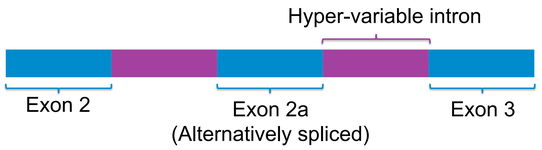
Genetic Variation in Introns as a Cause of Disease: The "Other" Genetic Code

Abigail Neininger[1], William Stewart, Ph.D.[1,2,3], David A. Greenberg, Ph.D.[1,3]

1. Battelle Center for Mathematical Medicine, Nationwide Children's Hospital, Columbus OH
2. Statistics Department, College of Arts and Sciences, The Ohio State University, Columbus OH
3. Pediatrics Department, College of Medicine, The Ohio State University, Columbus OH

background

- Disease-causing variants in exons have been extensively studied but intronic disease-related variation has not.
- *BRD2* is a chr. 6 gene linked and associated with a common form of epilepsy called juvenile myoclonic epilepsy (JME) but *BRD2* exonic variants **do no**t cause JME[2].
- However, in *BRD2* an **intron with many variants** has been associated with JME. This intron follows an alternatively spliced exon that leads to nonsense-mediated decay (NMD), altering the amount of BRD2 in the cell. This suggests that highly variable NMD-related introns play a role in gene control and potentially in disease. Is this variability a feature of NMD-related introns?



- We asked: **Are introns surrounding alternatively spliced (AS) and NMD-causing exons more variable than other introns?**

methods

- We examined all introns from 96 genes that contained at least one NMD-causing, alternatively-spliced exon. We compared the frequency of variants in those introns.
- Introns were categorized into three groups:

Table 1: Three intron types within the studied genes

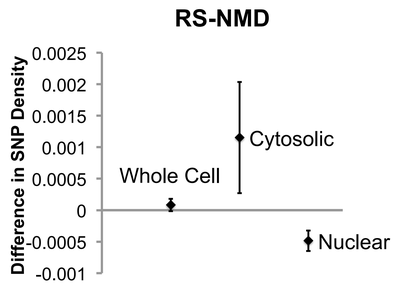
Abbreviation	Intron Type
RS introns	Introns surrounding regularly spliced exons
AS introns	Introns surrounding non NMD-causing AS exons
NMD introns	Introns surrounding NMD-causing AS exons

- We used sequence data from 1000 Genomes[1].
- We used a 2-sample test for equality of proportions without continuity to compare SNP density (variants/base pair) of the introns.
- We also compared densities by noting where in the cell the gene products localized by splitting genes into "cytosolic" and "nuclear". "Nuclear" genes were further split into transcription factors (like *BRD2*) and non-transcription factors.

results

We found highly significant differences in SNP Density depending on the role of the exon and localization of the product.

Figure 1: 95% confidence interval for the difference between RS SNP Density and NMD SNP Density



- Genes whose products localize to the nucleus have significantly *more* variants in introns surrounding NMD-causing alternatively spliced exons compared to introns around other types of exons.
- In contrast, for cytosolic protein genes, the opposite is true: Introns around NMD-causing exons have significantly *fewer* variants than introns around "regular" exons.

Figure 2: 95% confidence interval for the difference between RS SNP Density and AS SNP Density

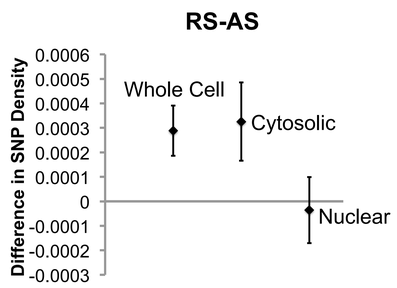


Table 2: SNP density for each gene type (Variants per 10,000 BP)

Protein Properties	RS introns	AS introns	NMD introns
Cytosolic	24.28	21.03	20.35
Nuclear	16.69	17.05	21.55
Transcription Factors	17.28	18.69	23.45
Nuclear non-Transcription Factors	16.47	15.80	20.96
Whole Cell	21.61	18.73	20.80

discussion

We know that NMD is a mechanism for controlling the amount of gene product. In *BRD2*, for example, the amount of NMD can change depending on whether or not the cell is proliferating. Our results suggest the following:

- The results from Figure 1 suggest that intronic variants surrounding NMD exons might be less important for control of cytoplasmic-destined protein production than for nuclear-destined gene products.
- Because NMD can play a crucial role as a gene expression control mechanism, the variant structure of introns surrounding NMD-causing exons could be critical in pathogenesis, especially in genes for nuclear-destined products. Such genes include transcription factors, which strongly influence gene expression. BRD2 is a prime example of such a gene.
- In *nuclear-destined genes*, there is no significant difference between RS introns and non NMD-causing AS introns (Figure 2). However, there is a difference in rates for AS vs. RS for cytosolic-destined genes. Thus, differences in NMD introns in *nuclear genes* may be entirely NMD-related, rather than strictly related to alternative splicing.

One of the conclusions from our work is that SNP *density* and other such non-obvious genome variation is likely to play a critical role in genetic disease susceptibility. Looking *only* for exonic single nucleotide polymorphisms, as is the current fashion, and ignoring subtle changes in non-coding regions probably means we are not looking in the right places for so-called disease mutations.

references

1. McVean et Al, An integrated map of genetic variation from 1,092 human genomes, Nature, 2012, 491, 56-65
2. Pal, DK & Greenberg, DA. Major Susceptibility Genes for Common Idiopathic Epilepsies: ELP4 in Rolandic Epilepsy and BRD2 in Juvenile Myoclonic Epilepsy, Jaspers Basic Mechanisms of the Epilepsies, 4th edition.

acknowledgements

- Our thanks to Ryan Subaran, Rajeswari Swaminathan, William Ray, and grant NS061829 for their assistance in this project.