SPLICING DOUBLE: INSIGHTS FROM THE SECOND SPLICEOSOME

Abhijit A. Patel and Joan A. Steitz

Almost 20 years after the discovery of introns and RNA splicing, a second spliceosome was uncovered. Although this new spliceosome is structurally and functionally analogous to the well-characterized major-class splicing apparatus, it mediates the excision of a minor class of evolutionarily conserved introns that have non-canonical consensus sequences. This unanticipated diversity in the splicing machinery is refining both the mechanistic understanding and evolutionary models of RNA splicing.

INTRON
An intervening non-coding sequence that interrupts two exons and that must be excised from pre-messenger RNA transcripts before translation.

EXON
The segment of a pre-messenger RNA transcript that contains protein-coding sequence and/or the 5′ or 3′ untranslated sequences, which must be spliced together with other exons to produce a mature messenger RNA.

Department of Molecular Biophysics and Biochemistry, Howard Hughes Medical Institute, Yale University School of Medicine, 295 Congress Avenue, New Haven, Connecticut 06536, USA. Correspondence to J.A.S. e-mail: js228@email.med.yale.edu
doi:10.1038/nrm1259

Most genes in higher eukaryotes are interrupted by non-coding sequences (INTRONS) that must be precisely excised from pre-messenger RNA (pre-mRNA) molecules to yield mature, functional mRNAs. It has long been known that intron removal and the ligation of flanking sequences (EXONS) occurs through two sequential *trans*-esterification reactions that are carried out by a multicomponent complex that is known as the SPLICEOSOME (FIG. 1). Most introns have common consensus sequences near their 5′ and 3′ ends that are recognized by spliceosomal components and are required for spliceosome formation. The assembly of a spliceosome onto a pre-mRNA is an ordered process that involves five SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES (snRNPs; U1, U2, U4, U5 and U6), as well as an array of protein factors. Catalysis of the splicing reaction proceeds by a coordinated series of RNA–RNA, RNA–protein and protein–protein interactions, which lead to exon ligation and release of the intron LARIAT.

Recently, METAZOAN genes have been found to contain a new class of rare introns that have non-canonical consensus sequences and are excised by a distinct splicing machinery. The low-abundance spliceosome that is responsible for the excision of these introns includes four snRNPs (U11, U12, U4atac and U6atac) that are different from, but functionally analogous to, the well-characterized U1, U2, U4 and U6 snRNPs, respectively. The U5 snRNP is shared by both spliceosomes. Because the excision of these minor-class introns is dependent on the U12 snRNP, they are referred to as U12-type introns, whereas the canonical, major-class introns that require the analogous U2 snRNP are referred to as U2-type introns. Although there are few U12-type introns in the genome of any given species, their persistence throughout virtually all of metazoan evolution indicates that they have an ancient origin and an important cellular function. The two spliceosomes have remarkable mechanistic similarities that highlight their common ancestry and enhance our understanding of the inner workings of the spliceosome.

Here, we summarize the present knowledge of U12-type introns and the spliceosome that is responsible for their excision (several excellent reviews have explored aspects of this topic[1–6]). The similarities and differences between the U12-type introns and spliceosome and the well-studied major-class U2-type introns and spliceosome raise several provocative questions that are addressed in this review. Does the incompatibility of U12-type and U2-type donor and acceptor splice sites yield unique patterns of alternative splicing? Is there communication between spliceosomes that are assembled on U12-type introns and their U2-type neighbours (as has been shown for adjacent U2-type spliceosomes) to prevent unintended skipping of exons or introns? What does analysis of the sequences, genomic distribution and phylogeny of U12-type introns tell us about their origins and evolution? Finally, if introns are merely junk DNA that must be pruned from pre-mRNAs, why have metazoan cells maintained two distinct, but seemingly redundant, systems to excise introns throughout a billion years of evolution?

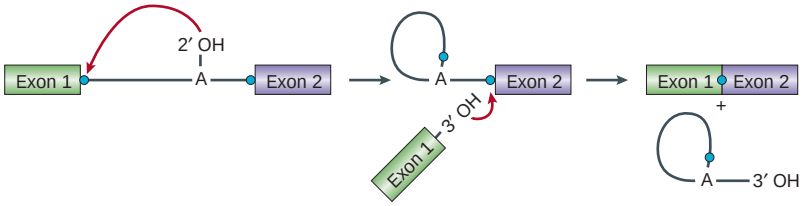


Figure 1 | **Pre-mRNA splicing occurs by two sequential *trans*-esterification reactions.** A schematic pre-messenger RNA is shown on the left as a single intron (solid line) flanked by two exons. The first and second steps of splicing involve nucleophilic attacks (red arrows) on the terminal phosphodiester bonds (blue dots) by the 2′ hydroxyl of the branch-point adenosine (A) and by the 3′ hydroxyl of the upstream exon, respectively. The ligated exons and the lariat intron products are shown on the right.

Splicing by the major-class spliceosome

The spliceosome is a dynamic machine that orchestrates the sequential binding and release of numerous snRNPs and protein factors by the formation and disruption of RNA helices, RNA–protein and protein–protein interactions, of which many require ATP hydrolysis[1,7–9]. It catalyses two *trans*-esterification reactions (FIG. 1). In the first step, cleavage of the 5′ exon–intron junction occurs on nucleophilic attack by the 2′ hydroxyl group of a conserved adenosine residue at the INTRON BRANCH SITE, upstream of the 3′ splice site. This generates a free 3′ hydroxyl group on the upstream exon, as well as a branched lariat intermediate. In the second step, the 3′ intron–exon junction is attacked by the 3′ hydroxyl of the 5′ exon, displacing a lariat intron and ligating the exons.

For major-class introns, spliceosome assembly (FIG. 2a) is thought to begin with the association of the U1 and U2 snRNPs by base-pairing interactions with conserved sequences at the 5′ splice site and intron branch site, respectively[10]. The U2 branch-site duplex 'bulges' the adenosine residue, the 2′ hydroxyl group of which participates in the first nucleophilic attack. The tri-snRNP complex of U5 and the base-paired U4–U6 then stably joins the pre-spliceosome[11], although there is evidence to suggest that U5 interacts upstream of the 5′ splice site at a much earlier stage[12]. Next, rearrangements that are promoted by ATP-hydrolysing protein factors juxtapose the 5′ and 3′ splice sites and form the catalytic core. Specifically, the U4–U6 duplexes unwind[13], and the U4 and U1 snRNPs are displaced, which allows U6 to form base-pairing interactions with the 5′ splice site[14] and with a region of U2 that is near to the U2 branch-site duplex[15–18]. The U5 snRNP has been shown to base-pair with sequences in both the 5′ and 3′ exons, and is believed to position the ends of the two exons for the second step of splicing[12,14,19–21]. After the second step has been completed, the ligated exons and a lariat intron are released, and the spliceosomal components dissociate and are recycled for further rounds of splicing.

Two general properties of the spliceosome are remarkable. First, it is conserved from yeast to humans, both in its protein make-up and in its small nuclear (sn)RNAs, which have short, almost universally conserved sequences that are known to be juxtaposed to the reaction centre (or centres) during catalysis. Second, it is extraordinarily flexible, as it can excise introns of many different lengths and many different sequences. It is also subject to regulation, giving rise to alternatively spliced products in different cells or at different stages of development. So, the discovery of a second spliceosome in the mid-1990s was completely unanticipated.

A new class of introns

The first intron sequences ever characterized revealed highly conserved dinucleotides at the 5′ and 3′ termini (GT and AG, respectively), which were later found to be parts of longer consensus sequences at the 5′ splice site and the 3′ splice site (FIG. 3). The presence of non-consensus splice sites was first recognized by Jackson in 1991 (REF. 22), but it was not until 1994 that Hall and Padgett[23] proposed that there was a distinct minor class of introns. They noted that four introns shared unusual consensus sequences, and predicted that their excision was mediated by a distinct spliceosome that involved low-abundance snRNPs (~10^4 copies per cell) of then unknown function, U11 and U12 (REF. 24). Indeed, U11 and U12 have base-pairing potential with the 5′ splice-site and branch-site sequences, whereas their secondary structures mimic those of U1 and U2, respectively (FIG. 4).

Because these new introns had AT and AC termini, which deviates from the nearly invariant GT-AG rule, they were named AT-AC introns. However, more extensive genomic database surveys revealed that AT-AC termini are not a defining feature of minor-class introns[25–27]. In fact, most minor-class introns have canonical GT-AG termini, and, very rarely, major-class introns have AT-AC termini[26]. Biochemical studies showed that mutation of AT-AC to GT-AG termini did not interfere with splicing by the U12-dependent pathway. Instead, U12-dependent splicing is determined by the longer and more tightly constrained consensus sequences at the 5′ splice site and branch site of minor-class introns, as well as by the lack of a polypyrimidine tract upstream of the 3′ splice site[25,26,28]. Therefore, the more suitable 'U12-type' nomenclature was adopted for this new class of introns[25].

Phylogeny and organization of U12-type introns

Examples of U12-type introns are found in plants and most of the metazoan taxa that have been examined, including vertebrates, insects and cnidarians (jellyfish)[28]. There is no evidence of these introns, or of the U12-type spliceosome, in the genomes of simple eukaryotes such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans* and protists[28]. A bioinformatics search of available genomic sequences in 1998 found a total of 60 non-redundant U12-type introns[28], and a more recent search of the draft version of the human genome sequence found 404 (REF. 29). So, the frequency of occurrence of U12-type introns is in the range of 0.15–0.34% relative to U2-type introns in vertebrates, but is lower in other metazoan taxa[28,29]. An updated phylogenetic distribution of U12-type introns is shown in FIG. 5.

SPLICEOSOME
A large complex that consists of five splicing small nuclear ribonucleoprotein particles as well as numerous protein factors. It mediates the excision of introns from pre-messenger RNA transcripts and ligates exon ends to produce mature mRNAs.

SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLE (snRNP)
A particle that is found in the cell nucleus and consists of a tight complex between a short RNA molecule (<~300 nucleotides) and one or more proteins. SnRNPs are involved in pre-mRNA processing and transfer RNA biogenesis.

LARIAT
An RNA, the 5′ end of which is joined by a phosphodiester linkage to the 2′ hydroxyl of an internal nucleotide, thereby creating a lasso-shaped molecule.

METAZOAN
Refers to all animal species that contain multiple cells differentiated into tissues and organs.

INTRON BRANCH SITE
The adenosine residue near the 3′ end of an intron the 2′ hydroxyl group of which becomes linked to the 5′ end of the intron during the first step of splicing.

U12-type introns almost always coexist with neighbouring U2-type introns in a host gene, and do not show any positional bias relative to other introns in the gene. Although U12-type introns most often occur singly in any given gene, there are several genes that have two, and one gene (human *NHE-6*) that has three[28,29]. There is only one gene that is known to contain only a U12-type intron, and that is *Drosophila melanogaster* transcript CT36969 (REF. 30).

The length of U12-type introns has a similar distribution to that of U2-type introns in humans (U12: mean = 3,600 base pairs (bp), standard deviation = 3,300 bp; U2: mean = 4,130 bp, standard deviation = 3,720 bp)[29]. However, the U2-type distribution has a significant

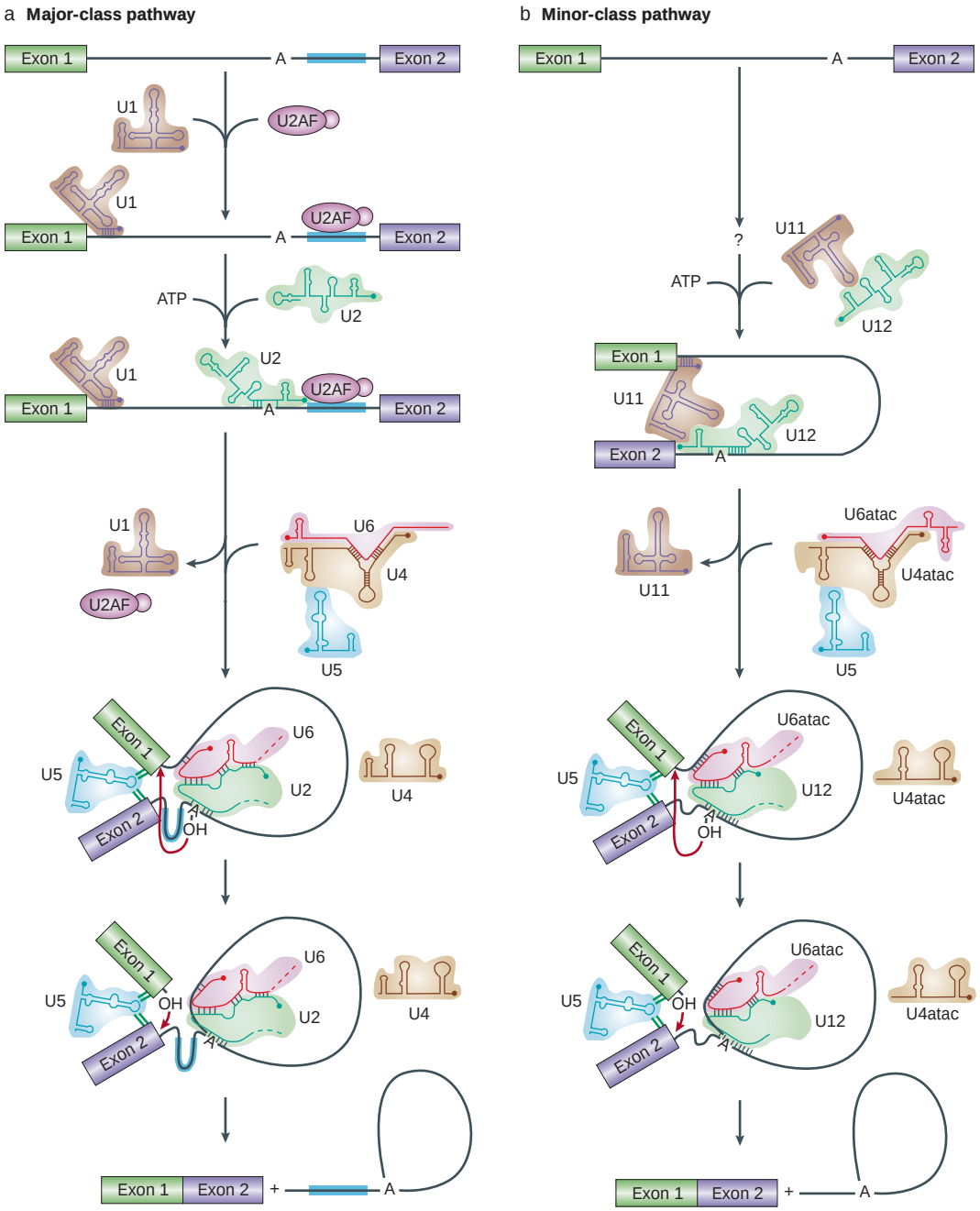


Figure 2 | **Pathways of assembly and catalysis of the major-class and minor-class spliceosomes. a** | The major-class and **b** | the minor-class splicing pathways are shown side by side, highlighting their similarities and differences. The two pathways are mechanistically very similar. Primary differences occur during the early steps of spliceosome formation. The two *trans*-esterification reactions are indicated by red arrows. Each schematic small ribonucleoprotein particle is shown as a small nuclear RNA (not drawn to scale, with the 5′ terminus denoted by a dot) with the surrounding shaded area representing proteins. The polypyrimidine tract of the major-class intron is shaded blue. The double green bars represent interactions between the conserved loop of U5 and exon termini. U2AF, U2 auxillary factor.

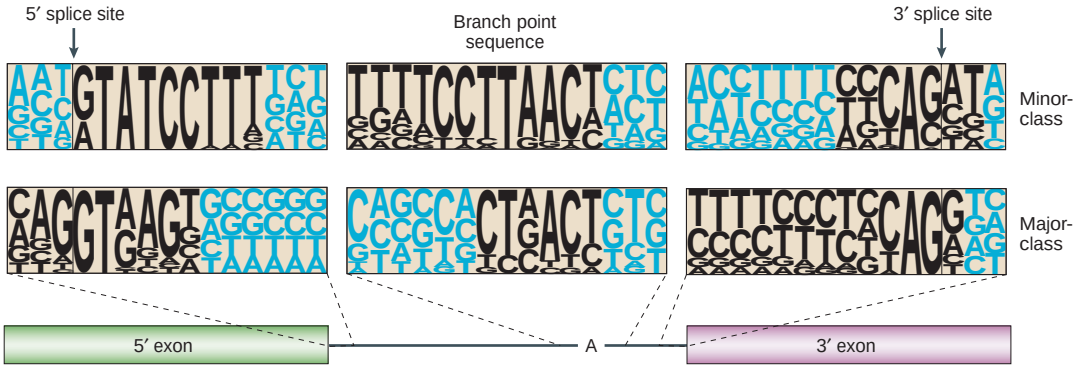


Figure 3 | **Consensus sequences of major-class and minor-class introns.** The consensus sequences of the 5′ splice site, branch site and 3′ splice site are shown from left to right for minor-class introns (upper row) and for major-class introns (lower row). The letter heights at each position represent the frequency of occurrence of the corresponding nucleotides at that position. The positions that are thought to be involved in intron recognition are shown in black; other positions are shown in blue. Frequencies were derived from a set of U12-type introns from various plant and animal species and from a set of mammalian U2-type introns. Modified with permission from REF. 1 © Cold Spring Harbor Laboratory (1999).

subset of short introns that peaks at ~90 bp, whereas the U12-type distribution lacks such a peak[29]. The precision in the pairing of 5′ and 3′ splice sites for short U2-type introns has been proposed to occur through the communication of splicing factors across introns (INTRON-DEFINITION MODEL), whereas that of longer introns involves communication across exons (EXON-DEFINITION MODEL)[31,32]. The dearth of short U12-type introns, as well as experimental evidence for exon-definition interactions[33], indicates that U12-type introns might be spliced exclusively by the exon-definition mechanism. This apparent lack of intronic bridging could be explained by the di-snRNP nature of U11 and U12, which interact poorly with isolated 5′ and 3′ splice sites, as discussed in greater detail below.

U12-type introns are not restricted to a particular functional class of genes, although a large proportion of the genes function in the processing of genetic information[28], such as DNA replication and repair, transcription, RNA processing and translation. Few U12-type intron-containing genes are involved in metabolism or small-molecule biosynthesis[28].

A second spliceosome

The mechanism of U12-type splicing has been characterized *in vitro*. Accurate splicing of an AT-AC intron-containing substrate was achieved in a standard HeLa-cell nuclear extract by adding an antisense oligonucleotide to prevent the major-class U2 snRNP from promoting competitive spliceosome formation on cryptic major-class splice sites within the substrate[34]. It is worth noting that blocking major-class splicing is not always essential to obtain U12-type splicing *in vitro*, but rather depends on the substrate and the quality of the nuclear extract. Psoralen crosslinking studies provided evidence that U12 indeed forms a duplex with the minor-class branch site, apparently bulging the branch-point adenosine[34], which can reside at either of two positions within the consensus site[35]. The minor-class splicing reaction *in vitro* proceeds through the same two-step pathway as the major-class reaction, which involves formation of a lariat intermediate[34]. Native gel electrophoresis of spliceosomal complexes allowed the initial characterization of the assembly pathway (FIG. 2b), and indicated that U11, U12 and U5 were components of the minor-class spliceosome[34]. Interaction of U11 with the 5′ splice site was later confirmed by site-specific crosslinking[36].

The absence of U4 and U6 from the minor-class spliceosome was initially perplexing, given the central role of U6 in catalysis in the major-class spliceosome. This enigma was resolved by the discovery of human U4atac and U6atac, two low-abundance snRNPs (with copy numbers similar to those of U11 and U12), which could be precipitated as a complex using antibodies that recognize the distinctive cap structure of U6 snRNA[37]. Although the sequences of U4atac and U6atac diverge significantly from those of U4 and U6 (FIG. 4), they predict analogous secondary structures and interactions with the pre-mRNA and other snRNAs. Crosslinking studies confirmed the predicted interactions between U4atac and U6atac[37], between U6atac and the minor-class 5′ splice site[36], and between U6atac and U12 (REF. 37), showing that the two spliceosomes undergo comparable dynamic rearrangements in which the snRNAs assume equivalent architectures (FIG. 2b). As U4 does for U6, U4atac seems to chaperone U6atac into the spliceosome, shielding U6atac from interactions with U12 and the 5′ splice site until the U4atac/U6atac helices are unwound during catalytic activation[37].

In vivo evidence of the requirement of U12 in minor-class splicing came from genetic suppression experiments, in which the deficient splicing of a minor-class intron containing two point mutations at the branch site was rescued by co-expression of a U12 snRNA with compensatory mutations[38]. Similar genetic suppression experiments later provided evidence for the *in vivo* interaction of the minor-class 5′ splice site with U11 (REF. 39), as well as with U6atac[40]. Fruitflies that are homozygous for disruptions in U12

INTRON-DEFINITION MODEL
A model that proposes the initial pairwise interaction of spliceosomal components across introns, defining intron units that subsequently interact to promote spliceosome assembly and catalysis.

EXON-DEFINITION MODEL
A model in which exon units, rather than intron units, are initially defined by pairing of spliceosomal components across exons.

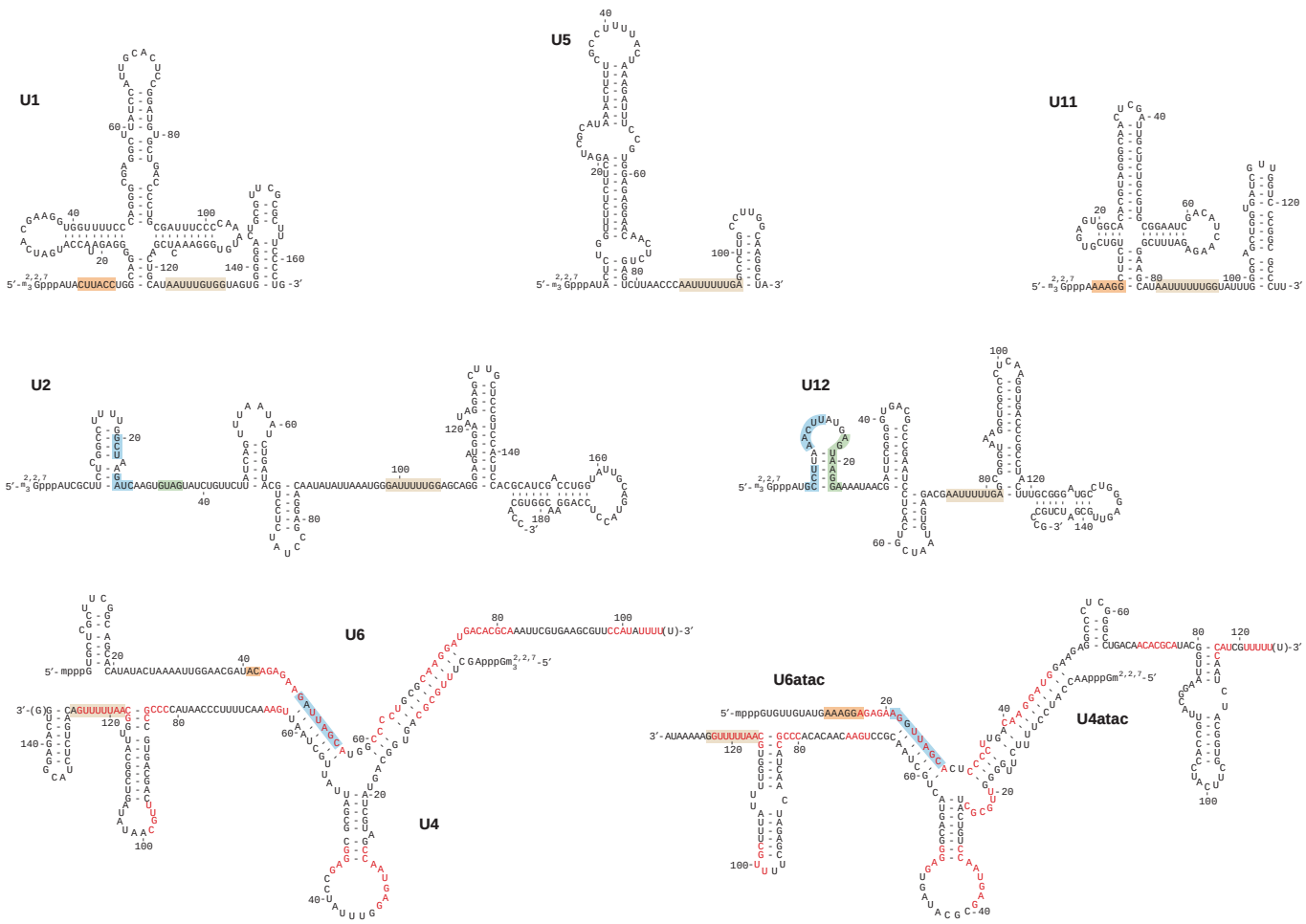


Figure 4 | **Sequences and predicted secondary structures of the human spliceosomal snRNAs.** Similarities in secondary structure are apparent between the major- and minor-class small nuclear RNA counterparts (U1 and U11, U2 and U12, and U4–U6 and U4atac–U6atac) despite substantial sequence divergence. The Sm-binding sites are shaded in light yellow. Coloured boxes indicate sequences that are predicted to be involved in intermolecular RNA–RNA base-pairing interactions: orange for interactions with the 5′ splice site, green for interactions with the branch site, and blue for U2–U6 or U12–U6atac helix I interactions. Sequences in red represent stretches of four or more identical nucleotides between U4–U6 and U4atac–U6atac. Modified with permission from REF. 37 © (1996) The American Association for the Advancement of Science, REF. 119 © Cold Spring Harbor Laboratory (1993) and REF. 120 © Cambridge University Press (2002).

or U6atac genes do not survive early development, which indicates that the minor-class spliceosome is essential for organisms that harbour U12-type introns[41]. Indeed, the presence of U12-type introns within most metazoan genomes indicates that an active U12-type splicing system is indispensable for the cells of most multicellular organisms.

Two sets of spliceosomal snRNPs

Although the minor-class snRNAs have obvious similarities in their secondary structures to the major-class snRNAs, their sequences reveal that they are not simply variants of the major-class snRNAs, but rather, distinct species that function analogously (FIG. 4). Whereas U11 and U12 have virtually no sequence similarity to U1 or U2, U4atac and U6atac have ~40% identity with human U4 and U6, respectively[37]. Remarkably, human U6atac is more divergent from human U6 than yeast and human U6 are from one another (~60% conserved)[42]. The regions of perfect complementarity between U6atac and the minor-class 5′ splice site and between U12 and the branch site, respectively (FIG. 4), argue that the snRNAs of the two distinct spliceosomes coevolved with their respective intron substrates.

The U5 snRNP is unique in serving as a component of both spliceosomes, which indicates that it does not base-pair with sequences that differentiate the two types of introns. Although its role in the major-class spliceosome can involve base-pairing[12,14,19–21], proteins are known to support the juxtaposition of exons for the second step of splicing. Recent evidence that the protein components of U5 undergo marked remodelling during spliceosome activation[43] indicates that U5 has a pivotal role in recruiting common protein factors to the two spliceosomes.

Post-transcriptional modifications of the spliceosomal snRNAs, primarily PSEUDOURIDYLATION and *2′O*-methylation, are clustered in regions that are involved in

PSEUDOURIDYLATION
The conversion of a uridine residue within an RNA chain into a pseudouridine residue, which requires the scission and reattachment of the base to the sugar.

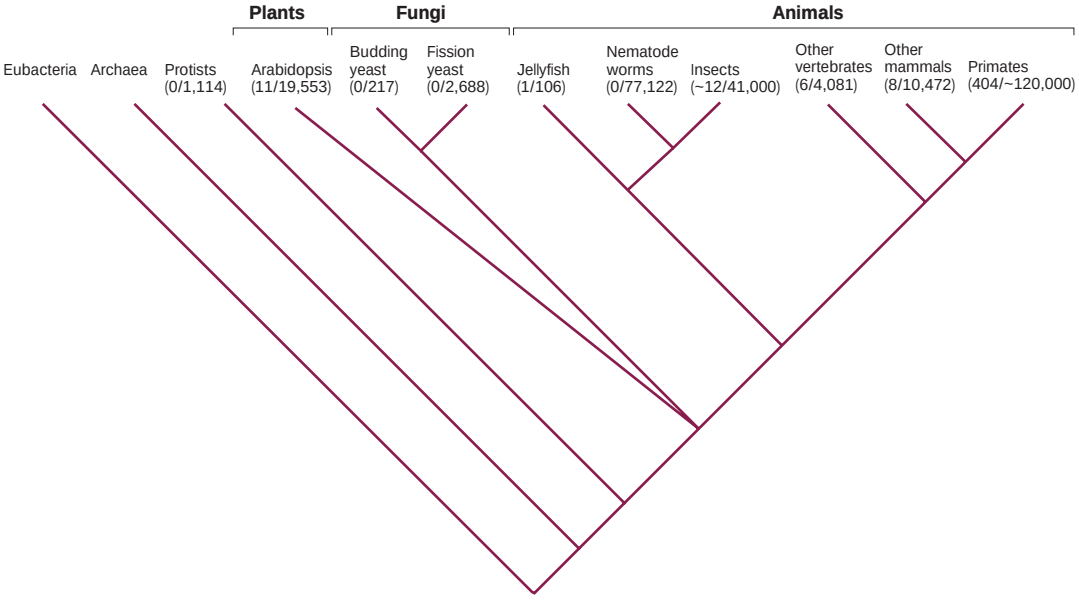


Figure 5 | **The phylogenetic distribution of minor-class introns.** The frequency of occurrence of U12-type introns in different taxa are shown. Below each taxon, the number of U12-type introns identified from the total number of introns analysed is indicated in brackets. The taxa are organized according to evolutionary relationship trees that are based on 18S ribosomal RNA conservation (not drawn to scale). Eubacteria and archaea do not have spliceosomal introns. Only a handful of proposed U12-type introns have been verified to be spliced by the U12-type pathway. They are in the following genes: *Drosophila melanogaster NHE3* (sodium hydrogen exchange channel 3)[114], *CT23545* (REFS 30,41), *CT36969* (REFS 30,41) and *prospero*[30,41], human *p120* (nucleolar protein 20)[34,38], *SCN4A* (voltage-gated sodium channel 4A)[33], *ADPRP* (ADP ribose polymerase)[25] and *SmE* (Sm protein E)[35], and *Xenopus laevis TFIIS* (transcription factor IIS)[35] and *RPL1a* (ribosomal protein L1a)[35]. Modified with permission from REF. 28 © Cell Press (1998) and updated with information from REF. 29 and from personal communications with C. B. Burge and S. Mount.

intermolecular interactions[44], which is consistent with the importance of these modifications in spliceosome assembly and/or function[45–47]. Interestingly, a recent study detected a total of only 4 pseudouridine residues among the 4 minor-class-specific snRNAs, whereas there are 20 in their major-class counterparts[48]. One possibility is that differences in modification confer specificity to the two splicing systems.

Approximately 100 different proteins participate in spliceosome formation, not counting the intron-specific factors that are involved in alternative splicing. Many are stable snRNP components (including the seven common Sm or Sm-like proteins, as well as several particle-specific proteins)[49], whereas others are non-snRNP splicing factors. It is striking that many of the snRNP-associated proteins are shared by the minor-class and major-class spliceosomes[50]. Evaluation of the U11–U12 di-snRNP (see below) showed that the canonical SM PROTEINS, as well as seven subunits of the essential U2-associated splicing factor SF3b[51] (including p14, a protein that is known to interact with the intron's branch-point adenosine[52,53]), are identical to equivalent factors in the major-class spliceosome, whereas eight proteins are U11–U12 specific[54]. Comparative analysis of immunoprecipitated U4–U6–U5 and U4atac–U6atac–U5 tri-snRNPs showed that they have a similar, if not identical, protein content[55]. Notably, the evolutionarily conserved 15.5-kD U4 5′ stem-loop-binding protein also binds to U4atac *in vitro*, which indicates that it is a component of both tri-snRNP complexes[56]. The U5-associated pre-RNA processing 8 (Prp8), a large protein of remarkably high conservation that is believed to have a fundamental role at the spliceosomal active site, immunoprecipitates with both major- and minor-class spliceosomes[57].

SM PROTEIN
A protein that belongs to a group of seven core proteins that are common to the splicing small nuclear ribonucleoprotein particles (except for U6 and U6atac, which have Sm-like proteins). Several are recognized by anti-Sm antibodies that are produced by patients with the autoimmune disease systemic lupus erythematosus.

GROUP II INTRONS
A rare class of autocatalytic introns, the excision of which is assisted by, but does not require, *trans*-acting protein factors.

RIBOZYME
An enzyme that consists of RNA.

Mechanisms of the two spliceosomes

It has been proposed that spliceosomal introns and spliceosomal snRNAs arose as *trans*-acting fragments that were derived from ancestors of the self-splicing GROUP II INTRONS[58–62], which are presently found in the genomes of bacteria, mitochondria and chloroplasts[58,63–65]. Although the similarities could have resulted from convergent evolution[66], the case for common ancestry is substantiated by the mechanistic similarities of spliceosomal and group-II self-splicing systems[67–69]. The catalytic active sites of these systems share numerous RNA structural motifs[70–72], giving rise to the hypothesis that the spliceosome is a protein-assisted RIBOZYME.

Most importantly, both spliceosomes contain an intramolecular stem-loop element (present in U6 and U6atac; FIG. 6) that corresponds to a structure in the self-splicing group II intron, which is known as domain 5 (D5)[73–77]. Analogous catalytic roles for these two elements have recently been shown by functionally replacing the U6atac intramolecular stem-loop with a slightly modified, catalytically active D5 sequence in an *in vivo* splicing assay[78]. The importance of this stem-loop

element in the presumptive catalytic core of both spliceosomes is underscored by the similarities in structure and sequence of this U6 region in organisms as diverse as humans, plants and fission yeast[42], and in human U6atac[37] (FIG. 6). In a similar functional replacement study, the stem-loop from both human and budding yeast U6 was able to substitute for the U6atac stem-loop *in vivo*[79]. Although human and plant U6atac differ in almost half the residues in this region, the stem-loop structure is preserved, and splicing activity is also maintained on transplantation of the plant stem-loop into human U6atac[80].

Another highly conserved region of U6 base-pairs with U2 in the major-class spliceosome to form an intermolecular bipartite helix, which is known as helix I (FIG. 6). Genetic suppression experiments in both yeast and mammalian cells have shown that helix I is crucial for splicing, positioning it at the spliceosomal catalytic core[17,81]. Indeed, the presence of helix I in the minor-class spliceosome — as predicted by the conservation of several residues and base-pairing interactions in this region — was confirmed by mapping crosslinks between U6atac and U12 (REF. 37). Several other U6 and U2 residues in this vicinity have also been identified as vital for major-class splicing[77,82–87], and most of these positions are conserved in the U6atac and U12 snRNAs (FIG. 6).

Parallels between the configuration of the catalytic cores of the two spliceosomes are also apparent during the second step[4]. This includes postulated U5 interactions with exon sequences and non-canonical base-pairing between the terminal intron nucleotides, which has been proposed to occur for either G-G or A-C[88].

The main mechanistic differences between major- and minor-class splicing are likely to occur at the stage of intron recognition rather than during catalysis. Indeed, greater rigidity in the early recognition process of U12-type introns is indicated by the presence of U12 in nuclear extracts as a di-snRNP complex with the more abundant U11 (REFS 24,89), and by evidence that initial intron recognition is carried out by the pre-formed U11–U12 complex[90]. The more constrained consensus sequences at the minor-class 5′ splice site and branch site, as well as the absence of a polypyrimidine tract (which facilitates recognition of major-class introns by binding of the splicing factor U2 auxillary factor (U2AF)[91]), indicates that specification of the sites of spliceosome assembly is more dependent on snRNA-based interactions than is the case for the major-class system.

Another potential difference in the initial recognition of U12-type versus U2-type introns is the requirement for 5′ exon sequences to establish U6atac–5′-splice-site interactions in the minor-class spliceosome[92]. Even very short 5′ exons support the first step of major-class splicing[93]. Recognition of the 5′ splice site twice in the major-class spliceosome, first by U1 and then by U6, is thought to be a proof-reading mechanism that improves splicing specificity[9,94–96]. This mechanism might be less important in the minor-class spliceosome. Whereas

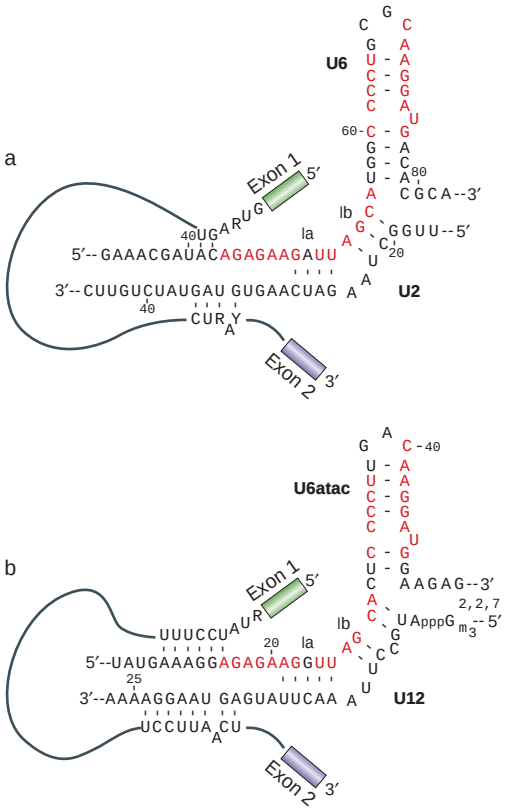


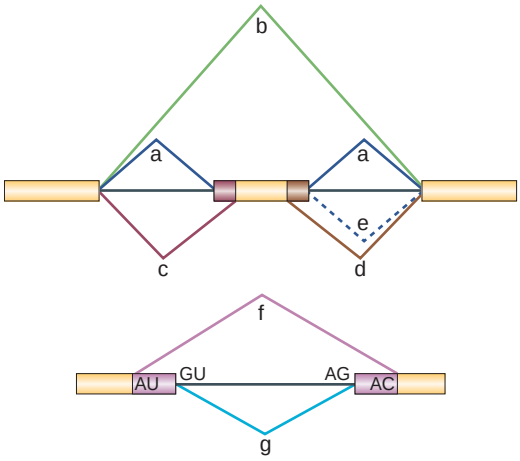
Figure 6 | **Comparison of U2–U6 and U12–U6atac interactions at the catalytic core of the two spliceosomes.** Highly conserved regions of human U2–U6 and U12–U6atac that are essential for catalysis are shown. These include helices Ia and Ib and the intramolecular stem-loops of U6 (nucleotides 57–78) and U6atac (nucleotides 31–48). Red nucleotides indicate residues that are conserved between major-class and minor-class sequences. For both spliceosomes, the branch-point adenosine is bulged from an RNA duplex. Modified with permission from REF. 1 © Cold Spring Harbor Laboratory (1999) and REF. 37 © The American Association for the Advancement of Science (1996).

recognition of the major-class 5′ splice site by U6 has been shown to precede U4–U6 unwinding and the formation of catalytically important U2–U6 interactions[95,97], the analogous U6atac–5′-splice-site and U6atac–U12 interactions can occur independently in either order[92]. Given the virtually identical protein content of the U4atac–U6atac–U5 and U4–U6–U5 tri-snRNPs[55], it is puzzling how each is specifically recruited to its corresponding spliceosome.

Specifying a U12-type intron

Analysis of the information content of major-class and minor-class splicing consensus sequences by Burge *et al.*[1] hinted at fundamental differences in the way the two types of introns are recognized by their respective spliceosomes. The information content of a particular consensus sequence describes the likelihood of finding that sequence in a tract of random sequences — it is roughly determined by the length of the consensus and the degree of conservation at each position[98]. So,

Box 1 | Patterns of alternative splicing



Alterations in splice-site choice determine the inclusion or exclusion of different segments of any given gene, and allow the production of protein isoforms with varying biological activity. Alternative splicing is therefore a source of increased protein diversity and is an important mode of genetic regulation. Splicing patterns can often be specific for tissue type, developmental stage or other physiologic conditions.

Some common alternative splicing patterns are shown in the figure: (**a**) constitutive splicing, (**b**) exon skipping, (**c**) alternative 3′ splice-site choice, (**d**) alternative 5′ splice-site choice and (**e**) intron retention. Major-class and minor-class splice sites are incompatible with each other, and as yet no hybrid introns have been identified. Therefore, unique patterns of alternative splice-site choice can arise. One such example is the 'intron-within-an-intron' architecture of the *prospero* gene of *Drosophila melanogaster*, in which minor-class splice sites (see figure, part **f**) flank an internal major-class intron (see figure, part **g**). In this case, splicing occurs by one or the other pathway, but not both, which produces important changes in the protein-coding sequence[41].

information content can be used to predict the ability of splicing signals to specify intron locations in transcripts accurately. For vertebrate U2-type introns, which can range from ~100 nucleotides to >100 kb, Burge *et al.* found that the relatively degenerate 5′ and 3′ splice signals (including the branch site and polypyrimidine tract) had inadequate information content to specify accurate intron recognition. They concluded, therefore, that the specification of U2-type splice junctions in most vertebrate transcripts is dependent on the recognition of other intronic and/or exonic sequences. By contrast, the longer and more stringently conserved 5′-splice-site and branch-site sequences of U12-type introns seem to contain enough information to specify intron location in a single transcript of average length. However, as human U12-type introns are found in only 1 out of approximately 200 transcripts, all of which are presumably accessible to the U12-type machinery, the information content of the splice signals is insufficient to specify uniquely their location within such a large amount of RNA sequence. So, the accurate recognition of U12-type introns, like U2-type introns, must require other sequences.

The dissimilar distribution of information content among the splice signals of the two intron types also indicates differences in the way they are recognized. The polypyrimidine tract confers greater information content to the 3′ splice site of U2-type introns, whereas the 5′-splice-site and branch-site sequences contain most of the information in U12-type introns. Moreover, the distance betwween the branch site and the 3′ splice site is functionally constrained in U12-type introns[99], which indicates that those two elements might constitute a single recognition unit.

The specification of most exon–intron boundaries is thought to be aided by exon-spanning interactions between factors that are bound to adjacent introns, as proposed by the exon-definition model[31]. Such cooperative protein–protein or protein–snRNP interactions are thought to form functional bridges across exons, which defines them as recognition units and promotes the accurate splicing of adjacent introns. Although considerable evidence supports this model for transcripts that are interrupted by only U2-type introns[100,101], the presence of transcripts containing a U12-type intron amidst U2-type neighbours raises the question of whether similar cooperative interactions can occur between the two different spliceosomes. Indeed, evidence for such interactions was provided when it was observed that the splicing of a U12-type intron could be stimulated when the downstream exon was followed by a U2-type intron 5′ splice site[33]. In addition, exons that flank U12-type introns are generally short, conforming to the size rules that are thought to be important for exon definition, which provides further support for the idea of cross-exon bridging interactions. Finally, the presence of purine-rich exonic enhancers that stimulate U12-type splicing by binding SR PROTEIN FACTORS[34,102] is compatible with the exon-definition model. So, the excision of U12-type introns *in vivo* is likely to be spatially and temporally coordinated with the splicing of neighbouring U2-type introns.

Alternative splicing of U12-type introns

Although the removal of most introns occurs constitutively, regulated splice-site choice can generate different protein isoforms from the same gene and represents an important source of increased genomic diversity[103–106] (BOX 1). The incompatibility of U12-type 5′ and 3′ splice sites with U2-type sites[25,26] prevents U12-type introns from participating in alternative splicing with neighbouring major-class introns. Nonetheless, a recent bioinformatics scan of the human genome uncovered 13 cases of alternative splicing among 404 identified U12-type introns (~3.2% compared with ~14% in a similar analysis of U2-type introns)[29]. Interestingly, most of these cases (11 out of 13) involved an alternative choice of the 3′ rather than the 5′ splice site.

A completely different type of alternative splicing is found in the *prospero* gene of *D. melanogaster*, which contains a U2-type intron that is flanked by U12-type intron consensus sequences (BOX 1). This 'intron-within-an-intron' architecture can be spliced by either the U2-dependent or U12-dependent pathway, which results in a change of five amino acids at the amino terminus of the protein's functionally important homeodomain[41]. The choice of U12-type versus U2-type splicing in this transcript seems to be modulated during embryogenesis, and might be regulated by the binding of protein factors to an intronic purine-rich element (P. Scamborova, A. Wong and J.A.S., manuscript in preparation).

SR FAMILY OF PROTEINS
A group of essential protein splicing factors with one or more RNA-recognition motif and a region containing arginine/serine (S/R) dipeptide repeats, which facilitate spliceosome assembly onto a pre-messenger RNA.

Similar purine-rich enhancer elements have been shown to function as stimulators of U12-type splicing *in vitro*[102], as well as to modify alternative splicing patterns of U12-type introns *in vivo*[107]. The action of such elements in U12-type splicing is thought to be mediated by the binding of *trans*-acting protein factors that belong primarily to the SR family[108]. So, as in the major-class spliceosome[109,110], both intronic and exonic purine-rich elements contribute to exon definition and splice-site choice.

Evolutionary origins and transformations

The remarkable structural and mechanistic similarities of the two spliceosomes indicates that they share a common ancestry. Phylogenetic comparisons of introns at homologous positions (in terms of codon location and phase) in homologous genes led Burge *et al.*[28] to conclude that U12-type introns must have occurred much more frequently earlier in evolution, and were either lost or converted to U2-type introns over time. Further, they proposed a fission/fusion model in which the U12-type and U2-type splicing systems evolved from a common ancestor independently in separate lineages, which later fused to become a progenitor of higher eukaryotes. The merging of genetic material in the progenitor would result in a genome that contained both intron types and the components for both types of spliceosome. Supporting this model is the observation that the fraction of U12-type intron-containing genes that have more than one U12-type intron is much greater than expected, given the extreme rarity of these introns. If the proposed progenitor organism had a fused genome containing U2-only genes and U12-only genes, after numerous conversions from U12-type to U2-type, the remaining U12-type introns would be clustered within the latter subset of genes.

Lynch and Richardson[111] have objected to this fission/fusion hypothesis. They argue that if the splicing systems of the two lineages diverged so much that they were able to function independently when reunited, then it is unlikely that the genomes would have drifted so little that the hybridizing lineages would have remained reproductively compatible. Rather, they propose an alternative model in which the genome was seeded with the descendents of two different self-splicing group-II-like introns. This model does not invoke the presence of two separate lineages, as crosstalk would be unlikely between the internally interacting elements of two self-splicing introns. Fragmentation of the RNA elements of these two self-splicing introns could have given rise to two distinct *trans*-acting splicing systems that structurally converged under the evolutionary pressure of a common pool of splicing proteins.

Despite the noted tendency for conversion and loss of U12-type introns, Burge *et al.*[28] have identified several remarkable instances in which U12-type introns are conserved between homologous genes from highly diverged species. One such example is the Huntington's disease gene (Huntingtin) in which the last intron of 66 nucleotides is a U12-type intron in both pufferfish and humans. Another particularly striking example is the second intron of the gene encoding the sodium channel α-subunit that is conserved between humans and jellyfish, which are organisms that diverged over 600–800 million years ago[112,113]. Perhaps most surprising, given the rarity of U12-type introns, is the observation of several examples of these introns at non-homologous positions in PARALOGOUS GENES[28].

These observations indicate that the few U12-type introns that have resisted conversion or loss could have indispensable roles in metazoan cells. Although such roles have yet to be defined, one hypothesis is that the splicing of U12-type introns might be the rate-limiting step in the processing of the pre-mRNAs that harbour them, and could therefore be a target in the post-transcriptional regulation of gene expression. This idea is supported by recent work from our group, which showed that U12-type introns are removed more slowly than U2-type introns from pre-mRNA transcripts *in vivo*, and that conversion of a U12-type intron to a U2-type intron can markedly increase (~sixfold) the amount of mature mRNA and protein expression *in vivo*[114].

Future directions

A comparison of the structural and mechanistic features of the major-class (U2-type) and minor-class (U12-type) spliceosomes has provided many valuable insights into the essential catalytic elements of the splicing reaction. Further refinements to our understanding of this remarkable RNA machine are sure to emerge from future mutational studies (such as those described in the recent article by Shukla *et al.*[115]) and functional substitutions.

The rate-limiting excision of U12-type introns and their use in the alternative expression of proteins *in vivo* indicates that they might be potential targets of gene regulation. It remains to be established whether the activity of the U12-type spliceosome is modulated in a developmental-stage-specific or tissue-specific manner. Assessing gene expression patterns in transgenic organisms with U12-type to U2-type intron mutations should provide vital evidence and help to rationalize the continued presence of these rare introns in metazoan genomes.

Recent evidence that links splicing to transcription has led to a proposal for co-transcriptional, RNA-polymerase-assisted spliceosome assembly[116–118]. Given the rarity of U12-type introns, it would be surprising if the polymerase carried and deposited both minor-class-specific and major-class-specific components during transcription. Perhaps these links have been important for defining differences in initial intron recognition during assembly of the two spliceosomes.

Finally, the existence of a second spliceosome raises the possibility that a third or fourth might be awaiting discovery. The degeneracy of the major-class consensus sequences would make yet another class of introns difficult to detect. Indeed, the GT-AG U12-type introns might well have been ignored were it not for the initial focus on AT-AC introns. Clearly, we must be alert to the tantalizing possibility that other deviant splicing machines lurk within genomic information.

PARALOGOUS GENES
Genes for which sequence similarity is the result of gene duplication within the same species and that encode proteins that carry out similar, but not identical functions.

1. Burge, C. B., Tuschl, T. & Sharp, P. A. in *The RNA World* 2nd edn (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 525–560 (Cold Spring Harbor Laboratory Press, New York, 1999).
2. Kreivi, J. P. & Lamond, A. I. RNA splicing: unexpected spliceosome diversity. *Curr. Biol.* **6**, 802–805 (1996).
3. Mount, S. M. AT-AC introns: an ATtACk on dogma. *Science* **271**, 1690–1692 (1996).
4. Nilsen, T. W. A parallel spliceosome. *Science* **273**, 1813 (1996).
5. Tarn, W. Y. & Steitz, J. A. Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* **22**, 132–137 (1997).
6. Wu, Q. & Krainer, A. R. AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. *Mol. Cell. Biol.* **19**, 3225–3236 (1999).
7. Moore, M. J., Query, C. C. & Sharp, P. A. in *The RNA World* 1st edn (eds Gesteland, R. F. & Atkins, J. F.) 303–357 (Cold Spring Harbor Laboratory Press, New York, 1993).
8. Nilsen, T. W. in *RNA Structure and Function* (eds Simmons, R. W. & Grunberg-Menago, M.) 279–307 (Cold Spring Harbor Laboratory Press, New York, 1998).
9. Staley, J. P. & Guthrie, C. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell* **92**, 315–326 (1998).
10. Reed, R. Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr. Opin. Genet. Dev.* **6**, 215–220 (1996).
11. Konarska, M. M. & Sharp, P. A. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* **49**, 763–774 (1987).
12. Wyatt, J. R., Sontheimer, E. J. & Steitz, J. A. Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. *Genes Dev.* **6**, 2542–2553 (1992).
13. Lamond, A. I., Konarska, M. M., Grabowski, P. J. & Sharp, P. A. Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proc. Natl Acad. Sci. USA* **85**, 411–415 (1988).
14. Wassarman, D. A. & Steitz, J. A. Interactions of small nuclear RNAs with precursor messenger RNA during *in vitro* splicing. *Science* **257**, 1918–1925 (1992).
15. Datta, B. & Weiner, A. M. Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing. *Nature* **352**, 821–824 (1991).
16. Hausner, T. P., Giglio, L. M. & Weiner, A. M. Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev.* **4**, 2146–2156 (1990).
17. Madhani, H. D. & Guthrie, C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**, 803–817 (1992).
18. Wu, J. A. & Manley, J. L. Base pairing between U2 and U6 snRNAs is necessary for splicing of a mammalian pre-mRNA. *Nature* **352**, 818–821 (1991).
19. Newman, A. & Norman, C. Mutations in yeast U5 snRNA alter the specificity of 5′ splice-site cleavage. *Cell* **65**, 115–123 (1991).
20. Newman, A. J. & Norman, C. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell* **68**, 743–754 (1992).
21. Sontheimer, E. J. & Steitz, J. A. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**, 1989–1996 (1993).
22. Jackson, I. J. A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* **19**, 3795–3798 (1991).
23. Hall, S. L. & Padgett, R. A. Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.* **239**, 357–365 (1994).
The first proposal that a distinct spliceosome might excise a variant class of pre-mRNA introns, the splice-site sequences of which diverge from the established consensus. The participation of low-abundance snRNAs U11 and U12 was suggested on the basis of their complementarity to common sequences at the 5′ end and putative branch point of four of the new introns.
24. Montzka, K. A. & Steitz, J. A. Additional low-abundance human small nuclear ribonucleoproteins: U11, U12, etc. *Proc. Natl Acad. Sci. USA* **85**, 8885–8889 (1988).
The discovery of the low-abundance U11 and U12 snRNPs and their association to form the U11–U12 di-snRNP in extracts of human cells.
25. Dietrich, R. C., Incorvaia, R. & Padgett, R. A. Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol. Cell* **1**, 151–160 (1997).
This study showed that the terminal dinucleotides of U12-type introns can be /GT and AG/, as in the major-class introns, and that U12-type introns can instead be defined on the basis of longer consensus sequences at the 5′ end and the branch point.
26. Sharp, P. A. & Burge, C. B. Classification of introns: U2-type or U12-type. *Cell* **91**, 875–879 (1997).
27. Wu, Q. & Krainer, A. R. Splicing of a divergent subclass of AT-AC introns requires the major spliceosomal snRNAs. *RNA* **3**, 586–601 (1997).
28. Burge, C. B., Padgett, R. A. & Sharp, P. A. Evolutionary fates and origins of U12-type introns. *Mol. Cell* **2**, 773–785 (1998).
A genomic database analysis identified 56 different genes containing U12-type introns in various eukaryotic taxa and detected switching of AT-AC to GT-AG termini in U12-type introns, as well as conversion of U12-type to U2-type introns in homologous genes. A fission/fusion model for the evolutionary origin of the two different splicing systems was proposed.
29. Levine, A. & Durbin, R. A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res.* **29**, 4006–4013 (2001).
30. Otake, L. R. Thesis, Yale Univ. (2001).
31. Berget, S. M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414 (1995).
32. Talerico, M. & Berget, S. M. Intron definition in splicing of small *Drosophila* introns. *Mol. Cell. Biol.* **14**, 3434–3445 (1994).
33. Wu, Q. & Krainer, A. R. U1-mediated exon definition interactions between AT-AC and GT-AG introns. *Science* **274**, 1005–1008 (1996).
The first evidence that splicing of pre-mRNAs containing both U12-type and U2-type introns can be coordinated by exon-definition interactions.
34. Tarn, W. Y. & Steitz, J. A. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell* **84**, 801–811 (1996).
The first description of an *in vitro* system that could remove an AT-AC intron provided evidence for the predicted interaction of U12 with the branch site and for the additional participation of the U11 and U5, but not the U4 and U6 snRNPs.
35. McConnell, T. S., Cho, S. J., Frilander, M. J. & Steitz, J. A. Branchpoint selection in the splicing of U12-dependent introns *in vitro*. *RNA* **8**, 579–586 (2002).
36. Yu, Y. T. & Steitz, J. A. 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 (1997).
37. Tarn, W. Y. & Steitz, J. A. Highly diverged U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* **273**, 1824–1832 (1996).
The identification and characterization of two new low-abundance snRNPs, U4atac and U6atac, allowed documentation of interactions between U6atac and U12 parallel to those previously proposed for U6 and U2 to occur in the spliceosomal active site and to juxtapose an intron's 5′ splice site and branch point for the first step of the splicing reaction.
38. Hall, S. L. & Padgett, R. A. Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science* **271**, 1716–1718 (1996).
Mutations in the conserved branch-point sequence of an AT-AC intron were suppressed by compensatory mutations in U12, providing the first *in vivo* evidence for the U12-type splicing machinery.
39. Kolossova, I. & Padgett, R. A. U11 snRNA interacts *in vivo* with the 5′ splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA* **3**, 227–233 (1997).
40. Incorvaia, R. & Padgett, R. A. Base pairing with U6atac snRNA is required for 5′ splice site activation of U12-dependent introns *in vivo*. *RNA* **4**, 709–718 (1998).
41. Otake, L. R., Scamborova, P., Hashimoto, C. & Steitz, J. A. The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in *Drosophila*. *Mol. Cell* **9**, 439–446 (2002).
P-element-mediated disruptions of the single loci encoding U12 and U6atac in *D. melanogaster* caused lethality during the embryonic and third instar larval stages, respectively, and defects in the excision of U12-type introns from several transcripts.
42. Brow, D. A. & Guthrie, C. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* **334**, 213–218 (1988).
43. Makarov, E. M. *et al.* Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* **298**, 2205–2208 (2002).
44. Massenet, S., Mougin, A. & Branlant, C. in *The Modification and Editing of RNA* (eds Grosjean, H. & Benne, R.) 201–228 (ASM Press, Washington D.C., 1998).
45. Massenet, S. *et al.* Pseudouridine mapping in the *Saccharomyces cerevisiae* spliceosomal U small nuclear RNAs (snRNAs) reveals that pseudouridine synthase pus1p exhibits a dual substrate specificity for U2 snRNA and tRNA. *Mol. Cell. Biol.* **19**, 2142–2154 (1999).
46. Szkukalek, A., Myslinski, E., Mougin, A., Luhrmann, R. & Branlant, C. Phylogenetic conservation of modified nucleotides in the terminal loop 1 of the spliceosomal U5 snRNA. *Biochimie* **77**, 16–21 (1995).
47. Yu, Y. T., Shu, M. D. & Steitz, J. A. Modifications of U2 snRNA are required for snRNP assembly and pre-mRNA splicing. *EMBO J.* **17**, 5783–5795 (1998).
48. Massenet, S. & Branlant, C. A limited number of pseudouridine residues in the human atac spliceosomal UsnRNAs as compared to human major spliceosomal UsnRNAs. *RNA* **5**, 1495–1503 (1999).
49. Will, C. L. & Luhrmann, R. in *Eukaryotic mRNA Processing* (ed. Krainer, A. R.) 130–173 (IRL Press, Oxford, 1997).
50. Will, C. L. & Luhrmann, R. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**, 290–301 (2001).
51. Golas, M. M., Sander, B., Will, C. L., Luhrmann, R. & Stark, H. Molecular architecture of the multiprotein splicing factor SF3b. *Science* **300**, 980–984 (2003).
52. Query, C. C., Strobel, S. A. & Sharp, P. A. Three recognition events at the branch-site adenine. *EMBO J.* **15**, 1392–1402 (1996).
53. Will, C. L. *et al.* A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J.* **20**, 4536–4546 (2001).
54. Will, C. L., Schneider, C., Reed, R. & Luhrmann, R. Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* **284**, 2003–2005 (1999).
Analysis of 20 proteins associated with the U11–U12 di-snRNP showed that not only the U5 snRNP, but several protein components are shared by the major-class and minor-class spliceosomes.
55. Schneider, C., Will, C. L., Makarova, O. V., Makarov, E. M. & Luhrmann, R. Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol. Cell. Biol.* **22**, 3219–3229 (2002).
56. Nottrott, S. *et al.* Functional interaction of a novel 15.5kD [U4/U6.U5] tri-snRNP protein with the 5′ stem-loop of U4 snRNA. *EMBO J.* **18**, 6119–6133 (1999).
57. Luo, H. R., Moreau, G. A., Levin, N. & Moore, M. J. The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA* **5**, 893–908 (1999).
58. Cavalier-Smith, T. Intron phylogeny: a new hypothesis. *Trends Genet.* **7**, 145–148 (1991).
59. Cech, T. R. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* **44**, 207–210 (1986).
60. Jacquier, A. Self-splicing group II and nuclear pre-mRNA introns: how similar are they? *Trends Biochem. Sci.* **15**, 351–354 (1990).
61. Sharp, P. A. On the origin of RNA splicing and introns. *Cell* **42**, 397–400 (1985).
62. Sharp, P. A. 'Five easy pieces'. *Science* **254**, 663. (1991).
63. Bonen, L. & Vogel, J. The ins and outs of group II introns. *Trends Genet.* **17**, 322–331 (2001).
64. Dai, L. & Zimmerly, S. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res.* **30**, 1091–1102 (2002).
65. Mattick, J. S. Introns: evolution and function. *Curr. Opin. Genet. Dev.* **4**, 823–831 (1994).
66. Weiner, A. M. mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell* **72**, 161–164 (1993).
67. Moore, M. J. & Sharp, P. A. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* **365**, 364–368 (1993).
68. Padgett, R. A., Podar, M., Boulanger, S. C. & Perlman, P. S. The stereochemical course of group II intron self-splicing. *Science* **266**, 1685–1688 (1994).
69. Sontheimer, E. J., Gordon, P. M. & Piccirilli, J. A. Metal ion catalysis during group II intron self-splicing: parallels with the spliceosome. *Genes Dev.* **13**, 1729–1741 (1999).
70. Hetzer, M., Wurzer, G., Schweyen, R. J. & Mueller, M. W. Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature* **386**, 417–420 (1997).
71. Parker, R., Siliciano, P. G. & Guthrie, C. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* **49**, 229–239 (1987).
72. Schmelzer, C. & Schweyen, R. J. Self-splicing of group II introns *in vitro*: mapping of the branch point and mutational inhibition of lariat formation. *Cell* **46**, 557–565 (1986).
73. Abramovitz, D. L., Friedman, R. A. & Pyle, A. M. Catalytic role of 2′-hydroxyl groups within a group II intron active site. *Science* **271**, 1410–1413 (1996).

74. Jarrell, K. A., Dietrich, R. C. & Perlman, P. S. Group II intron domain 5 facilitates a trans-splicing reaction. *Mol. Cell. Biol.* **8**, 2361–2366 (1988).
75. Konforti, B. B. *et al.* Ribozyme catalysis from the major groove of group II intron domain 5. *Mol. Cell* **1**, 433–441 (1998).
76. Peebles, C. L., Zhang, M., Perlman, P. S. & Franzen, J. S. Catalytically critical nucleotide in domain 5 of a group II intron. *Proc. Natl Acad. Sci. USA* **92**, 4422–4426 (1995).
77. Yu, Y. T., Maroney, P. A., Darzynkiwicz, E. & Nilsen, T. W. U6 snRNA function in nuclear pre-mRNA splicing: a phosphorothioate interference analysis of the U6 phosphate backbone. *RNA* **1**, 46–54 (1995).
78. Shukla, G. C. & Padgett, R. A. A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol. Cell* **9**, 1145–1150 (2002).
Evidence that the catalytic cores of the group-II self-splicing intron and the major- and minor-class spliceosomes are related was provided by functionally replacing a U6atac stem-loop with D5 of the group-II intron in an *in vivo* splicing assay.
79. Shukla, G. C. & Padgett, R. A. The intramolecular stem-loop structure of U6 snRNA can functionally replace the U6atac snRNA stem-loop. *RNA* **7**, 94–105 (2001).
80. Shukla, G. C. & Padgett, R. A. Conservation of functional features of U6atac and U12 snRNAs between vertebrates and higher plants. *RNA* **5**, 525–538 (1999).
81. Sun, J. S. & Manley, J. L. A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing. *Genes Dev.* **9**, 843–854 (1995).
82. Fabrizio, P. & Abelson, J. Thiophosphates in yeast U6 snRNA specifically affect pre-mRNA splicing *in vitro*. *Nucleic Acids Res.* **20**, 3659–3664 (1992).
83. Fortner, D. M., Troy, R. G. & Brow, D. A. A stem/loop in U6 RNA defines a conformational switch required for pre-mRNA splicing. *Genes Dev.* **8**, 221–233 (1994).
84. Madhani, H. D. & Guthrie, C. Randomization-selection analysis of snRNAs *in vivo*: evidence for a tertiary interaction in the spliceosome. *Genes Dev.* **8**, 1071–1086 (1994).
85. McPheeters, D. S. & Abelson, J. Mutational analysis of the yeast U2 snRNA suggests a structural similarity to the catalytic core of group I introns. *Cell* **71**, 819–831 (1992).
86. Wolff, T. & Bindereif, A. Conformational changes of U6 RNA during the spliceosome cycle: an intramolecular helix is essential both for initiating the U4-U6 interaction and for the first step of splicing. *Genes Dev.* **7**, 1377–1389 (1993).
87. Wolff, T., Menssen, R., Hammel, J. & Bindereif, A. Splicing function of mammalian U6 small nuclear RNA: conserved positions in central domain and helix I are essential during the first and second step of pre-mRNA splicing. *Proc. Natl Acad. Sci. USA* **91**, 903–907 (1994).
88. Parker, R. & Siliciano, P. G. Evidence for an essential non-Watson–Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* **361**, 660–662 (1993).
89. Wassarman, K. M. & Steitz, J. A. The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. *Mol. Cell. Biol.* **12**, 1276–1285 (1992).
90. Frilander, M. J. & Steitz, J. A. Initial recognition of U12-dependent introns requires both U11/5′ splice-site and U12/branchpoint interactions. *Genes Dev.* **13**, 851–863(1999).
Strong cooperativity was documented for binding of the 5′ splice site and branch-point sequence by U11 and U12 snRNPs during pre-spliceosome formation on a U12-type intron, indicating greater rigidity in the intron recognition process for minor- versus major-class splicing.
91. Ruskin, B., Zamore, P. D. & Green, M. R. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**, 207–219 (1988).
92. Frilander, M. J. & Steitz, J. A. Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Mol. Cell* **7**, 217–226 (2001).
93. Hertel, K. J. & Maniatis, T. Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. *Proc. Natl Acad. Sci. USA* **96**, 2651–2655 (1999).
94. Kandels-Lewis, S. & Seraphin, B. Involvement of U6 snRNA in 5′ splice site selection. *Science* **262**, 2035–2039 (1993).
95. Kuhn, A. N., Li, Z. & Brow, D. A. Splicing factor Prp8 governs U4/U6 RNA unwinding during activation of the spliceosome. *Mol. Cell* **3**, 65–75 (1999).
96. Lesser, C. F. & Guthrie, C. Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science* **262**, 1982–1988 (1993).
97. Staley, J. P. & Guthrie, C. An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* **3**, 55–64 (1999).
98. Stephens, R. M. & Schneider, T. D. Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.* **228**, 1124–1136 (1992).
99. Dietrich, R. C., Peris, M. J., Seyboldt, A. S. & Padgett, R. A. Role of the 3′ splice site in U12-dependent intron splicing. *Mol. Cell. Biol.* **21**, 1942–1952 (2001).
100. Kuo, H. C., Nasim, F. H. & Grabowski, P. J. Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* **251**, 1045–1050 (1991).
101. Robberson, B. L., Cote, G. J. & Berget, S. M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84–94 (1990).
102. Wu, Q. & Krainer, A. R. Purine-rich enhancers function in the AT-AC pre-mRNA splicing pathway and do so independently of intact U1 snRNP. *RNA* **4**, 1664–1673 (1998).
Enhancer sequences located in exons were shown to stimulate splicing of U12-type introns, as they do for U2-type introns.
103. Black, D. L. Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* **103**, 367–370 (2000).
104. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
105. Grabowski, P. J. & Black, D. L. Alternative RNA splicing in the nervous system. *Prog. Neurobiol.* **65**, 289–308 (2001).
106. Smith, C. W. & Valcarcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
107. Dietrich, R. C., Shukla, G. C., Fuller, J. D. & Padgett, R. A. Alternative splicing of U12-dependent introns *in vivo* responds to purine-rich enhancers. *RNA* **7**, 1378–1388 (2001).
By constructing artificial minigenes, the authors showed that U12-type introns can participate in alternative splicing and that the splicing pattern can be influenced by exonic enhancer elements.
108. Hastings, M. L. & Krainer, A. R. Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA* **7**, 471–82. (2001).
109. Hastings, M. L., Wilson, C. M. & Munroe, S. H. A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. *RNA* **7**, 859–874 (2001).
110. Lam, B. J. & Hertel, K. J. A general role for splicing enhancers in exon definition. *RNA* **8**, 1233–1241 (2002).
111. Lynch, M. & Richardson, A. O. The evolution of spliceosomal introns. *Curr. Opin. Genet. Dev.* **12**, 701–710 (2002).
A theoretical framework for the phylogenetic and physical distribution of introns is presented, including the idea that the progenitor eukaryote had spliceosomal introns that diverged early on into two distinct classes.
112. Spafford, J. D., Spencer, A. N. & Gallin, W. J. A putative voltage-gated sodium channel α-subunit (PpSCN1) from the hydrozoan jellyfish, *Polyorchis penicillatus*: structural comparisons and evolutionary considerations. *Biochem. Biophys. Res. Commun.* **244**, 772–780 (1998).
113. Spafford, J. D., Spencer, A. N. & Gallin, W. J. Genomic organization of a voltage-gated Na^+ channel in a hydrozoan jellyfish: insights into the evolution of voltage-gated Na^+ channel genes. *Recept. Channels* **6**, 493–506 (1999).
114. Patel, A. A., McCarthy, M. & Steitz, J. A. The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J.* **21**, 3804–3815 (2002).
Data from a quantitative RT-PCR assay showed that U12-type introns are removed more slowly than neighbouring U2-type introns from pre-mRNAs in both human and *Drosophila* cells, and that minor-class introns thereby serve to lower the abundance of the RNA and protein products of a gene.
115. Shukla, G. C., Cole, A. J., Dietrich, R. C. & Padgett, R. A. Domains of human U4atac snRNA required for U12-dependent splicing *in vivo*. *Nucleic Acids Res.* **30**, 4650–4657 (2002).
116. Hirose, Y. & Manley, J. L. RNA polymerase II and the integration of nuclear events. *Genes Dev.* **14**, 1415–1429 (2000).
117. Misteli, T. & Spector, D. L. RNA polymerase II targets pre-mRNA splicing factors to transcription sites *in vivo*. *Mol. Cell* **3**, 697–705 (1999).
118. Proudfoot, N. Connecting transcription to messenger RNA processing. *Trends Biochem. Sci.* **25**, 290–293 (2000).
119. Baserga, S. J., & Steitz, J. A. in *The RNA World* 1st edn (eds Gesteland, R. F. & Atkins, J. F.) 359–381 (Cold Spring Harbor Laboratory Press, New York, 1993).
120. Padgett, R. A. & Shukla, G. C. A revised model for U4atac/U6atac snRNA base pairing. *RNA* **8**, 125–128 (2002).

Acknowledgements

We thank C. B. Burge, R. A. Padgett, R. Luhrmann, M. Frilander, A. Krainer and R. Durbin for sharing unpublished information. We appreciate the critical reading of the manuscript by R. A. Padgett, K. Tycowski, L. Szewczak, T. Hirose and R. Lytle.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

FlyBase: http://flybase.bio.indiana.edu/*prospero*

LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink/Huntingtin | p14 | SF3b | U2AF

Swiss-Prot: http://ca.expasy.org/sprot/Prp8

Access to this interactive links box is free online.