Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?

Nicholas J. McGlincy and Christopher W.J. Smith

Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge, CB2 1GA, UK

Alternative splicing (AS) strongly affects gene expression by generating protein isoform diversity. However, up to one-third of human AS events create a premature termination codon that would cause the resulting mRNA to be degraded by nonsense-mediated mRNA decay (NMD). The extent to which such events represent functionally selected post-transcriptional gene control, as opposed to noise in the splicing process, has been a contentious issue. Recent analyses indicate that many splicing regulatory proteins are themselves regulated by AS-NMD. Intriguingly, many of these AS-NMD events are coincident with ultraconserved genomic elements, which indicates their importance to vertebrate biology. Examination of these highly conserved events has led to new insights into the functions of AS-NMD and the role it can have in physiological circumstances.

Alternative splicing and nonsense-mediated mRNA decay: partners in regulating global gene expression?

Alternative splicing (AS) is a prevalent form of post-transcriptional gene regulation that is estimated to affect 75%–90% of human genes [1,2]. By creating mRNA isoforms with distinct patterns of the linear arrangement of exons, AS not only creates functionally distinct protein isoforms but also influences gene expression in a quantitative manner [3,4]. Most simply, this can occur through sequence alterations in the 5′ and 3′ untranslated regions (UTRs) and leads to differential translational control, mRNA stability or localization [3,4]. However, some mRNA isoforms contain premature termination codons (PTCs) (Figure 1). Although it was originally predicted that they encode truncated protein isoforms, in light of our current knowledge of the nonsense-mediated mRNA decay (NMD) pathway, such PTC-containing mRNAs are expected to be degraded [5] (Figure 2). NMD is one of several RNA surveillance pathways that ensure the fidelity of gene expression by degrading mRNAs that lack the proper arrangement of translational signals [5]. The primary function of NMD is thought to be the removal of such errors in gene expression that might otherwise lead to the accumulation of potentially toxic truncated proteins [6]. However, this idea seemed to conflict with the widely varying phenotypes arising from inactivation in different species of UPF1 (up-frameshift mutation 1), a central component of the NMD pathway. For example, *Saccharomyces cerevisiae* lacking *upf1* show only a slight growth defect under aerobic conditions [7,8], whereas *Mus musculus UPF1* knockouts are embryonic lethal [9]. This finding led many to posit that UPF1–NMD also has a role in regulating physiological gene expression; indeed, we now know that AS can result in NMD. Through the examination of individual AS events, it became evident that conserved regulated AS can introduce PTCs [10–15]. These studies led to the suggestion that some AS events had evolved to exploit NMD to achieve quantitative post-transcriptional regulation (AS-NMD, also termed regulated unproductive splicing and translation, or RUST) [6,16,17]. Although examples of AS-NMD were originally reported sporadically based upon chance discovery, computational approaches and the use of newly developed microarray platforms that can detect AS isoforms have enabled more systematic large-scale surveys of AS-NMD. Despite their broader scope, these studies have generated conflicting views of the prevalence of functional AS-NMD. Here, we focus on recent results elucidating the function and biological significance of select groups of highly conserved AS-NMD events and on the association of some of these events with genomic ultraconserved elements. We refer readers to more detailed reviews on the mechanisms of NMD [5,18] and AS [19,20], and more comprehensive accounts of examples of AS-NMD [21].

Much ado about nothing? The mixed messages of bioinformatics and microarrays

The mechanism of NMD has been studied in a variety of model organisms. Although the core NMD machinery is conserved from yeast to humans, the precise mechanism for identifying PTCs varies [22]. In mammals, a stop codon is recognized as premature if it is located >50–55 nucleotides (nt) upstream of an exon–exon junction [5] (Figures 1 and 2). Although the generality of this rule has been questioned [5,18,23], it seems to be widely applicable and has facilitated computational analyses designed to investigate globally the extent to which AS is coupled to NMD. By combining cDNA-based prediction of AS with the 55 nt rule, Brenner and colleagues found that ~35% of reliably inferred human alternative mRNA isoforms in the RefSeq database were predicted to be subject to NMD [17,24] (Figure 3). A subsequent examination of the curated SWISSPROT database indicated that 144 (5.8% of 2483) isoforms, most of which were annotated as encoding truncated proteins, were derived from mRNAs

Corresponding author: Smith, C.W.J. (cwjs1@cam.ac.uk).

0968-0004/$ – see front matter © 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.tibs.2008.06.001 Available online 11 July 2008

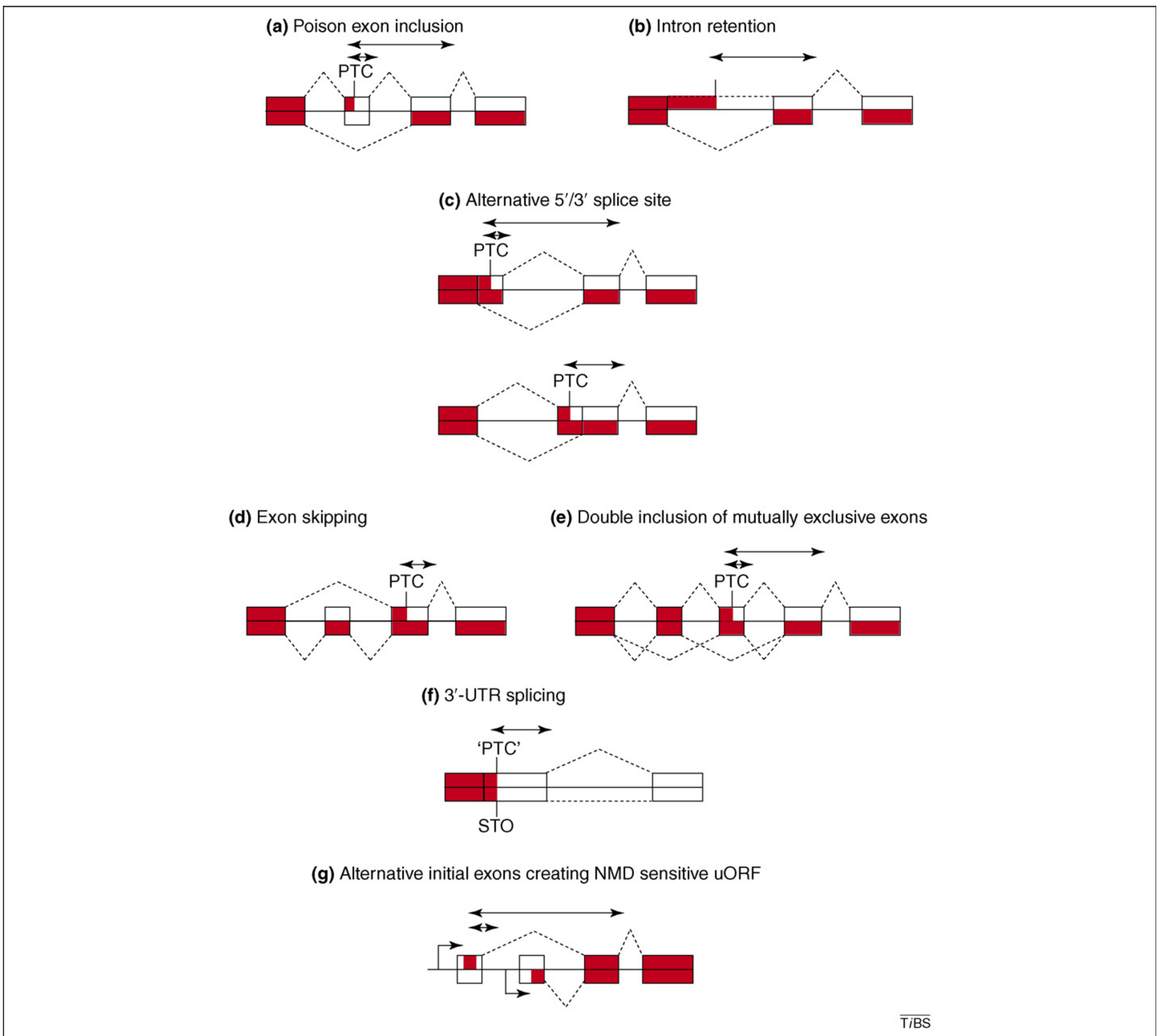

Figure 1. Splicing patterns that create PTCs. PTCs can be introduced by alternatively spliced segments of RNA. **(a)** Inclusion of a 'poison' cassette exon, an alternative exon that contains in-frame PTCs, **(b)** intron retention or **(c)** alternative 5′ or 3′ splice sites. Alternatively, frameshifts leading to PTC can occur by: **(d)** skipping of a cassette exon or **(e)** inclusion of both of a pair of exons that are normally mutually exclusive. **(f)** The natural termination codon can be detected as a PTC if introns are present >55 nt into the 3′ UTR. **(g)** The insertion of uORFs can lead to NMD because the uORF termination codon lies upstream of exon–exon junctions. Sequences that are part of an open reading frame (ORF) are red, whereas untranslated sequences are white. Different shading above and below the central line indicates the different ORFs engendered by the AS event. NMD-sensitive splicing patterns are shown above the exons. Arrows denote alternative promoters. Double-headed arrows denote a PTC to exon–exon junction distance of >50 nt, which constitutes the NMD-activating feature.

containing a PTC in a position that was predicted to lead to NMD [16]. Indeed, in several cases, prior experimental evidence supported predictions of NMD-sensitive isoforms, e.g. Calpain-10, a protease implicated in type-two diabetes; the cell-division cycle (CDC)-like kinases CLK1, CLK2 and CLK3, which are thought to regulate the activity of SR (serine arginine) proteins; and the apoptosis death-domain receptor, tumour necrosis-factor receptor superfamily, member 25 [16].

Many splicing events that result in the creation of a PTC bear the sequence hallmarks of functional AS. One-fifth of cassette-exon AS events conserved between *Homo sapiens* and *M. musculus* result in the creation of a PTC predicted to be subject to NMD [25]. Moreover, one-quarter of the conserved cassette exons that are predicted to cause NMD by their inclusion contain stop codons in all three reading frames, suggesting that their sole purpose is to induce NMD [25]. With the exception of being less likely to have a length divisible by three, the conserved cassette exons share similar characteristics with other alternative exons, including shorter length, weaker splice sites and increased conservation in the surrounding intronic sequence [25–28].

Owing to the limitations of computational AS predictions, especially when the isoforms of interest are preferentially degraded, these studies likely underestimate the

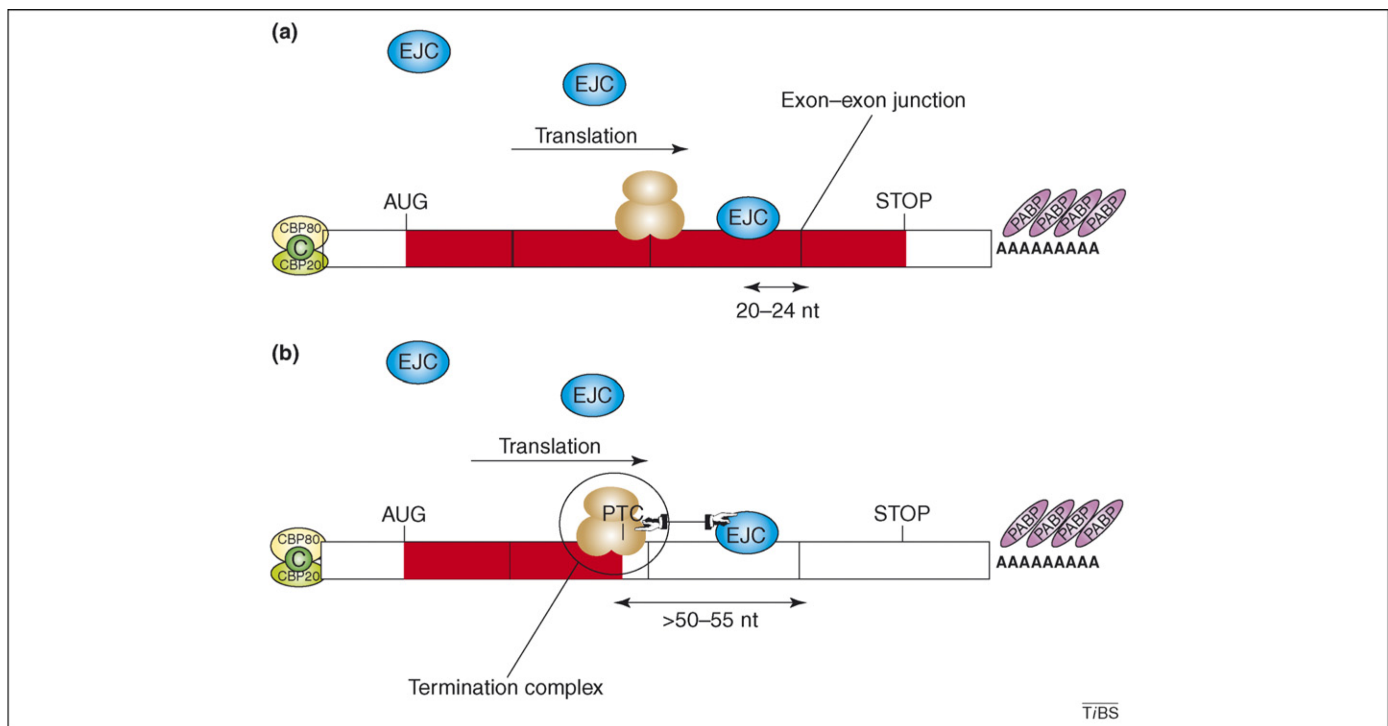

Figure 2. A summary of the molecular mechanism of mammalian NMD. NMD is a translation-dependent process. In mammals, NMD is thought to occur during the first, or 'pioneer', round of translation. At this point, mRNA is still bound by the nuclear cap-binding complex (CBC) of CBP20 (light green) and CBP80 (yellow). In mammals, a stop codon is generally recognized as premature by its spatial relation to exon–exon junctions, specifically, if one is found more than 50–55 nt downstream. The position of the exon–exon junction is communicated by the exon-junction complex, or EJC (blue), a large multiprotein complex that is deposited on the mRNA as a result of splicing. EJC deposition is sequence independent and occurs 20–24 nt upstream of the exon–exon junction. In a normal mRNA, **(a)** the ribosome (brown) proceeds along the mRNA, displacing EJCs. When it reaches the stop codon, a termination complex forms and the mRNA then goes on to direct protein synthesis. In a PTC-containing mRNA, **(b)**, the ribosome encounters the PTC and, recognizing it as a stop codon, forms a termination complex. The location of an exon–exon junction >50 nucleotides downstream then sets the stage for the interaction between the termination complex and EJC that provokes NMD. For greater detail on the molecular mechanism of NMD, the reader is referred to recent reviews [5,18]. Sequence that is part of an open reading frame (ORF) is shaded red, whereas untranslated sequence is white. The cap (C) is green. The poly-A-binding protein (PABP) is shown in purple. The ribosome is brown, and the EJCs are light blue.

true frequency of AS-NMD [25,29]. Combined with the pre-existing serendipitous reports, the potential frequency of such events led Brenner and colleagues to posit that AS-NMD is a widespread post-transcriptional regulatory strategy [16,17]. Regulating the AS decision between protein coding and NMD-sensitive isoforms would enable fine-tuning of the amount of mRNA that would go on to direct protein synthesis and, hence, control protein expression levels.

This view was challenged by the microarray analysis of HeLa cells in which the essential NMD factor UPF1 had been depleted by RNAi (RNA interference). Using microarrays designed to measure total transcript levels, Mendell *et al.* [30] and Wittmann *et al.* [31] obtained similar results to those obtained by examining *S. cerevisiae* lacking Upf1p [32–34] and estimated that 4.9% of expressed genes were regulated by NMD [30,31]. The affected genes possessed diverse NMD-activating features, including not only examples of AS-NMD but also upstream open-reading frames (uORFs), transcripts encoding selenocysteine-containing proteins, and transcripts arising from transposon and endogenous retroviral sequences. Moreover, UPF1 has also been implicated in other mRNA decay pathways [35–37], so some of the observed changes upon UPF1 knockdown [30] will have been due to the targets of these respective pathways. Therefore, AS-NMD is only one of several mechanisms by which NMD and UPF1 regulate physiological gene expression [30,31].

The question of the functional importance of AS-NMD was addressed directly by Pan *et al.* [38] using DNA microarrays sensitive to ~3000 cassette-exon AS events [38–40]. Their results provide the most compelling evidence against a widespread role for AS-NMD in regulating gene expression. Across ten mouse tissues they found that PTC-containing isoforms were expressed at relatively constant low levels, with little evidence of tissue-specific expression. Furthermore, across individual genes, increased expression of PTC-containing isoforms did not correlate with reduced total transcript levels, a correlation that might be expected if AS-NMD is generally utilized to downregulate gene expression. Upon siRNA-mediated depletion of UPF1 in HeLa cells, the levels of only a small proportion of PTC-containing isoforms showed pronounced increases. Whereas 30%–40% of genes with a PTC-containing isoform showed a >5% increase in PTC-containing isoform expression, <10% of such genes showed a >15% increase. Although this finding alone might not be inconsistent with some mode of quantitative regulation, only 6% of genes with PTC-containing isoforms showed significant increases in overall RNA levels. Strikingly, in many cases, compensatory decreases in the non-PTC-containing isoform were observed. This observation is difficult to reconcile with the function of UPF1 in NMD, which should not affect isoforms that lack PTCs, although in some cases indirect feedback could affect splicing in this manner [41]. The overriding message from Pan *et al.* [38] was that

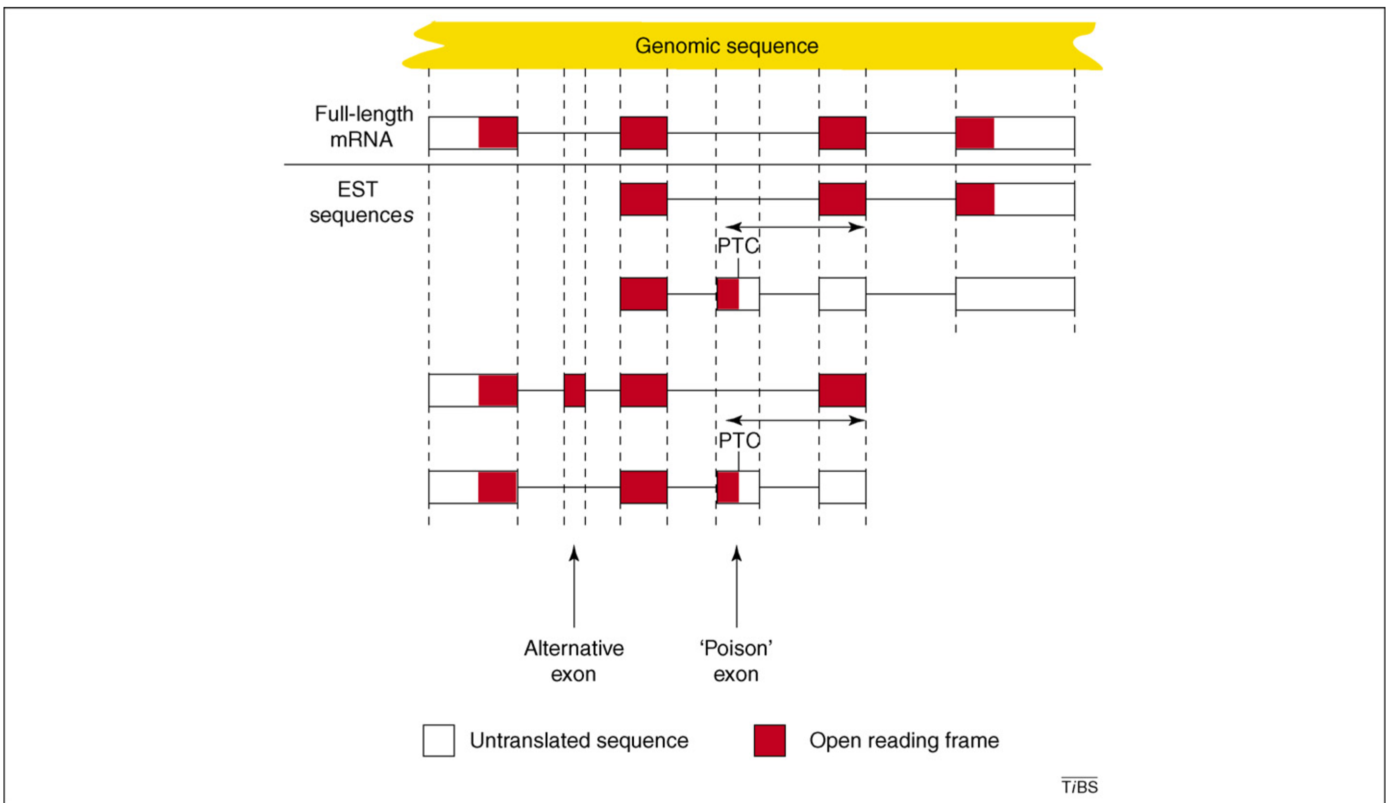

Figure 3. Computational prediction of AS-NMD from cDNA sequences. Gapped alignment of cDNA sequences from ESTs and full-length mRNAs to the genomic sequence enables the inference of the exonic structure of a gene [75]. Exons that are not always in all aligned cDNAs are deemed to be alternatively spliced. In this schematic only exon skipping is considered, but this method of inference equally applies to the other forms of AS. Alignment of a large number of cDNA sequences leads to a cluster of overlapping alignments. Equivalent alignments can be subsumed to predict the sequence of potential alternative mRNA isoforms [76]. The coding ORF of each isoform can then be calculated to determine whether the stop codon is >50 nt upstream of an exon–exon junction, making the isoform a potential target for NMD. Sequence that is part of an ORF is shown in red, and untranslated sequence is white. Double-headed arrows denote a PTC to exon–exon junction distance of >50 nt. Poison exon refers to an alternative exon that contains in-frame PTCs.

functional AS-NMD events are relatively uncommon and vastly outnumbered by events that represent background noise in gene expression.

Ultra-conserved nonsense in the SR and hnRNP protein families

Although functional AS-NMD might be less prevalent than was indicated by initial computational surveys, recent work has clearly demonstrated a function for this process in the regulation of certain gene families. AS-NMD events were known to occur within individual members of both the SR and hnRNP (heterogeneous nuclear ribonucleoprotein) families of mammalian splicing regulators, and within the SR-protein genes from *Caenorhabditis elegans* and the plants *Oryza sativa* and *Arabidopsis thaliana* [10–13,42]. Nevertheless, it came as a surprise when Lareau and coworkers showed that all members of the human and mouse SR-protein gene family undergo AS-NMD and, furthermore, that these events coincide with regions of extreme sequence conservation, termed 'ultraconserved elements' (UCEs) [41] (Box 1). These results were confirmed by Ni *et al.* [43] and Saltzman *et al.* [44], who examined the prevalence of AS-NMD by using a combination of AS-sensitive microarrays and bioinformatics. Both studies, in contrast to Pan *et al.* [38], sought to identify cases of cassette-exon AS-NMD that were highly conserved and, therefore, very likely to have functional consequences. Ni *et al.* [43] identified a set of 66 exons that were enriched for association with UCEs (Box 1), whereas their resident genes were preferentially involved in AS (including those encoding both SR and hnRNP proteins) [43]. Interestingly, reverse-transcription–PCR-based validation of 12 of the identified exons that created PTCs by their inclusion showed wide-ranging increases in inclusion (ranging from 2% to 75%) in response to NMD inhibition [45]. Saltzman *et al.* [44] not only confirmed the observation of AS-NMD within the SR and hnRNP genes but also observed conserved (although not ultra-conserved) AS-NMD of many core spliceosomal proteins, indicating that AS-NMD might regulate proteins involved in both constitutive and alternative splicing.

AS-NMD-mediated regulation

The examination of individual AS-NMD events has led to a deeper understanding of the function of AS-NMD and the role that such regulation can have in physiological circumstances. Much of this work has involved the highly conserved events within the SR and hnRNP families of splicing regulators. It seems that, although AS-NMD is essentially repressive in nature, it has been utilized during evolution to achieve several different functional outcomes.

Negative feedback

The AS-NMD events within the *SFRS2* (splicing factor, arginine/serine rich 2; SC35), *SFRS3* (splicing factor, arginine/serine rich 3; SRp20), *SFRS7* (splicing factor,

Box 1. UCEs in the human genome

Bejerano *et al.* [63] identified 481 UCEs in the human genome, defined as sequence segments of 200 bp or longer that are 100% conserved, with no insertions or deletions, in *M. musculus* and *Rattus norvegicus*. The majority of these elements are also present with a high degree of similarity in the *Gallus gallus* genome, and two-thirds are also present, albeit at a lower similarity, in the pufferfish *Takifugu rubripes*. However, in stark contrast to rRNA or most human coding regions, only 5% of UCEs have identifiable orthologues in non-vertebrate organisms, such as the sea squirt *Ciona intestinalis*, *Drosophila melanogaster* or *C. elegans*. This finding indicates that the functions encoded by UCEs have arisen during vertebrate evolution.

UCEs are either partially or wholly non-protein coding. Of 481 UCEs, 111 overlap the mRNA of known human protein-coding genes. These genes are significantly enriched for the functions of RNA binding and AS regulation and often contain the RNA-recognition motif protein domain. Furthermore, for the two-thirds of UCEs that overlap coding sequence, bioinformatic evidence indicates that the coding sequence is alternatively spliced, which indicates that some UCEs regulate AS decisions. This class of elements includes the examples of AS-NMD coincident with UCEs identified within the genes for the SR and hnRNP splicing regulators by Lareau *et al.* and Ni *et al.* The remaining elements lie wholly within introns (100) or intergenic regions (270) and tend to cluster near genes encoding transcription factors and genes that have important roles in early development. UCEs that are located in intergenic regions are often far from genes. This finding prompted the suggestion that these UCEs might represent distal enhancers of early developmental genes; indeed, this idea has been confirmed in many cases [79–81].

In conclusion, UCEs represent highly conserved, yet predominantly vertebrate-specific, non-coding elements. They have been implicated in both the regulation of AS of genes involved in RNA processing and the transcriptional regulation of genes involved in early development. However, it seems that we have some way to go before wholly understanding the function of these elements. For example, why do the AS events associated with UCEs require such extensive conservation when many important AS events do not (see Figure 4a in the main text)? Moreover, a recent study failed to detect any crucial abnormalities in transgenic mice lacking one of four UCEs that have been shown to act as developmental enhancers [82].

arginine/serine-rich 7; 9G8) and *PTBP1* (polypyrimidine tract binding protein 1) genes are promoted by their respective proteins as part of autoregulatory negative-feedback loops [10–13]. This negative-feedback motif is also evident in some other examples of AS-NMD, such as the genes encoding the ribosomal proteins L3 (*RPL3*) and L12 (*RPL12*) [14,15]. The common occurrence of AS-NMD within SR and hnRNP genes and their frequent coincidence with UCEs indicates that AS-NMD-mediated negative feedback could be an important mechanism for the homeostatic regulation of splicing factors. Indeed, a concordance is seen between the polarity of the AS-NMD events and the type of splicing factor affected [43]. That is, the AS-NMD events of SR proteins, which tend to activate splicing, comprise inclusion of exons possessing stop codons (so-called ‘poison exons’) (Figure 1a). By contrast, for the hnRNP proteins, which commonly function as splicing repressors, AS-NMD events are associated with exon skipping. This idea was further corroborated by a computational search for conserved intronic splicing-regulatory sequences by Yeo *et al.* [46]. They found that several splicing-factor genes possessed highly conserved exons predicted by their sequence characteristics to be alternatively spliced [46]. Many of these genes had already been identified as undergoing AS-NMD by microarray analysis [43,44]. In a subset of these, for which the protein-binding-site sequence is well characterised, the predicted alternative exon was surrounded by a high density of sequences resembling binding site of the factor, which is suggestive of autoregulation [46].

Negative-feedback loops limit the expression of a protein and prevent its over-accumulation, but the strength of feedback determines the dynamic response of the system. The study of transcriptional negative-feedback loops has shown that low-to-intermediate repression levels reduce gene-expression noise and shift such variations to higher frequencies that are expected to have minimal impact on subsequent processes [47–49]. Such homeostatic control would result in proteins maintaining more constant concentrations [49,50]. This effect might be expected to benefit the regulation of AS patterns, which are sensitive to the concentration of regulatory factors [19,20]. A recent study of protein variability in human cell populations corroborates this idea [51]; the SR-like protein SFRS11 (transformer 2 homologue, *D. melanogaster*), which regulates its own expression through a putative AS-NMD event [52], showed one of the lowest degrees of cell-to-cell variability of the proteins studied [51].

By contrast, strongly repressive negative-feedback loops can produce oscillatory behaviours. An example of this is the RNA-binding protein AtGRP7 (*A. thaliana* glycine-rich RNA-binding protein 7). Whereas its circadian-clock-driven promoter is responsible for basal oscillation in AtGRP7 mRNA and protein, negative feedback mediated by AtGRP7 is required for lasting high-amplitude oscillation [53,54]. This negative-feedback loop is mediated by AtGRP7 promoting an AS-NMD event within its own mRNA [55,56]. The crucial concentration at which AtGRP7 triggers this event is likely to determine the amplitude of the oscillation. Furthermore, the negative-feedback loop probably causes the observed delay between oscillations in AtGRP7 mRNA and protein expression [53–55]. Hence, although negative feedback is a feature of many characterised AS-NMD events, it might promote different functionalities, depending on the properties of the system.

Cross-repression of splicing regulators

The functional importance of AS-NMD is not restricted to negative-feedback loops. AS-NMD can also be utilized to repress production of a protein until it is required by the cell, functioning as a post-transcriptional on/off switch. In particular, certain groups of splicing regulators cross-regulate each other using AS-NMD. This process has been shown for TIA1 (T-cell-restricted intracellular antigen 1) and TIAR (TIA-1-related protein) [57–59] and also for PTBP1 (polypyrimidine-tract-binding protein 1) and its paralogues, PTBP2 (more commonly known as nPTB or neuronal PTB) and ROD1 (regulator of differentiation 1, *Schizosaccharomyces pombe*) [45,60,61] (Figure 4b). The cross-regulation of PTBP1 and PTBP2 is of particular importance because it is involved in regulating a set of AS events that occurs during neuronal differentiation, a vital process in the developing mammalian brain [45,60]. PTBP1 and PTBP2 expression patterns are negatively correlated. In the mammalian brain, PTBP2 protein is

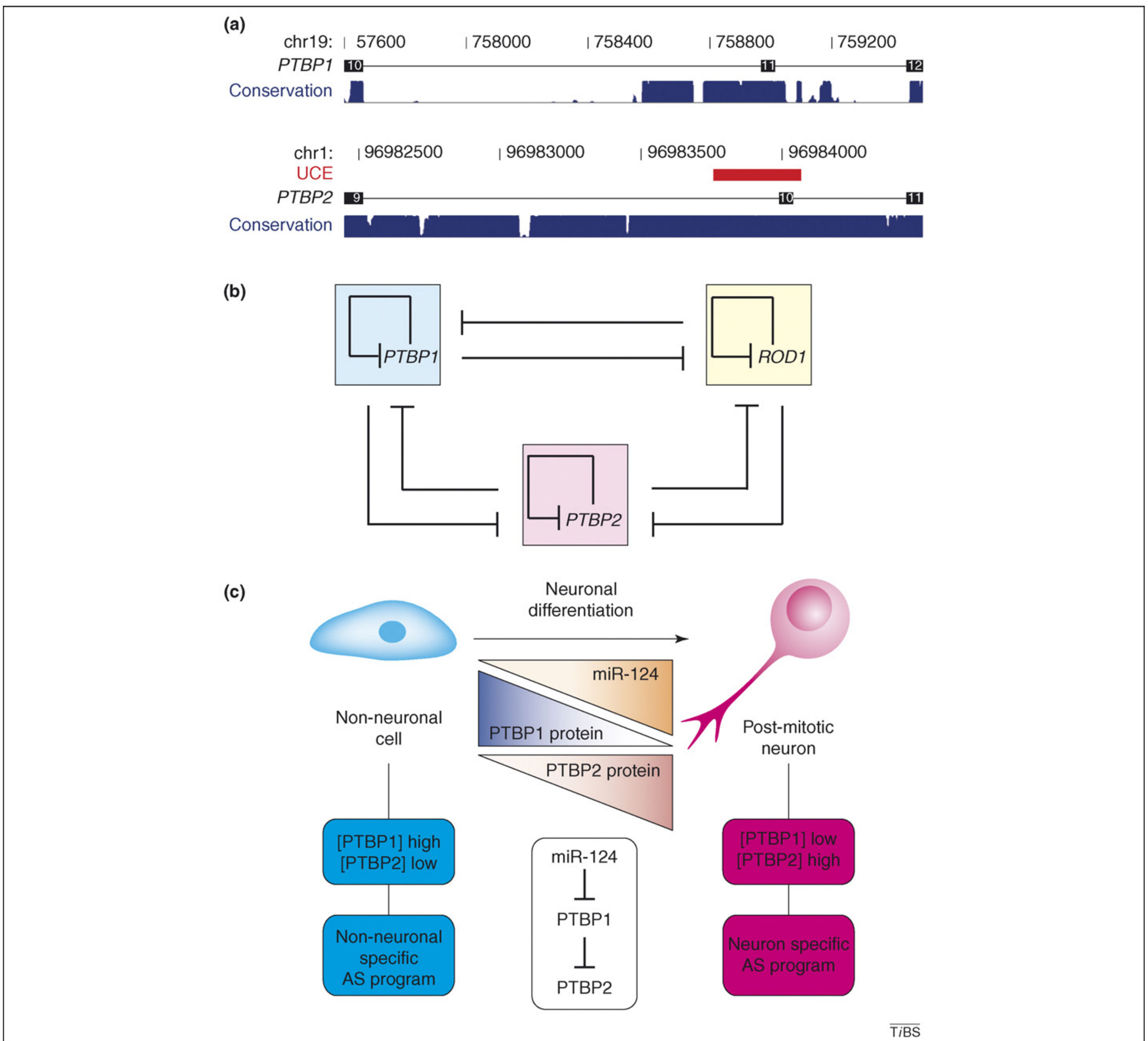

Figure 4. AS-NMD-mediated cross-repression of PTBP1 paralogues and its role in neuronal differentiation. Cross-repression and autoregulation of *PTBP1* paralogues provides a post-transcriptional switch controlling global changes in AS during neuronal differentiation. **(a)** Schematic of the genomic organization of the AS-NMD events within *PTBP1* (top) and *PTBP2* (bottom) produced using the UCSC (University of California Santa Cruz) genome browser [77]. Exons are represented by numbered black boxes and introns by black lines. A red box indicates a UCE as identified in Ref. [63]. The lower blue histogram represents conservation across 17 vertebrate species as calculated in Ref. [78]. A distance scale is shown above each schematic. Although *PTBP1* exon 11 and the surrounding intronic sequence is highly conserved, as might be expected of an important AS event, it pales in comparison to that observed surrounding *PTBP2* exon 10 and the associated UCE. **(b)** Cross-repression and autoregulation of *PTBP1* paralogues. Bold lines and thin lines indicate established and non-established associations, respectively. Both PTBP1 (blue) and PTBP2 (pink) autoregulate their expression by promoting AS-NMD events within their own pre-mRNA [13,45]. Likewise, PTBP1 represses PTBP2 expression by promoting AS-NMD within its pre-mRNA [45,61]. It is unclear whether PTBP2 also regulates PTBP1 in this way. Furthermore, it remains to be seen whether ROD1 (yellow) regulates its own expression or that of PTBP1 and/or PTBP2 through similar mechanisms. **(c)** The role of PTBP1–PTBP2 cross-repression in neuronal differentiation. Increased miR-124 expression during differentiation represses PTBP1, causing a relief in its AS-NMD-mediated repression of nPTB. This process results in increased PTBP2 expression and decreased PTBP1 expression and results in large-scale changes in AS patterns.

exclusively expressed in differentiated post-mitotic neurons, whereas PTB expression is restricted to neuronal precursor cells, glia and other non-neuronal cells [45]. Both *PTBP1* and *PTBP2* genes contain AS-NMD events. *PTBP1* exon-11 skipping results in a frame-shift and the subsequent creation of a PTC in exon 12 [13] (Figure 4a). Exon-11 skipping is promoted by PTBP1 itself as part of an autoregulatory negative-feedback loop [13] (Figure 4b). The orthologous event in *PTBP2* involves skipping of exon 10, which lies within a UCE [45,62,63] (Figure 4a).

Within neuronal progenitor cells, PTBP1 represses PTBP2 protein expression by promoting *PTBP2* exon-10 skipping, which results in the majority of PTBP2 mRNA being degraded by NMD [45,60] (Figure 4c). During neuronal differentiation, an increase in microRNA miR-124 expression results in the repression of PTBP1 protein expression [60] (Figure 4c). This, in turn, results in the

predominant inclusion of *PTBP2* exon 10, thereby leading to increased PTBP2 protein expression [45,60] (Figure 4c). Boutz *et al.* [45] showed that PTBP1 and PTBP2 regulate distinct but overlapping sets of exons and that 83% of the exons that demonstrated altered expression because of siRNA-mediated depletion of PTBP1 showed equivalent changes upon the differentiation of P19 cells (murine embryonic carcinoma cells that differentiate into postmitotic neurons after retinoic acid treatment). By taking these exons as a proportion of the total number of exons changing in expression upon P19 differentiation, the authors estimate that as many as 25% of the changes in AS that occur during neuronal differentiation result from the switch from PTBP1 to PTBP2 expression [45]. The genes regulated by PTBP1 and/or PTBP2 are involved in diverse cellular processes but are enriched for genes involved in cytoskeletal rearrangement and vesicular or protein transport [45]. These processes are essential for the remodelling of cell morphology during differentiation into a neuron. Makeyev *et al.* [60] confirmed the role of miR-124 in initiating this transition and showed that neuron-specific splicing of several genes could be induced by both siRNA-mediated depletion of PTBP1 and miR-124 overexpression. The tissue distribution of isoforms induced by miR-124 expression mirrors that of miR-124 itself and negatively correlates with PTBP1 [60].

Repression of PTBP2 expression by PTBP1 extends beyond neuronal progenitor cells. Even in HeLa cells, siRNA-mediated PTBP1 knockdown promotes increased PTBP2 production owing to the switch in *PTBP2* exon-10 splicing [61]. Numerous AS changes are observed after the knockdown of both PTBP1 and PTBP2, including, intriguingly, within the PTB paralogue ROD1 [61]. PTBP1 and PTBP2 repress *ROD1* exon-2 inclusion, which leads to the creation of a uORF with a PTC only 20 codons after the initiating AUG [61] (Figure 1g). Despite possessing a PTC, the exon-2-skipped mRNA is not sensitive to NMD, perhaps owing to translational reinitiation [61]. ROD1 is mainly expressed in haematopoietic cells. It remains to be seen whether a transition analogous to that from PTBP1 to PTBP2 expression in neuronal differentiation occurs during haematopoiesis, directing a set of haematopoietic-specific AS events (Figure 4b).

Splicing of conserved 3′ UTR introns

Splicing of introns located >50 nucleotides into the 3′ UTR causes the authentic termination codon to appear premature and, hence, targets the mRNA for NMD. Such splicing events can serve a regulatory function whether they are constitutive or alternative. An interesting example that affects mRNA stability in a complex manner is provided by the ARE-binding protein AUF1 [64,65]. AREs (AU-rich elements) within the 3′ UTR are a major *cis*-acting determinant of rapid cytoplasmic mRNA decay [66], which can be initiated by the binding of proteins such as AUF1 [66]. AUF1 expression is regulated by AS of the highly conserved exon 9 and flanking introns within the 3′ UTR. Retention of the two introns preserves two ARE sequences within the 3′ UTR, which exposes *AUF1* mRNA to negative autoregulation or cross-regulation by other ARE-binding proteins [64,65]. Splicing the intron downstream of exon 9, however, creates an exon–exon junction at sufficient distance from the natural stop codon to render the transcript sensitive to NMD [64,65]. Conversely, complete skipping of exon 9 produces an mRNA that is insensitive to both decay processes. An analysis of EST (expressed sequence tag) sequences indicates the presence of exon 9 included, NMD-sensitive isoforms in embryonic, but not adult, tissues from *H. sapiens* and *M. musculus* [65], which indicates that AS-NMD might mediate a developmental transition in AUF1-expression patterns. Indeed, pathological conditions appear to result from misregulation of AUF1 expression [65]. For example, patients with congestive heart failure overexpress AUF1 in their cardiac tissue [65]. This results in the decreased expression of AUF1-target mRNAs, including the β_1-adrenergic receptor, which contributes to regulating normal cardiac output [65].

An additional example of NMD resulting from 3′-UTR splicing is provided by ARC (activity-regulated cytoskeleton-associated protein), an important protein for the processes that result in neuronal plasticity and the consolidation of long-term memory [67]. After NMD inhibition, *ARC* mRNA and protein levels are increased owing to the presence of two conserved introns within the 3′ UTR [67]. It is currently unknown whether these introns are alternatively spliced or whether *ARC* mRNA is NMD-sensitive by default. A bioinformatic analysis, however, revealed several mRNAs, many of which encoding other proteins involved in synaptic function, that also seem to be deliberately subjected to NMD by 3′-UTR introns conserved in rodents. Many of these mRNAs are subjected to translational regulation to prevent their expression before they are appropriately localized. The function of this layer of control could be to destroy the mRNA rapidly should it be translated in an inappropriate location. Alternatively, a short pulse of expression programmed by translation-dependent mRNA degradation might be sufficient to deliver an effective concentration of active protein in restricted cellular compartments [67].

AS-NMD and the evolution of novel protein function

Although the above discussion of AS-NMD has focused on highly conserved events, many AS-NMD events are not well conserved [68]. Some of these events represent *de novo* skipping of formerly constitutive exons [68], whereas others arise from AS of newly ‘exonized’ sequences, which often derive from repetitive *Alu* elements in primates [28,68,69]. Although a lack of conservation does not preclude function [70], many of these events are thought to represent noise resulting from inaccuracy in the splicing process. Far from being inconsequential, however, this noise likely provides the raw material for molecular evolution. In this instance, AS-NMD is thought to provide a permissive environment for the evolution of novel protein functions. The expression of PTCs in low-frequency isoforms that are degraded by NMD will obscure them from the full scrutiny of selection [71,72]. This will allow the sequence to drift and possibly acquire new functions, in a process analogous to gene duplication and divergence [71,72]. Essentially, the low-frequency isoform functions as a paralogue [71,72]. This idea is corroborated by the observation that alternatively spliced exons that preserve

the reading frame have a lower mutation rate than those that alter the reading frame, which indicates that they are likely to be under greater selection [73]. New functions could appear in different forms. Genes evolving in this manner might acquire either novel protein isoforms or AS-NMD-mediated quantitative regulation.

Concluding remarks and future perspectives

Whereas initial bioinformatic surveys and microarray experiments provided conflicting views about the extent of functional AS-NMD, recent, more focused experiments demonstrate that subsets of these events have clear and important functions and have provided hints about further revelations in this area [67]. The frequent coincidence of these AS events with UCEs indicates that they have an important role in mammalian physiology. However, the reason for the association of UCEs with these AS-NMD events remains an intriguing but unresolved issue. Conserved AS exons contain a greater extent of conserved flanking intronic sequence than constitutive exons [27], which reflects the presence of regulatory elements. Yet splicing regulatory elements show high degrees of redundancy, and most conserved highly regulated AS events do not require such high levels of conservation [20,74]. A possible approach to investigate the physiological relevance of these events – and, indeed, other AS-NMD events not associated with UCEs – would be to generate transgenic animals in which the NMD-sensitive splicing event is disabled.

The proven functionalities of AS-NMD events include autoregulatory negative-feedback loops and cross-regulation. Future work should question whether AS-NMD negative-feedback loops function similarly to those at the transcriptional level to attenuate noise in expression and what impact this has on the processes controlled by these proteins, especially AS. We should also question whether any advantage exists for the negative-feedback loop occurring at the post-transcriptional level or whether this simply reflects the most permissive evolutionary route for an RNA-binding protein. In addition, AS-NMD can be used to suppress the production of a protein in the absence of the proper biological context. In particular, PTBP1 and PTBP2 utilize this mechanism to coordinate changes in global AS patterns during neuronal differentiation. It will be interesting to examine whether similar cross-regulation exists between other groups of splicing regulators and whether similar switches occur at other developmental transitions.

In conclusion, although initial estimates of the prevalence of functional AS-NMD events might have been overzealous, the characterisation of individual instances has proven their importance in regulating gene expression. AS-NMD-mediated regulation can function in several ways and provide repression, oscillation or reduced variability in gene expression. Indeed, even those events that might be regarded as 'non-functional' likely represent molecular evolution in action.

Acknowledgements

We apologise to those authors whose work was not cited directly because of length constraints. We also thank Martina Hallegger, Lit Yeen Tan and Claudio Alonso for comments on the manuscript. Work in the authors' laboratory is funded by the Wellcome Trust (077877) and EURASNET. N.J.M. was supported by a studentship from the Medical Research Council.

References

1 Harrow, J. *et al.* (2006) GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* 7 (Suppl. 1), S41–S49
2 Kapranov, P. *et al.* (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296, 916–919
3 Hughes, T.A. (2006) Regulation of gene expression by alternative untranslated regions. *Trends Genet.* 22, 119–122
4 Stamm, S. *et al.* (2005) Function of alternative splicing. *Gene* 344, 1–20
5 Chang, Y.F. *et al.* (2007) The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 76, 51–74
6 Neu-Yilik, G. *et al.* (2004) Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife. *Genome Biol.* 5, 218
7 Altamura, N. *et al.* (1992) NAM7 nuclear gene encodes a novel member of a family of helicases with a Zn-ligand motif and is involved in mitochondrial functions in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 224, 575–587
8 Leeds, P. *et al.* (1991) The product of the yeast *UPF1* gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev.* 5, 2303–2314
9 Medghalchi, S.M. *et al.* (2001) Rent1, a trans-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum. Mol. Genet.* 10, 99–105
10 Jumaa, H. and Nielsen, P.J. (1997) The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.* 16, 5077–5085
11 Lejeune, F. *et al.* (2001) Alternative splicing of intron 3 of the serine/arginine-rich protein 9G8 gene. Identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. *J. Biol. Chem.* 276, 7850–7858
12 Sureau, A. *et al.* (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J.* 20, 1785–1796
13 Wollerton, M.C. *et al.* (2004) Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* 13, 91–100
14 Cuccurese, M. *et al.* (2005) Alternative splicing and nonsense-mediated mRNA decay regulate mammalian ribosomal gene expression. *Nucleic Acids Res.* 33, 5965–5977
15 Mitrovich, Q.M. and Anderson, P. (2000) Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in *C. elegans*. *Genes Dev.* 14, 2173–2184
16 Hillman, R.T. *et al.* (2004) An unappreciated role for RNA surveillance. *Genome Biol.* 5, R8
17 Lewis, B.P. *et al.* (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U. S. A.* 100, 189–192
18 Shyu, A.B. *et al.* (2008) Messenger RNA regulation: to translate or to degrade. *EMBO J.* 27, 471–481
19 Black, D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336
20 Matlin, A.J. *et al.* (2005) Understanding alternative splicing: towards a cellular code. *Nat. Rev. Mol. Cell Biol.* 6, 386–398
21 Soergel, D.A.W. *et al.* (2006) Regulation of gene expression by the coupling of alternative splicing and nonsense-mediated mRNA decay. In *Nonsense-Mediated mRNA Decay* (Maquat, L.E., ed.), pp. 173–196, Landes Bioscience
22 Conti, E. and Izaurralde, E. (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* 17, 316–325
23 Isken, O. and Maquat, L.E. (2007) Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev.* 21, 1833–1856
24 Green, R.E. *et al.* (2003) Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. *Bioinformatics* 19 (Suppl. 1), i118–i121
25 Baek, D. and Green, P. (2005) Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12813–12818
26 Sugnet, C.W. *et al.* (2006) Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* 2, e4

27 Sorek, R. and Ast, G. (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* 13, 1631–1637
28 Magen, A. and Ast, G. (2005) The importance of being divisible by three in alternative splicing. *Nucleic Acids Res.* 33, 5574–5582
29 Modrek, B. and Lee, C. (2002) A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19
30 Mendell, J.T. *et al.* (2004) Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* 36, 1073–1078
31 Wittmann, J. *et al.* (2006) hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol. Cell. Biol.* 26, 1272–1287
32 Guan, Q. *et al.* (2006) Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. *PLoS Genet.* 2, e203
33 He, F. *et al.* (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5′ to 3′ mRNA decay pathways in yeast. *Mol. Cell* 12, 1439–1452
34 He, W. and Parker, R. (1999) Analysis of mRNA decay pathways in *Saccharomyces cerevisiae*. *Methods* 17, 3–10
35 Kim, Y.K. *et al.* (2005) Mammalian Staufen1 recruits Upf1 to specific mRNA 3′ UTRs so as to elicit mRNA decay. *Cell* 120, 195–208
36 Kim, Y.K. *et al.* (2007) Staufen1 regulates diverse classes of mammalian transcripts. *EMBO J.* 26, 2670–2681
37 Kaygun, H. and Marzluff, W.F. (2005) Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. *Nat. Struct. Mol. Biol.* 12, 794–800
38 Pan, Q. *et al.* (2006) Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. *Genes Dev.* 20, 153–158
39 Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell* 126, 37–47
40 Pan, Q. *et al.* (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* 16, 929–941
41 Lareau, L.F. *et al.* (2007) Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* 446, 926–929
42 Kalyna, M. *et al.* (2006) Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* 34, 4395–4405
43 Ni, J.Z. *et al.* (2007) Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* 21, 708–718
44 Saltzman, A.L. *et al.* (2008) Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol. Cell. Biol.* 28, 4320–4330
45 Boutz, P.L. *et al.* (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev.* 21, 1636–1652
46 Yeo, G.W. *et al.* (2007) Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet.* 3, e85
47 Austin, D.W. *et al.* (2006) Gene network shaping of inherent noise spectra. *Nature* 439, 608–611
48 Kaern, M. *et al.* (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* 6, 451–464
49 Becskei, A. and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature* 405, 590–593
50 Raser, J.M. and O'Shea, E.K. (2005) Noise in gene expression: origins, consequences, and control. *Science* 309, 2010–2013
51 Sigal, A. *et al.* (2006) Variability and memory of protein levels in human cells. *Nature* 444, 643–646
52 Stoilov, P. *et al.* (2004) Human tra2–β1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum. Mol. Genet.* 13, 509–524
53 Heintzen, C. *et al.* (1997) AtGRP7, a nuclear RNA-binding protein as a component of a circadian-regulated negative feedback loop in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 8515–8520
54 Staiger, D. and Apel, K. (1999) Circadian clock-regulated expression of an RNA-binding protein in *Arabidopsis*: characterisation of a minimal promoter element. *Mol. Gen. Genet.* 261, 811–819
55 Staiger, D. *et al.* (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. *Plant J.* 33, 361–371
56 Schoning, J.C. *et al.* (2007) Auto-regulation of the circadian slave oscillator component AtGRP7 and regulation of its targets is impaired by a single RNA recognition motif point mutation. *Plant J.* 52, 1119–1130
57 Izquierdo, J.M. and Valcarcel, J. (2007) Two isoforms of the T-cell intracellular antigen 1 (TIA-1) splicing factor display distinct splicing regulation activities. Control of TIA-1 isoform ratio by TIA-1-related protein. *J. Biol. Chem.* 282, 19410–19417
58 Le Guiner, C. *et al.* (2003) TIA-1 or TIAR is required for DT40 cell viability. *J. Biol. Chem.* 278, 10465–10476
59 Le Guiner, C. *et al.* (2001) TIA-1 and TIAR activate splicing of alternative exons with weak 5′ splice sites followed by a U-rich stretch on their own pre-mRNAs. *J. Biol. Chem.* 276, 40638–40646
60 Makeyev, E.V. *et al.* (2007) The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* 27, 435–448
61 Spellman, R. *et al.* (2007) Crossregulation and functional redundancy between the splicing regulator PTB and its paralogs nPTB and ROD1. *Mol. Cell* 27, 420–434
62 Rahman, L. *et al.* (2004) Evolutionary conservation of a 2-kb intronic sequence flanking a tissue-specific alternative exon in the *PTBP2* gene. *Genomics* 83, 76–84
63 Bejerano, G. *et al.* (2004) Ultraconserved elements in the human genome. *Science* 304, 1321–1325
64 Banihashemi, L. *et al.* (2006) Upf1/Upf2 regulation of 3′ untranslated region splice variants of AUF1 links nonsense-mediated and A+U-rich element-mediated mRNA decay. *Mol. Cell. Biol.* 26, 8743–8754
65 Wilson, G.M. *et al.* (1999) Regulation of AUF1 expression via conserved alternatively spliced elements in the 3′ untranslated region. *Mol. Cell. Biol.* 19, 4056–4064
66 Wilusz, C.J. *et al.* (2001) The cap-to-tail guide to mRNA turnover. *Nat. Rev. Mol. Cell Biol.* 2, 237–246
67 Giorgi, C. *et al.* (2007) The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell* 130, 179–191
68 Sorek, R. *et al.* (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet.* 20, 68–71
69 Sorek, R. *et al.* (2002) Alu-containing exons are alternatively spliced. *Genome Res.* 12, 1060–1067
70 Pan, Q. *et al.* (2005) Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet.* 21, 73–77
71 Xing, Y. and Lee, C. (2006) Alternative splicing and RNA selection pressure – evolutionary consequences for eukaryotic genomes. *Nat. Rev. Genet.* 7, 499–509
72 Xing, Y. and Lee, C.J. (2004) Negative selection pressure against premature protein truncation is reduced by alternative splicing and diploidy. *Trends Genet.* 20, 472–475
73 Zhang, C. *et al.* (2007) Evolutionary impact of limited splicing fidelity in mammalian genes. *Trends Genet.* 23, 484–488
74 Singh, R. and Valcarcel, J. (2005) Building specificity with nonspecific RNA-binding proteins. *Nat. Struct. Mol. Biol.* 12, 645–653
75 Modrek, B. *et al.* (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* 29, 2850–2859
76 Haas, B.J. *et al.* (2003) Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* 31, 5654–5666
77 Karolchik, D. *et al.* (2003) The UCSC genome browser database. *Nucleic Acids Res.* 31, 51–54
78 Siepel, A. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050
79 Pennacchio, L.A. *et al.* (2006) *In vivo* enhancer analysis of human conserved non-coding sequences. *Nature* 444, 499–502
80 Visel, A. *et al.* (2008) Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat. Genet.* 40, 158–160
81 Woolfe, A. *et al.* (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 3, e7
82 Ahituv, N. *et al.* (2007) Deletion of ultraconserved elements yields viable mice. *PLoS Biol.* 5, e234