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

Liana F. Lareau, Maki Inada, Richard E. Green, Jordan C. Wengrod, Steven E. Brenner

I. Supplementary methods and results

Identification of alternative splice forms

Reference forms of the 11 human and 10 mouse SR genes were identified from GenBank (Table S1). When a RefSeq mRNA encoding the full-length protein was available, it was used as the reference form[1]. When visual inspection indicated that no RefSeq form encoded the full-length protein, other mRNAs were identified from GenBank. The genomic loci were identified in Entrez Gene and were then expanded by 500 nt upstream and downstream to allow for slight variation in the transcriptional start and polyadenylation sites. UniGene clusters corresponding to the SR genes were identified in Entrez Gene (Table S1). The reference mRNA and the ESTs from each cluster were aligned to their genomic locus with `spidey` v1.4[1]; ESTs aligning to the opposite strand were discarded. Each `spidey` alignment was parsed into a set of splice junction coordinates. Junctions occurring in only one EST and junctions located within 10 nt on either side of a more prevalent splice junction were ignored, as these junctions were likely to represent splicing, sequencing, or alignment errors. Junctions that were seen in at least two ESTs but not in the reference splice form were considered to be alternative. These EST counts are indicated in Figure S2 and Table S3, and alternative splicing coordinates are provided in Table S2. Interestingly, the SRp38 reference splice junction had AT/AC donor/acceptor splice sites, which are expected to be recognized by the minor U12 spliceosome, but the alternative splice junctions had the canonical GT/AG sites recognized by the U2 spliceosome.

Quantitative RT-PCR detection of SR alternative isoforms in HeLa cells

HeLa cells were grown, transfected with control or UPF1 siRNA-expressing plasmids, and harvested as previously described with the following modification[2]. Plasmids were selected by treating cells with 1.5ug/ml puromycin for 72 or 96 hours and all cells were harvested 5 days post-transfection. RNAs were extracted and converted to cDNAs using gene specific primers. The primers for quantitative RT-PCR analysis were designed to detect the specific junctions and exons for total, reference, and PTC^+ transcript levels using the ArrayOligoSelector program[3] (Table S4). To assay total mRNA levels for each SR gene, a constitutive exon was selected. For reference mRNA levels, reference junctions were selected according to known reference mRNAs (Table S1). Alternative junctions were selected from

the EST-inferred gene structures to assay PTC-containing alternative isoforms (Table S2). Junction probes were designed with 10 nucleotides on the side of the junction to be extended and 15 nucleotides on the other side of the junction to minimize mispriming. Using SYBR green-based quantitative PCR, we measured the fold change for each primer pair between control and NMD-inhibited cells. We used primer-specific standard curves with total mRNA from untreated HeLa cells to assure that measurements were within linear range of detection and control for different amplification efficiencies.

To measure the fold change for a given primer pair, we first obtained raw fluorescence amplification curves for each sample and identified the cycle at which the curve crosses a threshold (C_T). We then made a standard curve of total mRNA vs $C_T$cycle threshold for the primer pair and plotted the control and Upf1-knockdown samples on the standard curve to determine the relative abundance. All data were normalized to the average of three control genes: beta actin, SDHA and TBP. The relative fold change was then calculated between the control and Upf1-knockdown samples.

Three PCR replicates were measured and averaged for each independent sample. The error was calculated as the standard deviation for technical replicates and propagated[4] for calculations of four biological replicates (two replicates each of 72 and 96-hour puromycin treatment). The data are plotted as $\log_2$ means $\pm$ SD. Melting curve analysis and agarose gel electrophoresis of PCR products were conducted to confirm that only a single product was present.

The signal from reference junction probes could include other unknown isoforms, some of which may contain alternative splice events that introduce PTCs at other locations in the same gene. Also, the c cassette exon probes could detect nuclear pre-mRNA that is not subject to NMD, decreasing the apparent fold change for these probes. The reference forms for SRp38 and SRp46 were not detected, so these genes were excluded from our analysis. Finally, the SRp30c cassette exon sequence was inferred by homology with mouse. The homologous exon ends with a canonical GT splice site, but other sites could be used. If the predicted splice site is not used in human, the d junction primer would not accurately detect the cassette exon.

Cycloheximide inhibition of NMD

To further determine whether the alternative isoforms of SR proteins are degraded by NMD, we treated HeLa cells with cycloheximide, which inhibits translation and hence NMD. Quantitative RT-PCR analyses of reference and PTC-containing mRNAs for each SR gene were conducted as described above. We measured relative levels of transcripts between HeLa cells treated with 10ug/ml cycloheximide (CHX) or DMSO during an 8 hour time course. For nearly all SR proteins, the levels of the PTC^+ isoforms increased within 30 minutes of cycloheximide treatment and reached 2- to 16-fold within 8 hours, supporting the expectation that the stabilization we observe in Upf1-depleted cells is a direct effect of NMD (Figure S4). We did not observe consistent stabilization for ASF/SF2; the PTC^+ isoform was stabilized by Upf1 depletion but not by cycloheximide treatment.

Fraction of unproductive isoforms for human SR genes

Using quantitative RT-PCR, we determined the fold changes in reference, PTC^+ and total transcript levels upon NMD inhibition in HeLa cells for each SR gene as described above. Our quantitative PCR measurements are relative to a standard curve of total untreated RNA for each primer pair and do not represent absolute amounts. Data from different primer pairs are not directly comparable, but data from the same primer pair in different conditions can be used to measure the fold change. Therefore, we cannot directly measure what fraction of transcripts in a given condition are PTC^+. However, knowing the relative fold changes of reference, PTC^+, and total transcripts, we were able to estimate the fraction of PTC^+ isoforms vs. total transcripts in both the control and NMD-inhibited conditions, as described below. The results are shown in Table S5.

We measure the the relative fluorescence of transcripts with the reference junction (***A***), transcripts with the PTC^+ junction (***B***) and transcripts with a constitutive exon (***C***) in NMD-inhibited ($_1$) and control conditions ($_0$). The measured amounts are proportionate to the unknown true abundances *a, b, c* with unknown proportions *α, β, γ*:

$$A_0 = \alpha a_0 \qquad A_1 = \alpha a_1$$
$$B_0 = \beta b_0 \qquad B_1 = \beta b_1$$
$$C_0 = \gamma c_0 \qquad C_1 = \gamma c_1$$

The measured fold change, *y*, for each splice form in NMD-inhibited vs. control conditions is:

Reference isoform fold change: $y_a = \frac{A_1}{A_0} = \frac{\alpha a_1}{\alpha a_0} = \frac{a_1}{a_0}$

PTC^+ isoforms fold change: $y_b = \frac{B_1}{B_0} = \frac{\beta b_1}{\beta b_0} = \frac{b_1}{b_0}$

Constitutive exon isoforms fold change: $y_c = \frac{C_1}{C_0} = \frac{\gamma c_1}{\gamma c_0} = \frac{c_1}{c_0}$.

By substituting the observed fold changes, we can solve for the fraction of PTC^+ isoforms, *r*, in control and NMD-inhibited conditions,

$$r_0 = \frac{b_0}{(a_0 + b_0)} \qquad r_1 = \frac{b_1}{(a_1 + b_1)},$$

without knowing the absolute quantities of the splice forms (e.g., a_0). The calculation depends on the assumption that

$$c = a + b$$

(see note below).

$$y_c = \frac{c_1}{c_0} = \frac{a_1 + b_1}{a_0 + b_0} = \frac{a_0 y_a + b_0 y_b}{a_0 + b_0} = (1 - r_0) y_a + r_0 y_b$$

$$\Rightarrow r_0 = \frac{y_c - y_a}{y_b - y_a}$$

and for the NMD-inhibited cells,

$$r_1 = \frac{b_1}{(a_1 + b_1)} = \frac{b_1}{(a_1 + b_1)} \cdot \frac{b_0}{b_0} \cdot \frac{a_0 + b_0}{a_0 + b_0} = \frac{b_0}{(a_0 + b_0)} \cdot \frac{b_1}{b_0} \cdot \frac{a_0 + b_0}{a_1 + b_1} = r_0 \cdot \frac{y_b}{y_c}$$

The fraction of transcripts with the PTC$^+$ junction in control conditions is then

$$r_0 = \frac{b_0}{(a_0 + b_0)} = \frac{(y_c - y_a)}{(y_b - y_a)}$$

and the fraction of transcripts with the PTC$^+$ junction in NMD-inhibited cells is

$$r_1 = \frac{b_1}{(a_1 + b_1)} = \frac{y_b(y_c - y_a)}{y_c(y_b - y_a)}.$$

Note: The estimation of the PTC$^+$ fraction depends on the assumption that the total mRNA, as detected by the primers to the constitutive exon, is the sum of the mRNA detected by the reference and PTC$^+$ primers ($c = a + b$). The measured PTC$^+$ isoform levels are based on splice-junction primers and do not include mRNAs with flanking intron retention. This calculation thus estimates a lower bound for the fraction of PTC$^+$ transcripts. While intron retention is present in some ESTs, our quantitative PCR analysis was unable to detect such pre-mRNAs for two assayed SR genes, 9G8 and SC35, so the estimated lower bound may represent the true level.

Repetitive elements in human SR genes

We searched for known repeat elements within the alternatively-spliced (AS) regions of the SR genes using the RepeatMasker track of the UCSC human genome browser in August 2006[5]. None of the AS regions contained annotated repetitive elements. We next searched for possible novel repetitive elements, which might be visible as short, highly similar matches between the AS region of an SR gene and multiple other locations in the genome.

Using the chained self-assembly track of the genome browser, we found that none of the SR poison cassette exons contained short repetitive elements. All but two SR genes (SRp40 and 9G8) aligned with other SR paralogs or with processed pseudogenes elsewhere in the genome, but these alignments included multiple constitutive exons rather than specifically the AS region. Three short sequences elsewhere in the genome matched within the ultraconserved elements in the 3'UTR of ASF/SF2. Similarly, the alternatively-spliced 3' UTRs of SC35 and SRp46, closely-related paralogs, matched two short sequences elsewhere in the genome. It is possible that these sequences are repetitive elements, but the one well-characterized retroposon previously identified as an ultraconserved element has multiple hits in the genome and is clearly delimited[6].

Sequence similarity between human SR alternatively-spliced regions

As described above, many of the SR genes have detectable nucleotide sequence similarity to the other SR paralogs. However, the detectable similarity was usually limited to the constitutive exons of the genes, not their introns or alternative exons. We searched for similarity between the AS regions of different human SR paralogs, which would indicate that the common ancestor of the paralogs might have had the same alternative splicing. We

compared the poison cassette exons plus their flanking introns between the different human SR paralogs, using SSEARCH v3.4 to perform full Smith-Waterman local alignment[7]. The only significant similarity was between the AS regions of SRp55 and SRp75, with an e-value of 1.5e-13 calculated by comparison to a database size of 114148068 residues in 50421 sequences. SRp55 reference intron 2 aligned with SRp75 reference intron 2 with 59% identity in a 687-nt overlap that included the poison cassette exon of each gene (Figure 1).

Exonic splicing enhancer analysis

The ESEfinder webserver (release 2.0) was used to search each SR gene with a poison cassette exon for sites scoring higher than the default thresholds of the ASF/SF2, SC35, SRp40, and SRp55 binding motifs[8]. We compared the number of hits in the poison exon of each gene to the number of hits in other exons of that gene, and the number of hits in the flanking introns to the number of hits in other introns, using a χ^2 test with no multiple-test correction (data not shown). Binding sites were somewhat underrepresented ($p \leq 0.05$) in the flanking introns of all genes except SRp55. Binding sites were slightly overrepresented ($p \leq 0.05$) in poison exons of SRp20, SRp54, and SRp55, although this overrepresentation disappeared when the poison exons were compared only to coding exons, excluding UTR exons.

The RESCUE-ESE webserver was used to search each poison-exon SR gene for known exonic splicing enhancer (ESE) sequences[9]. Comparisons were performed as above. ESEs were underrepresented ($p \leq 0.05$) in the poison exons of all genes except SRp38 and SRp55, and were overrepresented ($p \leq 0.05$) in the flanking introns of SRp20, SRp30c, and SRp75.

miRNA target site identification

We addressed the possibility that microRNA binding sites may be present in the highly conserved alternative regions of SR genes. miRNAs are thought to bind to 3' UTRs of mRNAs, but they may bind elsewhere in the mRNA. Such miRNA binding could affect the function of the alternative mRNAs or perhaps even the regulation of their alternative splicing. miRanda v1.0b was used to identify potential miRNA target sites in the human SR mRNAs[10]. We did not consider conservation in the predictions to avoid biasing the analysis towards the ultraconserved regions. Using the default cutoffs (score = 50, energy = -20) and a set of 470 mature human miRNAs downloaded from miRBase release 9.0[11], we identified many putative target sites throughout the mRNA of each SR gene. We compared the number of putative targets between the alternative region of each SR mRNA and the constitutive regions of that mRNA, using a χ^2 test with no multiple-test correction. miRNA target sites were overrepresented in the poison cassette exon of SRp30c ($p \leq 0.001$) and the 3'UTR cassette exon of SC35 ($p \leq 0.025$) and were underrepresented in the 3'UTR regions that are removed in the PTC^+ isoforms of ASF and SRp46 ($p \leq 0.01$). The remaining genes had no statistically significant difference in miRNA target sites between their alternative and constitutive regions. From these results, we conclude that miRNA binding is not a substantial contribution to the unusual sequence conservation throughout the family. In fact, the SRp30c cassette exon is the least-conserved of the SR cassette exons. However, some of the predicted miRNA targets may indeed be biologically important.

Conservation of unproductive splicing in *Ciona intestinalis*

We identified SR protein orthologs in release 1.0 of the *Ciona intestinalis* genome[12] by searching for protein sequences with SR-like domain structures. Sequences that contained one or two regions matching the Pfam RRM profile[13], had no Pfam domains other than those found in SR proteins, and had at least one C-terminal RSRS or SRSR repeat were collected. A phylogenetic tree incorporating the human and *C. intestinalis* SR proteins was constructed as described in Figure S1. We identified 13 probable SR proteins, some with unique orthologs (data not shown). Based on this result, we examined the alternative splicing for one example, the *C. intestinalis* gene ci0100137777, which appeared to be a unique ortholog of human SRp20. We used BLAST to identify corresponding ESTs from dbEST (downloaded June 2005). Analysis of alternative splice forms was performed as described for human and mouse. 94 ESTs aligned to the ci0100137777 locus. 12 ESTs supported the existence of a 260-nt cassette exon between exons 2 and 3 of the reference form, including an in-frame stop codon at nt 50 of the exon (Figure S5). ESTs also showed retention of the flanking introns. No sequence similarity was detectable with the human SRp20 poison cassette exon, but the poison exons in both species interrupt the normal protein sequence at approximately the same location

II. Supplementary Figures

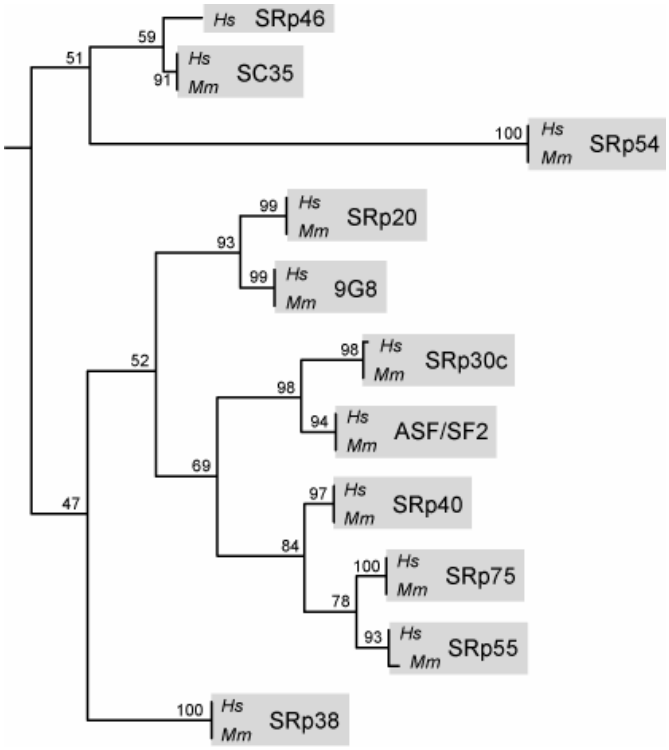
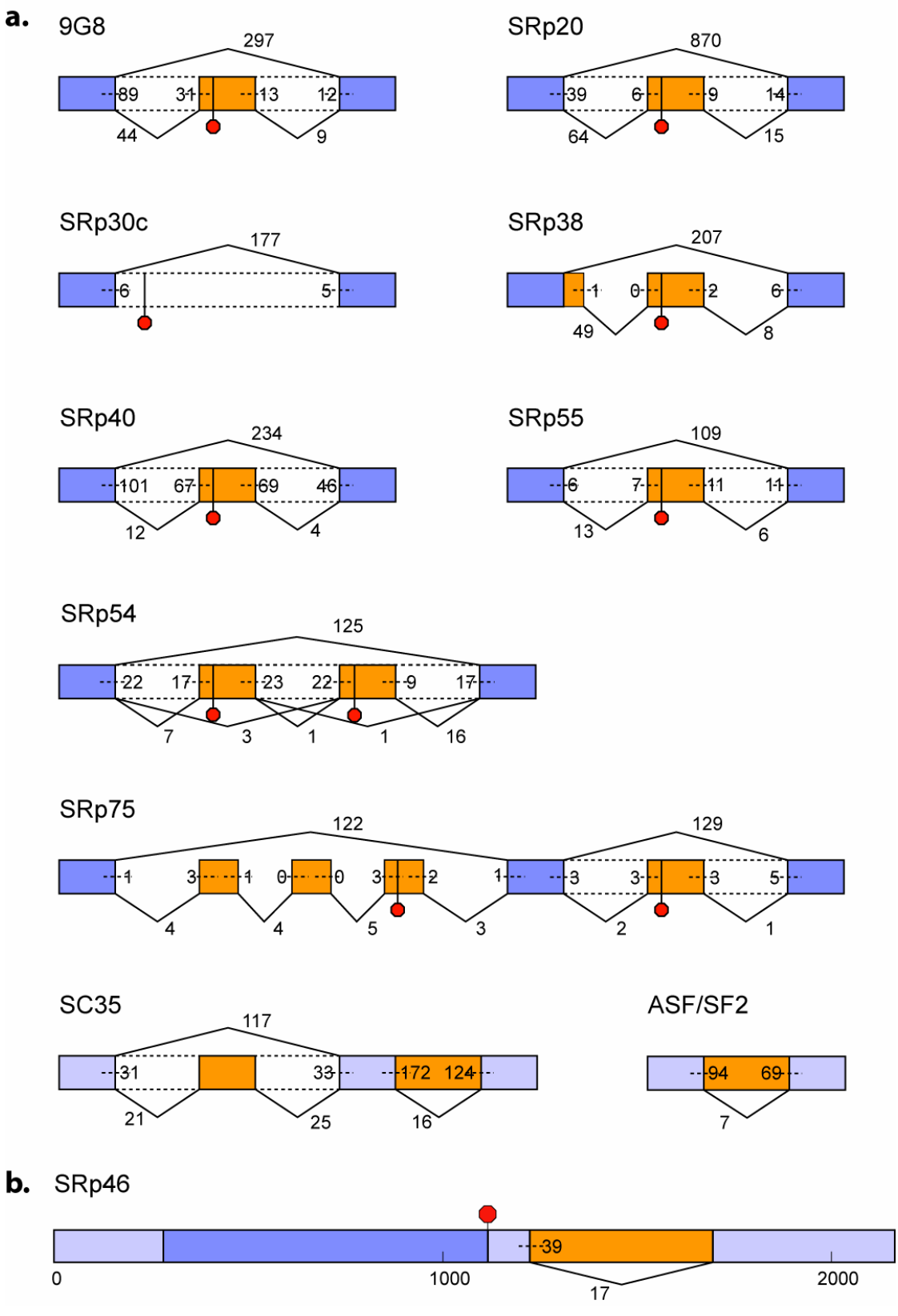


Figure S1 – The human SR protein family and their mouse orthologs.
A phylogenetic tree shows the 11 human (Hs) SR proteins and their 10 mouse (Mm) orthologs; human SRp46 is an expressed retropseudogene of SC35[14]. The N-terminal RNA-binding domain (RRM) sequence of each human and mouse SR protein were aligned using MUSCLE v3.6 with default options[15]. A maximum likelihood tree was constructed and bootstrapped with the Phylip `proml` program and rooted using the RRM of human PTB (NP_002810.1) as an outgroup[16].

Figure S2 (following page) – EST evidence for alternative splicing.
a) The counts indicate the number of ESTs supporting each splice junction: reference (top), alternative (bottom), and intron retention (dashed lines). The counts excluded ESTs in which the junction location is shifted due to alignment or sequencing ambiguities; this excluded some legitimate "NAGNAG" tandem acceptor splice sites that are shifted by 3nt. ESTs were counted as intron retention if they extended at least 10nt on both sides of the splice site, into the adjacent exon and intron. This method was chosen because ESTs are often too short to show retention of an entire intron. Exon and intron sizes and stop codon (red octagon) locations are not to scale.
b) SRp46, an expressed retropseudogene of SC35[14], is alternatively spliced in its 3' UTR. The reference mRNA is a single exon. The untranslated regions, coding region, and alternative region are shown (to scale) in light blue, dark blue, and orange, respectively. ESTs are counted as described in part a).



(**Figure S2** – caption on previous page)

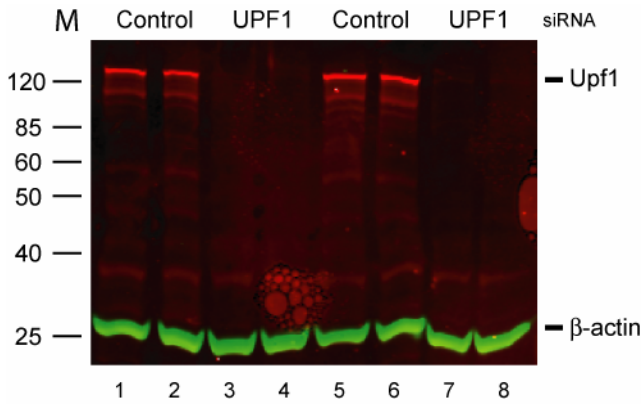
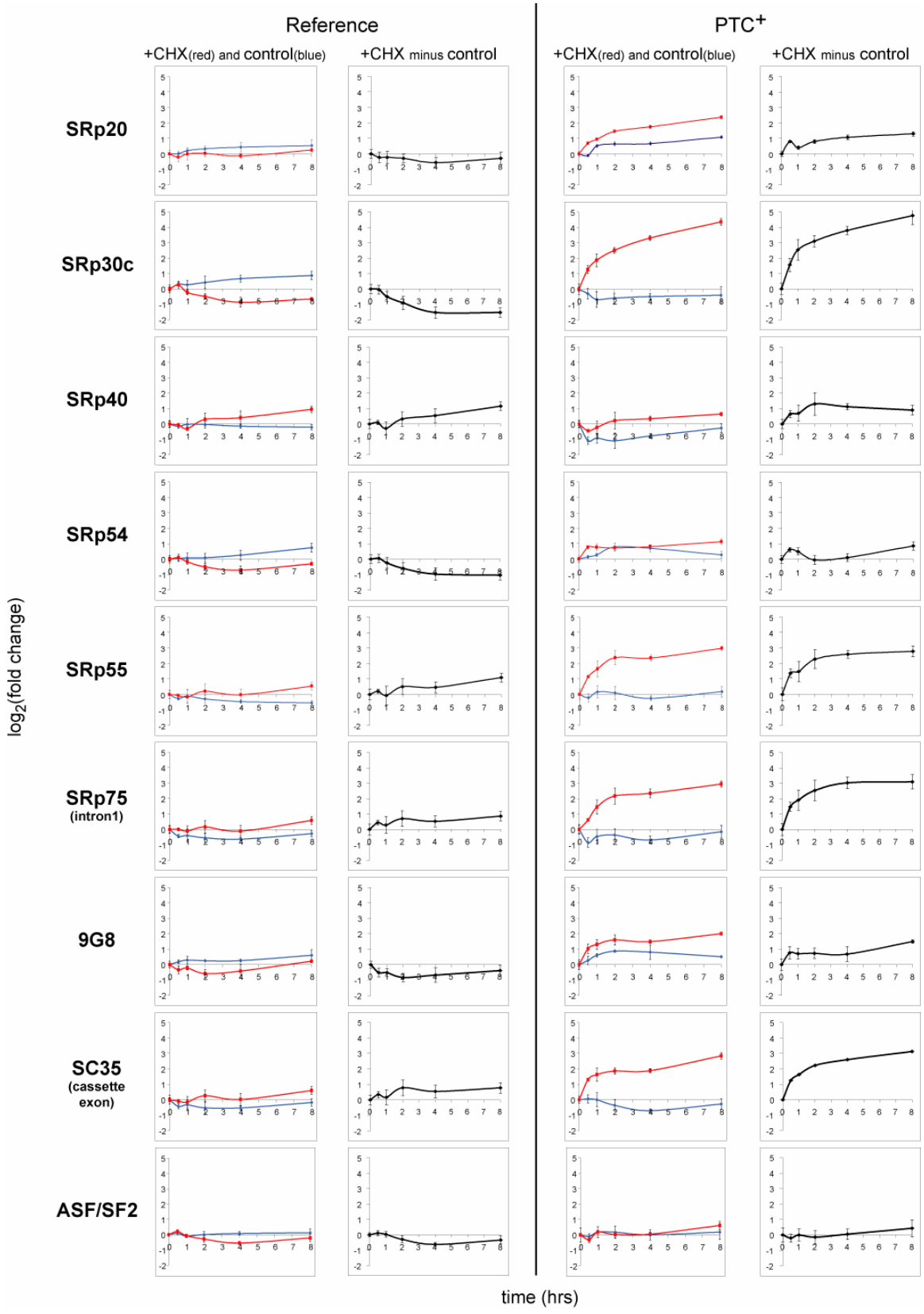


Figure S3 – Western blot of siRNA-mediated Upf1 depletion in HeLa cells
Protein lysates from HeLa cells expressing control (lane 1, 2, 5, 6) or UPF1 (lane 3, 4, 7, 8) siRNAs were electrophoresed in 7.5% SDS-polyacrylamide, transferred to nitrocellulose and probed with anti-Upf1 (red) (Bethyl) and anti-beta actin (green) (Sigma) antibodies. Immunoblots were scanned and analyzed (LI-COR Odyssey). HeLa cells were treated with 1.5ug/ml puromycin to select for the shRNA expressing plasmids for 72hrs (lane 1, 2, 3, 4) or 96hrs (lane 5, 6, 7, 8) and harvested 5 days post-transfection. Upf1 levels were depleted to less than 5% of control cells (lane 3, 4, 7, 8), normalizing to beta actin levels as a loading control. Molecular weight (kD) markers are indicated.

Figure S4 (following page) – Cycloheximide treatment of HeLa cells results in rapid stabilization of PTC$^+$ containing isoforms in the SR family
Quantitative RT-PCR analyses of reference ('a' or 'e' probe) and PTC-containing ('b' or 'f' probe) mRNAs for each SR gene were conducted. We measured relative levels of transcripts between HeLa cells treated with 10ug/ml cycloheximide (CHX) or DMSO during an 8 hour time course. The data are averages for 3 PCR replicates normalized by the average of 3 control genes: beta actin, SDHA and TBP. The $\log_2$ fold changes in SR transcript levels were plotted in two ways: fold change across individual cycloheximide-treated (red) and control (blue) time courses normalized to the 0hr time point or fold change in cycloheximide-treated cells minus fold change in the control cells (black). Reference isoforms of SRp38 and SRp46 were not detected and these genes were therefore excluded from our analyses.



(**Figure S4** – caption on previous page)

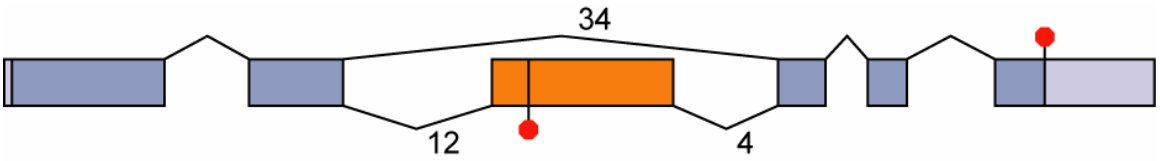


Figure S5 – Alternative splicing of the *Ciona intestinalis* SRp20 ortholog
A poison cassette exon is included in a fraction of *C. intestinalis* SRp20 transcripts. The number of ESTs including the reference splice event (top) or the alternative poison cassette exon (bottom) is displayed. The locations of the normal (top) and premature (bottom) termination codons are indicated (red octagon).

III. Supplementary Tables

Supplementay Table 1 - SR Gene Sequences

Gene	Unigene Cluster	Reference mRNA
SRP20	Hs.405144	NM_003017.3
	Mm.358634	NM_013663.3
SRP30c	Hs.567375	NM_003769.2
	Mm.287826	NM_025573.2
SRP38	Hs.3530	NM_054016.1
	Mm.10229	AB015894
SRP40	Hs.166975	CR595330
	Mm.43331	BC082593
SRP46	Hs.476680	NM_032102.1
SRP54*	Hs.479693	NM_004768.2
SRP55	Hs.6891	NM_006275.4
	Mm.24042	NM_026499.2
SRP75	Hs.469970	NM_005626.3
	Mm.2478	AK156989
9G8	Hs.309090	NM_001031684.1
	Mm.292016	NM_146083.1
ASF/SF2	Hs.68714	NM_006924.3
	Mm.45645	BC058627
SC35	Hs.73965	NM_003016.2
	Mm.21841	NM_011358.1

*The Entrez Gene entry for mouse SRp54 was truncated at the 5' end and no single mRNA spanned the entire locus, based on EST alignments and comparison with human SRp54. AK146719 and BC037053 were combined into a full-length reference mRNA. The UniGene clusters Mm.223946, Mm.381290, and Mm.381291 corresponded to different parts of the full locus and all three were included in the splicing analysis.

Supplementary Table 2 - Human SR alternative splicing coordinates

Gene	Genomic Locus	Alternative exon(s) or intron, relative to locus
SRP20	NC_000006.9:36669644-36679687	5933-6388
SRP30c	NC_000012.9:c119370778-119361692	5434-5523
SRP38	NC_000001.8:c24052627-24037745	2120-2179, 2559-2921, 5752-5754
SRP40	NC_000014.7:69303063-69308975	2919-3450
SRP46	NC_000011.8:94439257-94442434	1740-2210
SRP54	NC_000001.8:70382886-70430205	25914-26022, 26678-26794
SRP55	NC_000020.9:41519432-41526155	1776-2043
SRP75	NC_000001.8:c29329477-29294481	13712-13847, 15438-15621, 16286-16460, 22292-22486
9G8	NC_000002.9:c38890602-38881894	2464-2912
ASF/SF2	NC_000017.9:c53440087-53435354	2806-3726
SC35	NC_000017.9:c72245507-72241292	1956-2059, 2771-3268

Coordinates indicate cassette exons except in the case of ASF/SF2, SC35, and SRp46, whose coordinates indicate alternative 3' UTR introns, and SRp38, whose coordinates indicate a cassette exon and alternative ends of the upstream and downstream exons (see figure 1 for gene structures). Genomic locus coordinates are based on the position defined in Entrez Gene plus 500 nt on either side. Splicing event coordinates are relative to that locus.

Supplementary Table 3 – EST evidence for PTC$^+$ alternative splice forms of human SR genes

Gene	Alternatively-spliced ESTs			Intron retention ESTs	
	event	number	%	number	%
SRp20	cassette exon	64	7%	39	4%
SRp30c	cassette exon	0	0%	6	3%
SRp38	cassette exon (alt 5' ss)	49	19%		
SRp40	cassette exon	12	5%	101	29%
SRp54	cassette exons (2)	7, 3	5%, 2%	22	14%
SRp55	cassette exon	13	11%	6	5%
SRp75	cassette exons (3)	4 ,4, 5	3%, 3%, 4%		
	cassette exon (Hs only)	2	2%	3	2%
9G8	cassette exon	44	13%	89	21%
ASF/SF2	3' UTR intron	7	7%		
SC35	3' UTR cassette exon	21	15%	31	18%
	3' UTR intron	16	9%		
SRp46	3' UTR intron	17	30%		

Percent ESTs is calculated from the number of ESTs supporting alternative event divided by total number of ESTs that encompass the region. For cassette exons, the percent ESTs is based on ESTs including the upstream junction of the cassette exon. For intron retention, the percent ESTs is based on ESTs covering the upstream 5' splice site and extending into the intron. See Supplementary Information. SRp54 and SRp75 have 2 and 3 tandem cassette exons, respectively.

Supplementary Table 4 - Quantitative RT-PCR Oligos

Gene	Detected Isoform	Probe Type	Sequence
SRP20	Constitutive	Exon	GTGAAAAAAGAAGTAGAAATCGTGG
		Exon	ACTCCTCCTACGATAATCATCTCGA
	Reference (a)	Exon	GTGAAAAAAGAAGTAGAAATCGTGG
		Junction	CTCCTTCTTGGAGATCTGCGACGAG
	PTC+ (b)	Junction	TCCACCTCGTCGCAGAGTCACCATC
		Exon	TCATGTGAAACGACACCAGCCAAGC
	PTC+ (c)	Exon	CAACTAGCCCTTTCAGCGTCATGTG
		Exon	TCATGTGAAACGACACCAGCCAAGC
SRP30c	Constitutive	Exon	CCGCTCCGCGTTCCCACAATGCAGT
		Exon	ACGAGGCCGTGCCGGTTCTTGAGCT
	Reference (a)	Junction	GAGTTCTTGTTTCAGGACTTCCTCC
		Exon	CATGAGAGCGGAATTTGGTGTCATC
	PTC+ (b)	Junction	GAGTTCTTGTTTCAGAAAACTATTA
		Exon	GTGTTTACCGCTCACAGCTCAGGAC
	PTC+ (d)	Junction	CTTTCATGACATCAGGACTTCCTCC
		Exon	CATGAGAGCGGAATTTGGTGTCATC
SRP40	Constitutive	Exon	TAAGTGCGTCAGTTGTGGAGTGGCG
		Exon	GACCTCCACGCAGACCAGGACTTAA
	Reference (a)	Junction	AGAGTCAGCTGGCAGGATCTCAAAG
		Exon	TAGGTCGGTGTGCATCCGCAAACGT
	PTC+ (b)	Junction	AGAGTCAGCTGGCAGCCTGTCTGTG
		Exon	ATCACTGTAGGAGCTGACTGGCAAA
SRP54	Constitutive	Exon	CCAGATGTCGTCAGCACCAGCGCCT
		Exon	GATTGAAGTAATTCCAGGACGCTAG
	Reference (a)	Exon	CTCATCTCGTGTCTGCTTTGTTAAG
		Junction	CATCAGGAATAACTCCTTCTGCATA
	PTC+ (b)	Junction	TACCATATGCAGAAGTTCTTTTCCC
		Exon	AATATGTACAGGGCACTTAAGAGAA
	PTC+ (d)	Exon	AGAACCGACTGTGCTTGCTTACGTT
		Junction	CATCAGGAATAACTCCTGTACTCGA
SRP55	Constitutive	Exon	CCCCTGTGGTGTGAGGCGCGTGTTC
		Exon	CCTTCTCCCGGACGTTGTAGCTCAG
	Reference (a)	Junction	GCTACGGAAGCCGCAGTGGTGGAGG
		Exon	ATCTTGCCAACTGCACCGACTAGAA
	PTC+ (b)	Junction	GCTACGGAAGCCGCATGACCAATGG
		Exon	GGCCACAAAACACGCAAGGTAACAG
	PTC+ (d)	Junction	ACCTCACACTGTGATGTGGTGGAGG
		Exon	ATCTTGCCAACTGCACCGACTAGAA
	PTC+ (b/d)	Junction	GCTACGGAAGCCGCATGACCAATGG
		Junction	ATCCACCTCCACCACATCACAGTGT
SRP75	Constitutive	Exon	GTTGCCGGGATGCCGCGGGTGTACA
		Exon	GGATCTTCCCGTAGCCCTTAAAGAA
	Reference (a) reference intron 1	Junction	GGATCTGAAGAACGGATATGGTTTT
		Exon	GCCGCGGGCATGCTCAACAATTACT
	PTC+ (b)	Junction	GGATCTGAAGAACGGTCTGTTATGT

		Exon	TCACTCGTCTTTTGGTTCCCATTAG
	PTC+ (d)	Junction	AGTTTTAGGACAAAGATATGGTTTT
		Exon	GCCGCGGGCATGCTCAACAATTACT
	Reference (a) reference intron 2	Junction	ACGGTTCTGGACGCAGTGGATATGG
		Exon	CAGCTGCACCGACTTGACAAATTCT
	PTC+ (b)	Junction	ACGGTTCTGGACGCAAATCCTTTGA
		Exon	CCAGGTCTAATGGGGAAAGGTTGCG
9G8	Constitutive	Exon	TCAGCTTTTGCGTCACTCGAGCCCT
		Exon	ACCCGCGTGCTCGGCTCTTTAGCAA
	Reference (a)	Junction	GCGAAGAAGAAGCAGGTCACGGTCT
		Exon	TGCGTGAGCGAGAGTATCGCCTTCC
	PTC+ (b)	Junction	GCGAAGAAGAAGCAGGTTTCTTCGT
		Exon	CAAGGTCTAGGAGGATCTTAATTGT
	PTC+ (c)	Exon	GACCACATTCCCAACTAGAAAAATC
		Exon	AGATGTTTTCTTCAATCTAACGACG
ASF/SF2	Constitutive	Exon	AGGGAACAACGATTGCCGCATCTAC
		Exon	ATGTCGCGGATAGCGCCGTATTTGT
	Reference (e)	Intron	ACAGATGAAATTGGCAGTATTGACC
		Intron	ATTTTGCCACAATTGCCAAGGTTTA
	PTC+ (f)	Junction	ATAATGGAGGCAATGGTTTGGATTG
		Exon	TAAAAAAATCCACACGAATGCGGTT
SC35	Constitutive	Exon	AGTTGTTACTCAGGTGCGCTAGCCT
		Exon	AGGCCTTGCCGCAGAACAGCACGGA
	Reference (a) cassette exon	Junction	CCTCTTAAGAAAATGATGTATCGGC
		Exon	TCCAAGGACTCTTCTTCGATGGACT
	PTC+ (b)	Junction	CCTCTTAAGAAAATGCTGCGGTCTC
		Exon	ATCAGCCAAATCAGTTAAAATCTGC
	PTC+ (f) intron	Junction	CTGGCTATTGAAAAGGCGCAGTTGT
		Exon	CATCTTTACACAGGAGCAATCGGGA
	PTC+ (a/f)	Junction	CCTCTTAAGAAAATGATGTATCGGC
		Junction	GCTACACAACTGCGCCTTTTCAATA
β-actin	Normalization	Exon	CTGGAACGGTGAAGGTGACA
		Exon	AAGGGACTTCCTGTAACAATGCA
SDHA	Normalization	Exon	TGGGAACAAGAGGGCATCTG
		Exon	CCACCACTGCATCAAATTCATG
TBP	Normalization	Exon	TGTTTCTTGGCGTGTGAAGATAACC
		Exon	AGAAACCCTTGCGCTGGAACTCGTC

Exon and intron primers were picked using the ArrayOligoSelector program[3]. Each primer pair is listed with the 5' primer first then the 3' primer. Junction primers were designed with 10 nucleotides on the side of the splice junction to be extended and 15 nucleotides on the other side of the junction. Constitutive exon primers are in exon 1 of each gene except for SRp20, where the constitutive exon primers are in exon 3. Respective locations of primer pairs in each SR gene are described (a-f) according to the key in Figure 2. Primers to the control genes beta actin, SDHA, and TBP were selected as previously described[17].

Supplementary Table 5 - Calculated PTC+ isoform levels as a percentage of total

Gene	Isoform (probe)	log_2(fold change)	control (%)	UPF1 depleted (%)
SRp20	Total	0.4		
	Refseq (a)	-0.6		
	PTC+ (b)	3.5	6	52
SRp30c	Total	0.9		
	Refseq (a)	0.0		
	PTC+ (b) *	5.4	2	49
SRp40	Total	-0.1		
	Refseq (a)	-0.7		
	PTC+ (b)	1.8	12	43
SRp54	Total	0.2		
	Refseq (a)	-0.4		
	PTC+ (b)	1.9	13	43
SRp55	Total	0.5		
	Refseq (a)	-1.1		
	PTC+ (b)	4.3	5	69
SRp75	Total	0.9		
	Refseq (a)	-0.2		
	PTC+ (b)	3.7	8	58
9G8	Total	0.4		
	Refseq (a)	-1.2		
	PTC+ (b)	3.2	10	71
ASF/SF2	Total	-0.6		
	Refseq (e)	-1.3		
	PTC+ (f)	2.0	7	41
SC35	Total	0.2		
	Refseq (a)	-1.0		
	PTC+ (b)	3.2	8	62

PTC^+ isoform levels were estimated as a percentage of each gene's total spliced mRNA, as described in supplementary methods. The values provided for SRp75 represent the tandem cassette exons in the first reference intron. The values provided for SC35 represent the cassette exon within the 3'UTR. Both genes had multiple PTC^+ isoforms and the observed fold change may include the effects of NMD on the other isoforms.
*The increase in the PTC^+ isoform levels of SRp30c could not be reliably determined because the PTC^+ isoform levels in control cells were at the lower limit of detection.

IV. Supplementary Notes

Literature Cited

1. Wheeler, D.L. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **34**, D173-80 (2006).
2. 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).
3. Bozdech, Z. et al. Expression profiling of the schizont and trophozoite stages of Plasmodium falciparum with a long-oligonucleotide microarray. *Genome Biol* **4**, R9 (2003).
4. http://www.rit.edu/~uphysics/uncertainties/Uncertaintiespart1.html.
5. Kent, W.J. et al. The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
6. Bejerano, G. et al. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**, 87-90 (2006).
7. Pearson, W.R. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. *Genomics* **11**, 635-50 (1991).
8. 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**, 3568-71 (2003).
9. Fairbrother, W.G., Yeh, R.F., Sharp, P.A. & Burge, C.B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007-13 (2002).
10. John, B. et al. Human MicroRNA targets. *PLoS Biol* **2**, e363 (2004).
11. Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A. & Enright, A.J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140-4 (2006).
12. Dehal, P. et al. The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. *Science* **298**, 2157-67 (2002).
13. Finn, R.D. et al. Pfam: clans, web tools and services. *Nucleic Acids Res* **34**, D247-51 (2006).
14. Soret, J. et al. Characterization of SRp46, a novel human SR splicing factor encoded by a PR264/SC35 retropseudogene. *Mol Cell Biol* **18**, 4924-34 (1998).
15. Edgar, R.C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-7 (2004).
16. Felsenstein, J. Phylip (Phylogeny Inference Package). 3.6 edn (Seattle, WA, 2005).
17. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, RESEARCH0034 (2002).