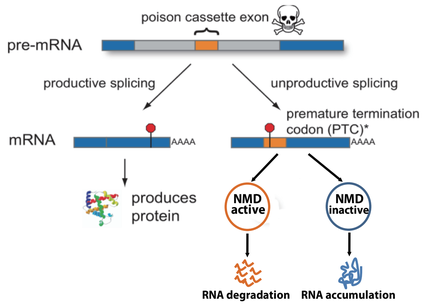
Thousands of nonsense-mediated mRNA decay targets revealed by transcriptome analysis offer clues to the mechanism of degradation

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.

*These authors contributed equally to this work [$]brenner@compbio.berkeley.edu

INTRODUCTION:

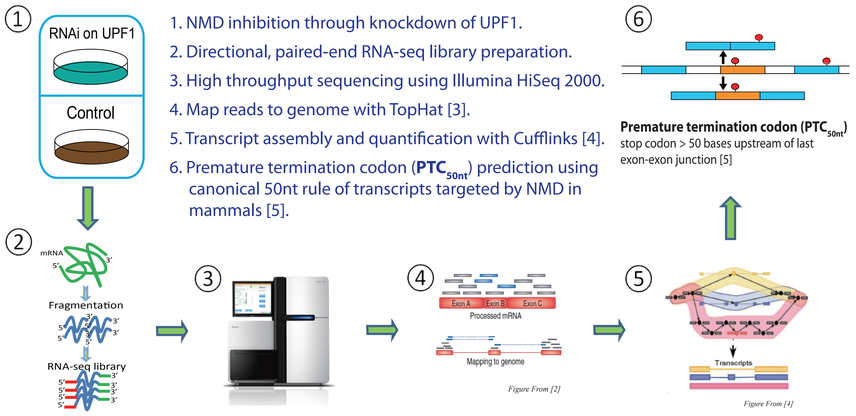


Alternative splicing plays a major role in the generation of proteomic diversity. However, mistakes in this process can introduce a premature termination codon (PTC) and result in non-functional proteins that are harmful to the cell. Such transcripts are usually degraded by nonsense-mediated mRNA decay (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?
How does NMD recognize early termination codons?

APPROACH:



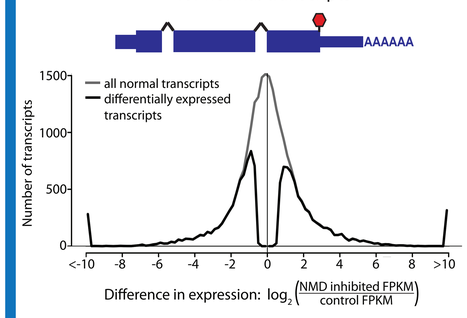
1. NMD inhibition through knockdown of UPF1.
2. Directional, paired-end RNA-seq library preparation.
3. High throughput sequencing using Illumina HiSeq 2000.
4. Map reads to genome with TopHat [3].
5. Transcript assembly and quantification with Cufflinks [4].
6. Premature termination codon (**PTC$_{50nt}$**) 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).

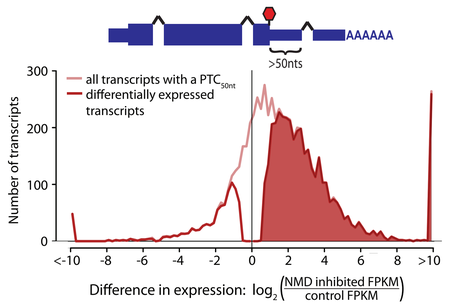
RESULTS:

Almost 2,800 robustly expressed transcripts, from a fifth of expressed genes, were identified as putative NMD targets

A No bias in direction of abundance changes for normal transcripts



B Transcripts with a PTC$_{50nt}$ are more likely to increase in abundance

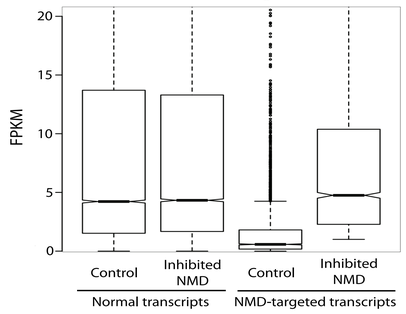


Altogether, 23,888 normal transcripts and 6,429 PTC$_{50nt}$ transcripts were expressed at FPKM>1 in at least one sample. A PTC$_{50nt}$ transcript has an exon-exon junction >50nts downstream of the termination codon. The abundance fold change distribution was symmetric for normal transcripts (A), but PTC$_{50nt}$ transcripts showed a strong bias toward increased abundance when NMD was inhibited (B).

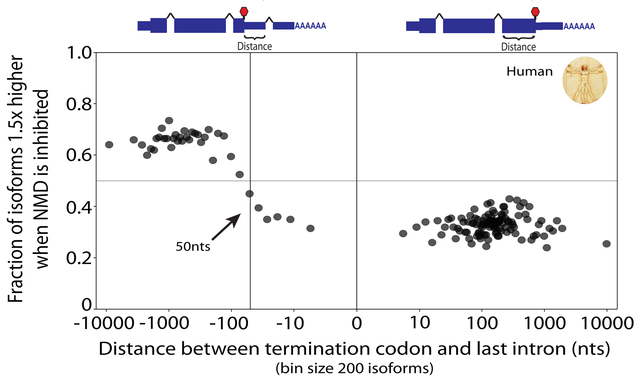
2,793 PTC$_{50nt}$ transcripts were significantly increased and defined as putative NMD targets (9% of expressed transcripts). They were derived from 2,116 genes, 19% of genes expressed in HeLa cells.

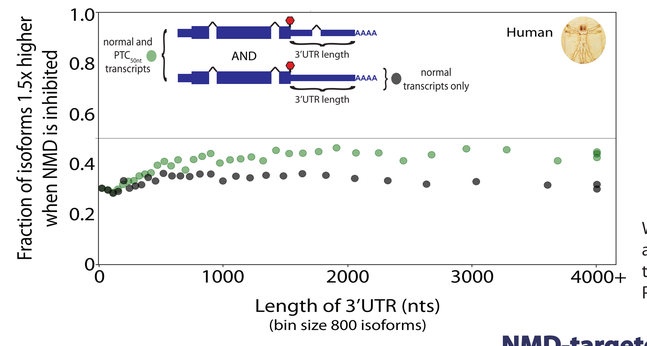
Before degradation, NMD-targeted transcripts can be produced at as high a level as normal transcripts (C).

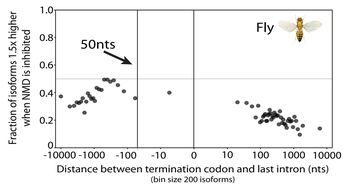
C Many transcripts degraded by NMD are transcribed at high levels



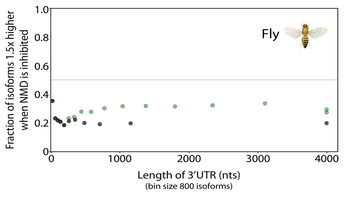
A PTC$_{50nt}$ is a strong predictor of degradation through NMD while a long 3'UTR has little effet





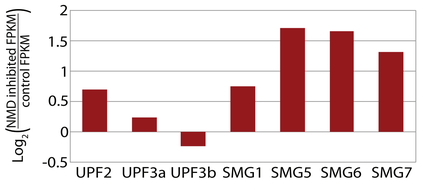


The 50nt rule is a powerful predictor of transcript stabilization and thus NMD degradation in human cells. There is a clear phase change in the fraction of isoforms that increase when NMD is inhibited between those with and without a PTC$_{50nt}$. Not all transcripts follow this rule, perhaps due to other NMD mechanisms or secondary effects of the UPF1 knockdown. Analogous analysis with fly RNA-seq data [modENCODE Fly Transcriptome Group] indicates the 50nt rule may not play as strong a role outside mammals.



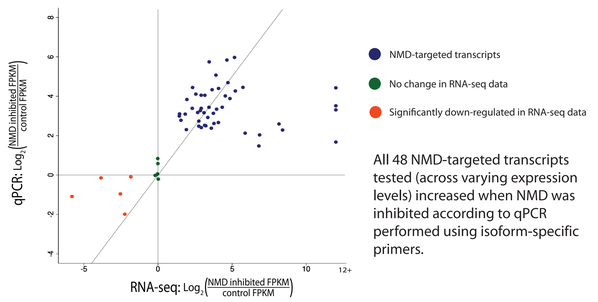
While we see a slight correlation between 3' UTR length and liklihood of increased expression across all transcripts (green), this effect largely disappears when PTC$_{50nt}$ transcripts are excluded (gray).

Gene expression of many NMD factors increased upon UPF1 knockdown



As previously reported [11], the NMD factors UPF2, SMG1, SMG5, SMG6, and SMG7 had increased gene expression levels when NMD is inhibited, while UPF3a and UPF3b do not.

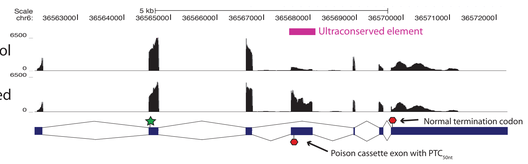
NMD-targeted transcripts are validated by qPCR



All 48 NMD-targeted transcripts tested (across varying expression levels) increased when NMD was inhibited according to qPCR performed using isoform-specific primers.

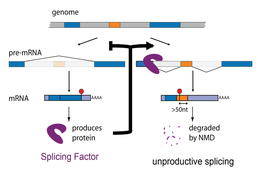
Previously inferred splicing events resulting in NMD targets were confirmed...

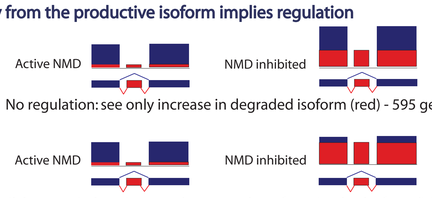
SRSF3



SRSF3 (SRp20), a SR protein, has been identified to have an NMD-targeted isoform [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



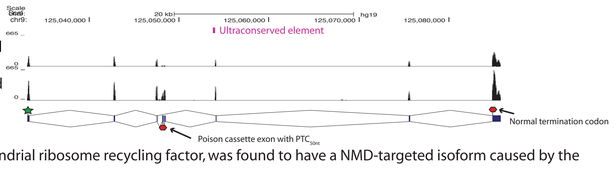


No regulation: see only increase in degraded isoform (red) - 595 genes

Splicing regulation: decrease in generation of productive isoform (blue) indicates shift towards producing unproductive isoform (red) - 355 genes)

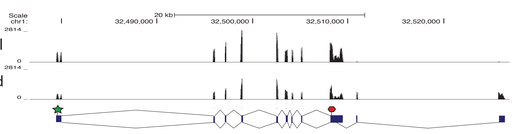
...and new ones were discovered, for example:

MRRF



MRRF, a mitochondrial ribosome recycling factor, was found to have a NMD-targeted isoform caused by the inclusion of two cassette exons.

KHDRBS1

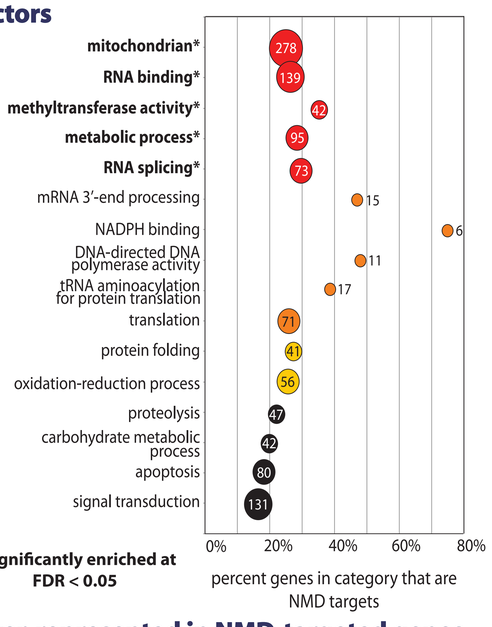


KHDRBS1, a signal transduction associated protein, was found to have a NMD-targeted isoform caused by alternative last exons.

Genes with an NMD-targeted isoform fall into diverse functional groups and are enriched for splicing factors

Splicing factor category	Genes with isoforms targeted by NMD
SR proteins	SRSF1, SRSF2, SRSF3, SRSF4, SRSF5, SRSF6, SRSF7, SRSF8, SRSF10, SRSF11
hnRNP	CIRBP, HNRNPA2B1, HNRPDL, HNRNPH1, HNRNPH3, HNRNPK, PTBP1, PTBP2, RBM3
snRNP	PPIH, PRPF3, SART1, SNRNP40, SNRNP48, SNRNP70, TXNL4A, U2AF1, U2AF2
DEAD	DDX5, DDX46, DHX9, DHX15, DHX35, INTS6
Sm	SNRPA1, SNRPB, SNRPN
Other	ACIN1, AKAP17A, CCAR1, C16orf80, CDK12, CIR1, CLASRP, CLK1, CPSF1, CRNKL1, CSTF1, DGCR14, DNAJC8, EIF2S2, FUBP3, FUS, GCFC1, GTF2F2, IVNS1ABP, LUC7L3, MAGOHB, MOV10, NCBP2, POLR2E, PRPF4, PRPF4B, PRPF18, PRPF38B, RBM25, SCAF8, SF3B1, SFPQ, SFSWAP, SMNDC1, SRPK1, SREK1, SRRM1, SRRM2, SUGP1, SUGP2, TCERG1, THOC2, THOC4, TIA1, TIAL1, TOP1MT, TRA2A, TRA2B, TXNL4B, U2SURP, USP39, YTHDC1, ZCCHC8, ZFR, ZNF207, ZRSR2

Splicing factors have been described as producing isoforms targeted by NMD [6,7,8]. We found 20 previously reported (red) and many more (list of factors from [9]). RNA splicing genes were significantly enriched for NMD targets (FDR corrected p<0.05).



number of genes in category that are NMD targets

0-20 21-50 51-100 101-200 201-300

raw p-value: <0.0001 0.0001-0.01 0.01-0.05 >0.05

*** significantly enriched at FDR < 0.05**

Ultraconserved elements are over-represented in NMD-targeted genes

Functional Category		Genes with isoforms targeted by NMD overlapping ultraconserved element
RNA processing	16	DDX5, DHX15, HNRNPH1, HNRNPK, HNRPDL, PRPF38B, PTBP2, SRSF1, SRSF3, SRSF6, SRSF7, SRSF11, TIAL1, TRA2A, TRA2B, ZFR
Transcriptional regulation	4	CCAR1, MED1, MGA, NFAT5
Other	6	HIRA, RC3H2, FAM98A, MRRF, STRN3, DLG2

26 of 73 genes that overlap an exonic ultraconserved element are NMD targets (significantly enriched by Fisher's exact test, p<6.8e-4). Ultraconserved elements are defined as >200 bp of 100% sequence identity between human, mouse, and rat [10].

CONCLUSIONS:

- Over 2,700 robustly expressed isoforms from over 2,100 genes (19% of expressed genes) are degraded by NMD.
- 1,548 of NMD-targeted transcripts are novel isoforms (55%).
- Splicing regulators are significantly enriched for NMD targets.
  - Genes from many other functional categories also produce NMD targets.
- Transcripts targeted by NMD are significantly enriched for exonic ultraconserved elements.
- Coupling of alternative splicing and NMD appears to regulate the expression of hundreds of genes.
- There is strong support for the 50nt rule in NMD degradation in human cells.
  - Support for a role for 3'UTR length in NMD is limited.

ACKNOWLEDGEMENTS:

The fly UPF1 knockdown RNA-seq data were generated by the **modENCODE Fly Transcriptome Group** (Sue Celniker, Brenton Graveley, Steven Brenner, Gemma May, Li Yang, Angela Brooks). Read alignments were done by Mike Duff, Sandrine Dudoit, Kasper Hansen.

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

This work was funded by an NIH Grant R01-GM071655 to S.E.B, G.W is also supported by a Tang Distinguished Scholarship from QB3 at UC Berkeley, and C.E.F. was supported by an NIH GGD training grant.

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. *Nature Biotechnology*. 2009. 7:457
3. Trapnell C, Pachter L, Salzberg SL. *Bioinformatics*. 2009. 25:1105-11
4. Trapnell, et al. *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. *BMC Genomics*. 2010. 11:565
9. Barbosa-Morais NL, Carmo-Fonseca M, Aparicio, S. *Genome Research*. 2006. 16:66-77
10. Bejerano G, et al. Ultraconserved elements in the human genome. *Science*. 2004. 304:1321-5
11. Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Mühlemann O. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA*. 2011. 17:2108-18.