

## MINIREVIEW

# AT-AC Pre-mRNA Splicing Mechanisms and Conservation of Minor Introns in Voltage-Gated Ion Channel Genes

QIANG WU† AND ADRIAN R. KRAINER\*

*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724*

Shortly after the discovery of split genes in 1977, a conserved sequence feature at both ends of cellular and viral introns was recognized, i.e., the presence of GT at the 5' splice site and AG at the 3' splice site, giving rise to the so-called GT-AG rule (8). This rule holds in most cases, but exceptions have been found. For example, GC is occasionally found at the 5' end of certain introns (Table 1) (38). GC-AG introns are processed by the same splicing pathway as conventional GT-AG introns (3). It had long been assumed that removal of all introns from eukaryotic pre-mRNAs took place by the same splicing pathway, until recent developments demonstrated the existence of a second pre-mRNA splicing pathway.

Intron 6 of the gene encoding human P120, a proliferating cell nucleolar antigen, and intron 7 of the gene encoding human CMP, a cartilage matrix protein (matrilin 1), were the first reported examples of introns with AT and AC at the intron ends, instead of GT and AG (38). Intron 6 of the gene encoding Rep-3, a DNA repair protein, and intron 2 of the gene encoding Prospero, a *Drosophila melanogaster* homeodomain transcription factor, also have AT-AC ends (28). In addition to their distinctive dinucleotide ends, these and other AT-AC introns have highly conserved 5'-splice-site and presumptive branch site 8-nucleotide sequence elements that are not present in the major class of introns, ATATCCTY and TCCT TRAY, respectively. On the basis of these sequence features, it was proposed that the minor U11 and U12 snRNAs, which have regions of complementarity to these elements, are required for splicing of AT-AC introns (28). Important aspects of this prediction were soon verified experimentally (29, 100), and two additional minor snRNAs involved in the novel pre-mRNA splicing pathway were discovered (101).

The AT-AC splicing pathway was originally named after the distinctive sequences of the intron ends (100). It was later found that a few introns with AT-AC ends are processed by the major pathway (120) and, conversely, that some introns with GT-AG ends are spliced by the minor pathway (18). Therefore, the name AT-AC no longer reflects the dinucleotide intron ends per se, but rather it refers to the minor pathway itself. An alternative designation for the two pathways—U2 dependent and U12 dependent—reflects their observed or expected requirements for one of the four snRNAs specific to each pathway (86).

### DISTRIBUTION OF AT-AC INTRONS

AT-AC introns exist in a variety of organisms ranging from *Arabidopsis thaliana* to *Drosophila*, *Xenopus laevis*, and mammals (86, 102, 120). However, AT-AC introns are absent in budding yeast. There are no obvious structural relationships or common expression patterns among the genes or gene families that contain AT-AC introns. The lengths of known AT-AC introns range from less than 100 bases to more than 3,000 bases (120). The position of AT-AC introns is not conserved among unrelated genes. However, AT-AC introns are conserved phylogenetically and within gene families. For example, the last intron of the CMP gene is an AT-AC intron that is conserved in human, mouse, and chicken genes (4). Two genes with an AT-AC intron, mouse *Rep-3* (a homologue of bacterial MutS) and *XPG* (a gene defective in xeroderma pigmentosum), are thought to be involved in DNA repair (120). The significance of the presence of AT-AC introns in DNA repair genes may be borne out as more sequences of DNA repair genes are determined. To date, three gene families have been noted to have AT-AC introns: the E2F transcription factor genes, the voltage-gated sodium and calcium channel  $\alpha$  subunit genes (reference 120 and references therein), and the cartilage matrix protein (matrilin) family genes (4, 109). The voltage-gated ion channel genes are especially interesting because they have multiple minor introns; the conservation of minor introns in these genes is therefore reviewed in detail below.

### IN VITRO AND IN VIVO SYSTEMS TO STUDY AT-AC PRE-mRNA SPLICING

Two different approaches have been used to study the mechanisms of AT-AC pre-mRNA splicing. The in vivo approach consists of transfecting wild-type or mutant minigenes containing an AT-AC intron and analyzing the splicing patterns after transient expression. Suppressor AT-AC snRNAs can be co-transfected with AT-AC intron mutants to test the functional significance of proposed base-pairing interactions between conserved intron elements and complementary regions in the snRNAs (29, 37, 45), as originally done for the conventional pathway (129).

The development of in vitro systems for AT-AC pre-mRNA splicing made it possible to begin to study the biochemical mechanisms of the reaction. To date, in vitro splicing conditions have been established for processing the AT-AC introns of pre-mRNAs from two different genes: the gene encoding P120 and that encoding SCN4A, the voltage-gated skeletal muscle sodium channel  $\alpha$  subunit (100, 119). In vitro splicing of both P120 and SCN4A pre-mRNAs in HeLa cell nuclear extract results in generation of a lariat intermediate and release of the intron as a lariat. Therefore, AT-AC splicing occurs in two steps involving *trans*-esterification reactions similar to

\* Corresponding author. Mailing address: Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724. Phone: (516) 367-8417. Fax: (516) 367-8453. E-mail: krainer@cshl.org.

† Present address: Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138.

TABLE 1. Compilation of GC-AG introns<sup>a</sup>

Gene product	Intron no.	5' splice site	Size (nt)
SCA2 <sup>c</sup>	9	AGgcaagt	6,600
Protein Z <sup>d</sup>	2	CTgcaagt	NA <sup>b</sup>
AGL <sup>e</sup>	17	AGgcaaga	400
LAMA4 <sup>f</sup>	7	AGgcatgg	NA
s-Laminin <sup>g</sup>	1	AGgcagag	86
HAL <sup>h</sup>	20	AGgcaagc	2,000
MNK <sup>i</sup>	9	AGgcaagt	1,200
FAA <sup>j</sup>	16	AGgcaagc	NA
FAA	37	AGgcaaga	NA
FAA	41	AGgcaagt	NA
CACNL1A3 <sup>k</sup>	32	AGgcacgc	1,000
RYR1 <sup>l</sup>	57	AGgcacgc	596
FE65 <sup>m</sup>	10	AGgcacgg	144
XPG <sup>n</sup>	3	AGgcaaga	NA
Ftnb <sup>o</sup>	13	AGgcaaga	3,100
Reln <sup>p</sup>	30	AGgcaagt	1,000
TFIIS.oA <sup>q</sup>	8	AGgcaagt	NA
TGG1 <sup>r</sup>	1	AGgcatgt	NA

<sup>a</sup> 5'-splice-site sequences are shown in lowercase letters, with the GC dinucleotide in boldface type, and the last two exon nucleotides are shown in uppercase letters. Six of these genes (those encoding HAL, MNK, CACNL1A3, XPG, Reln, and TFIIS.oA) also have AT-AC and/or U12-type GT-AG introns (Table 3) (120). Although nonconsensus introns are very rare, they are unusually prevalent in these genes, especially those encoding HAL, XPG, FAA, and ion channels. Twenty-six GC-AG introns described in a previous compilation (38) are not shown. Some intron sizes are only approximate.

<sup>b</sup> NA, not available.

<sup>c</sup> Human spinocerebellar ataxia type 2 protein (83).

<sup>d</sup> Human vitamin K-dependent plasma glycoprotein (22).

<sup>e</sup> Human glycogen debranching enzyme (5).

<sup>f</sup> Human laminin  $\alpha 4$  (79).

<sup>g</sup> Human laminin  $\beta 2$  (20).

<sup>h</sup> Human histidase (95).

<sup>i</sup> Human Menkes disease protein (ATP7A) (17, 107).

<sup>j</sup> Human Fanconi anemia group A (FAA) protein, whose gene has three GC-AG introns (36).

<sup>k</sup> Human skeletal muscle voltage-gated calcium channel  $\alpha$  subunit (33).

<sup>l</sup> Human skeletal muscle ryanodine receptor (75).

<sup>m</sup> Human protein that interacts with  $\beta$ -amyloid precursor protein (34).

<sup>n</sup> Mouse DNA excision repair protein, defective in xeroderma pigmentosum (54).

<sup>o</sup> Mouse fertilin beta (15).

<sup>p</sup> Mouse reelin (81).

<sup>q</sup> *Xenopus* transcription elongation factor (76).

<sup>r</sup> *A. thaliana* myosinase; the intron is conserved in other plant homologues (122, 123).

those of the major splicing pathway. The results are consistent in the two AT-AC in vitro splicing systems, but there is one major difference between them. Inactivation of U1 or U2 snRNAs in the nuclear extract is required to detect P120 AT-AC splicing (100; reviewed in reference 102). This observation led to the suggestion that U2 may compete with U12, which is about a hundred times less abundant (64), for binding to the AT-AC branch site (47). However, this is unlikely to be the case for SCN4A AT-AC splicing in vitro, which does not require inactivation of the major splicing pathway (119). In fact, SCN4A AT-AC splicing and cryptic splicing via conventional splice sites occur in the same reaction.

#### FOUR MINOR snRNAs ARE REQUIRED FOR AT-AC PRE-mRNA SPLICING

AT-AC introns have unique and highly conserved 5'-splice-site and branch site elements, which are recognized by a unique set of minor snRNAs, U11, U12, U4atac, and U6atac. These snRNAs lack extensive sequence homology to the major snRNAs, but they appear to have related secondary structures, and more importantly, they play analogous roles in splice site

recognition and perhaps in splicing catalysis. Despite the remarkable parallels, there are some significant differences between major and minor snRNAs. For example, U11 and U12 form a stable di-snRNP particle (111) that probably enters the AT-AC spliceosome as a single entity, whereas U1 and U2 are discrete snRNP mono-particles. Although the major U5 snRNA appears to be involved in both splicing pathways (100), this snRNA assembles onto the major spliceosome as part of a U4/U6 · U5 tri-snRNP particle, whereas an analogous U4atac/U6atac · U5 tri-snRNP particle has not yet been described. Notwithstanding the fact that the AT-AC and major spliceosomes have different snRNA constituents, the catalytic core of the AT-AC spliceosome is thought to resemble that of the major spliceosome (101; reviewed in references 9, 69, 70, and 102).

**U11 and U12 snRNAs.** Human U11 and U12 are rare snRNAs that have Sm antigen binding sites but exhibit no sequence homology to other snRNAs (64). U11 and U12 snRNP particles presumably contain all the Sm core proteins. U12 interacts with a fraction of the more abundant U11 to form a di-snRNP complex (111). A 65-kDa protein of the U11/U12 complex, identified by virtue of its reactivity with a scleroderma patient antiserum has been described previously, although its sequence is not known (25). The predicted secondary structures of U11 and U12 snRNAs are similar to those of U1 and U2, respectively (64, 126). U11 and U12 localize in the nucleoplasm and are concentrated in coiled bodies and nuclear speckles, but they are excluded from nucleoli (58). This distribution is very similar to that of the major spliceosomal snRNAs U1 and U2 (reviewed in reference 89). U12 orthologues have been cloned from mouse, chicken, and frog (103, 126), species in which AT-AC introns are known to exist (reviewed in references 86, 102, and 120).

The role of U11 and U12 snRNAs in AT-AC splicing has been firmly established (29, 45, 100, 119, 125; reviewed in references 66, 70, and 102). U11 snRNA is present in in vitro-assembled P120 AT-AC spliceosomes (100) and can be cross-linked to the P120 AT-AC 5' splice site (125). U11 also interacts with the P120 AT-AC 5' splice site in vivo through base pairing (45). U12 snRNA is essential for AT-AC splicing in vitro and in vivo (29, 100, 119). U12 functions in the AT-AC splicing pathway by base pairing with the highly conserved branch site sequences (29, 100). Thus, U11 is analogous to U1 snRNA in the major pathway, whereas U12 is analogous to U2.

U11 and U12 snRNPs were also shown to be part of a negative regulator of splicing complex that inhibits splicing of Rous sarcoma virus pre-mRNA via the major pathway (26). This finding suggests a regulatory role for the U11/U12 di-snRNP particle in addition to its general role in AT-AC pre-mRNA splicing.

**U4atac and U6atac snRNAs.** U4 and U6 snRNAs are not required for AT-AC splicing. U6 is highly conserved between yeasts and mammals and is thought to function at the catalytic core of the major spliceosome. This critical role of U6 in the conventional pathway suggested the existence of an analogous molecule for the AT-AC pathway (100). The spliceosomal U6 snRNA has a  $\gamma$ -monomethyl guanosine triphosphate (meGTP) cap structure. Several low-abundance snRNAs with a meGTP cap structure but otherwise structurally distinct from U6 were identified by immunoprecipitation with antibody specific for the meGTP cap structure (27). Using the same antibody, two novel minor snRNAs termed U4atac and U6atac were identified in affinity-purified AT-AC spliceosomes assembled in vitro on P120 pre-mRNA (101). U4atac has a trimethylguanosine cap but coprecipitated because of its tight association with U6atac, analogous to the interaction between U4 and U6. U4atac, like U4, is an Sm snRNA. U6atac, like U6, has a

meGTP cap structure and terminates with a stretch of U residues, suggesting its transcription by RNA polymerase III. Both U4atac and U6atac were shown to be essential for *in vitro* splicing of the P120 AT-AC intron and subsequently also for the SCN4A AT-AC intron (101, 120).

The predicted secondary structure of the U4atac/U6atac snRNA complex is strikingly similar to that of the U4/U6 snRNA complex, although U4atac and U6atac exhibit only about 40% overall sequence homology to U4 and U6, respectively (101). In the central region of U6atac, the homology to U6 is much greater, about 80%, and this region includes sites that can be cross-linked to U12 with psoralen. U6atac, which has an AAGGAGA box at the region corresponding to the ACAGA box of U6, base pairs with the AT-AC 5' splice site, replacing U11 (37, 125). The U6atac-U12 helix identified by psoralen cross-linking is analogous to a U6-U2 helix thought to be part of the major spliceosome active site, which strongly suggests that U6atac likewise functions at the catalytic core of the AT-AC spliceosome (101). U4atac appears to act as a chaperone for U6atac, similar to the role of U4 vis-à-vis U6 (101). U5 is apparently required for splicing of both AT-AC and GT-AG introns (100, 120). Whether U5 base pairs with AT-AC exon borders, as has been shown for conventional exon sequences (reviewed in reference 68), remains to be determined.

#### AT-AC SPLICE SITE RECOGNITION

The mechanisms by which conventional splice sites are selected with high fidelity in metazoan pre-mRNAs remain largely unknown (reviewed in reference 7). The short, degenerate splice site and branch site elements are clearly required for proper splice site recognition, but they are not sufficient (93). The arrangement, spacing, and sequence context of the splice sites probably also contribute to accurate splice site selection.

In the case of AT-AC introns, even though the 5'-splice-site and branch site sequences are highly conserved 8-nucleotide elements, they probably lack sufficient information to specify AT-AC splice site recognition. Genomic sequences contain many pairs of sequences that match the AT-AC 5'-splice-site and branch site elements, but probably only a few of these are authentic AT-AC intron elements. Therefore, the sequence complementarity between the minor snRNAs and the AT-AC 5'-splice-site and branch site sequences is not sufficient to accurately select AT-AC splice sites. As with conventional introns, additional mechanisms and/or auxiliary signals for AT-AC splicing may help identify the authentic splice sites. For example, the arrangement, spacing, and sequence context of AT-AC splice sites within the entire pre-mRNA, as well as the presence of intronic and exonic elements, are likely to contribute to accurate splice site selection.

**Exon definition interactions between the AT-AC and major spliceosomes.** Most vertebrate genes have multiple introns, which are usually very large, whereas the exons are relatively small (30, 94). Sequences within large introns can match the degenerate splice site consensus elements, but they are not recognized by the spliceosome, at least in the presence of the wild-type splice sites. The exon definition model (reviewed in reference 6) proposes that in pre-mRNAs with large introns, the splicing machinery initially recognizes a pair of splice sites around a short exon and assembles on the exon; subsequently, neighboring exons are juxtaposed (80). Thus, the sequences within large introns need not be recognized. In lower eukaryotes, or in the case of small introns, a pair of splice sites at the ends of the short intron is directly recognized by the spliceosome, by an intron definition mechanism (97). Exon definition and intron definition are probably not mutually exclusive.

Both types of interactions may occur simultaneously to facilitate the recognition of multiple splice sites.

The exon definition model predicts that in genes with multiple introns, exons and their flanking introns cannot both be large. Indeed, large introns are usually flanked by small exons and large exons are usually flanked by small introns (94, 128). Statistical analyses of the length of vertebrate internal regulated exons and primate internal exons showed that very large exons and very small exons are very rare (6, 91). Several lines of experimental evidence have given strong support to the exon definition model. First, mutation of a downstream 5' splice site inhibits splicing of the upstream intron (97). Second, strengthening the downstream 5' splice site increases the splicing efficiency of the upstream intron (49). Finally, joining a 5' splice site to the end of the downstream exon increases the splicing efficiency of the upstream intron *in vitro* (48). Numerous cases of splice site mutations in vertebrate genes cause exon skipping rather than intron retention (46, 67). These phenomena are consistent with the exon definition model, although exon skipping can also be explained by *cis*-competition between splice sites or, in some cases, by instability of the retained-intron RNAs because of blocked mRNA export and/or nonsense-mediated mRNA decay. Evidence for the recognition of terminal exons also supports the exon definition model. Components that bind to the 5' cap cooperate with the splicing machinery to facilitate the recognition of the first exon (52). Definition of the last exon involves cross talk between the splicing and polyadenylation machineries (71).

After exon definition, the correct exons must be ligated. The mechanisms responsible for the correct juxtaposition of exons are poorly understood. Members of the SR protein family, a group of essential pre-mRNA splicing factors with characteristic arginine-serine C-terminal repeats (RS domain) and one or two N-terminal RNA-recognition motifs, may play a bridging role for exon juxtaposition, since they can bind to exon sequences and also can interact with themselves (14, 92, 116). SR proteins can interact with U1-70K, a U1 snRNP-associated protein, and U2AF, a U2 auxiliary factor (116). It is thought that SR proteins can bridge partially assembled spliceosomes on neighboring exons through U1-70K and U2AF. Splicing factor 1 (SF1) may also play a bridging role for exon juxtaposition. SF1 interacts with both U2- and U1-associated factors (1, 78).

Because AT-AC introns always coexist with multiple major introns, the question arises whether exon definition interactions take place between the two different classes of introns. Indeed, *in vitro* splicing of the SCN4A AT-AC intron 2 is strongly stimulated when exon 3 is followed by the conventional 5' splice site of intron 3. More importantly, the stimulatory effect is dependent on intact U1 snRNP (119, 121). Therefore, U1 bound at the downstream 5' splice site interacts with the upstream AT-AC splicing machinery *in vitro* (Fig. 1). A 4-base deletion in the conventional 5' splice site of SCN8A intron 3 results in complete skipping of exons 2 and 3 *in vivo* (43). This is probably due to disruption of exon definition interactions between the AT-AC intron 2 and the conventional intron 3. The fact that exon 2 does not join to exon 4 implies that an AT-AC 5' splice site and a conventional 3' splice site are incompatible, i.e., that the AT-AC 5' spliceosomal components and the conventional 3' spliceosomal components cannot be bridged by intron definition. Consistent with this idea, replacement of the SCN4A AT-AC intron 2 branch site element with a conventional branch site abrogates AT-AC splicing *in vitro*, resulting in stimulation of the conventional pathway via a pair of cryptic splice sites (118).

Exon definition interactions between the major and minor

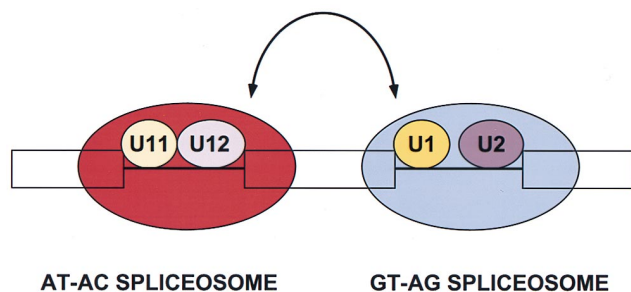


FIG. 1. Exon definition interactions between consecutive minor and major introns. Components of the minor and major spliceosomes interact across the intervening exon. The spliceosomes are denoted by the ellipses. U11 and U12 are shown bound to the AT-AC 5' splice site and branch site, respectively, and U1 and U2 are shown bound to the conventional 5' splice site and branch site, respectively. The interaction across the exon requires U1 snRNA base pairing at the conventional 5' splice site. Bound U1 snRNP probably interacts indirectly with U12 snRNP components, perhaps through bridging by SR proteins. U11 and U12 form a di-snRNP particle, although it is not known if their interactions are maintained in the spliceosome.

spliceosomes presumably involve indirect contacts between the major and minor snRNPs (Fig. 1). Whether exon definition can also take place between a downstream AT-AC intron and an upstream conventional intron is not known. The question also arises how the terminal exons are defined in the context of AT-AC introns, which are occasionally the first or the last intron. For example, the AT-AC intron in the calcium channel *CACNL1A4* gene is the first intron (73), whereas the AT-AC introns in the matrilin family genes (4, 109) and *GT335*, a gene of unknown function (50), are the last intron of their respective genes. Therefore, it will be of interest to determine whether terminal exons can be defined by interactions between the AT-AC splicing machinery with 5'-cap recognition components or the polyadenylation machinery.

**Purine-rich enhancers contribute to AT-AC splice site recognition.** In addition to splice site interactions that take place via intron or exon definition, splice site selection can also be facilitated by exonic and/or intronic sequences. Splicing enhancers are elements that contribute to the recognition of authentic splice sites. The actual prevalence of these elements is not yet known. Purine-rich sequences characteristic of most natural exonic splicing enhancers (ESEs) characterized to date are present in a wide variety of cellular and viral genes in

metazoans. Purine-rich sequences have also been shown to modulate plant 5' splice-site selection (61). In metazoans, purine-rich ESEs are recognized by members of the SR protein family. Purine-rich enhancers are usually composed of GAR repeats (R represents a purine nucleotide) but not runs of G or A (35, 98). However, G-rich repeats typical of small intron sequences in vertebrate genes may facilitate the recognition of splice sites in small introns (60). Non-purine-rich splicing enhancers can also stimulate splicing (16, 53, 104, 105).

Some exons downstream of known AT-AC introns have purine-rich sequences that closely resemble those present in natural enhancers of the major splicing pathway (Table 2), raising the question of whether purine-rich sequences can act as AT-AC splicing enhancers. Indeed, heterologous purine-rich sequences that function as enhancers in the conventional pathway also strongly stimulate AT-AC splicing when placed in the context of a downstream exon that naturally lacks such sequences (121). The purine-rich sequences that are present in many exons flanking AT-AC introns may be natural AT-AC splicing enhancers, or they may influence splicing of both conventional and AT-AC introns on either side. It will be interesting to determine whether non-purine-rich exonic enhancers or intronic enhancers can also function in the AT-AC splicing pathway. The finding that purine-rich enhancers can function in the AT-AC splicing pathway suggests that the enhancer-specific functions of SR proteins are relevant to the AT-AC splicing pathway. If this is the case, an important question is whether SR proteins interact with overlapping or with distinct components of the major and minor spliceosomes to mediate splicing enhancement.

In the major splicing pathway, both the downstream 5' splice site and purine-rich enhancers contribute to the selection of the upstream splice site and appear to involve the action of SR proteins (reviewed in references 6, 7, 10, and 57). In addition, the U1 snRNP has also been implicated in enhancer function: U1 snRNA has been cross-linked to a purine-rich enhancer (112) and is present in an *in vitro*-assembled enhancer complex (90, 92). SR proteins facilitate U1 binding to 5' splice sites (21, 44, 127), and it follows that SR proteins bound to purine-rich enhancers may also recruit U1 (reviewed in references 10, 57, and 108). However, intact U1 snRNP is not required for the stimulation of AT-AC splicing by purine-rich enhancers *in vitro* (121). Therefore, enhancer function and exon definition

TABLE 2. Exonic purine-rich sequences present downstream of known AT-AC introns<sup>a</sup>

Gene product	Exon no.	Sequence
P120 <sup>b</sup>	7	...AGC <b>GGGAGGAAGGG</b> CGG...CTC <b>CAAGAAGG</b> ATCT...
CMP <sup>c</sup>	8	...CCT <b>GGAGAAC</b> AC...
GT335 <sup>d</sup>	7	...GGT <b>GAGGAAGG</b> TGCTGGA <b>ACTCACTGGA</b> AGTGA...
CDK5 <sup>e</sup>	10	...GAC <b>GAGGA</b> ACAA...
REP-3 <sup>f</sup>	7	...TAC <b>GAAAGAAAAGGAGAA</b> CAT...GAC <b>AAAAAGAAGGGGAA</b> CCT...
HPS <sup>g</sup>	16	CTACCT <b>GGAGG</b> ATTT...AAC <b>GAAAAG</b> ATGT...GG <b>CAAGGG</b> CCCC...
TFIIS.oA <sup>h</sup>	7	...CCT <b>GAGAA</b> GAAATGT...

<sup>a</sup> Stretches of six or more consecutive purines are shown in boldface type. Each set of three dots represents a flanking or connecting exon segment longer than three nucleotides that is not shown. The exon sizes range from 44 to 200 nucleotides, except for the last exons. The spacing between the indicated purine segments and the preceding 3' splice site ranges from 6 to 136 nucleotides.

<sup>b</sup> Human proliferating cell nucleolar antigen (GenBank accession no. M33132).

<sup>c</sup> Human cartilage matrix protein (GenBank accession no. M55682).

<sup>d</sup> Human protein of unknown function (GenBank accession no. U53007).

<sup>e</sup> Mouse neuronal cyclin-dependent kinase (GenBank accession no. D29678).

<sup>f</sup> Mouse protein involved in DNA mismatch repair (GenBank accession no. L10301).

<sup>g</sup> Mouse protein defective in Hermansky-Pudlak syndrome (GenBank accession no. U78964).

<sup>h</sup> Frog transcription elongation factor (GenBank accession no. X97662).

are mechanistically different, at least in terms of a requirement for U1.

In the major pathway, ESEs and the downstream 5' splice site facilitate recognition of the upstream 3' splice site by specific interactions with SR proteins. These proteins apparently interact with the essential splicing factor U2AF, facilitating its binding to the 3'-splice-site polypyrimidine tract, while U2AF in turn promotes binding of U2 to the branch site (31, 130). The 3' region of AT-AC introns differs from that of conventional introns: the former lack a polypyrimidine tract, and the branch site is much closer to the 3' splice junction. These differences suggest that the AT-AC spliceosomal components involved in recognition of the 3' splice site are different from those of the conventional spliceosome. Surprisingly, despite these differences, both a downstream 5' splice site and purine-rich enhancers stimulate AT-AC splicing, suggesting that both contribute to AT-AC splice-site selection, as they do in the major splicing pathway. Although in the AT-AC splicing pathway both exonic enhancers and a downstream conventional 5' splice site are also expected to interact with spliceosomal components at the upstream 3' splice site, some of these components may be unique to the AT-AC splicing pathway.

#### A DIVERGENT SUBCLASS OF AT-AC INTRONS

In vitro studies showed that splicing of the SCN4A intron 2 occurs by the AT-AC splicing pathway, and this is probably also the case for the homologous introns in the voltage-gated sodium and calcium channel genes (119, 120). Interestingly, the SCN4A gene has another unusual intron, intron 21, which has 5'-AT and AC-3' ends (24, 59, 120), as does the corresponding intron 25 of the SCN5A gene (110, 120). However, the 5'-splice-site and branch site elements of these two introns do not match the highly conserved AT-AC 5'-splice-site and branch site consensus sequences, suggesting that these divergent introns belong to a distinct subclass of AT-AC introns (119). Indeed, in vitro splicing of these two sodium channel introns, termed AT-AC II, requires the major U1, U2, U4, U5, and U6 spliceosomal snRNAs, rather than the minor AT-AC snRNAs (120). Other AT-AC II introns have previously been described in the chicken parvalbumin gene and in xylanase genes of several filamentous fungi (120).

The vast majority of introns have GT and AG boundaries at the intron 5' and 3' ends, respectively. The importance of the intron ends is underscored not only by in vitro and in vivo mutational analyses (3) but also by the fact that splice site mutations impair gene expression and cause numerous human genetic diseases (46, 67). On the other hand, previous mutational analyses showed that certain splice site mutations are compatible with accurate splicing. For example, in the yeast actin intron and a human tropomyosin intron, mutation of G to A at the first position or G to C at the last position compromises splicing; however, the double mutation at both ends of these introns allows accurate splicing in vivo, albeit less efficiently (13, 74, 85). The double mutations generate 5'-AT and AC-3' intron boundaries. Although these observations preceded the discovery of a unique AT-AC splicing pathway, a possible explanation of these results is that the double-mutant pre-mRNAs were processed by the minor pathway. This appears unlikely, because the double-mutant intron sequences do not match the highly conserved AT-AC 5'-splice-site and branch site consensus (66). Although the snRNA requirements for splicing of the mutant introns have not been determined, it is likely that they are processed by the major pathway, by analogy to the natural AT-AC II introns, which require the major snRNAs (120). The mutational results were interpreted

to suggest a non-Watson-Crick interaction between the first and last bases of the introns. Analysis of the splicing of introns with inosine inserted at the intron ends supports this model (85, 99). The non-Watson-Crick interaction between the intron ends probably exists in the natural AT-AC II introns and probably also exists in all AT-AC introns. On the other hand, the observed lack of specificity in selection of the last nucleotide of a yeast intron argues against a direct interaction between the first and last bases of introns (55).

Although the splice site sequences and the positions of the SCN4A AT-AC II intron 21 and SCN5A AT-AC II intron 25 are conserved, the lengths of these two introns are different. The longer SCN4A AT-AC II intron 21 contains an Alu repeat insertion at nucleotides 358 to 660 (numbering as per GenBank entry AF007782) (117). Alu sequences are repetitive elements that are unique to primates and are thought to be derived from the 7SL RNA (reviewed in reference 62). The functions of Alu sequences are unclear, although in some cases they can influence gene expression. Alu sequences contain several regions that differ from either 5' or 3' splice site consensus elements by only one or two nucleotides. When Alu sequences are present within conventional introns, they can have dramatic effects on splicing (reviewed in reference 56). Point mutations can activate cryptic splice sites in intronic Alu elements and result in abnormal protein formation and clinical disease, e.g., Alport syndrome and ornithine delta-aminotransferase deficiency (41, 63). Mutations in the SCN4A gene cause hyperkalemic periodic paralysis and paramyotonia congenita (reviewed in reference 32). Whether there are natural mutations of this gene that involve activation of cryptic splice sites within the Alu insertion in intron 21 or whether this Alu insertion has any functional consequences remains to be seen.

#### U12-TYPE GT-AG INTRONS

AT-AC II introns have 5'-AT and AC-3' boundaries, yet their splicing requires the major spliceosomal snRNAs (120). Conversely, some introns that have 5'-GT and AC-3' boundaries are spliced by the minor spliceosomal snRNAs (18). With the exception of the first and last nucleotides, which are both G, the sequences of these introns match the AT-AC splice site and branch site consensus elements (Tables 3 and 4). This subclass of GT-AG introns has been termed U12-type GT-AG introns (86). Recent compilations showed that U12-type GT-AG introns are more prevalent than AT-AC introns (18, 86). Additional examples of this type of intron are shown in Tables 3 and 4. Interestingly, XPG- and CDK5-encoding genes contain both an AT-AC intron and a U12-type GT-AG intron (54, 72). Likewise, the members of the voltage-gated ion channel  $\alpha$  subunit gene family have several unconventional introns (see below). Intron 2 of the gene encoding human CACNLB3, the voltage-gated calcium channel  $\beta$  subunit, also belongs to the U12-type GT-AG intron class.

#### OTHER NATURAL INTRON BOUNDARIES

**AT-AA introns.** Previous studies showed that a G-to-A mutation at the last position of a major class intron can partially suppress a G-to-A mutation at the first position of the intron in vivo (11, 74). To date, three natural AT-AA introns are known: intron 6 of the gene encoding the DNA excision repair protein hMSH3 (113); intron 7 of the gene encoding *Arabidopsis* AtG5 (115); and intron 3 of the gene encoding pig succinyl-coenzyme A (CoA) synthetase (82) (Table 5). Intron 6 of the gene encoding hMSH3 and intron 7 of the gene encoding AtG5 have the AT-AC 5'-splice-site consensus. Intron 6 of the gene en-

TABLE 3. Compilation of U12-type GT-AG introns<sup>a</sup>

Gene product	Intron no.	5' splice site	Presumptive branch site	3' splice site	Distance (nt)	Size (nt)
MDC <sup>c</sup>	11	AG <b>g</b> ta <b>tc</b> ctc	... ctc <b>ac</b>	ccctc <b>ag</b> GG	9	109
HAL <sup>d</sup>	3	CT <b>g</b> ta <b>tc</b> ctc	... NA <sup>b</sup>	ctggata <b>ag</b> AT	NA	100
HAL	6	AG <b>g</b> ta <b>tc</b> ctt	... NA	tttattc <b>ag</b> TT	NA	100
HAL	15	CT <b>g</b> ta <b>tc</b> ctt	... NA	cttttgaa <b>ag</b> AT	NA	3,000
MNK <sup>e</sup>	11	AG <b>g</b> ta <b>tc</b> ttat	... atg <b>tt</b> a <b>ac</b>	ttatatc <b>ag</b> TG	12	1,000
BMP1 <sup>f</sup>	15	AG <b>g</b> ta <b>tc</b> cc	... NA	ttctgttgctc <b>ag</b> TC	NA	NA
BMP1	15	...	... NA	ctctctcgttt <b>ag</b> AA	NA	NA
BMP1	15	...	... NA	ttgctcccc <b>ag</b> AG	NA	NA
CACNL1A1 <sup>g</sup>	15	CT <b>g</b> ta <b>tc</b> ctc	... NA	acaca <b>ag</b> AT	NA	1,600
CACNL1A2 <sup>h</sup>	16	CT <b>g</b> ta <b>tc</b> ctc	... cc <b>ct</b> t <b>aa</b> a	aagttgaa <b>ag</b> AT	15	3,400
CACNL1A3 <sup>i</sup>	13	CT <b>g</b> ta <b>tc</b> ctc	... tc <b>ct</b> t <b>ag</b> c	taaaccgct <b>ag</b> AC	15	900
CACNL1A4 <sup>j</sup>	16	CT <b>g</b> ta <b>tc</b> ctc	... tc <b>ct</b> g <b>ag</b> ct	cagacatt <b>ag</b> AC	15	155
CACNLB3 <sup>k</sup>	2	AG <b>g</b> ta <b>tc</b> act	... gc <b>act</b> a <b>at</b>	gggcaaat <b>ag</b> CA	17	216
MSH5 <sup>l</sup>	6	T <b>G</b> gta <b>tc</b> ctc	... cc <b>ct</b> g <b>aa</b> a	t <b>ag</b> GT	5	733
Rehn <sup>m</sup>	27	AG <b>g</b> ta <b>tc</b> t	... tc <b>ct</b> t <b>ca</b> c	t <b>ag</b> CT	5	2,000
SCN10A <sup>n</sup>	8	AG <b>g</b> ta <b>tc</b> ctt	... cc <b>ct</b> t <b>g</b> aa	tctc <b>ag</b> AC	9	300
XPG <sup>o</sup>	1	GG <b>g</b> ta <b>tc</b> ctc	... tc <b>ct</b> t <b>ta</b> c	tggttcccc <b>ag</b> AT	14	NA
CDK5 <sup>p</sup>	6	GG <b>g</b> ta <b>tc</b> ctg	... NA	ctccc <b>ag</b> AA	NA	470

<sup>a</sup> Sequences around the 5' splice site, 3' splice site, and presumptive branch site of U12-type introns are shown. The U12-type GT-AG introns described in previous compilations (86) are not shown. See also the SCN10A intron 2 in Table 4. Intron sequences are shown in lowercase letters, flanked by two exon nucleotides on each side, shown in uppercase letters. The consensus positions at the 5' splice site, 3' splice site, and presumptive branch site are shown in boldface type for nearly invariant nucleotides and are underlined for conserved purines or pyrimidines. The three dots indicate sequences that are either not shown or not known. Distance refers to the number of nucleotides from the branch nucleotide to the 3' splice junction. Some of the intron sizes are only approximate.

<sup>b</sup> NA, not available.

<sup>c</sup> Human metalloprotease-like, disintegrin-like, cysteine-rich protein (39).

<sup>d</sup> Human histidase protein, whose gene has three U12-type GT-AG introns (95).

<sup>e</sup> Human Menkes disease protein (ATP7A) (17, 107).

<sup>f</sup> Human bone morphogenic protein 1/tolloid. The 5' splice site is alternatively spliced to three different 3' splice sites (96).

<sup>g</sup> Human fibroblast voltage-gated L-type calcium channel  $\alpha$  subunit (87).

<sup>h</sup> Human pancreatic voltage-gated calcium channel  $\alpha$  subunit (124).

<sup>i</sup> Human skeletal muscle voltage-gated calcium channel  $\alpha$  subunit (33).

<sup>j</sup> Human brain-specific P/Q-type voltage-gated calcium channel  $\alpha$  subunit (73).

<sup>k</sup> Human voltage-gated calcium channel  $\beta$  subunit (124).

<sup>l</sup> Human homologue of the *Saccharomyces cerevisiae* MSH5 protein (114).

<sup>m</sup> Mouse reelin (81).

<sup>n</sup> Mouse sensory neuronal voltage-gated sodium channel  $\alpha$  subunit (88).

<sup>o</sup> Mouse DNA excision repair protein defective in xeroderma pigmentosum (54).

<sup>p</sup> Mouse neuronal cyclin-dependent kinase (72).

coding hMSH3 also has the AT-AC branch site consensus and is the homologous intron of the AT-AC intron 6 of the *Rep-3* gene. Therefore, this intron is almost certainly processed via the AT-AC pathway. The corresponding intron 6 of the related hMSH5 gene is a U12-type GT-AG intron (Table 3). Interestingly, in alternative splicing of *hMSH2*, another related DNA repair gene, intron 12 was reported to have unusual TA-TT ends (65). Although the AT-AA intron 7 of the gene encoding AtG5 may be processed by the AT-AC pathway, this assumption needs to be verified experimentally, since the intron lacks a good match to the AT-AC branch site consensus. Intron 3 of the gene encoding pig succinyl-CoA synthetase is alternatively spliced and lacks the AT-AC 5'-splice-site and branch site consensus. Therefore, splicing of this intron likely requires the major snRNAs.

**AT-AG introns.** Mutation of the first nucleotide from G to A of a major-class intron inhibits the second step of splicing in vitro (13, 74, 85). Many natural mutations giving rise to various genetic diseases consist of G-to-A mutation at the first position of an intron, with the consequent impairment of gene expression (46, 67). For example, this type of mutation in intron 1 of the human  $\beta$ -globin gene prevents normal splicing, activates three cryptic 5' splice sites, and causes thalassemia (106); in intron 5 of the human adenosine deaminase gene, it results in severe immunodeficiency (84); in intron 5 or intron 18 of the human myosin VIIA gene, it causes Usher syndrome type IB (2). These in vitro and in vivo observations suggested that 5'-AT and AG-3' boundaries within the same introns are in-

compatible. This is likely to be true in the context of the major splicing pathway, so the few natural examples of AT-AG introns are likely to be processed via the minor splicing pathway.

Three natural introns have 5'-AT and AG-3' ends: intron 38 of the myosin VIIa gene (40, 51), which codes for a long-tailed unconventional myosin; intron 2 of *CACNL1A1* (87); and intron 24 of *SCN10A* (88). Intron 38 of the myosin VIIa gene has the 5'-splice-site and presumptive branch site sequences ATA TCCGT and TCCTTGAC, respectively, as independently reported by two different groups (40, 51). These sequences are close matches to the AT-AC consensus elements. The introns corresponding to intron 2 of *CACNL1A1* in other voltage-gated calcium channel  $\alpha$  subunit genes are either of the AT-AC class or of the U12-type GT-AG class. Thus, splicing of all three types of introns presumably requires the minor snRNAs, although this assumption needs to be confirmed experimentally. Splicing via the minor pathway would account for the tolerance to 5'-AT and AG-3' intron ends (see below).

#### DIFFERENTIAL TOLERANCE TO MUTATION OF THE LAST INTRON NUCLEOTIDE IN AT-AC AND CONVENTIONAL SPLICING

Most natural AT-AG and AT-AA introns have close matches to the highly conserved AT-AC 5'-splice-site and branch site consensus sequences. The distances between the branch site and the 3'-splice junction are all very short. Therefore, it appears that the AT-AC spliceosome tolerates changes

TABLE 4. Minor intron conservation in voltage-gated sodium and calcium channel  $\alpha$  subunit genes<sup>a</sup>

Gene product	Intron no.	5' splice site	Presumptive branch site	3' splice site	Distance (nt)	Size (nt)	
SCN4A <sup>c</sup>	2	GCat <b>at</b> tcctg	...	tccttgac	cctgccc <b>ac</b> GC	12	126
SCN5A <sup>d</sup>	3	TCat <b>at</b> tc	...	NA <sup>b</sup>	cccacg <b>ac</b> GC	NA	NA
SCN8A <sup>e</sup>	2	TCat <b>at</b> tcct	...	ccctt <b>aac</b>	tcctctc <b>ac</b> AG	12	NA
SCN10A <sup>f</sup>	2	TCGt <b>gt</b> tcct	...	tcctt <b>aac</b>	atggacctc <b>ac</b> agCT	15	4200
CACNL1A1 <sup>g</sup>	2	AAat <b>at</b> tcct	...	tccttgac	tcctttct <b>ac</b> agAC	14	5000
CACNL1A2 <sup>h</sup>	2	AAg <b>t</b> atcct	...	acctt <b>aac</b>	acattttt <b>ac</b> agAC	14	4800
CACNL1A3 <sup>i</sup>	1	AAg <b>t</b> atcct	...	tcctt <b>aac</b>	cctgctc <b>ac</b> agGC	12	600
CACNL1A4 <sup>j</sup>	1	CCat <b>at</b> tcct	...	tcctt <b>aat</b>	tcccca <b>ac</b> acTC	12	NA

<sup>a</sup> Sequences around the 5' splice site, 3' splice site, and presumptive branch site of the minor introns are shown. The minor introns interrupt a homologous position of the coding sequence in all eight members of the voltage-gated sodium and calcium channel  $\alpha$  subunit genes. Smaller subsets of these sequences were shown in previous compilations (18, 119, 120). Intron sequences are shown in lowercase letters, flanked by two exon nucleotides on each side, shown in uppercase letters. The consensus positions at the 5' splice site, 3' splice site, and presumptive branch site sequences are shown in boldface for nearly invariant nucleotides and are underlined for conserved purines or pyrimidines. The three dots indicate sequences that are either not shown or not known. Distance refers to the number of nucleotides from the branch nucleotide to the 3' splice junction. Some of the intron sizes are only approximate.

<sup>b</sup> NA, not available.

<sup>c</sup> Human skeletal muscle voltage-gated sodium channel  $\alpha$  subunit (24, 59).

<sup>d</sup> Human cardiac muscle voltage-gated sodium channel  $\alpha$  subunit (110).

<sup>e</sup> Mouse brain and spinal cord voltage-gated sodium channel  $\alpha$  subunit (43).

<sup>f</sup> Mouse sensory neuronal voltage-gated sodium channel  $\alpha$  subunit (88).

<sup>g</sup> Human fibroblast L-type voltage-gated calcium channel  $\alpha$  subunit (87).

<sup>h</sup> Human pancreatic voltage-gated calcium channel  $\alpha$  subunit (124).

<sup>i</sup> Human skeletal muscle voltage-gated calcium channel  $\alpha$  subunit (33).

<sup>j</sup> Human brain-specific P/Q-type voltage-gated calcium channel  $\alpha$  subunit (73).

in the last intron nucleotide, which can be either C, A, or G. Indeed, a recent in vivo mutational analysis of the last nucleotide of the AT-AC intron in the P120 gene showed that although C-to-G mutation results in activation of a pair of conventional cryptic splice sites, accurate splicing via the AT-AC pathway can still occur (18).

The 5'-AT and AG-3' boundaries of the *CACNL1A1* intron 2 may represent an evolutionary transition between the AT-AC introns in the voltage-gated sodium channel  $\alpha$  subunit genes and the U12-type GT-AG introns within the homologous domain I of the voltage-gated calcium channel  $\alpha$  subunit genes. Intron 2 of mouse *SCN10A*, which has a GTGTCC 5' splice site and a canonical AT-AC branch site (Table 4), may represent a further evolutionary transition from a U12-type GT-AG intron to a conventional intron. The 5'-AT and AA-3' boundaries of *hMSH3* intron 6 represent a drift from the corresponding AT-AC intron 6 of the homologous mouse *Rep-3* gene. The 5'-GT and AG-3' boundaries of *hMSH5* represent a further drift. The major spliceosome has a more stringent sequence requirement for the last intron nucleotide, which can only be a G, unless the first nucleotide is simultaneously mutated. Conversely, in AT-AC II introns, the last nucleotide presumably has to be a C, unless the first nucleotide changes simultaneously. It is probably difficult for an AT-AC II intron

to evolve into a conventional intron, or vice versa, because this would require simultaneous mutations at both ends of the intron. In contrast, AT-AC introns may more easily evolve gradually into conventional introns. The last nucleotide, C, can be first mutated to G, then the first nucleotide, A, can be mutated to G, and splicing would still be catalyzed by the minor spliceosome. Then the +5 position of the 5' splice site can change from C to G. This last mutation may be sufficient for switching from the minor splicing pathway to the major splicing pathway and would likely constitute an irreversible switch, since AT-AC introns have highly conserved 5'-splice-site and branch site elements, whereas conventional introns have much more degenerate elements.

#### COMMITMENT TO A SPECIFIC SPLICING PATHWAY

The dinucleotides at the intron ends are not responsible for commitment of a pre-mRNA to the major or the minor splicing pathway (18, 120). Likewise, the highly conserved AT-AC branch site consensus sequence (UCCUURAY) recognized by U12 snRNA is probably not the major determinant of commitment to the minor pathway, since it also matches the degenerate consensus for conventional branch sites (YNYURAY), and therefore it should also be recognized by the abun-

TABLE 5. Compilation of known AT-AA introns<sup>a</sup>

Gene product	Intron no.	5' splice site	Presumptive branch site	3' splice site	Distance (nt)	Size (nt)	
hMSH3 <sup>c</sup>	6	AGat <b>at</b> tcct	...	tcctt <b>aat</b>	tattatt <b>aa</b> AT	11	2,000
AtG5 <sup>d</sup>	7	AAat <b>at</b> tcct	...	atatt <b>aac</b>	caaggct <b>aa</b> GT	12	138
SCS <sup>e</sup>	3	TGat <b>a</b> agt	...	ttttt <b>gaa</b>	ttctccttctt <b>aa</b> GG	16	NA <sup>b</sup>

<sup>a</sup> Sequences around the 5' splice site, 3' splice site, and presumptive branch site of the AT-AA introns are shown. Intron sequences are shown in lowercase letters, flanked by two exon nucleotides on each side, in uppercase letters. The consensus positions at the 5'-splice-site, 3'-splice-site, and presumptive branch site sequences are shown in boldface for nearly invariant nucleotides and are underlined for conserved purines or pyrimidines. The three dots indicate sequences that are either not shown or not known. The distance refers to the number of nucleotides from the branch nucleotide to the 3' splice junction. The hMSH3 intron size is only approximate.

<sup>b</sup> NA, not available.

<sup>c</sup> Human homologue of the bacterial DNA mismatch repair protein MutS (113).

<sup>d</sup> *A. thaliana* putative transmembrane protein (115).

<sup>e</sup> Pig heart succinyl-CoA synthetase (82).

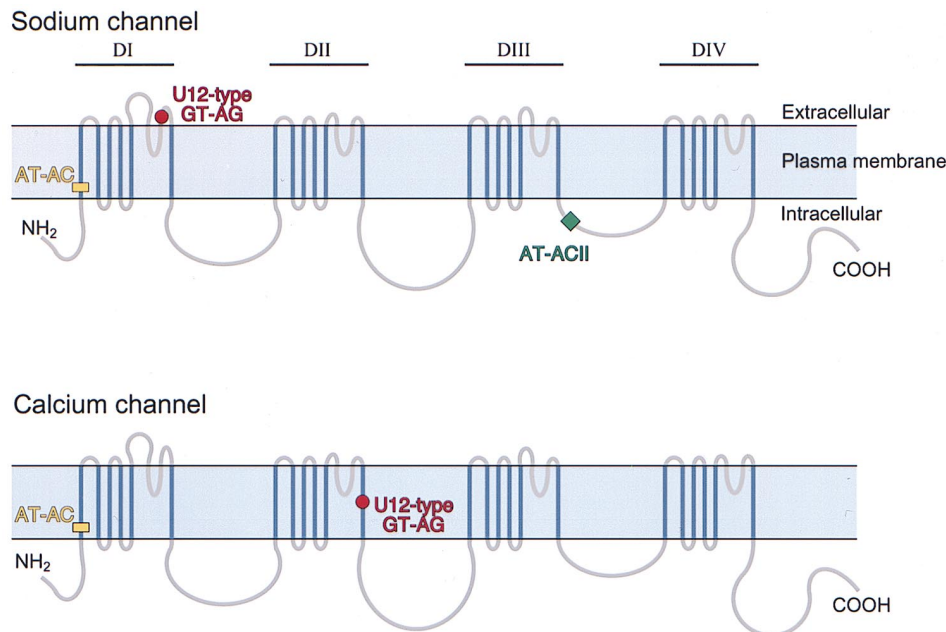


FIG. 2. Topology of voltage-gated ion channel proteins and positions of the unusual introns. The diagram shows the proposed topology of the voltage-gated sodium and calcium channel  $\alpha$  subunit protein families (12, 42). Domain repeats I through IV (DI through DIV) are indicated at the top. Within each domain repeat, the six membrane-spanning segments are shown in blue, and the connecting loops are shown in gray. The positions of unusual introns that interrupt the coding sequences of the corresponding genes are indicated. For some of these introns, the first and last nucleotides are not invariant among different family members (Tables 4 and 6). The conservation of the position of the AT-AC introns (yellow rectangles) in domain I of the sodium and calcium channel genes is indicative of the presence of this intron prior to the divergence of these two gene families. The U12-type GT-AG introns (red circles) are present at different positions of the sodium and calcium channel genes. The AT-AC II intron (green square) is only present in the sodium channel genes. Some family members have lost one or more of these introns or have conventional introns at the same position. Each type of intron interrupts a different part of the domain repeat coding sequence, and hence each intron at the four indicated locations in the voltage-gated ion channel genes probably arose independently.

dant U2 snRNA. Intron 7 of the xylanase *xylP* gene in the filamentous fungus *Penicillium chrysogenum* has a sequence identical to the AT-AC branch site consensus, but it is probably processed via the major splicing pathway (120). It is likely that the highly conserved AT-AC 5' splice site, which is recognized by U11 and U6atac snRNAs, is the major determinant of commitment to the AT-AC splicing pathway. An AT-AC branch site sequence is of course also required but is probably compatible with either pathway. The lack of an extensive polypyrimidine tract and the short distance between the branch site and the 3' splice site in AT-AC introns probably also contribute to their specific commitment to the AT-AC splicing pathway.

#### MULTIPLE MINOR INTRONS IN VOLTAGE-GATED ION CHANNEL GENES

The voltage-gated sodium channel and calcium channel  $\alpha$  subunit genes belong to the voltage-gated ion channel gene superfamily. Mutations in these genes cause numerous neuromuscular and neurological diseases (reviewed in reference 19). The genes have four internal homologous domains (I to IV) that are thought to arise from two rounds of duplication from a single ancestral gene (Fig. 2) (12, 42). Interestingly, these genes have two or sometimes three nonconsensus introns (Fig. 2; Tables 3, 4, and 6). Multiple sequence alignments of the

TABLE 6. AT-AC II intron conservation in voltage-gated sodium channel  $\alpha$  subunit genes<sup>a</sup>

Gene product	Intron no.	5' splice site	Presumptive branch site	3' splice site	Distance (nt)	Size (nt)
SCN4A <sup>c</sup>	21	AGatgagtat	... acctgac ...	ccactataacTT	21	822
SCN5A <sup>d</sup>	25	AGatacgtgg	... ctctgag ...	tctttgacTT	37	680
SCN8A <sup>e</sup>	21	AGataggtct	... NA <sup>b</sup> ...	cctcctttaacTT	NA	NA
SCN10A <sup>f</sup>	24	AGataagtg	... cgttaat	tctcccccttagTT	15	900

<sup>a</sup> Sequences around the 5' splice site, 3' splice site, and presumptive branch site of the minor introns are shown. The minor introns interrupt a homologous position of the coding sequence in all four voltage-gated sodium channel  $\alpha$  subunit genes. The sequences of the first two introns were shown in a previous compilation (120). Intron sequences are shown in lowercase letters, flanked by two exon nucleotides on each side, in uppercase letters. The consensus positions at the 5'-splice-site, 3'-splice-site, and presumptive branch site sequences are shown in boldface for nearly invariant nucleotides and are underlined for conserved purines or pyrimidines. The three dots indicate sequences that are either not shown or not known. Distance refers to the number of nucleotides from the branch nucleotide to the 3' splice junction. The sizes of the SCN4A intron 21 and the SCN5A intron 25 are according to GenBank accession no. AF007782 and AF007781, respectively. The size of SCN10A is only approximate.

<sup>b</sup> NA, not available.

<sup>c</sup> Human skeletal muscle voltage-gated sodium channel  $\alpha$  subunit (24, 59, 120). Reference 59 reported TA as the 3' splice site dinucleotide.

<sup>d</sup> Human cardiac muscle voltage-gated sodium channel  $\alpha$  subunit (110, 120).

<sup>e</sup> Human neuronal voltage-gated sodium channel  $\alpha$  subunit (77).

<sup>f</sup> Mouse sensory neuron voltage-gated sodium channel  $\alpha$  subunit (88).



**A**

SCN4A	exon 2	TFIVLNK GKAI FRFSATPALYLSPF SVVRRGATKVL IHA	LFSMFI MITITLNCVFM TMSDPPW SKNV	exon 3
SCN5A	exon 3	TFIVLNK GKTIFRF SATNALYV LSPFHPV RRAAVKILVHS	LFNMLIMCTILTNCVFM AQHDPPWTKYV	exon 4
SCN8A	exon 2	TFVVLNRGKTLFRFSATPALYLSPFNLIRRIATKIL IHS	VFSMIIMCTILTNCVFM TFSNPP EW SKNV	exon 3
SCN10A	exon 2	TFILLNKSRTISRFSATWALWLFSPFNLIRRTATKVS VHS	WFSIFITVTITLNCVYCMTRTDLPEKL	exon 3
CACNLIA1	exon 2	QYGKPKKGGSTTATRPPRALLCITLKNPIRRACISIV EWK	PFEITILLTIFANCVALAIYIPFEDDSNATNSNL	exon 3
CACNLIA3	exon 1	RKKQPKKPVEILPSPRALFCLTLENPLRKACISIV EWK	PFEITILLTIFANCVALAVYLPMPEDDNNSLNLGL	exon 2
CACNLIA2	exon 2	QYAKSKKGGNSSNSRPARALFCLSLNPIRRACISIV EWK	PFDIFILLATFANCVALAIYIPFEDDSNSTNHNL	exon 3
CACNLIA4	exon 1	MALYNPIVRQNC LTVNRSEFLFSEDNVVRRKYAKKIT EWP	PFEYMLATITIANCVL LALEQHLDDDKT PMSERL	exon 2

**B**

CACNLIA1	exon 15	ILTGEDWNSVMYDGI MAYGGPSFPGMLVCTYFII LFCGN	YILLENVFLAIAVDNLADAE SLTSAQKEEE EKERK KLA	exon 16
CACNLIA2	exon 16	ILTGEDWNAVMIYDGI MAYGGPSSGMI VCTYFII LFCGN	YILLENVFLAIAVDNLADAE SLTSAQKEE EAEKERK KIA	exon 17
CACNLIA3	exon 13	VLTGEDWTSMMYNGI MAYRQPSYPGMLVCTYFII LFCGN	YILLENVFLAIAVDNLAEAE SLTSAQKAKAE EKKRKRK MS	exon 14
CACNLIA4	exon 16	ILTGEDWNEVMYDGI KSGGQVQGGMVFSIYFIVL TLFGN	YTLLENVFLAIAVDNL ANAQELTK	exon 17

**C**

SCN4A	exon 21	KEEQPQYEVNLYMYLYFVFIIFG SFFTLNLFIGV IIDNFNQKKK	LGGQDIFMTEE QKKY YNAMKKL GSKKPQKPIR PPO	exon 22
SCN5A	exon 25	YEEQPQWBYNLYMYIYFVFIIFG SFFTLNLFIGV IIDNFNQKKK	LGGQDIFMTEE QKKY YNAMKKL GSKKPQKPIR RPI	exon 26
SCN8A	exon 21	PDEQPKYEDNIYMYIYFVFIIFG SFFTLNLFIGV IIDNFNQKKK	FGGQDIFMTEE QKKY YNAMKKL GSKKPQKPIR RPL	exon 22
SCN10A	exon 24	INSQPNWEE SLYMYLYFVFIIFG SFFTLNLFVGV IIDNFNQKKK	LGGQDIFMTEE QKKY YNAMKKL GSKKPQKPIR PPI	exon 25

FIG. 3. Conserved positions of unusual introns in voltage-gated ion channels. Multiple protein sequence alignments of the relevant regions of the voltage-gated sodium and calcium channel  $\alpha$  subunits are shown. The complete amino acid sequences encoded by the flanking exons were aligned by using the Genetics Computer Group Pileup program, with a gap creation penalty of 12 and a gap extension penalty of 4 (23). The positions of the unusual introns in the corresponding genes are indicated by the vertical black bar. Amino acid identities in more than half of the sequences are indicated by gray shading. Complete exon sequences are shown, except where indicated. For each alignment, the intron position is also conserved at the nucleotide level (117). (A) Conserved positions of the minor AT-AC introns in eight ion channel genes. Translations of the 5' portions of the calcium channel upstream exons are not shown in the alignment. The human SCN5A sequence is from reference 110; human SCN4A, mouse SCN8A, mouse SCN10A, and human CACNLIA1, CACNLIA2, CACNLIA3, and CACNLIA4 have GenBank accession no. L04216, U59964, Y09108, AJ224873, D43706, U30666, and X99897, respectively. (B) Conserved positions of the minor U12-type GT-AG introns in four calcium channel genes. Human CACNLIA1, CACNLIA2, CACNLIA3, and CACNLIA4 sequences have GenBank accession no. AJ224873, D43718, U30677, and X99897, respectively. (C) Conserved positions of the AT-AC II introns in four sodium channel genes. Human SCN5A and SCN8A sequences are from references 110 and 77, respectively; human SCN4A and mouse SCN10A sequences have GenBank accession no. L04233 and Y09108, respectively.

different family members show that these introns interrupt the genes at exactly homologous positions of the coding sequence (Fig. 3).

SCN4A, SCN5A, and SCN8A, the voltage-gated sodium channel  $\alpha$  subunit in skeletal muscle, cardiac muscle, and brain and spinal cord, respectively, have two rare introns: an AT-AC intron and an AT-AC II intron (43, 77, 120; Fig. 2). The AT-AC intron interrupts the coding sequence of internal domain I, while the AT-AC II intron is located between domains III and IV. Interestingly, the corresponding AT-AC intron 2 of the mouse gene encoding SCN10A, the sensory neuron voltage-gated sodium channel  $\alpha$  subunit, has the sequence GTGT CCT at the 5' splice site and AG as the 3'-splice-site dinucleotide (Table 4) (88). The voltage-gated calcium channel  $\alpha$  subunit genes also have an unusual intron at precisely the same position within domain repeat I (Fig. 2 and 3A). However, these introns have different boundaries (Table 4): intron 2 of the gene encoding CACNLIA1, the fibroblast L-type voltage-gated calcium channel  $\alpha$  subunit, has 5'-AT and AG-3' boundaries; intron 1 of the gene encoding CACNLIA4, the brain-specific P/Q-type voltage-gated calcium channel  $\alpha$  subunit, has 5'-AT and AC-3' boundaries; and intron 2 of the gene encoding CACNLIA2, the pancreatic voltage-gated calcium channel  $\alpha$  subunit, and intron 1 of the gene encoding CACNLIA3, the skeletal muscle voltage-gated calcium channel  $\alpha$  subunit, have 5'-GT and AG-3' boundaries. All eight introns have the highly conserved AT-AC branch site element. The same kind of intron is expected to be present in other voltage-gated sodium and calcium channel  $\alpha$  subunit genes and to be processed via the AT-AC splicing pathway (120). Splicing of an intron of the ADP ribose polymerase gene, which has the features of U12-type GT-AG introns, requires the minor U12 and U6atac snRNAs (18). It is therefore very likely that the GT-AG in-

trons in CACNLIA2 (CACN4) and CACNLIA3 are also processed via the minor pathway.

Interestingly, the CACNLIA2 and CACNLIA3 genes have a second minor intron—a U12-type GT-AG intron—which falls within domain repeat II (Table 3; Fig. 2). This type of intron also has a precisely conserved position in all four known calcium channel  $\alpha$  subunit gene sequences (Fig. 3B). These observations led to the prediction that analogous U12-type GT-AG introns may exist in other voltage-gated calcium channel  $\alpha$  subunit genes. Alignment of the sequences of domain repeats I and II shows that the two minor introns do not interrupt a homologous position. The different location of these introns within their respective domain repeats suggests that they arose independently.

The SCN10A gene is also unusual in having three rare introns out of a total of 26 introns (Fig. 2): the U12-type GT-AG introns 2 and 8 (Tables 3 and 4) and the AT-AC II intron 24 (Table 6). The SCN4A intron 8 and SCN5A intron 9, which correspond to SCN10A intron 8, are conventional GT-AG introns. The voltage-gated calcium channel  $\alpha$  subunit genes lack an intron at the corresponding position. This sequence comparison strongly suggests a shift from a U12-type GT-AG intron to a conventional GT-AG intron by mutational drift after the divergence of sodium and calcium channels from a common ancestral gene.

The SCN4A, SCN5A, SCN8A, and SCN10A genes are paralogues and belong to the voltage-gated sodium channel  $\alpha$  subunit gene family. The genomic organization of these four genes is very similar (77, 88, 110). SCN4A, SCN5A, and SCN8A have an AT-AC II intron at a homologous position (Table 6). Interestingly, the corresponding intron 24 of SCN10A has 5'-AT and AG-3' boundaries (Table 6) (88). All four introns interrupt a homologous position of the coding sequence of their

respective gene (Fig. 3C). However, the voltage-gated calcium channel  $\alpha$  subunit genes lack an intron at the homologous position (Fig. 2). Therefore, similar introns may exist in other members of the voltage-gated sodium channel  $\alpha$  subunit genes but not in the voltage-gated calcium channel  $\alpha$  subunit genes. Because splicing of the corresponding AT-AC II introns of *SCN4A* and *SCN5A* requires the major snRNAs, this is also expected to be the case for *SCN8A* intron 21 and *SCN10A* intron 24. The 5'-AT and AG-3' boundaries of *SCN10A* intron 24 may represent an evolutionary transition between the corresponding AT-AC II introns and the major GT-AG introns.

In summary, the voltage-gated sodium and calcium  $\alpha$  subunit genes, which are derived from a common ancestral gene, show an unusually high frequency of noncanonical introns. The distribution of introns among the two gene families and individual genes, the position of the introns within domain repeats of the coding sequence, and the patterns of sequence divergence at the intron ends provide interesting information about the evolutionary history of minor introns.

### CONCLUSION

Although the first examples of AT-AC intron sequences were discovered only recently (28, 38, 100), a considerable amount of information about the AT-AC pre-mRNA splicing pathway has already been obtained. It is now known that AT-AC introns are widespread in higher eukaryotes (86, 102, 120). The removal of AT-AC introns from pre-mRNA requires the four minor U11, U12, U4atac, and U6atac snRNAs, which play roles analogous to those of U1, U2, U4, and U6 snRNAs in the major splicing pathway, and the major U5 snRNA, which is the only snRNA component shared with the major spliceosome (reviewed in references 70 and 102). Both exon definition interactions and purine-rich exonic enhancers stimulate AT-AC splicing *in vitro* (119, 121). Therefore, they may contribute to AT-AC splice site recognition *in vivo*.

Introns exist in the majority of eukaryotic cellular or viral genes. The mechanisms of splicing, including both catalysis and splice site selection, and the regulation of splicing are incompletely understood at present. However, the identification and characterization of some of the many protein factors involved in splicing continue to provide important clues about splicing and splice site selection mechanisms. The unexpected discovery of the AT-AC splicing pathway provides additional challenges and opportunities for understanding splicing mechanisms and specificity. Whereas the snRNAs involved in AT-AC splicing have been identified, no information is currently available about snRNP and non-snRNP proteins involved in this pathway. It will be especially interesting to determine whether some protein components are shared between the two pathways, and/or whether novel components—which may or may not resemble splicing factors in the conventional pathway—are uniquely required for AT-AC splicing. Determining which components are involved in interactions between the two pathways, such as in exon definition, should also be an important priority. Identification and characterization of these components should provide interesting insights into the mechanisms, regulation, and evolution of pre-mRNA splicing.

### ACKNOWLEDGMENTS

We thank M. Hastings, M. Murray, and B. Graveley for helpful comments on the manuscript and M. Meisler for sharing unpublished sequence information.

The work on AT-AC pre-mRNA splicing in our laboratory is supported by grant GM42699 from the NIH to A.R.K.

### ADDENDUM

Shortly after this review was written, Burge et al. provided an extensive classification of U12-type and U2-type introns on the basis of a statistical analysis of 5'-splice-site and branch site sequences (9a). They describe instances of apparent intron conversion from U12 type to U2 type during evolution by examining introns of homologous or paralogous genes, and they discuss models for the evolution of the major and minor spliceosomes.

Comparison of the cDNA and genomic sequences (GenBank no. AF051782, AC005366, and AC005368) of *HDLA1* (human diaphanous 1), a nonsyndromic deafness susceptibility gene, reveals the presence of two AT-AC introns and one GC-AG intron (118).

### REFERENCES

1. Abovich, N., and M. Rosbash. 1997. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**:403–412.
2. Adato, A., D. Weil, H. Kalinski, Y. Pel-Or, H. Ayadi, C. Petit, M. Korostishevsky, and B. Bonne-Tamir. 1997. Mutation profile of all 49 exons of the human myosin VIIA gene, and haplotype analysis, in Usher 1B families from diverse origins. *Am. J. Hum. Genet.* **61**:813–821.
3. Aebi, M., H. Hornig, and C. Weissmann. 1987. 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. *Cell* **50**:237–246.
4. Aszodi, A., D. R. Beier, L. Hiripi, Z. Bosze, and R. Fassler. 1998. Sequence, structure and chromosomal localization of *Crtm* gene encoding mouse cartilage matrix protein and its exclusion as a candidate for murine achondroplasia. *Matrix Biol.* **16**:563–573.
5. Bao, Y., T. L. Dawson, Jr., and Y. T. Chen. 1996. Human glycogen debranching enzyme gene (*AGL*): complete structural organization and characterization of the 5' flanking region. *Genomics* **38**:155–165.
6. Berget, S. M. 1995. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**:2411–2414.
7. Black, D. L. 1995. Finding splice sites within a wilderness of RNA. *RNA* **1**:763–771.
8. Breathnach, R., C. Benoist, K. O'Hare, F. Gannon, and P. Chambon. 1978. Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl. Acad. Sci. USA* **75**:4853–4857.
9. Burge, C. B., T. H. Tuschl, and P. A. Sharp. Splicing of precursors to messenger RNAs by the spliceosome, p. 561–584. *In* R. F. Gesteland, and J. F. Atkins (ed.), *The RNA world II*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- 9a. Burge, C. B., R. A. Padgett, and P. A. Sharp. 1998. Evolutionary fates and origins of U12-type introns. *Mol. Cell* **2**:773–785.
10. Cáceres, J. F., and A. R. Krainer. 1997. Mammalian pre-mRNA splicing factors, p. 174–212. *In* A. R. Krainer (ed.), *Eukaryotic mRNA processing*. IRL Press, Oxford, United Kingdom.
11. Carothers, A. M., G. Urlaub, D. Grunberger, and L. A. Chasin. 1993. Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol. Cell. Biol.* **13**:5085–5098.
12. Catterall, W. A. 1988. Structure and function of voltage-sensitive ion channels. *Science* **242**:50–61.
13. Chanfreau, G., P. Legrain, B. Dujon, and A. Jacquier. 1994. Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3' splice site selection in *S. cerevisiae*. *Nucleic Acids Res.* **22**:1981–1987.
14. Chew, S., H.-X. Liu, A. Mayeda, and A. R. Krainer. An exonic splicing enhancer facilitates the second catalytic step of pre-mRNA splicing. Submitted for publication.
15. Cho, C., L. Turner, P. Primakoff, and D. G. Myles. 1997. Genomic organization of the mouse fertilin beta gene that encodes an ADAM family protein active in sperm-egg fusion. *Dev. Genet.* **20**:320–328.
16. Coulter, L. R., M. A. Landree, and T. A. Cooper. 1997. Identification of a new class of exonic splicing enhancers by *in vivo* selection. *Mol. Cell. Biol.* **17**:2143–2150.
17. Dierick, H. A., L. Ambrosini, J. Spencer, T. W. Glover, and J. F. Mercer. 1995. Molecular structure of the Menkes disease gene (*ATP7A*). *Genomics* **28**:462–469.
18. Dietrich, R. C., R. Incorvaia, and R. A. Padgett. 1997. Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol. Cell* **1**:151–160.
19. Doyle, J. L., and L. Stubbs. 1998. Ataxia, arrhythmia and ion-channel gene defects. *Trends Genet.* **14**:92–98.
20. Durkin, M. E., M. Gautam, F. Loechel, J. R. Sanes, J. P. Merlie, R. Albrechtsen, and U. M. Wewer. 1996. Structural organization of the human

- and mouse laminin beta2 chain genes, and alternative splicing at the 5' end of the human transcript. *J. Biol. Chem.* **271**:13407-13416.
21. **Eperon, I. C., D. C. Ireland, R. A. Smith, A. Mayeda, and A. R. Krainer.** 1993. Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *EMBO J.* **12**:3607-3617.
  22. **Fujimaki, K., T. Yamazaki, M. Taniwaki, and A. Ichinose.** 1998. The gene for human protein Z is localized to chromosome 13 at band q34 and is coded by eight regular exons and one alternative exon. *Biochemistry* **37**: 6838-6846.
  23. **Genetics Computer Group.** 1994. Program manual for the Wisconsin Package, version 8. Genetics Computer Group, Madison, Wis.
  24. **George, A. L., Jr., G. S. Iyer, R. Kleinfield, R. G. Kallen, and R. L. Barchi.** 1993. Genomic organization of the human skeletal muscle sodium channel gene. *Genomics* **15**:598-606.
  25. **Gilliam, A. C., and J. A. Steitz.** 1993. Rare scleroderma autoantibodies to the U11 small nuclear ribonucleoprotein and to the trimethylguanosine cap of U small nuclear RNAs. *Proc. Natl. Acad. Sci. USA* **90**:6781-6785.
  26. **Gontarek, R. R., M. T. McNally, and K. Beemon.** 1993. Mutation of an RSV intronic element abolishes both U11/U12 snRNP binding and negative regulation of splicing. *Genes Dev.* **7**:1926-1936.
  27. **Gupta, S., R. K. Busch, R. Singh, and R. Reddy.** 1990. Characterization of U6 small nuclear RNA cap-specific antibodies. Identification of gamma-monomethyl-GTP cap structure in 75K and several other human small RNAs. *J. Biol. Chem.* **265**:19137-19142.
  28. **Hall, S. L., and R. A. Padgett.** 1994. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239**:357-365.
  29. **Hall, S. L., and R. A. Padgett.** 1996. Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* **271**:1716-1718.
  30. **Hawkins, J. D.** 1988. A survey on intron and exon lengths. *Nucleic Acids Res.* **16**:9893-9908.
  31. **Hoffman, B. E., and P. J. Grabowski.** 1992. U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev.* **6**:2554-2568.
  32. **Hoffman, E. P.** 1995. Voltage-gated ion channelopathies: inherited disorders caused by abnormal sodium, chloride, and calcium regulation in skeletal muscle. *Annu. Rev. Med.* **46**:431-441.
  33. **Hogan, K., R. G. Gregg, and P. A. Powers.** 1996. The structure of the gene encoding the human skeletal muscle alpha 1 subunit of the dihydropyridine-sensitive L-type calcium channel (CACNL1A3). *Genomics* **31**:392-394.
  34. **Hu, Q., W. A. Kukull, S. L. Bressler, M. D. Gray, J. A. Cam, E. B. Larson, G. M. Martin, and S. S. Deeb.** 1998. The human FE65 gene: genomic structure and an intronic biallelic polymorphism associated with sporadic dementia of the Alzheimer type. *Hum. Genet.* **103**:295-303.
  35. **Humphrey, M. B., J. Bryan, T. A. Cooper, and S. M. Berget.** 1995. A 32-nucleotide exon-splicing enhancer regulates usage of competing 5' splice sites in a differential internal exon. *Mol. Cell. Biol.* **15**:3979-3988.
  36. **Ianzano, L., M. D'Apollito, M. Centra, M. Savino, O. Levran, A. D. Auerbach, A. M. Cleton-Jansen, N. A. Doggett, J. C. Pronk, A. J. Tipping, R. A. Gibson, C. G. Mathew, S. A. Whitmore, S. Apostolou, D. F. Callen, L. Zelante, and A. Savoia.** 1997. The genomic organization of the Fanconi anemia group A (FAA) gene. *Genomics* **41**:309-314.
  37. **Incorvaia, R., and R. A. Padgett.** 1998. Base pairing with U6atac snRNA is required for 5' splice site activation of U12-dependent introns in vivo. *RNA* **4**:709-718.
  38. **Jackson, I. J.** 1991. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19**:3795-3798.
  39. **Katagiri, T., Y. Harada, M. Emi, and Y. Nakamura.** 1995. Human metalloprotease/disintegrin-like (MDC) gene: exon-intron organization and alternative splicing. *Cytogenet. Cell. Genet.* **68**:39-44.
  40. **Kelley, P. M., M. D. Weston, Z. Y. Chen, D. J. Orten, T. Hasson, L. D. Overbeck, J. Pinnit, C. B. Talmadge, P. Ing, M. S. Mooseker, D. Corey, J. Sumegi, and W. J. Kimberling.** 1997. The genomic structure of the gene defective in Usher syndrome type Ib (MYO7A). *Genomics* **40**:73-79.
  41. **Knebelmann, B., L. Forestier, L. Drouot, S. Quinones, C. Chuet, F. Benessy, J. Saus, and C. Antignac.** 1995. Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. *Hum. Mol. Genet.* **4**:675-679.
  42. **Koester, J.** 1991. Voltage-gated ion channels and generation of the action potential, p. 104-118. *In* E. Kandel, J. Schwartz, and T. Jessell (ed.), Principles of neural science. Elsevier Science Publishing Co., New York, N.Y.
  43. **Kohrman, D. C., J. B. Harris, and M. H. Meisler.** 1996. Mutation detection in the med and medJ alleles of the sodium channel *Scn8a*. Unusual splicing due to a minor class AT-AC intron. *J. Biol. Chem.* **271**:17576-17581.
  44. **Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Lührmann, M. A. Garcia-Blanco, and J. L. Manley.** 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**: 119-124.
  45. **Kolosova, I., and R. A. Padgett.** 1997. U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA* **3**:227-233.
  46. **Krawczak, M., J. Reiss, and D. N. Cooper.** 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum. Genet.* **90**:41-54.
  47. **Kreivi, J. P., and A. I. Lamond.** 1996. RNA splicing: unexpected spliceosome diversity. *Curr. Biol.* **6**:802-805.
  48. **Kreivi, J. P., K. Zefrivitz, and G. Akusjärvi.** 1991. A U1 snRNA binding site improves the efficiency of in vitro pre-mRNA splicing. *Nucleic Acids Res.* **19**:6956.
  49. **Kuo, H. C., F. H. Nasim, and P. J. Grabowski.** 1991. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* **251**:1045-1050.
  50. **Lafrenière, R. G., D. L. Rochefort, Z. Kibar, E. A. Fon, F. Han, J. Cochiu, X. Kang, S. Baird, R. G. Korneluk, E. Andermann, J. M. Rommens, and G. A. Rouleau.** 1996. Isolation and characterization of GT335, a novel human gene conserved in *Escherichia coli* and mapping to 21q22.3. *Genomics* **38**:264-272.
  51. **Lévy, G., F. Levi-Acobas, S. Blanchard, S. Gerber, D. Larget-Piet, V. Chenal, X. Z. Liu, V. Newton, K. P. Steel, S. D. Brown, A. Munnich, J. Kaplan, C. Petit, and D. Weil.** 1997. Myosin VIIA gene: heterogeneity of the mutations responsible for Usher syndrome type IB. *Hum. Mol. Genet.* **6**:111-116.
  52. **Lewis, J. D., E. Izaurralde, A. Jarmolowski, C. McGuigan, and I. W. Mattaj.** 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev.* **10**:1683-1698.
  53. **Liu, H. X., M. Zhang, and A. R. Krainer.** 1998. Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev.* **12**:1998-2012.
  54. **Ludwig, D. L., J. S. Mudgett, M. S. Park, A. V. Perez-Castro, and M. A. MacInnes.** 1996. Molecular cloning and structural analysis of the functional mouse genomic XPG gene. *Mamm. Genome* **7**:644-649.
  55. **Luukkonen, B. G., and B. Séraphin.** 1997. The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in *Saccharomyces cerevisiae*. *EMBO J.* **16**:779-792.
  56. **Makalowski, W., G. A. Mitchell, and D. Labuda.** 1994. Alu sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet.* **10**:188-193.
  57. **Manley, J. L., and R. Tacke.** 1996. SR proteins and splicing control. *Genes Dev.* **10**:1569-1579.
  58. **Matera, A. G., and D. C. Ward.** 1993. Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells. *J. Cell Biol.* **121**:715-727.
  59. **McClatchey, A. I., C. S. Lin, J. Wang, E. P. Hoffman, C. Rojas, and J. F. Gusella.** 1992. The genomic structure of the human skeletal muscle sodium channel gene. *Hum. Mol. Genet.* **1**:521-527.
  60. **McCullough, A. J., and S. M. Berget.** 1997. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol. Cell. Biol.* **17**:4562-4571.
  61. **McCullough, A. J., and M. A. Schuler.** 1997. Intronic and exonic sequences modulate 5' splice site selection in plant nuclei. *Nucleic Acids Res.* **25**: 1071-1077.
  62. **Mighell, A. J., A. F. Markham, and P. A. Robinson.** 1997. Alu sequences. *FEBS Lett.* **417**:1-5.
  63. **Mitchell, G. A., D. Labuda, G. Fontaine, J. M. Saudubray, J. P. Bonnefont, S. Lyonnet, L. C. Brody, G. Steel, C. Obie, and D. Valle.** 1991. Splice-mediated insertion of an Alu sequence inactivates ornithine delta-aminotransferase: a role for Alu elements in human mutation. *Proc. Natl. Acad. Sci. USA* **88**:815-819.
  64. **Montzka, K. A., and J. A. Steitz.** 1988. Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc. Natl. Acad. Sci. USA* **85**:8885-8889.
  65. **Mori, Y., H. Shiwaku, S. Fukushima, S. Wakatsuki, M. Sato, T. Nukiwa, and A. Horii.** 1997. Alternative splicing of hMSH2 in normal human tissues. *Hum. Genet.* **99**:590-595.
  66. **Mount, S. M.** 1996. AT-AC introns: an ATtACK on dogma. *Science* **271**: 1690-1692.
  67. **Nakai, K., and H. Sakamoto.** 1994. Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* **141**:171-177.
  68. **Newman, A. J.** 1997. The role of U5 snRNP in pre-mRNA splicing. *EMBO J.* **16**:5797-5800.
  69. **Nilsen, T. W.** 1996. A parallel spliceosome. *Science* **273**:1813.
  70. **Nilsen, T. W.** 1998. RNA-RNA interactions in nuclear pre-mRNA splicing, p. 279-307. *In* R. W. Simons, and M. Grunberg-Manago (ed.), RNA structure and function. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
  71. **Niwa, M., and S. M. Berget.** 1991. Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev.* **5**:2086-2095.
  72. **Ohshima, T., J. W. Nagle, H. C. Pant, J. B. Joshi, C. A. Kozak, R. O. Brady, and A. B. Kulkarni.** 1995. Molecular cloning and chromosomal mapping of the mouse cyclin-dependent kinase 5 gene. *Genomics* **28**:585-588.
  73. **Ophoff, R. A., G. M. Terwindt, M. N. Vergouwe, R. van Eijk, P. J. Oefner,**

- S. M. Hoffman, J. E. Lamerdin, H. W. Mohrenweiser, D. E. Bulman, M. Ferrari, J. Haan, D. Lindhout, G. J. van Ommen, M. H. Hofker, M. D. Ferrari, and R. R. Frants. 1996. Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the  $Ca^{2+}$  channel gene CACN1A4. *Cell* **87**:543–552.
74. Parker, R., and P. G. Siliciano. 1993. Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361**:660–662.
75. Phillips, M. S., J. Fujii, V. K. Khanna, S. DeLeon, K. Yokobata, P. J. de Jong, and D. H. MacLennan. 1996. The structural organization of the human skeletal muscle ryanodine receptor (RYR1) gene. *Genomics* **34**:24–41.
76. Plant, K. E., A. Hair, and G. T. Morgan. 1996. Genes encoding isoforms of transcription elongation factor TFIIS in *Xenopus* and the use of multiple unusual RNA processing signals. *Nucleic Acids Res.* **24**:3514–3521.
77. Plummer, N. W., J. Galt, J. M. Jones, D. L. Burgess, L. K. Sprunger, D. C. Kohrman, and M. H. Meisler. 1998. Exon organization, coding sequence, physical mapping, and polymorphic intragenic markers for the human neuronal sodium channel gene SCN8A. *Genomics* **54**:287–296.
78. Rain, J. C., Z. Rafi, Z. Rhani, P. Legrain, and A. Krämer. 1998. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* **4**:551–565.
79. Richards, A., C. Luccarini, and F. M. Pope. 1997. The structural organization of LAMA4, the gene encoding laminin alpha4. *Eur. J. Biochem.* **248**:15–23.
80. Robberson, B. L., G. J. Cote, and S. M. Berget. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**:84–94.
81. Royaux, L., C. Lambert de Rouvroit, G. D'Arcangelo, D. Demirov, and A. M. Goffinet. 1997. Genomic organization of the mouse reelin gene. *Genomics* **46**:240–250.
82. Ryan, D. G., T. Lin, E. Brownie, W. A. Bridger, and W. T. Wolodko. 1997. Mutually exclusive splicing generates two distinct isoforms of pig heart succinyl-CoA synthetase. *J. Biol. Chem.* **272**:21151–21159.
83. Sahba, S., A. Nechiporuk, K. P. Figueroa, T. Nechiporuk, and S. M. Pulst. 1998. Genomic structure of the human gene for spinocerebellar ataxia type 2 (SCA2) on chromosome 12q24.1. *Genomics* **47**:359–364.
84. Santisteban, I., F. X. Arredondo-Vega, S. Kelly, M. Debre, A. Fischer, J. L. Perignon, B. Hilman, J. elDahr, D. H. Dreyfus, E. W. Gelfand, et al. 1995. Four new adenosine deaminase mutations, altering a zinc-binding histidine, two conserved alanines, and a 5' splice site. *Hum. Mutat.* **5**:243–250.
85. Scadden, A. D. J., and C. W. Smith. 1995. Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J.* **14**:3236–3246.
86. Sharp, P. A., and C. B. Burge. 1997. Classification of introns: U2-type or U12-type. *Cell* **91**:875–879.
87. Soldatov, N. M. 1994. Genomic structure of human L-type  $Ca^{2+}$  channel. *Genomics* **22**:77–87.
88. Souslova, V. A., M. Fox, J. N. Wood, and A. N. Akopian. 1997. Cloning and characterization of a mouse sensory neuron tetrodotoxin-resistant voltage-gated sodium channel gene, Scn10a. *Genomics* **41**:201–209.
89. Spector, D. L. 1993. Nuclear organization of pre-mRNA processing. *Curr. Opin. Cell Biol.* **5**:442–447.
90. Staknis, D., and R. Reed. 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.* **14**:7670–7682.
91. Stamm, S., M. Q. Zhang, T. G. Marr, and D. M. Helfman. 1994. A sequence compilation and comparison of exons that are alternatively spliced in neurons. *Nucleic Acids Res.* **22**:1515–1526.
92. Stark, J. M., D. P. Bazett-Jones, M. Herfort, and M. B. Roth. 1998. SR proteins are sufficient for exon bridging across an intron. *Proc. Natl. Acad. Sci. USA* **95**:2163–2168.
93. Stephens, R. M., and T. D. Schneider. 1992. Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.* **228**:1124–1136.
94. Sterner, D. A., T. Carlo, and S. M. Berget. 1996. Architectural limits on split genes. *Proc. Natl. Acad. Sci. USA* **93**:15081–15085.
95. Suchi, M., H. Sano, H. Mizuno, and Y. Wada. 1995. Molecular cloning and structural characterization of the human histidase gene (HAL). *Genomics* **29**:98–104.
96. Takahara, K., S. Lee, S. Wood, and D. S. Greenspan. 1995. Structural organization and genetic localization of the human bone morphogenetic protein 1/mammalian tollid gene. *Genomics* **29**:9–15.
97. Talerico, M., and S. M. Berget. 1994. Intron definition in splicing of small *Drosophila* introns. *Mol. Cell. Biol.* **14**:3434–3445.
98. Tanaka, K., A. Watakabe, and Y. Shimura. 1994. Polypurine sequences within a downstream exon function as a splicing enhancer. *Mol. Cell. Biol.* **14**:1347–1354.
99. Tarn, W. Y. 1996. Site-specific substitution of inosine at the terminal positions of a pre-mRNA intron: implications for the configuration of the terminal base interaction. *Biochimie* **78**:1057–1065.
100. Tarn, W. Y., and J. A. Steitz. 1996. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* **84**:801–811.
101. Tarn, W. Y., and J. A. Steitz. 1996. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**:1824–1832.
102. Tarn, W. Y., and J. A. Steitz. 1997. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22**:132–137.
103. Tarn, W. Y., T. A. Yario, and J. A. Steitz. 1995. U12 snRNA in vertebrates: evolutionary conservation of 5' sequences implicated in splicing of pre-mRNAs containing a minor class of introns. *RNA* **1**:644–656.
104. Tian, H., and R. Kole. 1995. Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.* **15**:6291–6298.
105. Tian, M., and T. Maniatis. 1994. A splicing enhancer exhibits both constitutive and regulated activities. *Genes Dev.* **8**:1703–1712.
106. Treisman, R., S. H. Orkin, and T. Maniatis. 1983. Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. *Nature* **302**:591–596.
107. Tumer, Z., B. Vural, T. Tonnesen, J. Chelly, A. P. Monaco, and N. Horn. 1995. Characterization of the exon structure of the Menkes disease gene using vectorette PCR. *Genomics* **26**:437–442.
108. Valcárcel, J., R. Singh, and M. R. Green. 1995. Mechanism of regulated pre-mRNA splicing, p. 97–112. In A. I. Lamond (ed.), *Pre-mRNA processing*. Springer-Verlag, Heidelberg, Germany.
109. Wagener, R., B. Kobbe, and M. Paulsson. 1998. Genomic organisation, alternative splicing and primary structure of human matrilin-4. *FEBS Lett.* **438**:165–170.
110. Wang, Q., Z. Li, J. Shen, and M. T. Keating. 1996. Genomic organization of the human SCN5A gene encoding the cardiac sodium channel. *Genomics* **34**:9–16.
111. Wassarman, K. M., and J. A. Steitz. 1992. The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. *Mol. Cell. Biol.* **12**:1276–1285.
112. Watakabe, A., K. Tanaka, and Y. Shimura. 1993. The role of exon sequences in splice site selection. *Genes Dev.* **7**:407–418.
113. Watanabe, A., M. Ikejima, N. Suzuki, and T. Shimada. 1996. Genomic organization and expression of the human MSH3 gene. *Genomics* **31**:311–318.
114. Winand, N. J., J. A. Panzer, and R. D. Kolodner. 1998. Cloning and characterization of the human and *Caenorhabditis elegans* homologs of the *Saccharomyces cerevisiae* MSH5 gene. *Genomics* **53**:69–80.
115. Wu, H. J., P. Gaubier-Comella, M. Delseny, F. Grellet, M. Van Montagu, and R. Rouzé. 1996. Non-canonical introns are at least 10 (9) years old. *Nat. Genet.* **14**:383–384.
116. Wu, J. Y., and T. Maniatis. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**:1061–1070.
117. Wu, Q. 1998. Mechanisms of AT-AC pre-mRNA splicing in vitro. Ph.D. thesis. State University of New York at Stony Brook, Stony Brook.
118. Wu, Q., and A. R. Krainer. Unpublished observations.
119. Wu, Q., and A. R. Krainer. 1996. U1-mediated exon definition interactions between AT-AC and GT-AG introns. *Science* **274**:1005–1008.
120. Wu, Q., and A. R. Krainer. 1997. Splicing of a divergent subclass of AT-AC introns requires the major spliceosomal snRNAs. *RNA* **3**:586–601.
121. Wu, Q., and A. R. Krainer. 1998. Purine-rich enhancers function in the AT-AC pre-mRNA splicing pathway and do so independently of intact U1 snRNP. *RNA* **4**:1664–1673.
122. Xue, J., M. Jorgensen, U. Pihlgren, and L. Rask. 1995. The myrosinase gene family in *Arabidopsis thaliana*: gene organization, expression and evolution. *Plant Mol. Biol.* **27**:911–922.
123. Xue, J., and L. Rask. 1995. The unusual 5' splicing border GC is used in myrosinase genes of the *Brassicaceae*. *Plant Mol. Biol.* **29**:167–171.
124. Yamada, Y., K. Masuda, Q. Li, Y. Ihara, A. Kubota, T. Miura, K. Nakamura, Y. Fujii, S. Seino, and Y. Seino. 1995. The structures of the human calcium channel alpha 1 subunit (CACNL1A2) and beta subunit (CACNLB3) genes. *Genomics* **27**:312–319.
125. Yu, Y. T., and J. A. Steitz. 1997. Site-specific crosslinking of mammalian U11 and U6atac to the 5' splice site of an AT-AC intron. *Proc. Natl. Acad. Sci. USA* **94**:6030–6035.
126. Yu, Y. T., W. Y. Tarn, T. A. Yario, and J. A. Steitz. 1996. More Sm snRNAs from vertebrate cells. *Exp. Cell Res.* **229**:276–281.
127. Zahler, A. M., and M. B. Roth. 1995. Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. *Proc. Natl. Acad. Sci. USA* **92**:2642–2646.
128. Zhang, M. Q. 1998. Statistical features of human exons and their flanking regions. *Hum. Mol. Genet.* **7**:919–932.
129. Zhuang, Y., and A. M. Weiner. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**:827–835.
130. Zuo, P., and T. Maniatis. 1996. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* **10**:1356–1368.