Alternative splicing and evolution: diversification, exon definition and function

Hadas Keren, Galit Lev-Maor and Gil Ast

Abstract | Over the past decade, it has been shown that alternative splicing (AS) is a major mechanism for the enhancement of transcriptome and proteome diversity, particularly in mammals. Splicing can be found in species from bacteria to humans, but its prevalence and characteristics vary considerably. Evolutionary studies are helping to address questions that are fundamental to understanding this important process: how and when did AS evolve? Which AS events are functional? What are the evolutionary forces that shaped, and continue to shape, AS? And what determines whether an exon is spliced in a constitutive or alternative manner? In this Review, we summarize the current knowledge of AS and evolution and provide insights into some of these unresolved questions.

Spliceosome
A ribonucleoprotein complex that is involved in splicing of nuclear precursor mRNA (pre-mRNA). It is composed of five small nuclear ribonucleoproteins (snRNPs) and more than 50 non-snRNPs, which recognize and assemble on exon–intron boundaries to catalyse intron processing of the pre-mRNA.

Nucleosome
The basic unit of chromatin, containing ~147 bp of DNA wrapped around a histone octamer (which is composed of two copies each of histone 3 (H3), H4, H2A and H2B).

Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel.
Correspondence to H.K.
e-mail:
hadasker@post.tau.ac.il
doi:10.1038/nrg2776
Published online 8 April 2010

Splicing of precursor mRNA (pre-mRNA) is a crucial regulatory stage in the pathway of gene expression: introns are removed and exons are ligated to form mRNA. The inclusion of different exons in mRNA — alternative splicing (AS) — results in the generation of different isoforms from a single gene and is the basis for the discrepancy between the estimated 24,000 protein-coding genes in the human genome and the 100,000 different proteins that are postulated to be synthesized[1]. Splicing in general, and AS in particular, is also important for regulation of the levels and tissue specificity of gene expression and, if disrupted, can lead to disease[2–5].

The importance of splicing is emphasized by its presence in species throughout the phylogenetic tree. However, we still have much to learn about how diverse intron–exon structures are generated and recognized. Comparing species to see what has changed and what is conserved is proving valuable in addressing these issues and has recently yielded substantial progress. For example, new high-throughput sequencing technology has revealed that >90% of human genes undergo AS — a much higher percentage than anticipated[6]. Such technological progress is providing more comprehensive studies of splicing and genomic architecture in an increasing number of species, and these studies have extended our evolutionary understanding.

The basis of splicing is the recognition of introns and exons by the splicing machinery (FIG. 1). It can be regulated at many different levels, often in a tissue- or developmental stage-specific manner. At a basic level, regulation includes splice-site recognition by the spliceosome, which is mediated by various proteins. Additional regulatory levels include environmental changes that affect splice-site choice and the relationship among transcription by RNA polymerase II (RNAPII), nucleosome occupancy and splicing[7–10]. New insights into the regulation of AS that have emerged from evolutionary studies include the findings that the origin of an exon can influence how frequently it is spliced into an mRNA and that the evolution of an RNA degradation mechanism might have facilitated intron gain. The definition of alternative exons is also important for understanding the links between splicing and evolution.

In this Review, we look at the relationship between AS and evolution at various levels and consider progress in understanding the biological importance and control of AS. We start by discussing AS from an evolutionary perspective — comparing splicing and genomic architecture among species. We then evaluate the three major ways that alternative exons emerge: exon shuffling, exonization and transition. This is followed by discussion of the characteristics of alternative exons and how they are distinguished from constitutive exons at the RNA and DNA levels, including the recent discovery of a role for chromatin. We conclude with our perspective on the potential direction of future research.

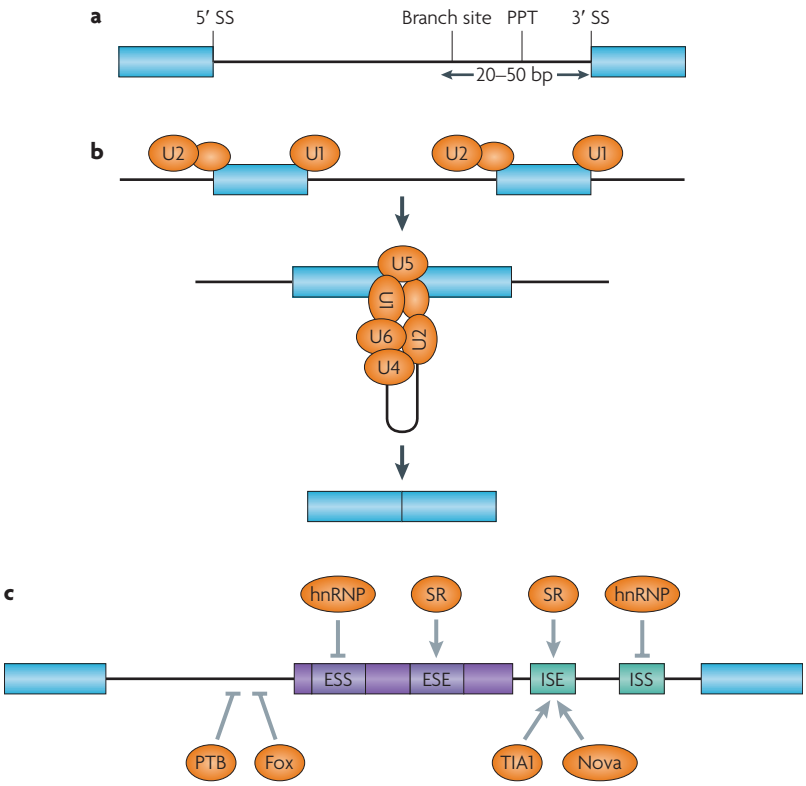


Figure 1 | **The splicing machinery.** Splicing is a conserved mechanism controlled by the spliceosome — a complex composed of many proteins and five small nuclear RNAs (U1, U2, U4, U5 and U6) that assemble with proteins to form small nuclear ribonucleoproteins (snRNPs). **a** | The four conserved signals that enable recognition of RNA by the spliceosome are: the exon–intron junctions at the 5′ and 3′ ends of introns (the 5′ splice site (5′ SS) and 3′ SS), the branch site sequence located upstream of the 3′ SS and the polypyrimidine tract (PPT) located between the 3′ SS and the branch site. **b** | The key steps in splicing are shown. Regulation of splicing can occur at the basic level of splice-site recognition by the spliceosome through the facilitation or interference of the binding of U1 and U2 snRNPs to the splice sites[7]. The unlabelled orange ovals represent other, unspecified components of the spliceosome. **c** | Exons and introns contain short, degenerate binding sites for splicing auxiliary proteins. These sites are called exonic splicing enhancers (ESEs), intronic splicing enhancers (ISEs), exonic splicing silencers (ESSs) and intronic silencing silencers (ISSs). Splice-site recognition is mediated by proteins that bind specific regulatory sequences, such as the serine/arginine (SR) proteins, heterogenous nuclear ribonucleoproteins (hnRNPs), polypyrimidine tract-binding (PTB) proteins, the TIA1 RNA-binding protein, Fox proteins, Nova proteins, and more[7,9,10]. Constitutive exons are shown in blue, alternatively spliced regions in purple, and introns are represented by solid lines.

Changes in alternative splicing during evolution

There are two models for the mechanism of exon and intron selection: intron definition and exon definition. In intron definition, the splicing machinery recognizes an intronic unit and places the basal splicing machinery across introns. This model is regarded as the ancient mechanism and these introns are probably under evolutionary selection to remain short. Exon definition takes place when the basal machinery is placed across exons; this constrains the length of exons. The higher GC content in exons relative to their flanking introns[11] is presumed to be the signal that allows exons to be identified. Exon definition probably evolved later and is considered to be the main mechanism in higher eukaryotes[12,13].

Basal splicing
A conserved mRNA splicing mechanism. It is composed of the splicing signals and the core of the machinery is formed by five spliceosomal small nuclear ribonucleoproteins and an unknown number of proteins.

Prevalence and type of alternative splicing. Splicing is ubiquitous in eukaryotes, but there are few examples of splicing in bacteria[14] and archea[15,16]. The importance of AS can be inferred from inspecting its prevalence throughout the eukaryotic evolutionary tree. AS is more abundant in higher eukaryotes than in lower eukaryotes, and the percentage of genes and exons that undergo AS is higher in vertebrates than in invertebrates[17,18].

There are several types of AS (BOX 1) and the type varies among species. Intron retention is most common in lower metazoans and is also common in fungi and protozoa[19]. The prevalence of exon skipping gradually increases further up the eukaryotic tree[19]. This might suggest that exon skipping is the type of AS that contributes most to phenotypic complexity. Alternative 5′ and 3′ splice sites are believed to be subfamilies of exon skipping and might represent an intermediate evolutionary stage[20]. Plants show low levels of alternatively spliced genes, with a high level of intron retention (~30%) and a very low level of exon skipping (<5%)[19]. This suggests that AS did not have a substantial role in plant evolution, perhaps because plants can enhance their transcriptomic and proteomic diversity by whole-genome duplication events[19].

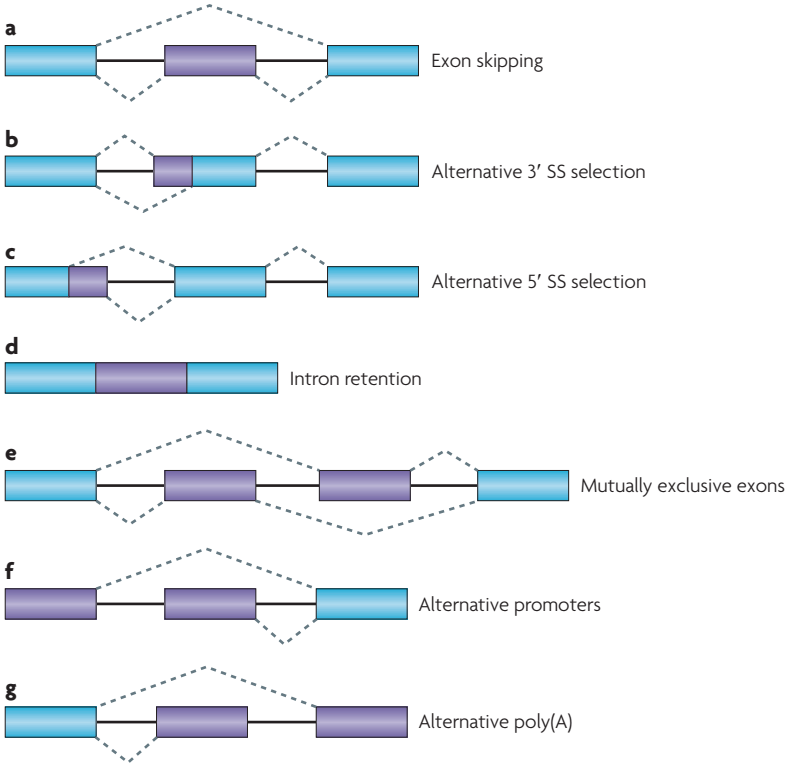
Exon–intron architecture. The relative length of introns and exons — the exon–intron architecture — varies across the eukaryotic kingdom. Intron and exon lengths can reflect the constraints imposed by splicing recognition — for example, based on whether the exon is identified through the intron or exon definition mechanism. In vertebrates, there are relatively long introns and short exons, whereas in lower eukaryotes, introns are short and exons are long.

In *Drosophila melanogaster*, most of the introns flanking alternatively spliced exons are long, whereas constitutively spliced exons are flanked by short introns. The length of the upstream intron was found to have a greater influence on exon selection than that of the downstream intron[21]. This shows a large contribution of the exon–intron architecture in *D. melanogaster* to the frequency of AS. In humans, exons flanked by long introns are subject to exon skipping more often than those flanked by short introns. This correlation was stronger in humans than in *D. melanogaster*, which might be explained by the presence in mammals of factors that regulate splicing that are absent or less important in non-mammalians[21].

An additional illustration of the effect of exon–intron architecture on AS is revealed when the size of mammalian exons is examined. Usually, enlarged exons lead to exon skipping, but if the flanking introns are short, the enlarged exon is included[22]. Also, it seems that exon length has decreased during evolution[11,21].

The number of introns in a genome is determined by the relative rates of intron gain and intron loss over an evolutionary period. Intron gain is believed to be a rare event that has declined over the past 1.3 billion years in most eukaryotes[23]. In contrast to these findings, extensive intron gain was recently observed in *Daphnia pulex*, a water flea[24], and in *D. melanogaster*[25]. Also, several

Box 1 | **Different types of alternative splicing**



There are several different types of alternative splicing (AS) events, which can be classified into four main subgroups. The first type is exon skipping, in which a type of exon known as a cassette exon is spliced out of the transcript together with its flanking introns (see the figure, part **a**). Exon skipping accounts for nearly 40% of AS events in higher eukaryotes[17,111] but is extremely rare in lower eukaryotes. The second and third types are alternative 3′ splice site (3′ SS) and 5′ SS selection (parts **b** and **c**). These types of AS events occur when two or more splice sites are recognized at one end of an exon. Alternative 3′ SS and 5′ SS selection account for 18.4% and 7.9% of all AS events in higher eukaryotes, respectively. The fourth type is intron retention (part **d**), in which an intron remains in the mature mRNA transcript. This is the rarest AS event in vertebrates and invertebrates, accounting for less than 5% of known events[17,19,98,111]. By contrast, intron retention is the most prevalent type of AS in plants, fungi and protozoa[19]. Less frequent, complex events that give rise to alternative transcript variants include mutually exclusive exons (part **e**), alternative promoter usage (part **f**) and alternative polyadenylation (part **g**)[12,19,112]. Another rare form of AS involves reactions between two primary transcripts in *trans*[113] (not shown).

In the figure, constitutive exons are shown in blue and alternatively spliced regions in purple. Introns are represented by solid lines, and dashed lines indicate splicing options.

examples were found in human genes in which insertion of *Alu*, a primate-specific retroelement, into an exon created a new intron in the 3′ UTR[26].

Splicing signals and machinery. Splicing signals are major contributors to evolutionary change and have evolved dramatically[27]. Moreover, there has been a selective expansion of splicing regulatory proteins, such as serine/arginine proteins (SR proteins) in metazoans and heterogenous nuclear ribonucleoproteins (hnRNPs) in vertebrates[28], which may have assisted the basal splicing machinery in finding short exons in large intronic sequences. The ability of the SR proteins to support the basal splicing machinery has enabled alternative isoforms to exist, and the proliferation of SR proteins during evolution has increased the abundance of AS. Recently, it was found that nonsense-mediated decay (NMD) is necessary for normal splicing regulation in *D. melanogaster*, and its emergence presumably enabled intron gain[25].

The common ancestor of alternative splicing. Compared with extant species, early eukaryotic ancestors had high intron densities in their genes[29–35]. This information, together with the observation that ancestral splicing signals are degenerate[27], the presence of complex spliceosomes in lower eukaryotes[29,32,33], the presence of NMD pathways in animals, fungi, plants, excavates and chromalveolates[32,36,37], and the homology of splicing factors in different species[27,28,38,39], suggests that, in terms of splicing, the ancestral eukaryote was probably more similar to the mammalian eukaryote than previously anticipated. AS could even have existed early in eukaryotic evolution. A genome-wide study of 12 eukaryotic genomes revealed similarities among AS patterns in different eukaryotic lineages, which is consistent with an early eukaryotic origin of AS[40].

What is the origin of alternatively spliced exons?

To gain insights into the importance of AS, we need to understand how alternatively spliced exons have evolved and fixated in different genomes. Human–mouse comparative analyses have revealed that AS is often associated with recent exon creation and/or loss[41–43]. We currently know of three different evolutionary mechanisms for the appearance of an alternatively spliced exon: exon shuffling, exonization of intronic sequences and transition of a constitutive exon to an alternative exon.

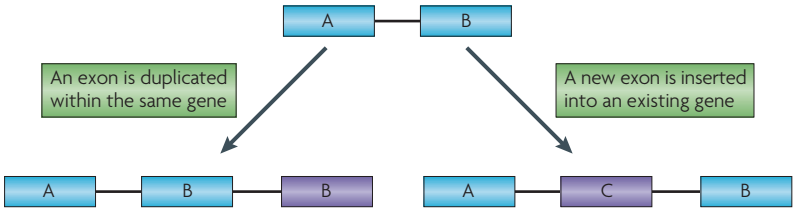
Exon shuffling. Exon shuffling is a process in which a new exon is inserted into an existing gene or an exon is duplicated in the same gene[44–46] (BOX 2). The exon shuffling theory was first proposed by Walter Gilbert in 1978, who suggested that shuffling of exons creates a new chimeric protein that gives an evolutionary advantage to the organism[45]. Many studies have attempted to prove or disprove this theory. In concordance with the exon shuffling theory, a correlation has been found between borders of exons and protein domains[47–49]. In a study of nine species, including invertebrates and vertebrates, the more complex organisms showed a stronger exon–domain correlation. This study suggests that exon shuffling has been common in eukaryotic evolution and has contributed substantially to the complexity of the proteome[49].

Recent analysis of exons that are conserved between humans and mice revealed that newly duplicated exons tend to preserve the splicing status of their original exons: that is, they remain alternatively or constitutively spliced[50]. Therefore, AS seems to have preceded tandem duplication, so duplication propagates rather than creates AS[50].

Tandem exon duplication was found to be the origin of ~10% of cases of substitution alternative splicing, and this duplication was dated to the radiation of mammalian

Alu
An interspersed DNA sequence of 300 bp that belongs to the short interspersed element (SINE) family and is found in the genome of primates. *Alu* elements are composed of a head-to-tail dimer in which the first monomer is 140 bp long and the second is 170 bp long. In humans, there are ~1.1 million copies of *Alu* elements, of which ~500,000 copies are located in introns.

Box 2 | **Exon shuffling**



Exon shuffling creates new combinations of exons by intronic recombination — referred to as illegitimate recombination (IR) — which is recombination between two non-homologous sequences or between short homologous sequences that induce genomic rearrangements[114,115]. Over 30% of unequal homologous recombination is thought to occur through crossovers between *Alu* elements[115]. A possible mechanism for exon shuffling is referred to as the 'modularization hypothesis'. The mechanism includes the insertion of introns at positions that correspond to the boundaries of a protein domain, tandem duplications resulting from recombination in the inserted introns, and the transfer of introns to a different, non-homologous gene by intronic recombination. These three stages were reported for a variety of domains, such as the EGF-like domain and the C-type lectin domain[48]. Exon shuffling involves modules with introns of the same phase class (their position relative to the reading frame of the gene) at both their 5′ and 3′ ends. The insertion of introns has to be in the same phase class or the recombination will cause a shift in the reading frame and lead to loss of protein information[48,55,116]. The mechanism of exon shuffling can also be deduced from side products of DNA transfections in cell culture, which mimics exon shuffling[114,117].

In the figure, constitutive exons are shown in blue, alternatively spliced regions are shown in purple and introns are represented by solid lines.

orders or even the radiation of vertebrate classes[46]. Letunic *et al.* found that ~10% of the genes in humans, flies and worms contain tandemly duplicated exons. In 60% of cases, mutually exclusive alternative splicing of the duplicated exons is likely, which enables modification of protein activity[51]. Competition between two exons is restricted to a distance of less than 70 bp[52]. Another study found a correlation between intron phases — the position of the intron in a codon — and estimated that 19% of eukaryotic intron-containing genes contain shuffled exons[53].

It seems that the importance of exon shuffling increased with the evolution of complex genomes. As the genome increases in size, number of introns and proportion of repetitive elements, the chances of exon shuffling by intronic recombination increase. The evolutionary distribution of modular proteins suggests that exon shuffling became important when multicellular organisms appeared. Therefore, as exon shuffling appeared at a relatively late stage in evolution, it could not have had a major role in the construction of ancient proteins[54,55]. This analysis also suggested that exon shuffling might have contributed to the rapid metazoan radiation[54], which is consistent with the observation that almost all examples of modular proteins are in metazoans and not in bacteria, archaea, plants, protists or fungi[54]. Most modular proteins produced by exon shuffling are associated with multicellularity, such as the extracellular matrix membrane-associated proteins that mediate cell–cell and cell–matrix interactions and the receptor proteins that regulate cell–cell communications. All such proteins are essential for the organism to function as an integrated unit[54].

The diversity that can be generated by duplicated exons is strikingly shown by the Down syndrome cell adhesion molecule (*Dscam*) gene in *D. melanogaster*. The multiple, mutually exclusive exons of *Dscam* lead to enormous numbers of splice variants[56]. Recently, it was found that in humans, tandem repeats can modify the structure of genes in segmental duplications, which influences their coding sequence, splicing pattern and tissue expression[57].

Exonization. A way to gain an exon 'out of nothing' is through genomic sequences becoming exons. Exonization was first suggested as a mechanism for the acquisition of the 5′ region of the bovine thyroglobulin (*TG*) gene[58]. Since then, many exonization events have been reported in humans[59–61], as well as in other vertebrate genomes[17,62,63]. Recently, exonization outside vertebrates was documented in *D. melanogaster* in the RNA-binding bruno 3 (*bru3*) gene[64]. Exonization may also have occurred in sugar receptor genes in insects[65] and in the 5S ribosomal RNA genes in plants[66].

About half of the human genome is derived from transposable elements (TEs)[67], and these repeat-forming elements — particularly *Alu* elements (BOX 3) — can become exonized. About 4% of human genes contain TE motifs in their coding regions, which suggests an exonization event[19,60,61]. The exonization process can be tissue- or tumour-specific, and several genes that include TE sequences have tissue- or tumour-specific isoforms[68–70]. Interestingly, exonization of TEs is observed in 53% of 'orphan genes', which shows the involvement of TE exonization in species-specific adaptive processes[71]. To date, we know of two main mechanisms that create a true 3′ or 5′ splice site, which can lead to spliceosome recognition and exonization: random mutations in intronic sequences and RNA editing. Both mechanisms have acted on TEs in mammals, especially in *Alu* elements. The molecular mechanisms that lead to the exonization of *Alu* elements have been studied in detail[72–74] and are discussed further in BOX 3.

The formation of alternative exons from *Alu* elements permits new functions to be established without eliminating the original function of a protein[75]. In some cases, the insertion of *Alu* into an exon or the formation of a constitutive exon from *Alu* can be deleterious and can lead to human genetic diseases. However, most constitutively spliced *Alu* exons have been inserted into UTRs and therefore do not affect the protein[70]. An interesting case of selection against exonization in single copy genes was identified in *TIF1A* (also known as *TRIM24*), which underwent triplication: *Alu* exonization only occurred in one of the non-coding transcripts or in cancerous cells[68].

Whether an *Alu*-derived exon is included in an mRNA is influenced by various factors. The 5′ and 3′ splice sites are particularly important: in general, *Alu* exons are flanked by stronger splice sites than other exons[76]. At the 3′ side of the exon, two adjacent AG dinucleotides compete with each other; only one is selected as the 3′ splice site and the other suppresses the selection of that exon, leading to sub-optimization of that site and therefore

Retroelement
A mobile genetic element. Its DNA is transcribed into RNA, which is reverse-transcribed into DNA and then inserted into a new location in the genome.

Serine/arginine proteins
A group of highly conserved serine- and arginine-rich splicing regulatory proteins in metazoans.

Heterogenous nuclear ribonucleoproteins
A large set of proteins that bind the precursor mRNA and regulate splicing.

Nonsense-mediated decay
The process by which the cell destroys mRNAs that are untranslatable due to the presence of a premature stop codon in the coding region.

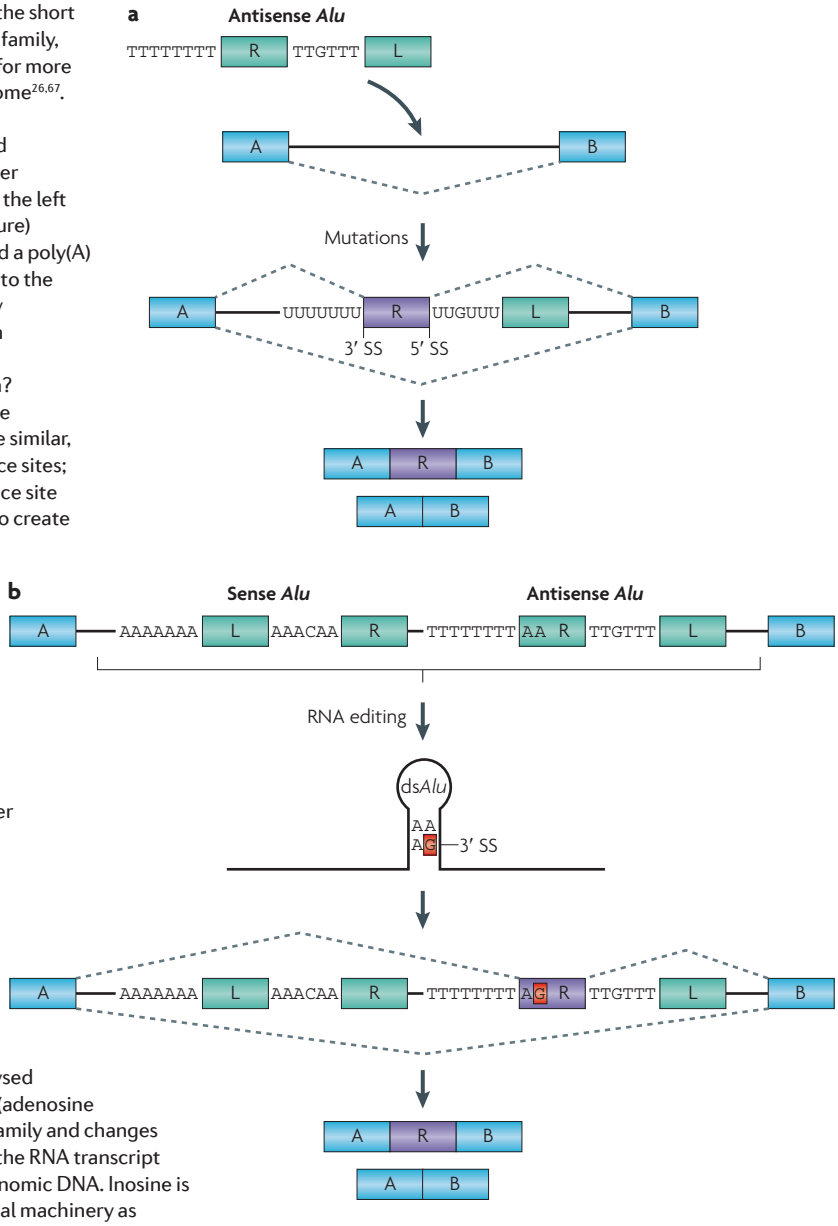
Excavates
A major kingdom of unicellular eukaryotes, often known as Excavata. The phylogenetic category Excavata contains a variety of free-living and symbiotic forms, and also includes some important parasites of humans.

Box 3 | **Exonization of *Alu* elements**

The *Alu* element belongs to the short interspersed element (SINE) family, and *Alu* sequences account for more than 10% of the human genome[26,67]. A typical *Alu* is around 300 nucleotides (nt) long and contains two similar monomer segments (the right arm and the left arm, green R and L in the figure) joined by an A-rich linker and a poly(A) tail-like region. *Alu*s insert into the introns of primate genes by retrotransposition, usually in the antisense orientation.

So how do *Alu* exons form? The consensus *Alu* sequence carries multiple sites that are similar, but not identical, to real splice sites; a few mutations in the 3′ splice site (3′ SS) or 5′ SS are required to create a new exon[41,74,77] (see the figure, part **a**). 85% of exonizations occur from the right arm in the antisense orientation[26,60]. The poly(A) tract of this arm in the antisense orientation creates a strong polypyrimidine tract (PPT). Downstream from this PPT a 3′ SS is selected, and further downstream from that site (approximately 120 nt) a 5′ SS is recognized[39].

A second mechanism for creation of splice sites relies on RNA secondary structure, and involves the enzymatic editing of adenosine (A) to inosine (I) (part **b**). This editing is catalysed by enzymes from the ADAR (adenosine deaminase acting on RNA) family and changes the nucleotide sequence of the RNA transcript from that of the encoded genomic DNA. Inosine is recognized by most biological machinery as guanosine (G)[118]. Two adjacent intronic *Alu* sequences in opposite orientations, sense and antisense, can form a dsRNA structure that serves as a template for the ADAR enzyme, which can create a functional 3′ SS (AA to AG) by RNA editing[119].



An example of an *Alu* exon that is exonized through RNA editing is exon 8 of nuclear prelamin A recognition factor (NARF)[120,121]. It is important to note that in this exon, RNA editing also eliminates a stop codon, therefore allowing the newly added *Alu* exon to maintain the reading frame. As editing is higher in certain tissues, such as the brain, RNA editing modulates the inclusion level of this exon in a tissue-specific manner. RNA editing was suggested to be a major contributor to the evolution of phenotypic complexity in mammals, particularly in the brain[122].

The reason for the high exonization level of *Alu*s is that they are the only transposable elements with two arms that originate from the same sequence and therefore share high sequence identity. The left *Alu* arm functions as a pseudoexon that competes with the right arm for the affinity of the splicing machinery. Deletion of the left arm or insertion of a spacer of more than 150 nt between the two arms shifts splicing from alternative to constitutive inclusion. Also, insertion of the left arm (which is not exonized) downstream, but not upstream, to a constitutive exon shifts the splicing pattern to alternative, indicating that this *Alu* arm functions as a pseudoexon[52]. The monomeric form of *Alu* — called B1 — exists in rodents, but its level of exonization is much lower than that of *Alu* elements in humans (0.07% compared with 0.2%, respectively)[26]. We assume that during rodent evolution these B1 elements were exonized, but such exonization events were mostly selected against as they seem to be constitutively spliced.

In the figure, constitutive exons are shown in blue, alternatively spliced regions are shown in purple, introns are represented by solid lines and dashed lines indicate splicing options.

Chromalveolates
A hypothetical 'supergroup' of protists, including apicomplexa, dinoflagellates, ciliates, heterokonts, haptophytes and cryptomonads, all of which are suggested to have diverged from an ancient common ancestor that acquired a plastid by secondary endosymbiosis with a red alga.

Substitution alternative splicing
An alternative splicing pattern in which one of two amino acid sequences is included in the protein.

Mutually exclusive alternative splicing
Only one of a set of two or more exons in a gene is included in the final transcript.

Modular proteins
Proteins created by intronic recombination. According to the exon shuffling theory, each exon encodes a single protein domain (a 'module'), and the process of shuffling creates a new chimeric protein from the combination of domains (or 'modules').

Transposable elements
Segments of genetic material that are capable of changing their location in the genome of an organism.

Orphan genes
Genes that do not share any homology with genes from other species.

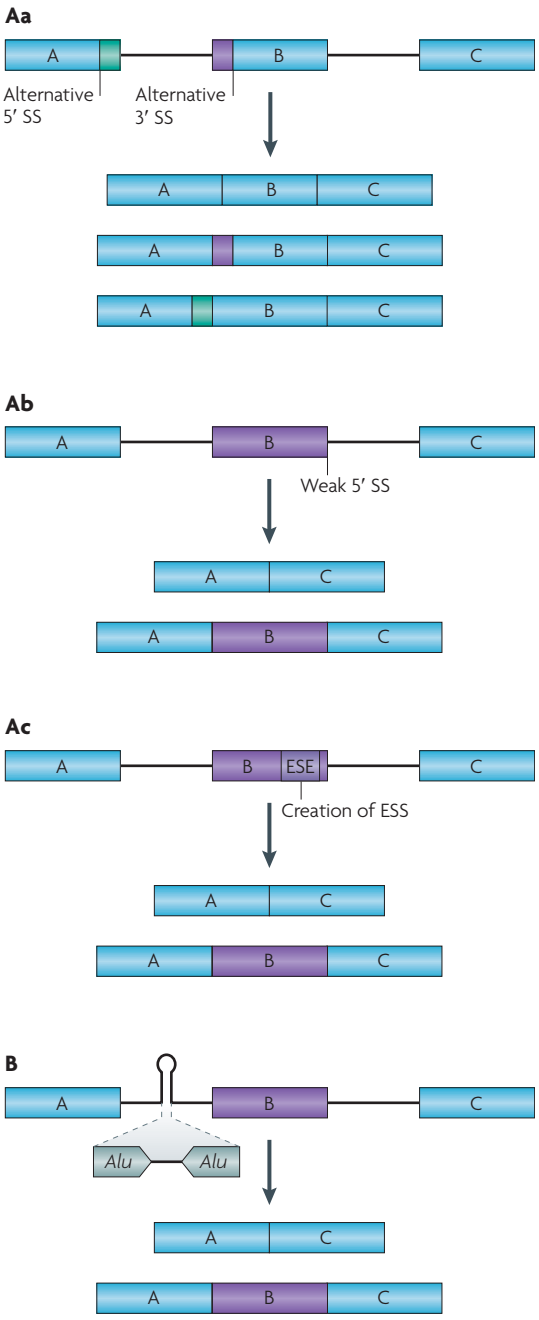


Figure 2 | **Transition from constitutive to alternative splicing.** There are two mechanisms by which a constitutive exon can become an alternative exon. **A** | Mutations that lead to suboptimal recognition of the exon and result in exon skipping. **Aa** | Mutations can lead to a new alternative 5′ splice site (5′ SS) or 3′ SS. **Ab** | Mutations can lead to suboptimal recognition of the 5′ SS. **Ac** | Mutations in exons (or introns) can disrupt an exonic splicing enhancer (ESE) (or intronic splicing enhancer (ISE)) or may create an exonic splicing silencer (ESS) (or intronic silencing silencer (ISS)). **B** | A secondary structure, usually formed between two *Alu* elements in opposite orientation, can interrupt exon recognition. The resulting isoforms are represented for each pathway. Constitutive exons are shown in blue, alternatively spliced regions are shown in purple, and introns are represented by solid lines.

Splicing regulatory elements
Specific *cis*-acting RNA sequence elements that are present in introns or in exons. They are bound by *trans*-acting splicing regulatory proteins (repressors and activators), which regulate alternative splicing.

to exon skipping[41]. At the 5′ side, a strong 5′ splice site leads to full exon inclusion and a weak one to full exon skipping. Intermediate 5′ splice site strengths lead to different levels of exon inclusion[77]. A cryptic 5′ splice site located downstream of the *Alu* exon can enhance its selection through a process involving the binding of U1 small nuclear RNA (snRNA) to the cryptic splice site[78]. Splicing regulatory elements can also influence the selection of *Alu* exons. For example, the presence of specific exonic splicing regulatory elements (ESRs) in the *Alu* element determines specific splice-site selection in *Alu* exons[78]. Also, *Alu* exons are enriched in exonic splicing enhancers (ESEs)[76] and depleted in exonic splicing silencers (ESSs)[76,79]. The exon–intron architecture also has a large impact on the selection of *Alu* exons: *Alu* exons are approximately 10 nucleotides longer than non-exonizing ones and are flanked by introns that are almost 50% shorter[76]. The selection of an *Alu* exon is also affected by the flanking genomic sequence (that is, different genomic environments lead exons to differ in their susceptibility to exonization)[78]. Another novel feature of *Alu* exons is their secondary structure: *Alu* exons in general, and their 5′ splice sites in particular, have a less stable secondary structure than non-exonizing exons[76].

The importance of exonization is reflected through its exaptation during evolution to provide important cellular functions. An example of this is the human cathelicidin antimicrobial peptide (*CAMP*) gene — this gene is involved in innate immunity in humans and primates, and an ancient *Alu* insertion caused it to be regulated by the vitamin D pathway[80]. Another example is surfactant protein B (*SFTPB*), in which a TE insertion resulted in a lung tissue-specific expression[81].

Transition. In the two mechanisms described above, the alternatively spliced exons are generated *de novo*. In a third mechanism, transition, alternative cassette exons are derived from constitutively spliced ones (FIG. 2). Transition can be accomplished by two mechanisms: accumulation of mutations in the splice sites or in ESRs, which leads to suboptimal recognition of the exon and hence its skipping; and formation of a dsRNA secondary structure from two *Alu*s in opposite orientation to each other in the upstream exon, which influences the downstream exon and changes its mode of splicing.

Conserved exons can maintain different modes of splicing in different species[82]. Reconstruction of the evolution of such exons has revealed that, in the specific sets of exons tested, all were constitutively spliced in the common ancestor and shifted their mode of splicing from constitutive to alternative cassette exons during evolution. This shift was usually due to mutations that reduced the affinity of U1 snRNA binding to the 5′ splice site. This shift was also found to be associated with the fixation of ESRs that control the alternative exon inclusion level[12,83].

Alternative exons of the alternative 3′ or 5′ splice-site selection type can also originate from ancestral constitutive exons. Indeed, it was shown that mutations inside an exon or in flanking introns are responsible for the creation of new splice signals that compete with the authentic ones, leading to alternative splice-site selection[20].

Negative selection pressure on these new splice sites causes them to be the minor splice sites; the ancestral site remains the major splice site, so the original protein is not disturbed. To maintain the reading frame, the distance between the two alternative splice sites tends to be divisible by three.

Mutations in putative ESRs can also shift the splicing pattern from constitutive to alternative. In a minigene model system, ESRs affected splicing only when the recognition of the exon was suboptimal[83]. Strong negative selection was found for synonymous substitutions that disrupt predicted ESEs or create predicted ESSs[84].

Introns flanking alternatively spliced exons tend to contain more *Alu* sequences than constitutively spliced ones, and this is also true of exons that have changed their mode of splicing from constitutive to alternative during human evolution[85]. Furthermore, intronic *Alu*s probably affect the selection of constitutive exons. Several reports indicate that *de novo Alu* insertion into intronic sequences in the antisense orientation and in close proximity to the adjacent exon leads to exon skipping or AS, as shown in endogenous genes in the context of disease[86,87] and in a minigene model system[85]. This regulation probably involves the formation of a double-stranded region between two *Alu*s or regulatory sequences in the antisense *Alu* that act as intronic splicing silencers (ISSs)[85].

In summary, the pathways that lead to the creation of alternative exons show the complexity of AS regulation and the role of AS in transcriptome enrichment.

Conservation and function of alternative exons

There has been much discussion of what proportion of alternative transcripts are functional[88]. So how can we predict whether an AS event confers a function? Conservation of a specific AS pattern throughout evolution provides strong evidence of biological function, as a non-functional isoform is likely to be subject to negative selection.

Alternative exons that are conserved between humans and mice are enriched in genes expressed in the brain and in genes involved in transcriptional regulation, RNA processing and development[89]. But, despite the importance of AS conservation, the fraction of alternatively spliced exons that is conserved is smaller than the fraction of constitutive exons, and no case of conservation of a specific AS pattern throughout eukaryotic kingdoms has been found[32,90]. Nonetheless, it is possible that conserved alternative transcripts are difficult to detect owing to high rates of change of AS patterns and/or turnover of different mechanisms[32].

Another important indication of functionality is that the alternative exon can be divided by three (that is, the exon is symmetrical). Symmetrical alternative exons preserve the reading frame of the protein, do not introduce premature stop codons and do not tend to disrupt protein domain structures. There is a very low inclusion level of alternative exons that are non-symmetrical, suggesting that such transcripts are non-functional, have low stability or are degraded by the NMD pathway. Non-symmetrical exons that are conserved tend to reside in the 5′ end of the coding sequence, presumably increasing the potential of the transcript to activate the NMD mechanism[91]. However, a low inclusion level is not indicative that the alternative product is not functional. For example, the low skipping level in the solute carrier family 35, member B3 (*SLC35B3*) gene[83] is conserved in all mammals that have been tested, which suggests functionality. It is also possible that a high rate of new transcript generation results in a higher rate of emergence of functional alternative transcripts[90,92].

A new alternatively spliced isoform can lead to the generation of a novel protein, possibly harbouring different domains to the original transcript. Another option is that the new mRNA isoform has a regulatory role[18] through balancing levels of mRNAs that produce functional proteins relative to levels of mRNAs that produce non-functional proteins[19]. An alternative isoform can also be the result of stochastic noise in the splicing machinery[93], and a non-functional AS isoform can have the potential to be a useful splicing isoform. A possible model might be that a new isoform can acquire a function through several steps. Perhaps an isoform is first created through mutations and has no apparent function, but the low abundance of this isoform ensures that it does not harm the cell. If this isoform causes a deleterious effect, it is eliminated through purifying selection. If the isoform is relatively inert, the cell will 'tolerate' its presence. With time, this isoform might accumulate mutations, without altering the activity of the original isoform. In fact, exons of low inclusion level are associated with increased evolutionary changes[90]. If the new transcript acquires a function, random mutations that strengthen its regulatory sequences (splice sites, for example) will increase its inclusion level or give it tissue-specific attributes. The final fate of the isoform depends on the benefit of the new function.

Defining alternative exons

Genomic features of alternatively spliced exons. To understand how splicing affects genome evolution and how sequences evolve to become alternative exons, one needs to understand what defines an alternative exon. Different evolutionary constraints act on alternative cassette exons and constitutive exons. As exon skipping is the most prevalent type of AS in higher eukaryotes, much analysis has been done to understand its regulation. For example, bioinformatics has been used to distinguish alternative cassette exons from constitutive exons. Interspecies comparative analysis is a useful tool for evaluating the importance of different elements because conserved sequences are under purifying selection and therefore conservation presumably indicates important function[12]. Also, comparative studies help us to identify which factors are involved in exon definition.

Human–mouse comparative genomic analysis of constitutive and alternative cassette exons reveals that alternatively spliced exons are under different selective pressures from constitutive ones and that various features distinguish the two groups[18,19,94–96]. Firstly, the sequences of alternatively spliced cassette exons are more conserved than constitutive exons. The conservation is higher towards the edges of the exon; these sequences presumably direct the basal splicing machinery to the

Purifying selection
Selection against deleterious alleles that arise in a population, preventing their increase in frequency and assuring their eventual disappearance from the gene pool.

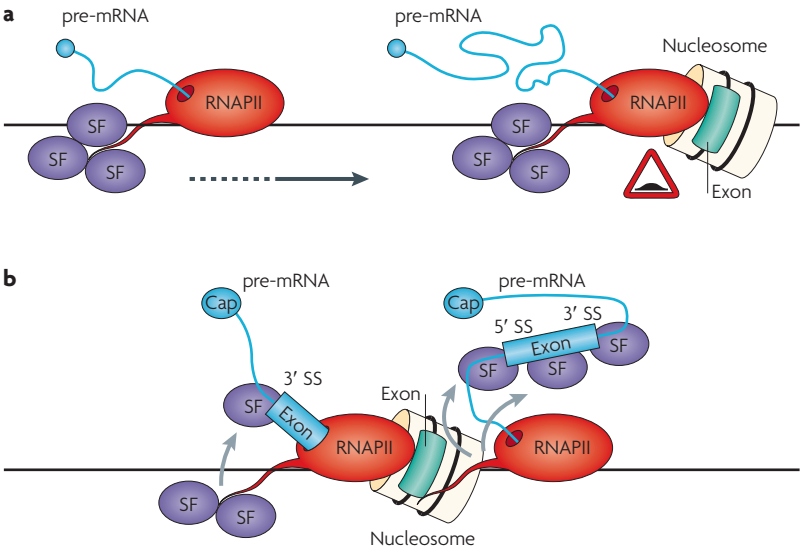


Figure 3 | **The 'speed bump' model.** Nucleosome occupancy marks exons and is coupled to transcription. **a** | RNA polymerase II (RNAPII), associated with different splicing factors (SFs), travels along the gene and transcribes it. When RNAPII reaches an area with high nucleosome occupancy and encounters specific histone modifications that mark an exon, it is slowed down. **b** | This panel shows RNAPII and the nucleosome at the point at which their coupling marks the exon boundaries for the splicing machinery. RNAPII transcribes the exon and SFs detach from the carboxy-terminal domain of RNAPII and bind to the 3′ splice site (3′ SS) region of the precursor mRNA (pre-mRNA). During transcription elongation, additional SFs bind intronic and exonic splicing regulatory elements and the 5′ SS.

correct exon–intron junctions. Conservation extends 80–100 bases into the introns flanking the alternatively spliced exons. The ratio of non-synonymous mutations (Ka) to synonymous mutations (Ks) is higher in alternatively spliced exons than in constitutively spliced exons. This is thought to be due to low Ks in cassette exons as a result of their high conservation. However, this is still a topic of ongoing research[95,97].

Secondly, ESRs, which can act as enhancers or suppressors, are substantially more conserved in alternatively spliced exons than in constitutively spliced ones. This reflects the reliance of cassette exons on ESRs for exon selection. Cassette exons also usually have weaker splice sites than constitutive exons. Weak splice sites are suboptimal sites for the splicing machinery, which determines the level of inclusion or skipping in AS. Also, alternative cassette exons are usually shorter than constitutively spliced ones and alternative cassette exons are flanked by longer introns. Long intronic sequences that flank short cassette exons presumably obstruct the recognition of these exons by the splicing machinery. Also, as noted above, there seems to be a stronger evolutionary selection for symmetrical cassette exons compared with symmetrical constitutive exons. The characteristics that distinguish conserved alternative exons from constitutive ones are likely to point to factors that regulate their mode of splicing.

These characteristics are also related to types of AS other than exon skipping: alternative conserved exons of the 3′ and 5′ splice-site selection types have similar characteristics to alternative cassette exons at the side of the exon that is subject to AS. The 'alternative side' of these exons has weak splicing signals and the flanking introns are highly conserved. Suboptimal recognition of the splicing machinery, for example by weaker splicing signals, is partially compensated for by additional splicing signals, such as ESRs in the flanking introns. The constitutively selected side of alternative 3′ or 5′ splice-site exons is similar to constitutive exons: they have strong splicing signals and lower conservation of the flanking introns. This suggests that alternative splice site use is an intermediate between constitutive and alternative cassette exons[20].

The least common form of AS in higher eukaryotes, intron retention, involves weaker splice sites, shorter introns, a higher level of expression and lower density of splicing enhancers[98] compared with non-retained introns. Intron retention might therefore reflect mis-splicing owing to the failure of an intron definition mechanism, as the splicing machinery fails to recognize weak splice sites flanking short introns[12].

An additional factor that has an effect on the selection of exons is the polypyrimidine tract (PPT) sequence. Recent research indicates that the selection of exons is enhanced by the presence of a strong PPT. However, expansion of the exon selection process is restricted by the presence of a second PPT located further downstream. For example, in *Alu* elements, an internal PPT sequence separates the two *Alu* arms[39]. Therefore, the PPT has an effect on the selection of exons at both the RNA level, at which it serves as a binding site for splicing factors and limits the exonization process, and at the DNA level, at which it possesses characteristics that disfavour nucleosome positioning[11].

Higher-order features. Recently, it has become apparent that chromatin structure is another factor that influences the decision as to whether an exon is alternatively or constitutively spliced. Recent studies have found that exons have increased nucleosome occupancy levels compared with introns and that exons are bound by histones enriched in certain modifications[11,99,100].

Nucleosome positioning can affect the selection of exons through two possible scenarios. Firstly, the nucleosome might act as a 'speed bump' on the exon (FIG. 3), which slows RNAPII elongation and leads to increased inclusion of that exon. Indeed, in humans, nucleosome occupancy levels correlate with inclusion levels: introns have the lowest nucleosome occupancy, followed by alternative exons that are rarely included, followed by alternative exons that are frequently included, with constitutive exons having highest occupancy[11]. Furthermore, nucleosomes are depleted from pseudoexons[99]. The speed bump model is strongly supported by a recent study showing that the nucleosome behaves as a fluctuating barrier that results in pausing of RNAPII[101].

Another possibility is that nucleosomes in exons have a specific set of histone modifications that lead to interaction with the splicing machinery and enable more efficient recognition of the exon[11,99]. In *Caenorhabditis elegans*, exons are preferentially marked, relative to introns, with the chromatin modification histone 3 lysine 36 trimethylation (H3K36me3)[102]. A peak

Pseudoexons
Precursor mRNA sequences that resemble exons — both in their size and in the presence of flanking splice-site sequences — but that are not normally recognized by the splicing machinery.

of H3K36me3 was also observed in human exons[11,99]. In addition, human exons are enriched in histones that are modified with three other post-translational modifications: H3K79me1, H4K20me1 and H2BK5me1 (REF. 11). However, this enrichment probably reflects the higher level of nucleosome occupancy[11,99]. Very recently, a direct link was found between histone modifications and AS: the modulation of histone modifications resulted in splice-site switching[103]. The preferential nucleosome occupancy of exons is conserved through evolution, as shown by the analysis of seven organisms[11]. Intriguingly, nucleosome enrichment at exons was also observed in human sperm and in medaka (Japanese killifish)[104], indicating that the preference of nucleosomes for exons is present in diverse species.

It is worth noting that the average length of metazoan exons (125–165 bp) is similar to the length of DNA that wraps around a nucleosome (147 bp), which suggests that nucleosome occupancy might confer purifying selection on exon length[11,99]. However, the length of an average human exon is only 126 bp[11]. What might be the reason for this shorter length? There are several nucleosome structures, each with unique properties and specific role in chromatin function[105]. One structure is the subnucleosome, which is the (H3–H4)2 histone tetramer, also known as the tetrasome. The di-tetrasome has an important role *in vivo* in nucleosome dynamics, transcription and replication[106,107]. The tetrasome is wrapped with ~120 bp of DNA[105,106], which is close to the average 126-bp human exon. It can be speculated that human exons require tighter regulation than those of other species, so the dynamic tetrasome is recruited and enables AS regulation through transcriptional control. Future experiments will be necessary to test these suggestions.

What drives nucleosomes to exons? Exons contain a higher GC content than flanking intron sequences, and regions with higher GC content favour nucleosome occupancy sequences. By contrast, the intron sequences that immediately flank exons (such as the PPT) contain sequences that disfavour nucleosome occupancy[11]. Therefore, these adjacent favourable and disfavourable sequences would cause a well-defined 'exonic mononucleosome'. What might be the biological advantages of an 'exonic nucleosome'? It is possible that nucleosomes inhibit recombination or protect the DNA from UV irradiation and other lesions, which would lead to a reduced mutation rate and therefore increase the conservation of exonic sequences[108]. Also, the 'exonic nucleosome' might mark the location of exons for the splicing machinery. This might have caused the splicing machinery to shift from intron to exon definition during evolution — that is, from recognizing short introns as the spliced units in lower eukaryotes to recognizing short exons in higher eukaryotes. Such a model could explain why introns have become longer over the course of evolution, especially during mammalian evolution. Overall, nucleosome positioning in exons seems to encourage the proper location of molecular interactions across the exon, which contributes to the exon definition mechanism and suggests another level of complexity in eukaryotic splicing regulation.

Conclusions

Over the past decade it has become clear that AS is a key process that contributes to the creation of phenotypic complexity among higher eukaryotes by increasing transcriptomic and proteomic diversity, and the prevalence of AS throughout evolution shows its importance. In 2004, we suggested that AS might have originated from multi-intron genes with no AS, through DNA mutations that weakened splice sites[12]. If early eukaryotes had strong splice sites, constitutive splicing would have taken place, whereas if splice sites were weak, AS was likely to have occurred. Surprisingly, reconstruction of the evolution of these signals has shown that the 5′ splice site of early eukaryotes was degenerate, not conserved[27,40], which implies the creation of alternatively spliced exons early in evolution. This suggests that AS emerged together with the common ancestor of eukaryotes.

Evolutionary studies, which have revealed the formation of *de novo* alternative exons and the evolution of exon–intron architecture, highlight the importance of AS in the diversification of the transcriptome, especially in humans. These studies also show the endless 'attempt' of genomes to search for new functional products. Moreover, they have revealed how exons and introns are defined and the dynamics of this definition process during evolution. The evolution of gene structure is coupled to the evolution of *trans*-acting splicing factors and their respective binding sites. In general, it was shown that *trans*-acting splicing factors (such as SR and hnRNP proteins) provide plasticity, as their binding sites and binding affinity are quite diverse. The manner in which these factors operate together to regulate tissue- and developmental stage-specific AS is slowly becoming clearer, and will probably be understood in the next few years. Due to these advances, the regulation of AS is being widely used in gene therapy, revealing new therapeutic targets[109,110].

Considerable progress was made over the past decade in finding the origin of alternative exons. The potential of an alternative exon to generate a functional transcript has begun to be uncovered. Evolutionary work has suggested that a new transcript is tested after it is first generated, but there is no pre-screening for the selection of those transcripts that will eventually generate a functional product. Out of the new transcripts, presumably only a small fraction will eventually gain functions. The evolutionary forces that select functional transcripts ensure that the new transcript maintains the original coding sequence without inserting a premature stop codon. Once a new functional transcript is established, its inclusion level increases. The functionality of transcriptomes is a crucial issue. It is still unknown which transcripts will synthesize a protein and what regulatory role other transcripts will have as RNA molecules. The next decade will probably be devoted to functional transcriptomic and proteomic analyses linking the suggested new transcripts with their emerging new roles in the cell. Such information is crucial if we are to better understand AS and its contribution to the unique traits that make us human.

1. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature Genet.* **30**, 13–19 (2002).
2. Tazi, J., Bakkour, N. & Stamm, S. Alternative splicing and disease. *Biochim. Biophys. Acta* **1792**, 14–26 (2009).
3. Wang, G. S. & Cooper, T. A. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Rev. Genet.* **8**, 749–761 (2007).
4. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Rev. Genet.* **3**, 285–298 (2002).
5. Venables, J. P. Aberrant and alternative splicing in cancer. *Cancer Res.* **64**, 7647–7654 (2004).
6. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
7. Chen, M. & Manley, J. L. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nature Rev. Mol. Cell Biol.* **10**, 741–754 (2009).
8. Hartmann, B. & Valcarcel, J. Decrypting the genome's alternative messages. *Curr. Opin. Cell Biol.* **21**, 377–386 (2009).
9. Hui, J. Regulation of mammalian pre-mRNA splicing. *Sci. China C Life Sci.* **52**, 253–260 (2009).
10. Licatalosi, D. D. & Darnell, R. B. RNA processing and its regulation: global insights into biological networks. *Nature Rev. Genet.* **11**, 75–87 (2010).
11. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon–intron structure. *Nature Struct. Mol. Biol.* **16**, 990–995 (2009).
This paper shows that exons have increased nucleosome occupancy levels compared with introns, and four specific post-translational histone modifications are enriched in exons. This article, together with Tilgner *et al.* and Andersson *et al.* (references 98 and 99, respectively), presents evidence that the positioning and modifications of nucleosomes might help to define the exon–intron architecture of genes.
12. Ast, G. How did alternative splicing evolve? *Nature Rev. Genet.* **5**, 773–782 (2004).
13. Ram, O. & Ast, G. SR proteins: a foot on the exon before the transition from intron to exon definition. *Trends Genet.* **23**, 5–7 (2007).
14. Edgell, D. R., Belfort, M. & Shub, D. A. Barriers to intron promiscuity in bacteria. *J. Bacteriol.* **182**, 5281–5289 (2000).
15. Watanabe, Y. *et al.* Introns in protein-coding genes in archaea. *FEBS Lett.* **510**, 27–30 (2002).
16. Yokobori, S. *et al.* Gain and loss of an intron in a protein-coding gene in archaea: the case of an archaeal RNA pseudouridine synthase gene. *BMC Evol. Biol.* **9**, 198 (2009).
17. Alekseyenko, A. V., Kim, N. & Lee, C. J. Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes. *RNA* **13**, 661–670 (2007).
18. Artamonova, I. I. & Gelfand, M. S. Comparative genomics and evolution of alternative splicing: the pessimists' science. *Chem. Rev.* **107**, 3407–3430 (2007).
19. Kim, E., Goren, A. & Ast, G. Alternative splicing: current perspectives. *Bioessays* **30**, 38–47 (2008).
20. Koren, E., Lev-Maor, G. & Ast, G. The emergence of alternative 3′ and 5′ splice site exons from constitutive exons. *PLoS Comput. Biol.* **3**, e95 (2007).
21. Fox-Walsh, K. L. *et al.* The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc. Natl Acad. Sci. USA* **102**, 16176–16181 (2005).
22. Sterner, D. A., Carlo, T. & Berget, S. M. Architectural limits on split genes. *Proc. Natl Acad. Sci. USA* **93**, 15081–15085 (1996).
23. Carmel, L., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Patterns of intron gain and conservation in eukaryotic genes. *BMC Evol. Biol.* **7**, 192 (2007).
24. Li, W., Tucker, A. E., Sung, W., Thomas, W. K. & Lynch, M. Extensive, recent intron gains in *Daphnia* populations. *Science* **326**, 1260–1262 (2009).
25. Farlow, A., Meduri, E., Dolezal, M., Hua, L. & Schlötterer, C. Nonsense-mediated decay enables intron gain in *Drosophila*. *PLoS Genet.* **6**, e1000819 (2010).
26. Sela, N. *et al.* Comparative analysis of transposed element insertion within human and mouse genomes reveals *Alu*'s unique role in shaping the human transcriptome. *Genome Biol.* **8**, R127 (2007).
27. Schwartz, S. H. *et al.* Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* **18**, 88–103 (2008).
28. Barbosa-Morais, N. L., Carmo-Fonseca, M. & Aparicio, S. Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Res.* **16**, 66–77 (2006).
29. Fedorov, A., Merican, A. F. & Gilbert, W. Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc. Natl Acad. Sci. USA* **99**, 16128–16133 (2002).
30. Csuros, M., Rogozin, I. B. & Koonin, E. V. Extremely intron-rich genes in the alveolate ancestors inferred with a flexible maximum-likelihood approach. *Mol. Biol. Evol.* **25**, 903–911 (2008).
31. Nguyen, H. D., Yoshihama, M. & Kenmochi, N. New maximum likelihood estimators for eukaryotic intron evolution. *PLoS Comput. Biol.* **1**, e79 (2005).
32. Roy, S. W. & Irimia, M. Splicing in the eukaryotic ancestor: form, function and dysfunction. *Trends Ecol. Evol.* **24**, 447–455 (2009).
This review covers important aspects of eukaryotic evolution.
33. Rogozin, I. B., Wolf, Y. I., Sorokin, A. V., Mirkin, B. G. & Koonin, E. V. Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* **13**, 1512–1517 (2003).
34. Roy, S. W. & Gilbert, W. Rates of intron loss and gain: implications for early eukaryotic evolution. *Proc. Natl Acad. Sci. USA* **102**, 5773–5778 (2005).
35. Carmel, L., Wolf, Y. I., Rogozin, I. B. & Koonin, E. V. Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res.* **17**, 1034–1044 (2007).
36. Jaillon, O. *et al.* Translational control of intron splicing in eukaryotes. *Nature* **451**, 359–362 (2008).
37. Kerenyi, Z. *et al.* Inter-kingdom conservation of mechanism of nonsense-mediated mRNA decay. *EMBO J.* **27**, 1585–1595 (2008).
38. Plass, M., Agirre, E., Reyes, D., Camara, F. & Eyras, E. Co-evolution of the branch site and SR proteins in eukaryotes. *Trends Genet.* **24**, 590–594 (2008).
39. Gal-Mark, N., Schwartz, S., Ram, O., Eyras, E. & Ast, G. The pivotal roles of TIA proteins in 5′ splice-site selection of *Alu* exons and across evolution. *PLoS Genet.* **5**, e1000717 (2009).
40. Irimia, M., Rukov, J. L., Penny, D. & Roy, S. W. Functional and evolutionary analysis of alternatively spliced genes is consistent with an early eukaryotic origin of alternative splicing. *BMC Evol. Biol.* **7**, 188 (2007).
41. Lev-Maor, G., Sorek, R., Shomron, N. & Ast, G. The birth of an alternatively spliced exon: 3′ splice-site selection in *Alu* exons. *Science* **300**, 1288–1291 (2003).
42. Nurtdinov, R. N., Artamonova, I. I., Mironov, A. A. & Gelfand, M. S. Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum. Mol. Genet.* **12**, 1313–1320 (2003).
43. Thanaraj, T. A., Clark, F. & Muilu, J. Conservation of human alternative splice events in mouse. *Nucleic Acids Res.* **31**, 2544–2552 (2003).
44. Kondrashov, F. A. & Koonin, E. V. Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences. *Trends Genet.* **19**, 115–119 (2003).
45. Gilbert, W. Why genes in pieces? *Nature* **271**, 501 (1978).
46. Kondrashov, F. A. & Koonin, E. V. Origin of alternative splicing by tandem exon duplication. *Hum. Mol. Genet.* **10**, 2661–2669 (2001).
47. Doolittle, R. F. The multiplicity of domains in proteins. *Annu. Rev. Biochem.* **64**, 287–314 (1995).
48. Kolkman, J. A. & Stemmer, W. P. Directed evolution of proteins by exon shuffling. *Nature Biotech.* **19**, 423–428 (2001).
49. Liu, M. & Grigoriev, A. Protein domains correlate strongly with exons in multiple eukaryotic genomes — evidence of exon shuffling? *Trends Genet.* **20**, 399–403 (2004).
The authors found a strong correlation between borders of exons and protein domains in multiple eukaryotic genomes. This is consistent with the principles of exon shuffling.
50. Peng, T. & Li, Y. Tandem exon duplication tends to propagate rather than to create *de novo* alternative splicing. *Biochem. Biophys. Res. Commun.* **383**, 163–166 (2009).
51. Letunic, I., Copley, R. R. & Bork, P. Common exon duplication in animals and its role in alternative splicing. *Hum. Mol. Genet.* **11**, 1561–1567 (2002).
52. Gal-Mark, N., Schwartz, S. & Ast, G. Alternative splicing of *Alu* exons — two arms are better than one. *Nucleic Acids Res.* **36**, 2012–2023 (2008).
53. Long, M., Rosenberg, C. & Gilbert, W. Intron phase correlations and the evolution of the intron/exon structure of genes. *Proc. Natl Acad. Sci. USA* **92**, 12495–12499 (1995).
54. Patthy, L. Genome evolution and the evolution of exon-shuffling — a review. *Gene* **238**, 103–114 (1999).
55. Patthy, L. Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett.* **214**, 1–7 (1987).
56. Schmucker, D. *et al. Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684 (2000).
57. De Grassi, A. & Ciccarelli, F. D. Tandem repeats modify the structure of human genes hosted in segmental duplications. *Genome Biol.* **10**, R137 (2009).
58. Parma, J., Christophe, D., Pohl, V. & Vassart, G. Structural organization of the 5′ region of the thyroglobulin gene. Evidence for intron loss and 'exonization' during evolution. *J. Mol. Biol.* **196**, 769–779 (1987).
59. Makalowski, W., Mitchell, G. A. & Labuda, D. *Alu* sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet.* **10**, 188–193 (1994).
60. Sorek, R., Ast, G. & Graur, D. *Alu*-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067 (2002).
61. Nekrutenko, A. & Li, W. H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **17**, 619–621 (2001).
62. Wang, W. & Kirkness, E. F. Short interspersed elements (SINEs) are a major source of canine genomic diversity. *Genome Res.* **15**, 1798–1808 (2005).
63. Wang, W. *et al.* Origin and evolution of new exons in rodents. *Genome Res.* **15**, 1258–1264 (2005).
64. Kandul, N. P. & Noor, M. A. Large introns in relation to alternative splicing and gene evolution: a case study of *Drosophila bruno-3*. *BMC Genet.* **10**, 67 (2009).
65. Kent, L. B. & Robertson, H. M. Evolution of the sugar receptors in insects. *BMC Evol. Biol.* **9**, 41 (2009).
66. Fu, Y. *et al.* Alternative splicing of anciently exonized 5S rRNA regulates plant transcription factor TFIIIA. *Genome Res.* **19**, 913–921 (2009).
67. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
68. Amit, M. *et al.* Biased exonization of transposed elements in duplicated genes: a lesson from the *TIF-IA* gene. *BMC Mol. Biol.* **8**, 109 (2007).
69. Mersch, B., Sela, N., Ast, G., Suhai, S. & Hotz-Wagenblatt, A. SERpredict: detection of tissue- or tumor-specific isoforms generated through exonization of transposable elements. *BMC Genet.* **8**, 78 (2007).
70. Lin, L. *et al.* Diverse splicing patterns of exonized *Alu* elements in human tissues. *PLoS Genet.* **4**, e1000225 (2008).
71. Toll-Riera, M. *et al.* Origin of primate orphan genes: a comparative genomics approach. *Mol. Biol. Evol.* **26**, 603–612 (2009).
72. Sorek, R. The birth of new exons: mechanisms and evolutionary consequences. *RNA* **13**, 1603–1608 (2007).
73. Singer, S. S., Mannel, D. N., Hehlgans, T., Brosius, J. & Schmitz, J. From 'junk' to gene: curriculum vitae of a primate receptor isoform gene. *J. Mol. Biol.* **341**, 883–886 (2004).
74. Krull, M., Brosius, J. & Schmitz, J. *Alu*-SINE exonization: *en route* to protein-coding function. *Mol. Biol. Evol.* **22**, 1702–1711 (2005).
75. Kreahling, J. & Graveley, B. R. The origins and implications of *Alu*ternative splicing. *Trends Genet.* **20**, 1–4 (2004).
76. Schwartz, S. *et al. Alu* exonization events reveal features required for precise recognition of exons by the splicing machinery. *PLoS Comput. Biol.* **5**, e1000300 (2009).
This article describes the features that are required for precise recognition of exons by the splicing machinery by analysing *Alu* exonization events.
77. Sorek, R. *et al.* Minimal conditions for exonization of intronic sequences: 5′ splice site formation in *Alu* exons. *Mol. Cell* **14**, 221–231 (2004).
78. Ram, O., Schwartz, S. & Ast, G. Multifactorial interplay controls the splicing profile of *Alu*-derived exons. *Mol. Cell. Biol.* **28**, 3513–3525 (2008).
79. Corvelo, A. & Eyras, E. Exon creation and establishment in human genes. *Genome Biol.* **9**, R141 (2008).
The authors show that specific sequence environments are required for exonization and that these can change with time.

80. Gombart, A. F., Saito, T. & Koeffler, H. P. Exaptation of an ancient *Alu* short interspersed element provides a highly conserved vitamin D-mediated innate immune response in humans and primates. *BMC Genomics* **10**, 321 (2009).
81. Lee, J. R. *et al.* Lineage specific evolutionary events on *SFTPB* gene: *Alu* recombination-mediated deletion (ARMD), exonization, and alternative splicing events. *Gene* **435**, 29–35 (2009).
82. Pan, Q. *et al.* Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet.* **21**, 73–77 (2005).
83. Lev-Maor, G. *et al.* The 'alternative' choice of constitutive exons throughout evolution. *PLoS Genet.* **3**, e203 (2007).
This paper shows that exons shift from constitutive to alternative splicing during evolution, and relaxation of the 5′ splice site sequence is one of the molecular mechanisms that leads to this shift.
84. Ke, S., Zhang, X. H. & Chasin, L. A. Positive selection acting on splicing motifs reflects compensatory evolution. *Genome Res.* **18**, 533–543 (2008).
85. Lev-Maor, G. *et al.* Intronic *Alu*s influence alternative splicing. *PLoS Genet.* **4**, e1000204 (2008).
This article shows that *Alu* insertions into introns change the mode of splicing of the flanking exons.
86. Tappino, B., Regis, S., Corsolini, F. & Filocamo, M. An *Alu* insertion in compound heterozygosity with a microduplication in *GNPTAB* gene underlies Mucolipidosis II. *Mol. Genet. Metab.* **93**, 129–133 (2008).
87. Mola, G., Vela, E., Fernandez-Figueras, M. T., Isamat, M. & Munoz-Marmol, A. M. Exonization of *Alu*-generated splice variants in the survivin gene of human and non-human primates. *J. Mol. Biol.* **366**, 1055–1063 (2007).
88. Sorek, R., Shamir, R. & Ast, G. How prevalent is functional alternative splicing in the human genome? *Trends Genet.* **20**, 68–71 (2004).
89. Yeo, G. W., Van Nostrand, E., Holste, D., Poggio, T. & Burge, C. B. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl Acad. Sci. USA* **102**, 2850–2855 (2005).
90. Modrek, B. & Lee, C. J. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nature Genet.* **34**, 177–180 (2003).
91. Magen, A. & Ast, G. The importance of being divisible by three in alternative splicing. *Nucleic Acids Res.* **33**, 5574–5582 (2005).
92. Irimia, M. *et al.* Widespread evolutionary conservation of alternatively spliced exons in *Caenorhabditis*. *Mol. Biol. Evol.* **25**, 375–382 (2008).
93. Melamud, E. & Moult, J. Stochastic noise in splicing machinery. *Nucleic Acids Res.* **37**, 4873–4886 (2009).
94. Xing, Y. & Lee, C. Alternative splicing and RNA selection pressure — evolutionary consequences for eukaryotic genomes. *Nature Rev. Genet.* **7**, 499–509 (2006).
95. Ermakova, E. O., Nurtdinov, R. N. & Gelfand, M. S. Fast rate of evolution in alternatively spliced coding regions of mammalian genes. *BMC Genomics* **7**, 84 (2006).
96. Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008).
97. Plass, M. & Eyras, E. Differentiated evolutionary rates in alternative exons and the implications for splicing regulation. *BMC Evol. Biol.* **6**, 50 (2006).
98. Sakabe, N. J. & de Souza, S. J. Sequence features responsible for intron retention in human. *BMC Genomics* **8**, 59 (2007).
99. Tilgner, H. *et al.* Nucleosome positioning as a determinant of exon recognition. *Nature Struct. Mol. Biol.* **16**, 996–1001 (2009).
The authors found stronger nucleosome occupancy in exons than in exons with weak splice sites and in pseudoexons.
100. Andersson, R., Enroth, S., Rada-Iglesias, A., Wadelius, C. & Komorowski, J. Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res.* **19**, 1732–1741 (2009).
The authors found higher nucleosome occupancy in exons. The exons were enriched with specific histone modifications.
101. Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M. & Bustamante, C. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* **325**, 626–628 (2009).
102. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.* **41**, 376–381 (2009).
103. Luco, R. F. *et al.* Regulation of alternative splicing by histone modifications. *Science* **327**, 996–1000 (2010).
The authors show the first direct link between histone modification and AS: the modulation of AS resulted in splice-site switching.
104. Nahkuri, S., Taft, R. J. & Mattick, J. S. Nucleosomes are preferentially positioned at exons in somatic and sperm cells. *Cell Cycle* **8**, 3420–3424 (2009).
105. Lavelle, C. & Prunell, A. Chromatin polymorphism and the nucleosome superfamily: a genealogy. *Cell Cycle* **6**, 2113–2119 (2007).
106. Sivolob, A. & Prunell, A. Nucleosome conformational flexibility and implications for chromatin dynamics. *Philos. Transact. A Math. Phys. Eng. Sci.* **362**, 1519–1547 (2004).
107. Alilat, M., Sivolob, A., Revet, B. & Prunell, A. Nucleosome dynamics. Protein and DNA contributions in the chiral transition of the tetrasome, the histone (H3–H4)2 tetramer–DNA particle. *J. Mol. Biol.* **291**, 815–841 (1999).
108. Kaplan, C. D. Revealing the hidden relationship between nucleosomes and splicing. *Cell Cycle* **8**, 3633–3634 (2009).
109. Garcia-Blanco, M. A., Baraniak, A. P. & Lasda, E. L. Alternative splicing in disease and therapy. *Nature Biotech.* **22**, 535–546 (2004).
110. Wood, M., Yin, H. & McClorey, G. Modulating the expression of disease genes with RNA-based therapy. *PLoS Genet.* **3**, e109 (2007).
111. Sugnet, C. W., Kent, W. J., Ares, M. Jr & Haussler, D. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.* **9**, 66–77 (2004).
112. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
113. Labrador, M. & Corces, V. G. Extensive exon reshuffling over evolutionary time coupled to *trans*-splicing in *Drosophila*. *Genome Res.* **13**, 2220–2228 (2003).
114. van Rijk, A. & Bloemendal, H. Molecular mechanisms of exon shuffling: illegitimate recombination. *Genetica* **118**, 245–249 (2003).
115. Babushok, D. V., Ostertag, E. M. & Kazazian, H. H. Jr. Current topics in genome evolution: molecular mechanisms of new gene formation. *Cell. Mol. Life Sci.* **64**, 542–554 (2007).
116. Patthy, L. Exon shuffling and other ways of module exchange. *Matrix Biol.* **15**, 301–310; discussion 311–312 (1996).
117. van Rijk, A. A., de Jong, W. W. & Bloemendal, H. Exon shuffling mimicked in cell culture. *Proc. Natl Acad. Sci. USA* **96**, 8074–8079 (1999).
118. Bass, B. L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
119. Athanasiadis, A., Rich, A. & Maas, S. Widespread A-to-I RNA editing of *Alu*-containing mRNAs in the human transcriptome. *PLoS Biol.* **2**, e391 (2004).
120. Lev-Maor, G. *et al.* RNA-editing-mediated exon evolution. *Genome Biol.* **8**, R29 (2007).
121. Moller-Krull, M., Zemann, A., Roos, C., Brosius, J. & Schmitz, J. Beyond DNA: RNA editing and steps toward *Alu* exonization in primates. *J. Mol. Biol.* **382**, 601–609 (2008).
122. Gommans, W. M., Mullen, S. P. & Maas, S. RNA editing: a driving force for adaptive evolution? *Bioessays* **31**, 1137–1145 (2009).

Acknowledgements
We thank D. Hollander for preparing the figures.

Competing interests statement
The authors declare no competing financial interests.

DATABASES
Entrez Gene: http://www.ncbi.nlm.nih.gov/gene
CAMP | *SFTPB* | *SLC35B3* | *TG* | *TRIM24*
FlyBase: http://flybase.org
bru3 | *Dscam*
UniProtKB: http://www.uniprot.org
NARF | TIA1

FURTHER INFORMATION
Authors' homepage: http://www.tau.ac.il/~gilast

ALL LINKS ARE ACTIVE IN THE ONLINE PDF