

Genes & Development

Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay

Julie Z. Ni, Leslie Grate, John Paul Donohue, Christine Preston, Naomi Nobida, Georgeann O'Brien, Lily Shiue, Tyson A. Clark, John E. Blume and Manuel Ares, Jr.

Genes & Dev. 2007 21: 708-718

Access the most recent version at doi:[10.1101/gad.1525507](https://doi.org/10.1101/gad.1525507)

Supplementary data

"Supplemental Research Data"

<http://www.genesdev.org/cgi/content/full/21/6/708/DC1>

References

This article cites 56 articles, 26 of which can be accessed free at:

<http://www.genesdev.org/cgi/content/full/21/6/708#References>

Article cited in:

<http://www.genesdev.org/cgi/content/full/21/6/708#otherarticles>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Notes

To subscribe to *Genes and Development* go to:
<http://www.genesdev.org/subscriptions/>



Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay

Julie Z. Ni,¹ Leslie Grate,¹ John Paul Donohue,¹ Christine Preston,² Naomi Nobida,² Georgeann O'Brien,² Lily Shiue,¹ Tyson A. Clark,³ John E. Blume,³ and Manuel Ares Jr.^{1,2,4}

¹Center for Molecular Biology of RNA and Department of Molecular, Cell, and Developmental Biology, University of California at Santa Cruz, Santa Cruz, California 95064, USA; ²Hughes Undergraduate Research Laboratory, University of California at Santa Cruz, Santa Cruz, California 95064, USA; ³Affymetrix, Inc., Santa Clara, California 95051, USA

Many alternative splicing events create RNAs with premature stop codons, suggesting that alternative splicing coupled with nonsense-mediated decay (AS-NMD) may regulate gene expression post-transcriptionally. We tested this idea in mice by blocking NMD and measuring changes in isoform representation using splicing-sensitive microarrays. We found a striking class of highly conserved stop codon-containing exons whose inclusion renders the transcript sensitive to NMD. A genomic search for additional examples identified >50 such exons in genes with a variety of functions. These exons are unusually frequent in genes that encode splicing activators and are unexpectedly enriched in the so-called "ultraconserved" elements in the mammalian lineage. Further analysis show that NMD of mRNAs for splicing activators such as SR proteins is triggered by splicing activation events, whereas NMD of the mRNAs for negatively acting hnRNP proteins is triggered by splicing repression, a polarity consistent with widespread homeostatic control of splicing regulator gene expression. We suggest that the extreme genomic conservation surrounding these regulatory splicing events within splicing factor genes demonstrates the evolutionary importance of maintaining tightly tuned homeostasis of RNA-binding protein levels in the vertebrate cell.

[*Keywords:* SR proteins; splicing microarray; hnRNP proteins; splicing factor; autogenous regulation; epigenetics]

Supplemental material is available at <http://www.genesdev.org>.

Received December 22, 2006; revised version accepted February 2, 2007.

The nonsense-mediated decay (NMD) pathway is an mRNA surveillance mechanism that limits the translation of mRNAs with premature termination codons (PTCs) (Conti and Izaurralde 2005; Lejeune and Maquat 2005; Maquat 2005). In mammals, PTCs are partially determined by their position relative to the activity of the splicing machinery. The spliceosome leaves a protein mark called the exon-junction complex (EJC) at each junction of the spliced mRNA (Le Hir et al. 2000, 2001). During the first (pioneer) round of translation, the ribosome removes the EJCs, and the mRNA is stabilized (Ishigaki et al. 2001; Chiu et al. 2004; Lejeune et al. 2004). If an EJC remains on the mRNA as the ribosome terminates, proteins within the EJC, notably the Upf3 or Upf3x and Upf2 proteins, recruit Upf1, which triggers

NMD-mediated destruction of the mRNA (Lykke-Andersen et al. 2000, 2001; Kim et al. 2001; Wang et al. 2001; Schell et al. 2003; Kashima et al. 2006). This relationship between NMD and splicing has consequences for gene structure: The most 3' exon-exon junction can be no >50–55 nucleotides (nt) downstream from the stop codon. This means that normal stop codons are nearly always found in the last exon, or within 55 nt of the end of the second-to-last exon of the gene (Nagy and Maquat 1998; Maquat 2005).

A second consequence of the relationship between NMD and splicing is that errors of splicing in which normal exons are skipped, or inappropriate intron segments are included will be destroyed by NMD if they generate a premature stop codon (Lejeune and Maquat 2005). Studies of human and mouse EST data predict that about one-fifth of the alternatively spliced human and mouse genes produce PTC-containing RNAs that should be subject to NMD (Lewis et al. 2003; Baek and Green 2005). How many of these are splicing errors is not clear

⁴Corresponding author.

E-MAIL ares@biology.ucsc.edu; FAX (831) 459-3737.

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.1525507>.

(Sorek et al. 2004). Besides removing splicing errors, NMD may also play an important role in regulating the expression of many genes by coupling alternative splicing (AS) to decay (Lewis et al. 2003; Baek and Green 2005).

The regulatory role of coupling between AS and NMD is supported by several lines of experimental evidence. The expression profile in Upf1-depleted HeLa cells shows that ~5% of expressed transcripts are up-regulated (Mendell et al. 2004). Splicing factors like polypyrimidine-binding protein (PTB) and serine/arginine-rich 2 protein (SC35) are known to autoregulate their own expression by binding to their own mRNA to increase splicing of the NMD-sensitive isoform (Sureau et al. 2001; Wollerton et al. 2004). In these cases the splicing regulation represses an exon, whose skipping creates a PTC by frame shifting (Wollerton et al. 2004), or activates the removal of cryptic introns in the 3' untranslated region (UTR), in essence introducing an EJC downstream from the normal stop codon (Sureau et al. 2001). Genes such as neural PTB (nPTB) (Rahman et al. 2002), TRA2- β (Stoilov et al. 2004), and SRp20 (Jumaa and Nielsen 1997), have been reported to create mRNA isoforms potentially encoding truncated proteins by alternative splicing, but their relationship to NMD has not been determined.

Recent work measuring PTC-containing RNAs by microarrays showed that a large proportion of them are expressed at low levels and are not sensitive to inhibition of NMD (Pan et al. 2006). This result goes against the hypothesis that AS-NMD is a widespread mechanism of gene regulation in vertebrates (Lewis et al. 2003; Mendell et al. 2004; Baek and Green 2005), and raises more questions concerning which NMD-coupled splicing isoforms in the database represent regulated splicing events, and which represent splicing errors (Lewis et al. 2003; Sorek et al. 2004; Baek and Green 2005; Pan et al. 2006). Since the levels of NMD-sensitive isoforms are kept low due to their efficient degradation, the database counts of ESTs for them is unlikely to reflect the frequency at which they are generated by splicing (Baek and Green 2005). It has thus been a challenge to reveal functionally important NMD-coupled splicing events genome-wide and study their regulation mechanisms.

In this paper we use splicing-sensitive microarrays to identify a group of alternative RNA isoforms that accumulate after inhibition of NMD. We identify a biologi-

cally relevant class of NMD-sensitive isoforms that arises by inclusion of highly conserved exons that contain stop codons in all three reading frames. We used the properties of the array-identified exons to create a program to scan the genome for other examples and found many. These conserved stop codon exons are particularly common in genes encoding splicing activators known to activate exon inclusion, and frequently overlap ultraconserved elements in mammalian genomes (Bejerano et al. 2004). Unlike previous examples of autogenous negative regulation of alternative splicing coupled to NMD (Sureau et al. 2001; Wollerton et al. 2004), the inclusion of these novel exons must be activated to block gene expression. We suggest that the exonic class of ultraconserved elements is critically required for homeostatic maintenance of splicing factor expression levels.

Results

Detection of AS-NMD using splicing-sensitive microarrays

Public databases of EST sequences hinted at the existence of coupling between alternative splicing and NMD. But discrimination of functionally important splicing from splicing errors in these ESTs has been a challenge (Sorek et al. 2004; Pan et al. 2006). To experimentally detect biologically relevant examples of AS-NMD, we blocked NMD and used splicing-sensitive microarrays (Clark et al. 2002; Wang et al. 2003; Ule et al. 2005; Pan et al. 2006; Sugnet et al. 2006), seeking alternatively spliced RNAs whose levels increased compared with controls. To identify effects resulting from NMD inhibition, rather than secondary effect specific to a particular treatment, we used two different methods of NMD inhibition. In one method, we transfected mouse N2A cells with Upf1 small interfering RNA (siRNA) to down-regulate Upf1, a core component of the NMD machinery (Mendell et al. 2002). Upf1 protein was reduced to at least 25% of its level in untreated cells or control siRNA-treated cells (Fig. 1A). In the second method, we treated N2A cells with the translation inhibitor emetine (Noensie and Dietz 2001), since translation is a prerequisite for NMD-mediated degradation. To test the effectiveness of these approaches, we measured accumulation of the alternative isoform resulting from skipping of

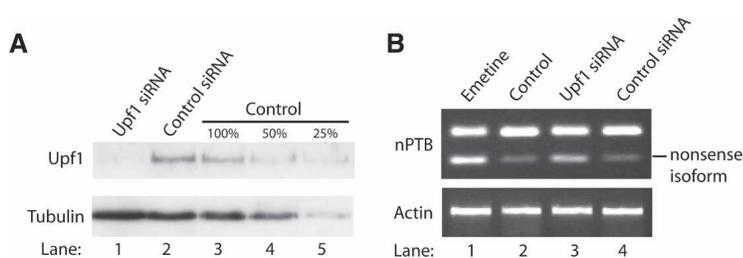


Figure 1. NMD pathway is blocked by Upf1 RNA interference (RNAi) and emetine treatment. (A) Western blot using antibodies against Upf1 and α -Tubulin. Lanes 1, 2, and 3 contain equal amounts of total cell protein 2 d after siRNA transfection. Lanes 4 and 5 contain 50% and 25% of lane 3. (B) RT-PCR of nPTB mRNA isoforms. nPTB primers are in constitutive exons 9 and 11 of nPTB mRNA. Exon 10 skipping will generate NMD-targeted nonsense isoforms. Nonsense isoforms are accumulated in Upf1 siRNA-transfected cells and emetine-treated cells. RT-PCR of β -actin mRNA was used as control.

Ni et al.

exon 10 of neural polypyrimidine tract-binding protein (nPTB) mRNA (Rahman et al. 2002) by RT-PCR. This isoform is predicted to create a PTC, and the data shows that both treatments effectively blocked NMD (Fig. 1B).

RNA from treated and untreated cells was converted to cDNA, labeled, and applied to Affymetrix alternative splicing microarrays as previously described (Wang et al. 2003; Ule et al. 2005; Sugnet et al. 2006). We analyzed the data using the method of Sugnet et al. (2006). Since the EST/mRNA sequences used for array design were not filtered by frequency (D. Kulp and M. Ares, unpubl.), many NMD-sensitive (and hence rarely present in EST libraries) RNAs can be measured by this array. Changes in levels of alternatively spliced isoforms were observed for all splicing modes (e.g., cassette exon, alternative 5' splice site, etc.) in each individual class of treatment (data not shown). We focused on the 229 cassette exons for which isoform representation in the RNA pool appeared to change under either of the NMD-inhibiting treatments (Supplementary Table S1). Using the University of California at Santa Cruz (UCSC) Genome Browser (Karolchik et al. 2003), we identified 28 cassette exons

for which the predicted NMD isoform accumulated when NMD was blocked (Table 1). We tested 25 of these by RT-PCR, and verified that 24 (96%) of them change as indicated by the array (Supplementary Table S1; see also Figs. 2, 3). We conclude that both Upf1 siRNA knock-down and emetine treatment lead to many changes in isoform representation, of which a little more than 10% can directly be attributed to NMD. The nearly 90% of changes in isoform representation that do not conform to known rules of NMD could be due to perturbation of other translation and Upf1-dependent decay mechanisms (Kim et al. 2005) or could be secondary effects of blocking NMD.

A class of conserved alternative exons brings stop codons into mRNA to trigger NMD

Previous studies of splicing and NMD have suggested a relationship between conserved and nonconserved alternative splicing events and NMD (Baek and Green 2005; Pan et al. 2006). To distinguish which of the 28 NMD isoforms detected by microarray might be generated by

Table 1. The 28 alternative cassette exon-splicing-induced NMD events from array data

NMD-inducing splicing event	Gene name	Description	Conservation score	RT-PCR tests			
Exon-skipping-induced NMD	Frameshift	<i>NFYB</i>	Nuclear transcription factor Y subunit β	0.99	True		
		<i>DLGH4</i>	Post-synaptic density protein 95	0.97	True		
		<i>MRPL13</i>	60S ribosomal protein L13mitochondrial	0.98	True		
		<i>WBSCR22</i>	Putative methyltransferase HUSSY-03	0.91	True		
		<i>CCAR1</i>	Death inducer with SAP domain DIS	0.99	True		
		<i>AK035230</i>	Protein phosphatase 1 regulatory subunit 12A	0.99	ND		
		<i>CDC16</i>	Cell division cycle protein 16 homolog	0.99	NI		
		<i>CCT8</i>	T-complex protein 1, θ subunit	0.97	ND		
		<i>BRD8</i>	Bromodomain-containing protein 8	0.99	True		
		<i>RPN1</i>	Oligosaccharide protein glycosyltransferase	0.97	ND		
		<i>HSF1</i>	Heat-shock factor protein 1	0.98	True		
		<i>RPA1</i>	Replication protein A1	0.78	True		
		<i>NIPSNAP1</i>	NipSnap1 protein	0.99	True		
		Exon-inclusion-induced NMD	Inclusion of stop codon exon	<i>BRD2</i>	Female sterile homeotic-related protein Frg-1	0.93	True
				<i>FARSIB</i>	Phenylalanyl-tRNA synthetase β chain	No align	True
<i>NOL5</i>	Nucleolar protein 5, NOP58 homolog			0.98	True		
<i>SFRS4</i>	Splicing factor, arginine/serine-rich 4			0.98	True		
<i>NKIRAS1</i>	KAPPAB-RAS1			No align	True		
<i>RNPC2</i>	RNA-binding domain protein 2 (CAPER α)			0.99	True		
<i>1300007B12RIK</i>	Hypothetical protein			0.99	True		
<i>SRRM1</i>	Ser/Arg-related nuclear matrix protein			0.99	True		
<i>ZCCHC6</i>	Zn finger CCHC containing protein 6			0.99	True		
<i>D4ERTD196E</i>	Hypothetical cytochrome c family protein			0.94	True		
Exon-inclusion-induced NMD	New EJC in 3' UTR	<i>AK031374^B</i>	Hypothetical Ub-associated protein	0.92	True		
		<i>TLK1</i>	Serine/threonine-protein kinase touselled-like 1	0.99	True		
		<i>SFRS9^B</i>	Splicing factor, arginine/serine-rich 9	0.98	True		
		<i>G430041M01RIK</i>	Tra-2 protein homolog (TRA-2 α)	0.99	True		
		<i>HNRPD</i>	Heterogeneous nuclear ribonucleoprotein D0	0.99	True		

All stop codon exons have stop codons in all three reading frames except those indicated by ^B, which have stops in two frames including the coding frame of the protein. (ND) Not determined; (NI) not interpretable.

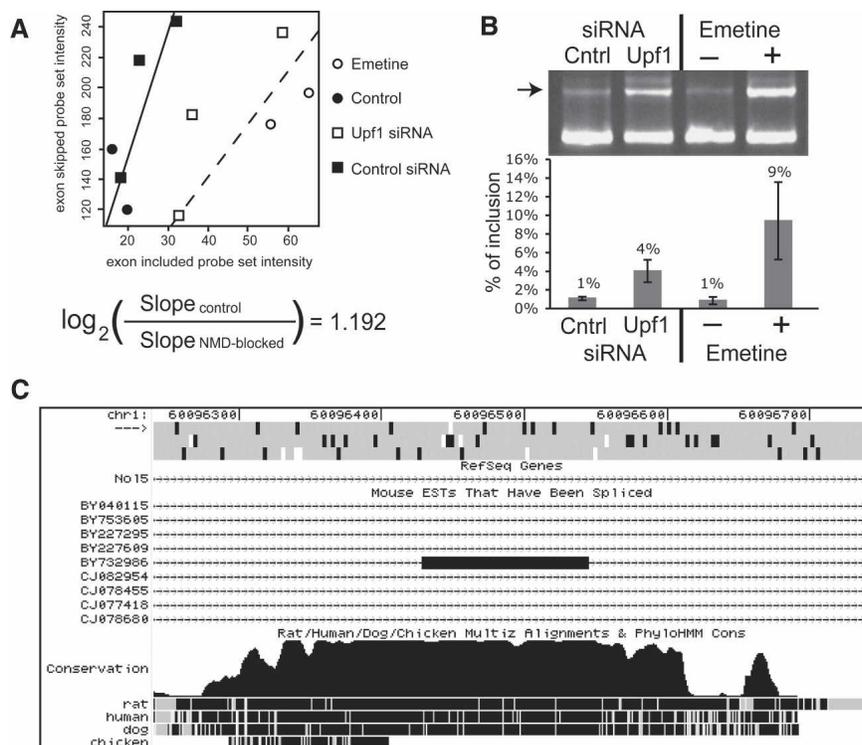


Figure 2. *NOL5*-splicing isoform with inclusion of stop codon exon is subject to NMD. (A) Microarray data of skip probe set intensity versus include probe set intensity for *NOL5* stop codon exon before and after blocking NMD. The lines of the plot represent the robust regression coefficient (constrained to go through the origin) for NMD-blocked sample groups (open circles, emetine; open squares, Upf1 siRNA treated) or non-NMD-blocked sample groups (filled circles, untreated cells; filled squares, control siRNA treated). The \log_2 difference in the slopes is 1.1922, indicating 2.3-fold inclusion in NMD-blocked cells relative to non-NMD-blocked cells for this exon (Sugnet et al. 2006). (B) RT-PCR validation of accumulation of *NOL5* stop codon exon after blocking NMD. The stop codon exon-inclusion isoform (nonsense isoform) is specifically accumulated upon Upf1 siRNA transfection or emetine treatment as compared with control cell samples. Percentage of exon inclusion increases fourfold by Upf1 siRNA transfection or ninefold by emetine treatment, as measured by an Agilent Bioanalyzer. (C) *NOL5* stop codon exon as seen in the UCSC Genome Browser (Karolchik et al.

2003). Base position shows three reading frames. In each reading frame, black squares represent stop codons and the white squares represent methionine codons, whereas the gray squares represent nonmethionine amino acid codons. The *NOL5* exon has stop codons in all three reading frames. The conservation track at the bottom shows high conservation in exonic and flanking intronic sequences of the *NOL5* stop codon exon.

functionally relevant splicing rather than by splicing errors, we grouped the 28 splicing events into four classes according to whether exon skipping or inclusion triggered NMD, and whether the exon was conserved in vertebrate genomes or not. All of the exons whose skipping triggered NMD were >78% conserved as indicated by their conservation score (Table 1). One known AS-NMD mechanism is the regulated skipping of a conserved protein coding exon to trigger NMD, which has been reported in the PTB gene (Wollerton et al. 2004). However, the skipping of conserved protein coding exons has been frequently observed, and can be interpreted as a splicing error (Sorek et al. 2004; Baek and Green 2005). Unfortunately, in the exon-skipping NMD class of events it is difficult to separate exon conservation that could be required for splicing regulation from that required to code for protein. This creates doubt about whether regulated exon-skipping NMD or splicing error is the underlying cause of this alternative splicing. Some of the exon-skipping NMD exons in Table 1 are embedded in conserved intron sequences, similar to those in the exon-skipping NMD-regulated PTB (Wollerton et al. 2004) and nPTB genes (Fig. 1; Rahman et al. 2004). This suggests that these conserved exons could be new examples of genes regulated by the coupling of alternative splicing and NMD, but additional experimentation is required to test this.

The situation is different for the exon-inclusion NMD

class (Table 1). The majority of the exon-inclusion NMD events we found belong to an unexpected class of highly conserved, nonprotein coding exons containing in-frame stop codons (Table 1). A detailed example is shown for *NOL5*, a nucleolar protein homologous to yeast Nop58p, component of C/D-Box methylation guide snoRNPs (Fig. 2). A plot of the skip/include ratios calculated from the plots of data from each array for the NMD inhibited (Fig. 2A, filled characters) and control groups (Fig. 2A, open figures) shows that the inclusion of the *NOL5* exon increases when NMD is blocked (Fig. 2A). The \log_2 of the difference in the slopes derived from the NMD-inhibited (Fig. 2A, solid line) and control groups (Fig. 2A, dotted line) indicates a slightly more than twofold increase in spliced forms containing the *NOL5* stop codon exon (Fig. 2A). RT-PCR followed by quantitation of the PCR products using an Agilent Bioanalyzer indicates that the levels of stop codon exon-containing transcripts increases from ~1% to between 4% and 9% depending on the treatment used to block NMD (Fig. 2B). Inspection of the *NOL5* exon in the UCSC Genome Browser (Fig. 2C) reveals that it and its adjacent intron sequences are highly conserved (Fig. 2C, conservation track) despite the inability to encode protein (stop codons in all reading frames, black bars in the sequence track at top in Fig. 2C).

The *NOL5* exon is typical (Table 1). Most members of this class of exons are >90% conserved in human,

42 RT-PCR tests for stop codon exons

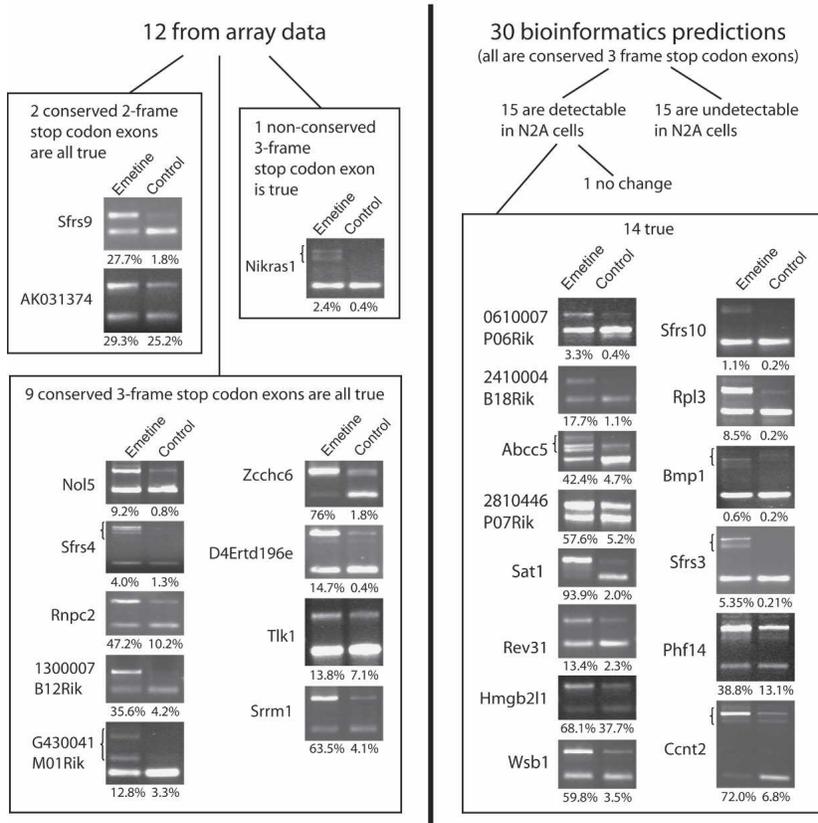


Figure 3. RT-PCR tests of stop codon exons. Twelve stop codon exons found by array and 30 conserved three-frame stop codon exons were tested by RT-PCR. The upper bands of each test are stop codon exon-inclusion isoforms. Bracket shows more than one stop codon exon-inclusion isoform was detected. Percentage of exon inclusion is labeled at the bottom of the corresponding exon and treatment and is a molar percent as determined using an Agilent Bioanalyzer. True validation result means that stop codon exon-inclusion isoforms increased when blocking NMD by emetine. All 12 stop codon exons found by array are true. Fourteen of 15 (94%) detectable bioinformatics-predicted stop codon exons are true.

mouse, rat, and dog, as indicated by the conservation probability calculated and displayed in the UCSC Genome Browser (Fig. 2C; Supplementary Table S2; Karolchik et al. 2003; Siepel et al. 2005). These stop codon exons often maintain high conservation well into flanking intronic regions (Fig. 2C; Supplementary Table S4), as is seen for exons whose alternative splicing is conserved (Sorek and Ast 2003; Sugnet et al. 2004, 2006; Yeo et al. 2005). The stop codon exons found in *RNPC2* and *TRA2A* overlap the so-called "ultraconserved" elements in the vertebrate genome, previously defined to be the 481 regions of ≥ 200 base pair (bp) without variation in the human, mouse, and rat genomes (Bejerano et al. 2004). Unlike the exon-skipping NMD exons, conservation of the stop codon exons cannot be due to protein coding, and therefore strongly implies a regulatory function for these exons.

Within this group of exons we found two that are present only in the mouse and do not appear in other vertebrates (*FARSLB* and *NKIRAS1*). These could represent important species-specific splicing (Modrek and Lee 2003), but could also be recently exonized short interspersed repetitive elements (SINEs) (see Lev-Maor et al. 2003; Bejerano et al. 2006; also see below). There are many examples of such exons in the genome and both exons show homology mouse SINE elements (Supplementary Table S2). These two SINE-derived exons were presumably identified along with the conserved class be-

cause they are among the most highly expressed and actively decayed SINE-derived exons in the mouse N2A cells we tested.

A class of conserved stop codon exons in mammalian genomes

Our array experiment is limited to splicing events that are represented on our array and are expressed in mouse neuroblastoma cells at a level that meets our stringent criteria for detection and analysis (Sugnet et al. 2006). To find additional conserved stop codon exons in mammalian genomes, we searched mouse and other mammalian EST and cDNA databases for internal (i.e., not 5' or 3' UTR) exons with $\geq 80\%$ sequence conservation in mouse, human, rat, and dog carrying stop codons in all three reading frames (Supplementary Table S2). We found 55 more such exons in the mouse genome. Together with the 11 conserved stop codon exons found by the NMD experiment (most of which were also detected in the bioinformatic search), this group of 66 exons resides in 65 different genes (Supplementary Table S2). Using the mRNA and spliced EST alignments in the UCSC Genome Browser (Karolchik et al. 2003) we found evidence for skipping of 45 of the 55 new stop codon exons. The remaining 10 are internal exons that have alternative 5' or 3' splice sites that add new exonic segments containing stop codons in all three frames. In principle,

these alternative 5' or 3' splice sites would function identically to a cassette stop codon exon to trigger NMD.

We tested 30 of the 55 new stop codon exons found in the computational search by RT-PCR of N2A cell RNA, and 15 were detectably expressed in both the emetine-treated and untreated control N2A cells (Fig. 3). Of these, 14 (93%) clearly show accumulation of nonsense isoforms upon blocking NMD with emetine, and one showed no change (Fig. 3). We conclude that the majority of conserved stop codon exons are actively included in pre-mRNA by alternative splicing, and destabilize the RNA isoforms that contain them through NMD (Fig. 3). From the conservation of these exons and the consequence of their inclusion for expression of the genes that carry them, we infer that stop codon exons are an important negative regulatory element for post-transcriptional gene regulation. We define a stop codon exon as a conserved, alternatively spliced, noncoding exon internal to a transcript that triggers NMD.

Stop codon exons are enriched in RNA-splicing factor genes and ultraconserved elements

To ask whether conserved stop codon exons are found more often in certain functional gene classes than others, we analyzed the functional annotations associated with the 65 stop codon exon-containing genes in the mouse genome using FuncAssociate (Berriz et al. 2003). Of these, 11 genes function in RNA splicing (GO Biological Process GO: 0,008,380) (Supplementary Table S3). The likelihood that the association of stop codon exons with splicing factors is due to chance is small (p value = 3.60×10^{-10}) (Supplementary Table S3).

Because of the unexpectedly high conservation of stop codon exons and their association with RNA-splicing proteins, we examined the distribution of ultraconserved element among stop codon exons. There are ~136,000 exons in the UCSC Genome Browser Known Genes track (Karolchik et al. 2003), and 111 exonic ultracon-

served elements (Bejerano et al. 2004). Of the 66 conserved stop codon exons (in 65 different genes) we identified in the mouse, the human orthologs of nine overlap or are entirely contained within ultraconserved elements. This is an association statistically unlikely to be due to chance (p value = 4.1×10^{-18} , Fisher's exact test). Of the nine ultraconserved elements that overlap stop codon exons, eight are found in seven different RNA-splicing-associated genes (*RNPC2*, *TRA2A*, *SFRS3*, *SFRS6*, *SFRS7*, *SFRS10*, and *TIAL1*). Previously, Bejerano et al. (2004) demonstrated that the "exonic" class of ultraconserved elements is associated with RNA-splicing and -binding protein genes. Combining this finding with our data showing that stop codon exons are strongly enriched in splicing factor genes and frequently contained within ultraconserved elements, we conclude that the exonic class of ultraconserved elements contains members essential for the proper regulation of splicing factor genes by AS-NMD (Table 2).

Exonic ultraconserved elements within splicing factor genes generally link alternative splicing to NMD with a polarity consistent with autogenous regulation

Given the link between stop codon exons, ultraconserved elements, and exon-inclusion NMD, we searched all the human exonic ultraconserved elements in RNA-binding protein genes (Bejerano et al. 2004) using the UCSC Genome Browser (Karolchik et al. 2003) for more instances of ultraconserved elements connecting alternative splicing to NMD. Of the 29 exonic ultraconserved elements in RNA-binding protein genes in human (Bejerano et al. 2004), 15 have human and/or mouse EST evidence suggesting the presence of AS-NMD in those regions. Of the 15, eight are stop codon exons, and the other seven appear to link alternative splicing to NMD regulation via other kinds of splicing events such as exon skipping or activation of a 3' UTR intron (Table 2). One example is the ultraconserved element of the splicing

Table 2. Human ultraconserved elements in splicing factor genes host-splicing-NMD regulation

Splicing regulator	Ultra element	Gene	Protein	Splicing-NMD event	
<u>SR proteins and splicing activators</u>	uc.28	<i>SFRS11</i>	Splicing factor p54	Stop codon exon	
	uc.50	<i>SFRS7</i>	Splicing factor 9G8	Stop codon exon	
	uc.138	<i>SFSR10</i>	Tra2- β (Stoilov et al. 2004)	Stop codon exon	
	uc.189	<i>SFRS3</i>	SRp20 (Jumaa and Nielsen 1997)	Stop codon exon	
	uc.208, 209	<i>TRA2A</i>	Tra2- α	Stop codon exon	
	uc.418	<i>SFRS1</i>	SF2/ASF, SRp30a	3' UTR intron activation	
	uc.456	<i>SFRS6</i>	SRp55	Stop codon exon	
	uc.455	<i>RNPC2</i>	CAPER α	Stop codon exon	
	uc.313	<i>TIAL1</i>	TIAR (Le Guiner et al. 2001)	Stop codon exon	
	<u>hnRNP proteins</u>	uc.33	<i>PTBP2</i>	Neural hnRNP I, nPTB (Rahman et al. 2004)	Exon-skipping frameshift
		uc.144	<i>HNRPDL</i>	hnRNP D-like	3' UTR exon activation
uc.186		<i>HNRPH1</i>	hnRNP H1	Exon-skipping frameshift	
uc.263		<i>HNRPK</i>	hnRNP K	Exon-skipping frameshift	
uc.443		<i>HNRPM</i>	hnRNP M	Exon-skipping frameshift	
<u>Other</u>		uc.151	<i>ZFR</i>	Zinc-finger RNA-binding protein	Exon-skipping frameshift

Ni et al.

repressor gene *PTBP2* (nPTB) (Bejerano et al. 2004; Rahman et al. 2004), which contains a 34-nt exon whose skipping would be predicted to cause a frame shift and NMD (Fig. 1B; Table 2). In another example, the ultraconserved element uc.418 in *SFRS1*, encoding the splicing activator ASF/SF2, overlaps an alternative intron in the 3' UTR whose removal causes NMD by placing an EJC downstream from the normal stop codon (Table 2; Wang et al. 1996; Sureau et al. 2001).

A striking aspect of the findings in Table 2 is that for splicing factors known mostly to be activators, such as SR proteins, the splicing event that triggers NMD within its pre-mRNA is a splicing activation event, whereas for factors known mostly to be repressors the splicing event that triggers NMD is a splicing-repression event. This polarity is identical to that expected for homeostatic auto- or cross-regulatory maintenance of proper splicing factor levels in the cell. Thus, in general, NMD is triggered for splicing activator mRNAs by activating a splicing event (exon inclusion or intron activation), and for splicing repressor mRNAs by repressing a splicing event (exon skipping), in order to bring levels of splicing factor back toward a set point (Fig. 4; see Discussion). We suggest this polarity is the key to maintaining an appropriate balance between positively acting and negatively acting general splicing factors to ensure proper levels of basal exon inclusion.

Retrotransposons, stop codon exons, and ultraconserved elements associated with AS-NMD

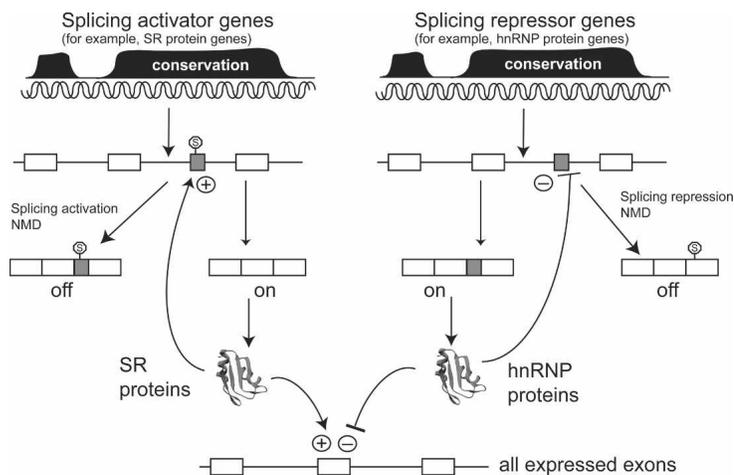
Retrotransposons that land in introns can become "exonized" or spliced into mRNA of the gene into which they have inserted, often introducing stop codons into mRNA (Lev-Maor et al. 2003; Mendell et al. 2004; Bejerano et al. 2006). Recent transposition events will appear nonconserved; however, ancient retrotransposons that became exonized and subject to purifying selection early in the vertebrate lineage would be conserved in extant

vertebrates and could achieve ultraconserved status despite their humble beginnings (Bejerano et al. 2004, 2006). The clearest example is the LF-SINE still active in the Coelocanth genome, which gave rise to two distinct ultraconserved elements, one of which overlaps an ultraconserved coding exon of the *PCBP2* gene (Bejerano et al. 2006). To investigate whether retrotransposon exonization generally plays a role in creating stop codon exons for exon-inclusion NMD regulation during evolution, we used the UCSC Genome Browser repeating elements track (Jurka 2000; Karolchik et al. 2003) and found that four of our 66 conserved stop codon exons clearly overlap members of a SINE family that predates our last common ancestor with the dog (MIR/MIRb family) (Supplementary Table S2; Smit 1999). Three others appear derived from long interspersed repetitive elements (LINEs) and long terminal repeat elements (LTR). None of the retrotransposon-derived stop codon exons overlap with the ultraconserved class. Two detected by arrays (*NKIRAS1* and *FARSLB*) are derived from more recent SINE element insertions found only in the mouse (Table 1; Supplementary Table S2). We conclude that retrotransposons contribute to, but do not entirely explain the origin of stop codon exons in the vertebrate genome (Supplementary Table S2).

Discussion

In this study we have used splicing-sensitive microarrays to identify a class of alternative splicing events that actively produce NMD-sensitive isoforms. A striking class of highly conserved stop codon exons comprises a large fraction of the events we found (Table 1; Fig. 3). Using the features of these experimentally defined exons, we searched genomic data, finding and then validating more examples (Fig. 3; Supplementary Table S2). The abundance of alternative splicing events that give rise to NMD isoforms (Lewis et al. 2003; Baek and Green 2005) prompted the suggestion that coupled alternative splic-

Figure 4. Model for homeostatic auto- or cross-regulatory maintenance of proper splicing factor levels in the cell. Expression of splicing activator genes (such as SR protein genes) that carry stop codon exons are regulated by AS-NMD. When the stop codon exon is skipped, functional activator protein mRNA is on. Too many splicing activator proteins can turn off their own expression by activating inclusion of their stop codon exon, triggering NMD. In addition, splicing activator protein level activates splicing of their many target exons in the genome to counteract the effect of the negative (hnRNP protein) splicing-repressor factors. Expression of splicing repressors (such as hnRNP protein) that carry a skipped coding exon can be regulated by AS-NMD. When the coding exon is included, functional hnRNP protein mRNA is on. Too many splicing repressor proteins can turn off their own expression by repressing the inclusion of the coding exon (not a multiple of three) that creates a frameshift and triggers NMD. In addition, splicing repressor protein levels also repress splicing at incorrect splice sites at many target exons in the genome and counteract the effects of the positive splicing factors (see text).



ing and NMD might regulate many genes (Lewis et al. 2003), although a recent study concluded that such regulation was not likely to be widespread, and no clear examples of such regulation were reported (Pan et al. 2006). The conserved stop codon exons identified in our experiments represent the most obvious examples of such regulation. Because of their high evolutionary conservation in the absence of protein coding potential (Fig. 2; Supplementary Table S2), and their accumulation upon a block to NMD (Figs. 2, 3), we deduce that this class of actively included exons contributes to the negative regulation of the genes that carry them. Although enriched in splicing factor genes (see below) conserved stop codon exons are found in a variety of functional gene classes (Supplementary Table S2), indicating the general utility of such a regulatory arrangement.

Splicing regulation of stop codon exons

Regulation of alternative splicing of each stop codon exon is likely to be dependent on the nature of the gene that carries it. One conserved stop codon exon we found in the SSAT (*SAT1*) gene (Fig. 3; Supplementary Table S2) was recently shown to contribute to control of polyamine levels (Hyvonen et al. 2006). This exon plus its adjacent intron sequences can be placed in a heterologous gene and its inclusion responds appropriately to perturbation of polyamine pools by inhibitors of polyamine metabolism (J. Ni and M. Ares, unpubl.), indicating that the conserved region containing the exon carries polyamine responsive elements. In addition, both SRp20 (*SFRS3*) and Tra2- β (*SFRS10*), when overexpressed, activate the inclusion of stop codon exons in their own pre-mRNAs (Jumaa and Nielsen 1997; Stoilov et al. 2004), and we found that the resulting splice forms are subjected to NMD (Fig. 3). These earlier studies often considered a possible function for the truncated protein produced by the PTC-containing isoform. Because NMD requires translation, such truncated proteins may be produced in small amounts even as NMD is actively destabilizing their mRNAs. Although regulatory functions for these truncated proteins are not excluded by our observations, destabilization of their mRNAs by NMD still constitutes a major regulatory influence.

Based on these examples we propose that for most of the 66 conserved stop codon exons we have identified in the mouse, there exists regulation of alternative splicing that controls how much functional mRNA is produced through NMD. This idea is supported by the observation that intronic sequences near these stop codon exons are also conserved (Supplementary Table S4), and that such conservation is a hallmark of alternatively spliced exons (Sorek and Ast 2003; Sugnet et al. 2004, 2006; Yeo et al. 2005).

Regions that control AS-NMD in splicing factor genes are ultraconserved

Ultraconserved elements are genomic regions that are longer than 200 bp and identical in sequence in the hu-

man, mouse, and rat genomes (Bejerano et al. 2004). In our data, eight stop codon exon-containing genes (*RNPC2*, *TRA2A*, *SFRS3*, *SFRS6*, *SFRS7*, *SFRS10*, *TIAL1*, and *2900045N06Rik*) overlap nine ultraconserved elements, and are statistically overrepresented in the 65 stop codon exon-containing genes relative to all genes. We also showed that more than half (15 of 29) of exonic ultraconserved elements found in RNA-splicing and RNA-binding protein genes appear to be associated with AS-NMD regulation (Table 2). Besides stop codon exon-inclusion NMD, many (eight of 15) of these ultraconserved exons are subject to other modes of AS-NMD regulation, such as exon-skipping NMD (Table 2; Fig. 4). In addition, even those stop codon exons whose conservation is extremely high but falls just short of ultraconserved may have exceedingly important or constrained functions. Together, these observations argue that the unusual evolutionary conservation of the special regions where alternative splicing leads to NMD in certain genes such as those encoding splicing factors could be due to a strong functional demand for combined AS-NMD auto-regulation and cross-regulation.

A model for homeostasis of splicing factor gene expression to maintain proper basal levels of exon recognition

In general, SR proteins are considered splicing activators that promote exon recognition through exonic splicing enhancers (ESEs), whereas hnRNP proteins act antagonistically to SR proteins, providing a regulatory counterweight through exonic or intronic splicing silencers (ESSs or ISSs) (Black 2003; Matlin et al. 2005). The ESE and ESS/ISS sequences that bind splicing factors are remarkably short and degenerate, and present in bewildering combinations, even in ordinary exons (Fairbrother et al. 2002; Black 2003; Wang et al. 2004; Matlin et al. 2005). Although this generality has its exceptions, the expression of two sets of opposing splicing regulators with potential to influence the recognition of every one of ~136,000 exons in the vertebrate cell must reflect a grand compromise. Such a compromise is also the ground state on which alternative splicing regulation must operate, since a regulated exon must be neither too strongly repressed nor too strongly activated in this ground state; otherwise, it may not be subject to the regulatory influence of specialized splicing factors. It therefore follows that the levels of the two sets of opposing splicing regulators must themselves be stably, as well as jointly, regulated.

We found that the polarity (activation or repression) of the splicing event that triggers NMD in the opposing groups of splicing factor genes has a strong positive correlation with the polarity of each group's action in splicing (Table 2). This creates a general negative feedback on the mRNA levels of the splicing factors in accordance with their activity and contribution to the basal state (Fig. 4). If positive regulators are in general too high, they act to lower the level of splicing activator mRNA by exon-inclusion NMD, or intron-activation NMD (Table

2; Fig. 4; Le Guiner et al. 2001; Sureau et al. 2001). As the level of negatively acting splicing factors becomes too high, they act to lower the levels of splicing repressor mRNAs by exon-skipping NMD (Table 2; Fig. 4; Wollerton et al. 2004). We envision this stable balance in the activities of splicing activator and repressor proteins to be a self-perpetuating, stable state of gene expression, in essence, an RNA-based epigenetic state. Although there are many other mechanisms by which individual splicing factor activities are regulated (e.g., see Huang and Steitz 2005), we propose that the general demand for regulating a stable balance between antagonistic splicing factor activities at the mRNA level explains the ultra-conserved nature of the genomic sequences that host AS-NMD.

The idea that autogenous control by RNA-binding proteins can maintain a stably inherited state of gene expression has precedence in studies of the *Drosophila* splicing regulator Sxl. Once activated early in development, Sxl maintains an epigenetic female state by activating its own splicing in a positive feedback loop (Bell et al. 1991). In the case of opposing splicing regulators in vertebrates, the feedback we illustrate in Figure 4 is primarily negative, but it seems likely there will be cross-regulatory influences that add layers to this simple model, and that these will be focused on the alternative splicing events that trigger NMD. For example, the repression of the stop codon exon of a splicing activator mRNA could occur if splicing repressor proteins become too high, and this would boost the levels of activators to counteract the increased level of repressors. Conversely, increasing levels of splicing activators could increase recognition of the key regulatory exon in splicing repressor mRNAs that would lead to increases in splicing repressor mRNA. There are many examples of alternative splicing regulation of splicing factor mRNAs by other splicing factors. For example, SRp30c (*SFRS9*) controls alternative splicing of hnRNP A1 (Simard and Chabot 2002), ASF/SF2 (*SFRS1*) represses the SRp20 (*SFRS3*) stop codon exon and antagonizes SRp20 autoregulation (Jumaa and Nielsen 1997). It will be important to understand the regulatory influences of other splicing factors on the regions where AS-NMD is taking place in order to understand the influences that maintain and regulate the activities of the many splicing regulators.

Materials and methods

Cell culture, transfection, and drug treatment

Dr. Douglas L. Black (University of California at Los Angeles, Los Angeles, CA) generously provided mouse N2A cells. The cells were maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% fetal bovine serum (FBS). For transfection, siRNA or minigenes were transfected into the cells using Lipofectamine 2000 reagent (Invitrogen) according to the manufacturer's instructions. Sequences of Upf1 and negative control siRNA were 5'-GAUGCAGUCCGUCCAUCdTdT-3' and 5'-UAGUUCGACUAUCCUGCCGdTdT-3', respectively. For emetine treatment, cells were exposed to 100 µg/mL emetine dihydrochloride hydrate (Fluka) 10 h before harvesting.

Western blotting and RT-PCR

Proteins from extracts of treated cells were separated in SDS-polyacrylamide gels, transferred to nitrocellulose, and probed with polyclonal goat anti-Rent1 (Upf1) antibody (Bethyl) or monoclonal anti- α -Tubulin as loading control. Bound antibodies were detected with ECL+plus Western Blotting Detection System (Amersham) as instructed by the manufacturer.

Total RNA was extracted using Trizol reagent (Invitrogen). cDNA was generated from 1 µg of total RNA using SuperScript II Reverse Transcriptase (Invitrogen) using oligo-dT primer, according to the manufacturer's instructions. For PCR, ~50 ng of cDNA was used as a template with primer pairs designed to measure splicing of the target exons (available on request). Reactions used Taq polymerase (Promega) and were run for 25–35 cycles at annealing temperatures appropriate for the primer pairs used. PCR products were checked using agarose gels stained with ethidium bromide, and were quantitated on an Agilent Bioanalyzer running the DNA 1000 chip.

Mouse splicing microarray

The Affymetrix mouse splicing microarray is described in Sugnet et al. (2006). Because EST/mRNA sequences used for array design were not filtered by frequency, many rare NMD-sensitive isoforms can be measured by this array. Three individual total RNA samples from Upf1 siRNA-transfected cells and control siRNA-transfected cells, and two individual total RNA samples each from emetine-treated cells and control cells were reverse transcribed, fragmented, end-labeled with biotin, and hybridized to arrays as described (Sugnet et al. 2006). Arrays were then stained and scanned, normalized, and analyzed as described (Sugnet et al. 2006). We analyzed the resultant microarray data using the method of Sugnet et al. (2006), reanalyzing the data three times, each time grouping the samples in one of three ways: Emetine versus Control, Upf1 siRNA versus Control siRNA, and Upf1 siRNA plus Emetine versus All controls. To simplify the identification of splicing isoforms that might be NMD substrates, we focused our analysis on the cassette exon mode of alternative splicing. Microarray data is deposited in the Gene Expression Omnibus (GEO) under accession number GSE6611 at <http://www.ncbi.nlm.nih.gov/geo>.

Calculation of conservation score

The multiple alignment and conservation score of each alternatively spliced exon comes from data in the UCSC Genome Browser conservation track (Karolchik et al. 2003). This track is made using the phastCons program (Siepel et al. 2005), which is based on a phylo-HMM, a type of probabilistic model that describes both the process of DNA substitution at each site in a genome and the way this process changes from one site to the next. The score is a probability that the alignment in question was derived by evolutionary conservation.

Bioinformatic identification of stop codon exon

Genome-wide identification of alternatively spliced exons relied on mouse ExonWalk data (Sugnet 2005). The ExonWalk method (Karolchik et al. 2003; Sugnet 2005) merges cDNA evidence together to predict full-length isoforms, including alternative transcripts. To predict transcripts that are biologically functional rather than the result of technical or biological noise, ExonWalk requires that every intron and exon either (1) be present in cDNA libraries of another organism (and also present in mouse), (2) have three separate cDNA GenBank (Benson et al.

2005) entries supporting it, or (3) be evolving like a coding exon as determined by Exoniphy (Siepel and Haussler 2004). Once the transcripts are predicted, an ORF finder (BESTORF from Softberry, <http://www.softberry.com>) is used to find the best ORF. To adapt this data for the current study, transcripts that are targets for NMD were not filtered out. We looked in both ExonWalk database and alternative exons used for the splicing array design for conserved three-frame stop exons. Exons that have >0.8 conservation score among human, mouse, rat, dog, and chicken (when available) were scanned in all three reading frames for stop codons by a custom program. The initial set of conserved three-frame stop exons was 669 from ExonWalk and 114 from the splicing array (where there was detectable expression in any mouse tissue) (Sugnet et al. 2006). There are 70 exons in common between these sets, because the filtering of ExonWalk is more stringent than that for the array, and because not all exons that pass the ExonWalk filters are detectable on the array. These 783 exons were examined by manual curation using the UCSC Genome Browser on the mm5 assembly (Karolchik et al. 2003). Exons in the 5' UTR or 3' UTR or that place a PTC <50 nt upstream of the normal stop codon were filtered out. FuncAssociate (Berriz et al. 2003) was used to identify over-represented classes of genes based on annotations in the gene ontology (GO) database (Ashburner et al. 2000). Fisher's exact test was used to determine statistical significance of the representation of ultraconserved elements in stop codon exons.

Acknowledgments

We thank Doug Black and Paul Boutz for N2A cells and advice on siRNA transfection. Thanks also to Gill Bejerano, Doug Black, Benoit Chabot, and Xiang-Dong Fu for comments on the manuscript. Thanks also to David Kulp for insight and encouragement. This project was supported by funds from NIH grants R01 GM040478 (supporting J.Z.N. and L.G.) and R24 GM 070857 (supporting J.P.D. and L.S.), and an HHMI "Professor" Award to M.A.

References

- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. 2000. Gene ontology: Tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.* **25**: 25–29.
- Baek, D. and Green, P. 2005. Sequence conservation, relative isoform frequencies, and nonsense-mediated decay in evolutionarily conserved alternative splicing. *Proc. Natl. Acad. Sci.* **102**: 12813–12818.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. 2004. Ultraconserved elements in the human genome. *Science* **304**: 1321–1325.
- Bejerano, G., Lowe, C.B., Ahituv, N., King, B., Siepel, A., Salama, S.R., Rubin, E.M., Kent, W.J., and Haussler, D. 2006. A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**: 87–90.
- Bell, L.R., Horabin, J.I., Schedl, P., and Cline, T.W. 1991. Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in *Drosophila*. *Cell* **65**: 229–239.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. 2005. GenBank. *Nucleic Acids Res.* **33** (Data-base issue): D34–D38.
- Berriz, G.F., King, O.D., Bryant, B., Sander, C., and Roth, F.P. 2003. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19**: 2502–2504.
- Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**: 291–336.
- Chiu, S.Y., Lejeune, F., Ranganathan, A.C., and Maquat, L.E. 2004. The pioneer translation initiation complex is functionally distinct from but structurally overlaps with the steady-state translation initiation complex. *Genes & Dev.* **18**: 745–754.
- Clark, T.A., Sugnet, C.W., and Ares Jr., M. 2002. Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**: 907–910.
- Conti, E. and Izaurralde, E. 2005. Nonsense-mediated mRNA decay: Molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* **17**: 316–325.
- Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**: 1007–1013.
- Huang, Y. and Steitz, J.A. 2005. SRprises along a messenger's journey. *Mol. Cell* **17**: 613–615.
- Hyvonen, M.T., Uimari, A., Keinänen, T.A., Heikkinen, S., Pellinen, R., Wahlfors, T., Korhonen, A., Narvanen, A., Wahlfors, J., Alhonen, L., et al. 2006. Polyamine-regulated unproductive splicing and translation of spermidine/spermine N1-acetyltransferase. *RNA* **12**: 1569–1582.
- Ishigaki, Y., Li, X., Serin, G., and Maquat, L.E. 2001. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. *Cell* **106**: 607–617.
- 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.
- Jurka, J. 2000. Repbase update: A database and an electronic journal of repetitive elements. *Trends Genet.* **16**: 418–420.
- Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. 2003. The UCSC Genome Browser Database. *Nucleic Acids Res.* **31**: 51–54.
- Kashima, I., Yamashita, A., Izumi, N., Kataoka, N., Morishita, R., Hoshino, S., Ohno, M., Dreyfuss, G., and Ohno, S. 2006. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. *Genes & Dev.* **20**: 355–367.
- Kim, V.N., Kataoka, N., and Dreyfuss, G. 2001. Role of the nonsense-mediated decay factor hUpf3 in the splicing-dependent exon-exon junction complex. *Science* **293**: 1832–1836.
- Kim, Y.K., Furic, L., Degroiseurs, L., and Maquat, L.E. 2005. Mammalian Staufen 1 recruits Upf1 to specific mRNA 3' UTRs so as to elicit mRNA decay. *Cell* **120**: 195–208.
- Le Guiner, C., Lejeune, F., Galiana, D., Kister, L., Breathnach, R., Stevenin, J., and Del Gatto-Konczak, F. 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.
- Le Hir, H., Moore, M.J., and Maquat, L.E. 2000. Pre-mRNA splicing alters mRNP composition: Evidence for stable association of proteins at exon-exon junctions. *Genes & Dev.* **14**: 1098–1108.
- Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. 2001. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20**: 4987–4997.
- Lejeune, F. and Maquat, L.E. 2005. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in

Ni et al.

- mammalian cells. *Curr. Opin. Cell Biol.* **17**: 309–315.
- Lejeune, F., Ranganathan, A.C., and Maquat, L.E. 2004. eIF4G is required for the pioneer round of translation in mammalian cells. *Nat. Struct. Mol. Biol.* **11**: 992–1000.
- Lev-Maor, G., Sorek, R., Shomron, N., and Ast, G. 2003. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science* **300**: 1288–1291.
- Lewis, B.P., Green, R.E., and Brenner, S.E. 2003. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci.* **100**: 189–192.
- Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. 2000. Human Upf proteins target an mRNA for nonsense-mediated decay when bound downstream of a termination codon. *Cell* **103**: 1121–1131.
- Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. 2001. Communication of the position of exon–exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293**: 1836–1839.
- Maquat, L.E. 2005. Nonsense-mediated mRNA decay in mammals. *J. Cell Sci.* **118**: 1773–1776.
- Matlin, A.J., Clark, F., and Smith, C.W. 2005. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**: 386–398.
- Mendell, J.T., ap Rhys, C.M., and Dietz, H.C. 2002. Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. *Science* **298**: 419–422.
- Mendell, J.T., Sharifi, N.A., Meyers, J.L., Martinez-Murillo, F., and Dietz, H.C. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat. Genet.* **36**: 1073–1078.
- Modrek, B. and Lee, C.J. 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat. Genet.* **34**: 177–180.
- Nagy, E. and Maquat, L.E. 1998. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem. Sci.* **23**: 198–199.
- Noensie, E.N. and Dietz, H.C. 2001. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. *Nat. Biotechnol.* **19**: 434–439.
- Pan, Q., Saltzman, A.L., Kim, Y.K., Misquitta, C., Shai, O., Maquat, L.E., Frey, B.J., and Blencowe, B.J. 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.
- Rahman, L., Bliskovski, V., Reinhold, W., and Zajac-Kaye, M. 2002. Alternative splicing of brain-specific PTB defines a tissue-specific isoform pattern that predicts distinct functional roles. *Genomics* **80**: 245–249.
- Rahman, L., Bliskovski, V., Kaye, F.J., and Zajac-Kaye, M. 2004. Evolutionary conservation of a 2-kb intronic sequence flanking a tissue-specific alternative exon in the PTBP2 gene. *Genomics* **83**: 76–84.
- Schell, T., Kocher, T., Wilm, M., Seraphin, B., Kulozik, A.E., and Hentze, M.W. 2003. Complexes between the nonsense-mediated mRNA decay pathway factor human upf1 (up-frame-shift protein 1) and essential nonsense-mediated mRNA decay factors in HeLa cells. *Biochem. J.* **373**: 775–783.
- Siepel, A. and Haussler, D. 2004. Computational identification of evolutionarily conserved exons. In *Proceedings of the eighth annual international conference on research in computational biology (RECOMB)* (eds. P.E. Bourne and D. Gusfield), pp. 177–186. ACM Press, New York.
- Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**: 1034–1050.
- Simard, M.J. and Chabot, B. 2002. SRp30c is a repressor of 3' splice site utilization. *Mol. Cell Biol.* **22**: 4001–4010.
- Smit, A.F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**: 657–663.
- Sorek, R. and Ast, G. 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13**: 1631–1637.
- Sorek, R., Shamir, R., and Ast, G. 2004. How prevalent is functional alternative splicing in the human genome? *Trends Genet.* **20**: 68–71.
- Stoilov, P., Daoud, R., Nayler, O., and Stamm, S. 2004. Human tra2- β 1 autoregulates its protein concentration by influencing alternative splicing of its pre-mRNA. *Hum. Mol. Genet.* **13**: 509–524.
- Sugnet, C.W. 2005. "Discovery and detection of alternative splicing," p. 132. Computer Science Department, University of California at Santa Cruz, Santa Cruz, CA.
- Sugnet, C.W., Kent, W.J., Ares Jr., M., and Haussler, D. 2004. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.* **2004**: 66–77.
- Sugnet, C.W., Srinivasan, K., Clark, T.A., O'Brien, G., Cline, M.S., Wang, H., Williams, A., Kulp, D., Blume, J.E., Haussler, D., et al. 2006. Unusual intron conservation near tissue-regulated exons found by splicing microarrays. *PLoS Comput. Biol.* **2**: e4.
- Sureau, A., Gattoni, R., Dooghe, Y., Stevenin, J., and Soret, J. 2001. SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J.* **20**: 1785–1796.
- Ule, J., Ule, A., Spencer, J., Williams, A., Hu, J.S., Cline, M., Wang, H., Clark, T., Fraser, C., Ruggiu, M., et al. 2005. Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.* **37**: 844–852.
- Wang, J., Takagaki, Y., and Manley, J.L. 1996. Targeted disruption of an essential vertebrate gene: ASF/SF2 is required for cell viability. *Genes & Dev.* **10**: 2588–2599.
- Wang, W., Czaplinski, K., Rao, Y., and Peltz, S.W. 2001. The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. *EMBO J.* **20**: 880–890.
- Wang, H., Hubbell, E., Hu, J.S., Mei, G., Cline, M., Lu, G., Clark, T., Siani-Rose, M.A., Ares, M., Kulp, D.C., et al. 2003. Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics* **19** (Suppl. 1): i315–i322.
- Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* **119**: 831–845.
- Wollerton, M.C., Gooding, C., Wagner, E.J., Garcia-Blanco, M.A., and Smith, C.W. 2004. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13**: 91–100.
- Yeo, G.W., Van Nostrand, E., Holste, D., Poggio, T., and Burge, C.B. 2005. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc. Natl. Acad. Sci.* **102**: 2850–2855.