Thousands of targets of nonsense-mediated mRNA decay revealed by transcriptome analysis offer clues to the mechanism in human, fish, and fly

Courtney E. French[1], Anna Desai[2], James P. B. Lloyd[3], Gang Wei[4,5], Angela N. Brooks[6], Thomas L. Gallagher[7], Li Yang[8], Brenton R. Gravely[9], Sharon L. Amacher[7] and Steven E. Brenner[1,4*]

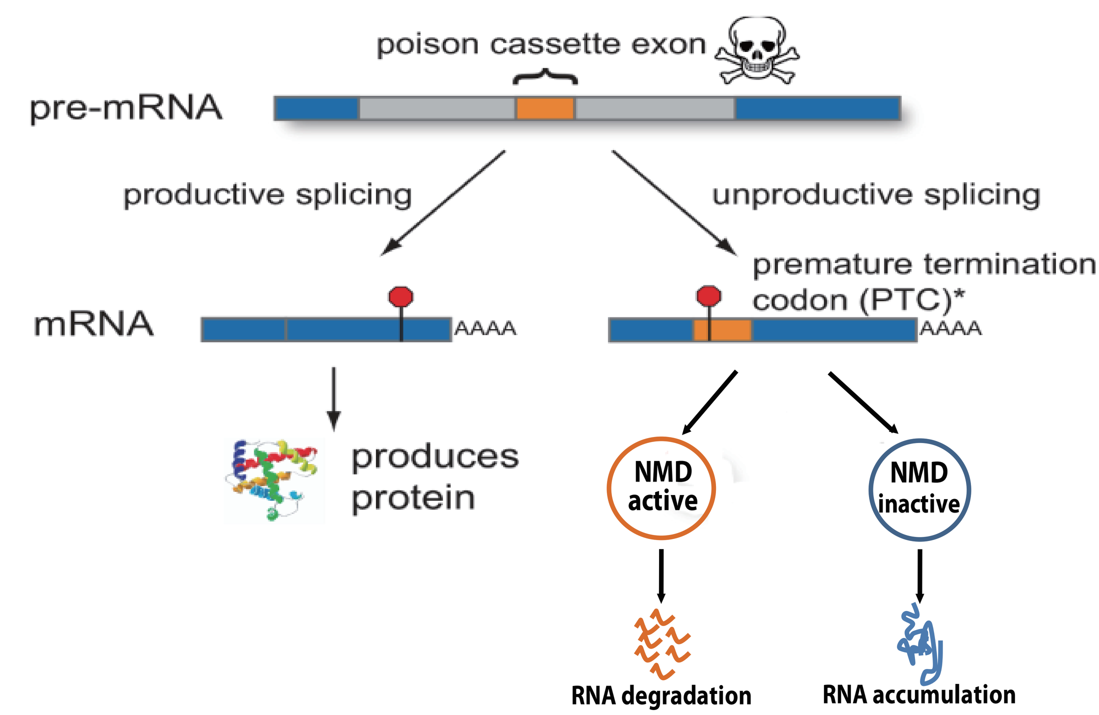
[1]Department of Molecular and Cell Biology, [2]Department of Comparative Biochemistry, [3]Center for RNA Systems Biology, [4]Department of Plant and Microbial Biology, University of California, Berkeley, CA, 94720, USA. [5]Now at Fudan University, Shanghai, China. [6]Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA. [7]Department of Molecular Genetics, Ohio State University, Columbus, OH, 43210, USA. [8]Partner Institute of Computational Biology, Shanghai, China. [9]Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT, 06030, USA. *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 thousands of genes in human.

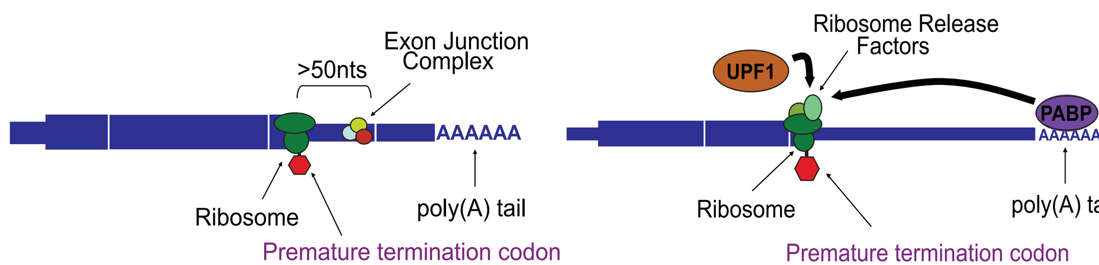
Numerous RNA-binding proteins, including all the human SR splicing factors, are regulated by alternative splicing coupled to NMD, in conjunction with highly- or ultra-conserved elements. This suggests a complex auto- and cross-regulatory network exists, controlling the expression of splicing factors.



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. 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.

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 inhibited NMD to determine the features associated with degradation in human and in fly.

50nt rule **longer 3' UTR**



GOALS:

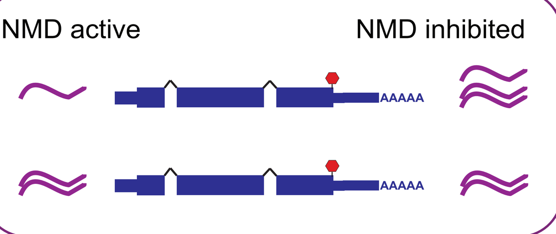
How pervasive is alternative splicing coupled with NMD across eukaryotes?

What features define a premature termination codon in different species?

What is the architecture of the network of alternative splicing coupled with NMD regulation for splicing factors?

Thousands of genes produce alternative isoforms degraded by NMD in human and fish and fly have hundreds

Potential NMD target gene set

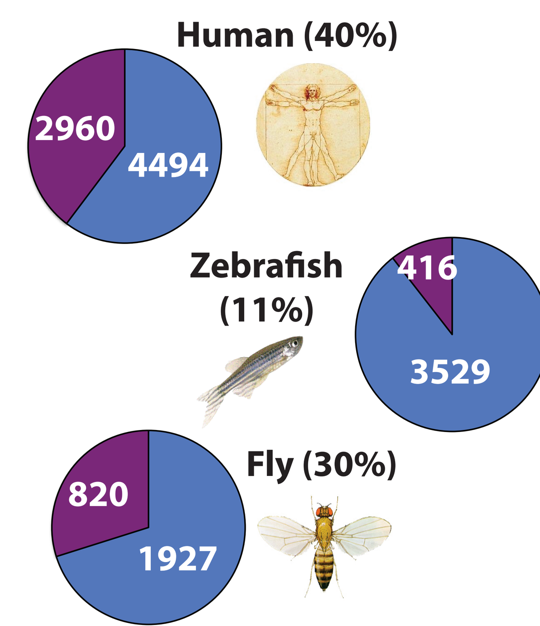


We infer that a gene produces an alternative isoform that may be degraded by NMD if the gene has at least one isoform that increases >2x when NMD is inhibited and at least one isoform that does not increase.

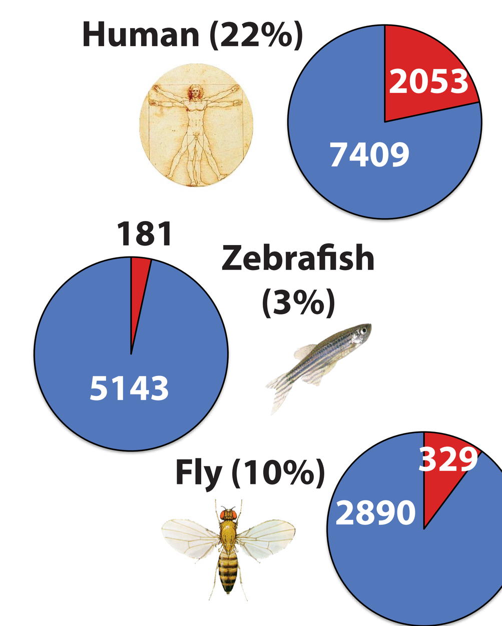
Thousands of genes fall into this category for human. Zebrafish and fly have hundreds (**PURPLE**).

Ribosomal and translation genes are enriched in these NMD targeted genes for fly, and intracellular signaling genes are enriched in zebrafish (Fisher's exact test, FDR<0.05).

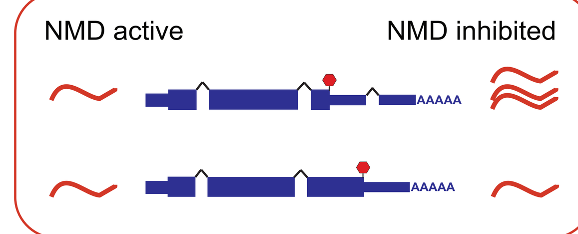
Fraction of genes with >1 isoform (FPKM>1) that are **potential NMD targets**



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



Strict NMD target set

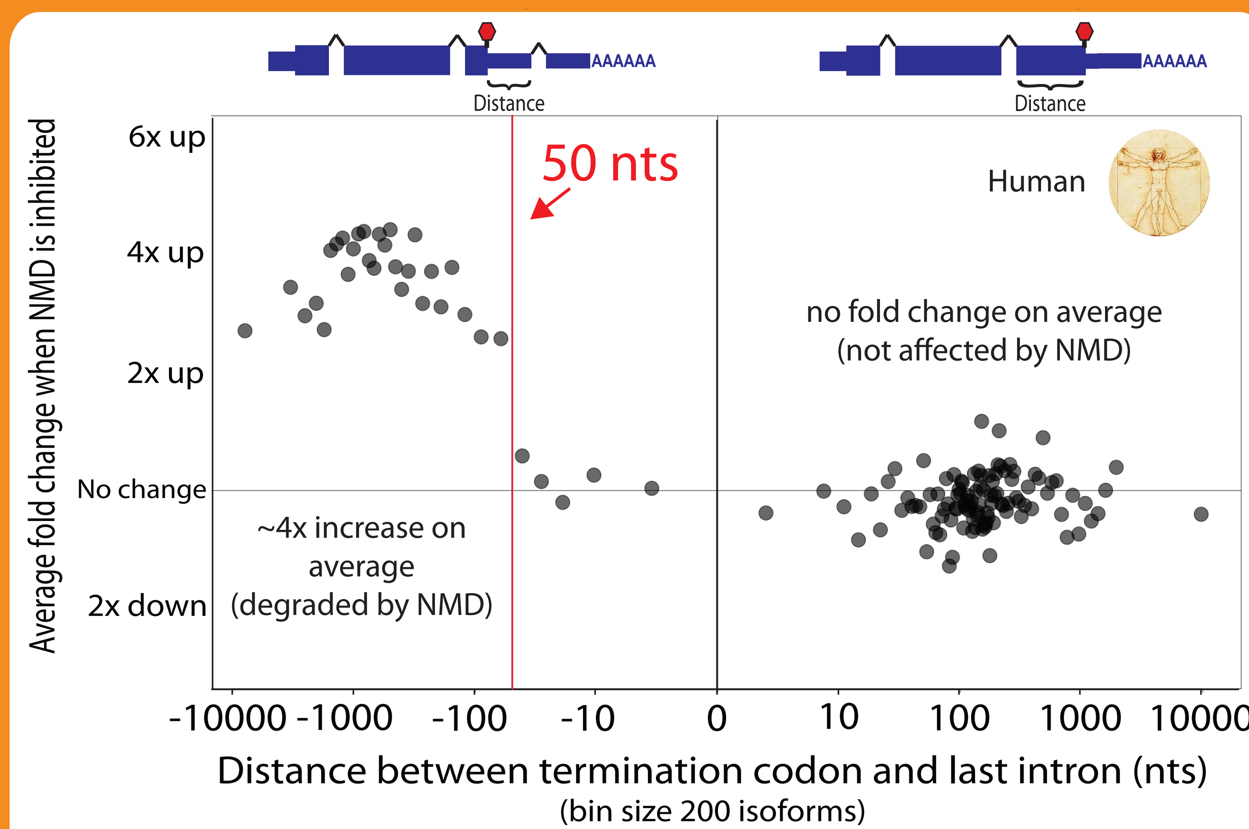


Since an intron sufficiently downstream of the termination codon is known to trigger NMD in human (50nt rule), we defined a strict set of probable NMD targeted genes that are required to have an isoform that follows the 50nt rule (contains a 3' UTR intron) and increases >2x (after controlling for transcriptional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

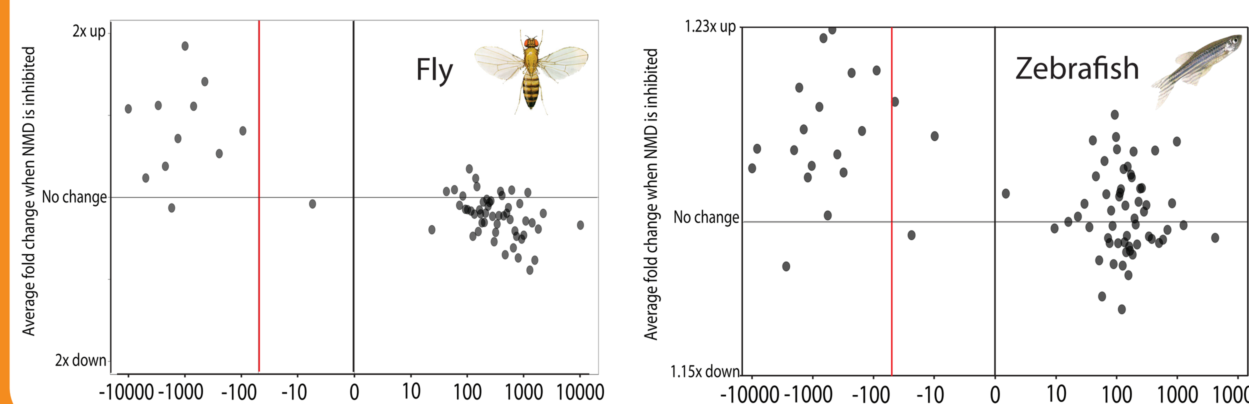
Over 20% of alternatively spliced genes are targeted by NMD in human (**RED**). These genes are enriched for splicing genes in human and in fly (Fisher's exact test, FDR<0.05).

The 50nt rule is a strong predictor of NMD in human, zebrafish and fly while a longer 3' UTR has a limited effect

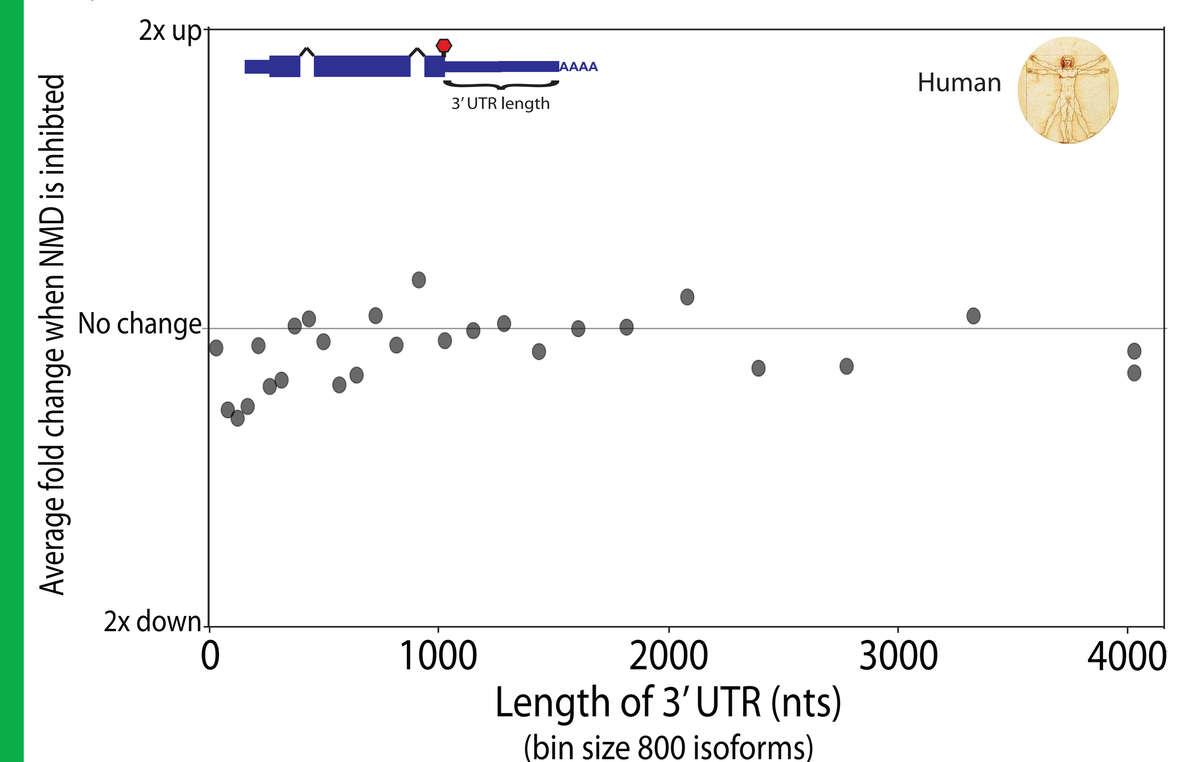
50nt rule



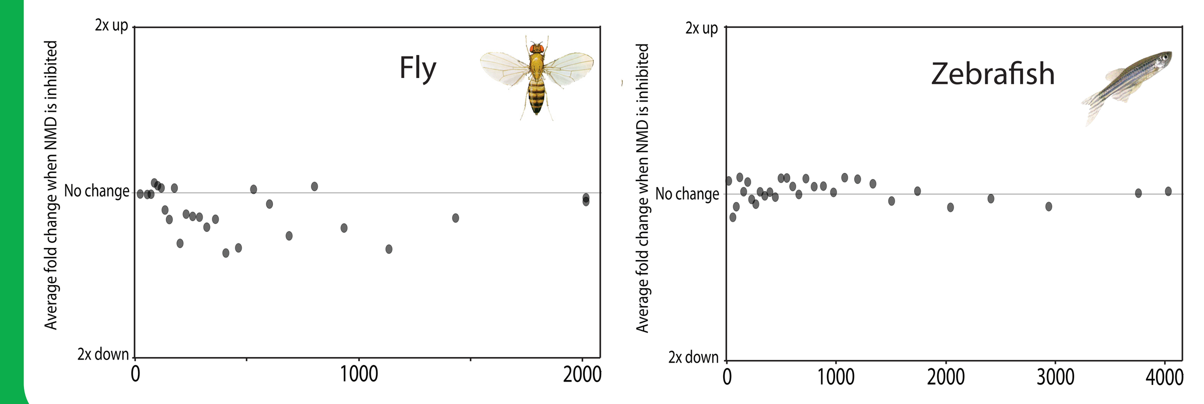
Transcripts with an exon-exon junction over 50nts downstream of a stop codon (left) are significantly more likely to increase when NMD is inhibited in human, fly, and zebrafish (K-S test: $p < 2x10^{-308}$, $p = 2x10^{-21}$, $p = 1x10^{-79}$, respectively).
Each point is the mean distance and fold change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



3' UTR length



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 $p = 0.81$). Each point is the mean length and fold change of 800 isoforms.



Highly connected splicing factor network suggests extensive auto- and cross-regulation by alternative splicing coupled to NMD

Literature-based splicing factor interaction network

Blue edges: Negative regulation by alternative splicing coupled to NMD, observed in human
(**Purple edges:** Observed in mouse)

Red text: Gene produces NMD-targeted isoform

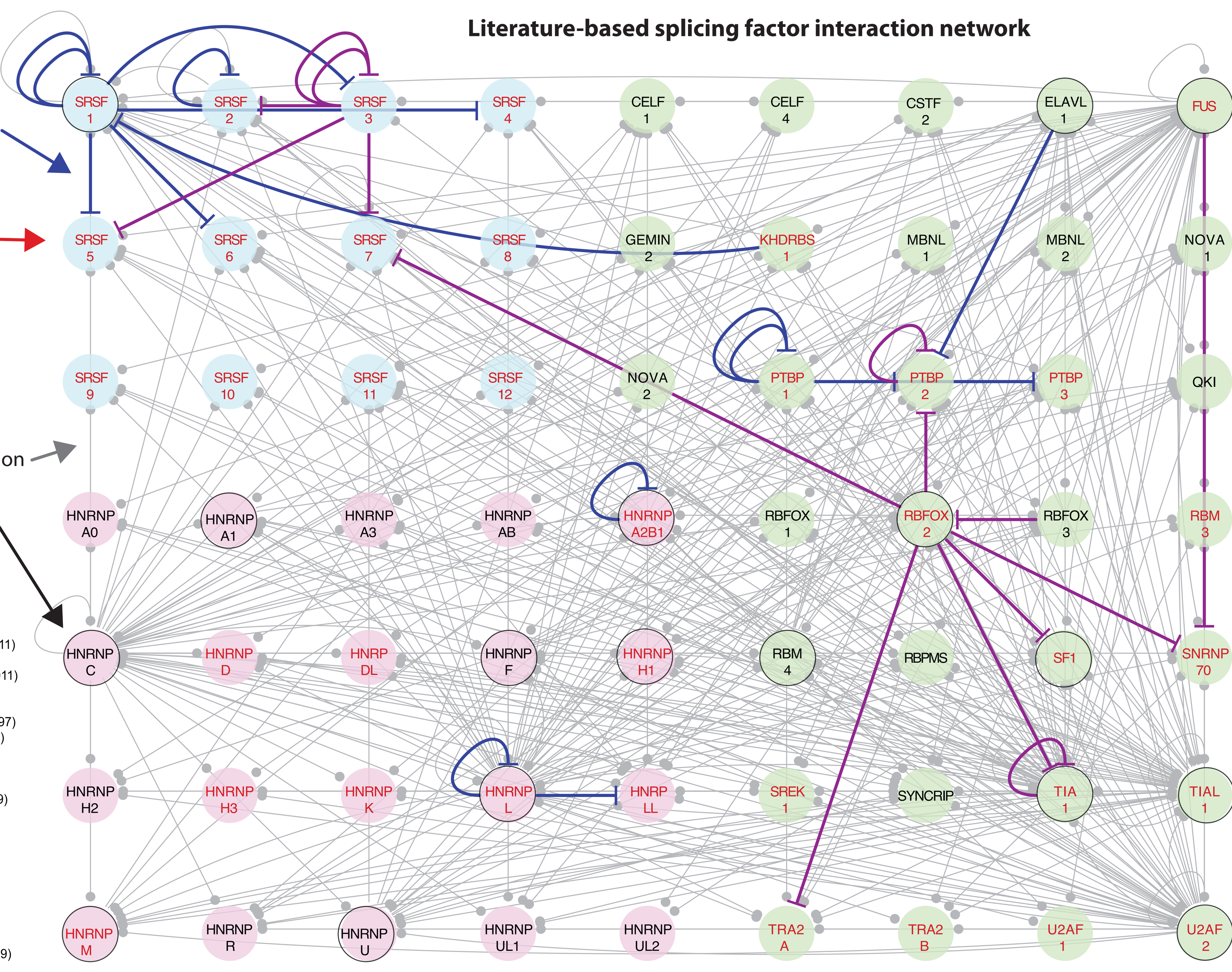
SR family of splicing factors

hnRNP splicing factors

Other splicing factors

Grey edges: protein-mRNA interaction

Black outline: CLIP-seq performed on this factor



REFERENCES:

Änkö, M.-L. *et al. Genome Biol.* **13,** R17 (2012)
Corioni, M. *et al. Nucleic Acids Res.* **39,** 1868–1879 (2011)
Dredge B. K. and Jensen K. B. *PLoS One* **6** e21585
Hoell, J. I. *et al. Nat. Struct. Mol. Biol.* **18** 1428-1431 (2011)
Huelga, S. C. *et al. Cell Reports* **1,** 167–178 (2012)
Jangi, M. *et al. Genes & Dev* **28**, 637–651 (2014)
Jumaa, H. & Nielsen, P. J. *EMBO J.* **16,** 5077–5085 (1997)
König, J. *et al. Nat. Struct. Mol. Biol.* **17,** 909–915 (2010)
Lebedeva, S. *et al. Mol. Cell* **43** 340-352 (2011)
McGlincy, N. J. *et al. BMC Genomics* **11,** 565 (2010)
Nakaya, T. *et al*. RNA **19** 498-509 (2013)
Rossbach, O. *et al. Mol. Cell. Biol.* **29,** 1442–1451 (2009)
Sanford, J. R. *et al. Genome Res.* **19,** 381–394 (2009)
Shankarling, G. *et al. Mol. Cell Biol.* **34** 71-83 (2014)
Spellman, R. *et al. Mol. Cell* **27,** 420–434 (2007)
Sun, S. *et al. Nat. Struct. Mol. Biol.* **17,** 306–312 (2010)
Sureau, A. *et al. EMBO J.* **20,** 1785–1796 (2001)
Uniacke, J. *et al. Nature* **486** 126-129 (2012)
Valacca, C. *et al. J. Cell Biol.* **191,** 87–99 (2010)
Wang, Z. *et al. PLoS Biol.* **8** e1000530 (2010)
Wollerton, M. C. *et al. Mol. Cell* **13,** 91–100 (2004)
Yeo, G. W. *et al. Nat. Struct. Mol. Biol.* **16** 130–137 (2009)
Zarnack, K. et al. Cell 152 453-466 (2013)

CONCLUSIONS:

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

Hundreds of alternatively spliced genes (10-30%) produce transcripts possibly degraded by NMD in fly and zebrafish.

The 50nt rule is a strong predictor of NMD in human and also appears to have a role in fly and zebrafish.

3' UTR length has little correlation with NMD in human, fly, and zebrafish.

Splicing factor genes are enriched in human NMD targets.

Extensive protein-mRNA interactions reveal the potential for pervasive regulation of splicing factors through alternative splicing coupled with NMD.

There is little evidence of a hierarchy with "master regulators" of splicing factors.

ACKNOWLEDGEMENTS:
This work was funded by NIH Grant R01 GM071655 to S.E.B., G.W. was supported by a Tang Distinguished Scholarship from QB3 at UC Berkeley, C.E.F. was supported by a DOD National Defense Science and Engineering Graduate Fellowship, J.P.B.L. was supported by the Center for RNA Systems Biology at UC Berkeley, and A.D. was supported by NIH Grant 5F31 GM108462.