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

[a]The 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 (aag**CAG**) 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

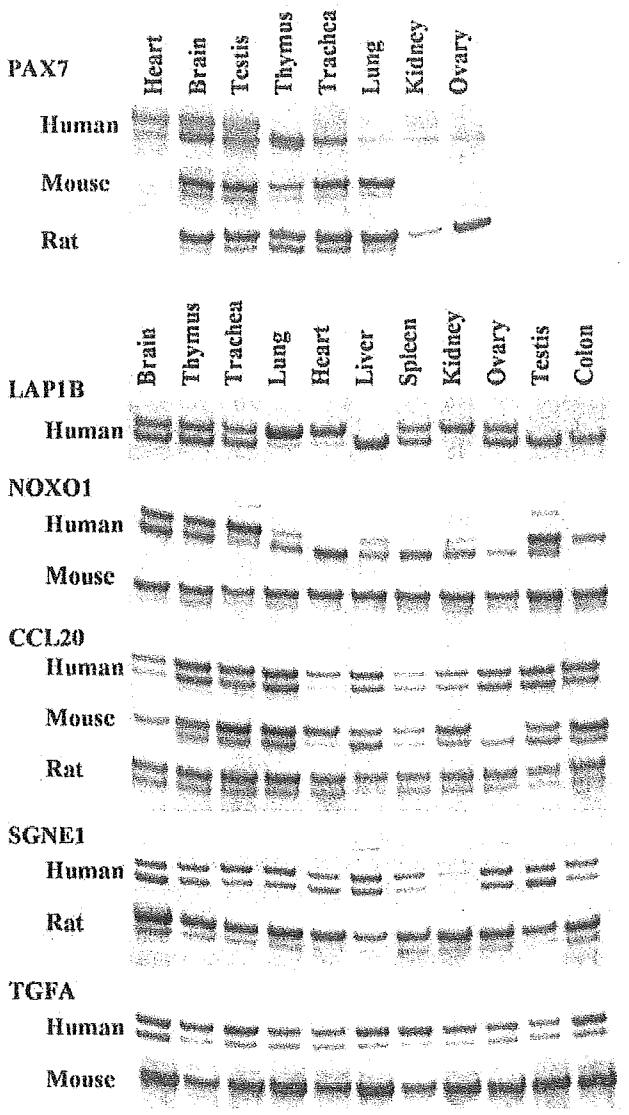

Fig. 4 Representative profiles of RT-PCR showing a variable ratio of two transcripts with 3 nucleotide (nt) difference by subtle alternative splicing among human tissues. Similar analyses of rodent's orthologues are also illustrated

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	
aagAAG	6	1	1	1							1	1		1
aagCAG	49	33v	25	8	2	12	3	3	9	5	8	22		10
aagGAG	2	1	1			1			1			1		1
aagTAG	11	6	6	1		5				2	4	4	2	1
cagAAG	67	22	15	13		1	1	1	3	2	9	13	1	4
cagCAG	206	123	110	60	5	34	11	20	34	19	37	86	4	24
cagGAG	17	5	3	3							3	3		1
cagTAG	28	13	12	8		3	1	4	1	3	4	8		2
gagAAG	3	1	1	1					1			1		1
gagCAG	5	2	1				1				1		1	
gagGAG	4	3	2			2		1	1			1		
gagTAG	1	1	1	1						1		1		
tagAAG	31	14	9	7		2		3			6	4	2	3
tagCAG	87	47	43	24	3	9	7	5	17	7	14	37	1	14
tagGAG	4	1	1	1					1			1		
tagTAG	15	4	4	2			2	1		3		3		3
Total	536	277	235	130	10	69	26	38	68	42	87	186	11	65
aagNAG	68	41	33	10	2	18	3	3	10	7	13	28	2	13
cagNAG	318	163	140	84	5	38	13	25	38	24	53	110	5	31
gagNAG	13	7	5	2	0	2	1	1	2	1	1	3	1	1
tagNAG	137	66	57	34	3	11	9	9	18	10	20	45	3	20
nagAAG	107	38	26	22	0	3	1	4	4	2	16	19	3	9
nagCAG	347	205	179	92	10	55	22	28	60	31	60	145	6	48
nagGAG	27	10	7	4	0	3	0	1	3	0	3	6	0	2
nagTAG	55	24	23	12	0	8	3	5	1	9	8	16	2	6

[a]The 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 *Noxo1* are not available, and mouse *Sgne1* 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 *Dmnt1*) 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 nag**NAG** 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 nag**NAG** 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 unfavorableness 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-flame situation not so high (68/197 = 34.5%). Thus, the biased occurrence of nucleotide species at the "n" and "N" positions in nag**NAG** 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 nag**NAG** 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 nag**NAG** structure. The selection 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 slicing 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 tag**CAGCC** ATG in *STK38* and cag**AAGCC** 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_2**NAG**N_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 *IGF1R* 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-1) 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 Bhiophys 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 NH_2-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

Human Molecular Genetics, 2005, Vol. 14, No. 22 *3379–3388*
doi:10.1093/hmg/ddi369
Advance Access published on October 3, 2005

Detecting tissue-specific alternative splicing and disease-associated aberrant splicing of the *PTCH* gene with exon junction microarrays

Kazuaki Nagao[1], Naoyuki Togawa[2], Katsunori Fujii[3], Hideki Uchikawa[1,3], Yoichi Kohno[3], Masao Yamada[1] and Toshiyuki Miyashita[1,*]

[1]Department of Genetics, National Research Institute for Child Health and Development, Tokyo 157-8535, Japan, [2]Yokohama Research Laboratories, Mitsubishi Rayon Co., Ltd, Yokohama 230-0053, Japan and [3]Department of Pediatrics, Graduate School of Medicine, Chiba University, Chiba 260-8670, Japan

Received June 8, 2005; Revised and Accepted September 23, 2005 GenBank accession numbers[†]

Mutations in the human ortholog of *Drosophila patched* (*PTCH*) have been identified in patients with autosomal dominant nevoid basal cell carcinoma syndrome (NBCCS), characterized by minor developmental anomalies and an increased incidence of cancers such as medulloblastoma and basal cell carcinoma. We identified many isoforms of *PTCH* mRNA involving exons 1–5, exon 10 and a novel exon, 12b, generated by alternative splicing (AS), most of which have not been deposited in GenBank nor discussed earlier. To monitor splicing events of the *PTCH* gene, we designed oligonucleotide arrays on which exon probes and exon–exon junction probes as well as a couple of intron probes for the *PTCH* gene were placed in duplicate. Probe intensities were normalized on the basis of the total expression of *PTCH* and probe sensitivity. Tissue-specific regulation of AS identified with the microarrays closely correlated with the results obtained by RT–PCR. Of note, the novel exon, exon 12b, was specifically expressed in the brain and heart, especially in the cerebellum. Additionally, using these microarrays, we were able to detect disease-associated aberrant splicings of the *PTCH* gene in two patients with NBCCS. In both cases, cryptic splice donor sites located either in an exon or in an intron were activated because of the partial disruption of the consensus sequence for the authentic splice donor sites due to point mutations. Taken together, oligonucleotide microarrays containing exon junction probes are demonstrated to be a powerful tool to investigate tissue-specific regulation of AS and aberrant splicing taking place in genetic disorders.

INTRODUCTION

Alternative splicing (AS) is one of the major mechanisms by which humans produce the complexity of the proteome. It has been estimated that greater than 55% of all genes and at least 74% of multi-exon genes are alternatively spliced in humans (1,2). Protein isoforms produced by AS can have antagonistic functions, such as anti-apoptotic Bcl-x_L versus pro-apoptotic Bcl-x_S (3), or completely different amino-acid compositions due to different reading frames, such as p16(INK4a) versus p19(ARF) (4). In addition, AS is also implicated in pathophysiological processes and it has been estimated that at least 15% of point mutations that cause human genetic diseases affect splicing (5).

Nevoid basal cell carcinoma syndrome (NBCCS), also called Gorlin's syndrome, is an autosomal dominant neurocutaneous disorder characterized by large body size, developmental and skeletal abnormalities, radiation sensitivity, basal

*To whom correspondence should be addressed at: Department of Genetics, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan. Tel: +81 334160181; Fax: +81 354947035; Email: tmiyashita@nch.go.jp

[†]The nucleotide sequence data of human and mouse isoforms, +12b, have been deposited with the GenBank Library under Accession Nos AB214500 and AB214501, respectively. Human isoforms, −3, 4,5, +4′ and −10, have been deposited with the GenBank Library under Accession Nos AB233423, AB233424 and AB233422, respectively.

cell carcinoma (BCC) and an increased incidence of medulloblastoma (6). NBCCS is caused by inactivating mutations in the *patched* (*PTCH*) gene (7,8). The human *PTCH* gene contains 23 exons spanning ~ 65 kb and is predicted to encode a protein of 1447 amino-acid residues containing 12 transmembrane-spanning domains and two large extracellular loops (7). Heterozygous loss of *PTCH* found in certain sporadic and familial cases of BCC indicates that *PTCH* is also a tumor suppressor gene (9,10). In vertebrates, a second *patched* gene (*PTCH2* in humans) was identified (11,12). So far, no mutations in *PTCH2* have been reported in NBCCS although a limited number of mutations were found in BCC and medulloblastoma (11).

Recently, others and we have identified that *PTCH* undergoes complex AS between multiple first exons and a second exon (13–16). In addition, we also have identified additional mRNA isoforms downstream of exon 2. Therefore, to survey tissue-specific AS and disease-associated aberrant splicing of *PTCH*, we have developed oligonucleotide microarrays designed for profiling AS. Current conventional microarray technologies are limited in their ability to distinguish and analyze mRNA isoforms. Several groups have reported a survey of human AS using exon junction microarrays that can circumvent this problem (2,17–20). In this study, considering the volume and complexity of the data produced by genome-wide studies and the cost–benefit ratio of these arrays, we have developed microarrays focused on the *PTCH* gene in which probes were designed for individual exons and splice junctions including junctions derived from rare AS. Using these arrays, we demonstrate a detailed evaluation of the data on tissue-specific regulation of AS. In addition, we describe the use of DNA microarrays to identify the aberrant splicing taking place in a genetic disorder.

RESULTS

Detection and validation of AS by oligonucleotide microarrays

So far, at least five alternatively used first exons have been reported in the *PTCH* gene (13–16). Recently, using RNA from various human tissues, we have identified additional mRNA isoforms generated by AS involving downstream exons (exons 2, 3, 4, 5 and 10) and an alternative exon named 12b by RT–PCR and sequencing (Fig. 1A and B). The numbering of exons is according to Johnson *et al.* (7) with GenBank accession no. U59464. Among these isoforms, the one skipping exon 10 has been briefly described previously (21) and the one skipping exons 4 and 5 has been deposited in Genbank (accession no. AB209495) but not discussed in a prior publication. The rest of the isoforms are novel. Wicking *et al.* (22) reported our exon 13 as exon 12b. However, as this is a constitutive exon, it is named as exon 13 in most publications and databases and the following exons are numbered by counting from 5′ to 3′. No AS involving exon 13 or exons further downstream could be identified in any tissues examined. If alternative exons can be spliced independently, then the *PTCH* locus potentially encodes 483 isoforms. To monitor splicing events of the *PTCH* gene, we designed oligonucleotide arrays on which 17 exon probes and 23 exon–exon junction probes as well as a couple of intron probes were placed in duplicate (Fig. 1A). Before analyzing the data obtained from various tissues and NBCCS patients, we validated each probe using plasmids encoding various *PTCH* isoforms. For this purpose, each cDNA sequence encoding six *PTCH* isoforms was amplified by PCR followed by *in vitro* transcription and labeling with Cy-5 and then applied onto the microarrays. The intensity data prior to normalization represented as an intensity heatmap is shown in Figure 1B. Each probe's intensity was then adjusted in two ways: first, for the mean probe intensity for each construct and then for each probe's sensitivity. Finally, mean probe intensities that should be positive according to the plasmid and probe sequences were adjusted to 1, and those that should be negative were adjusted to zero. As depicted in Figure 1C, all constructs showed profiles expected from exon composition and could be clearly discriminated from one another. No intensities in-between were observed. However, when placing probes at exon–exon junctions, there is little choice as to the underlying nucleotide composition. So, two junction probes, exon 9–11 and exon 12–13, were non-informative, e.g. saturated, constant intensity and could lead to erroneous predictions, therefore excluded from the figure.

Tissue-specific regulation of AS

To investigate tissue-specific regulation of AS using array data, we normalized probe intensities as described in Materials and Methods and in Figure 4. If no AS takes place at the exons or exon–exon junctions for which probes are designed, then the normalized relative probe intensities are near 1.0 in all tissues. In contrast, probes involved in AS should give intensities with large standard deviations. As expected from mRNA isoforms we have identified so far, a marked variation in normalized probe intensities was observed in exons 2–5, 10 and 12b (Fig. 2A). The tissue-specific accuracy of these array data was confirmed by RT–PCR. For example, when we focused on exon 12b, in tissues in which elevated probe intensities for exons 12–12b, 12b and 12b–13 were observed (i.e. the brain and heart), RT–PCR products containing exon 12b were also evident (Fig. 2B, lanes 1 and 8). We next addressed the question of where in the brain exon 12b was highly expressed. The subsequent investigation demonstrated that exon 12b was particularly expressed in the cerebellum among various brain tissues (Fig. 2B, lane 2). Additionally, two independent methods, microarray analysis and RT–PCR, demonstrated a strong correlation (Fig. 2C), validating the method of array data normalization. Array data as to inclusion or exclusion of exon 10 was similarly compared with the RT–PCR results. As shown in Figure 2D, these two methods again correlated well with each other. Isoforms skipping exon 10 were the most highly expressed in the thymus and the lung based on both these methods.

As shown in Figure 1A, AS involving exons 2–5 is relatively complicated and cannot be explained simply by inclusion or exclusion of a single exon. When we focused on isoforms skipping exons 4 and 5, the data from the two methods still correlated (Fig. 2E). However, the degree of correlation represented by R^2 was weaker than that obtained with

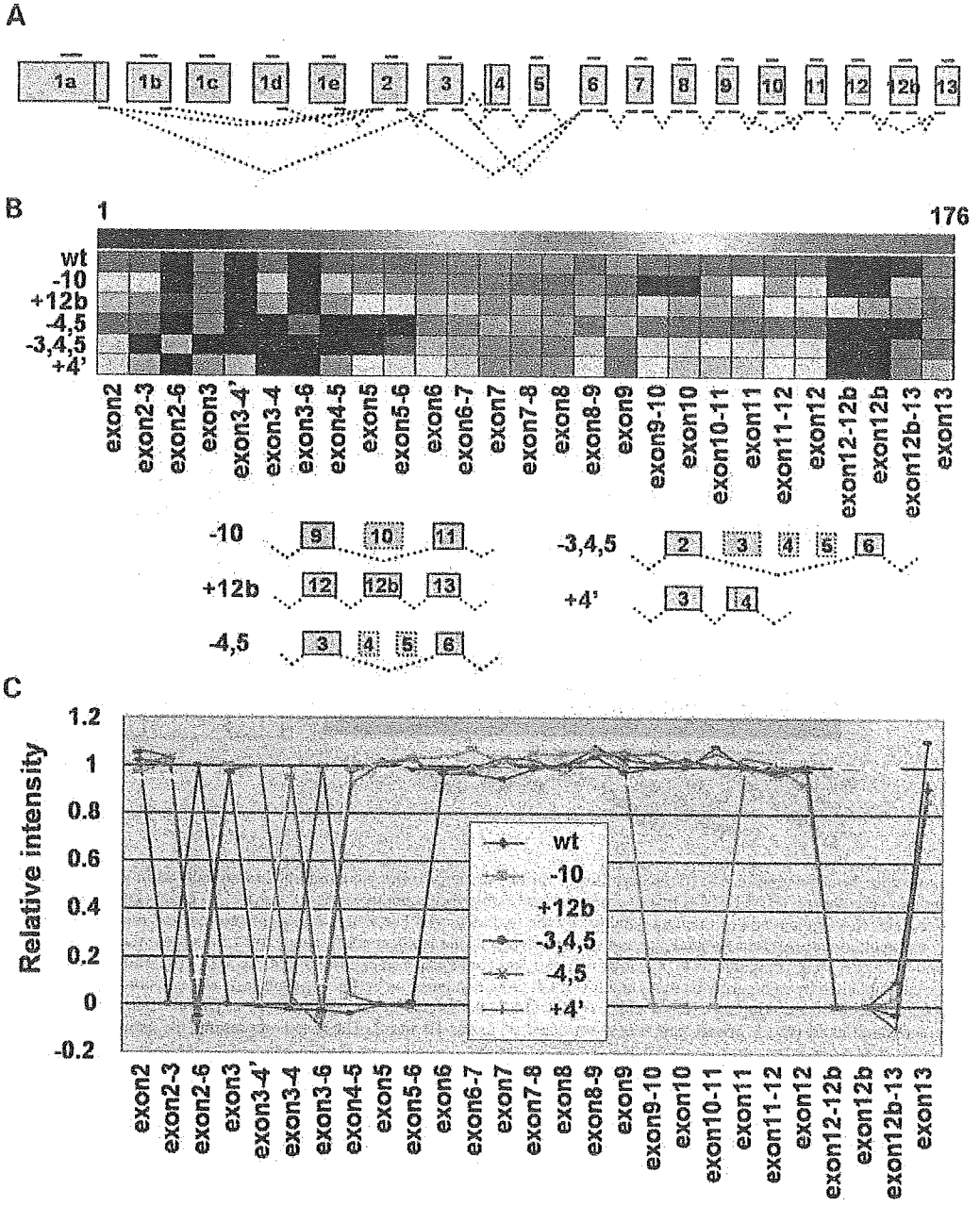

Figure 1. Detection and validation of the known isoforms of *PTCH*. (**A**) Exon structure of the *PTCH* gene. Positions of the exon probes and the exon junction probes are indicated. (**B**) Probe intensities obtained with Cy-5-labeled RNA from constructs encoding various isoforms of *PTCH*. The labeling reaction was performed with T7 RNA polymerase by using the PCR product as a template obtained from the constructs. Each matrix point shows the S/N ratio of one exon or exon junction probe in one construct. Probes are ordered horizontally, 5′–3′, together with some probes for infrequently spliced junctions such as exon 2–6. Exon 3–4′ is an exon junction probe between exon 3 and the alternative splice acceptor site located in intron 3. Probe sequences for most 5′ exons were not included in some of the constructs and were excluded from this figure. Hybridization samples form the vertical axis of each matrix. Exon compositions of the expression plasmids are briefly depicted at the bottom. (**C**) Normalized probe intensities. The mean positive intensity and the mean negative intensity are adjusted to 1 and 0, respectively.

probes for exon 12b or 9 and 10, because AS skipping exons 4 and 5 is a rare event [up to 3% of the authentic isoform (shown by *X*-axis in Fig. 2E)].

Next, we evaluated the microarray data regarding the usage of the alternative first exons. RT–PCR was performed using the same forward primers for each alternative exon 1, as those used for the microarray analysis. As shown in Figure 2F and G, the two methods showed a good correlation regarding the usage of exon 1b. However, isoforms starting from exons 1a, 1d and 1e failed to demonstrate significant

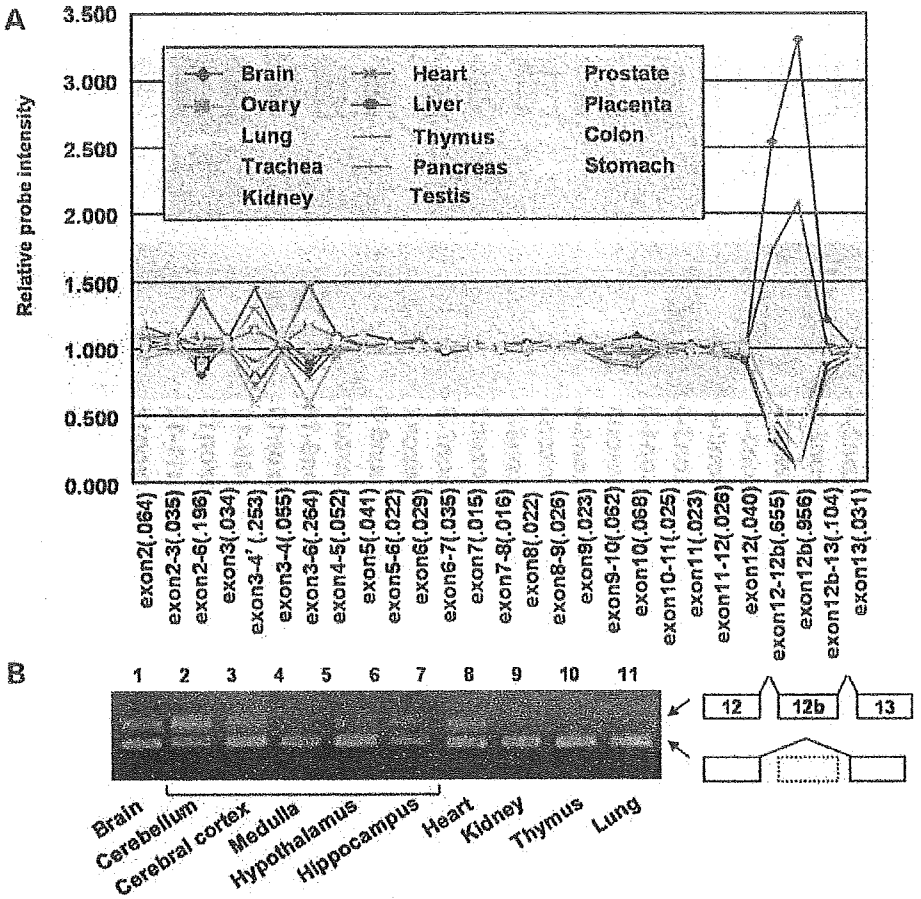

Figure 2. Detection of tissue-specific regulation of AS. **(A)** AS profiles in various tissues. Probes were ordered horizontally as in Figure 1B. Standard deviations for each probe are indicated in parentheses. **(B)** RT–PCR results from a primer pair hybridizing to exons 11 and 14. AS events corresponding to each RT–PCR product are depicted at the right. **(C)** Comparison of the data obtained by the two methods. The *Y*-axis represents normalized relative probe intensity shown in (A). RT–PCR products were applied onto Agilent Bioanalyzer and the percentage of the isoform containing exon 12b is presented by the *X*-axis. **(D)** Two data obtained by microarray and RT–PCR were compared as in (C). A primer pair was constructed on exons 8 and 13. *X*-axis represents the percentage of the isoform skipping exon 10. **(E)** The data obtained by microarray analysis and RT–PCR were compared as in (C). A primer pair was constructed on exons 1b and 6. The *X*-axis represents the percentage of the isoform skipping exons 4 and 5. **(F, G)** The data obtained by microarray analysis [probe 1b in (F) and junction probe 1b-2 in (G)] and RT–PCR were compared as in (C). A primer pair was constructed on exons 1b and 2. The *X*-axis represents the estimated molarity of the isoform starting from exon 1b.

correlations (data not shown). This is probably because the expression of isoforms starting from exons 1a and 1e is much lower than that starting from exon 1b (average expression of exons 1a and 1e is 3.4 and 13.8% of exon 1b, respectively, when pooled S/N ratios are compared). In addition, the PCR efficiency of the exon 1d primer is significantly lower than that of the exon 1b primer (37% of the exon 1b primer) (Supplementary Material, Fig. S1).

Detection of aberrant splicing in NBCCS patients

We have been investigating *PTCH* mutations in NBCCS patients (23) and have detected mutations in 13 out of 17 cases analyzed so far. A list of all cases is presented in Supplementary Material, Table S1. Among 13 cases, G17 had a mutation, c.584G>A (as per GenBank entry NM_000264.2: the A of the ATG of the initiator Met codon is counted as nucleotide +1), on exon 3. This raised two possibilities that explain the effects on the coding for the PTCH protein. One is a missense mutation, p.R195K (as per GenBank entry NP_000255.1). However, as this point mutation was located at the 3′ end of exon 3, and potentially disrupts a splice donor site, we sought the second possibility that the mutation may affect splicing. The data of the microarray analysis showed a significant decreased intensity for the junctional probe exon 3–4, but otherwise the data looked normal indicating an abnormal splicing between exons 3 and 4 (Fig. 3A). The electrophoretogram of the RT–PCR product revealed an additional band which had a larger molecular weight indicating the presence of an aberrant splicing (Fig. 3B, left panel). Sequencing of the additional product demonstrated that abnormal splicing was indeed taking place in which a cryptic splice donor site located in intron 3 was activated, resulting in the insertion of a 37-bp

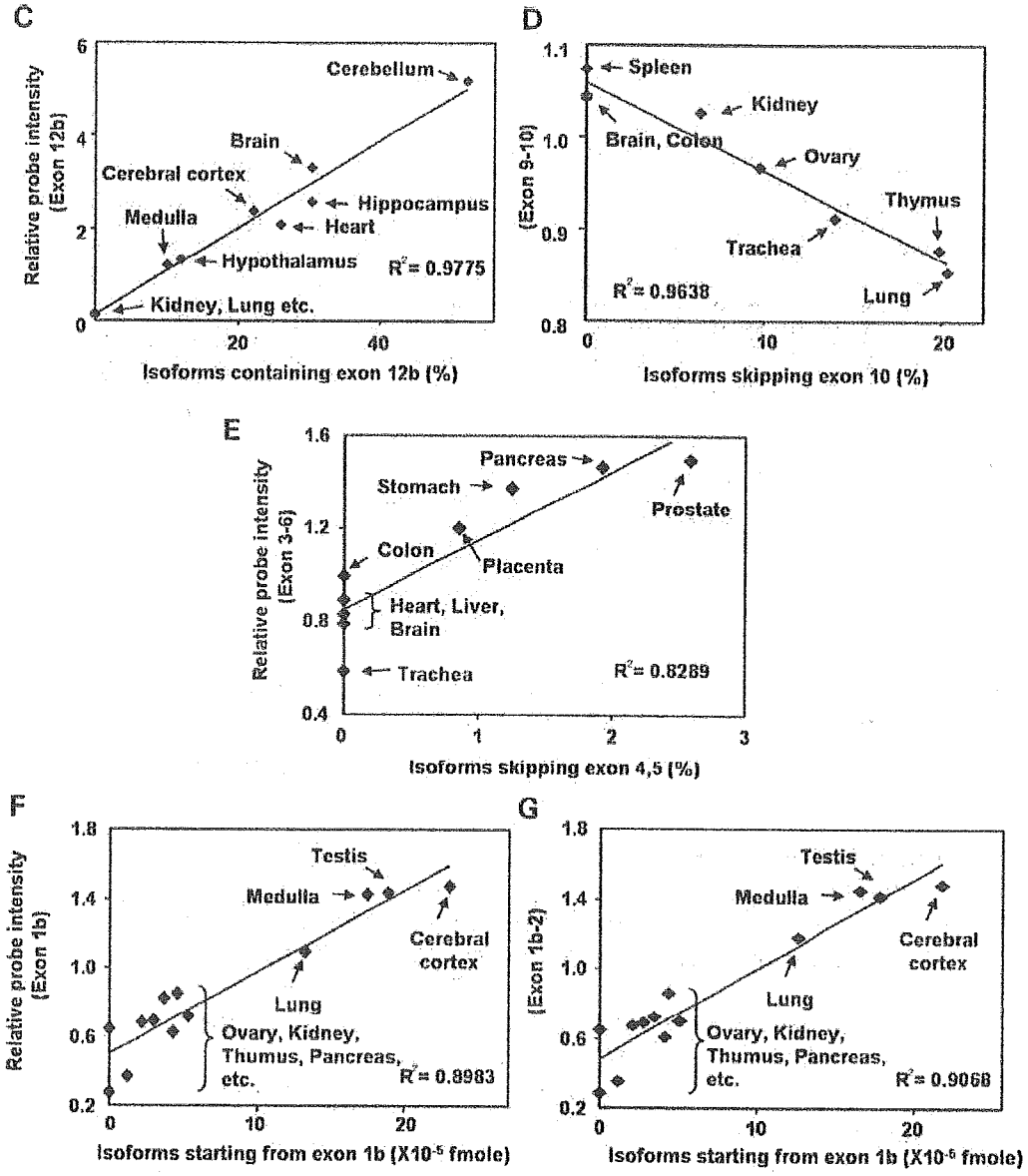

Figure 2. Continued.

intronic sequence between exons 3 and 4 causing a premature termination of the PTCH protein (Fig. 3C). We therefore concluded that it was the aberrant splicing rather than the missense mutation that caused the disease phenotype in this patient.

Another case, G9, did not contain mutations in any of the coding exons of *PTCH*. However, the array data demonstrated a markedly decreased intensity for the junctional probe exon 6–7 (Fig. 3A). The detection of the smaller RT–PCR product in G9 (Fig. 3B, right panel) prompted us to sequence this RT–PCR product, which again revealed the presence of an aberrant splicing. But in this case, a cryptic splice donor site located in exon 6 was activated generating the deletion of an exonic sequence of 87 bp (Fig. 3C). As a result, 29 amino-acid residues located in the first extracellular loop important for Shh binding were deleted. Mutational analysis in intron 6 identified a point mutation, c.945+5G>T, that partially disrupts the consensus sequence for the splice donor site (Fig. 3C).

The third case, G8, had a mutation, c.1526G>A, on exon 11. This mutation presumably results in a missense mutation, p.G509D. However, a growing body of evidence indicates that some single base changes may be capable of switching regulation from positive to negative or vice versa by disrupting or creating exonic splicing enhancers or exonic splicing silencers that are just beginning to be understood (24). Therefore, we investigated such possibilities using microarrays. As shown in Figure 3A, the splicing profile of G8 was similar

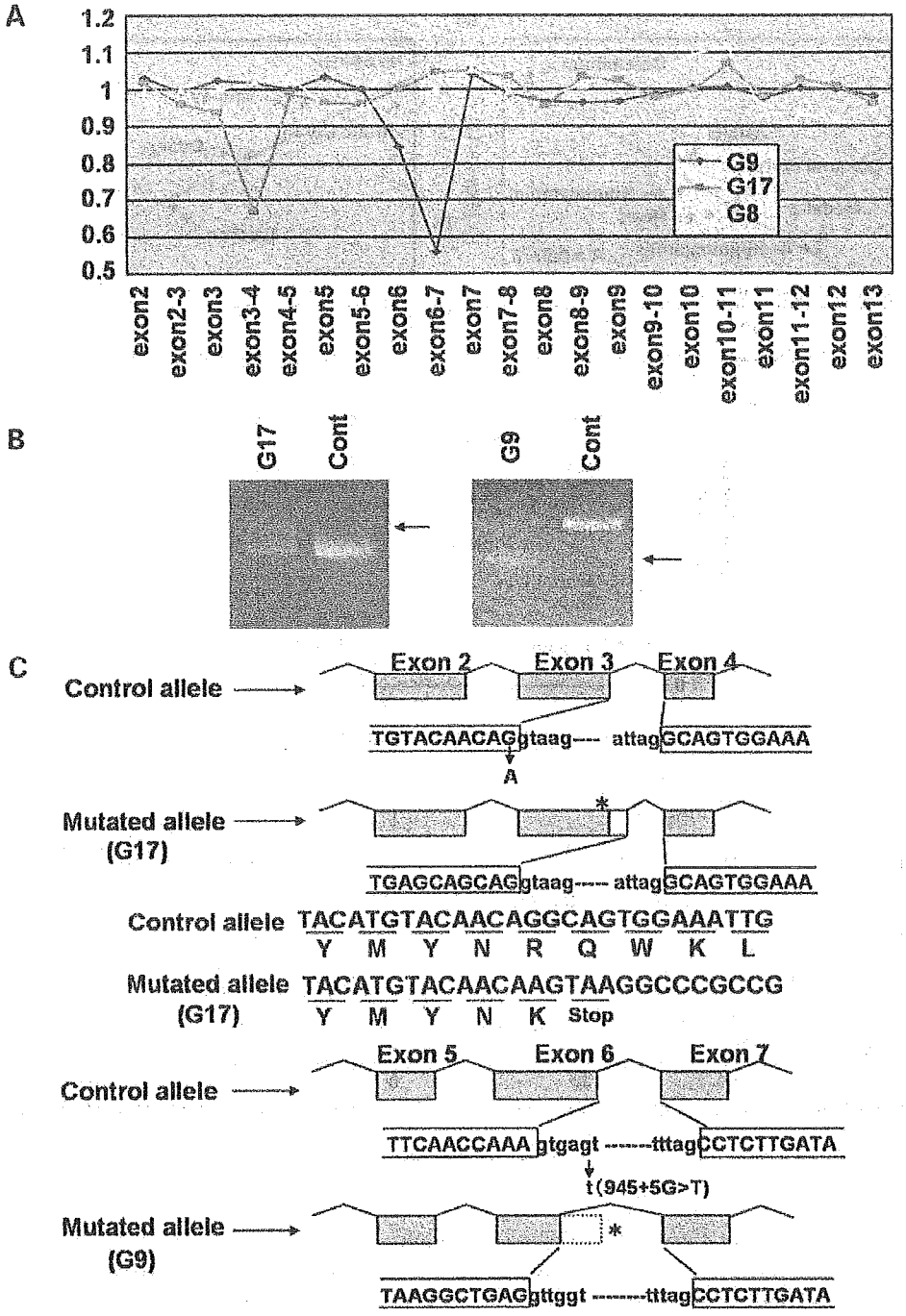

Figure 3. Detection of aberrant splicing in NBCCS patients. (**A**) Normalized probe intensities obtained from NBCCS patients, G8, G9 and G17. Data were analyzed as described in Figure 2A. (**B**) RT–PCR analysis of AS in the *PTCH* gene. Exon 3 (G17) and exon 6 (G9) and flanking exons were amplified by RT–PCR and subjected to agarose gel electrophoresis. The positions of extra bands not observed in a control healthy individual (Cont) are indicated by arrows. (**C**) Schematic representation of abnormal splicing identified in G17 and G9. The positions of the point mutations are indicated by asterisks.

to that of the control and no significant changes in probe intensity were observed. So, we concluded that the mutation found in G8 was a bona fide missense mutation. Interestingly, substitutions of the same amino-acid residue have been reported in NBCCS (p.G509R and p.G509V) (25,26) and these mutations are predicted to disrupt the sterol-sensing domain of the PTCH protein (27). It has also been observed that the mutation p.G509V results in dominant negative activity *in vivo* in *Drosophila* (28). Two out of 17 patients, G7 and G13, did not have a mutation in the region we

have sequenced, nor have any abnormal splicing profile by microarray analysis (Supplementary Material, Table S1). They may have a mutation in non-coding region important for the transcription of *PTCH*, or have a mutation in a gene other than *PTCH*. Altogether, these results demonstrate that exon junction microarrays can be used for detecting the mutations which result in aberrant splicing.

DISCUSSION

In this paper, we described the use of oligonucleotide microarrays containing exon junction probes to investigate tissue-specific AS and to detect disease-associated aberrant splicing. Our microarrays have been demonstrated to be a particularly powerful tool to detect the exon skipping/inclusion type of AS, the most common type (38%) of AS (29), in which even a rare AS event (no more than 3%) can be quantitatively detected using microarrays. Theoretically, previously unrecorded mRNA isoforms can be predicted with this method. If unexpectedly high variation is detected in certain probes, then AS is suggested in this region.

In this study, we found a novel alternative exon named exon 12b. High expression of exon 12b in the brain, particularly in the cerebellum, is intriguing, because proliferative effects of sonic hedgehog (Shh), the ligand of the PTCH protein, in the external granular layer of the cerebellum are well characterized (30) and patients with NBCCS are prone to develop the cerebellar tumor medulloblastoma (6). As exon 12b has an in-frame stop codon, the mRNA isoform containing this exon encodes a truncated PTCH protein that is presumably non-functional. However, this protein may have a dominant negative effect on the wild-type protein and such a possibility needs to be ruled out. Importantly, exon 12b was also found in mice and was also expressed in a brain- and heart-specific manner (data not shown). Recently, the splicing-regulatory element, UGCAUG, was reported to be evolutionarily conserved in introns that flank brain-specific alternative exons (31). Such an element was indeed identified in the intron between exons 12b and 13 both in humans and in mice (data not shown). In contrast, skipping of exon 10 leads to a 52 amino-acid in-frame deletion in the second and third transmembrane domains. The effect on the function of the PTCH protein is currently unknown and this PTCH isoform may be functional in some context.

Detection of AS using exon junction microarrays is limited in several ways. First, as the detection requires probes that match specific exons and splice junctions, probe selection is tightly constrained and some probes are non-informative due to cross-hybridization with somewhere else in human genes. In our study, two out of 42 probes, exon 9–11 and exon 12–13, gave constitutive high intensities and therefore were excluded from further study. Secondly, detection is based on differential expression. Therefore, if two isoforms are present in the same proportion in every tissue, no prediction will result, because normalized probe intensities will behave similarly to the pool. Third, precaution should be taken when interpreting the data on alternative usage of 5′-terminal exons, because not only expression levels but also other factors such as the sequences of forward primers constructed for 5′-terminal exons can have an effect on probe sensitivities.

In this study, we monitored a relatively small number of AS events of one gene in a relatively small set of samples. It should be noted, however, that this is the first report of microarrays used for the identification of aberrant splicings in a genetic disorder. In NBCCS, a considerable number of the patients do not have mutations within the coding region of *PTCH* (23,32). In such patients, exon junction microarrays will be a valuable tool for detecting mutations affecting the splicing event. According to the *PTCH* mutation database (http://www.cybergene.se/cgi-bin/w3-msql/ptchbase/index.html), at least 20 mutations have been reported to potentially result in abnormal splicing in *PTCH* and some of them have been proven experimentally (21,26). Thus, mutations having an effect on splicing events do not seem to be uncommon. Not only NBCCS, but also an increasing number of genetic diseases are known to be caused by mutations that alter splicing in *cis* (at least 15% of point mutations) (5 and reviewed in 24). Some of these mutations weaken or activate *cis*-acting element such as intronic splicing enhancers or intronic splicing silencers that are sometimes located in intronic sequences distant from exons. In such cases, mutations are not identified by the sequencing of coding regions, warranting the use of exon junction microarrays like the ones described in this study to rapidly search the candidate regions to be sequenced.

Apart from *PTCH*, germ-line mutations in the genes encoding Shh signaling components such as Shh or Gli3 are responsible for a variety of genetic disorders, most of which are accompanied by developmental anomalies in the central nervous system (33). In addition, somatic mutations of the genes involved in this signaling pathway, such as *PTCH2* or *Smoothened*, are associated with various sporadic cancers (33). Moreover, distinct roles of *PTCH2* splice variants in Shh signaling have recently been proposed (34). Therefore, oligonucleotide microarrays containing exon and exon junction probes widely covering these genes would be an attractive tool to study the pathogenesis of these disorders. In addition, these microarrays may also be useful to detect mutations in the dystrophin gene, because the removal of exon(s) is found in 60–65% of patients with Duchenne and Becker muscular dystrophies (35).

MATERIALS AND METHODS

Plasmids

The plasmid encoding myc-tagged full-length PTCH protein (exons 1b–23) (pMyc-Ptc1) was kindly provided by Dr Jeffrey Ming (15). The plasmids encoding other isoforms of *PTCH* were created by PCR-mediated mutagenesis as described previously (36) using pMyc-Ptc1 as a template. The details of construction are available on request.

Oligonucleotide microarray construction

All probes 34–76 bp in length were designed to have approximately the same annealing temperature (T_m). Splice junction probes between two exons were designed so that the T_m for the first exon sequence is the same as that for the second

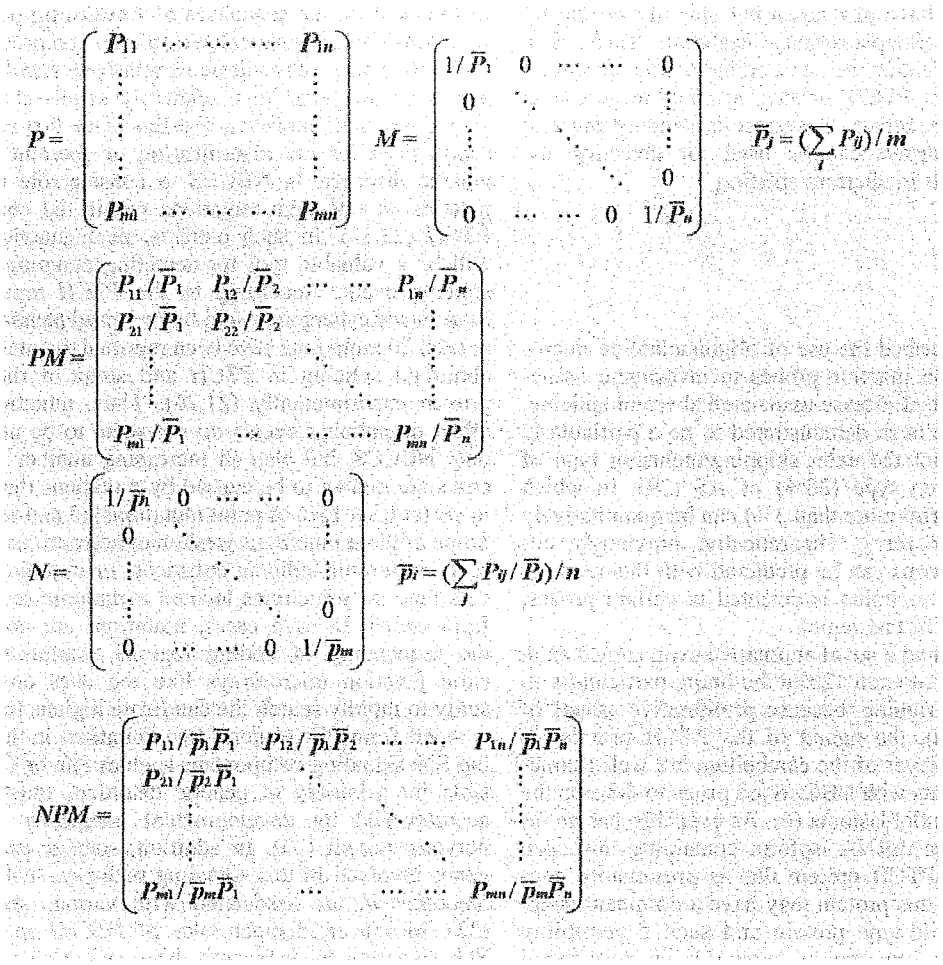

Figure 4. Matrix representation. The relationship between probes and tissues can be represented by an $m \times n$ matrix P. The total number of probes is m, and n is the total number of tissues. Let M be an $n \times n$ diagonal matrix where $\overline{P}j$ represents the mean P in tissue j. The probe intensities normalized by total *PTCH* expression in each tissue can then be expressed as PM. Let N be an $m \times m$ diagonal matrix where $\overline{p}i$ represents the mean normalized P in probe i. The matrix NPM therefore represents the probe intensities normalized by both factors described above and $P_{ij}/\overline{p}i\overline{P}j$ is used in Figures 2 and 3.

exon sequence (Supplementary Material, Table S2). The oligonucleotide microarray, GenoPal™ (Mitsubishi Rayon Co., Ltd), was made in the following manner. Plastic hollow fibers were bundled in an orderly arrangement, and hardened with resin to form a block (Supplementary Material, Fig. S2). Oligonucleotide-capture probes were chemically bonded inside each hollow fiber with hydrophilic gel. The block was then sliced to make thin chips, each of which was set into a holder (http://www.mrc.co.jp/genome/e/index.html for details).

Preparation of labeled probe

Total RNA from a panel of human tissues was purchased from Ambion. Lymphoblastoid cell lines from NBCCS patients immortalized by Epstein-Barr virus were also used to obtain total RNA. All studies using patient samples were approved by the local ethic committee. The template for *in vitro* transcription was generated through a modified procedure using a SuperScript One-Step RT–PCR System with Platinum Taq (Invitrogen). In brief, 2.5 μg of total RNA was reverse-transcribed with SuperScriptIII RT/Platinum Taq Mix and T7-exon15 reverse primer, 5′-TAATACGACTCACTATAGGGGTCATATTCTCTGGTTTCCCGAGGTACAATGTC-3′, for 30 min at 50°C. To evaluate the quality of RNA, T7-GAPDH reverse primer, 5′-TAATACGACTCACTATAGGGAGGAGGGGAGATTCAGTGTGGT-3′ was also added in the reaction. The RNA was degraded with the addition of RNaseH (Invitrogen) for 15 min at 37°C. After the addition of forward primers specific for each first exon (5′-AGCGCCTGTTTACCCAGGAG-3′ for exon 1a, 5′-GGACCGGGACTATCTGCACC-3′ for exon 1b, 5′-AAATGCCGCGCCGGGGAGCAGCCT-3′ for exon 1d and 5′-TTCTCGGCGGGGGTCCAGTT-3′ for exon 1e) and T7-exon15 reverse primer (final concentration 0.5 μM each) and the activation of Platinum Taq for 10 min at 94°C, PCR was run for 30 cycles of denaturation at 94°C for 30 s, annealing at 56°C for 30 s, and extension at 72°C for 2 min. Plasmid DNA encoding

various isoforms of *PTCH* was also subjected to PCR using the exon 2 forward primer, 5′-GCTGAGAGCGAAGTTTC AGA-3′, and T7-exon 15 reverse primer. PCR products generated from the reverse-transcribed cDNA and the plasmid DNA were purified with a PCR Purification kit (QIAGEN) according to the manufacturer's instructions. Then, a Cy-5-labeled probe was generated by using the PCR product as a template with a MEGAscript T7 kit (Ambion). The reaction contained a 1:1.5 mixture of uridine triphosphate (UTP) and Cy-5-UTP (Amersham Biosciences). The product was purified with an RNeasy mini kit (QIAGEN) following the manufacturer's instructions. One microgram of labeled probe was fragmented with RNA Fragmentation Reagents (Ambion) at 70°C for 3 min.

Hybridization and detection

Hybridization was carried out in a final volume of 0.1 ml injected into a hybridization chamber at 65°C for 16–24 h in 0.5× SSC with 0.2% SDS. The microarray chip was then washed twice in 0.5× SSC with 0.2% SDS at 55°C for 20 min and once in 0.5× SSC at 55°C for 10 min, before being slowly cooled to room temperature. GenoPal was then scanned and the image was captured with a cooled CCD-type Microarray Image Analyzer (Mitsubishi Rayon Co., Ltd). Fluorescent intensity was analyzed with software developed by Mitsubishi Rayon Co., Ltd. Fluorescence throughout the three-dimensional structure of each array feature can be efficiently captured due to the long focal depth of the optical system of the image analyzer (http://www.mrc.co.jp/genome/e/index.html).

Microarray data analysis

To analyze changes in the AS of a gene, changes in the total expression of a gene and differences in probe sensitivity should be separated and excluded. We used a simple and generalized pooling strategy presented by Le *et al.* (20) with modifications. The probe response P_{ij} (represented as an S/N ratio) for a specific probe i to a specific tissue sample j was normalized using the total expression of *PTCH* and probe sensitivity as described in Figure 4.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We thank Kaori Takeuchi-Inoue and Mayu Yamazaki-Inoue for technical support, and Kayoko Saito for preparing the manuscript. This work was supported by the Naito Foundation and Grants for Cancer Research and Child Health and Development from the Ministry of Health, Labour and Welfare; a Grant-in-Aid for Scientific Research and the Budget for Nuclear Research from the Ministry of Education, Culture, Sports, Science and Technology.

Conflict of Interest statement. None declared.

REFERENCES

1. Kan, Z., Rouchka, E.C., Gish, W.R. and States, D.J. (2001) Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.*, **11**, 889–900.
2. Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. and Shoemaker, D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science*, **302**, 2141–2144.
3. Boise, L.H., Gonzalez-Garcia, M., Postema, C.E., Ding, L., Lindsten, T., Turka, L.A., Mao, X., Nunez, G. and Thompson, C.B. (1993) bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell*, **74**, 597–608.
4. Quelle, D.E., Zindy, F., Ashmun, R.A. and Sherr, C.J. (1995) Alternative reading frames of the INK4a tumor suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell*, **83**, 993–1000.
5. Krawczak, M., Reiss, J. and Cooper, D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.*, **90**, 41–54.
6. Gorlin, R.J. (1987) Nevoid basal-cell carcinoma syndrome. *Medicine*, **66**, 98–113.
7. Johnson, R.L., Rothman, A.L., Xie, J., Goodrich, L.V., Bare, J.W., Bonifas, J.M., Quinn, A.G., Myers, R.M., Cox, D.R., Epstein, E.H., Jr *et al.* (1996) Human homolog of *patched*, a candidate gene for the basal cell nevus syndrome. *Science*, **272**, 1668–1671.
8. Hahn, H., Wicking, C., Zaphiropoulous, P.G., Gailani, M.R., Shanley, S., Chidambaram, A., Vorechovsky, I., Holmberg, E., Unden, A.B., Gillies, S. *et al.* (1996) Mutations of the human homolog of *Drosophila patched* in the nevoid basal cell carcinoma syndrome. *Cell*, **85**, 841–851.
9. Gailani, M.R., Stahle-Backdahl, M., Leffell, D.J., Glynn, M., Zaphiropoulos, P.G., Pressman, C., Unden, A.B., Dean, M., Brash, D.E., Bale, A.E. *et al.* (1996) The role of the human homologue of *Drosophila patched* in sporadic basal cell carcinomas. *Nat. Genet.*, **14**, 78–81.
10. Unden, A.B., Holmberg, E., Lundh-Rozell, B., Stahle-Backdahl, M., Zaphiropoulos, P.G., Toftgard, R. and Vorechovsky, I. (1996) Mutations in the human homologue of *Drosophila patched* (*PTCH*) in basal cell carcinomas and the Gorlin syndrome: different *in vivo* mechanisms of PTCH inactivation. *Cancer Res.*, **56**, 4562–4565.
11. Smyth, I., Narang, M.A., Evans, T., Heimann, C., Nakamura, Y., Chenevix-Trench, G., Pietsch, T., Wicking, C. and Wainwright, B.J. (1999) Isolation and characterization of human *patched 2* (*PTCH2*), a putative tumour suppressor gene in basal cell carcinoma and medulloblastoma on chromosome 1p32. *Hum. Mol. Genet.*, **8**, 291–297.
12. Zaphiropoulos, P.G., Unden, A.B., Rahnama, F., Hollingsworth, R.E. and Toftgård, R. (1999) *PTCH2*, a novel human patched gene, undergoing alternative splicing and up-regulated in basal cell carcinomas. *Cancer Res.*, **59**, 787–792.
13. Kogerman, P., Krause, D., Rahnama, F., Kogerman, L., Unden, A.B., Zaphiropoulos, P.G. and Toftgård, R. (2002) Alternative first exons of *PTCH1* are differentially regulated *in vivo* and may confer different functions to the PTCH1 protein. *Oncogene*, **21**, 6007–6016.
14. Ågren, M., Kogerman, P., Kleman, M.I., Wessling, M. and Toftgård, R. (2004) Expression of the *PTCH1* tumor suppressor gene is regulated by alternative promoters and a single functional Gli-binding site. *Gene*, **330**, 101–114.
15. Nagao, K., Toyoda, M., Takeuchi-Inoue, K., Fujii, K., Yamada, M. and Miyashita, T. (2005) Identification and characterization of multiple isoforms of a murine and human tumor suppressor, *Patched*, having distinct first exons. *Genomics*, **85**, 462–471.
16. Shimokawa, T., Rahnama, F. and Zaphiropoulos, P.G. (2004) A novel first exon of the *Patched1* gene is upregulated by Hedgehog signaling resulting in a protein with pathway inhibitory functions. *FEBS Lett.*, **578**, 157–162.
17. Castle, J., Garrett-Engele, P., Armour, C.D., Duenwald, S.J., Loerch, P.M., Meyer, M.R., Schadt, E.E., Stoughton, R., Parrish, M.L., Shoemaker, D.D. *et al.* (2003) Optimization of oligonucleotide arrays and RNA amplification protocols for analysis of transcript structure and alternative splicing. *Genome Biol.*, **4**, R66.
18. Wang, H., Hubbell, E., Hu, J.S., Mei, G., Cline, M., Lu, G., Clark, T., Siani-Rose, M.A., Ares, M., Kulp, D.C. *et al.* (2003) Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics*, **19** (Suppl. 1), i315–i322.

19. Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S. and Fu, X.D. (2002) Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.*, **20**, 353–358.
20. Le, K., Mitsouras, K., Roy, M., Wang, Q., Xu, Q., Nelson, S.F. and Lee, C. (2004) Detecting tissue-specific regulation of alternative splicing as a qualitative change in microarray data. *Nucleic Acids Res.*, **32**, e180.
21. Smyth, I., Wicking, C., Wainwright, B. and Chenevix-Trench, G. (1998) The effects of splice site mutations in patients with naevoid basal cell carcinoma syndrome. *Hum. Genet.*, **102**, 598–601.
22. Wicking, C., Gillies, S., Smyth, I., Shanley, S., Fowles, L., Ratcliffe, J., Wainwright, B. and Chenevix-Trench, G. (1997) *De novo* mutations of the *Patched* gene in nevoid basal cell carcinoma syndrome help to define the clinical phenotype. *Am. J. Med. Genet.*, **73**, 304–307.
23. Fujii, K., Kohno, Y., Sugita, K., Nakamura, M., Moroi, Y., Urabe, K., Furue, M., Yamada, M. and Miyashita, T. (2003) Mutations in the human homologue of *Drosophila patched* in Japanese nevoid basal cell carcinoma syndrome patients. *Hum. Mutat.*, **21**, 451–452.
24. Garcia-Blanco, M.A., Baraniak, A.P. and Lasda, E.L. (2004) Alternative splicing in disease and therapy. *Nat. Biotechnol.*, **22**, 535–546.
25. Chidambaram, A., Goldstein, A.M., Gailani, M.R., Gerrard, B., Bale, S.J., DiGiovanna, J.J., Bale, A.E. and Dean, M. (1996) Mutations in the human homologue of the *Drosophila patched* gene in Caucasian and African-American nevoid basal cell carcinoma syndrome patients. *Cancer Res.*, **56**, 4599–4601.
26. Pastorino, L., Cusano, R., Nasti, S., Faravelli, F., Forzano, F., Baldo, C., Barile, M., Gliori, S., Muggianu, M., Ghigliotti, G. *et al.* (2005) Molecular characterization of Italian nevoid basal cell carcinoma syndrome patients. *Hum. Mutat.*, **25**, 322–323.
27. Kuwabara, P.E. and Labouesse, M. (2002) The sterol-sensing domain: multiple families, a unique role? *Trends Genet.*, **18**, 193–201.
28. Hime, G.R., Lada, H., Fietz, M.J., Gillies, S., Passmore, A., Wicking, C. and Wainwright, B.J. (2004) Functional analysis in *Drosophila* indicates that the NBCCS/PTCH1 mutation G509V results in activation of smoothened through a dominant-negative mechanism. *Dev. Dyn.*, **229**, 780–790.
29. Sugnet, C.W., Kent, W.J., Ares, M., Jr and Haussler, D. (2004) Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.*, 66–77.
30. Dahmane, N. and Altaba, A. (1999) Sonic hedgehog regulates the growth and patterning of the cerebellum. *Development*, **126**, 3089–3100.
31. Minovitsky, S., Gee, S.L., Schokrpur, S., Dubchak, I. and Conboy, J.G. (2005) The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. *Nucleic Acids Res.*, **33**, 714–724.
32. Wicking, C., Shanley, S., Smyth, I., Gillies, S., Negus, K., Graham, S., Suthers, G., Haites, N., Edwards, M., Wainwright, B. *et al.* (1997) Most germ-line mutations in the nevoid basal cell carcinoma syndrome lead to a premature termination of the PATCHED protein, and no genotype–phenotype correlations are evident. *Am. J. Hum. Genet.*, **60**, 21–26.
33. Cohen, M.M., Jr (2003) The hedgehog signaling network. *Am. J. Med. Genet.*, **123A**, 5–28.
34. Rahnama, F., Toftgård, R. and Zaphiropoulos, P.G. (2004) Distinct roles of PTCH2 splice variants in Hedgehog signalling. *Biochem. J.*, **378**, 325–334.
35. Muntoni, F., Torelli, S. and Ferlini, A. (2003) Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol.*, **2**, 731–740.
36. Imai, Y., Matsushima, Y., Sugimura, T. and Terada, M. (1991) A simple and rapid method for generating a deletion by PCR. *Nucleic Acids Res.*, **19**, 2785.

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Genomics 85 (2005) 462–471

GENOMICS

www.elsevier.com/locate/ygeno

Identification and characterization of multiple isoforms of a murine and human tumor suppressor, *patched*, having distinct first exons☆

Kazuaki Nagao[a], Masashi Toyoda[a], Kaori Takeuchi-Inoue[a], Katsunori Fujii[b], Masao Yamada[a], Toshiyuki Miyashita[a,*]

[a]*Department of Genetics, National Research Institute for Child Health and Development, 2-10-1 Ohkura, Setagaya-ku, Tokyo 157-8535, Japan*
[b]*Department of Pediatrics, Graduate School of Medicine, Chiba University, 1-8-1 Inohana, Chuo-ku, Chiba 260-8670, Japan*

Received 3 September 2004; accepted 23 November 2004
Available online 11 January 2005

Abstract

Mutations in mouse and human *patched* (*PTCH*) genes are associated with birth defects and cancer. PTCH, a 12-pass transmembrane protein, is a receptor for Sonic hedgehog (Shh) signaling proteins. Shh proteins activate transcription of target genes, including *PTCH,* via GLI transcription factors. Here we identified seven and five isoforms of human and mouse *PTCH* mRNA, respectively, which are generated by the complex alternative use of five exons as the first exon (exons 1a to 1e in the 5′-to-3′ order). Although expression profiles of these isoforms were highly variable among human tissues, three of them, *PTCHa, PTCHb,* and *PTCHd,* were predominantly expressed in most tissues, *PTCHd* being most ubiquitous. In contrast, *PTCHb* was always predominant and reached a maximum at E10.5 during mouse development. These three mRNA isoforms encode three PTCH proteins with distinct N-termini, $PTCH_L$, $PTCH_M$, and $PTCH_S$. The expression of these three isoforms was regulated by GLI transcription factors, and at least two functional GLI-binding sequences were identified, one in exon 1a and the other between exon 1a and exon 1b. $PTCH_L$ and $PTCH_M$ were equally active in terms of suppressing GLI-mediated transcription and inducing apoptosis. $PTCH_S$ protein (encoded by *PTCHd*), lacking the first transmembrane domain, was more unstable than the other two, resulting in a reduced activity. This study may shed light on the mechanism whereby a single *PTCH* gene plays a role in both tumor cell growth and embryonic development.

Keywords: *Patched*; Sonic hedgehog; Basal cell carcinoma; Medulloblastoma; Alternative splicing

The Sonic hedgehog (Shh) signaling cascade is pivotal to embryonic development, because holoprosencephaly (HPE), characterized by a failure of the forebrain to separate completely into hemispheres, and HPE-like abnormalities are associated with a loss of Shh function in humans and in mice [1–3]. The role of the Shh pathway in tumorigenesis was also established with the discovery that inactivating mutations in the *Patched* (*PTCH*) gene, which encodes one component of the Shh receptor, are responsible for the inherited cancer predisposition disorder known as Gorlin's or nevoid basal cell carcinoma syndrome (NBCCS) [4,5], as well as sporadic basal cell carcinomas (BCCs) and medulloblastomas [6–8]. NBCCS is an autosomal dominant neurocutaneous disorder characterized by developmental abnormalities such as palmar and plantar pits, jaw cysts, calcification of the falx cerebri, and skeletal anomalies and also by a predisposition to cancers such as BCC and medulloblastoma [9]. Familial and sporadic BCCs display loss of heterozygosity in this region, consistent with *PTCH* being a tumor suppressor gene [6,10]. In addition, activating mutations in *Smoothened* (*Smo*), also encoding another component of the Shh receptor, have been detected in BCCs [11], further emphasizing the importance of this pathway in tumor development. More importantly, the recent finding that this pathway is essential for growth of a wide range of tumor types not associated with NBCCS, such as lung

☆ Sequence data from this article have been deposited with the GenBank Library under Accession Nos. AB164615, AB164616, and AB189436–AB189442.

* Corresponding author. Fax: +81 3 5494 7035.
E-mail address: tmiyashita@nch.go.jp (T. Miyashita).

0888-7543/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2004.11.014

cancers or digestive tract tumors, sheds light on potential new diagnostic and therapeutic approaches [12–14].

PTCH, a 12-pass transmembrane protein, is the ligand-binding component of the Shh receptor complex. In the absence of Shh binding, PTCH is thought to hold Smo, a 7-pass transmembrane protein, in an inactive state and thus inhibit signaling to downstream genes. Upon the binding of Shh, the inhibition of Smo is released and signaling is transduced, leading to the activation of target genes by the Gli family of transcription factors [15]. The transcription of *PTCH* itself is induced by Shh pathway activity [16], thus generating a negative feedback loop, which may play an important role in tumor suppression by inhibiting a sustained activation of the pathway.

Hahn et al. predicted that there are three different forms of the PTCH protein present in humans: the ancestral form and two human-specific forms [4]. Recently, a detailed characterization of three alternative first exons was reported [17]. However, our study using the 5′ rapid amplification of cDNA ends (5′RACE) technique revealed the existence of an additional first exon and unexpectedly complex splicing between the first and the second exons that is evolutionarily conserved across species. Therefore, the characterization of several potential forms of the PTCH protein may reveal the mechanism whereby a single *PTCH* gene could play a role in different pathways, and the determination of the regulation of different splice forms of *PTCH* mRNA may shed light on the apparent role of the gene in tumor cell growth as well as embryonic development. Here we characterize multiple isoforms of *PTCH* in humans and mice and discuss the functions of their products, expression profiles, and transcriptional regulation.

Results

Isolation of isoforms of human and mouse PTCH

PTCH is a multiexon gene comprising 23 exons distributed over a region of ~70 kb. To date, three cDNA sequences encoding the human *PTCH* gene's first exon have been reported and named exons 1, 1A, and 1B [17], and another exon has recently been deposited with GenBank (exon 1a described below, GenBank Accession No. BC043542). In contrast, only a single mRNA species of *PTCH* has been reported in mice [18] (GenBank Accession No. U46155). Due to the use of alternative exons, several mRNA isoforms are generated. On the basis of this background we performed a comprehensive analysis of the 5′ structure of mRNA species derived from the human *PTCH* gene employing the 5′RACE technique. Sequencing of 31 RACE clones revealed an additional alternative first exon (exon 1c described below, submitted to GenBank as Accession No. AB189438) and complex splicing between the first and the second exon. Using a genomic sequence containing the *PTCH* gene (GenBank Accession No. AL161729), the precise genomic organization of the human *PTCH* gene was determined as shown in Fig. 1. For the sake

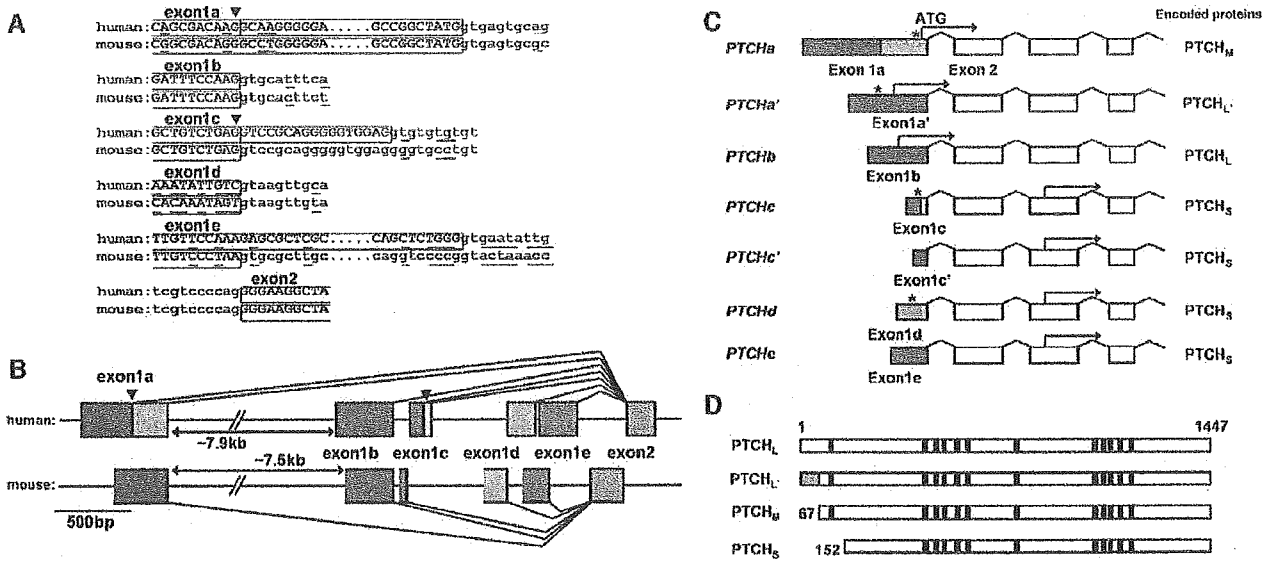

Fig. 1. Identification of human and mouse *PTCH* isoforms. (A) Comparison of human and mouse exon–intron boundaries. Upper- and lowercase letters indicate exon and intron sequences, respectively. Nucleotides not conserved between the two species are underlined. Alternative splice donor sites are indicated by arrowheads. (B) 5′ region of human and mouse *PTCH* gene structure. The 5′ ends of the mouse first exons have not been determined. (C) 5′ structure of *PTCH* isoforms. The positions of the first methionine codons and in-frame stop codons are indicated by arrows and asterisks, respectively. In four of seven mRNAs, in-frame stop codons were identified. The first in-frame methionine codon could be determined in the other three transcripts since the 5′RACE system we employed amplifies only full-length transcripts [47]. (D) PTCH protein isoforms encoded by mRNA species described in (C). Numbers refer to amino acid positions relative to the first methionine of $PTCH_L$. The positions of the 12 transmembrane regions are indicated by filled boxes. $PTCH_{L'}$ has 65 unique amino acid residues at the N-terminus depicted with a shaded box.

of simplicity, we named the first exons exon 1a to 1e on the basis of their 5′-to-3′ order. Thus, exons 1b, 1d, and 1e are the former exons 1B, 1, and 1A, respectively. In addition to multiple first exons, we found that alternative 5′ splice sites allow the shortening of exons 1a and 1c, generating exons 1a′ and 1c′ (Fig. 1C). The complex alternative splicing described above thus generates up to seven mRNA species, each with its own distinct 5′ sequence (Figs. 1B and 1C). RT-PCR using isoform-specific forward primers for each alternative exon 1 and a common reverse primer for exon 2 indeed validated the existence of the seven different mRNAs. These mRNA isoforms encode four PTCH proteins termed $PTCH_L$, $PTCH_{L'}$, $PTCH_M$, and $PTCH_S$ (Figs. 1C and 1D). $PTCH_S$ is an N-terminally truncated PTCH protein that lacks the first transmembrane domain (Fig. 1D). Although only a single species of *PTCH* mRNA has been reported in mice, a comparison of the human *PTCH* genomic sequence with the mouse sequence (NCBI Locus NT_039587) suggested the existence of multiple first exons. In this study, mouse and human *Patched* genes are collectively referred to by the human nomenclature (*PTCH,* whereas mouse *Patched* is often called *Ptc*). RT-PCR using the forward primers constructed at mouse putative first exons and reverse primers at exon 2 demonstrated that most of the *PTCH* isoforms found in humans are indeed conserved in mice. At least in mouse P19 cells and several mouse tissues from which total RNA was extracted, *PTCHa′* and *PTCHc* have not been identified and the splice donor site at exon 1e was different from that of humans (Fig. 1A). All exons were flanked by splice junctions that conformed to the consensus GT–rule except for exon 1a′–exon 2 in humans, in which the GC–AG intron was observed. GC–AG introns are occasionally found and processed by the same splicing pathway as conventional GT–AG introns [19].

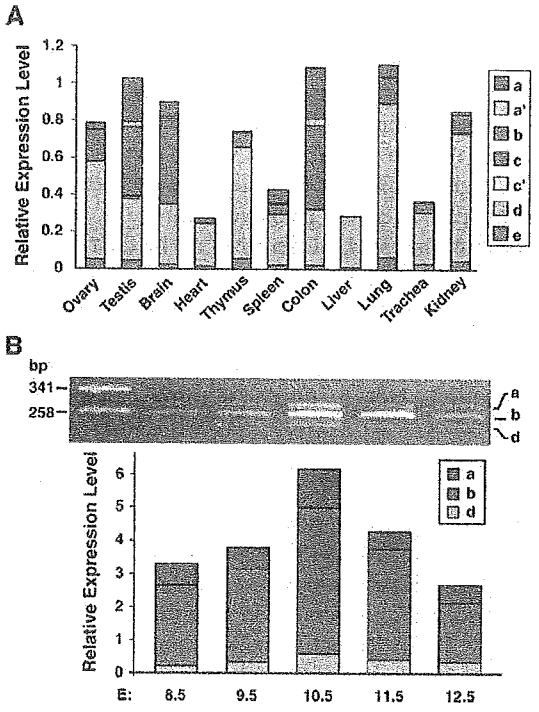

Fig. 2. Expression profiling of *PTCH* isoforms. (A) RT-PCR analysis of expression profiles in various tissues. Total RNA obtained from a panel of human tissues was subjected to RT-PCR. Forward primers specific to each of the first exons and a common reverse primer for exon 2 were synthesized and used for PCR. The RT-PCR products were quantified with an Agilent 2100 bioanalyzer. *PTCH* expression levels were normalized to those of *GAPDH*. Exons with relative expression levels lower than 0.007 do not appear in the graph. (B) RT-PCR analysis of expression profiles in various mouse developmental stages. Total RNA obtained from mouse embryos at various developmental stages was subjected to RT-PCR using mouse-specific primers. Mean *PTCH* expression levels normalized to β-*actin* expression are presented at the bottom (n = 2–4).

Expression profiles of three isoforms of PTCH in various tissues

Selective usage of the 5′-most exons suggests a complex tissue-specific transcriptional regulation. Therefore, to investigate the expression profiles of *PTCH* isoforms, RT-PCR was performed with isoform-specific primers for the first alternative exons using total RNA from a panel of human tissues, and profiles were analyzed with an Agilent 2100 bioanalyzer. As shown in Fig. 2A, *PTCH* was expressed in a wide range of human tissues. However, the levels of total *PTCH* RNA varied among human tissues. For example, the heart and liver showed low levels of expression, which is largely consistent with previous reports on human and mouse *PTCH* expression [18,20]. Expression profiles of the *PTCH* isoforms were also highly variable among tissues. While *PTCHd* (encoding $PTCH_S$) was widely expressed, the expression of *PTCHa* (encoding $PTCH_M$) and *PTCHb* (encoding $PTCH_L$) was relatively restricted. For example, *PTCHb* was expressed in all the analyzed tissues apart from liver, while the expression of *PTCHa* was more restricted, showing virtually no expression in the heart, thymus, liver, and trachea. The other *PTCH* isoforms using exons 1a′, 1c, 1c′, and 1e were found to be expressed at very low levels if at all throughout the tissues. Therefore, we focused on *PTCHa, PTCHb,* and *PTCHd* in further experiments. Since Shh signaling plays a key role in embryonic development, we next investigated the expression profile in mouse embryogenesis. Consistent with a previous report, the expression of *PTCH* reached a maximum at E10.5, at which point the limb buds become increasingly prominent, and declined thereafter [18]. Notably, in contrast to human adult tissues, the expression of *PTCHb* was always prevalent during embryonic development (Fig. 2B).

Transcriptional regulation of PTCH isoforms by GLI

It is well known that *PTCH* itself is one of the target genes in the Shh signaling network creating a negative feedback loop and a balance via the antagonism of Shh and