Transcriptome analysis reveals extensive alternative splicing coupled with nonsense-mediated mRNA decay in human cells

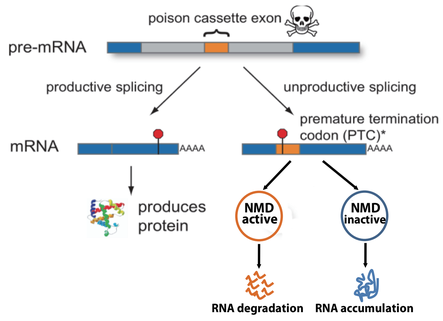
Courtney E. French[1*], Gang Wei[2*], Angela N. Brooks[1] and Steven E. Brenner[1,2]

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

brenner@compbio.berkeley.edu

*These authors contributed equally to this work

INTRODUCTION:

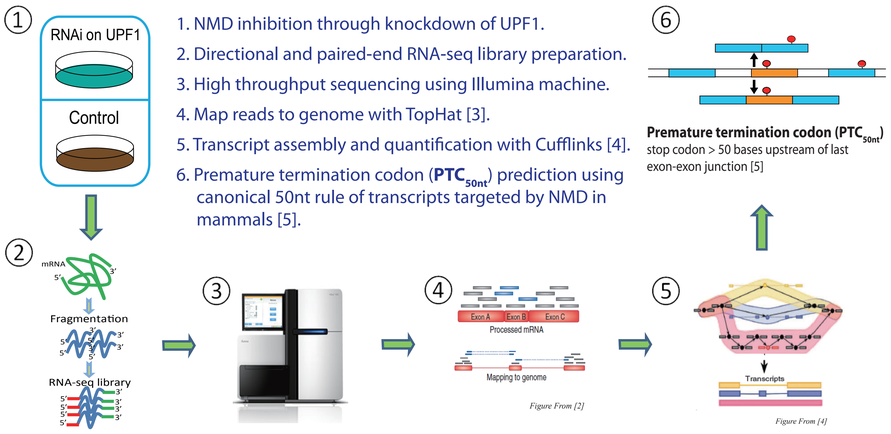


Alternative splicing plays a major role in the generation of proteomic diversity. However, mistakes in this process can introduce a premature termination codon (PTC) and result in non-functional proteins that are harmful to the cell. Such transcripts are usually degraded by nonsense-mediated mRNA decay (NMD). The coupling of alternative splicing and NMD has also been reported as an important regulatory mechanism for certain sets of genes [1]. 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 the essential NMD factor UPF1 and 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:



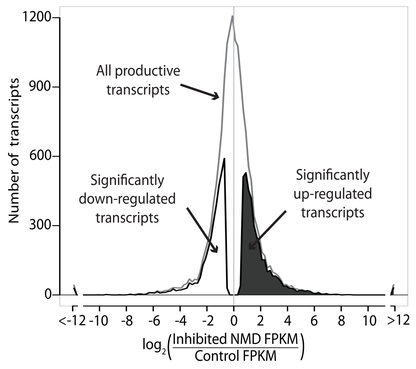
1. NMD inhibition through knockdown of UPF1.
2. Directional and paired-end RNA-seq library preparation.
3. High throughput sequencing using Illumina machine.
4. Map reads to genome with TopHat [3].
5. Transcript assembly and quantification with Cufflinks [4].
6. Premature termination codon (**PTC_{50nt}**)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

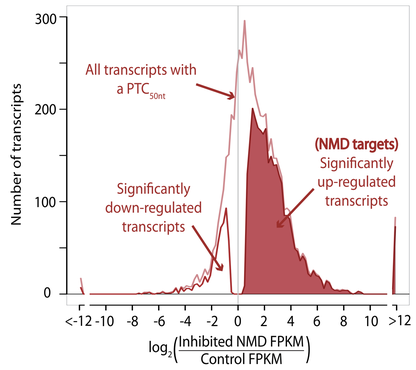
RESULTS:

Almost 2,500 transcripts identified as putative NMD targets

A No bias in direction of abundance changes for productive isoforms



B Isoforms with a PTC_{50nt} tend to increase in abundance

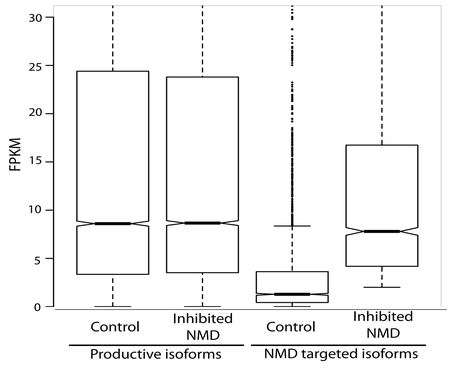


Altogether, 16,180 productive transcripts and 5,809 PTC_{50nt} transcripts were expressed in our data. The abundance fold change distribution was symmetric for productive transcripts (A), but PTC_{50nt} transcripts showed a strong bias toward increased abundance when NMD was inhibited (B).

2,443 PTC_{50nt} transcripts were significantly increased and defined as putative NMD targets.

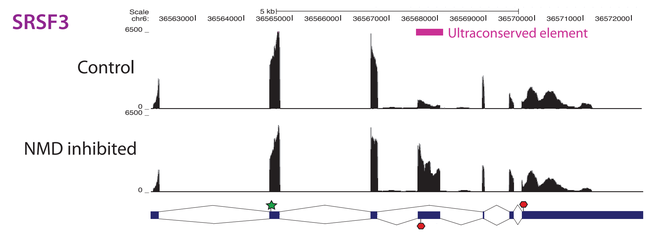
Transcripts degraded by NMD in normal cells can be expressed at as high a level as productive ones before degradation (C).

C Many isoforms degraded by NMD are expressed at high levels



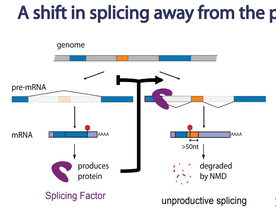
Previously inferred splicing events resulting in NMD targets were confirmed...

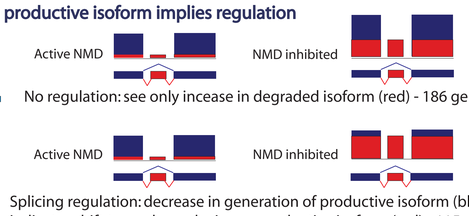
SRSF3



SRSF3 (SRp20), a SR protein, has been identified to have an NMD-targeted isoform whose abudance was greatly increased in UPF1-depleted cells [6,7]. This unproductive transcript is caused by cassette exon inclusion. Concurrent decrease of the productive isoform implies regulation.

A shift in splicing away from the productive isoform implies regulation



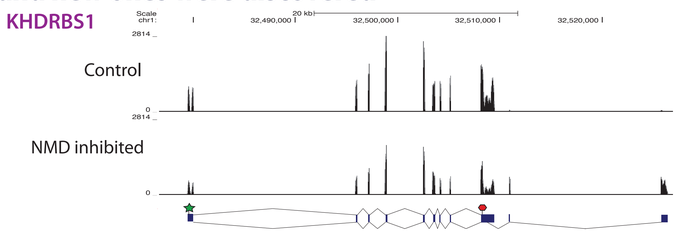


No regulation: see only incease in degraded isoform (red) - 186 genes

Splicing regulation: decrease in generation of productive isoform (blue) indicates shift towards producing unproductive isoform (red) - 115 genes)

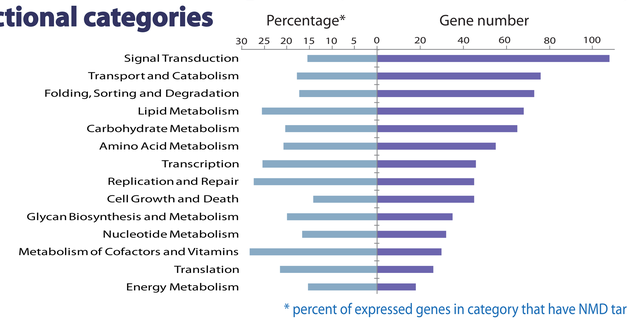
...and new ones were discovered

KHDRBS1



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

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



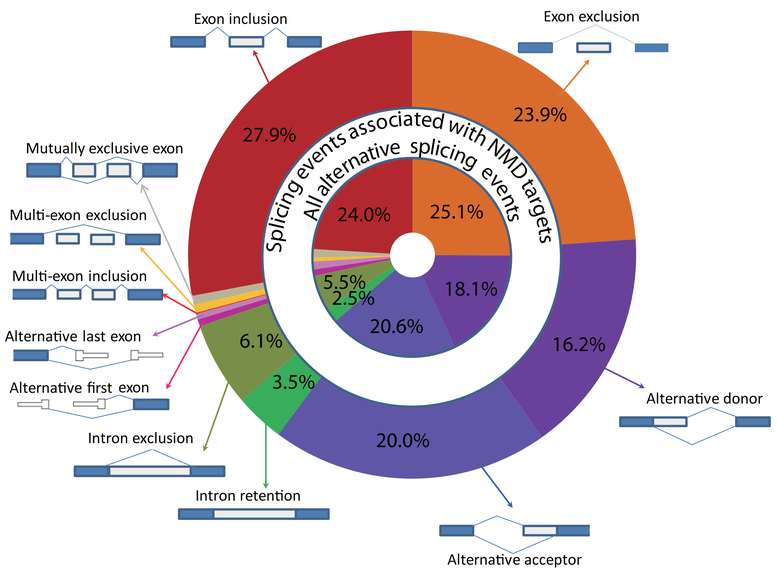
* percent of expressed genes in category that have NMD targeted isoforms

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

Splicing factor category	Genes expressed	Number 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	CIRBP, HNRNPA2B1, HNRPDL, HNRNPH1, HNRNPH3, HNRNPK, HNRNPL, PCBP2, PTBP1, PTBP2, RBM3, SYNCRIP
snRNP	39	10	SNRNP70, SNRNP48, SNRNP40, TXNL4A, SART1, PRPF3, PPIH, U2AF1, U2AF2, U2AFL4
DEAD	15	5	DDX5, DDX46, DHX9, DHX15, INTS6
Sm	18	2	SNRPB, SNRPN
Other	114	35	ACIN1, C16orf80, CDK12, CLASRP, CLK1, CRNKL1, DNAJC8, EIF2S2, FUBP3, FUS, GCFC1, ISY1, LUC7L3, MOV10, NCBP2, PPIE, PRPF4B, RBM5, RBM39, SFPQ, SMNDC1, SRPK1, SREK1, SRRM1, SRRM2, TCERG1, THOC2, THOC4, TIA1, TIAL1, TOP1MT, TRA2A, TRA2B, U2SURP, ZNF207

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]).

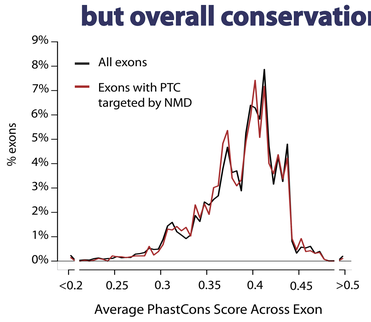
Cassette exon alternative splicing events are the dominant method of producing NMD targets



Prevalence of different categories of alternative splicing events found to be significantly different between samples [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: alternative splicing exon

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

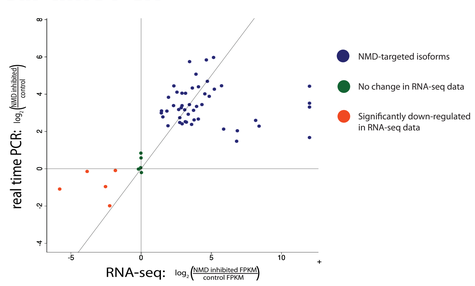


Distributions of the average conservation across an exon. The conservation score per base is the phastCons score from the 46 vertebrate alignment.

Functional Category	Genes with isoforms targeted by NMD overlapping UCE	
RNA processing	19	DDX5, DHX15, HNRNPH1, HNRNPK, HNRPDL, PCBP2, PRPF38B, PRPF39, PTBP2, RBM39, SRSF1, SRSF3, SRSF6, SRSF7, SRSF11, SYNCRIP, TRA2A, TRA2B, ZFR
Transcriptional regualtion	4	CCAR1, MED1, MGA, NFAT5
Other	3	FAM98A, MRRF, STRN3

27 of 79 genes that overlap an exonic ultraconserved element are NMD targets (significantly enriched by Fisher's exact test, p<6e-5). Ultraconserved regions are defined as >200 bp of 100% sequence identity between human, mouse, and rat [12].

NMD-targeted transcripts are validated by real-time PCR



All 48 NMD-targeted transcripts tested (with varying expression levels) increased when NMD was inhibited according to real-time PCR performed using isoform-specific primers on four biological replicates.

CONCLUSIONS:

- Almost 2,500 isoforms from 1,900 genes are degraded by NMD.
- Splicing regulators are significantly enriched for NMD targets.
  - Genes from many other functional categories also produce NMD targets.
- Exons of transcripts targeted by NMD are significantly enriched for ultraconserved elements.
  - Their overall conservation is similar to that of all exons genome-wide.
- NMD-targeted isoforms can be generated by various splicing events.
- 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 C and Salzberg SL. How to map billions of short reads onto genomes. *Nature Biotechnology.* 2009. 7:457
3. Trapnell C, Pachter L, Salzberg SL. Tophat: discovering splice junctions with RNA-seq. *Bioinformatics.* 2009. 25:1105-11
4. Trapnell, et al. Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology.* 2010. 28:511-5
5. Nagy E and Maquat LE. 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 LF, Inada M, Green RE, Wengrod JC, Brenner SE. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature.* 2007. 446:926-9
7. Ni JZ, 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, Tan LY, Paul N, Zavolan M, Lilley KS, Smith CW. 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 NL, Carmo-Fonseca M, Aparicio, S. 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 AN, et al. Conservation of an RNA regulatory map between Drosphila and mammals. *Genome Research.* 2011. 21:193-202
12. Bejerano G, 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.