Check for updates

BRIEF COMMUNICATION   OPEN

Estimating the proportion of nonsense variants undergoing the newly described phenomenon of manufactured splice rescue

Bushra Haque[1,2], David Cheerie[1,2], Saba Birkadze[1,2], Alice Linyan Xu [2], Thomas Nalpathamkalam[3], Bhooma Thiruvahindrapuram [3], Susan Walker[3] and Gregory Costain[1,2,4 ✉]

© The Author(s) 2023

A recent report described a nonsense variant simultaneously creating a donor splice site, resulting in a truncated but functional protein. To explore the generalizability of this unique mechanism, we annotated >115,000 nonsense variants using SpliceAI. Between 0.61% (donor gain delta score >0.8, for high precision) and 2.57% (>0.2, for high sensitivity) of nonsense variants were predicted to create new donor splice sites at or upstream of the stop codon. These variants were less likely than other nonsense variants in the same genes to be classified as pathogenic/likely pathogenic in ClinVar ($p < 0.001$). Up to 1 in 175 nonsense variants were predicted to result in small in-frame deletions and loss-of-function evasion through this "manufactured splice rescue" mechanism. We urge caution when interpreting nonsense variants where manufactured splice rescue is a strong possibility and correlation with phenotype is challenging, as will often be the case with secondary findings and newborn genomic screening programs.

INTRODUCTION

Stop-gain (nonsense) variants are typically assumed to result in loss-of-function, and assigned "very strong" evidence in favour of pathogenicity [1, 2]. A recent report described a nonsense variant in *BUD13* [NM_032725.4:c.688C>T; p.(Arg230*)] that simultaneously activated a new cryptic donor splice site in the same canonical isoform [3]. Surprisingly, the alternative splice product resulted in a truncated but functional protein product, and converted a loss-of-function into a hypomorphic allele [3]. Intrafamilial phenotypic severity of the associated progressive multisystem disease was correlated with the expression level of the truncated protein [3]. This molecular mechanism, which we will term "manufactured splice rescue", is distinct from nonsense-associated altered splicing (NAS) [4–6], and is not acknowledged in variant interpretation guidelines [1, 2, 7]. The nucleotide triplets TAA and TGA are both stop codons and highly conserved components of canonical splice sites (+2 to +4 positions; Fig. 1), meaning that these codons may be susceptible to cryptic splicing effects. The prevalence of nonsense variants potentially triggering manufactured splice rescue is unknown. We describe the predicted splicing effects of >115,000 single nucleotide nonsense variants, finding that ~1 in 40 variants (2.57%) potentially create new donor splice sites and that ~1 in 175 variants might result in small in-frame deletions rather than a definite loss-of-function.

METHODS

To investigate the generalizability of this "manufactured splice rescue" phenomenon, we used advanced in silico methods and large datasets. We extracted single nucleotide nonsense variants from three variant databases: gnomAD (v3.1.2 and v2.1.1) [8], ClinVar (download date: August 29, 2022) [9], and MSSNG, the largest genome sequencing database for autism with deep phenotyping (latest release: October 16, 2019) [10]. We restricted to canonical transcripts of protein-coding genes, and excluded nonsense variants in the last exon, as these would already be treated cautiously in their interpretation [1, 2]. The remaining 115,171 unique variants (gnomAD: $n = 84,891$; ClinVar: $n = 33,517$; MSSNG: $n = 5904$) were then annotated with SpliceAI using Ensembl Variant Effect Predictor and/or a custom script developed at The Centre for Applied Genomics (TCAG) [11, 12]. We used author-recommended cutoffs for SpliceAI donor gain (DG) delta scores: ≥0.2 (high recall), ≥0.5, and ≥0.8 (high precision) [11]. Recognizing that predicted splicing changes downstream to the variant stop codon would not prevent nonsense mediated decay (NMD), we considered only those variants with DG scores meeting pre-set cutoffs that also had (strand-corrected) pre-mRNA positions/delta positions [11] <3 as potentially resulting in manufactured splice rescue. We used Alamut Visual Plus (v1.7, © 2022 SOPHiA) to inspect the predicted splicing impact of a subset of variants using additional in silico tools [13]. Whether a partial exon deletion resulting from mis-splicing would be in-frame or out-of-frame was based on the difference between the DG position and the exon end position (determined using ExonCalculator; github.com/haqueb2/ExonCalculator). We considered in-frame deletions of less than 10% of the coding transcript to be those potentially resulting in loss-of-function evasion [7]. Protein domains were annotated using InterPro domains [14] with ANNOVAR. Statistical analyses, including Chi-squared, Mann–Whitney U, and Wilcoxon Rank-Sum tests, were performed using R statistical software, version 4.1.0 (R Foundation for Statistical Computing) with two-tailed statistical significance set at $p < 0.05$.

RESULTS

Across the 115,171 unique variants, 2.57% had DG scores ≥0.2 at DPs <3 and 0.61% had DG scores ≥0.8 at DPs <3 (Fig. 2A). Findings

[1]Program in Genetics & Genome Biology, SickKids Research Institute, Toronto, ON, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada. [3]The Centre for Applied Genomics (TCAG), SickKids Research Institute, Toronto, ON, Canada. [4]Division of Clinical & Metabolic Genetics, The Hospital for Sick Children (SickKids), and Department of Paediatrics, University of Toronto, Toronto, ON, Canada. ✉email: gregory.costain@sickkids.ca
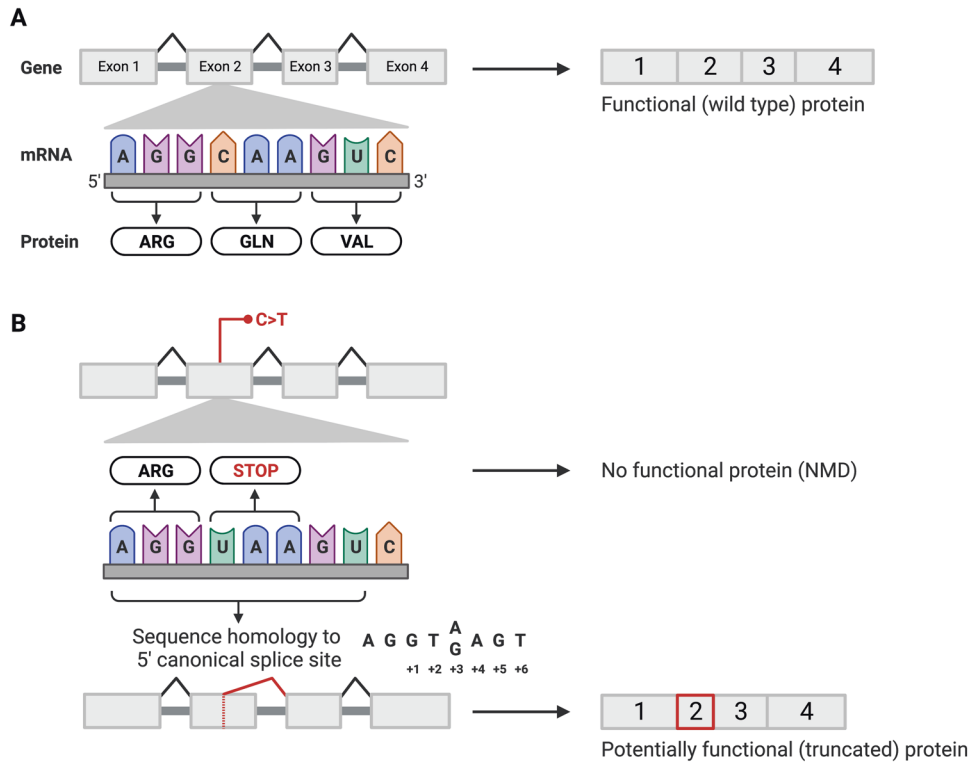
Fig. 1 **Diagram of proposed mechanism by which a nonsense variant could result in aberrant splicing and thus a potentially functional protein product. A** Grossly simplified depiction of the "normal" splicing of a 4-exon protein coding gene. **B** Example of a sequence variant that could simultaneously result in a stop-gain and in activation of a cryptic 5' (donor) splice site. If use of the latter results in a small in-frame deletion, there may be a truncated but functional protein product. In the example shown, the pre-mRNA position of the new splice site would be 2 nucleotides upstream of the variant (i.e., delta position = −2). Created with BioRender.com.

were similar across the three datasets (Fig. 2A). As expected (Fig. 1), nonsense variants with DG scores ≥0.2 at DPs <3 were significantly more likely to be TAA or TGA stop codons (63.1%) than the remaining nonsense variants in the overall dataset (56.9%; chi-square = 59.5, $p < 0.00001$). The proportion of nonsense variants that were TAA or TGA stop codons increased to 72.1% when restricting to the subset meeting the high precision threshold of DG scores ≥0.8. Also as expected (Fig. 1), the predicted new donor splice sites clustered at the -2 pre-mRNA position (Supplemental Fig. 1).

We then investigated whether this molecular mechanism could explain some instances of apparent incomplete penetrance and highly variable expression. Restricting to the ClinVar dataset, nonsense variants potentially triggering a manufactured splice rescue were significantly less likely than the remaining nonsense variants in the same genes to have a likely pathogenic or pathogenic (LP/P) classification (Fig. 2B). There were also differences in the confidence level (star rating) of those LP/P classifications (Fig. 2C), including a lower mean star rating in nonsense variants potentially triggering a manufactured splice rescue compared with the remaining nonsense variants in the same genes (SpliceAI ≥ 0.2: 1.10 vs. 1.34, respectively, W = 4944247, $p < 0.00001$; SpliceAI ≥ 0.8: 1.11 vs. 1.36, W = 584088, $p < 0.0001$).

Considering the subset of nonsense variants that met our SpliceAI cut-offs of DG delta score ≥0.2 at DPs <3 ($n = 2863$), and assuming partial exon deletion as a result of using the newly created donor splice site (Fig. 1), we predicted that 662 nonsense variants (23.1% of 2863, or ~1 in 175 of all 115,171 nonsense variants) would result in in-frame deletions accounting for <10% of the coding transcript (Fig. 3). There was a non-significant trend towards nonsense variants predicted to result in in-frame deletions being less likely than nonsense variants predicted to result in out-of-frame deletions to be classified as LP/P variants in

ClinVar (87.6% vs. 90.7%, $p = 0.78$). A proportion of the variants also impacted protein domains (Fig. 3), however whether small in-frame deletions in these domains would disrupt overall protein function could not be determined.

For example, a nonsense variant in *TSC2* (NM_000548.5: c.4081C > T) was reported in ClinVar (SCV000819981.3) as a variant of uncertain significance after it was identified in an individual without features of tuberous sclerosis complex. This variant's SpliceAI DG score is 0.80, and additional in silico tools also predict the creation of a donor splice site 2 bp upstream to the variant position in the pre-mRNA (Supplemental Fig. 2). In silico analysis of the variant suggests the outcome may be an in-frame deletion [GRCh38(Chr16):g.2084302_2084950del; p.(Glu1360_Ser1498delinsAsp)] that removes <10% of the total protein length and does not impact key functional protein domains.

DISCUSSION

Secondary sequence properties can alter the predicted impacts of variants [15]. However, consideration of "manufactured splice rescue" (in contrast to other mechanisms, like "naturally occurring candidate rescue transcripts" [7]) is not yet codified in variant classification criteria for nonsense variants [1, 2, 7]. We found only rare instances of it being acknowledged by clinical genetic testing laboratories during variant review (e.g., ClinVar Accession: SCV002216056.2). Inspired by a recent case report [3], we found evidence that this molecular mechanism could apply to a small but meaningful proportion of all nonsense variants.

Our preliminary study has several limitations. In silico prediction scores are imperfect [11, 16]. We did not confirm the splicing effect of specific nonsense variants in individuals by RNA sequencing [3, 13, 17] or other functional assays [18]. The creation of a splice site upstream of the nonsense variant might still result
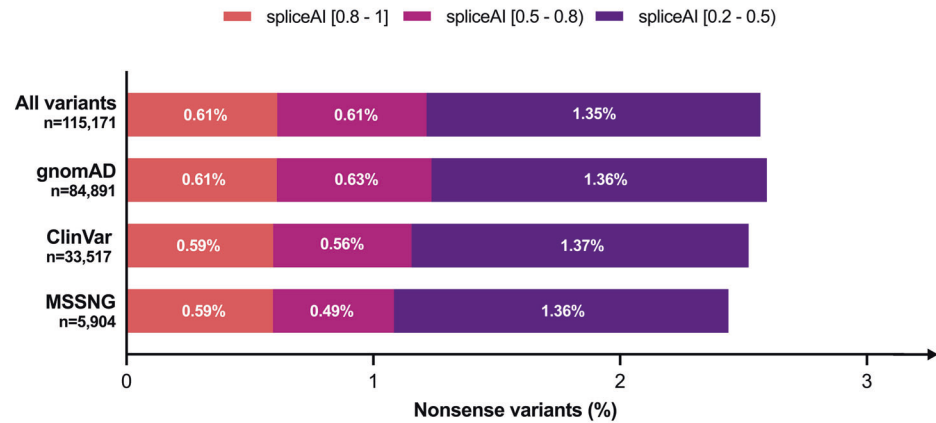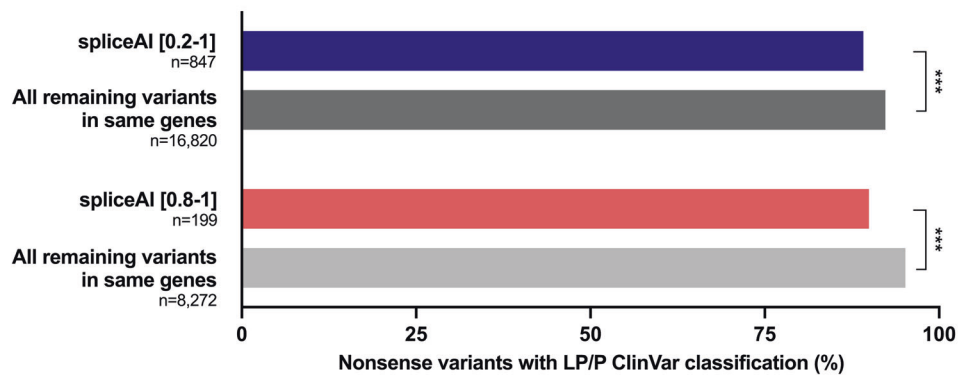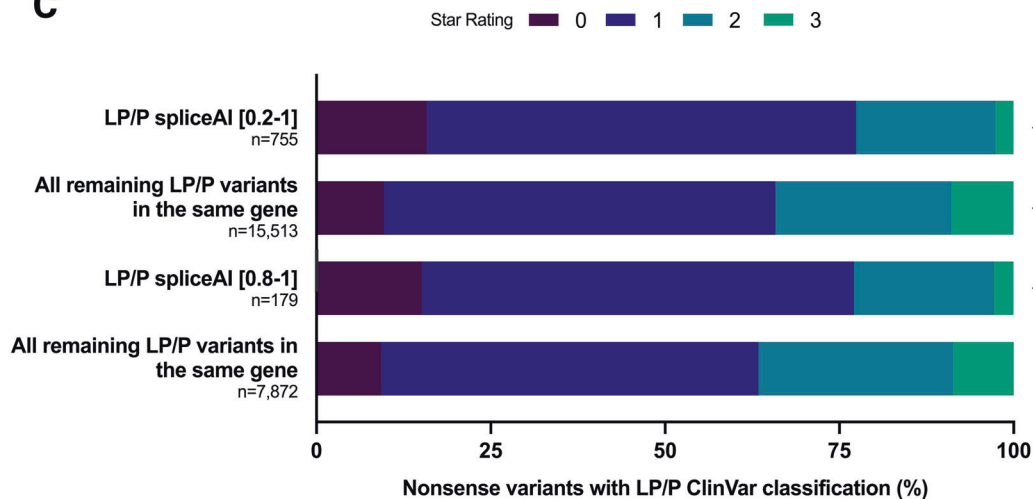
Fig. 2 A consistent proportion of nonsense variants across large-scale databases may create new donor splice sites. A Stacked bar chart with percentage of nonsense variants predicted to create donor gain sites using SpliceAI. Nonsense variants from three variant databases (gnomAD [8], ClinVar [9], MSSNG [10]) were annotated with DG SpliceAI scores and categorized into three score categories: [0.8–1], [0.5–0.8), and [0.2–0.5) (see Methods for additional details). **B** Bar chart including percentages of likely pathogenic/pathogenic (LP/P) nonsense variants in ClinVar with SpliceAI scores ≥0.2 or ≥0.8, compared to all other variants in the same genes. **C** Stacked bar chart with percentages of ClinVar "star ratings" for LP/P variants with SpliceAI scores ≥0.2 or ≥0.8, compared to all other variants in the same genes. Wilcoxon Rank-Sum test was used to evaluate statistical differences between the two groups. **p < 0.01, ***p < 0.001, ****p < 0.0001. Created with GraphPad Prism.

in a loss-of-function (e.g., from an indel that results in a frameshift). The predicted impact of an in-frame deletion within a protein domain on protein function is best determined on a gene-by-gene basis through manual curation of the literature and/or experimental (in vivo or ex vivo) approaches, and was beyond the scope

of this report. Conversely, nonsense variants may be rescued by different mechanisms unrelated to manufactured splice rescue [6, 7, 15, 19, 20]. Lastly, while we explored three different datasets (ClinVar, gnomAD, and MSSNG) to offset the ascertainment biases inherent in each and noted similar expected rates of
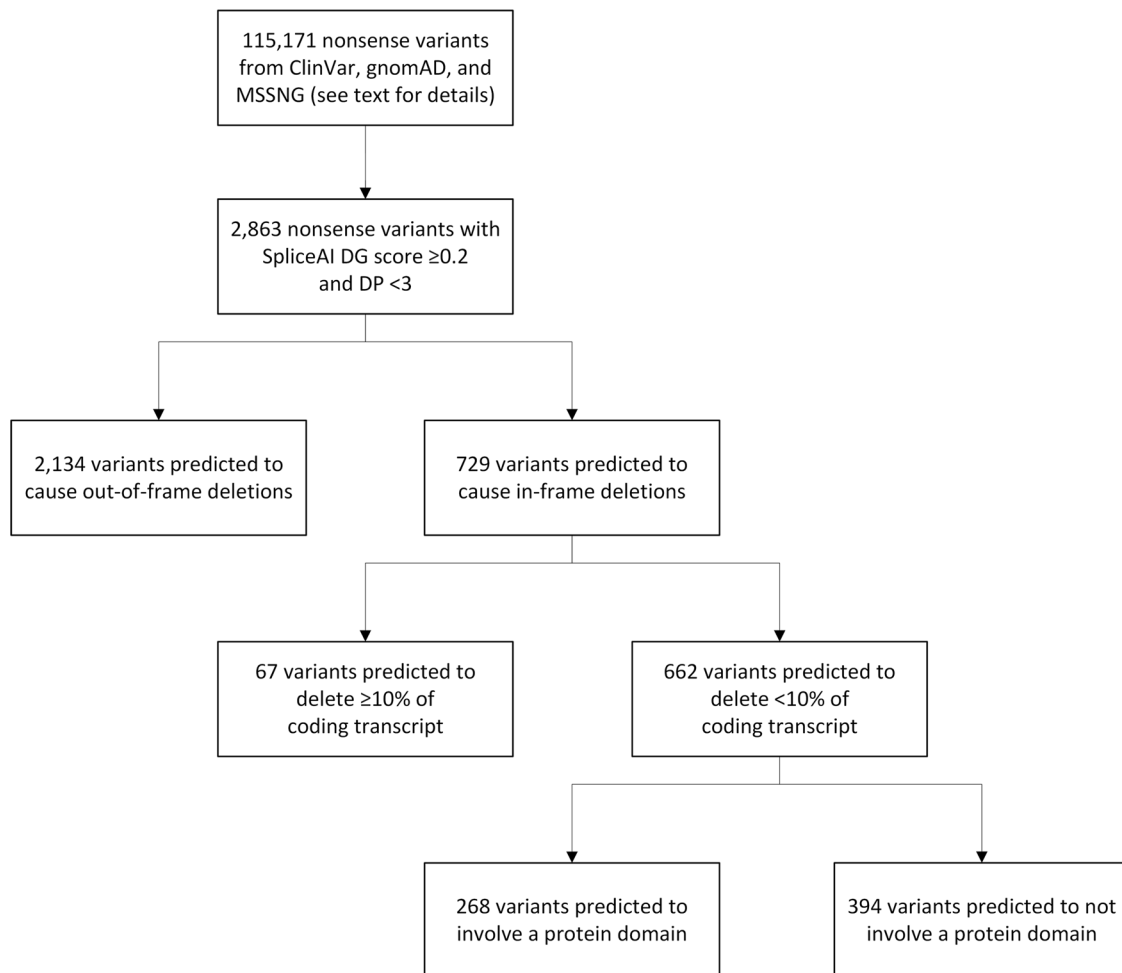
Fig. 3   In silico predictions of protein-level consequences indicate that a proportion of all nonsense variants may evade loss-of-function through "manufactured splice rescue". See text for details. With the assumption that creation and use of a new donor splice site within an exon will lead to deletion of the downstream component of that exon (Fig. 1), variants were categorized as causing in-frame deletions or out-of-frame deletions, then as deleting <10% or ≥10% of the coding transcript, and then as to whether the deletion did not or did involve a protein domain. DG donor gain, DP delta position.

manufactured splice rescue, none provides an unbiased sampling of germline human nonsense variants. The true prevalence in the genome of this phenomenon of manufactured splice rescue remains unknown.

In summary, we have assessed an underappreciated mechanism whereby unchallenged assumptions regarding variant impact could result in inaccurate variant interpretation. There is growing awareness that in silico tools like SpliceAI are invaluable for identifying deleterious cryptic splice variants within classes of variation often presumed to be benign (e.g., synonymous variants, deep intronic variants) [16], but the inverse scenario is rarely considered. We recommend against initially applying PVS1-level evidence to novel nonsense variants where manufactured splice rescue is a strong possibility and correlation with phenotype is challenging, as will often be the case with secondary findings and in the anticipated future wave of newborn genomic screening programs.

DATA AVAILABILITY
The datasets analysed during the current study are available in the ClinVar [ncbi.nlm.nih.gov/clinvar/], gnomAD [gnomad.broadinstitute.org/], and MSSNG [research.mss.ng/] repositories.

REFERENCES
1. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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. Genet Med. 2015;17:405–24.
2. Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. Hum Mutat. 2018;39:1517–24.
3. Kornak U, Saha N, Keren B, Neumann A, Taylor Tavares AL, Piard J, et al. Alternative splicing of BUD13 determines the severity of a developmental disorder with lipodystrophy and progeroid features. Genet Med. 2022;24:1927–40.
4. Hull J, Shackleton S, Harris A. The stop mutation R553X in the CFTR gene results in exon skipping. Genomics. 1994;19:362–4.
5. Aznarez I, Zielenski J, Rommens JM, Blencowe BJ, Tsui LC. Exon skipping through the creation of a putative exonic splicing silencer as a consequence of the cystic fibrosis mutation R553X. J Med Genet. 2007;44:341–6.
6. Sofronova V, Fukushima Y, Masuno M, Naka M, Nagata M, Ishihara Y, et al. A novel nonsense variant in ARID1B causing simultaneous RNA decay and exon skipping is associated with Coffin-Siris syndrome. Hum Genome Var. 2022;9:26.
7. Walker LC, Hoya M, Wiggins GAR, Lindy A, Vincent LM, Parsons MT, et al. Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: recommendations from the ClinGen SVI splicing subgroup. Am J Hum Genet. 2023;110:1046–67.
8. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.

9. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46:D1062–D1067.

10. Trost B, Thiruvahindrapuram B, Chan AJS, Engchuan W, Higginbotham EJ, Howe JL, et al. Genomic architecture of autism from comprehensive whole-genome sequence annotation. Cell 2022;185:4409–27.e4418.

11. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176:535–.e524.

12. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The ensembl variant effect predictor. Genome Biol. 2016;17:122.

13. Walker S, Lamoureux S, Khan T, Joynt ACM, Bradley M, Branson HM, et al. Genome sequencing for detection of pathogenic deep intronic variation: a clinical case report illustrating opportunities and challenges. Am J Med Genet A. 2021;185:3129–35.

14. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2023;51:D418–D427.

15. Singer-Berk M, Gudmundsson S, Baxter S, Seaby EG, England E, Wood JC, et al. Advanced variant classification framework reduces the false positive rate of predicted loss-of-function variants in population sequencing data. Am J Hum Genet. 2023;110:1496–508.

16. Ellingford JM, Ahn JW, Bagnall RD, Baralle D, Barton S, Campbell C, et al. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. Genome Med. 2022;14:73.

17. Deshwar AR, Yuki KE, Hou H, Liang Y, Khan T, Celik A, et al. Trio RNA sequencing in a cohort of medically complex children. Am J Hum Genet. 2023;110:895–900.

18. Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol Biol. 2010;653:249–57.

19. Teraoka SN, Telatar M, Becker-Catania S, Liang T, Onengut S, Tolun A, et al. Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. Am J Hum Genet. 1999;64:1617–31.

20. Dupont MA, Humbert C, Huber C, Siour Q, Guerrera IC, Jung V, et al. Human IFT52 mutations uncover a novel role for the protein in microtubule dynamics and centrosome cohesion. Hum Mol Genet. 2019;28:2720–37.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

S.W. and G.C. designed the study. B.H., D.C., S.B., A.L.X., T.N., and B.T. acquired the data. B.H., D.C., S.W., and G.C. analyzed and interpreted the data. B.H., D.C., and G.C. drafted the manuscript, and S.B., A.L.X., T.N., B.T., and S.W. were given the opportunity to revise it critically for important intellectual content. All authors give final approval of the submitted version and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

COMPETING INTERESTS

Dr. Walker is currently an employee of Genomics England Limited. The other authors declare no competing interests.

ETHICS APPROVAL

Informed consent was not required as this study only analyzed de-identified genetic variant data from large-scale databases.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41431-023-01495-6.

Correspondence and requests for materials should be addressed to Gregory Costain.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.