Title

Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements

Journal

Nature, 446(7138)

ISSN

0028-0836

Authors

Lareau, Liana F
Inada, Maki
Green, Richard E
et al.

Publication Date

2007-04-01

DOI

10.1038/nature05676

Vol 446|19 April 2007|doi:10.1038/nature05676

LETTERS

Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements

Liana F. Lareau[1]*, Maki Inada[2]*, Richard E. Green[1]†, Jordan C. Wengrod[1] & Steven E. Brenner[1,2]

The human and mouse genomes share a number of long, perfectly conserved nucleotide sequences, termed ultraconserved elements[1]. Whereas these regions can act as transcriptional enhancers when upstream of genes, those within genes are less well understood. In particular, the function of ultraconserved elements that overlap alternatively spliced exons of genes encoding RNA-binding proteins is unknown[1,2]. Here we report that in every member of the human SR family of splicing regulators, highly or ultraconserved elements are alternatively spliced, either as alternative 'poison cassette exons' containing early in-frame stop codons, or as alternative introns in the 3′ untranslated region. These alternative splicing events target the resulting messenger RNAs for degradation by means of an RNA surveillance pathway called nonsense-mediated mRNA decay. Mouse orthologues of the human SR proteins exhibit the same unproductive splicing patterns. Three SR proteins have been previously shown to direct splicing of their own transcripts, and one of these is known to autoregulate its expression by coupling alternative splicing with decay[3–5]; our results suggest that unproductive splicing is important for regulation of the entire SR family. We find that unproductive splicing associated with conserved regions has arisen independently in different SR genes, suggesting that splicing factors may readily acquire this form of regulation.

The 11 human SR proteins comprise a homologous family of splicing factors (Supplementary Fig. 1) with diverse roles in RNA processing including control of export, translation, stability, and constitutive and alternative splicing[6]. To determine the extent of alternative splicing of SR genes themselves, we mined expressed sequence tag (EST) libraries for SR transcripts, aligned these ESTs to the human genome to infer splice junctions, and identified alternative junctions by comparison with a full-length reference mRNA of each gene. We found that all SR splicing regulators are alternatively spliced (Fig. 1, panel 2).

Notably, we found that all human SR genes had alternative splice forms that are expected to be degraded by nonsense-mediated mRNA decay (NMD) (Fig. 1, panel 2). The alternative splice forms contain stop codons that are thought to be recognized as premature by the NMD machinery because they are located more than 50 nucleotides upstream of the final exon–exon junction[7]. In some cases, ESTs with premature termination codons (PTCs) comprised a sizeable fraction of the total ESTs for that gene (Supplementary Table 3 and Supplementary Fig. 2). Seven of the SR genes exhibited a distinctive splicing pattern: an alternative, or cassette, exon was skipped in the reference isoform and included in an alternative isoform. *SRp20* (also called *SFRS3*), *SRP40* (*SFRS5*), *p54* (*SFRS11*), *SRP55* (*SFRS6*), *SRP75* (*SFRS4*) and *9G8* (*SFRS7*) contained poison cassette exons that introduced in-frame PTCs. *SRp38* (*FUSIP1*) had two linked alternative splicing events: an alternative 5′ splice site followed by a cassette exon, with a PTC in the extended upstream exon (Fig. 1, panel 2). For most of these genes, we also observed ESTs in which the introns flanking the poison cassette exons were retained and introduced PTCs. *SRp30c* (*SFRS9*) had similar intron-retention ESTs, but no ESTs supporting the cassette exon (discussed below). The remaining genes, *ASF/SF2* (also called *SFRS1*), *SC35* (*SFRS2*) and *SRP46* (*SFRS2B*), had alternative splicing in their 3′ untranslated regions (UTRs) (Fig. 1, panel 2). Although splicing in the 3′ UTR has no effect on the protein-coding sequence, the introduction of a new exon–exon junction more than 50 nucleotides downstream of the original stop codon marks that stop codon as premature and targets the transcript for NMD.

The existence of PTC^+ isoforms for all of the SR genes led us to ask whether this alternative splicing is truly unproductive, leading to downregulation of the spliced mRNA. We therefore measured the effect of NMD inhibition on mRNA isoform levels in HeLa cells using quantitative polymerase chain reaction with reverse transcription (quantitative RT–PCR). We disrupted NMD using short interfering (si)RNAs against *UPF1* (ref. 8), a key effector of the NMD pathway. If the PTCs trigger NMD as expected, the PTC^+ isoforms would be relatively stabilized upon UPF1 depletion. In all nine cases where the reference isoform could be detected, UPF1 depletion resulted in a 4- to 40-fold increase in the relative abundance of PTC-containing mRNAs (Fig. 2). We estimate that PTC^+ isoforms comprised about 2–14% of the spliced mRNA population from each gene in control conditions and 40–70% when NMD was inhibited (Supplementary Table 5), suggesting that a substantial fraction of SR transcripts is spliced into a form that is degraded by NMD. Time course experiments using cycloheximide rather than RNA interference to inhibit NMD showed similar effects on PTC^+ isoforms, supporting the expectation that the stabilization is a direct effect of NMD (Supplementary Fig. 4). Overall, our results indicate that alternative splice forms of the human SR protein family are produced at appreciable levels and are subsequently downregulated by NMD.

The 11 human SR genes correspond to 10 mouse SR genes (Supplementary Fig. 1; human *SRP46* is an expressed retropseudogene of *SC35*). We used ESTs to identify alternative splicing of the mouse SR genes. All of the genes except *SRp30c* had ESTs supporting the same unproductive splicing patterns in mouse as in human (Fig. 1, panel 3); the PTCs were in identical positions except in *SRP75* (called *Sfrs4* in mouse). Mouse *SRp30c* (also called *Sfrs9*) had a poison cassette exon, but there was no EST evidence of an equivalent cassette exon in

[1]Departments of Molecular and Cell Biology and [2]Plant and Microbial Biology, University of California, Berkeley, California 94720, USA. †Present address: Max Planck Institute for Evolutionary Anthropology, Leipzig D-04103, Germany.
*These authors contributed equally to this work.

the human orthologue (Fig. 1, panels 2 and 3). We identified a sequence within the human intron that was homologous to the mouse cassette exon and experimentally detected the predicted PTC$^+$ splice form in HeLa cells (Fig. 2). Because alternative splice forms are not generally shared between human and mouse, those that are conserved may be functionally important[9].

We next aligned the full genomic locus of each human SR protein with its mouse counterpart. Remarkably, all but one SR gene with conserved unproductive splice patterns had long regions of 100% nucleotide conservation between the human and mouse orthologues, ranging from 118 to 618 nucleotides in length (Fig. 1, panel 1). The conserved regions corresponded closely to the alternatively spliced poison cassette exons or to sequences flanking the 3′ UTR introns. The remaining gene, *SRp30c*, had shorter conserved stretches of 42 and 44 nucleotides at the ends of its poison cassette exon. The conserved sequences found in *9G8* (also called *Sfrs7* in mouse), *ASF/SF2* (*Sfrs1*), *SRp20* (*Sfrs3*), *p54* (*Sfrs11*) and *SRP55* (*Sfrs6*) were previously identified as ultraconserved genomic elements, defined as regions of at least 200 nucleotides of perfect identity between the human, mouse and rat genomes[1]; the conserved sequences in the other SR genes were shorter than that threshold, although still striking.

We compared the highly conserved regions among the human paralogues in order to investigate whether these regions arose from a common source. The most parsimonious scenario would be that the highly conserved sequences and their unproductive splicing

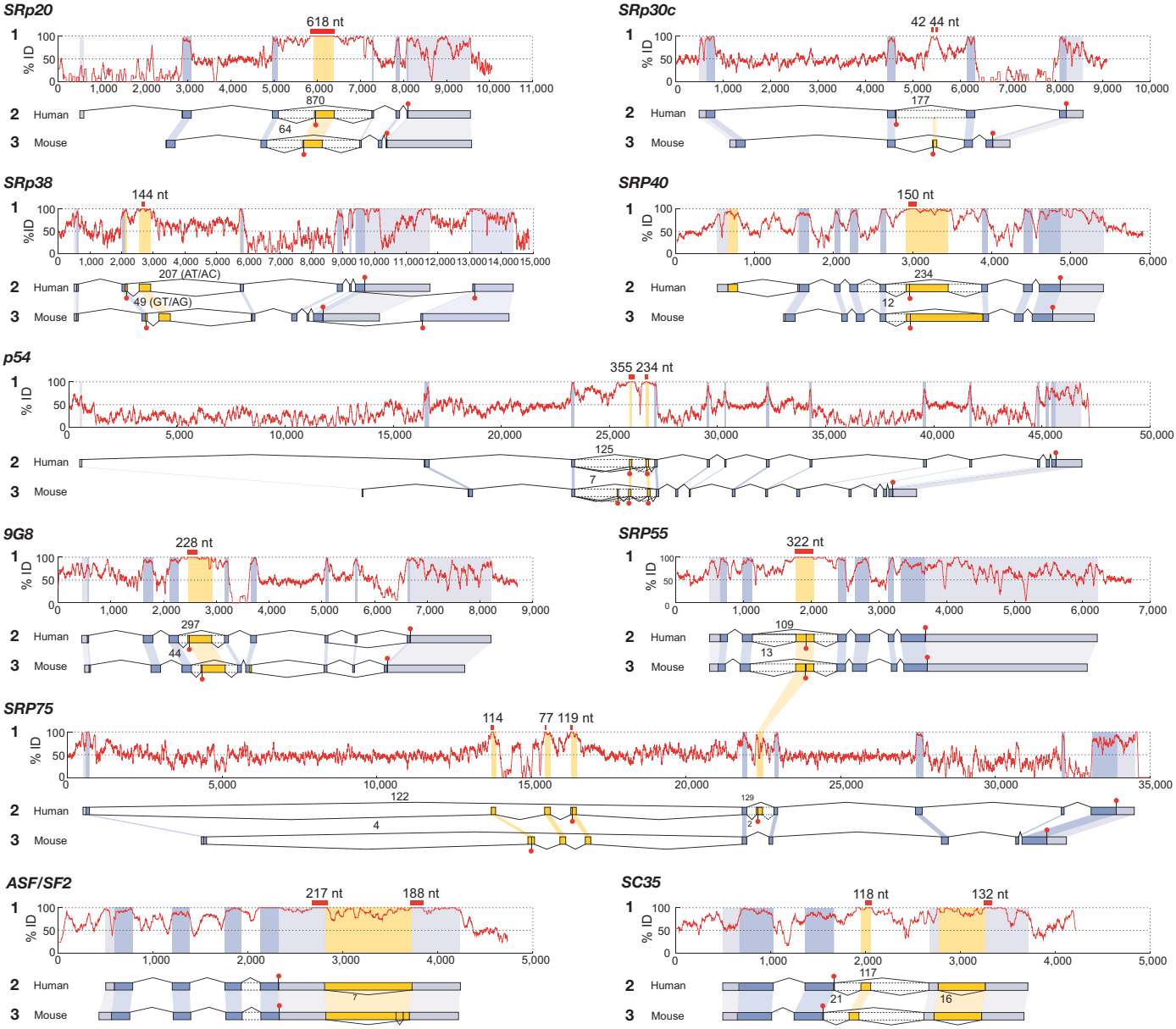

Figure 1 | SR gene alternative splice forms with PTCs are conserved between human and mouse and are associated with highly conserved and ultraconserved DNA elements. For each human/mouse SR gene pair, nucleotide conservation between human and mouse SR genomic loci is shown (panel 1). Per cent identity is shown in 50-nucleotide windows (*p54* in 100-nucleotide windows). DNA sequences with 100% conservation that are associated with alternatively spliced regions are marked (red bars). EST-inferred alternative splicing of human (panel 2) and mouse (panel 3) SR genes is shown with reference splice events (drawn above exon structure), alternative splicing (drawn below) and intron retention (dotted horizontal lines). Reference stop codons (above) and PTCs (below) are marked (red octagons). Gene regions are annotated as coding (blue), UTR (light blue) and alternatively spliced (orange). EST counts for alternative splice events and corresponding reference splice events are noted for human genes. The cassette exon in human *SRp30c* is shown by comparison to mouse. Sequence similarity between poison cassette exons of *SRP55* and *SRP75* is indicated with a connecting bar. nt, nucleotides.

originated in a common ancestor. Although some ultraconserved elements are known to have originated from a short interspersed element (SINE) retroposon[2], we found no evidence that the alternative regions of the SR genes contain known or novel repetitive elements (Supplementary Information). We also investigated whether a cassette exon was present in the ancestor of the SR paralogues. We found that the poison cassette exons were in non-homologous positions in different SR genes (Fig. 1, panel 2), implying independent rather than common origins. Moreover, local sequence alignments showed no significant nucleotide identity between the alternative regions of most human SR genes (Supplementary Information). The only significant relationship among alternative regions was found between the poison cassette exons in reference intron 2 of the most closely related human SR genes, *SRP55* and *SRP75*; this exon may therefore have been present in their common ancestor (Fig. 1, panel 2). This exon was ultraconserved between human and mouse in *SRP55* but not in *SRP75*, and it was observed rarely in *SRP75* ESTs, indicating that the exon may be under less selective pressure in *SRP75*. *SRP75* had additional, highly conserved poison cassette exons in its first reference intron that were not found in *SRP55*. Further study of the evolutionary dynamics of these ultraconserved regions may offer an insight into their origin and function. It is remarkable that similar splicing associated with highly conserved sequences seems to have arisen independently in different SR genes within this family.

Commonly measured constraints on sequence evolution do not explain the extreme conservation in these genes. Although a fraction

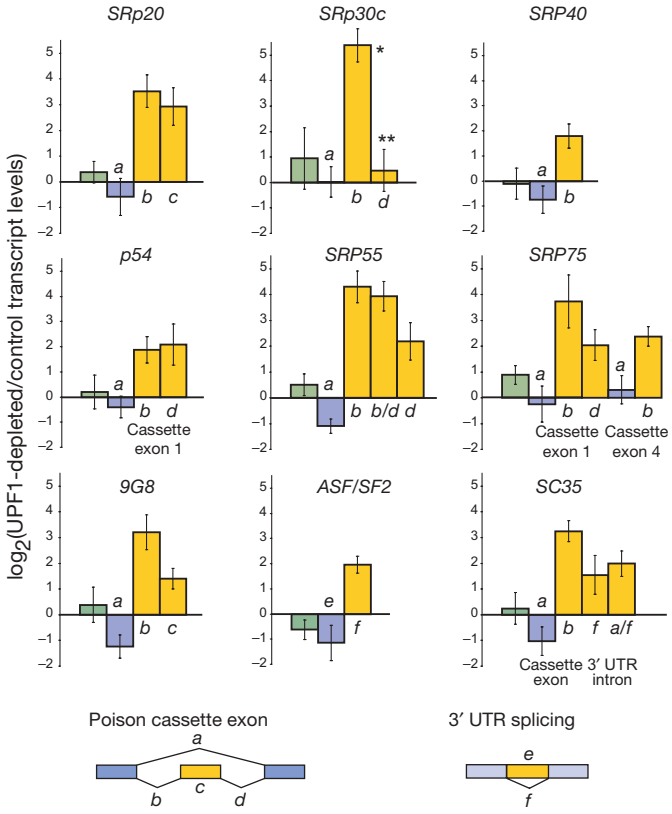

Figure 2 | SR gene alternative isoforms containing PTCs are degraded by NMD. Quantitative RT–PCR was used to assay levels of mRNAs in HeLa cells treated with control or *UPF1* siRNAs. The ratio of NMD-inhibited to control transcript levels was measured for constitutive (green), reference (blue; *a*, *e*) and PTC$^+$ (orange; *b*–*d*, *f*) regions. The data are $\log_2$ means of transcript ratios (±s.d.; three PCR replicates of four independent transfections). Asterisk, *SRp30c* PTC$^+$ isoform levels in control cells were at the lower limit of detection. Double asterisk, junction probes for *SRp30c* are predicted by homology to mouse ESTs and may not detect the actual junction used in human.

of the PTC$^+$ isoforms might escape NMD and be translated into truncated protein, selective pressure on the conserved regions is unlikely to arise from protein-coding constraints[3,10]. The stop codons are generally near the beginning of the conserved regions, or entirely upstream of the conserved region in the case of UTR splicing, leaving up to 422 nucleotides of non-coding, 100% conserved sequence downstream of the stop codon. A more probable explanation for the conservation is the presence of regulatory sequences that might control the use of the alternative splice sites. We found no consistent over-representation of binding sites for four well-characterized SR proteins in or near the alternative regions, and known exonic splicing enhancers were under-represented in the cassette exons (Supplementary Information); nonetheless, other binding sites may be present[11,12]. Similarly, *trans*-acting RNAs such as microRNAs might bind to the alternative regions. We identified putative microRNA-binding sites in the SR mRNAs[30], but the sites are not over-represented within the conserved alternative regions (Supplementary Information). The highly conserved sequences might also contribute to regulation through RNA structure, but most of the ultraconserved elements previously examined lack significantly low-energy secondary structure[1]. Extending the conservation analysis to other vertebrate genomes may help to pinpoint the functionally essential sequence elements[13]. We suspect that intertwined and as yet unknown constraints from RNA structure, RNA binding and protein binding may be at play.

The frequency of alternative splicing, the similar splice patterns and the exceptional sequence conservation of the SR genes strongly suggest that their unproductive splicing is functionally important. The human SR protein SC35 provides one model for this function. *SC35* pre-mRNAs can be spliced into either a reference isoform or PTC$^+$ isoforms. When SC35 protein is overproduced, it affects the splicing of its own pre-mRNAs, increasing the proportion of NMD-target isoforms[5]. This creates a negative feedback loop: high levels of SC35 protein cause a greater fraction of *SC35* transcripts to be spliced into the unproductive isoform and then degraded, thereby down-regulating SC35 protein production. Our data show that the entire SR family might couple alternative splicing and NMD to regulate protein production, a mechanism termed regulated unproductive splicing and translation (RUST)[14]. Consistent with this model, two other human SR proteins, 9G8 and SRp20, are known to affect splicing of their own transcripts[3,4]. Our data also lend support to this model: after NMD inhibition, some reference splice forms decrease in abundance (Fig. 2). Because inhibiting NMD should not directly affect the abundance of productive mRNAs, this decrease indicates a change in the underlying splicing ratio of productive to unproductive isoforms, consistent with negative feedback regulation. Other modes of regulation, such as cross-regulation, may also be involved. Notably, SR proteins are also regulated by Clk kinases, which themselves have PTC$^+$ splice forms. Modulation of SR proteins with a network of unproductive splicing might affect a wide range of downstream splicing targets[15–17].

The association between unproductive splicing and ultraconserved regions extends beyond the SR family to other RNA-binding proteins. Two SR-related genes and six heterogeneous nuclear ribonucleoprotein (hnRNP) genes also contain ultraconserved elements, and one hnRNP, PTB (also called PTBP1), autoregulates its expression by means of tissue-regulated unproductive splicing[1,18,19]. Furthermore, this mode of regulation may be phylogenetically widespread, as SR genes are alternatively spliced in other species including nematodes, flies and plants[20–24]. Some of these splice forms are homologous to human splice forms, such as a poison cassette exon found in the *Ciona intestinalis* orthologue of *SRp20* (Supplementary Fig. 5). Although the poison cassette splicing is conserved, the sequence ultraconservation does not extend from human to *Ciona*, showing that the regulatory phenomenon may be conserved but, if so, the specific sequence constraints are flexible.

What benefit would regulation by unproductive splicing confer, and why did this particular mechanism evolve in SR genes? The cell

may use unproductive splicing to maintain homeostasis of splicing factors. An SR-related splicing factor has been shown to have unusually low cell-to-cell variability in protein level[25]. Unproductive splicing may also allow tissue-specific expression of SR proteins[4]. If regulation would provide a selective advantage, alternative splicing linked with NMD is an evolutionarily accessible means of regulation for splicing factors, as they are inherently capable of binding mRNA. Whereas unproductive splice forms of many genes may represent cellular noise[26], the splicing events we have identified demonstrate how unproductive splicing can provide the capacity for regulation. Our observations paint a picture of unproductive splicing associated with highly conserved sequences as a regulatory process that has evolved repeatedly, affecting expression throughout an entire family of splicing factors.

METHODS

Detailed methods are supplied in Supplementary Information. To identify alternative splice forms of the SR genes, we aligned ESTs from UniGene and reference mRNAs to genomic loci using spidey v1.4, and compared the EST-inferred and reference splice junctions[27]. Alternative splice events seen in only one EST or occurring within ten nucleotides of a more prevalent splice event were discarded. To test whether alternative splice forms are degraded by NMD, we transfected HeLa cells with plasmids expressing siRNAs against *UPF1* (ref. 8) and reduced the level of UPF1 protein to $<5\%$ of the control level (Supplementary Fig. 3). For each SR gene, we measured relative mRNA levels using quantitative RT–PCR with primers to detect splice junctions or exons found in reference, PTC^+ and constitutive splice forms. The fold change in levels between control and experimental conditions for each primer pair was normalized to the average of three control genes: β-actin, *SDHA* and *TBP*.

Genomic alignments were constructed by splitting the human and mouse SR gene loci into individual exon and intron sequences and aligning each pair of orthologous sequences using the global alignment program from FASTA v2.1 (ref. 28). We searched for known and novel repeat elements within the genes using the RepeatMasker and chained self-assembly tracks of the UCSC human genome browser[29]. We performed local alignments to compare the poison cassette exons plus their flanking introns between the different human SR paralogues. The ESEfinder web server was used to search each SR gene containing a poison cassette exon for sites scoring higher than the default thresholds of the ASF/SF2, SC35, SRP40 and SRP55 binding motifs, and the RESCUE-ESE web server was used to search for known exonic splicing enhancer (ESE) sequences[11,12]. miRanda v1.0b was used to identify potential microRNA target sites[30].

Received 16 October 2006; accepted 28 February 2007.
Published online 14 March 2007.

1. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304,** 1321–1325 (2004).
2. Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441,** 87–90 (2006).
3. Jumaa, H. & Nielsen, P. J. The splicing factor SRp20 modifies splicing of its own mRNA and ASF/SF2 antagonizes this regulation. *EMBO J.* **16,** 5077–5085 (1997).
4. Lejeune, F., Cavaloc, Y. & Stevenin, J. 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 (2001).
5. Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J. & Soret, J. SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J.* **20,** 1785–1796 (2001).
6. Sanford, J. R., Ellis, J. & Caceres, J. F. Multiple roles of arginine/serine-rich splicing factors in RNA processing. *Biochem. Soc. Trans.* **33,** 443–446 (2005).
7. Nagy, E. & Maquat, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem. Sci.* **23,** 198–199 (1998).
8. Paillusson, A., Hirschi, N., Vallan, C., Azzalin, C. M. & Muhlemann, O. A GFP-based reporter system to monitor nonsense-mediated mRNA decay. *Nucleic Acids Res.* **33,** e54 (2005).
9. Lareau, L. F., Green, R. E., Bhatnagar, R. S. & Brenner, S. E. The evolving roles of alternative splicing. *Curr. Opin. Struct. Biol.* **14,** 273–282 (2004).
10. Ge, H., Zuo, P. & Manley, J. L. Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell* **66,** 373–382 (1991).
11. Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q. & Krainer, A. R. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res.* **31,** 3568–3571 (2003).
12. Fairbrother, W. G., Yeh, R. F., Sharp, P. A. & Burge, C. B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297,** 1007–1013 (2002).
13. Hughes, J. R. *et al.* Annotation of cis-regulatory elements by identification, subclassification, and functional assessment of multispecies conserved sequences. *Proc. Natl Acad. Sci. USA* **102,** 9830–9835 (2005).
14. Lewis, B. P., Green, R. E. & Brenner, S. E. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl Acad. Sci. USA* **100,** 189–192 (2003).
15. Hillman, R. T., Green, R. E. & Brenner, S. E. An unappreciated role for RNA surveillance. *Genome Biol.* **5,** R8 (2004).
16. Duncan, P. I., Stojdl, D. F., Marius, R. M. & Bell, J. C. *In vivo* regulation of alternative pre-mRNA splicing by the Clk1 protein kinase. *Mol. Cell. Biol.* **17,** 5996–6001 (1997).
17. Garcia-Sacristan, A., Fernandez-Nestosa, M. J., Hernandez, P., Schvartzman, J. B. & Krimer, D. B. Protein kinase clk/STY is differentially regulated during erythroleukemia cell differentiation: a bias toward the skipped splice variant characterizes postcommitment stages. *Cell Res.* **15,** 495–503 (2005).
18. Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A. & Smith, C. W. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13,** 91–100 (2004).
19. 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).
20. Kumar, S. & Lopez, A. J. Negative feedback regulation among SR splicing factors encoded by Rbp1 and Rbp1-like in *Drosophila*. *EMBO J.* **24,** 2646–2655 (2005).
21. Zorio, D. A., Lea, K. & Blumenthal, T. Cloning of *Caenorhabditis* $U2AF^{65}$: an alternatively spliced RNA containing a novel exon. *Mol. Cell. Biol.* **17,** 946–953 (1997).
22. Morrison, M., Harris, K. S. & Roth, M. B. *smg* mutants affect the expression of alternatively spliced SR protein mRNAs in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **94,** 9782–9785 (1997).
23. Kalyna, M., Lopato, S., Voronin, V. & Barta, A. Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res.* **34,** 4395–4405 (2006).
24. Iida, K. & Go, M. Survey of conserved alternative splicing events of mRNAs encoding SR proteins in land plants. *Mol. Biol. Evol.* **23,** 1085–1094 (2006).
25. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643–646 (2006).
26. Pan, Q. *et al.* 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 (2006).
27. Wheeler, D. L. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **34,** D173–D180 (2006).
28. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85,** 2444–2448 (1988).
29. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).
30. John, B. *et al.* Human microRNA targets. *PLoS Biol.* **2,** e363 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank present and past members of the Brenner laboratory and J. Pleiss for discussions and comments on the manuscript, Q. Meng and D. Rio for expertise, A. Fisher for assistance, and the Feldman, Luan, Rio and Tjian laboratories for use of their equipment. This work was supported by NIH grants and a Sloan fellowship to S.E.B. L.F.L. was supported by an NIH training grant. R.E.G. is supported by an NSF postdoctoral fellowship in Biological Informatics.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.E.B. (brenner@compbio.berkeley.edu).