The role of the 50nt rule in targeting a transcript for nonsense-mediated mRNA decay in human and in fly

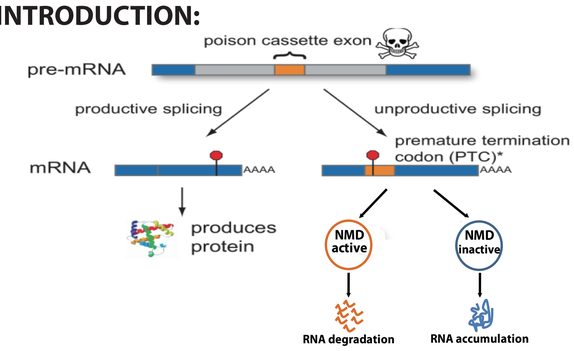
Courtney E. French[1*], Gang Wei[2*], Angela N. Brooks[1] and Steven E. Brenner[1,2]

[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

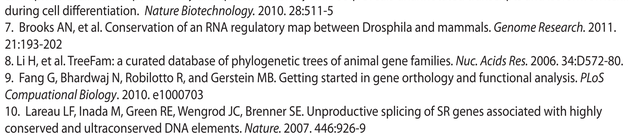
*These authors contributed equally to this work

INTRODUCTION:



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 (mammals) **long 3'UTR (fly)**



GOALS:

How conserved is alternative splicing coupled with NMD between human and fly?

What features define a premature termination codon in the two species?

APPROACH:

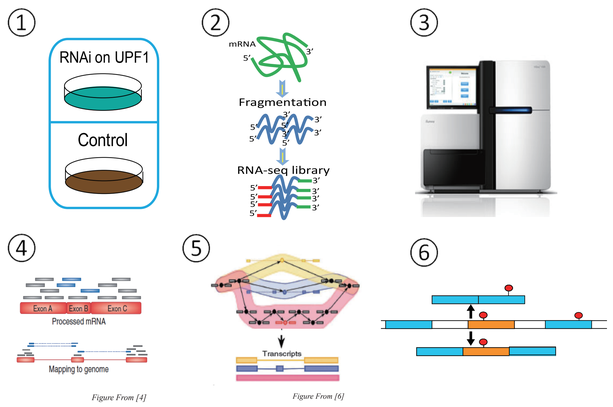


Figure From [4] *Figure From [6]*

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 [5] or juncBASE [7].
5. Transcript assembly and quantification with Cufflinks [6].
6. Premature termination codon prediction.

ACKNOWLEDGEMENTS:

The fly UPF1 knockdown RNA-seq data was 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.

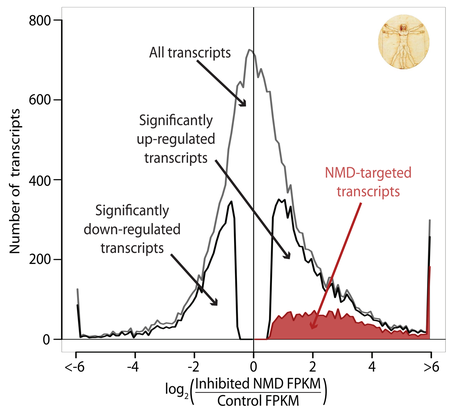
REFERENCES:

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 and Salzberg SL. How to map billions of short reads onto genomes. *Nature Biotechnology.* 2009. 7:457
5. Trapnell C, Pachter L, Salzberg SL. Tophat: discovering splice junctions with RNA-seq. *Bioinformatics.* 2009. 25:1105-11
6. 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
7. Brooks AN, et al. Conservation of an RNA regulatory map between Drosphila and mammals. *Genome Research.* 2011. 21:193-202
8. Li H, et al. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nuc. Acids Res.* 2006. 34:D572-80.
9. Fang G, Bhardwaj N, Robilotto R, and Gerstein MB. Getting started in gene orthology and functional analysis. *PLoS Computational Biology.* 2010. e1000703
10. 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

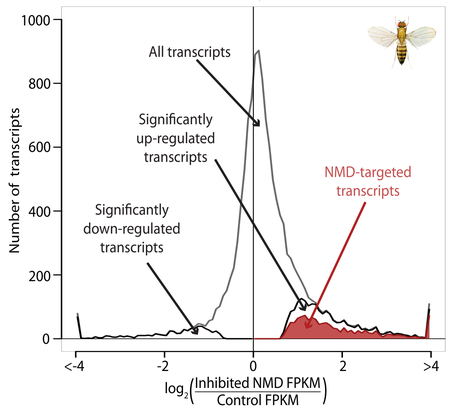
RESULTS:

Thousands of isoforms are degraded by NMD in human and in fly

A We find 2,443 isoforms that are degraded by NMD in human HeLa cells



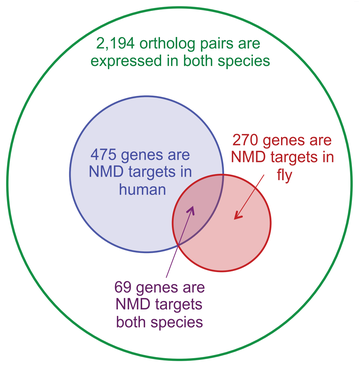
B We find 1,216 isoforms that are degraded by NMD in fly S2 cells



A. Of 21,989 expressed transcripts, 2,443 are NMD targets, defined as being significantly 1.5x higher when NMD is inhibited and containing a premature termination codon according to the 50nt rule [2]. Thus, 1,960 genes (18%) produce NMD-targeted isoforms.

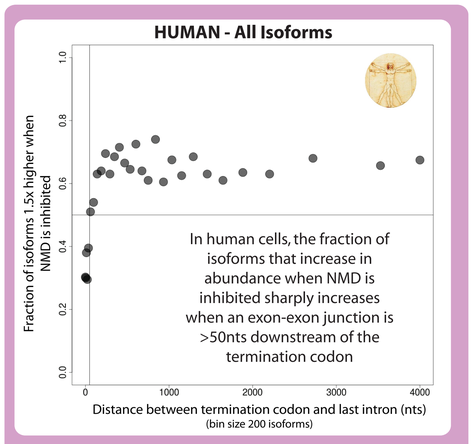
B. Of 12,689 expressed transcripts, 1,216 are NMD targets, defined as being significantly 1.5x higher when NMD is inhibited and coming from genes with an isoform that does not increase. Thus, 931 genes (13%) produce NMD-targeted isoforms.

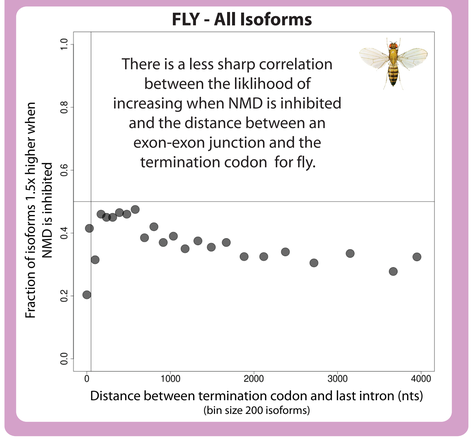
Genes involved in mRNA processing are targeted by NMD in human and fly cells

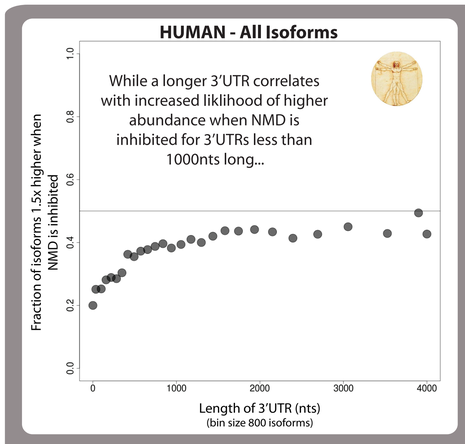


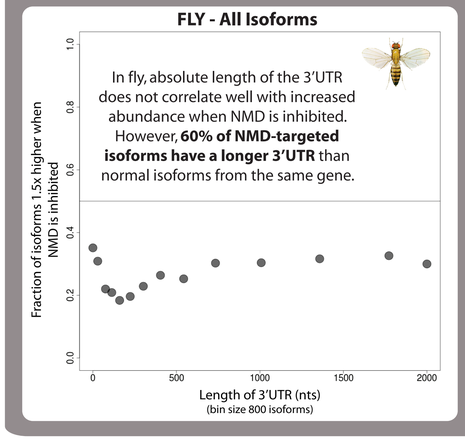
RNA processing genes targeted by NMD in both species	
CLASRP	CLK4-associating serine/arginine rich protein
DHX15	**DEAH (Asp-Glu-Ala-His) box polypeptide 15**
PRPF38B	**PRP38 pre-mRNA processing factor 38 domain containing B**
RBM39	**RNA binding motif protein 39**
SCAF4	SR-related CTD-associated factor 4
SCAF8	SR-related CTD-associated factor 8
SUGP1	SURP and G patch domain containing 1
TARBP1	TAR RNA binding protein 1
TRMT2A	TRM2 tRNA methyltransferase 2 homolog A
U2SURP	U2 snRNP-associated SURP domain containin

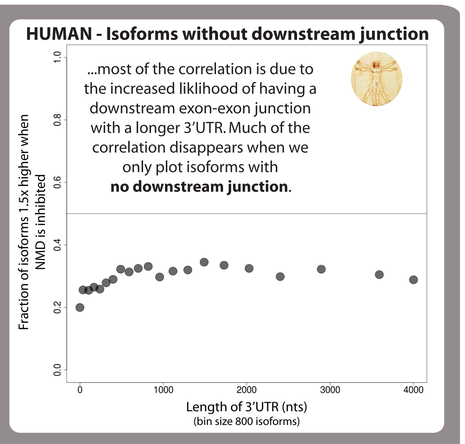
Of 2,194 orthologs expressed in both our human and fly data, 69 genes were NMD targets in both species. Of the five most enriched GO terms (uncorrected p<0.005), three involved RNA processing. Three of these genes overlap ultraconsered elements of the human genome (**PURPLE**). Orthologs from TreeFam 7.0 and collected by Gerstein, et al [8,9]

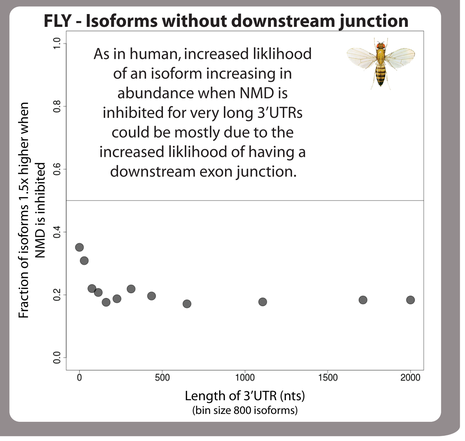


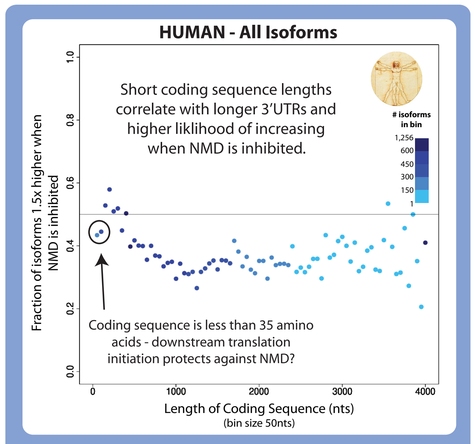


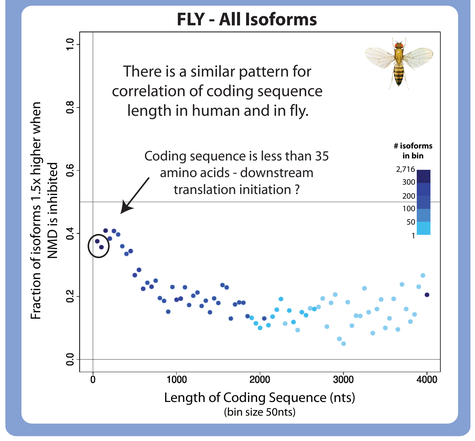








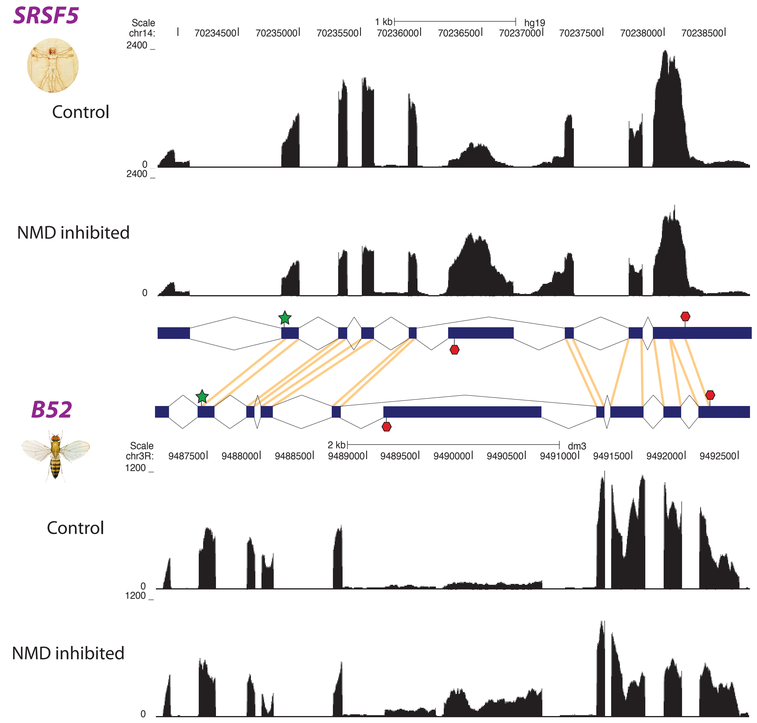




CONCLUSIONS:

- Over 1,200 alternative isoforms from over 900 genes are degraded by NMD in fly.
- The conserved NMD targets between human and fly include many mRNA processing factors.
- There is strong support for the 50nt rule of NMD degradation in human cells.
  - Support for a role for 3'UTR length in NMD is limited.
- Coding sequences less than 35 amino acids long have diminished NMD susceptibility.
- In fly, 60% of NMD-targeted isoforms have a longer 3'UTR than normal isoforms from the same gene.
  - A downstream exon junction may enhance NMD in fly.

SRSF5/B52 is an NMD target conserved between human and fly



Both the human gene SRSF5 and its fly ortholog B52 produce NMD-targeted alternative isoforms when the sixth exon is included. All SR genes have NMD-targeted isoforms in human and mouse [10].