between exons 4 and 5. As the boundary around the distal end of intron 4 and the proximal start of exon 5 has a structure of **cagCAGGAA**, this result strongly suggests that two positions just after the **cag** and **CAG** serve as splice acceptors. (Hereafter, nucleotides in intron and exon are indicated with lower and upper cases, respectively, and the target nucleotides that are included or excluded in respective isoforms are indicated by bold upper case.) However, the result may be explained by other type of alternative splicing, for example, inclusion or exclusion of cassette exon consisting of only 3 nt. There are several **cag** sequences in intron 4, and even one

Fig. 2 Expression ratio of two *DRPLA* mRNA isoforms in various tissues. Total RNA from indicated tissues was analyzed by RT-PCR followed by electrophoresis. The PCR products (73 bp and 70 bp) generated with each form of cDNA as a template were loaded in parallel. As *DRPLA* expression levels in tissues varied

extensively, the amount of transcribed RNA in RT-PCR was adjusted to give a similar level of PCR products. An almost constant expression ratio of the two transcripts was observed in repeated experiments, with the longer form being major, and representative results are illustrated

Fig. 3 Subcellular localization of DRPLA protein isoforms with or without a single glutamine residue. HeLa cells were transiently transfected, with one of the constructs carrying *DRPLA* cDNA (full length) tagged with EGFP at its N-terminus. *Q14* and *Q71* indicate the size of the polyglutamine chain (*Q14* is within the normal range while *Q71* is in a disease range). +*Q* and -*Q* indicate inclusion or exclusion of a single glutamine residue at the 94th position by subtle alternative splicing, respectively. **a** Expression levels were ascertained to be almost equal among various constructs by Western blotting with a mouse anti-GFP monoclonal antibody (B-2) or with a rabbit anti-DRPLA polyclonal antibody at 30 h after transfection. **b** Subcellular localization of the EGFP-DRPLA protein isoforms. We classified them into four categories: *NS* ± *C* predominant nuclear localization as a large speckled form (or aggregates) with minor cytoplasmic distribution, *ND* ± *C* predominant nuclear localization as small speckled or diffused forms with minor cytoplasmic distribution, *N* + *CS* almost equal distribution in the nucleus and cytoplasm, *CS* ± *ND* predominant cytoplasmic localization with minor nuclear distribution. Typical images are illustrated in the upper panels. Respective patterns were counted with 100–200 cells at 30–40 h after transfection and repeated in four independent experiments. Asterisks on error bars show a significant difference of the -*Q* isoform compared with the corresponding +*Q* isoform by statistical analyses with Student's *t* test

agcag sequence. We substituted the g nucleotide at the -1 position to generate the longer transcript (CAG-included form) with one of the other nucleotides (i.e.,

caaCAGGAA). The minigene system with such a substitution only generated the shorter transcript. Thus, we conclude that the alternative splicing between DRPLA exons 4 and 5 takes place utilizing two acceptor sites separated by 3 nt.

Expression ratio of two mRNA isoforms of DRPLA

We previously reported with Northern blotting analyses that DRPLA was expressed in a wide variety of tissues (ubiquitous expression), but the extent varied considerably, with the brain, testis, and ovary being relatively high (Nagafuchi et al. 1994b). To distinguish two transcripts including or excluding the CAG sequence, the primer pair was set in exons 4 and 5 to generate relatively short RT-PCR products. The relative amounts of RT-PCR products varied considerably, which was almost consistent to our previous Northern blot results (data not shown). Thus, the amount of reverse-transcribed DNA from each tissue was adjusted so as to generate an almost equal level to easily detect the ratio of the two forms of transcripts. The expression ratio was almost constant among tissues, with the longer form being major (81%–91%) (Fig. 2, left). We pursued the possibility that either isoform of transcripts was prominent in the specific area of the brain, and we chose the rat because the size of the brain was large enough to be finely dissected, and the *cagCAG* structure at the boundary was conserved. The expression ratio of the two isoforms was almost constant among various sections of rat brain, with the longer form being major (79%–88%, Fig. 2, right). The result indicates that neither form alone nor the ratio of the two isoforms directly determines the brain area or the subset of neurons in DRPLA pathogenesis. We also analyzed mouse tissue during developmental stages and found consistent results (data not shown).

Characterization of DRPLA protein isoforms

Among the characters of DRPLA protein we were especially interested in the subcellular localization. Many previous reports indicate that DRPLA protein is a shuttle flying across the nuclear membrane (Ross 1997; Miyashita et al. 1998; Igarashi et al. 1998; Okamura-Oho et al. 1999; Miyashita et al. 1999; Ellerby et al. 1999; Yanagisawa et al. 2000; U et al. 2001; Okamura-Oho et al. 2003; Nucifora et al. 2003). DRPLA protein has both the nuclear localization and exclusion signals and has implied functions both in the nucleus and cytoplasm. Nonetheless, our group consistently observed predominant nuclear localization of DRPLA protein with very rare cytoplasmic localization in a variety of culture cells with expression experiments as well as for endogenous protein in staining with antibodies (Miyashita et al. 1998; Okamura-Oho et al. 1999; Miyashita et al. 1999; Yanagisawa et al. 2000; U et al.

Table 2 Distribution of distances between two adjacent splice donors as well as two adjacent splice acceptors according to the exon isoforms in the AltSplice data set^a

Difference (nt)	3' end of exon = donor sites (no. of pairs)					5' end of exon = acceptor sites (no. of pairs)				
	A	B	C	D	E	F	G	H	I	J
2	17	4	2	3	2	36	9	4	5	3
3	30	16	11	8	5	600	310	258	199	176
3 nt, after cleaning up						(536	269	220	174	153)
4	108	33	18	13	8	146	57	31	36	21
5	28	13	8	5	3	89	41	27	24	20
6	35	11	9	4	4	71	26	15	12	10
7	14	7	5	2	2	30	7	4	5	3
8	14	6	5	5	4	17	6	2	3	1
9	48	23	18	15	13	35	10	4	3	2
10	19	9	5	4	1	17	5	3	4	3
11	21	12	8	6	5	16	4	0	1	0
12	47	27	22	13	11	51	19	18	12	12
13	13	6	2	2	1	18	6	3	5	2
14	10	2	1	0	0	21	8	3	4	1
15	26	9	7	5	4	50	19	10	7	3
16	13	5	3	4	2	18	5	4	2	2
17	16	8	8	4	4	24	5	3	3	2
18	35	11	7	6	4	61	17	15	11	9
19	16	6	6	5	5	28	8	3	3	1
20	9	2	2	1	1	29	12	9	9	7
21	15	5	3	2	2	44	17	11	8	8
22	14	7	4	5	2	24	7	3	4	1
23	6	3	1	2	0	15	3	0	2	0
24	18	8	7	3	3	41	17	11	7	7
25	11	2	1	0	0	18	5	3	4	3
Others	839	314	217	153	108	1496	586	410	323	239
Total	1422	549	380	270	194	2995	1209	854	696	536

^aThe data sets of human exon isoforms in the AltSplice database (<http://www.ebi.ac.uk/asd/altsplice/>, prerelease Version 2) were investigated. *A* From the exon file, pairs with an identical 5' end and respective 3' ends were selected and then it was confirmed whether the following intron shared the identical 3' end (= downstream acceptor). *B* Among *A*, pairs were selected when the target exon as well as the following exon was covered with at least two transcripts for both forms. *C* Among *B*, pairs were selected when the ratio of the numbers of covering transcripts was in a range from 1:9 to 9:1. *D* Among *A*, pairs were selected when the

target exon as well as the following exon was covered with at least three transcripts for both forms. *E* Among *D*, pairs were selected when the ratio of the numbers of covering transcripts was in a range from 1:9 to 9:1. *F* From the exon file, pairs with an identical 3' end and respective 5' ends were selected and then it was confirmed whether the preceding intron shared the identical 5' end (= upstream donor). *G–J* Pairs were further selected as in *B–E*. The numbers in the *A* and *F* columns are slightly different from those described in the Event file in the AltSplice database

2001; Okamura-Oho et al. 2003). In contrast, some other groups, including Ross's group, who reported the sequence excluding the CAG sequence, more or less emphasized cytoplasmic localization (Ross 1997; Igarashi et al. 1998; Ellerby et al. 1999; Nucifora et al. 2003). We carefully examined subcellular localization in transfection experiments with four expression constructs with or without the glutamine residue coupled with extended polyglutamine in a disease range or short polyglutamine in a normal range after tagging with green fluorescent protein (GFP). Nuclear predominant localization was observed with the glutamine-included form. In contrast, a considerable fraction of the glutamine-excluded form localized in the cytoplasm although the majority was still in the nuclei (Fig. 3). The speckled structure in the cytoplasm may reflect a high tendency of agglutination of DRPLA protein although it was much influenced by the expression level, as observed in the nuclei. For another protein feature, phosphorylation by JNK (Okamura-Oho et al. 2003), two forms did not show a significant difference (data not shown).

Other cases of protein isoforms with or without a single amino acid residue

During the course of the above experiments, we experienced other cases of inclusion or exclusion of 3 nt in a cloning process of *GHRHR* encoding the growth hormone releasing hormone receptor (Miki et al. 1996) and also in the determination of exon–intron boundaries for *BAIAP2* encoding IRSp53 (Okamura-Oho et al. 2003; Miyahara et al. 2003). Literature surveys detected a few previous reports that clearly depicted isoforms with a subtle difference due to alternative splicing utilizing two acceptor sites separated by 3 nt (Manrow and Berger 1993; Condorelli et al. 1994; Vogan et al. 1996; Oberkofler et al. 1997; Lin et al. 2000; from *PTMA* to *Dnmt1* in Table 1). The *PAX3* and *PAX7* cases were originally reported with mice, and we confirmed that the human genome conserved the structure (aagCAG) at the intron–exon boundary and generated two mRNA isoforms, thus showing them with the human sequence in Table 1. In contrast, the human genome does not have a

Fig. 4 Representative profiles of RT-PCR showing a variable ratio of two transcripts with 3 nucleotide (nt) difference by subtle alternative splicing among human tissues. Similar analyses of rodent's orthologues are also illustrated

corresponding sequence at the indicated intron–exon boundary of mice *Dnmt1*, thus only *Dnmt1* in Table 1 is shown with the mouse sequence. We then surveyed the ExPASy database for inconsistent human sequences with or without a single amino acid residue, where such cases are marked as “conflict.” After consultation of genome databases to see whether it occurred at an exon boundary, we experimentally tested by RT-PCR and found that the cases of *CAST*, *MAN2B1*, and *PSEN2* were explained by this type of alternative splicing. Thus, subtle alternative splicing using two acceptor sites separated by 3 nt became 12 cases at this stage, and relevant sequences and features are illustrated in Table 1 (from *DRPLA* to *PSEN2*).

Candidates in the AltSplice database

Alternative splicing is one of the central issues in the current field of genetics, as the human-genome-sequencing project has revealed a limited number of genes in the genome. Several databases on alternative splicing have been constructed in which many transcripts, including expression sequence tags (ESTs), were aligned with each other and with genomic sequences to find a difference. Current databases on alternative splicing do not include most of the cases we point out above although such databases are valuable resources. During database searches, we found that a source data set was released to the public at one of the sites, the AltSplice database (<http://www.ebi.ac.uk/asd/altsplice/>, prerelease Version 2; Thanaraj et al. 2004), which was constructed with 18,632 known human genes and 614,877 transcripts and included 12,470 and 5,214 events for cassette exon and exon isoforms, respectively. When the retrieved data set was subjected to our analysis, we were amazed to find that alternative splice acceptors as well as donors were frequently situated close to each other. This phenomenon was not described in the database nor in the accompanied report (Thanaraj et al. 2004). The exon isoform events peaked at the position separated by 4 nt from the 3' end of the exon (splice donor) and at the 3 nt position from the 5' end of the exon (acceptor), sharply decreased along with the distance from the exon ends, and distributed widely with a slight high at the positions with a multiple of 3 (Table 2). As many as 600 alternative splicing events utilizing two acceptor sites separated by 3 nt were included in the data set. Some events shared the acceptor pairs with diverse donors, and some others were not confirmed with the standard genomic sequences at NCBI. After removing such events, the number of unique acceptor pairs became 536. We further selected 269 events, which were covered by at least two transcripts for both forms, and subjected them to RT-PCR to examine whether two forms of transcripts with a 3 nt difference were easily detectable. For the source of mRNA, commercially available mRNA preparations from ten human adult tissues as well as mRNA isolated from several cell lines were used. We evaluated the expression ratio with a score of 11 grades from 0:10 to 10:0 (longer form/shorter form) after the figures were rounded out to the nearest whole number to bring the total to ten. When the minor form was detectable in images at 2%–4% levels, the score was indicated with f:10 or 10:f.

According to the database, 220 out of the 269 events were represented by the numbers of covering transcripts within a range of 1:9 to 9:1 ratio. In more than 90% of these cases (202/220), two forms of 3 nt difference were easily detectable by RT-PCR, at least in a certain tissue showing a reasonable level of expression. In most cases, the ratio of two transcripts was almost constant among RNA sources. In contrast, only one transcript was apparently observed in most of the remaining 49 events that were represented by the numbers of covering tran-

Table 3 Summary of subtle alternative splicing utilizing two acceptors separated by 3 nucleotides (nt)^a

Structure at boundary	Candidates in AltSp DB	Examined by RT-PCR	Confirmed two forms by RT-PCR	Expression ratio			Location in the coding region				Amino acid change		Others (aa-row)	
				$L > S$	$L = S$	$L < S$	Ext. varied	In 5' noncoding	In frame	Frame -1	Frame +1	Indel of single		Two/one change
6 aagAAG	49	1 33v	1 25	1 8	2	12	3	3	9	5	1 8	1 22	1 10	
2 aagGAG	11	6	6	1		5			1			1	1	
67 cagAAG	206	22 123	15 110	13 60		1 34	1 11	1 20	3 34	2 19	4 37	4 86	1 24	
17 cagGAG	28	5	3	3	5							3	1	
3 gagAAG	5	13	12	8		3	1	4	1	3	4	8	2	
4 gagCAG	3	1	1	1					1		1	1	1	
4 gagGAG	1	2	2			2	1		1		1			
31 tagAAG	87	14 47	9 43	7 24		2 9		3 5		1 7	6 14	4 37	3 14	
4 tagGAG	15	1	1	1	3				1			1		
536 Total	277	4	4	2			2	1		3		3	3	
68 aagNAG	41	277	235	130	10	69	26	38	68	42	87	186	65	
318 cagNAG	163	63	33	10	2	18	3	3	10	7	13	28	13	
13 gagNAG	7	163	140	84	5	38	13	25	38	24	53	110	31	
137 tagNAG	5	66	57	34	3	11	9	9	18	10	20	45	20	
107 nagAAG	38	205	26	22	0	3	1	4	4	2	16	19	9	
347 nagCAG	27	10	7	4	0	55	22	28	60	31	60	145	48	
27 nagGAG	27	10	7	4	0	3	0	1	3	0	3	6	2	
55 nagTAG	55	24	23	12	0	8	3	5	1	9	8	16	6	

^aThe numbers of respective cases are shown. The examined cases include 11 original and previous cases but not mouse *Dnmt1*. In the expression ratio, $L > S$, $L = S$, and $L < S$ indicate that the longer form is more abundant than, almost equal to, and less than the shorter form, respectively. *Indel* insertion or deletion, *aa* amino acid. Supplementary information, including the respective genes, sequences, and the expression ratio, will be available on our Web site at <http://www.nch.go.jp/genetics/subtlealtsp/>

scripts outside the range of 1:9 to 9:1 ratio. The RT-PCR may not be suitable to validate such unbalanced cases. However, images sometimes showed the other form with the 3 nt difference at a very low level in certain tissue and also revealed a variable case (see below). Altogether, we confirmed by RT-PCR two isoforms of a 3 nt difference in 236 cases, including our original and previously reported cases, after removing redundancies.

The expression ratio of two forms of transcripts sometimes varied among RNA sources, and we assigned as an “extremely varied” case when the ratio fluctuated by four or more grades in the score (26 cases). In the ultimate cases, the quantitative major form was reversed, and even the ratio changed from the 0:10 to 10:0 grade. Several “extremely varied” cases are illustrated in Fig. 4, together with similar analyses with mouse and rat orthologues, if the structure of the intron–exon boundary was conserved. Human *PAX7* was expressed in a limited range of tissues, and the ratio was extremely varied, from 9:1 to almost 0:10, among tissues. Mouse and rat orthologues were expressed in a slightly different expression range of tissues, and the ratio was also different from that observed with the human counterpart. Five other human genes in Fig. 4 were ubiquitously expressed, but the ratio of two forms considerably varied among tissues. In contrast, rodent orthologues were expressed at an almost constant ratio, except for the mouse *Ccl20*, which was one of the two “extremely varied” cases that we detected in rodents to date. Genomic sequences of rat *Noxol* are not available, and mouse *Sgnel* have a disrupted sequence (cagCTGATG). Orthologues of human *LAP1B* are probably mouse *MGC6357* and rat *Lap1b* and do not maintain the nagNAG sequence at the boundary. The ratio in human *TGFA* varied from 6:4 in the brain to 9:1 in the spleen, with the longer form being major, while both mouse (Fig. 4) and rat (data not shown) orthologues generated two forms in the f:10 ratio. Through these analyses, no correlation between predominance of one of the forms and the expression levels and also a particular tissue were detected.

The results are summarized in Table 3 where the events are classified in terms of the nucleotide species at the boundary (nagNAG) and include the expression ratio and resulting amino acid changes if inclusion or exclusion of 3 nt takes place in the coding region. Interesting features associated with this phenomenon are discussed in the following section.

Discussion

In this report, we have demonstrated that the subtle sequence difference in *DRPLA* cDNA was indeed due to alternative splicing utilizing two acceptor sites separated by 3 nt, and the resultant inclusion or exclusion of the single glutamine residue affected the subcellular localization of the product. With this study, together with a few previous studies on respective genes, we would call broad attention to the subtle differences in mRNA and

protein structures by alternative splicing. Meantime, we found a variable data set on alternative splicing, experimentally examined by RT-PCR, and finally concluded that protein isoforms with or without a single amino acid residue are quite frequently generated. When we were preparing our manuscript, Hiller et al. (2004) reported widespread occurrence of alternative splicing at NAGNAG acceptors. Their conclusion was based on bioinformatics analyses. Our major conclusion is similar to theirs; however, we communicate our view as we have taken a different course.

Most databases on alternative splicing have been constructed by comparison of transcripts, mostly ESTs, with each other and also with genomic sequences. EST sequences were usually obtained by single-pass sequencing with a reverse-transcribed source of RNA isolated from a variety of tissues or cultured cells, and the accuracy was not usually validated. Thus, the EST sequences should be regarded as samples roughly representing the transcriptome although they are a powerful resource for various purposes. Scientists in the field of bioinformatics sometimes refer to the events revealed with EST as “experimentally confirmed” matters; however, they must be regarded as candidates and should be validated by any other means. In this point of view, we were very much interested in alternative splicing events making a 2 nt difference. The source of the AltSplice database contained as many as 36 events for two acceptor sites separated by 2 nt. If this type of alternative splicing really occurred, 2 nt were the minimal difference in mRNA isoforms although downstream amino acid sequences would differ considerably. We examined by RT-PCR all the nine cases that were covered by at least two transcripts for both forms in the data set (including *IKBKAP* covered by 12 and 4 transcripts and *NDUFB7* by 85 and 4 transcripts), as well as several other cases. All the cases examined apparently generated one form corresponding to the ag-included form, except for *GLYCTK* that generated the ag-excluded form, and no bands were detectable at the position where products of 2 nt difference were migrated. The results do not completely exclude the possibility of such a type of alternative splicing, as the events may occur at a more biased ratio or in a particular tissue that we did not examine. However, the RT-PCR results for the events with 2 nt difference are remarkably different from those with 3 nt difference where two forms were easily detectable at about 90%. As repeatedly pointed out, ESTs as well as alternative splicing databases based on ESTs contain the results of aberrant splicing (noise) (Sorek et al. 2004). Another irregularity with EST should be pointed out. The expression ratio of two forms that we experimentally determined by RT-PCR, excluding the extremely varied cases, does not necessarily coincide with the ratio of the numbers of covering transcripts in the database. This is partly accounted for by a relatively small number of transcripts covering the particular event for most cases. The number of EST is extensively biased due to the level of expression, the

source of transcripts, and the position within the gene, as most ESTs are generated from the 5' or 3' terminal of transcripts. Thus, experimental validation with other techniques is essential.

A total of 235 cases for which we experimentally confirmed two forms of 3 nt difference (after removing mouse *Dmmt1*) were subjected to further analyses and revealed the following features (Table 3). Preceding intron of all the cases is the GT-AG type for both proximal and distal acceptors, which generate longer and shorter forms, respectively. Thus, a nagNAG sequence appears at the intron–exon boundary of target exon, and the NAG sequence flanked by two acceptor sites is the target of inclusion or exclusion in the longer or shorter isoform (optional 3 nt). The frequencies of each nucleotide at the “n” and “N” positions in nagNAG show an interesting feature. A preferable nucleotide at the –3 position of exon is pyrimidine, according to the consensus splice site sequence, and a large data set in the AltSplice database (140,508 GT-AG type intron excluding redundancy) shows the frequencies of “a,” “c,” “g,” and “t” at the position are 5.8%, 65.1%, 0.5%, and 28.6%, respectively. Compared with these frequencies, the nucleotide at the –3 position to generate the longer forms is somewhat unusual, with a slightly higher incidence of “a” and “g” (14.0% and 2.1%). Furthermore, the nucleotide at the –3 position to generate the shorter forms is extremely irregular, and appearance of “t” is significantly suppressed (9.8% versus 28.6%). This may be accounted for by the fact that the optional 3 nt is also served as the 5' end of the exon. The large data set in the AltSplice database shows the frequencies of “A,” “C,” “G,” and “T” at the +1 position of the exon are 25.7%, 14.4%, 48.6%, and 11.3%, respectively. Thus, the lower occurrence of TAG as the optional 3 nt may be explained by unfavorability for the start of the exon. However, “C” is frequently situated at the position although it is also unfavorable for the start of the exon. For another possibility, mRNA having in-frame TAG may be degraded by a similar mechanism, as observed in genetic diseases where unexpected translational stop codon sometimes causes degradation of not only protein products but also mRNA. While our original cases imply a high occurrence of the optional 3 nt in-frame (8/12 cases), the data set from AltSplice shows the in-frame situation not so high ($68/197 = 34.5\%$). Thus, the biased occurrence of nucleotide species at the “n” and “N” positions in nagNAG should be interpreted by other points.

The next issue is which adjacent acceptor is more frequently used? In general, the proximal acceptor (E acceptor in the report by Hiller et al. 2004) seems to be preferable, as the longer form is dominant in about 56% of cases. After classified in terms of nagNAG species, the distal acceptor (I acceptor) appears to be stronger when purine occurs at “n” and pyrimidine occurs at “N.” For this consideration, it is more appropriate to include other cases where only one type of transcript is apparently produced with the nagNAG structure. The selec-

tion of the two adjacent sites in this type of alternative splicing is not explained by the simple scanning model and may be accomplished by interaction of a splice factor(s) through the sequence context. U2AF35 and Slu7 were demonstrated to be involved in the recognition of “ag” and also SPF45 in the selection of two adjacent “ag” separated by 16 nt in *Drosophila Sex-lethal* (Wu et al. 1999, Chua and Reed 2001; Lallena et al. 2002). However, no obvious rules come up at a glance, for example, in the length of pyrimidine-rich sequence and in the distance to a branch point, and no current programs for splice-site prediction detect the two sites with an appropriate value reflecting the ratio we experimentally determined. Moreover, we have revealed variable cases, which may become good markers to define the alternative splicing status in disease-associated changes and also to examine a role of isolated splice factors. These studies may be accelerated when the segments covering two adjacent acceptors are assembled on a microchip for detection of alternative splicing. Thus, the issue of site selection in this type of alternative splicing provides an enormous opportunity to improve the algorithm to find splice sites and also to elucidate the molecular mechanism of regulation in splicing.

The third feature associated with this type of alternative splicing is the location and amino acid changes. In about 15% of cases, the optional 3 nt localizes in the 5' noncoding region and frequently occurs near the translational initiation site. In typical cases, the optional 3 nt is inserted just upstream of the initiation ATG, like tagCAGCC ATG in *STK38* and cagAAGCC ATG in *ORC1L*. These changes in mRNA may affect the efficiency of translation although we have yet only examined experimentally. For about 85% of cases, the optional 3 nt localizes in a coding region. The distribution of coding frames is not even, as previously pointed out (Hiller et al. 2004), but the biased distribution is milder in our study than the previous analysis. Despite the coding frame, inclusion and exclusion of optional 3 nt results in inclusion or exclusion of a single amino acid residue and exchange of one amino acid residue with two different amino acid residues (two/one type) is rare. This fact was already mentioned by Hiller et al. (2004) who proved this by counting the cases in a large data set and comparing with the events in an artificial null model. However, this is simply explained by the following facts: It is obvious when the optional 3 nt is situated in-frame. When the optional 3 nt is situated in coding frame –1 (intron phase 2 in the report by Hiller et al. 2004), the resultant frames are $N_1N_2NAGN_3$ and $N_1N_2N_3$ in the longer and shorter forms, respectively. The chance of an identical amino acid encoded by N_1N_2N at the proximal frame of the longer form and by $N_1N_2N_3$ in the shorter form is considerably high due to codon degeneracy, whatever nucleotide occupies N and N_3 . Moreover, “G,” especially “AG,” frequently occupies the exon end as a mononucleotide and dinucleotide, and the large data set of the AltSplice database shows the frequencies at 81.3% and 55.6%, respectively. When “AG” occupies

the end of the preceding exon, resultant frames become $AGNAGN_3$ and AGN_3 , resulting in simple inclusion of an amino acid encoded by AGN because the codon in the shorter form is identical to the distal frame of the longer form (AGN_3). Thus, coding frame -1 has double constraints on the amino acid changes. Indeed, only four out of 15 cases where the non-“G” nucleotide occupies the end of the preceding exon with coding frame -1 generate the two/one type of amino acid change. When the optional 3 nt is situated in coding frame $+1$ (intron phase 1), the resultant frames become $N_1NAGN_2N_3$ and $N_1N_2N_3$. As the “ N_1 ” is derived from the end of the preceding exon where “G” frequently occupies, the frames are highly expected to be $GNAGN_2N_3$ and GN_2N_3 , thus resulting in simple inclusion of an amino acid residue encoded by GNA in the proximal frame. The frequency of appearance of “G” at the preceding exon end is also high in this type of alternative splicing ($199/235 = 84.7\%$). Eleven out of the 87 cases with coding frame $+1$ have a different nucleotide from “G” at the preceding exon end. Four cases incidentally generate inclusion of an identical amino acid residue due to codon degeneracy. High occurrence of “c” at N, as described above, contributes to the incidence because “c” as the second nucleotide in the codon table shows high degeneracy. The remaining seven cases generate the two/one type of change. Thus, the fact of simple inclusion or exclusion of a single amino acid residue is primarily accounted for by the high incidence of “G” at the exon end and codon degeneracy. As a total, single, amino acid inclusion or exclusion occurs in 186 cases while only 11 cases cause the two/one type amino acid change. The included amino acid residues are Gln, Ala, Ser, Glu, Val, Arg, Lys, and Gly in the order of occurrence. It is also noted that the included single amino acid is frequently the same as one of the adjacent residues and results in generation of a row of identical amino acid residue at the junction. The double constraint in coding frame -1 , as discussed above, somewhat contribute to this phenomenon but even in-frame cases generate the row of identical amino acid.

We have demonstrated with several genes that the expression ratio of the two forms drastically varied among RNA sources, which clearly shows that a splice control system operates even in the narrow range of a few nucleotides but does not show a simple reflection of infidelity in splice machinery. The variability of the ratio seems to be more extensive in a human gene than in rodent genes and may contribute to functional complexity in the highly evolved organism. However, this fact must be ascertained in a broad analysis with other species, as we examined the rodent genes once after human counterparts showed the variability. It has been documented that alternative splicing associated with functional significance is sometimes regulated differently in temporal and spatial manners. With this connection, more important is the case in which the expression ratio is changed in a particular tissue rather than the extensive variability among RNA sources. We observed several

such variations; however, we did not fully ascertain the tissue specificity because of limitation of RNA sources and information. The pattern of alternative splicing may be affected by polymorphisms, gender, age, and physiological conditions.

In the case of *DRPLA*, the protein natures were changed by the inclusion or exclusion of the single amino acid by subtle alternative splicing. It was described that receptor activities of *IGFIR* and the transcriptional activities of *PAX3* and *PAX7* altered by the amino acid changes (Condorelli et al. 1994; Vogan et al. 1996), but no functional changes were detected with *LEP* (Oberkofler et al. 1997). Each function of respective protein isoforms must be experimentally examined.

This phenomenon, widespread occurrence of subtle alternative splicing, immediately raised a concern on annotation of human as well as other species' genome. Which is the “standard” transcript? Currently, NCBI assigns the CAG-excluded form of *DRPLA* as the standard, but it is inappropriate because of the expression levels, as demonstrated here and by the number of sequences deposited in databases. The standard must be reconsidered, and any transcripts representing more than a certain threshold of expression fraction (we tentatively propose more than 5% in RT-PCR) should not be lost in the standard set. Although RT-PCR has the limitation to reveal an event spanning a long distance, it is also important to determine whether the included and excluded forms are associated with a particular splice by other alternative splicing events in the gene, a similar phenomenon as a mutually exclusive exon.

Finally, we wish to bring broad attention to the phenomenon in which alternative splicing sometimes makes a subtle difference, like 3 nt as demonstrated here. This phenomenon contributes to enormous diversification of proteomes and provides an opportunity to understand the mechanism in splice-site selection and regulation.

Acknowledgements We thank A. Asaka for technical assistance and K. Saito for preparing the manuscript. This study was supported in part by Grants for Human Genome, for Paediatric Research from the Ministry of Health, Labour and Welfare, Japan.

References

- Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72:291–336
- Chua K, Reed R (2001) An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. *Mol Cell Biol* 21:1509–1514
- Condorelli G, Bueno R, Smith RJ (1994) Two alternatively spliced forms of the human insulin-like growth factor I receptor have distinct biological activities and internalization kinetics. *J Biol Chem* 269:8510–8516
- Cummings CJ, Zoghbi HY (2000) Trinucleotide repeats: mechanisms and pathophysiology. *Annu Rev Genomics Hum Genet* 1:281–328
- Ellerby LM, Andrusiak RL, Wellington CL, Hackam AS, Propp SS, Wood JD, Sharp AH, Margolis RL, Ross CA, Salvesen GS, Hayden MR, Bredesen DE (1999) Cleavage of atrophin-1 at caspase site aspartic acid 109 modulates cytotoxicity. *J Biol Chem* 274:8730–8736

- Forman MS, Trojanowski JQ, Lee VM (2004) Neurodegenerative diseases: a decade of discoveries paves the way for therapeutic breakthroughs. *Nat Med* 10:1055–1063
- Hiller M, Huse K, Szafranski K, Jahn N, Hampe J, Schreiber S, Backofen R, Platzer M (2004) Widespread occurrence of alternative splicing at NAGNAG acceptors contributes to proteome plasticity. *Nat Genet* 36:1255–1257
- Igarashi S, Koide R, Shimohata T, Yamada M, Hayashi Y, Takano H, Date H, Oyake M, Sato T, Sato A, Egawa S, Ikeuchi T, Tanaka H, Nakano R, Tanaka K, Hozumi I, Inuzuka T, Takahashi H, Tsuji S (1998) Suppression of aggregate formation and apoptosis by transglutaminase inhibitors in cells expressing truncated DRPLA protein with an expanded polyglutamine stretch. *Nat Genet* 18:111–117
- International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* 431:931–945
- Kanazawa I (1998) Dentatorubral-pallidoluysian atrophy or Naito-Oyanagi disease. *Neurogenetics* 2:1–17
- Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcarcel J (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* 109:285–296
- Lin MJ, Lee TL, Hsu DW, Shen CK (2000) One-codon alternative splicing of the CpG MTase Dnmt1 transcript in mouse somatic cells. *FEBS Lett* 469:101–104
- Lopez AJ (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* 32:279–305
- Love SJ, Margolis RL, Young WS, Li SH, Schilling G, Ashworth RG, Ross CA (1995) Cloning and expression of the rat atrophin-I (DRPLA disease gene) homologue. *Neurobiol Dis* 2:129–138
- Manrow RE, Berger SL (1993) GAG triplets as splice acceptors of last resort. An unusual form of alternative splicing in prothymosin alpha pre-mRNA. *J Mol Biol* 234:281–288
- Margolis RL, Li SH, Young WS, Wagster MV, Stine OC, Kidwai AS, Ashworth RG, Ross CA (1996) DRPLA gene (atrophin-I) sequence and mRNA expression in human brain. *Brain Res Mol Brain Res* 36:219–226
- Michalik A, Van Broeckhoven C (2003) Pathogenesis of polyglutamine disorders: aggregation revisited. *Hum Mol Genet* 12:R173–R186
- Miki N, Ono M, Murata Y, Ohsaki E, Tamitsu K, Yamada M, Demura H (1996) Regulation of pituitary growth hormone-releasing factor (GRF) receptor gene expression by GRF. *Biochem Biophys Res Commun* 224:586–590
- Miyahara A, Okamura-Oho Y, Miyashita T, Hoshika A, Yamada M (2003) Genomic structure and alternative splicing of the insulin receptor tyrosine kinase substrate of 53-kDa protein. *J Hum Genet* 48:410–414
- Miyashita T, Okamura-Oho Y, Mito Y, Nagafuchi S, Yamada M (1997) Dentatorubral pallidoluysian atrophy (DRPLA) protein is cleaved by caspase-3 during apoptosis. *J Biol Chem* 272:29238–29242
- Miyashita T, Nagao K, Ohmi K, Yanagisawa H, Okamura-Oho Y, Yamada M (1998) Intracellular aggregate formation of dentatorubral-pallidoluysian atrophy (DRPLA) protein with the extended polyglutamine. *Biochem Biophys Res Commun* 249:96–102
- Miyashita T, Matsui J, Ohtsuka Y, U M, Fujishima S, Okamura-Oho Y, Inoue T, Yamada M (1999) Expression of extended polyglutamine sequentially activates initiator and effector caspases. *Biochem Biophys Res Commun* 257:724–730
- Modrek B, Lee C (2002) A genomic view of alternative splicing. *Nat Genet* 30:13–19
- Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M, Takeda T, Tadokoro K, Kondo I, Murayama N, Tanaka Y, Kikushima H, Umino K, Kurosawa H, Furukawa T, Nihei K, Inoue T, Sano A, Komure O, Takahashi M, Yoshizawa T, Kanazawa I, Yamada M (1994a) Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat Genet* 6:14–18
- Nagafuchi S, Yanagisawa H, Ohsaki E, Shirayama T, Tadokoro K, Inoue T, Yamada M (1994b) Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA). *Nat Genet* 8:177–182
- Nucifora FC Jr, Ellerby LM, Wellington CL, Wood JD, Herring WJ, Sawa A, Hayden MR, Dawson VL, Dawson TM, Ross CA (2003) Nuclear localization of a non-caspase truncation product of atrophin-1, with an expanded polyglutamine repeat, increases cellular toxicity. *J Biol Chem* 278:13047–13055
- Oberkofler H, Beer A, Breban D, Hell E, Krempler F, Patsch W (1997) Human obese gene expression: alternative splicing of mRNA and relation to adipose tissue localization. *Obes Surg* 7:390–396
- Okamura-Oho Y, Miyashita T, Ohmi K, Yamada M (1999) Dentatorubral-pallidoluysian atrophy protein interacts through a proline-rich region near polyglutamine with the SH3 domain of an insulin receptor tyrosine kinase substrate. *Hum Mol Genet* 8:947–957
- Okamura-Oho Y, Miyashita T, Nagao K, Shima S, Ogata Y, Katada T, Nishina H, Yamada M (2003) Dentatorubral-pallidoluysian atrophy protein is phosphorylated by c-Jun NH₂-terminal kinase. *Hum Mol Genet* 12:1535–1542
- Onodera O, Oyake M, Takano H, Ikeuchi T, Igarashi S, Tsuji S (1995) Molecular cloning of a full-length cDNA for dentatorubral-pallidoluysian atrophy and regional expressions of the expanded alleles in the CNS. *Am J Hum Genet* 57:1050–1060
- Ozaki M, Itoh K, Miyakawa Y, Kishida H, Hashikawa T (2004) Protein processing and releases of neuregulin-1 are regulated in an activity-dependent manner. *Neurochem* 91:176–188
- Ross CA (1997) Intranuclear neuronal inclusions: a common pathogenic mechanism for glutamine-repeat neurodegenerative diseases? *Neuron* 19:1147–1150
- Sorek R, Shamir R, Ast G (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet* 20:68–71
- Tadokoro K, Oki N, Sakai A, Fujii H, Ohshima A, Nagafuchi S, Inoue T, Yamada M (1993) PCR detection of 9 polymorphisms in the WT1 gene. *Hum Mol Genet* 2:2205–2206
- Thanaraj TA, Stamm S, Clark F, Riethoven JJ, Le Texier V, Muilu J (2004) ASD: the alternative splicing database. *Nucleic Acids Res* 32:D64–D69
- U M, Miyashita T, Ohtsuka Y, Okamura-Oho Y, Shikama Y, Yamada M (2001) Extended polyglutamine selectively interacts with caspase-8 and -10 in nuclear aggregates. *Cell Death Differ* 8:377–386
- Vogan KJ, Underhill DA, Gros P (1996) An alternative splicing event in the Pax-3 paired domain identifies the linker region as a key determinant of paired domain DNA-binding activity. *Mol Cell Biol* 16:6677–6686
- Wu S, Romfo CM, Nilsen TW, Green MR (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* 402:832–835
- Yanagisawa H, Bundo M, Miyashita T, Okamura-Oho Y, Tadokoro K, Tokunaga K, Yamada M (2000) Protein binding of a DRPLA family through arginine-glutamic acid dipeptide repeats is enhanced by extended polyglutamine. *Hum Mol Genet* 9:1433–1442