Title: Don’t Judge a Variant by Its Position: +2T>C and the Splicing Surprise

Authors: Heather L Zimmermann, PhD and Shoji Ichikawa, PhD

Abstract:

BACKGROUND: The consensus donor splice site of U2 introns contains a GT dinucleotide at the +1 and +2 intronic positions. This GT dinucleotide is often considered invariant, and variants that impact these nucleotide positions usually completely disrupt RNA splicing. However, 0.87% of U2 introns in the human reference genome have a C at the +2 position instead of the T, suggesting that GC donor sites can support normal splicing in some genomic contexts. Recent minigene assay data have indicated that 15-18% of +2T>C variants generate varying amounts of normally spliced transcripts, but there is high discordance between different minigene assays for the same variant and limited information on how +2T>C variants impact pathogenicity in humans.

METHODS: To assess the splice impact of +2T>C variants, we evaluated RNA data derived from concurrent DNA + RNA multi-gene oncology panel testing at our laboratory. We compared RNAseq data across 29 native donor splice sites in 15 genes for which we had a +2T>C variant and at least one other canonical variant predicted or observed to result in the same splice defect. To judge the pathogenicity of +2T>C variants, we identified all splice donor sites in the *NF1* gene for which we had a patient with a +2T>C variant (n=9 out of 56 total *NF1* donor sites). We then reviewed available RNA (RNAseq and/or RT-PCRseq) and clinical data for patients heterozygous for the +2T>C variants and any other canonical variants at the identified donor sites.

RESULTS: Across the 29 donor splice sites that met our inclusion criteria, 21% (6/29) of the +2T>C variants had average aberrant splicing at least 15% less than the corresponding non+2T>C canonical variants. In *NF1* specifically, pathogenicity is unlikely for 22% (2/9) of +2T>C variants. Those variants display both reduced aberrant splicing and benign clinical data compared to other variants impacting the GT dinucleotide at the same donor site. Interestingly, the splicing prediction algorithm SpliceAI correctly predicted a reduced splicing impact for 71% (5/7) of the +2T>C variants that appeared to support normal splicing across both studies.

CONCLUSION: During variant classification, +2T>C variants must be interpreted cautiously, particularly when SpliceAI scores are low. The prior probability of pathogenicity for these variants is not as high as for other canonical splice site variants, with ~20% of +2T>C variants able to support a substantial amount of normal splicing.