Investigating the use of SpliceAI in Variant Reanalysis of a Cohort of Fetuses with Congenital Brain Abnormalities

Ben McLean BA, Tam Sneddon, PhD, FACMG, Scott Melville PhD, Neeta Vora MD, FACMG, and Bradford Powell MD, PhD, FACMG

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Background

- Splice-altering variants have been proposed as an important potential class of variants to explain the "missing heritability" of cases where a genetic etiology is strongly suspected but not revealed by current analysis.
- With improvements in machine learning, the ability to identify and characterize splice site variants has improved remarkably.
- One such model powered by artificial intelligence, SpliceAI, has substantially improved accuracy compared to prior models and has started to be incorporated into research and clinical bioinformatic pipelines.
- The current study used SpliceAI to re-analyze a cohort of fetuses with brain abnormalities.

Hypotheses

1. The incorporation of SpliceAI into clinical variant analysis is expected to add complexity and cost to analysis, from additional computational cost to the possibility of identifying false-positive splicing results.
2. The use of SpliceAI will increase diagnostic yield by identifying novel pathogenic variants in a cohort of fetuses with brain abnormalities.

Methods

- This is a retrospective study examining a cohort of fetuses with congenital brain abnormalities.
- Trio sequencing (exome or genome) with DNA collected from amniocytes or chorionic villi of 91 fetuses with parental comparator data
- Inclusion criteria: fetal brain abnormality detected on prenatal imaging
- Exclusion criteria: causative finding on prenatal microarray
- Initial analysis in this study was blinded to results of prior analysis
- Baseline analysis (matching prior analysis) includes annotation and prioritization of variants using:
  - Prior publicly-reported clinical classifications (ClinVar database)
  - Population allele frequencies (gnomAD 4.0)
  - Predicted molecular consequence (SnpEff 5.2c)
  - Allelic state (e.g., de novo status) based on parental data
  - Flagging of genes with potential relevance to fetal phenotype (PanelApp Fetal Anomalies gene list 4.33)
- SpliceAI scores were calculated for identified variants, with analysis stratified by score thresholds above 0.2, 0.5, and 0.8
- Descriptive statistics and graphs were generated in R Statistical software v4.1.2

Results of Initial Analysis

- Among 91 trios, 490 fetal variants met initial filtering criteria (population allele frequency <= 0.01, gene presence on PanelApp Fetal Anomalies list, SpliceAI score >= 0.2).
- 72 variants could not be excluded through gene-specific population allele frequency filters or by meeting one or more of the ACMG/AMP criteria for benign evidence: BS1, BA1, BS2.
- 5 were determined to be possibly disease-causing and presented to multidisciplinary team for further discussion with potential to return to families as a clinical or research report (Table 1).

HGVS Transcript	Gene	Variant Type	Already Presented?
NM_000284.4:c.586G>T (p.Asp196Tyr)	*PDHA1*	Missense	Yes
NM_000284.4:c.604-2A>G	*PDHA1*	Canonical Splice Site	Yes
NM_013382.7:c.1006+5G>A	*POMT2*	Intronic	Yes
NM_013382.7:c.1329_1332+5del	*POMT2*	Intronic	Yes
NM_002291.3:c.4280_4281dup (p.Pro1428GlyfsTer26)	*LAMB1*	Frameshift	No

Table 1. Potentially pathogenic variants with a SpliceAI score of at least 0.2 found on re-analysis.

Burden of Analysis

- We assessed a key component of the analytic burden of incorporating SpliceAI in variant analysis: additional variant data that needs to be considered when manually assessing variant classification.
- 4,476 variants met criteria for SpliceAI score >= 0.2 and population allele frequency <= 0.01.
- Variants were stratified by SpliceAI threshold and by presence on PanelApp Fetal Anomalies lists with different confidence (Table 2)
- Variants aggregated by prior reports in ClinVar are represented in Figure 1
- Of the variants with a SpliceAI score of just 0.2 or greater, 950/1312 (72%) had a benign or likely benign classification.

SpliceAI score:	Number of variants with a match for a gene on the PanelApp Fetal Anomalies list	Number of variants with a match for a gene on the PanelApp Any list	Total number of variants with an allele frequency less than 0.01
0.2 or greater	490	1825	4476
0.5 or greater	159	467	1291
0.8 or greater	52	192	582

Table 2. Characterizing the burden of analysis by different SpliceAI scores.

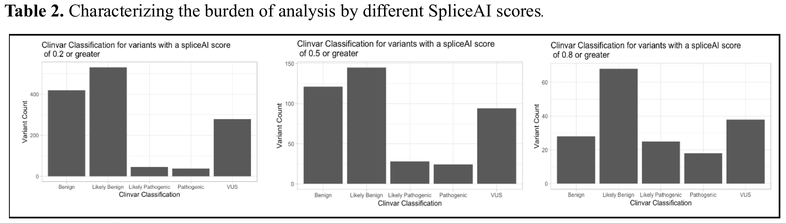

Figure 1. *Total number of variants with a SpliceAI score of 0.2, 0.5, or 0.8 or greater by ClinVar classification.*

Burden of Analysis

- SnpEff classifies variants by direct predicted sequence effect (e.g., nonsense, missense, presence in canonical splice site)
  - HIGH: expected to result in truncation or loss of expression
  - MODERATE: expected to change amino acid sequence, but retain expression
  - LOW/MODIFIER: synonymous or no clear effect
- A high proportion of variants classified as LOW/MODERATE by SnpEff have a SpliceAI score of at least 0.2 (2718/3815; 71%).

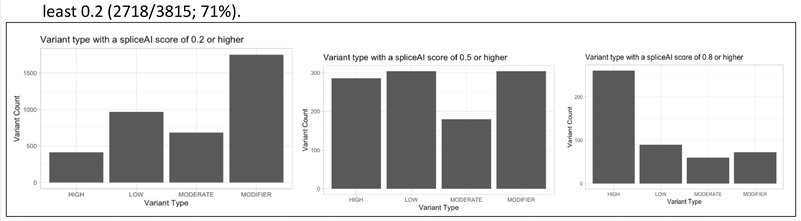

Figure 2. *Total number of variants with a SpliceAI score of 0.2, 0.5, or 0.8 or greater by predicted impact on protein function.*

- Out of a total of 1,850 intronic variants with a spliceAI score of at least 0.2, 82% (1,518) of them were located outside of the canonical splice site (Figure 3).
- Finally, 214 variants located within canonical splice sites had a SpliceAI donor or acceptor gain score of 0.2 or greater. Of the 214 variants, 15 of them had an alternative splicing site located 3, 6, 9, 12, or 15 base pairs away from the canonical splicing site.

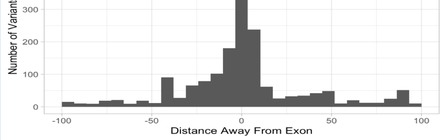

Figure 3. *Histogram showing the number of variants with a spliceAI score of 0.2 or greater according to how far each variant was located away from the exon-intron boundary. Negative values represents the distance before the exon, positive values represents the distance after the exon.*

Conclusions

- The results of our study suggest that SpliceAI adds little additional sensitivity in variant re-analysis.
- These results are particularly surprising given the higher pre-test probability of genetic disease in our fetal cohort compared to pediatric or adult cohorts.
- The added burden of systematically incorporating this tool into molecular analysis was not insignificant. In total, 490 variants were examined with our filtering scheme. This suggests that SpliceAI is most useful when applied narrowly to variant analysis, such as the interpretation of variants near canonical splicing sites, investigating cryptic splice sites, or when assessing the impact of de novo variants.
- Acknowledgements: Initial participant recruitment was performed with support from NIH (5R01HD105868, PI: Neeta Vora). The GENYSIS core facility (Director: Tam Sneddon) provided assistance for variant classification).

References

1. Ars, E., Serra, E., García, J., Kruyer, H., Gaona, A., Lázaro, C., & Estivill, X. (2000). Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. Human molecular genetics, 9(2), 237–247. https://doi.org/10.1093/hmg/9.2.237
2. Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Hum Genet*. 2017;136(9):1093-1111. doi:10.1007/s00439-017-1809-4
3. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019;176(3):535-548.e24. doi:10.1016/j.cell.2018.12.015
4. Zeng T, Li YI. Predicting RNA splicing from DNA sequence using Pangolin. Genome Biology. 2022 Apr 21;23(1):103.
5. Salfati EL, Spencer EG, Topol SE, et al. Re-analysis of whole-exome sequencing data uncovers novel diagnostic variants and improves molecular diagnostic yields for sudden death and idiopathic diseases. Genome Med. 2019;11(1):83. Published 2019 Dec 17. doi:10.1186/s13073-019-0702-2
6. Vora NL, Gilmore K, Brandt A, et al. An approach to integrating exome sequencing for fetal structural anomalies into clinical practice [published correction appears in Genet Med. 2020 Aug;22(8):1426]. Genet Med. 2020;22(5):954-961. doi:10.1038/s41436-020-0750-4
7. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
8. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in medicine : official journal of the American College of Medical Genetics, 17(5), 405–424. https://doi.org/10.1038/gim.2015.30
9. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3.", Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. Fly (Austin). 2012 Apr-Jun;6(2):80-92. PMID: 22728672
10. Walker, L. C., Hoya, M., Wiggins, G. A. R., Lindy, A., Vincent, L. M., Parsons, M. T., Canson, D. M., Bis-Brewer, D., Cass, A., Tchourbanov, A., Zimmermann, H., Byrne, A. B., Pesaran, T., Karam, R., Harrison, S. M., Spurdle, A. B., & ClinGen Sequence Variant Interpretation Working Group (2023). Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: Recommendations from the ClinGen SVI Splicing Subgroup. American journal of human genetics, 110(7), 1046–1067. https://doi.org/10.1016/j.ajhg.2023.06.002
11. Martin, A.R., Williams, E., Foulger, R.E. et al. PanelApp crowdsources expert knowledge to establish consensus diagnostic gene panels. Nat Genet 51, 1560–1565 (2019). https://doi.org/10.1038/s41588-019-0528-2