Elucidating the role of mRNA splicing in disease pathogenesis

Ankeeta Shah[1], Yang I Li[2,3]

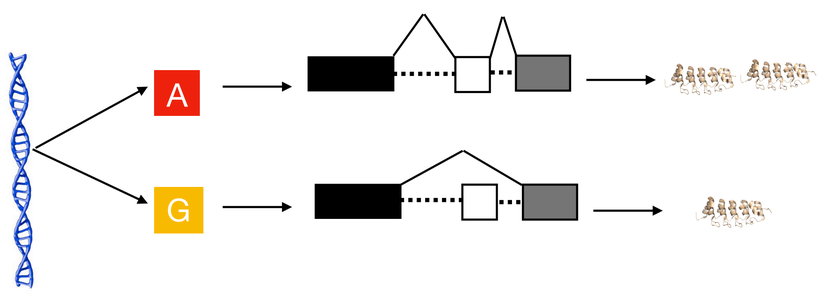
[1] Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL

[2] Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL

[3] Department of Human Genetics, University of Chicago, Chicago, IL

ankeetashah@uchicago.edu

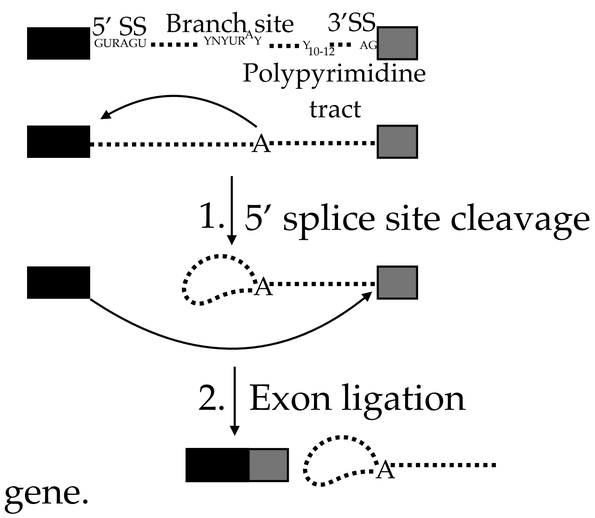
75% of disease-associated genetic variants may disrupt gene regulation in a manner independent of promoters and enhancers*.



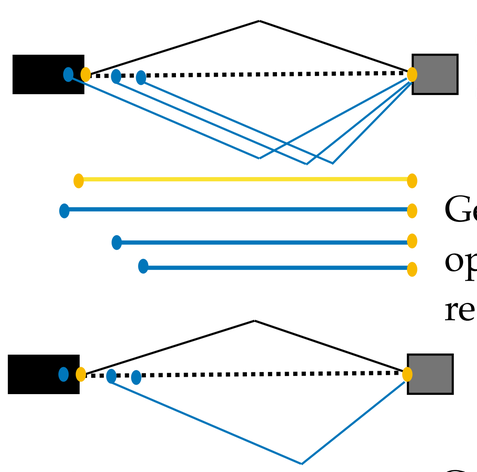
Trait-associated variant | mRNA Splicing | Phenotype

mRNA splicing can create multiple transcripts from the same gene. However, the function of many of these alternative transcripts is unknown (i.e. a result of splicing errors/noisy splicing, Pickrell et al., *PLoS Genetics*, 2010). *Chun et al., *Nature Genetics*, 2017.

mRNA splicing involves two transesterification reactions.



Hypothesis: genetic variants that increase risk of disease through splicing may do so by reducing splicing accuracy.



494,296 splicing junctions extracted from Yoruba GEUVADIS* lymphoblastoid cells lines (LCLs)

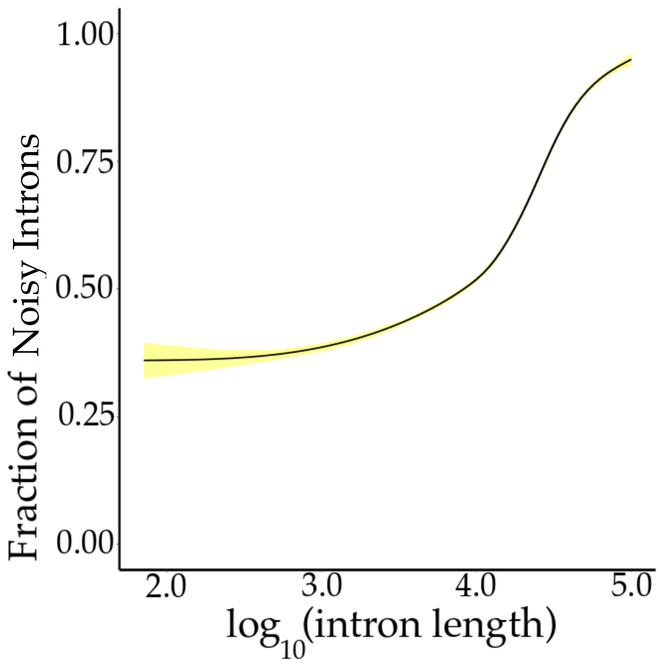
204,159 "noisy" splicing junctions**

290, 137 optimal splicing junctions

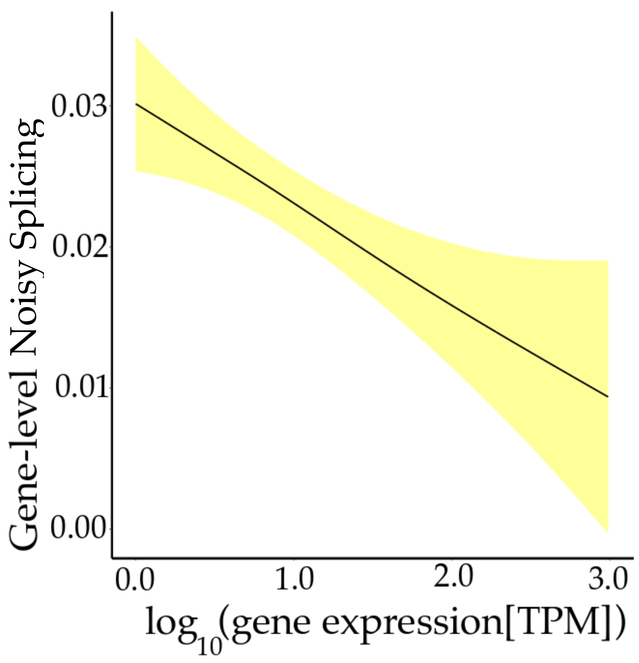
*Lappalainen et al., *Nature*, 2013

**Noisy splicing events called using an adaptation of LeafCutter (Li et al., *Nature Genetics*, 2018)

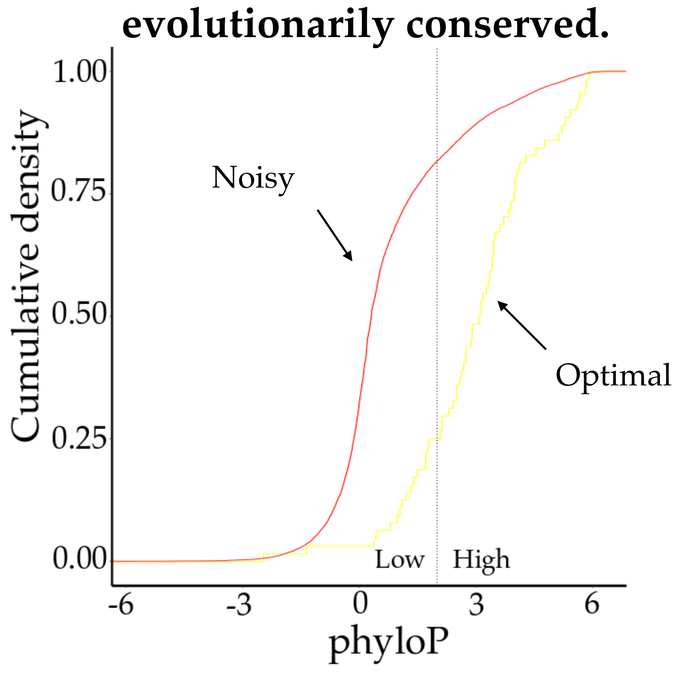
Splicing error correlates with intron length.



Highly expressed genes exhibit less "error-prone" splicing.



Suboptimal splicing junctions are less likely to be evolutionarily conserved.



Splicing quantitative trait loci (sQTLs) can be explained by SNPs that (a) disrupt recognition of splice sites or (b) enhance recognition of "noisy" splice sites.

Genotype for individual i

$$Y_i = X_i\beta + \epsilon_i$$

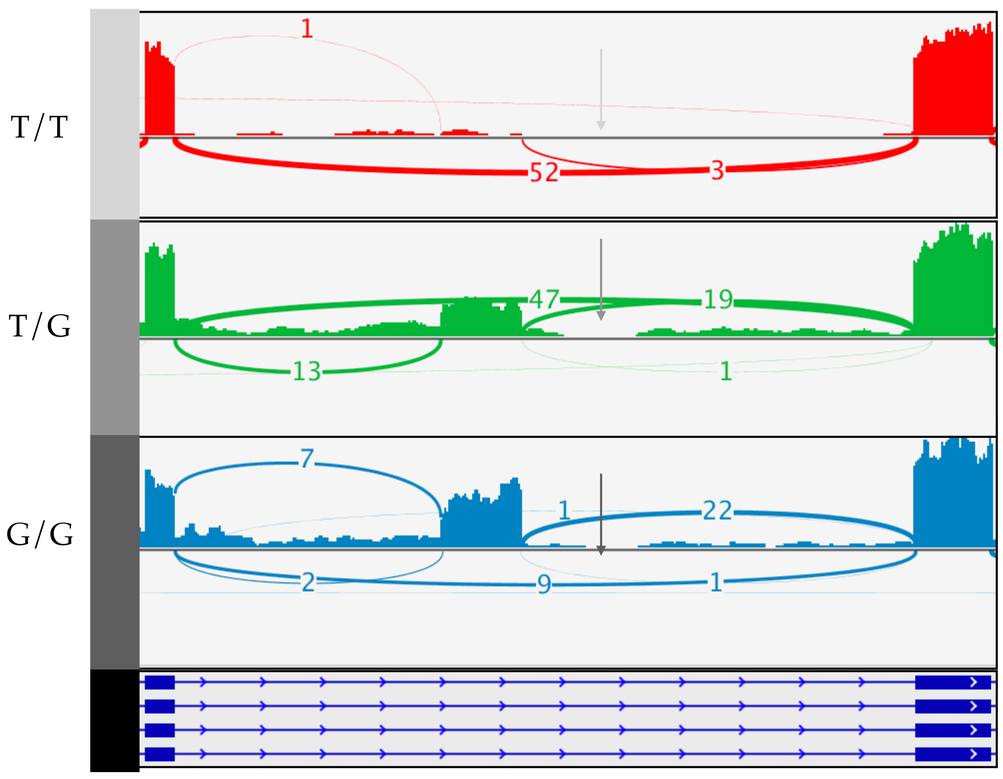
Aggregated "noisy" event usage (i.e. splicing nosiness) for individual i

Effect size

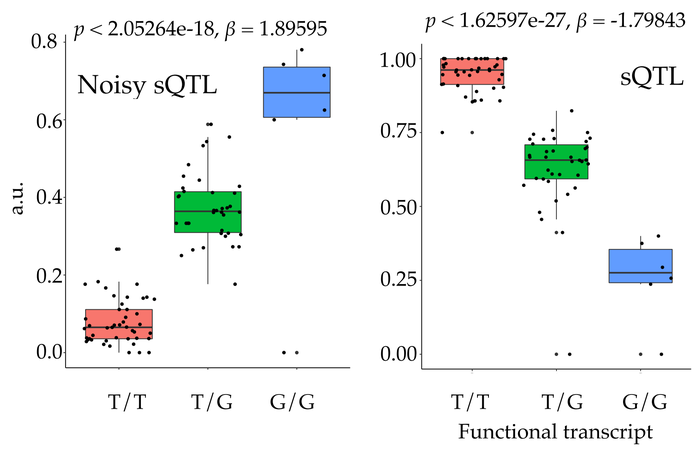
Error

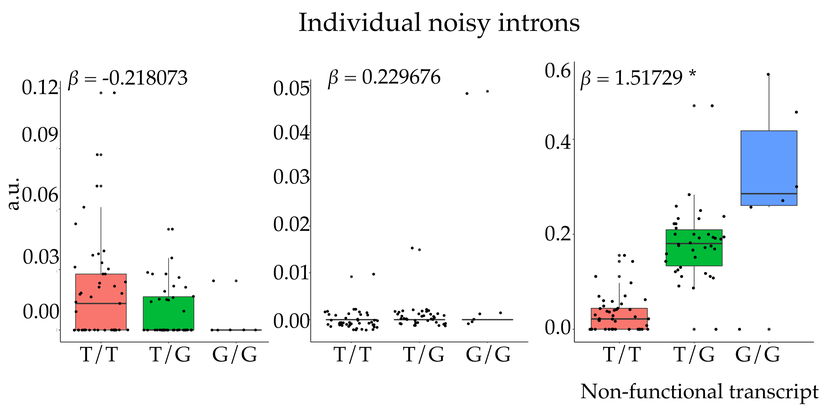
	eQTL	sQTL	Noisy sQTL
Yoruba LCLs (N = 89)	2,553	2,096	178

FastQTL, Ongen et al., *Bioinformatics*, 2016



G allele at rs6808944 increases use of a *single* cryptic SS pair, resulting in the inclusion of a pseudo-exon.





Conclusions and Future Directions

Having gained intuition about how noisy splicing can explain how some SNPs act to contribute to variation in splicing, we will test the hypothesis that SNPs that increase disease risk through splicing likely do so by reducing splicing accuracy (e.g. produce non-functional mRNA transcripts) by performing transcriptome-wide association studies (TWAS) for a number of diseases.

Acknowledgements
NIH T32 GM007197 (AS)
UChicago Research Computing Center