Systematic Survey of Human Targets of Nonsense-Mediated mRNA Decay

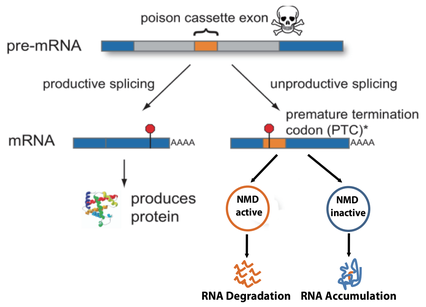
Gang Wei[1*], Courtney E. French[2*], Angela N. Brooks[2] and Steven E. Brenner[1,2,§]

[1] Department of Plant and Microbial Biology, [2] Department of Molecular and Cell Biology, University of California, Berkeley, CA, 94720-3102, USA.

(*These authors contributed equally to this work; §Corresponding author: brenner@compbio.berkeley.edu)

Get this poster as PDF

INTRODUCTION:

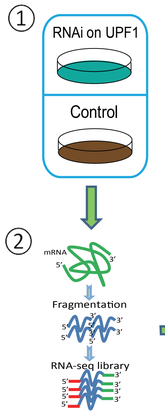


Alternative splicing plays a major role in the generation of proteomic diversity. However, mistakes in this process can introduce a premature termination codon (PTC) and result in non-functional proteins that are harmful to the cell. Such transcripts are usually degraded by nonsense-mediated mRNA decay (NMD). The coupling of alternative splicing and NMD has also been reported as an important regulatory mechanism for certain sets of genes [1,6,7]. Though many NMD targets have been identified in various species, we still lack a comprehensive view of the landscape of those transcripts degraded by NMD. Here, we characterize the transcripts normally degraded by NMD in human HeLa cells by inhibiting NMD through knockdown of core protein UPF1 and then performing RNA-seq analysis.

GOALS:

- How many genes produce isoforms that are targets for NMD in human cells?
- How highly transcribed are NMD targets before degradation?
- What is the functional role of NMD-related regulation?

APPROACH:



1. NMD inhibition through knockdown of UPF1.
2. Directional and paired-end RNA-seq library preparation.
3. High throughput sequencing using Illumina.
4. Map reads to genome with TopHat [3].
5. Transcript assembly and quantification with Cufflinks [4].
6. Early stop codon prediction using canonical 50nt rule of transcripts targeted by NMD in mammals [5].

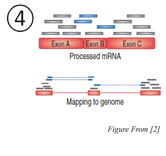


Figure From [2]

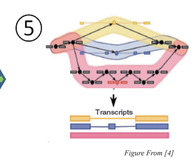


Figure From [4]

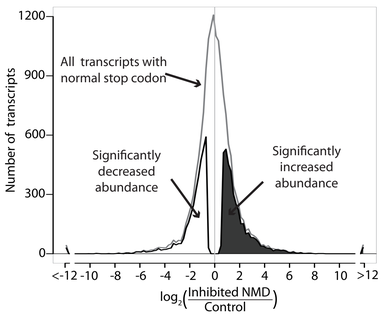
Premature termination codon (PTC): stop codon > 50 bases upstream of last exon-exon junction [5]

We define *NMD targets* as those transcripts that have a premature termination codon and are significantly up-regulated when NMD is inhibited (in two biological replicates).

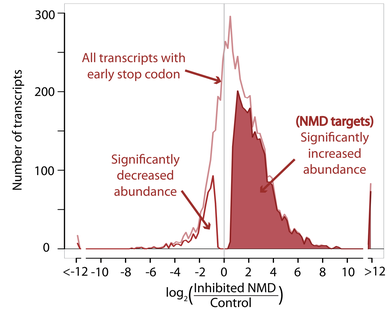
RESULTS:

Almost 2,500 transcripts identified as putative NMD targets

A No bias in direction of abundance changes for isoforms with normal stop codon



B Isoforms with early stop codon tend to be significantly increased when NMD is inhibited

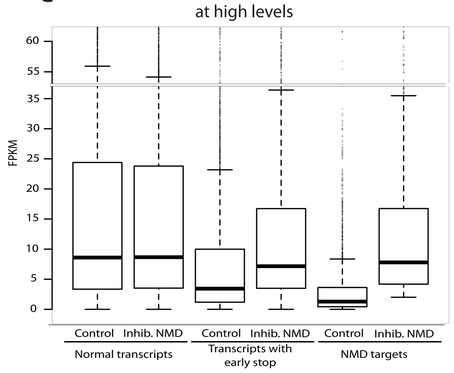


Altogether, 16,180 transcripts with normal stop codon and 5,809 transcripts with an early stop codon were detected in RNA-seq data. The abundance fold change distribution was log normal for productive transcripts (A), but transcripts with an early stop showed a strong bias toward increased abundance when NMD was inhibited (B).

2,443 transcripts with an early stop codon were significantly up-regulated and defined as putative NMD targets.

Transcripts degraded by NMD in normal cells can originally be expressed at as high a level as normal transcripts before degradation (C).

C Many isoforms degraded by NMD are expressed at high levels



Many genes encoding splicing factors were identified as having isoforms targeted by NMD

Splicing factor category	Expressed genes	Targeted by NMD	Genes with isoforms targeted by NMD
SR proteins	11	10	SRSF1, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF8, SRSF10, SRSF11
hnRNP	34	12	HNRNPH1, HNRNPH3, HNRNPA2B1, HNRNPK, HNRNPL, HNRPDL, SYNCRIP, PTBP1, PTBP2, PCBP2, RBM3, CIRBP
snRNP	39	10	SNRNP70, SNRNP48, SNRNP40, TXNL4A, SART1, PRPF3, PPIH, U2AF1, U2AF2, U2AF1L4
DEAD	15	5	DDX5, DDX46, DHX9, DHX15, INTS6
Sm	18	2	SNRPB, SNRPN
Other	114	35	ZNF207, LUC7L3, MOV10, CLASRP, RBM39, RBM5, ISY1, SMNDC1, SRRM1, SRRM2, FUS, DNAJC8, U2SURP, EIF2S2, GCFC1, TOP1MT, SFPQ, SRPK1, NCBP2, SREK1, C16orf80, ACIN1, THOC2, THOC4, TCERG1, CDK12, CRNKL1, PPIE, FUBP3, PRPF4B, TIA1, TIAL1, CLK1, TRA2A, TRA2B

Splicing factors have been described as producing isoforms targeted by NMD [6,7,8]. We found 17 previously reported (red) and many more (list of factors from [9]). RNA splicing genes were significantly enriched for NMD targets (p<4e-6; Goseq [10]).

Alternative splicing and NMD can contribute to changes in gene expression

Scenario A: Overall gene expression increased by accumulation of NMD targets

186 genes with isoforms targeted by NMD were significantly up-regulated without an increase in the abundance of productive isoforms.

Scenario B: A shift in splicing towards the NMD-targeted isoform as an effect of UPF1-knockdown

115 genes with isoforms targeted by NMD had little overall gene expression change because of a shift in splicing away from the productive isoform

Scenario C: Combination of decreased transcription and shift towards NMD-targeted isoform down-regulates gene expression.

75 genes with isoforms targeted by NMD were significantly down-regulated despite increased abundance of NMD target.

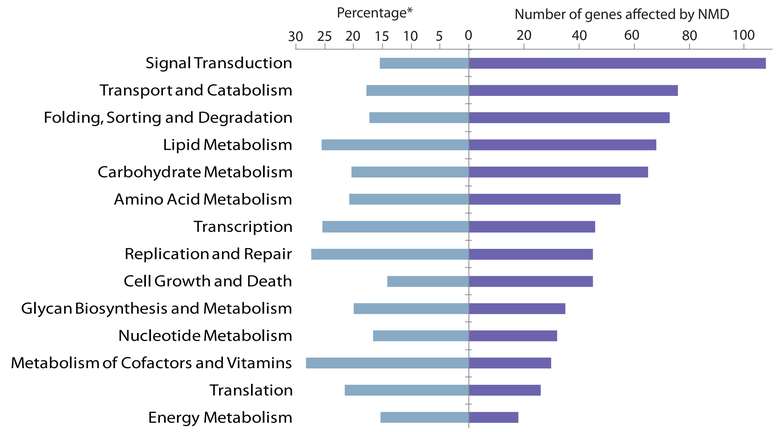
Productive transcript NMD target

Cassette exon alternative splicing events are the dominant method of producing NMD targets

Prevalence of different categories of alternative splicing events found to be significantly different between samples (juncBASE [11]).

Inner circle: all significant alternative splicing events
Outer circle: significant alternative splicing events associated with an NMD target
Blue box: constitutive exon
White box: alternatively spliced exon

Genes with isoforms targeted by NMD fall into a variety of KEGG functional categories

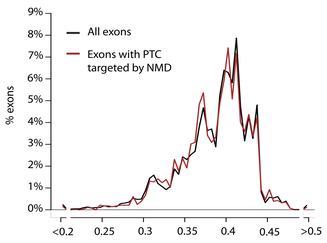


* percent of expressed genes in category that have NMD targeted isoforms

Ultraconserved elements (UCE) are over-represented in NMD targets, but overall conservation was not significantly different

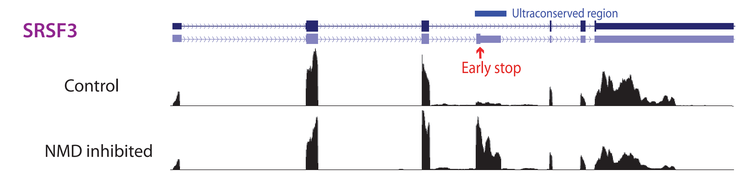
Category	Genes with UCE and isoforms degraded by NMD	
	Number	Gene Names
RNA processing	19	HNRNPH1, HNRNPK, PCBP2, SYNCRIP, PRPF38B, PRPF39, TRA2A, TRA2B, PTBP2, DDX5, DHX15, RBM39, TIAL1, ZFR, SRSF1, SRSF3, SRSF6, SRSF7,
Transcriptional regulation	5	CCAR1, MGA, NFAT5, TAF1, MED1
Other	8	STRN3, MRRF, OGT, CCDC45, FAM98A, unknown functions

32 of 82 genes that overlap an exonic ultraconserved element are NMD targets (significantly enriched by Fisher's exact test, p<7e-6). Ultraconserved regions are defined as >200 bp of 100% sequence identity between human, mouse, and rat [12].

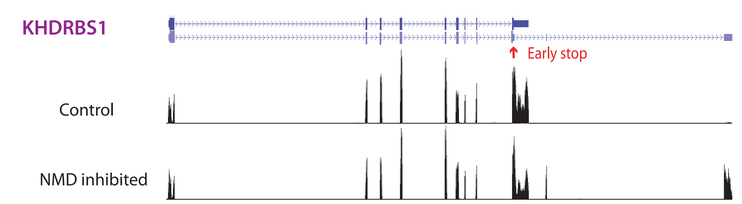


Distributions of the average conservation across an exon. The conservation score per base is the phastCons score from the 46 vertebrate alignment.

Previously inferred splicing events resulting in NMD targets were confirmed and new ones were discovered



SRSF3 (SRp20), a SR protein, has been identified to have NMD-targeted isoforms whose abundance was greatly increased when NMD was inhibited [6,7]. This transcript is caused by cassette exon inclusion.



KHDRBS1, a signal transduction associated protein, was found to have a novel NMD-targeted isoform caused by alternative last exons.

CONCLUSIONS:

1. Almost 2,500 transcripts transcribed from ~1,900 genes were discovered to be targeted by NMD.
2. Splicing regulators are significantly enriched for NMD targets, and genes from many other functional categories also produce NMD targets.
3. Exons of transcripts targeted by NMD are significantly enriched for ultraconserved elements although their overall conservation is similar to that of all exons genome-wide.
4. NMD-targeted isoforms can be generated by various splicing events, most of which are exon inclusion/exclusion and alternative donor/acceptor sites.
5. Coupling of alternative splicing and NMD appears to regulate the expression of hundreds of genes.

REFERENCES:

1. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated decay in humans. *PNAS*. 2003. 100:189-192
2. Trapnell, et al. *Nature Biotechnology*. 2009. 7:457
3. Trapnell, et al. *Bioinformatics*. 2009. 25:1105-11
4. Trapnell, et al. *Nature Biotechnology*. 2010. 28:511-5
5. Nagy, E. and L. E. Maquat. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends in Biochemical Science*. 1998. 23:198-9
6. Lareau, L. F., M. Inada, et al. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature*. 2007. 446:926-9
7. Ni, J. Z., L. Grate, et al. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes and Development*. 2007. 21:708-18
8. McGlincy, N. J., L. Y. Tan, et al. Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. *BMC Genomics*. 2010. 11:565
9. Barbosa-Morais, N. L., M. Carmo-Fonseca, et al. Systematic genome-wide annotation of spliceosomal proteins reveals differential gene family expansion. *Genome Research*. 2006. 16:66-77
10. Young, M. D., et al. *Genome Biology*. 11:R14
11. Brooks, A. N., L. Yang, et al. *Genome Research*. 2011. 21:193-202
12. Bejerano, G., M. Pheasant, et al. Ultraconserved elements in the human genome. *Science*. 2004. 304:1321-5

Acknowledgement:
We thank Adam Roberts and Lior Pachter of UC Berkeley for help with the optimization of Cufflinks.