Unlocking the Code: When SpliceAI Falls Short in Variant Assessment

Jessica Grzybowski, MS, CGC, Meghan C. Towne, MS, CGC, Brooklynn Gasser, MS, Melissa Samons, MS, CGC, Carolyn Horton, MS, CGC, Heather Zimmermann, PhD

INTRODUCTION

The clinical interpretation and classification of spliceogenic variants are challenging due to the immense complexity of splicing mechanisms. While advancements in splice prediction algorithms have improved the accuracy of in silico prediction tools like SpliceAI, relying solely on in silico predictors is inadequate, as evidenced by cases where variants with significant splicing impacts were miscategorized.

RNA analysis allows for the identification and quantification of splice impacts generated by genetic variants. To demonstrate the benefit of experimental RNA studies over reliance on splice prediction tools alone, this study describes variants with evidence of a substantial splice impact that have SpliceAI scores below the threshold widely accepted as indicative of benign splicing predictions (0.1). We present 5 distinct examples, although these examples do not represent the full extent of this phenomenon.

METHODS

We retrospectively queried results from individuals submitted for DNA and RNA clinical genetic testing for suspicion of an inherited disorder at a diagnostic laboratory (Ambry Genetics, Aliso Viejo, California). RNA, isolated from whole blood samples, underwent cDNA hybrid capture and deep sequencing (RNA CaptureSeq) and/or RT-PCR deep sequencing (RT-PCRSeq). Results were reviewed to identify variants with observed splicing impacts based on RNA transcript data that were not predicted by SpliceAI.

RESULTS

We identified 5 variants with benign SpliceAI predictions (maximum score <0.1) that had significant splice impacts observed via internal RNA studies. In the first example, a de novo variant within the consensus donor splice site of intron 6 of JAG1 (JAG1 c.886+3A>G) resulted in clear exon 6 skipping in a individual with a phenotype suggestive of Alagille syndrome, despite a SpliceAI donor loss (DL) score of 0. An intronic variant in APC located outside the standard reporting range for clinical laboratories (APC c.1744-22C>G) was demonstrated to result in coding exon 14 skipping, although the SpliceAI acceptor loss (AL) score for this variant was only 0.07. This variant has been detected in multiple probands with numerous tubular adenomatous colon polyps. NF1 c.6084G>A is synonymous variant detected in an individual who had a personal and family history consistent with features of neurofibromatosis type 1. Despite impacting the last nucleotide of exon 40 of NF1, this variant was not predicted to impact splicing by SpliceAI (DL score of 0.08). However, RNA studies revealed a clear exon skipping event. CDH1 c.1613A>T p.D538V is a mid-exonic missense change that has been detected in multiple patients with diffuse gastric cancers. Although BayesDel and SpliceAI in silico

predictions for this variant are tolerated (BayesDel –0.29 and SpliceAI 0.01 AL), the variant was classified as likely pathogenic based on phenotype data and exon 11 skipping. PMS2 c.1144G>C p.G382R is another last nucleotide change not predicted to impact splicing by SpliceAI (DL 0.02). We detected this variant in a family meeting Amsterdam criteria for Lynch syndrome, and RNA studies revealed that most transcripts derived from the variant allele lacked exon 10, with only 2% of normally spliced transcripts containing the variant allele.

CONCLUSION

While SpliceAI significantly enhances variant assessment by predicting splice effects, acknowledging its limitations is crucial. Incorporating RNA data as a functional line of evidence in real time enhances the accuracy of variant interpretation, particularly for variants that are miscategorized by in silico prediction tools. These case examples underscore the potential for misclassification without RNA functional data and emphasize the value of integrating multiple lines of evidence to ensure accurate variant classification. By leveraging the power of RNA analysis, we can refine our understanding of splicing mechanisms and improve patient care through more precise variant curation and assessment.