BJP British Journal of Pharmacology

DOI:10.1111/bph.12526
www.brjpharmacol.org

# INTERNATIONAL UNION OF BASIC AND CLINICAL PHARMACOLOGY REVIEW

# Should pharmacologists care about alternative splicing? IUPHAR Review 4

T I Bonner

*National Institute of Mental Health, Bethesda, MD, USA*

**Correspondence**
Tom I. Bonner, Office of the Scientific Director, NIMH, 49 Convent Drive, Bldg 49, Rm B1B80, MSC4405, Bethesda MD 20892-4405, USA.
E-mail: tibonner@mail.nih.gov

**Keywords**
alternative splicing; splice variant nomenclature; 5-$HT_4$ receptor; $PAC_1$ receptor; $Ca_V1.2$; $Ca_V3.1$

**Received**
26 August 2013
**Revised**
27 October 2013
**Accepted**
13 November 2013

Tom Bonner is a member of the Evolving Pharmacology Group of the Nomenclature Committee of the International Union of Basic and Clinical Pharmacology (NC-IUPHAR). The role of this group is to monitor the 'de-orphanisation' of orphan GPCRs, to classify the remaining orphan receptors and to define the criteria used to recommend the pairing of an orphan receptor with its cognate ligand(s). Further information can be found on the IUPHAR/BPS Guide to PHARMACOLOGY website (http://www.guidetopharmacology.org/). This has recently been updated to reflect the current state of orphan GPCRs in each of the three classes, and has the latest information on new pairings under the 'Recent receptor-ligand pairings' section.

Alternative splicing of mRNAs occurs in the majority of human genes, and most differential splicing results in different protein isoforms with possibly different functional properties. However, there are many reported splicing variations that may be quite rare, and not all combinatorially possible variants of a given gene are expressed at significant levels. Genes of interest to pharmacologists are frequently expressed at such low levels that they are not adequately represented in genome-wide studies of transcription. In single-gene studies, data are commonly available on the relative abundance and functional significance of individual alternatively spliced exons, but there are rarely data that quantitate the relative abundance of full-length transcripts and define which combinations of exons are significant. A number of criteria for judging the significance of splice variants and suggestions for their nomenclature are discussed.

## Introduction

The Nomenclature Committee of the International Union of Basic and Clinical Pharmacology (NC-IUPHAR) originally proposed guidelines for the naming of splice variants in 1996 (Vanhoutte *et al*., 1996). Specifically, it stated that 'new splice variants, if pharmacologically relevant, should be indicated by subscript letters in lowercase, in parentheses'. This format has not been widely adopted, especially in the subscripting aspect, and is not widely used outside of pharmacology, where various other naming systems are used. Furthermore, the meaning of 'pharmacologically relevant' is somewhat ambiguous and needs some clarification.

## Types of splice variants

Before considering the issues, let us define some terms. The possible types of alternative mRNA transcripts of a given gene are as follows: alternative first exons, cassette exons (which can be either skipped or included), a choice between two (or possibly more) mutually exclusive exons, a change in length at the 5′ end of an exon due to use of an alternative splice acceptor (often referred to as an alternative 3′ splice site, as it occurs at the 3′ end of the intron), a change in length at the 3′ end of an exon due to usage of an alternative splice donor (also referred to as an alternative 5′ splice site), the retention of an intron, and alternative last exons (Figure 1). Alternative

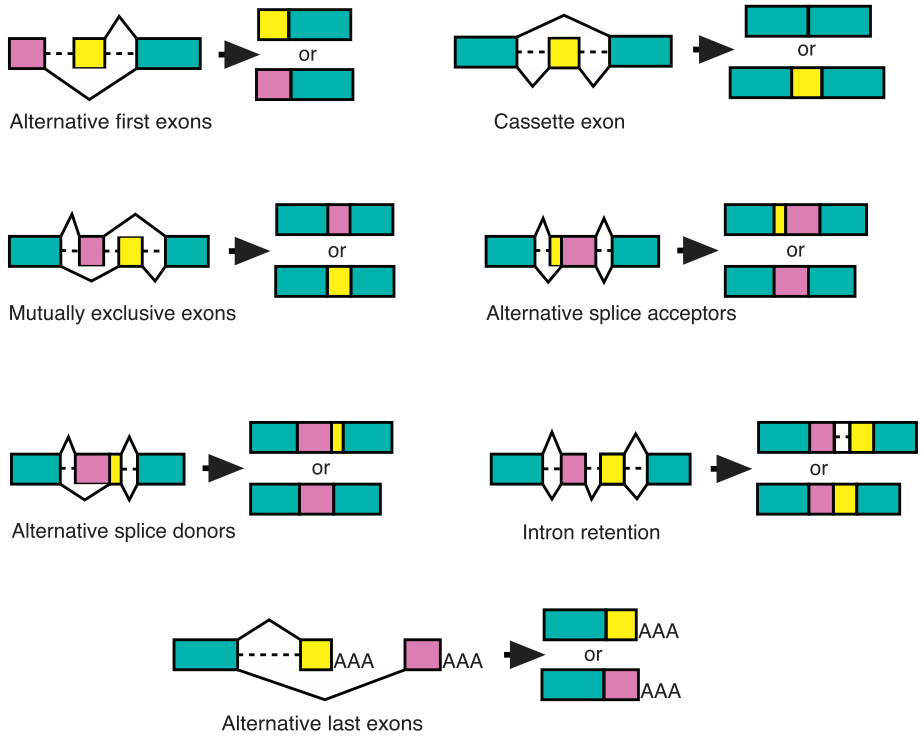



**Figure 1**

Types of splice variants. Exons are represented by coloured boxes and introns by dashed lines. Solid lines represent sequences spliced out by alternative splicing choices, and the resulting alternatively spliced products are shown to the right of the arrows.

first exons where one or a few of the first exons in a transcript are exclusive to a given promoter and are subsequently spliced to exons common to various transcripts are not precisely alternative splicing events but rather use of alternative promoters. One of the more common and plausibly significant splicing variations that occurs in GPCRs results in alternative carboxyl terminals for the protein, which could affect interaction with scaffolding and signalling proteins inside the cell.

A large-scale cDNA sequencing study (400 million reads of 32 bases; Wang *et al.*, 2008) looking at 10 diverse human tissues and five cell lines estimated the frequencies of alternative transcript types as cassette exon 38%, alternative acceptor 17%, alternative donor 15%, alternative first exon 14%, alternative last exon 9%, mutually exclusive exons 4% and retained intron 1%. The fraction of each of these types that are tissue-regulated was estimated as 50–75%. It was also estimated that 86% of human genes are alternatively spliced, with at least two variants having a frequency of 15% or more.

## Many possible variants may not make functional proteins

Much is made of the number of theoretically possible splicing products from genes with a substantial number of exons, yet only those alternatives that change the number of nucleotides by an integer multiple of 3 can be added or subtracted without altering the reading frame, and many fewer can be added or removed from transmembrane proteins without seriously altering the transmembrane topology.

Many possible transcripts (up to a third in humans) that alter reading frame lead to premature termination and may be eliminated by nonsense mediated decay (NMD) (McGlincy and Smith, 2008). Although the core NMD machinery is conserved from yeast to humans, the precise mechanism for identifying premature termination codons varies. In mammals, a stop codon is recognized as premature if it is located >50–55 nucleotides upstream of an exon–exon junction. Although the generality of this rule has been questioned, it seems to be widely applicable. A microarray hybridization study across 3126 cassette-type exons in 10 mouse tissues (Pan *et al.*, 2006) found that variants containing premature termination codons were expressed at relatively constant low levels, with little evidence of tissue-specific expression.

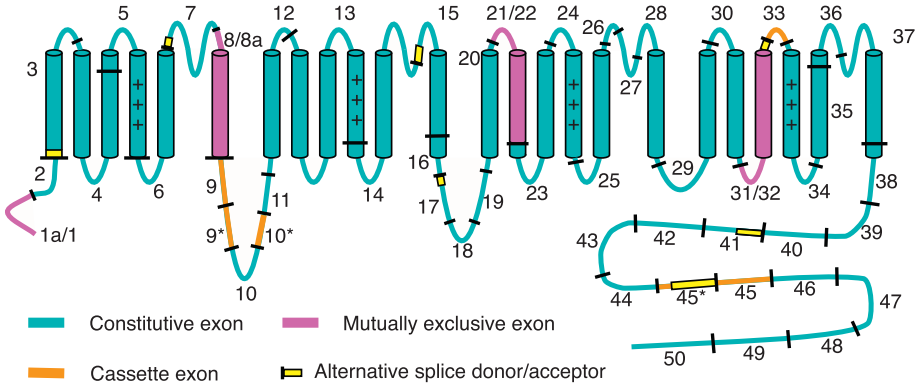
Many splice variants appear in databases such as GenBank because of publication bias; that is, any variant not already in the databases is considered novel even if it is based on a single rare cDNA. Such rare transcripts may well represent a catalogue of splicing errors.

## Some well-characterized examples

There have been a number of studies using microarray hybridization to catalogue splice variations in GPCRs (Einstein *et al.*, 2008) or in whole genomes (Castle *et al.*,

2008); however, this technique can sample only known or predicted exon junctions. Potentially more informative are high-throughput sequencing studies (Mortazavi *et al.*, 2008; Pan *et al.*, 2008; Wang *et al.*, 2008) of millions of short RNA sequences that are mapped to the genome to provide an unbiased catalogue of splice junctions in various tissues. Neither of these methods characterizes full-length cDNAs, which would allow assessment of how many of the combinatorially possible splicing variants actually occur at significant frequencies. Unfortunately, many of the genes of interest to pharmacologists, such as receptors and ion channels, are expressed at such low levels that they are not adequately sampled in these genome-wide surveys of splice variation. For example, the data of Wang *et al.* (2008) can be viewed as they map on the human genome in the UCSC genome browser (http://genome.ucsc.edu/) by displaying the Burge RNA-seq annotation track found in the Expression section. Doing so for two GPCRs that we discuss more fully below, the PACAP $PAC_1$ receptor and the 5-$HT_4$ receptor, detects a single splice junction in both cases, while for two calcium-gated ion channels, $Ca_V1.2$ and $Ca_V3.1$, there are four splice junctions defined for $Ca_V1.2$ and none for $Ca_V3.1$. For $PAC_1$ and $Ca_V3.1$ there are numerous RNA sequences, the vast majority of which map within known exons, but for both 5-$HT_4$ receptors and $Ca_V2.1$ channels, many of the sequences map within introns.

However, there have been a few studies that actually look at a significant number of full-length clones from various tissues that allow an assessment of what the major transcripts are and their tissue specificity. First, let us consider a voltage-gated calcium channel, as they have a large number of exons and are frequently presented as cases of potentially immense splicing complexity. The human $Ca_V1.2$ gene, *CACNA1C*, has 23 splice variants listed by the National Center for Biotechnology Information (NCBI) in Entrez Gene (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=775). The sites of alternative splicing were examined in heart and brain (both fetal and adult for both tissues) by PCR analysis of overlapping segments across the gene by Tang (2004), who found 19 of 55 exons to be alternatively spliced (Figure 2). There were seven cases of alternative splice donor/acceptor, three mutually exclusive exon pairs, six cassette exons (one of which also has an alternative splice donor) and one alternative first exon that leads to a different amino-terminal peptide. The alternative first exons, although mutually exclusive, are not counted in the mutually exclusive exon category. The three mutually exclusive pairs (exons 8/8a, 21/22 and 31/32) occur in transmembrane helices, but note that three of the Entrez Gene variants [1 (NM_199460.2), 2 (NM_01129827.1) and 6 (NM_01129832.1)] contain both exons 21 and 22 [in the numbering of Soldatov (1994)], and four of the Entrez Gene variants [1, 2, 5 (NM_01129831.1) and 10 (NM_01129836.1)] contain both exons 31 and 32, which would add an extra helical domain. Presumably, these insertions are structurally disruptive, and the RefSeqs are annotated as predicted to form non-functional channels. Of the seven alternative splice donor/acceptor changes, two changes, of four and 73 nucleotides, would cause frame shifts and premature termination near the beginning and middle of the protein; three would cause 2–4-amino acid insertions/deletions in loop regions; one would cause a 19-amino acid insertion/deletion in the carboxyl terminal region; and one would introduce a frame shift in the carboxyl terminal region. Four of the cassette exons are in loop regions, and two are in the carboxyl terminal region. Although this paper did not examine full length cDNAs, a subsequent paper (Tang *et al.*, 2007) did examine 92–98 full-length clones from each of three full-length cDNA libraries of $Ca_V$ 1.2 transcripts isolated from rat heart and aorta and found a total of 41 alternative transcripts. Compared with the 19 alternatively spliced exons in the human gene, there are only 11 alternatively spliced exons, with no alternatives in the carboxyl terminal. There were no exons 45, 45* or 10*; no alternative acceptors in exons 3, 7 and 15; and no alternative donors in exons 17 and 40. The exons were numbered as in humans [based on the 50 exons of Soldatov (1994), to which were added the subsequently discovered exons 1a, 8a, 9*, 10* and 45*], which facilitates




## Figure 2

Human $Ca_V1.2$ gene structure and splice variants. Diagram of alternative splicing of $Ca_V1.2$ transcripts adapted from Tang (2004). Exons are numbered based on the numbering of Soldatov (1994), with more recently discovered exons added as exons 1a, 8a, 9*, 10* and 45*, and are mapped onto the protein structure that they encode. Note that exon 8a precedes exon 8 in the genome and exon 1a precedes exon 1, which is also called exon 1b in rat. Exons 3, 7, 15, 17, 32, 41 and 45* use alternative splice acceptors or splice donors, resulting in length variations of 4, 12, 73, 9 or 12, 6, 57 and 187 nucleotides respectively.

comparison between the two papers but is not consecutive according to genome order.

Of the 11 variants in heart using exon 1a (the alternative promoter that gives a longer N-terminal peptide), two major variants accounted for 75% (50% and 25%) of the cDNAs, and the other variants ranged from 1% to 5%. Of the 19 variants in heart using the exon 1 promoter, four major variants accounted for 60% of the cDNAs (29%, 23%, 9% and 7%), and the rest ranged from 1% to 5%. Of the 20 variants found in aorta, all used the exon 1 promoter; four major variants accounted for 60% of the cDNAs (26%, 18%, 9% and 7%), four more for 5% each and the remainder for 1% to 3%. Comparing the exon 1 transcripts in heart to those in aorta shows a high degree of tissue specificity. The top four variants in heart make up 0–1% of transcripts in aorta, and the top four variants in aorta make up 1–2% of those in heart, largely due to tissue specificity in the choice of mutually exclusive exons 8 and 8a. The top eight variants in aorta all use exon 8, while four of the top six variants in heart using exon 1 and all of the variants in heart using exon 1a use exon 8a. The two major variants in heart using exon 1a differ in the choice of mutually exclusive exons 21 and 22, as is the case for the top two variants in heart using exon 1, and the two most abundant in aorta differ in the inclusion of cassette exon 33. Thus, most of the differences among the abundant variants are due to choice of mutually exclusive exons. A review by Liao *et al.* (2005) discusses exon specificity in human cardiac and smooth muscle, concentrating on the four mutually exclusive exons plus exons 9* and 45*. It is worth noting that of the 10 major rat variants described above, only two have homologues among the 23 human variants listed in Entrez Gene [17 (NM_01129843.1) and 18 (NM_000719.6)], due in large part to the fact that none of the listed human variants contain exon 1a and only four contain exon 8. Only the first two rat variants are represented by rat sequences in GenBank (AF394838 and NM_012517.2 respectively).

More recently, an additional first exon, 1c, which is contained in the majority of $Ca_V1.2$ mRNA isolated from cerebral artery myocytes, has been discovered in rat (Cheng *et al.*, 2007). Analysis of 27 full-length cDNAs from arterial smooth muscle cells resulted in 16 different transcripts (four of which contained premature terminators), of which 18 cDNAs contained exon 1c and nine contained exon 1b (Cheng *et al.*, 2009). All of the clones contained exon 8 instead of exon 8a. Half of the exon 1c-containing clones also contained exons 32 and 33 and differed in the presence or absence of exon 9* and the choice of exon 21 or 22. It is unclear whether exon 1c exists in humans. While the 5′ end of the rat cDNA AY974797 maps onto the human genome in an appropriate location, the splice donor site used in rat is not conserved in primates, mouse, dog or elephant, but there are possible splice donor sites located 6 bases 5′ of the rat donor site that could lead to an *N*-terminal peptide of MLHLILDC in humans, instead of MLCCALDCAC as in rat.

Turning to another calcium channel gene, Emerick *et al.* (2006) examined 287 full-length fetal clones and 758 full-length adult clones of human $Ca_V3.1$ from whole brain. The gene (*CACNA1G*) has 38 exons; four are cassette exons (14, 26, 34 and 35), three use alternative splice donors (25, 30 and 38) and three have alternative splice acceptors [25 (with three alternatives), 31 and 38]. The exon 38 variant splices out a 237-base internal segment of the exon and thus might properly be considered a retained short intron. These alternatives all occur in cytoplasmic loops, primarily in long loops following the four membrane domain clusters, although the exon 30–31 variants are within a short cytoplasmic loop between the fourth and fifth transmembrane helices in the fourth membrane domain. In addition, there are two alternative promoter sites and two alternative polyadenylation sites. In all, 30 distinct reading frames were observed among 1580 full length cDNAs from fetal (which included 535 from highly amplified libraries with anomalously low diversity) and adult whole brain. Ninety-two per cent of clones preserve all transmembrane domains. There are five abundant clones in the more representative libraries that differ slightly from a baseline clone, present at 3–5% in fetal and adult brain, which contains the shorter form of exon 25 resulting from splice donor choice and the long forms of exons 30, 31 and 38, as well as being marked by the absence of exons 14, 26, 34 and 35. The abundant clones include fetal clones that add exon 26 (22%), exon 14 (12%) or both 26 and 14 (38%), and adult clones that add the 3′ extension to exon 25 (33%) or the extension plus exon 14 (38%) to the baseline clone. Six different transcripts containing exon 34 and/or 35 total about 4% in both fetal and adult whole brain. The clones that use the short form of exon 31 cause a frame shift and premature termination. There are five different transcripts in fetal brain with this short exon, totalling 2%, and one other transcript in adult brain, with a frequency of 1.3%. There are two different transcripts in fetal brain using the short form of exon 38, totalling 0.7%, and two others in adult brain, totalling 1.7%. This gene has 33 different RefSeqs, of which variants 11 (NM_198382.2), 10 (NM_198386), 8 (NM_198384), 12 (NM_198387) and 9 (NM_198378) represent the five abundant clones in the order mentioned above.

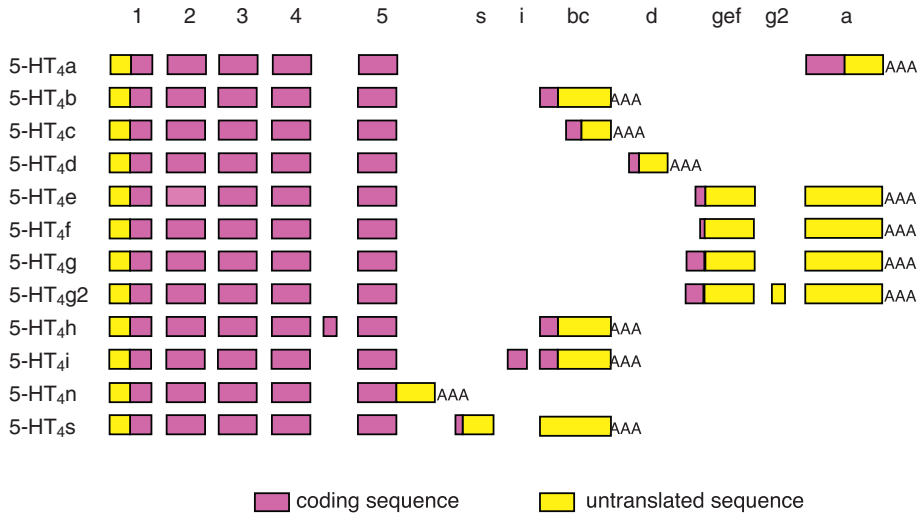
For both of these voltage-gated calcium channels, it is possible that there might be additional significant transcripts if additional tissues were examined or if specialized populations of cells were to be examined. It is reasonably clear from these examples that there are a small number of variants that predominate in a given tissue and that the major variants typically differ from one another in only one or two splice choices.

However, not all voltage-gated calcium channels have complex splicing patterns. $Ca_V3.3$ (*CACNA1I*) has 37 exons (Gomora *et al.*, 2002), of which only two are alternately spliced, exon 9 is a cassette exon, and exon 33 uses an alternative splice acceptor that deletes 13 amino acids (Mittman *et al.*, 1999). By PCR analysis with fetal brain mRNA, it appears that transcripts lacking exon 9 are approximately 10-fold more abundant than those with exon 9 and that those with the long form of exon 33 are about fivefold more abundant that those with the short form (Murbartián *et al.*, 2004).

Turning now to GPCRs, the first description of alternative splicing that resulted in functional differences in the expressed receptor was the finding of Spengler *et al.* (1993) that the rat PACAP receptor, $PAC_1$, has two cassette exons, named hip and hop, that can insert an additional sequence in the third cytoplasmic loop and that alter patterns of adenylate cyclase and phospholipase C stimulation in hip-containing variants. The hop exon actually appears in two forms differing in length by three bases due to alternative

splice acceptor usage. However, the variant lacking both hip and hop exons predominated in most adult rat tissues, with the hop-containing variant predominating in testes, olfactory bulb and adrenal gland. The hip exon was only detected at very low levels by exon-specific *in situ* hybridization. The subsequent cloning of the human cDNA (Pisegna and Wank, 1996) isolated 36 clones from a frontal cortex library and found that 33 (92%) contained neither the hip nor the hop exons [RefSeq variant 3 (NM_001118.4)] and 3 (8%) contained the hop exon in the longer hop1 form [RefSeq variant 1 (NM_01186564.1)]. They found little difference in signal transduction among the human splice variant clones, except for greater efficacy for stimulation of phospholipase C by variants containing the hop exon. Subsequently, splice variation was found within the extracellular amino-terminal domain in both mouse and human (Pantaloni *et al.*, 1996; Dautzenberg *et al.*, 1999; Lutz *et al.*, 2006), most commonly a 21-amino acid deletion resulting from the skipping of exons 5 and 6 but also including a 57-amino acid deletion created by skipping exons 4–6. The human gene has 18 exons, of which exons 8–17 encode the seven transmembrane domains. The variants with the 21-amino acid deletion were found to be the major receptor variants in several human neuroblastoma cell lines and also were more common in embryonic brain than in adult brain (Lutz *et al.*, 2006). Lutz *et al.* examined 81 full-length clones from SH-SY5Y neuroblastoma cells and found 53% with the 21-amino acid *N*-terminal deletion and no hip or hop [RefSeq variant 4 (NM_001199637.1)], 22% with the 21 amino acid deletion and hop insertion, and 12% with the 57-amino acid deletion and no hip or hop. There were a total of eight clones containing hip exon, in six (7%) of which it was coupled with a 21-amino acid deletion, while only three clones (4%) contained no *N*-terminal deletion. Cloned versions of most of the variants expressed in cell lines show modest variations in pharmacology. The most pronounced variation is greatly increased ability of VIP to activate production of cAMP in variants with the 21-amino acid deletion (Dautzenberg *et al.*, 1999; Lutz *et al.*, 2006).

A more complicated situation is presented by the 5-HT$_4$ receptor, for which there are a variety of splice variants that result in variations in the *C*-terminal cytoplasmic domain. Such variations in the *C*-terminal domain are fairly common in GPCRs and are potentially significant for differential interaction with intracellular proteins, and there are reports of such interactions (e.g. Joubert *et al.*, 2004). There are five exons containing coding sequence that generally encode the first 358 amino acids, which extend well into the *C*-terminal domain. They are followed by a variety of exons in various species that encode from 1 to 70 terminal residues (Figure 3). There is an additional exon commonly called 'h' which inserts a 14-amino acid segment in an extracellular loop at amino acid 170. The human receptor has 11 described variants: (a)–(g), (g2), (h), (i) and (n) (Bender *et al.*, 2000; Vilaró *et al.*, 2002; Brattelid *et al.*, 2004). Mice have four described variants: (a), (b), (e) and (f) (Hernandez and Janušonis, 2010). Rats have four: (a), (b), (e) and (c1) (Ray *et al.*, 2009). Pigs have 12 – (a), (b), (b2), (j)–(m) and (o)–(r) – as well as three containing the h exon insertion: (ha), (hr) and (hm) (De Maeyer *et al.*, 2008). The naming of the variants generally uses the same name for homologous transcripts across species. The (e) and (f) variants are derived from the exon encoding the (g) variation, commonly referred to as the g (or gef) exon, by use of alternative splice acceptor sites. Although this exon is well conserved, the splice acceptor giving rise to the (g) variant is not conserved in rat and mouse. The rat (c1) variant contains a sequence homologous to the human (c) but uses a different splice acceptor, the acceptor used in humans not being conserved in rat or mouse. Although the transcript has yet to be identified, the (c1) terminal peptide and splice acceptor are well conserved in the mouse genome.




## Figure 3

Human 5-HT$_4$ receptor splice variants. Constitutively spliced coding exons are numbered. Alternatively spliced coding exons are labelled h, s, i, bc, d, gef and a. The g2 exon is an exception, as it is untranslated, but the g2 transcript encodes the same isoform as the g variant. Exons and spacing between them are not to scale.

Examination of cDNA sequences in GenBank reveals two additional unrecognized splice variants. The human adrenal cDNA AK308471 encodes a variant with the *C*-terminal sequence 359-RAVRGNRNSTSSA-371, which we will call variant (s). The mouse brain cDNA BC145161 has been erroneously annotated as a frame shift error when it encodes the *C*-terminal sequence 359-RTSDYDTYLQSLGSDKVSI-377, but the exon is not well conserved in humans. In addition, the single bovine cDNA, BT020865, encodes an (hb) variant. Thus, of all the human variants, only (a), (b), (e), (f), (g) and perhaps (c) appear to be well conserved beyond primates. The exon for the (d) variant appears to be well conserved in some non-primate species such as cow and dog but not in rats and mice. Whether the other variants that are not evolutionarily well conserved are just noise in the system or important adaptations remains to be seen.

The 3′ ends of most of the human cDNAs are incomplete, based on the lack of a poly(A) signal sequence and comparison with the (a) and (b) variants from mouse and rat, but there appear to be five authentic 3′ ends for the 12 human variants. The (a), (g) and (g2) variants share a 3′ end, with the (g) and (g2) variants having one or two exons inserted before the 'a' exon and the g2 exon being entirely untranslated due to a termination codon in the g exon. The (n) variant is a read-through of the fifth coding exon (i.e. it does not recognize the splice donor) and has a unique 3′ end with an authentic poly(A) signal. The (d) variant also has a unique and authentic 3′ end. The remaining variants, (b), (c), (h), (i) and (s), all share the 'b' exon 3′ end. The (c) variant uses an alternative acceptor in the 'b' exon, the (h) variant inserts the 'h' exon between the fourth and fifth coding exons, and the (i) and (s) variants insert unique exons prior to the 'b' exon. cDNAs for the (e) and (f) variants have not been described, although (Bender *et al.*, 2000) reported different patterns of expression of the (e), (f) and (g) variants in various brain regions. However, interpretation of their results requires some care, as the design of their primers is such that the g primer would detect the (g) but not the (e) variant, but the e primer would detect both variants, with the (e) variant product being 21 nucleotides larger.

Some of the human variants have been quantitated by quantitative PCR. Medhurst *et al*. (Medhurst *et al*., 2001) used quantitative PCR to compare splice variants (a), (b), (c), (d) and (g) in caudate putamen, temporal cortex, hypothalamus, hippocampus, pituitary, heart and small intestine. The (b) variant predominated in all tissues at 86–98%. The (c) and (a) variants were second most abundant, reaching as much as 5–7% in several tissues. The (d) variant was detected only in small intestine, where it was the least abundant of the five. Lezoualc'h *et al*. (2007) did quantitative PCR on atrial tissue from atrial fibrillation patients with a similar set of variants; the results gave roughly 85% for (b) and roughly 3–6% for (g), (a) and (c) in normal sinus rhythm or chronic atrial fibrillation, but in acute fibrillation (b) and (g) became negligible and (a), although unaltered compared with normal, was roughly twice as abundant as (c). However, one should note that both studies used an antisense primer for the c exon based on a cDNA sequence that skipped a base found in other cDNAs (Bender *et al.*, 2000) and the genomic sequence. The imperfect match of the primer to the mRNAs might have impaired the detection of the (c) variant. Quantitative PCR was also used to follow the developmental expression of 5-$HT_4$ receptors in mouse telencephalon at embryonic days E13–E18 (Hernandez and Janušonis, 2010). All four variants, (a), (b), (e) and (f), were expressed at very low levels but increased fivefold to 11-fold by E17–E18, with variant (a) accounting for roughly half, while (b), (e) and (f) each accounted for 10–25%. The (a) and (b) variants were rapidly up-regulated between E14 and E15, while (e) and (f) showed a gradual, steady increase. In a comparison of numerous rat tissues (Ray *et al*., 2009) using quantitative PCR for the (a), (b) and (c1) variants, the abundances of the (a) and (b) variants were comparable in CNS samples, while (b) was several-fold more abundant than (a) in gastrointestinal tissue. The (c1) variant was insignificant in the CNS but accounted for 10% or more of the product in some gastrointestinal tissues. Thus, there is not much quantitative evidence for substantial levels of expression for anything but variant (b) in human tissues, but several variants and tissues have not been quantitated. In contrast, both rodent studies showed significant expression level (at least 10% of total receptor mRNA) for each of the tested variants in some samples.

## What defines a significant variant?

I suggest that there are a number of criteria, as shown in Table 1, that define various levels of significance. At the most basic level, stage 1, the variant must be completely defined by a full-length cDNA, minimally including the complete coding sequence. To be useful to others, the sequence should be in a public database. This is a necessary but not sufficient condition, as any sequence in the databases, even reference sequences such as the NCBI RefSeq, may encode a non-functional protein or may be very rare. Indeed, some RefSeqs are apparently created with as many exons as possible for mapping gene mutations, even when such a transcript has not been proven to exist (see the recommendations for

### Table 1

Six stages of significance for splice variants, in more or less ascending order towards fully significant

1. A full-length cDNA sequence in a public database encoding a complete coding sequence of a presumptively functional protein, for example one that has a full complement of transmembrane domains normally associated with the gene or has been demonstrated to be functional in transfected cells
2. Expression at a level of greater than 10–15% of the combined variants in at least one tissue, developmental stage, disease state or population of naturally occurring cells
3. Evolutionary conservation at the genome level across a reasonable range of species, for example human to rodent
4. Evidence of biological regulation
5. Differences in biochemical or pharmacological properties of variants when expressed in transfected cells
6. Functional differences that can be demonstrated for naturally expressed variants

coding DNA reference sequence in the document at http://www.hgvs.org/mutnomen/refseq.html#complete). All of the human 5-$HT_4$ variants except (e) and (f) meet this criterion, as do the the five most abundant variants of $Ca_V3.1$ and RefSeq variants 3 and 4 of $PAC_1$.

The second and probably most important threshold to cross on the way to significance is the demonstration of a significant level of expression in some population of naturally occurring cells. This is not difficult for simple genes, where a small number of closely spaced alternative exons, as in 5-$HT_4$ and $PAC_1$ receptors, lend themselves to PCR analysis. However, large genes with a variety of spliced exons often require the analysis of a large number of essentially full-length clones. The five most abundant variants of human $Ca_V3.1$ meet this criterion, as do several of the 5-$HT_4$ variants in rodents and the (b) variant in humans, four variants of $PAC_1$ and a variety of rat variants of $Ca_V1.2$, as discussed above.

A third factor is evolutionary conservation of the relevant exons along with their splice acceptor and splice donor sites. While conservation across a reasonable range of species creates confidence that the function of the individual exons is being conserved, it does not guarantee that all variants containing a specific exon will be conserved. However, lack of conservation does not exclude the possibility that a recent adaptive change has occurred. Examples of conserved and less well-conserved exons were discussed earlier in the case of the human 5-$HT_4$ receptor.

The fourth factor is evidence of biological regulation. A clear example is mutually exclusive exons, the existence of which is dependent on stringent regulation of splicing choice. A more common case would be differences in relative abundance of variants in different tissues or developmental stages. These changes can be quite dramatic, as discussed above for rat variants of $Ca_V1.2$, which are largely due to choices of mutually exclusive exons.

The fifth factor that strongly indicates significance is the demonstration of differences in biochemical or pharmacological properties of variants when expressed in transfected cells. Prime examples would be the $PAC_1$ variants with the 21-amino acid deletion in the amino-terminal domain, which have increased sensitivity to VIP, and the $PAC_1$ variants with the hip or hop exons, which show differences in coupling to phospholipase C. In both cases there are inconsistencies concerning the magnitude of these effects in different reports, which one suspects may be dependent on properties of the cell line used for expression and levels of receptor expression. Cells used for transfection typically do not naturally express the gene being tested, so that interpretation of the results can be simplified. However, one must be concerned as to whether all the proteins that normally interact with the gene product are expressed at appropriate levels.

The highest level of significance would be established by demonstration of functional differences for naturally expressed variants. For genes with more than one or two alternatively spliced exons, this may be more of an aspirational goal than an easily achievable one, as it is unusual to find cells that express only a single variant. However, naturally expressed variants, including disease-causing mutations, can provide evidence of the functional significance of individual spliced exons. For example, Timothy syndrome is due to amino-acid-changing mutations in exons 8 and 8a of the $Ca_V1.2$ gene that result in failure of channel inactivation and different phenotypes, depending on which exon is affected, due to the different patterns of expression of the mutually exclusive exons (Splawski *et al.*, 2005). Furthermore, as we gain more knowledge of the splicing regulatory code (Wang and Burge, 2008) it should be possible to manipulate the splicing pattern of naturally expressed variants. For the $Ca_V1.2$ gene, it has been shown that knockdown of the polypyrimidine tract-binding protein, PTB, in N2a mouse neuroblastoma cells using a short hairpin RNA dramatically reduced the inclusion of exon 8 from 82% to 7–10% and proportionally increased the inclusion of exon 8a (Tang *et al.*, 2011), although it is not known what effect this treatment has on the other alternatively spliced exons. A variety of more traditional drugs have been shown to correct the splicing defect, skipping of exon 20 in the *IKBKAP* gene, in familial dysautonomia. Most recently, digoxin has been shown to correct the defect by suppressing the expression of splicing factor SRSF3 (Liu *et al.*, 2013) as well as modulating the splicing of 3 of 35 other genes that were tested.

## Nomenclature usage

The primary reasons for a standard nomenclature are to facilitate literature searches and facilitate the comparison of results from different papers. Once the standardized nomenclature is used in a paper, a more idiosyncratic name could be used in cases such as that of the flip and flop transcripts of the GluA2 receptor where they might be more convenient or immediately recognizable. However, splicing alternatives are generally complex enough that a paper must precisely define the transcript by citing an appropriate cDNA sequence accession number, as well as provide text and/or diagrams that clearly indicate what the differences are between the variants being discussed. The punctuation of an abbreviated form or symbol such as 'gene name_v1' is clearly not as important as indicating in the abstract or keywords that, for example, splice variants 1, 2, 4 and 5 are described.

Another problem in describing the variants in terms of exon structure (as opposed to the precise sequence from a cDNA accession) is that exons are not uniformly numbered. Even disregarding the issue of renumbering exons when new ones are discovered, the presence of more than one first exon due to alternative promoters leads to several numbering possibilities. In most cases the first exons are numbered 1a, 1b and so on in chromosomal order, with the one most distant from the common downstream exons being named 1a. However, in other cases people have just numbered the exons 1, 2, 3 and so on in chromosomal order, where 1–3 might be alternative first exons. There is even a suggestion in a paper proposing nomenclature for P450 genes that proposes naming the first exons 1a and 1b in inverse chromosomal order on the supposition that the one closest to the common second exon is more often the more common transcript (Nelson *et al.*, 2004). All these variations require that there be a clear exon structure diagram in each paper, preferably with a clear indication of the location of important protein features such as transmembrane domains.

## Official nomenclatures

The question of what nomenclature to use and whether there is an official recommendation that should be followed is still an open question. I am unaware of any formal recommendation from any official organization with a broader mandate than NC-IUPHAR. The Human Genome Nomenclature Committee does not have an official recommendation but suggests that when referring to splice variants, the gene symbol can be followed by an underscore and the lower case letter 'v', then a consecutive number to denote which variant is which, for example G6PD_v1 (http://www.genenames.org/guidelines.html). The Human Genome Variation Society (HGVS) recommends the same format for transcript variants, but restricted to variants within the coding sequence, and uses a similar format of _i1 for the protein isoform (i.e. protein variant) encoded by transcript variant _v1 (http://www.hgvs.org/mutnomen/standards.html#traiso).

Turning to Entrez Gene to see if NCBI uses any standard nomenclature, one finds that the answer is no. They use a variety of nomenclatures, perhaps just what has been used in the primary literature. However, unlike the primary GenBank entries, which NCBI cannot change without submitter approval, the RefSeqs and the Entrez Gene entries represent NCBI syntheses of the primary entries where NCBI could choose a uniform nomenclature. Even here there is a wide range of usage. The convention of protein isoforms being encoded by mRNA splice variants is followed, but the isoforms or variants are either numbered or lettered with either capitals or lowercase, and either with or without descriptions of what is different about the isoforms or variants. In addition, mRNA variants are listed that encode the same protein, so that there is no difference in the protein name. For example, the human $M_2$ muscarinic receptor, *CHRM2*, has eight listed variants that all differ in splicing of three 5′ UTR exons and two different lengths (89 or 341 nt) of the first exon, possibly resulting from different but closely spaced promoters. As all the protein products are the same, it is not clear that pharmacologists would be interested in these variants, but if one were, a nomenclature based on proteins as in the HGVS suggestion would provide no distinction.

The nomenclature problem is compounded by the fact that other major databases may use different nomenclatures. For example, the Vertebrate Genome Annotation database (Vega) (http://vega.sanger.ac.uk/index.html) provides manual annotation of vertebrate genomes but gives different lists of splice variants from those in the NCBI databases. For example, for the human *CACNA1C* gene, which encodes $Ca_V1.2$, Vega lists 31 splice variants, nine of which do not encode full-length proteins (greater than 2100 amino acids). Nineteen of the variants are contained in the NCBI Gene list, but they have names of the form CACNA1C-001 where the numbering is entirely different. The Consensus Coding Sequence project (http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi) attempts to reach a consensus among four major genome annotation sites for the human and mouse genomes (Pruitt *et al.*, 2009; Harte *et al.*, 2012). It currently contains the 19 variants for the human *CACNA1C* gene that are common to the Vega and NCBI Gene lists with ID numbers 44787–44801 and 53733–53736, but with essentially no annotation, so that only by looking at the corresponding NCBI RefSeqs does one realize that CCDS IDs 44788.1, 44789.1, 44790.1 and 44797.1 are presumably non-functional, as they contain both members of a pair of mutually exclusive exons.

## Is an official NC-IUPHAR nomenclature needed?

It is clear that some splice variants have pharmacologically significant effects, altering signalling properties of GPCRs and voltage-gated ion channels and in some cases altering the binding properties of drugs. In addition, various studies indicate changes in splicing patterns in disease states. For example, in atherosclerosis, $Ca_V1.2$ transcripts in human vascular smooth muscle cells completely switch from exon 21 to exon 22 and lose the 5′ extension of exon 41 (commonly referred to as exon 41A) while simplifying the splicing pattern for exons 31–34 (Tiwari *et al.*, 2006). These are precisely the sorts of variants that NC_IUPHAR originally chose to focus on. If the nucleotide and protein sequence databases were reasonably complete in their lists of transcripts, it would be simplest for the IUPHAR database to annotate only the significant variants using our criteria to score their significance and referring to the sequence databases for the appropriate sequence identifier.

Many studies, especially of larger, more complicated genes such as $Ca_V1.2$, use PCR scanning methods to characterize variants, even if there are a large number of individual cDNA clones that could be sequenced. As such studies lack directly determined cDNA sequences, they are not eligible for normal submission to databases such as GenBank. However, in many cases the sequences could be inferred from the data and are essential to any clearly defined nomenclature. Rather than having IUPHAR attempt to create complete lists of all observed splice variants, I suggest that the most practical approach is for IUPHAR to facilitate the incorporation of these sequences, at least those expressed at significant levels, into the sequence databases by interacting with the primary sequence annotation groups at the databases. I can say from personal experience that the RefSeq group at NCBI are generally quite receptive to suggestions for improving their annotations.

## Acknowledgements

This work was supported in part by the Intramural Research Program of the National Institutes of Health (National Institute of Mental Health).

## Conflict of interest

None.

## References


Bender E, Pindon A, van Oers I, Zhang YB, Gommeren W, Verhasselt P *et al*. (2000). Structure of the human serotonin 5-$HT_4$ receptor gene and cloning of a novel 5-$HT_4$ splice variant. J Neurochem 74: 478–489.

Brattelid T, Kvingedal AM, Krobert KA, Andressen KW, Bach T, Hystad ME *et al*. (2004). Cloning, pharmacological characterisation and tissue distribution of a novel 5-$HT_4$ receptor splice variant, 5-$HT_4$(i). Naunyn Schmiedebergs Arch Pharmacol 369: 616–628.

Castle J, Zhang C, Shah J, Kulkarni A, Kalsotra A, Cooper T *et al*. (2008). Expression of 24,426 human alternative splicing events and predicted *cis* regulation in 48 tissues and cell lines. Nat Genet 40: 1416–1425.

Cheng X, Liu J, Asuncion-Chin M, Blaskova E, Bannister JP, Dopico AM *et al*. (2007). A novel $Ca_V1.2$ *N* terminus expressed in smooth muscle cells of resistance size arteries modifies channel regulation by auxiliary subunits. J Biol Chem 282: 29211–29221.

Cheng X, Pachuau J, Blaskova E, Asuncion-Chin M, Liu J, Dopico AM *et al*. (2009). Alternative splicing of $Ca_V1.2$ channel exons in smooth muscle cells of resistance-size arteries generates currents with unique electrophysiological properties. Am J Physiol Heart Circ Physiol 297: H680–H688.

Dautzenberg FM, Mevenkamp G, Wille S, Hauger RL (1999). *N*-terminal splice variants of the type I PACAP receptor: isolation, characterization and ligand binding/selectivity determinants. J Neuroendocrinol 11: 941–949.

De Maeyer JH, Aerssens J, Verhasselt P, Lefebvre RA (2008). Alternative splicing and exon duplication generates 10 unique porcine 5-HT4 receptor splice variants including a functional homofusion variant. Physiol Genomics 34: 22–33.

Einstein R, Jordan H, Zhou W, Brenner M, Moses EG, Liggett SB (2008). Alternative splicing of the G protein-coupled receptor superfamily in human airway smooth muscle diversifies the complement of receptors. Proc Natl Acad Sci U S A 105: 5230–5235.

Emerick MC, Stein R, Kunze R, McNulty MM, Regan MR, Hanck DA *et al*. (2006). Profiling the array of $Ca_v3.1$ variants from the human T-type calcium channel gene CACNA1G: alternative structures, developmental expression, and biophysical variations. Proteins 64: 320–342.

Gomora JC, Murbartián J, Arias JM, Lee J-H, Perez-Reyes E (2002). Cloning and expression of the human T-type channel $Ca_v3.3$: insights into prepulse facilitation. Biophys J 83: 229–241.

Harte RA, Farrell CM, Loveland JE, Suner M-M, Wilming L, Aken B *et al*. (2012). Tracking and coordinating an international curation effort for the CCDS Project. Database 2012: bas008.

Hernandez MC, Janušonis S (2010). Quantitative mRNA analysis of serotonin 5-HT4 and adrenergic β2 receptors in the mouse embryonic telencephalon. Dev Neurosci 32: 278–287.

Joubert L, Hanson B, Barthet G, Sebben M, Claeysen S, Hong W *et al*. (2004). New sorting nexin (SNX27) and NHERF specifically interact with the 5-$HT_{4a}$ receptor splice variant: roles in receptor targeting. J Cell Sci 117: 5367–5379.

Lezoualc'h F, Steplewski K, Sartiani L, Mugelli A, Fischmeister R, Bril A (2007). Quantitative mRNA analysis of serotonin 5-$HT_4$ receptor isoforms, calcium handling proteins and ion channels in human atrial fibrillation. Biochem Biophys Res Commun 357: 218–224.

Liao P, Yong TF, Liang MC, Yue DT, Soong TW (2005). Splicing for alternative structures of $Ca_V1.2$ Ca2+ channels in cardiac and smooth muscles. Cardiovasc Res 68: 197–203.

Liu B, Anderson SL, Qiu J, Rubin BY (2013). Cardiac glycosides correct aberrant splicing of IKBKAP-encoded mRNA in familial dysautonomia derived cells by suppressing expression of SRSF3. FEBS J 280: 3632–3646.

Lutz EM, Ronaldson E, Shaw P, Johnson MS, Holland PJ, Mitchell R (2006). Characterization of novel splice variants of the $PAC_1$ receptor in human neuroblastoma cells: consequences for signaling by VIP and PACAP. Mol Cell Neurosci 31: 193–209.

McGlincy NJ, Smith CWJ (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem Sci 33: 385–393.

Medhurst AD, Lezoualc'h F, Fischmeister R, Middlemiss DN, Sanger GJ (2001). Quantitative mRNA analysis of five *C*-terminal splice variants of the human 5-$HT_4$ receptor in the central nervous system by TaqMan real time RT-PCR. Brain Res Mol Brain Res 90: 125–134.

Mittman S, Guo J, Emerick MC, Agnew WS (1999). Structure and alternative splicing of the gene encoding α1I, a human brain T calcium channel α1 subunit. Neurosci Lett 269: 121–124.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621–628.

Murbartián J, Arias JM, Perez-Reyes E (2004). Functional impact of alternative splicing of human T-type $Ca_V3.3$ calcium channels. J Neurophysiol 92: 3399–3407.

Nelson DR, Zeldin DC, Hoffman SMG, Maltais LJ, Wain HM, Nebert DW (2004). Comparison of cytochrome P450 (CYP) genes from the mouse and human genomes, including nomenclature recommendations for genes, pseudogenes and alternative-splice variants. Pharmacogenetics 14: 1–18.

Pan Q, Saltzman AL, Kim YK, Misquitta C, Shai O, Maquat LE *et al*. (2006). Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. Genes Dev 20: 153–158.

Pan Q, Shai O, Lee L, Frey B, Blencowe B (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat Genet 40: 1413–1415.

Pantaloni C, Brabet P, Bilanges B, Dumuis A, Houssami S, Spengler D *et al*. (1996). Alternative splicing in the *N*-terminal extracellular domain of the pituitary adenylate cyclase-activating polypeptide (PACAP) receptor modulates receptor selectivity and relative potencies of PACAP-27 and PACAP-38 in phospholipase C activation. J Biol Chem 271: 22146–22151.

Pisegna JR, Wank SA (1996). Cloning and characterization of the signal transduction of four splice variants of the human pituitary adenylate cyclase activating polypeptide receptor. Evidence for dual coupling to adenylate cyclase and phospholipase C. J Biol Chem 271: 17267–17274.

Pruitt KD, Harrow J, Harte RA, Wallin C, Diekhans M, Maglott DR *et al*. (2009). The consensus coding sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. Genome Res 19: 1316–1323.

Ray AM, Kelsell RE, Houp JA, Kelly FM, Medhurst AD, Cox HM *et al*. (2009). Identification of a novel 5-$HT_4$ receptor splice variant (r5-$HT_{4c1}$) and preliminary characterisation of specific 5-$HT_{4a}$ and 5-$HT_{4b}$ receptor antibodies. Eur J Pharmacol 604: 1–11.

Soldatov NM (1994). Genomic structure of human L-type $Ca^{2+}$ channel. Genomics 22: 77–87.

Spengler D, Waeber C, Pantaloni C, Holsboer F, Bockaert J, Seeburg PH *et al*. (1993). Differential signal transduction by five splice variants of the PACAP receptor. Nature 365: 170–175.

Splawski I, Timothy KW, Decher N, Kumar P, Sachse FB, Beggs AH *et al*. (2005). Severe arrhythmia disorder caused by cardiac L-type calcium channel mutations. Proc Natl Acad Sci U S A 102: 8089–8096.

Tang ZZ (2004). Transcript scanning reveals novel and extensive splice variations in human L-type voltage-gated calcium channel, $Ca_V1.2$ 1 subunit. J Biol Chem 279: 44335–44343.

Tang ZZ, Hong X, Wang J, Soong TW (2007). Signature combinatorial splicing profiles of rat cardiac- and smooth-muscle $Ca_V1.2$ channels with distinct biophysical properties. Cell Calcium 41: 417–428.

Tang ZZ, Sharma S, Zheng S, Chawla G, Nikolic J, Black DL (2011). Regulation of the mutually exclusive exons 8a and 8 in the $Ca_V1.2$ calcium channel transcript by polypyrimidine tract-binding protein. J Biol Chem 286: 10007–10016.

Tiwari S, Zhang Y, Heller J, Abernethy DR, Soldatov NM (2006). Atherosclerosis-related molecular alteration of the human $Ca_V1.2$ calcium channel α1C subunit. Proc Natl Acad Sci U S A 103: 17024–17029.

Vanhoutte PM, Humphrey PPA, Spedding M (1996). X. International Union of Pharmacology recommendations for nomenclature of new receptor subtypes. Pharmacol Rev 48: 1–2.

Vilaró MT, Doménech T, Palacios JM, Mengod G (2002). Cloning and characterization of a novel human 5-$HT_4$ receptor variant that lacks the alternatively spliced carboxy terminal exon. RT-PCR distribution in human brain and periphery of multiple 5-$HT_4$ receptor variants. Neuropharmacology 42: 60–73.

Wang E, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C *et al*. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature 456: 470–476.

Wang Z, Burge CB (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14: 802–813.