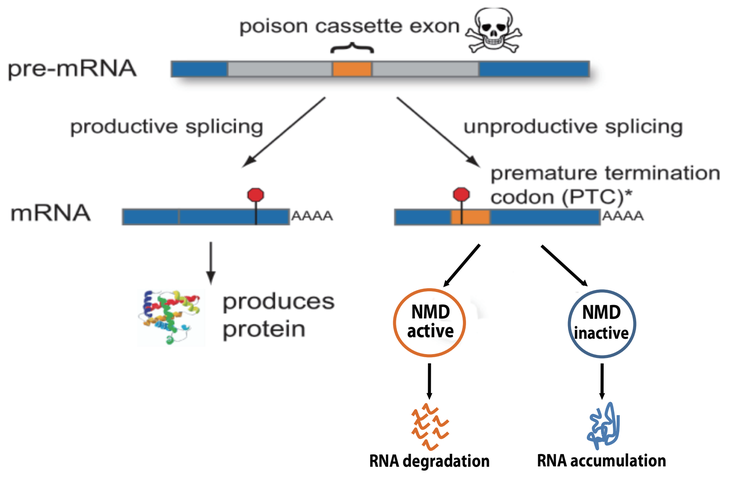
Thousands of targets of nonsense-mediated mRNA decay revealed by transcriptome analysis offer clues to the mechanism in multiple species

Courtney E. French[1], Anna Desai[2], Gang Wei[3], James P.B. Lloyd[3,4], Thomas L. Gallagher[5], Darwin S. Dichman[1], Maki Inada[6], Sharon L. Amacher[5], Richard M. Harland[1], and Steven E. Brenner[1,3*]

[1]Department of Molecular and Cell Biology, [2]Department of Comparative Biochemistry, [3]Department of Plant and Microbial Biology, [4]Center for RNA Systems Biology, University of California, Berkeley, CA. [5]Department of Molecular Genetics, Ohio State University, Columbus, OH, [5]Department of Biology, Ithaca College, Ithaca, NY *brenner@compbio.berkeley.edu

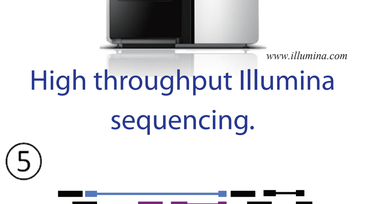
Download PDF

INTRODUCTION:

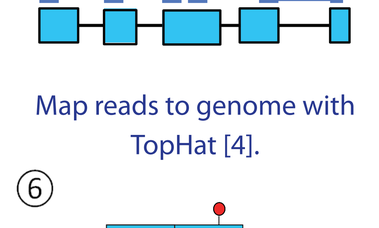


Nonsense-mediated mRNA decay (NMD) is an RNA surveillance system that degrades aberrant isoforms containing a premature termination codon. This pathway is conserved throughout eukaryotes and protects against the production of harmful truncated proteins. Additionally, NMD coupled with alternative splicing is a mechanism of post-transcriptional gene regulation that affects the mRNA levels of hundreds of genes in human [1]. The canonical model of defining a premature termination codon in mammals is the 50nt rule: a termination codon more than 50 nucleotides upstream of an exon-exon junction is premature and triggers degradation by NMD [2]. In other animals, a 3′ UTR intron is not required for NMD. There is also evidence that a longer 3′ UTR triggers NMD in plants, flies, and mammals [3]. The importance of each mechanism appears to vary between species, and it is currently unclear which is the major mechanism at work in human cells. We used RNA-seq analysis done on cells with inhibted NMD to determine the features associated with degradation in human and in fly.

50nt rule



longer 3′ UTR

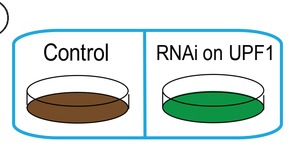


GOALS:

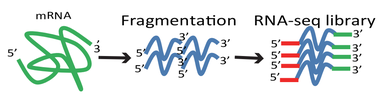
How conserved are the targets of alternative splicing coupled with NMD?

What features define a premature termination codon in different species?

APPROACH:

① 

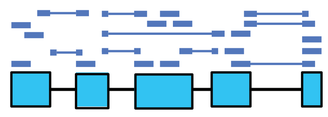
NMD inhibition through knockdown/knockout of UPF1.

② 

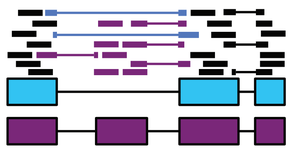
Directional and paired-end RNA-seq library preparation.

③ 

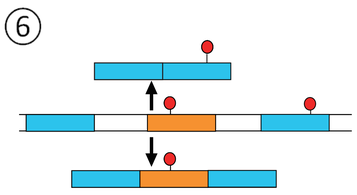
High throughput Illumina sequencing.

④ 

Map reads to genome with TopHat [4].

⑤ 

Transcript assembly and quantification with Cufflinks [5] or JuncBASE [6].

⑥ 

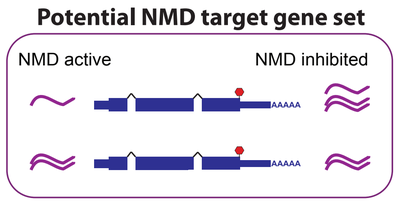
Premature termination codon prediction.

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. 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
3. Kerenyi Z, Merai Z, Hiripi L, Benkovics A, Gyula P, Lacomme C, Barta E, Nagy F, and Silhavt D. Inter-kingdom conservation of mechanisms of nonsense-mediated mRNA decay. *EMBO Journal.* 2008. 27:1585-95
4. Trapnell C, Pachter L, Salzberg SL. Tophat: discovering splice junctions with RNA-seq. *Bioinformatics.* 2009. 25:1105-11
5. 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
6. Brooks AN, et al. Conservation of an RNA regulatory map between Drosphila and mammals. *Genome Research.* 2011. 21:193-202
7. Hurt JA, Robertson AD, Burge CB. Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay. *Genome Research*. 2013. 10:1636-50
8. Brooks AN, et al. Regulation of alternative splicing in Drosophila by 56 RNA binding proteins. *Genome Research*. 2013.

Hundreds to thousands of genes produce alternative isoforms degraded by NMD in each of human, mouse, zebrafish, frog, fly, *Arabidopsis*, and *S. pombe*.

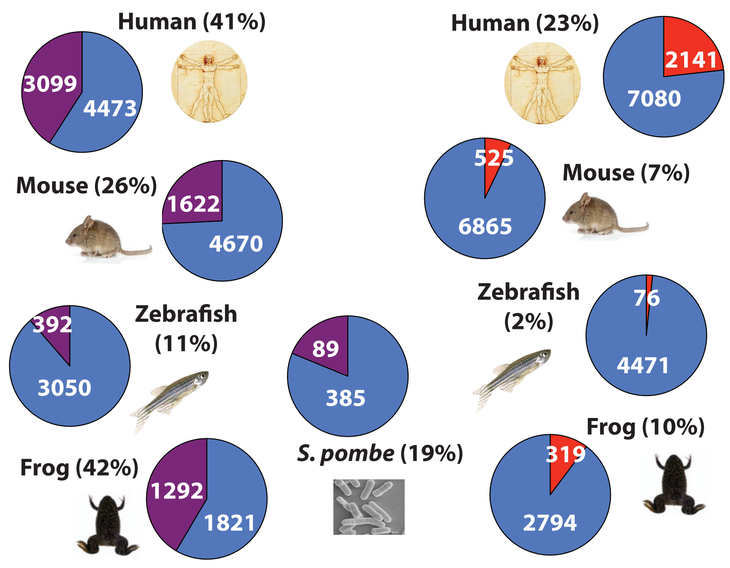
Potential NMD target gene set



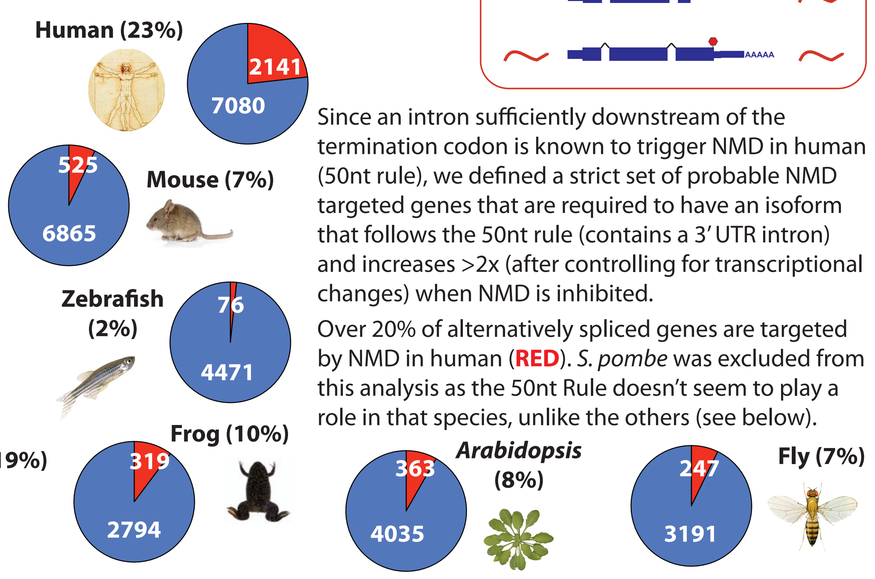
We infer that a gene produces an alternative isoform that may be degraded by NMD if the gene has at least one isoform that increases >2x when NMD is inhibited and at least one isoform that does not increase.

Thousands of genes fall into this category for human, mouse [7], frog, and *Arabidopsis*. Zebrafish and fly [8] have hundreds, while *S. pombe* has dozens (**PURPLE**).

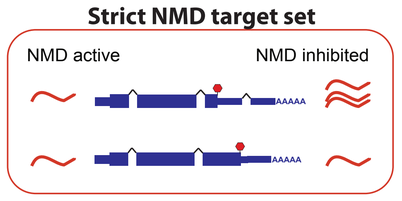
Fraction of genes with >1 isoform (FPKM>1) that are **potential NMD targets**



Fraction of alternatively spliced genes that are in our **strict set of NMD targets**



Strict NMD target set

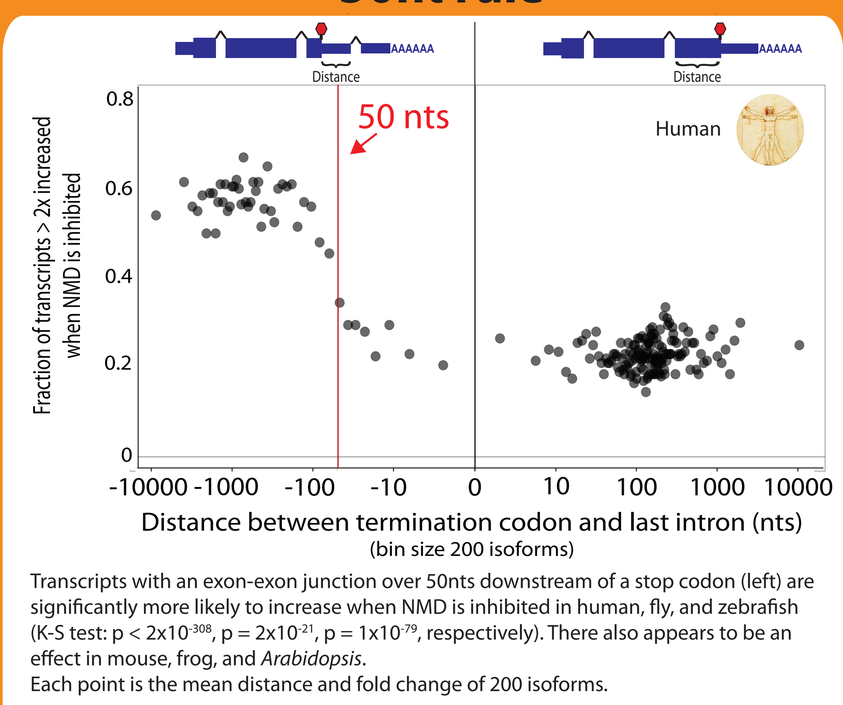


Since an intron sufficiently downstream of the termination codon is known to trigger NMD in human (50nt rule), we defined a strict set of probable NMD targeted genes that are required to have an isoform that follows the 50nt rule (contains a 3′ UTR intron) and increases >2x (after controlling for transcriptional changes) when NMD is inhibited.

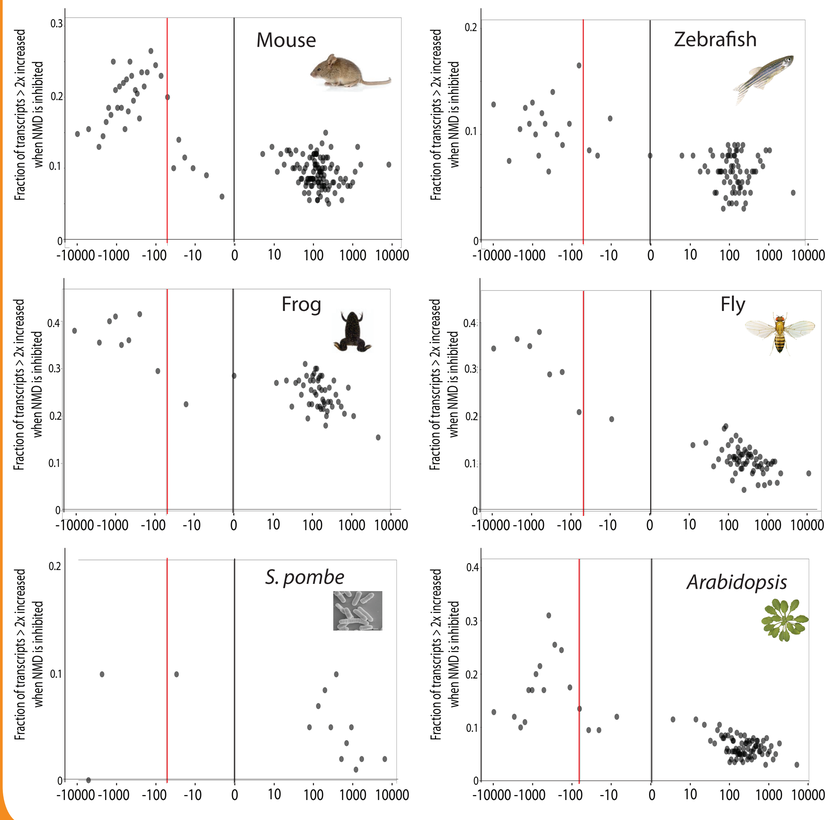
Over 20% of alternatively spliced genes are targeted by NMD in human (**RED**). *S. pombe* was excluded from this analysis as the 50nt Rule doesn't seem to play a role in that species, unlike the others (see below).

The 50nt rule is a strong predictor of NMD in human and plays a role in other species while a longer 3′ UTR has little to no effect in any species

50nt rule

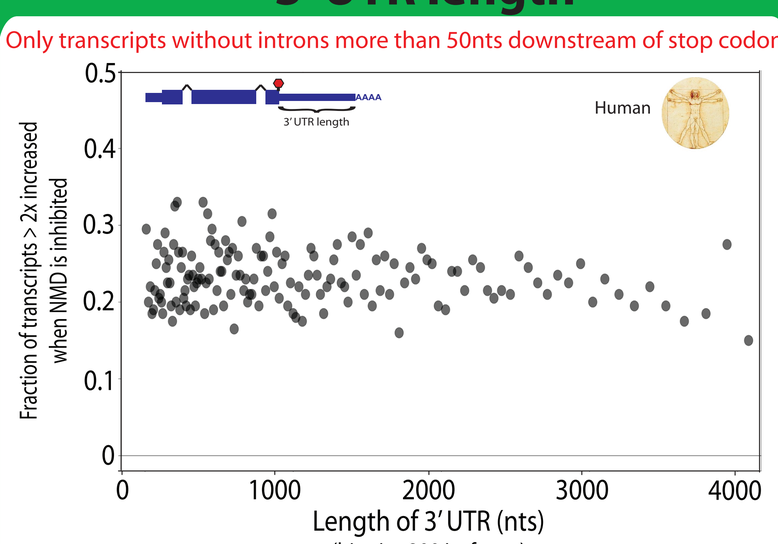


Transcripts with an exon-exon junction over 50nts downstream of a stop codon (left) are significantly more likely to increase when NMD is inhibited in human, fly, and zebrafish (K-S test: $p < 2x10^{-308}$, $p = 2x10^{-21}$, $p = 1x10^{-79}$, respectively). There also appears to be an effect in mouse, frog, and *Arabidopsis*.
Each point is the mean distance and fold change of 200 isoforms.

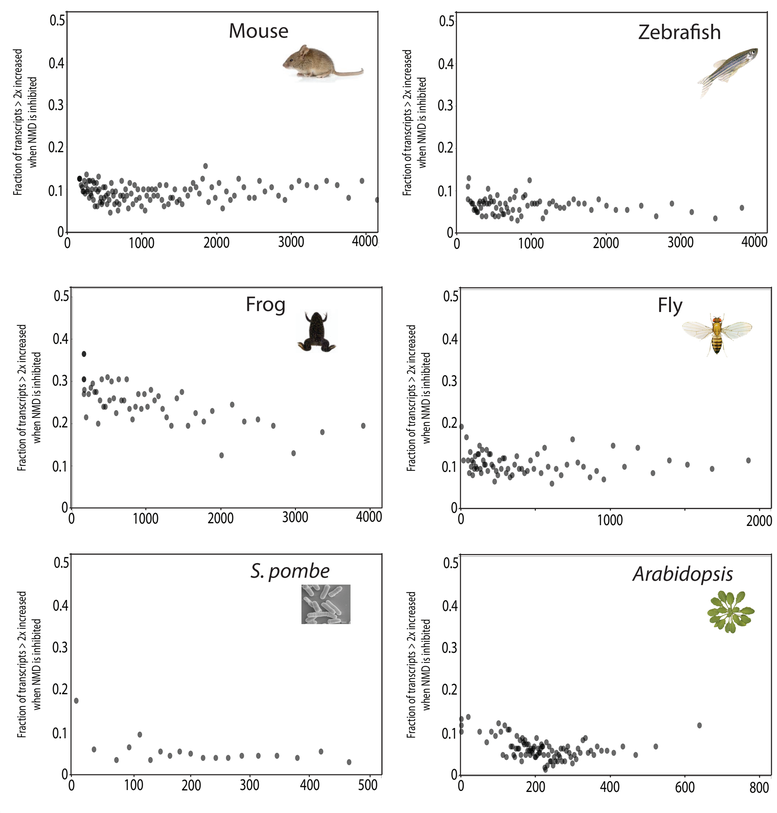


3′ UTR length

Only transcripts without introns more than 50nts downstream of stop codon



When only looking at transcripts that cannot be affected by 50nt rule, we see only a slight correlation between 3′ UTR length and an increase when NMD is inhibited in human and no correlation in fish or fly (K-S test: human $p = 3x10^{-11}$, fish $p = 0.13$, fly $p = 0.81$). There also appears to be no effect in mouse, frog, *S. pombe* and *Arabidopsis*. Each point is the mean length and fold change of 200 isoforms.



CONCLUSIONS:

Thousands of alternatively spliced genes (>20%) produce transcripts that fall into our strict set of NMD targets in human.

Hundreds to thousands of alternatively spliced genes (11-42%) produce transcripts possibly degraded by NMD in diverse eukaryotes.

The 50nt rule is a strong predictor of NMD in human and also appears to have an often more limited role in numerous other species, except *S. pombe*.

3′ UTR length has little correlation with NMD in any of the species checked.

ACKNOWLEDGEMENTS:
Human and *Arabidopsis* experiments done by **Gang Wei**. Zebrafish experiments done by **Thomas Gallagher** in Sharon Amacher's lab. Frog experiments done by **Darwin Dichmann** in Richard Harland's lab. *S. pombe* experiments done by **Maki Inada**. The fly data came from **modENCODE** with thanks to Angela Brooks, Li Yang, and Brenton Graveley.

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