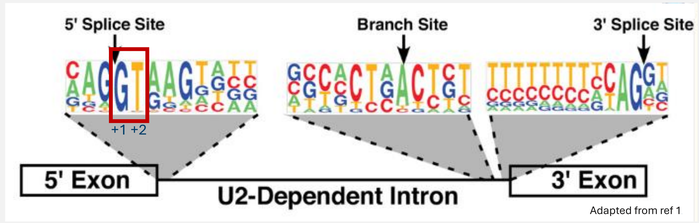
Don’t Judge a Variant by its Position: +2T>C and the Splicing Surprise

Heather L Zimmermann, PhD and Shoji Ichikawa, PhD

Contact: hzimmermann@ambrygen.com

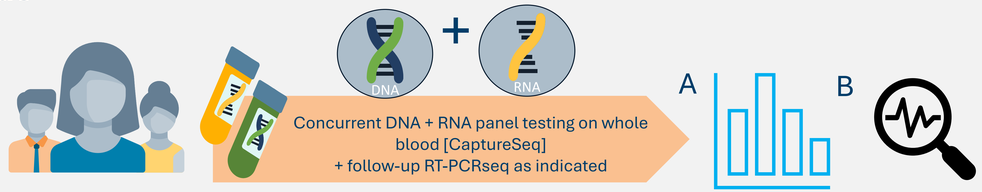
Background



- The GT dinucleotides at +1 and +2 intronic positions of U2-dependent introns are often considered invariant.
- 0.77% of U2 introns in the human reference genome have a C at +2 position[2].

- 15-18% of +2T>C variants support some degree of normal splicing in minigene assays[3].
  - There is a high degree of discordance between different assays for the same variant[4].
  - There is limited information on the clinically relevant impact in humans.

Methods



A. Splice impact: Compare the percent splicing index (PSI) at all native donor sites for which there is ≥1 individual with +2T>C variant and ≥1 individual with a different canonical variant with same expected splice event

B. Pathogenicity impact: Evaluate PSI and clinical data for all individuals with +2T>C variants in *NF1* to assess pathogenicity of +2T>C variants. NF1 pathogenic variants cause neurofibromatosis, a high penetrance disorder.

Results

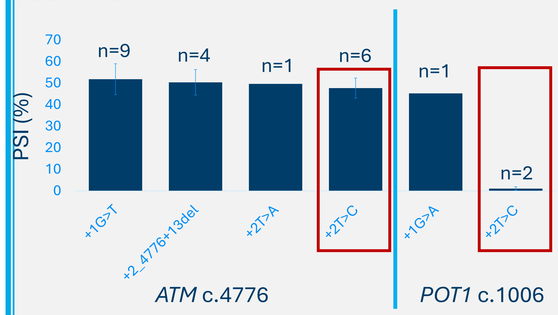


Figure 1. While many +2T>C variants cause splice impacts of similar magnitude to other canonical variants at the same splice site (*e.g.*, left), a subset of +2T>C variants look quite different (*e.g.*, right).

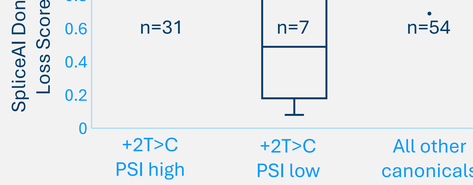


Figure 3. +2T>C variants exhibiting lower PSI than other canonicals (Fig. 2) at the same splice site are more likely to have lower SpliceAI donor loss scores.

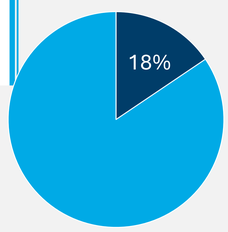


Figure 2. 38 native splice donor sites across 21 genes met inclusion criteria. At 18% (7/38) of donor sites across 7 genes, PSI for +2T>C variants was substantially lower than matched canonical variants.

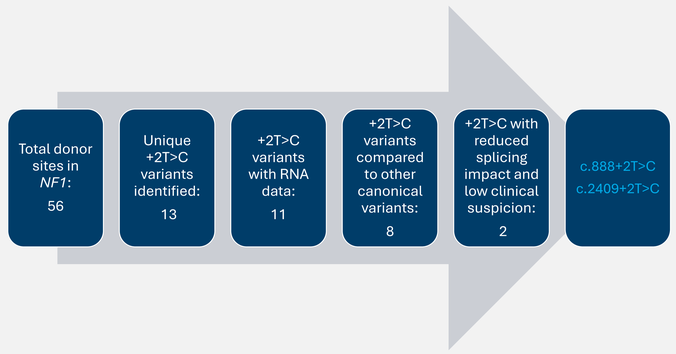


Figure 4. 8 *NF1* native donor sites met inclusion criteria. 25% (2/8) +2T>C variants had substantially lower PSI and less clinical suspicion of NF1 than matched canonical variants.

Table 1. Canonical variants in *NF1* intron 8 and 20 donor sites

NF1 variant	Lit Report?	SpliceAI	RNA analysis PSI†	gnomAD v4.1.0	ClinVar Classification
c.888+1G>A	Y	DL 0.95	14-34 (n=2)	-	5 Pathogenic
c.888+1G>C	Y	DL 1.00	-	-	4 (Likely) Pathogenic
c.888+2T>C	N	DL 0.18	0 (n=1)	-	1 Pathogenic, 1 VUS*
c.888+2T>G	Y	DL 1.00	36 (n=1)	-	4 Pathogenic
c.2409+1G>A	Y	DL 1.00; DG 0.27	36-37 (n=2)	2 alleles	7 Pathogenic
c.2409+1G>C	Y	DL 1.00; DG 0.20	-	1 allele	6 Pathogenic
c.2409+1G>T	Y	DL 1.00; DG 0.33	31-49 (n=2)	-	7 Pathogenic
c.2409+2T>C	N	DL 0.18; DG 0.28	6-15 (n=3)	4 alleles	1 Pathogenic, 1 VUS*
c.2409+2T>G	Y	DL 1.00; DG 0.13	-	-	4 Pathogenic

†r.888_889ins888+1_888+60 for c.888+1,2 and r.2326_2409del84 for c.2409+1,2. *Ambry classification

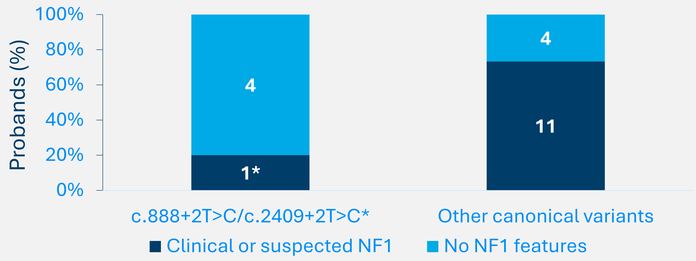


Figure 5. c.888+2T>C and c.2409+2T>C do not appear to be associated with NF1 features.

*c.2409+2T>C was detected in an infant reported to have 2 café-au-lait macules and Noonan syndrome-like dysmorphic features, but it was inherited from an unaffected parent.

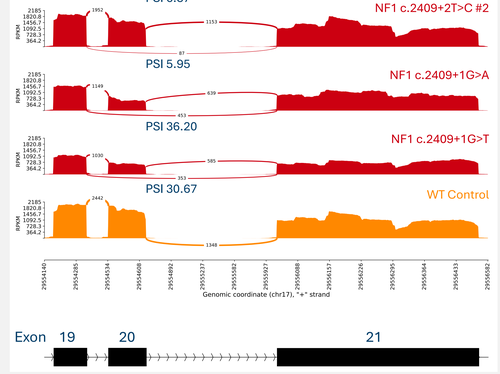


Figure 6. CaptureSeq RNA analysis showed that disruption of the donor site in intron 20 results in exon skipping (r.2326_2409del84). However, *NF1* c.2409+2T>C exhibits a substantially lower splicing impact than other canonical variants.

*Similar results observed with RT-PCRseq in another case (data not shown).

Take Home Points

- ~18% of +2T>C variants support significant normal splicing in humans
- +2T>C variants have a lower prior probability of pathogenicity than other canonical splice site variants
- +2T>C variants must be interpreted cautiously, particularly if SpliceAI donor loss scores are low

References

1. Padgett 2012. New connections between splicing and human disease. PMID 22397991
2. Moyer et al 2020. Comprehensive database and evolutionary dynamics of U12-type introns. PMID 32484558
3. Lin et al 2019. First estimate of the scale of canonical 5’ splice site GT>Gc variants capable of generating wild-type transcripts. PMID 31131953
4. Lin et al 2021. Splicing outcomes of 5’ splice site GT>GC variants that generate wild-type transcripts differ significantly between full-length and minigene splicing assays. PMID 34422003