BIOINFORMATICS

Vol. 19 Suppl. 1 2003, pages i118–i121
DOI: 10.1093/bioinformatics/btg1015

# Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes

Richard E. Green[1, 2], Benjamin P. Lewis[1, 3], R. Tyler Hillman[1], Marco Blanchette[2], Liana F. Lareau[2], Aaron T. Garnett[2], Donald C. Rio[2] and Steven E. Brenner[1, 2, 3,*]

[1]Departments of Plant and Microbial Biology, [2]Molecular and Cell Biology and [3]Biophysics Graduate Group, University of California, Berkeley, CA 94720, USA

Received on January 6, 2003; accepted on February 20, 2003

## ABSTRACT

We have recently shown that a third of reliably-inferred alternative mRNA isoforms are candidates for nonsense-mediated mRNA decay (NMD), an mRNA surveillance system (Lewis *et al.*, 2003, *Proc. Natl Acad. Sci. USA*, **100**, 189–192). Rather than being translated to yield protein, these transcripts are expected to be degraded and may be subject to regulated unproductive splicing and translation (RUST). Our initial experimental studies are consistent with these predictions and suggest an unappreciated role for NMD in several human diseases.
**Contact:** brenner@compbio.berkeley.edu

## INTRODUCTION

Alternative splicing is a process that enables a single genetic locus to generate multiple mRNA isoforms, and it affects over a third (Brett *et al.*, 2000; Kan *et al.*, 2002) of all human genes. Notably, functional explanations exist for only a handful of alternative isoforms. Using the human genome sequence, databases of expressed sequence tags (ESTs), and reliable mRNAs, we have discovered an apparent widespread coupling of alternative splicing with NMD.

### Nonsense-mediated mRNA decay

To date, all eukaryotes studied have the capacity to detect and degrade prematurely terminating transcripts. This NMD surveillance system is thought to rid the cell of potentially harmful truncated versions of proteins generated by mutations or by errors in mRNA processing. Biochemical, genetic and cell biological experiments have led to the following model for NMD (Fig. 1). During pre-mRNA processing, the spliceosome deposits exon junction complexes (EJCs) at sites of intron removal (Le Hir *et al.*, 2001). The EJCs serve the dual role of facilitating mRNA export and remembering gene structure. With the pioneering first round of translation, the ribosome displaces the EJCs in its path and subsequently disassociates from the mRNA at the site of termination (Ishigaki *et al.*, 2001). If the ribosome reaches a stop codon more than 50 nucleotides upstream of the final exon-exon junction, one or more EJCs will remain bound to the mRNA. Interactions between EJC proteins and release factors recruit a decapping enzyme that triggers rapid mRNA decay (Lykke-Andersen *et al.*, 2000). Thus, a general rule for specifying whether a transcript will be targeted by the NMD pathway has been stated as follows (Nagy and Maquat, 1998): if an intron is located more than 50 nucleotides downstream of the stop codon, then the termination codon is recognized as premature and the transcript will be down-regulated by NMD.

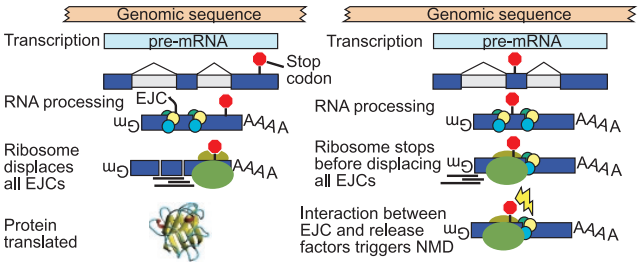


**Fig. 1.** Nonsense-mediated mRNA decay. Normal translation involves displacement of all exon junction complexes (EJCs). Any remaining EJCs will trigger NMD.

## RESULTS

### Alternative mRNAs of human genes are apparent targets of NMD

Just as the binding of transcription factors may turn on (or off) gene expression, the binding of splicing factors can turn on (or off) gene expression by splicing productive (or

*To whom correspondence should be adressed.

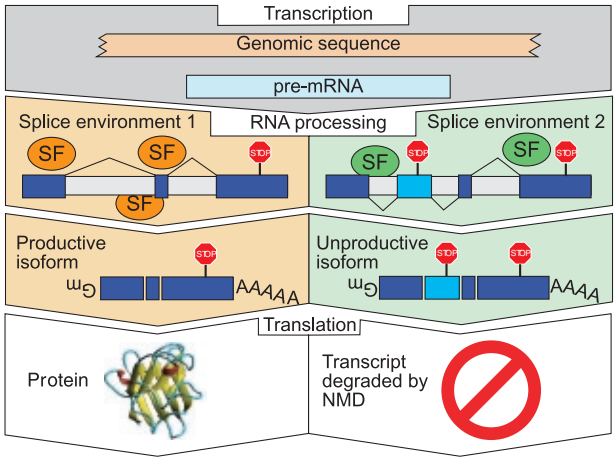


**Fig. 2.** Regulated unproductive splicing and translation (RUST). Protein expression is controlled by regulated variations in the splice environment, the presence and abundance of splicing factors (SF). This determines whether productive or prematurely terminating (unproductive) mRNAs are spliced.

unproductive) isoforms. We term this regulated unproductive splicing and translation (RUST) (Fig. 2) (Lewis *et al.*, 2003). This mode of regulation has been demonstrated for the splicing factor SC35, which auto-regulates its own expression through coupling alternative splicing and NMD (Sureau *et al.*, 2001). Additional cases of RUST have been observed for ribosomal proteins in *C.elegans* (Mitrovich and Anderson, 2000).

We previously applied the 50-nucleotide rule to alternatively-spliced transcripts evidenced within dbEST (Boguski *et al.*, 1993) to identify cases in which alternative splicing generates NMD-candidate isoforms (Fig. 3). By mapping human mRNA sequences from RefSeq (Pruitt and Maglott, 2001) to human genomic sequence (Lander *et al.*, 2001), we constructed sequences for whole loci, which we called RefSeq-coding genes (Fig. 3a). In this way, we determined the intron/exon structure of each RefSeq gene. We used the TAP (Kan *et al.*, 2001) and WU-BLASTN (Gish, 1996–2002) programs to perform high-stringency alignments between EST sequences and each RefSeq-coding gene. We identified alternative splice pairs and alternative isoforms, keeping track of the number of ESTs providing evidence for each alternative splice (Fig. 3b). We have higher confidence in the predicted alternative splices that are supported by alignments with multiple ESTs. Application of the 50-nucleotide NMD rule revealed that 35% of alternative isoforms are candidates for NMD (Fig. 3c). Additionally, 35% of alternatively-spliced RefSeq genes were found to generate at least one NMD-candidate isoform.

Though alternative splicing and NMD combine to make RUST a general means of gene regulation–any gene may be engineered or evolve to be susceptible to RUST by modifying splicing signals in the 3'UTR–the prevalence of RUST is still hypothetical. Preliminary phylogenetic sequence analyses hint that NMD might have evolved before alternative splicing. This allows the speculation that the predominant ancestral role of alternative splicing could possibly have been RUST, rather than the generation of protein diversity. Initial studies suggest that among the genes with NMD-candidate mRNA isoforms, certain functional classes are over-represented, including chaperone proteins.

## Nonsense-mediated decay and disease

The current understanding of NMD has impacted our understanding of numerous diseases. For example, the genotype-phenotype disconnect in many patients with muscular dystrophy (MD), a disease caused by mutations in the dystrophin locus, is only understandable in light of NMD (Gillard *et al.*, 1989; Roberts *et al.*, 1994). Mutations that change the reading frame and consequently introduce premature termination codons cause the severe form of MD (Duchenne), as the whole transcript is eliminated by NMD. On the other hand, mutations, even some large deletions, that do not introduce premature termination codons, result in the milder form of MD (Becker).

We have searched the SWISS-PROT database (O'Donovan *et al.*, 2002) for alternatively spliced human genes that generate NMD-candidate isoforms that might have been interpreted as truncated proteins. We further focused our search on disease related proteins in which a role for NMD may have been unappreciated. This screen has revealed several genes whose expression may be impacted by NMD (Hillman *et al.*, in preparation). For example, Calpain-10 was found to have four NMD-candidate isoforms described in SWISS-PROT. Mutations in Calpain-10 (Fig. 4), a ubiquitously expressed protease, have been strongly linked to susceptibility to Type-II diabetes in several populations (Horikawa *et al.*, 2000). Interestingly, one of these mutations lies in intronic sequence and may affect splicing. To determine if the isoforms of Calpain-10 are subjected to NMD in agreement with our hypothesis, we generated PCR primers that specifically amplify the alternative regions of Calpain-10 isoforms (Fig. 4).

We treated cells with the translational inhibitor cycloheximide, which has been shown to inhibit NMD, and performed RT-PCR on total cellular mRNA to determine the abundance of each isoform. This experiment revealed an increase in abundance of isoforms predicted to be NMD targets beyond that seen for other isoforms or the control gene, beta-actin. Other genes we are currently experimentally investigating include Presenilin1, which has been

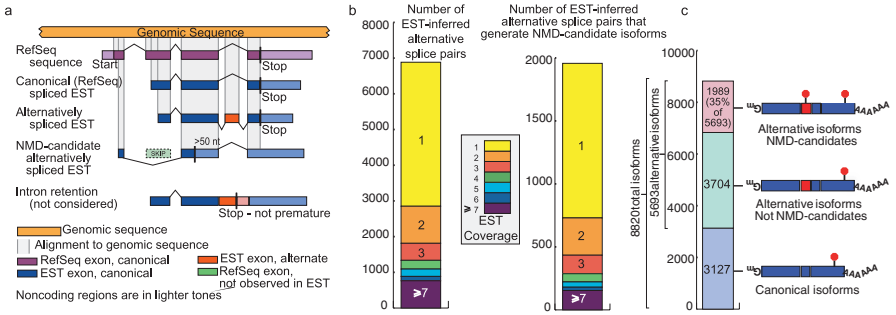


**Fig. 3.** Widespread coupling of alternative splicing with NMD (a) Alternative splicing was detected by aligning EST sequence to RefSeq-coding sequences. (b) The number of splice pairs demonstrating alternative splicing and the subset generating NMD-candidate alternative isoforms classified by degree of EST coverage. (c) 35% of alternative isoforms were NMD-candidates.

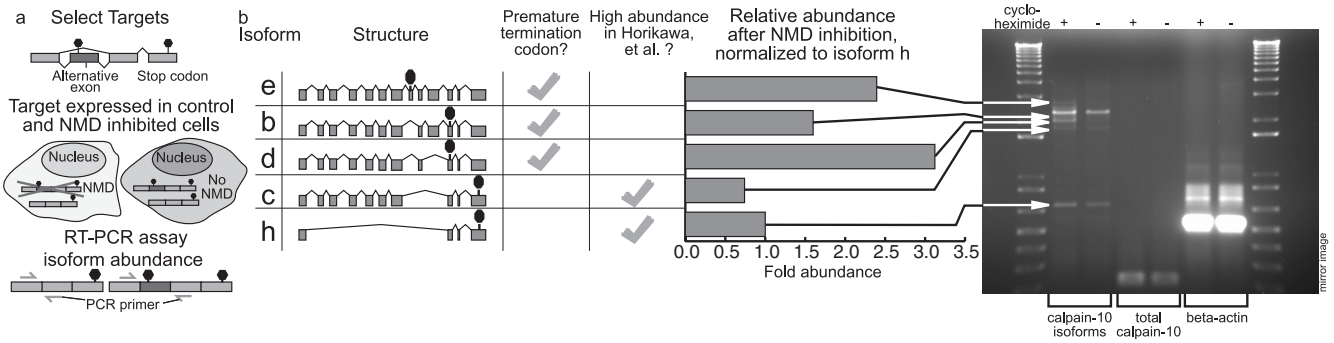


**Fig. 4.** Experimental validation of NMD-candidate mRNA isoforms of Calpain-10. (a) Isoform expression of selected alternatively spliced genes was assayed in control cells and NMD-inhibited cells using primers flanking the alternative regions. (b) In NMD-inhibited cells, the resolvable NMD-candidate isoforms of Calpain-10 increased in abundance relative to others isoforms.

linked to Alzheimer's disease susceptibility (Sherrington *et al.*, 1995). As with Calpain-10, at least one alternatively spliced isoform generated from this gene is a putative NMD-candidate (Sahara *et al.*, 1996). The human lymphocyte associated receptor of death (LARD) gene, associated with the appearance of lymphoid malignancies, is also being investigated with particular interest as the alternative splicing of the gene's pre-mRNA is closely linked with the proliferation of activated lymphocytes (Screaton *et al.*, 1997).

Recent experiments elucidating the mechanism of NMD have enabled us to investigate its role in regulating gene expression. It appears that NMD may function not only as a surveillance complex for the rare, wayward transcript but also as part of a pathway that controls expression of a large number of genes. The importance of NMD is underscored by its potential involvement in degrading many disease-linked gene transcripts. We hope that a more complete understanding of the interplay between alternative splicing and NMD will provide insights leading to effective treatments for some genetic diseases and a better understanding of gene regulation.

## ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health Grants K22 HG00056 and T32 HG0047, the Searle Scholars' Program (01-L-116) and an IBM Shared University Research grant. B.P.L. is supported by a Krell Research Fellowship.

## REFERENCES

Boguski,M.S., Lowe,T.M. and Tolstoshev,C.M. (1993) dbEST–database for expressed sequence tags. *Nature Genet.*, **4**, 332–333.

Brett,D., Hanke,J., Lehmann,G., Haase,S., Delbruck,S., Krueger,S. *et al.* (2000) EST comparison indicates 38% of human mRNAs contain possible alternative splice forms. *FEBS Lett.*, **474**, 83–86.

Gillard,E.F., Chamberlain,J.S., Murphy,E.G., Duff,C.L., Smith,B., Burghes,A.H. *et al.* (1989) Molecular and phenotypic analysis of patients with deletions within the deletion-rich region of the Duchenne muscular dystrophy (DMD) gene. *Am. J. Hum. Genet.*, **45**, 507–520.

Gish,W.R. (1996–2002) WU-BLAST (St Louis, Washington University).

Horikawa,Y., Oda,N., Cox,N.J., Li,X., Orho-Melander,M., Hara,M. *et al.* (2000) Genetic variation in the gene encoding calpain-10 is

associated with type 2 diabetes mellitus. *Nature Genet.*, **26**, 163–175.

Ishigaki,Y., Li,X.J., 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.

Kan,Z., Rouchka,E.C., Gish,W. and States,D.J. (2001) Gene structure prediction and alternative splicing analysis using genomically aligned ESTs. *Genome Res.*, **11**, 889–900.

Kan,Z., States,D. and Gish,W. (2002) Selecting for Functional Alternative Splices in ESTs. *Genome Res.*, **12**, 1837–1845.

Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

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.

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. USA*, **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 bounddownstream of a termination codon. *Cell*, **103**, 1121–1131.

Mitrovich,Q.M. and Anderson,P. (2000) Unproductively spliced ribosomal protein mRNAs are natural targets of mRNA surveillance in*C.elegans*. *Genes Dev.*, **14**, 2173–2184.

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.

O'Donovan,C., Martin,M.J., Gattiker,A., Gasteiger,E., Bairoch,A. and Apweiler,R. (2002) High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform.*, **3**, 275–284.

Pruitt,K.D. and Maglott,D.R. (2001) RefSeq and LocusLink: NCBI gene-centeredresources. *Nucleic Acids Res.*, **29**, 137–140.

Roberts,R.G., Gardner,R.J. and Bobrow,M. (1994) Searching for the 1 in 2 400 000: a review of dystrophin gene point mutations. *Hum. Mutat.*, **4**, 1–11.

Sahara,N., Yahagi,Y., Takagi,H., Kondo,T., Okochi,M., Usami,M. *et al.* (1996) Identification and characterization of presenilin I-467, I-463 and I-374. *FEBS Lett.*, **381**, 7–11.

Screaton,G.R., Xu,X.N., Olsen,A.L., Cowper,A.E., Tan,R., McMichael,A.J. *et al.* (1997) LARD: a new lymphoid-specific death domain containing receptor regulated by alternative pre-mRNA splicing. *Proc. Natl Acad. Sci. USA*, **94**, 4615–4619.

Sherrington,R., Rogaev,E.I., Liang,Y., Rogaeva,E.A., Levesque,G., Ikeda,M. *et al.* (1995) Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease. *Nature*, **375**, 754–760.

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.