HGG Advances

ARTICLE

Investigation of cryptic *JAG1* splice variants as a cause of Alagille syndrome and performance evaluation of splice predictor tools

Ernest Keefer-Jacques,[1,5] Nicolette Valente,[1,5] Anastasia M. Jacko,[1] Grace Matwijec,[1] Apsara Reese,[1] Aarna Tekriwal,[1] Kathleen M. Loomes,[2,3] Nancy B. Spinner,[1,4] and Melissa A. Gilbert[1,2,4,6,*]

Summary

Haploinsufficiency of *JAG1* is the primary cause of Alagille syndrome (ALGS), a rare, multisystem disorder. The identification of *JAG1* intronic variants outside of the canonical splice region as well as missense variants, both of which lead to uncertain associations with disease, confuses diagnostics. Strategies to determine whether these variants affect splicing include the study of patient RNA or minigene constructs, which are not always available or can be laborious to design, as well as the utilization of computational splice prediction tools. These tools, including SpliceAI and Pangolin, use algorithms to calculate the probability that a variant results in a splice alteration, expressed as a Δ score, with higher Δ scores (>0.2 on a 0–1 scale) positively correlated with aberrant splicing. We studied the consequence of 10 putative splice variants in ALGS patient samples through RNA analysis and compared this to SpliceAI and Pangolin predictions. We identified eight variants with aberrant splicing, seven of which had not been previously validated. Combining these data with non-canonical and missense splice variants reported in the literature, we identified a predictive threshold for SpliceAI and Pangolin with high sensitivity (Δ score >0.6). Moreover, we showed reduced specificity for variants with low Δ scores (<0.2), highlighting a limitation of these tools that results in the misidentification of true splice variants. These results improve genomic diagnostics for ALGS by confirming splice effects for seven variants and suggest that the integration of splice prediction tools with RNA analysis is important to ensure accurate clinical variant classifications.

Introduction

Precise splicing of mRNA is critical to generate mature, protein-coding cDNA, and the disruption of this process can lead to disease. RNA splicing is tightly regulated, with canonical consensus sequences within the introns providing docking instructions for the spliceosome, a large ribonucleoprotein complex that excises introns from pre-mRNA transcripts and joins exons.[1] Consensus sequences typically consist of a donor site, located at the 5′ end of the intron with the invariant nucleotide sequence GU at positions +1 and +2, and an acceptor site at the 3′ end of the intron with the invariant sequence AG at positions −2 and −1, along with a lariat branchpoint site deeper in the intron that may be more difficult to identify. Although these canonical nucleotides influence where splicing occurs, sequences both near and far from exon-intron boundaries can direct spliceosome docking and catalysis,[2] which can lead to altered splicing and result in disease.

Utilization of *in silico* splice prediction tools can be helpful in identifying putative splice variants. SpliceAI and Pangolin use deep learning-based algorithms to predict the consequences of sequence variation at the whole-transcript level, including exon skipping, cryptic splice site activation, and pseudo-exon creation.[3,4] The probability that a variant will result in a splice alteration is expressed as a Δ score (0–1 scale) for both tools, with higher Δ scores positively correlated with aberrant splicing. At a threshold of 0.2, SpliceAI performs with a predictive sensitivity of ∼90%, and therefore, this has been routinely used as a minimum Δ score cutoff, with a similar sensitivity reported for corresponding Pangolin Δ scores.[3–5] Although the sensitivity and specificity of these tools represent an improvement over earlier splice predictors,[6] analysis of RNA is necessary to confirm aberrant splicing.

Splicing variants in the disease gene Jagged1 (*JAG1* [MIM: 601920]) have been well documented to cause Alagille syndrome (ALGS [MIM: 118450]), a pediatric cause of liver disease, among other multisystemic features.[7,8] Nearly 100 splice variants have been described for *JAG1*, accounting for ∼12% of all *JAG1* variants.[7,9] Although many of these variants affect the canonical splice donor and splice acceptor sites, occasionally, variants have been reported in non-canonical intronic positions ($n = 13$ out of 62 [21%] splice variants reported in the Human Gene Mutation Database [HGMD]).[7,10–15] Moreover, there have been reports of nucleotide substitutions within the coding region that result in aberrant splicing through the

[1]Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [2]Division of Pediatric Gastroenterology, Hepatology, and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [3]Department of Pediatrics, The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA; [4]Department of Pathology and Laboratory Medicine, The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA
[5]These authors contributed equally
[6]Lead contact
*Correspondence: gilbertma@chop.edu
https://doi.org/10.1016/j.xhgg.2024.100351.

generation of a cryptic splice site, rather than missense variation, as initially predicted ($n = 4$ reported in HGMD).[7,15–18] Loss of function, most often occurring through protein truncation, is the known pathomechanism of ALGS; therefore, confirmation of aberrant splicing from non-canonical intronic variants or through the generation of novel splice acceptor or donor sites by variation within the coding region is immediately relevant to diagnostics. We sought to test the predictive ability of two deep learning-based splice tools, SpliceAI[4] and Pangolin,[3] in an ALGS cohort of individuals with either *JAG1* missense or non-canonical intronic variants predicted to affect splicing. Using RNA samples, we were able to test the splice tools' predictions, improve diagnostics, and provide guidance for clinical interpretation for putative *JAG1* splice variants.

Subjects, material, and methods

ALGS cohort

Individuals were ascertained through an institutional review board-approved ALGS research study at the Children's Hospital of Philadelphia (CHOP). We identified 479 probands and affected family members who were enrolled between May 1992 and June 2022 with *JAG1* variants (Figure 1). This cohort was filtered to include all participants with missense variants or variants in intronic regions captured with our Sanger sequencing primers, which contain on average ~100 bp of intronic sequence (Table S1). Intronic variants that alter the ±1 or 2 position (canonical splice site) were removed from the final cohort. A cohort of 90 individuals was reviewed for redundant variants, with 50 unique putative cryptic splice variants identified ($n = 43$ missense and $n = 7$ non-canonical intronic variants). Splice predictor filters (described below), including SpliceAI,[4] Pangolin,[3] and Alamut (SOPHiA Genetics, Lausanne, Switzerland) were applied to identify variants predicted to affect splicing ($n = 14$ out of 50). Biospecimens were available for 10 variants ($n = 5$ missense and $n = 5$ non-canonical intronic variants) (Figure 1).

Splice predictor filters

A total of 50 variants ($n = 43$ missense and $n = 7$ non-canonical intronic variants) were analyzed using SpliceAI,[4] Pangolin,[3] and Alamut (SOPHiA Genetics) to predict their effects on splicing. Variants were included for analysis if Alamut, Pangolin, or SpliceAI predicted a possible effect at the nearest splice site. Pangolin and SpliceAI predictions were provided as a Δ score, with higher Δ scores indicating a greater probability of a splicing effect.[19] Variants were retained if a Δ score for splice loss or gain was greater than 0.2 for either Pangolin or SpliceAI, as previously recommended.[19] Alamut integrates four splice prediction algorithms into their software, including SpliceSiteFinder-like,[20] MaxEntScan,[21] GeneSplicer,[22] and NNSPLICE.[23] Any variant annotated by Alamut to produce a possible effect at the nearest splice site was included for study, regardless of the strength of their score by each of the four algorithms. Ten individuals harboring 10 unique variants (missense, $n = 5$; non-canonical splice site, $n = 5$) with predicted effects on splicing had available lymphoblastoid cell lines, allowing for RNA analysis (Table S2).

Cell culture

Lymphoblastoid cell lines were cultured in RPMI 1640 media containing 20% fetal bovine serum, 1% penicillin-streptomycin, and 2% L-glutamine. Cycloheximide treatment was performed as described.[24] Briefly, cells were treated with 1:1,000 Ready-Made cycloheximide solution (stock 100 mg/mL, catalog no. C4859; Sigma-Aldrich, St. Louis, MO) to inhibit nonsense-mediated decay or with 1:1,000 DMSO (Sigma-Aldrich) as a control for 6 h. Cells were harvested by centrifugation at 1,500 rpm for 5 min and washed once with PBS before proceeding to RNA extraction.

DNA and RNA analysis

Genomic DNA (gDNA) was extracted from untreated cells using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), followed by amplification using AmpliTaq Gold (Applied Biosystems, Waltham, MA) and listed primers (Table S1). Clean-up of amplified gDNA was accomplished using ExoSAP-IT Express Reagent (Applied Biosystems) before sequencing.

RNA was extracted using the RNEasy Kit (Qiagen) and cDNA was transcribed using TaqMan Reverse Transcription Reagents (Invitrogen, Waltham, MA), followed by amplification using previously described primers[25] and primers designed using Primer3Plus (Thermo Fisher Scientific, Waltham, MA) (Table S3). Amplified cDNA was visualized on a 2% agarose gel. Bands were excised and purified using the Nucleospin Gel and PCR Purification Kit (Machery-Nagel, Duren, Switzerland) before sequencing. Sequence alignment and analysis were performed using Mutation Surveyor (SoftGenetics, State College, PA), SnapGene (Dotmatics, Boston, MA), and Alamut (SOPHiA Genetics).

Literature review

A literature search on PubMed for splice variants in *JAG1* was performed with a last search date of April 26, 2024. Queries included the search terms "jag1," "splice site," "cryptic splice," and "non-canonical splice."

Results

Within a cohort of 479 individuals with a *JAG1* variant within our ALGS research study at CHOP, we identified 14 variants predicted to affect splicing (Table S4), the biospecimens of which were available for 10 participants (Figure 1). These included 5/7 missense variants that were

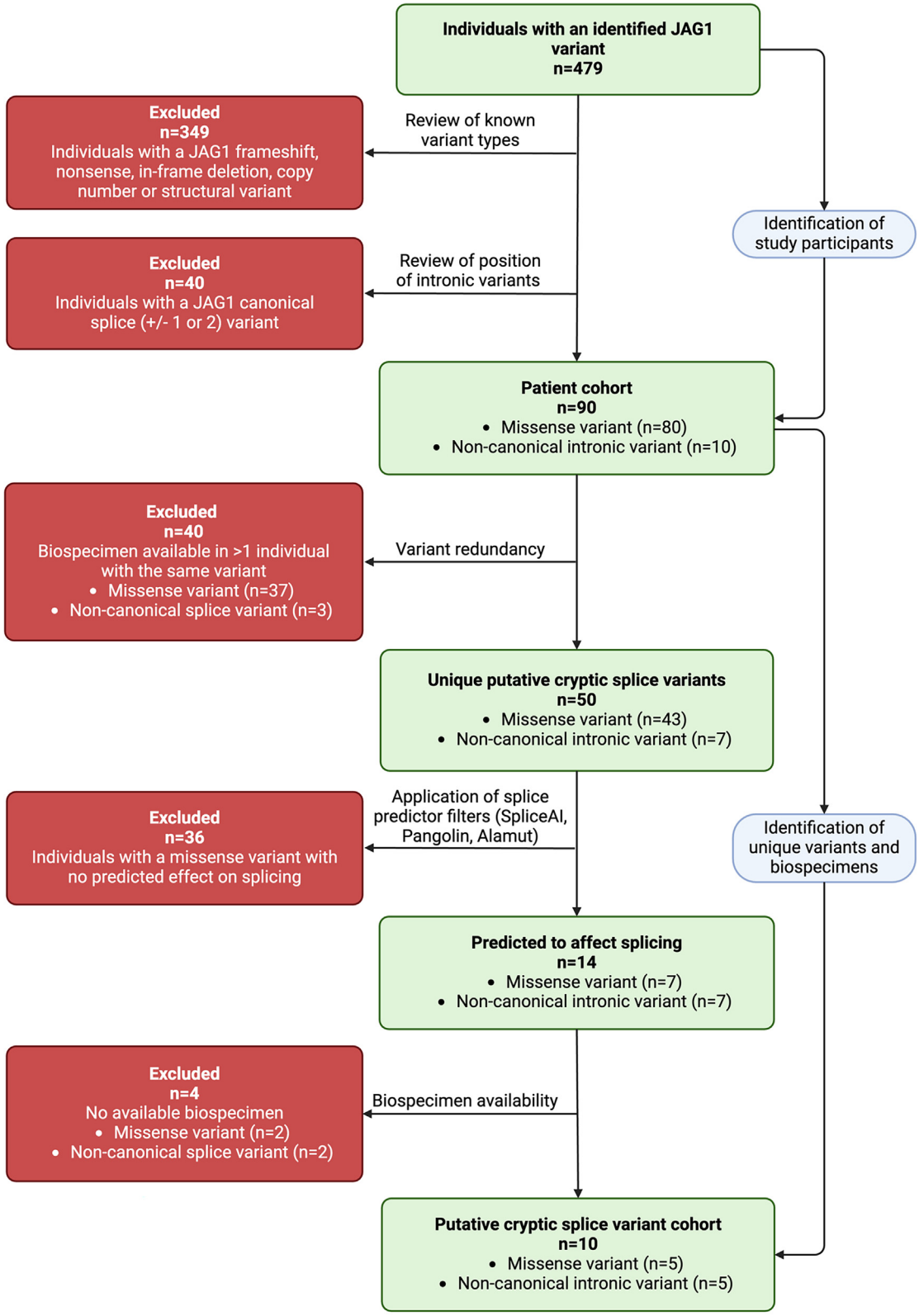

Figure 1. Flow diagram of the study population
RNA analysis was performed on a cohort of 10 individuals identified in our research study. Inclusion and exclusion criteria are indicated.

predicted to result in a cryptic splice site by *in silico* tools,[3,4] and four individuals who harbored intronic variants captured with our sequencing primers but outside of the ±1 or 2 canonical splice site, with a fifth variant resulting in a duplication of the splice donor (NM_000214.3:c.755+1_755+2dup). The final cohort included six males and four

females with a median age at enrollment of 12.6 years (range 0.5–32 years) and with clinical features presented in Table S5.

The five missense variants were found throughout the JAG1 protein, with no predominance for specific exons or functional domains (Table 1). Two intronic but non-canonical variants predicted to affect splicing were identified in adjacent bases within intron 3 (c.439+5G>A and c.439+6T>A; GenBank: NM_000214.3), while the remaining three were found within introns 5, 10, and 20. To study the effects of these variants on RNA, lymphoblastoid cell lines obtained from individuals with ALGS were treated with either DMSO or cycloheximide, and cDNA was generated from extracted RNA for study. Since many splicing alterations result in frameshift variants, with creation of a premature stop codon, these mutations would be predicted to be degraded via the nonsense-mediated decay pathway, and the use of cycloheximide interferes with translation, thereby blocking nonsense-mediated decay and allowing for the sequencing of the resultant mRNA.

Three of the five missense variants that were suspected to affect splicing were confirmed to produce altered splice products (Figure 2). The NM_000214.3:c.1156G>A (p.Gly386Arg) variant (individual 2) results in the formation of a cryptic splice site, producing a 37-nt deletion, which splices out the first 37 nucleotides of exon 9 and leads to a frameshift (p.Asn374Argfs*26). The NM_000214.3:c.1720G>C (p.Val574Leu) variant (individual 3) alters the last nucleotide of exon 13, introducing a cryptic splice site that results in splicing out of the entire exon 13 and leads to an early protein truncation (p.Leu524*). The NM_000214.3:c.2455A>G (p.Ile819Val) variant (individual 5) located within exon 20 introduces a cryptic splice site that results in splicing out of the last 4 nucleotides in exon 20 and leads to a frameshift (p.Ile819Thrfs*50). All three of these spliced products were stabilized by the addition of cycloheximide, an inhibitor of nonsense-mediated decay, which is expected when there is a premature stop codon. Two variants hypothesized to result in splicing abnormalities, NM_000214.3:c.686G>A (p.Cys229Tyr) (individual 1) and NM_000214.3:c.2231G>A (p.Arg744Gln) (individual 4), did not affect splicing and are predicted to result in missense substitutions (Figures 2B and 2E). These results are consistent with the *in silico* informatic prediction for the c.686G>A variant, which had SpliceAI and Pangolin Δ scores of 0 (Table S2). The c.686G>A variant was only predicted by Alamut to affect splicing. Alamut uses four splice prediction algorithms, and further review indicated that only one of the four algorithms (GeneSplicer[22]) predicted a moderate effect on splicing for the c.686G>A variant. The c.2231G>A variant, however, had a maximum Δ score of 0.48 (Pangolin, splice loss) (Table S2), which was well above the recommended Δ score threshold of 0.2.

All five of the non-canonical intronic variants were confirmed to produce aberrantly spliced products through the apparent introduction of cryptic splice sites after the analysis of cDNA (Figure 3). The NM_000214.3:c.439+5G>A and NM_000214.3:c.439+6T>C variants (individuals 6 and 7, respectively) both resulted in exon skipping and removal of the entire exon 3, leading to the same frameshift (p.Arg130Asnfs*14). The NM_000214.3:c.755+1_755+2dup variant (individual 8) identified in intron 5resulted in exon skipping and the removal of the entire exon 5, leading to a frameshift (p.Ala232Glyfs*160). The NM_000214.3:c.1349-12T>G variant (individual 9) identified in intron 10 resulted in exon skipping and removal of the entire exon 11, leading to a frameshift (p.Asn450Argfs*4). The NM_000214.3:c.2458+5G>A variant (individual 10) identified in intron 20 resulted in exon skipping and removal of the entire exon 20, leading to a premature stop (p.Cys791*). Two variants, c.755+1_755+2dup and c.2458+5G>A (GenBank: NM_000214.3), were incompletely sensitive to cycloheximide, indicating partial escape from nonsense-mediated decay for the aberrant splice products (Figure 3).

A literature review was performed to compile a comprehensive list of all reported *JAG1* non-canonical and cryptic splice variants, with a total of 23 identified variants, 8 of which were studied in our cohort (Figure 4; Table S6). Most of these variants (n = 18; 78.3%) were reported to be predicted splice variants, without laboratory confirmation, including 7 of 8 that we studied, while a minority (n = 5; 21.7%) were previously reported with RNA evidence supporting a splicing effect, including 1 of 8 that we studied (NM_000214.3:c.1156G>A; p.Asn374Argfs*26). We reviewed all 23 variants using SpliceAI and Pangolin to obtain *in silico* splicing predictions (Table S6). Two of 23 variants were not predicted to affect splicing by either SpliceAI or Pangolin (Δ score <0.2). Both variants (c.886G>A [p.Asp296Asn] and c.886+3A>G; GenBank: NM_000214.3) are located at the same exon/intron 6 boundary, and one of these variants, NM_000214.3:c 886+3A>G, had prior RNA evidence confirming an effect on splicing.[13] The c.886+3G>A variant was shown to cause skipping of exon 6 that resulted in a frameshift and premature termination (p.Arg252Argfs*17), while the c.886G>A variant was described as a putative splice variant due to its position near the splice site.[13,16] An additional two variants (c.1395G>T [p.Arg465=] and c.1395+3A>G; GenBank: NM_000214.3) had scores just above the 0.2 cutoff for SpliceAI (>0.2), but with higher scores for Pangolin (>0.6). Of these two variants, evidence supporting an RNA effect was previously published for the c.1395+3A>G variant (maximum SpliceAI/Pangolin Δ scores 0.26/0.64).[26] The second variant (c.1395G>T; maximum SpliceAI/Pangolin Δ scores 0.27/0.66) was originally published as a splice variant due to its position immediately upstream of the exon/intron 11 junction and its possible effect in generating a new splice donor site; however, confirmatory assays were not completed.[18] Of note, the NM_000214.3:c.1395G>T (p.Arg465=) variant would be predicted to result in a synonymous substitution at the amino acid level if it did not affect splicing.

Table 1. Splicing and protein effects of study variants

No.	Exon/intron	Nucleotide variant (NM_000214.3)	Predicted protein effect	Actual protein effect	Splicing effect	Max. SpliceAI score	Max. Pangolin score	Previous report(s)
1	exon 4	c.686G>A	(p.Cys229Tyr)	p.Cys229Tyr	no effect	0[a]	0[a]	none
2	exon 9	c.1156G>A	(p.Gly386Arg)	p.Asn374Argfs*26	37 bp del. in exon 9	0.98	0.99	Gilbert et al.,[7] Heritage et al.,[26] Lin et al.,[27] Tada et al.,[28] Wang et al.[29]
3	exon 13	c.1720G>C	(p.Val574Leu)	p.Leu524*	skipping of exon 13 (150 bp del.)	0.7	0.82	Gilbert et al.,[7] Samejima et al.[30]
4	exon 18	c.2231G>A	(p.Arg744Gln)	p.Arg744Gln	no effect	0.25	0.48	de Filippis et al.[31] *described in this report as a missense variant
5	exon 20	c.2455A>G	(p.Ile819Val)	p.Ile819Thrfs*50	4 bp del. in exon 20	0.72	0.87	Gilbert et al.[7]
6	intron 3	c.439+5G>A	(p.?)	p.Arg130Asnfs*14	skipping of exon 3 (51 bp del.)	0.97	0.8	Gilbert et al.[7]
7	intron 3	c.439+6T>A	(p.?)	p.Arg130Asnfs*14	skipping of exon 3 (51 bp del.)	0.77	0.65	Gilbert et al.,[7] Warthen et al.[25]
8	intron 5	c.755+1_755+2dup	(p.?)	p.Ala232Glyfs*160	skipping of exon 5 (100 bp del.)	0.92	0.98	Gilbert et al.[7]
9	intron 10	c.1349-12T>G	(p.?)	p.Asn450Argfs*4	skipping of exon 11 (54 bp del.)	0.64	0.72	Gilbert et al.,[7] Krantz et al.[12]
10	intron 20	c.2458+5G>A	(p.?)	p.Cys791*	skipping of exon 20 (85 bp del.)	0.49	0.64	Gilbert et al.,[7] Warthen et al.[25]

del., deletion; Max., maximum.
[a]Variant was predicted to affect splicing by Alamut.

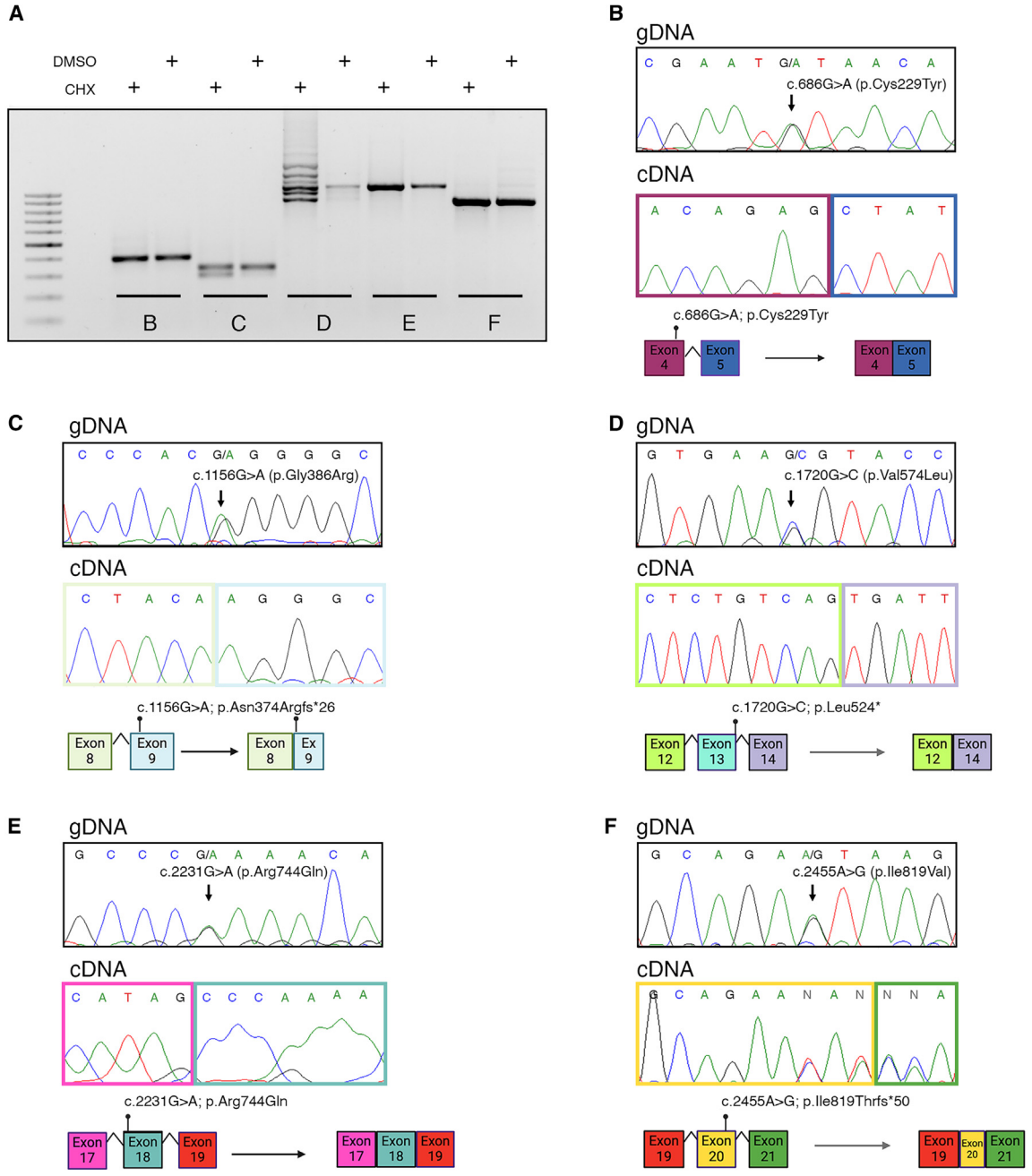

Figure 2. cDNA analysis of putative missense splice variants

(A) Gel of cDNA from cells treated with DMSO or cycloheximide. Bands are labeled according to the panel showing their respective gDNA and cDNA chromatograms along with a schematic of the splicing effect for (B) c.686G>A, (C) c.1156G>A, (D) c.1720G>C, (E) c.2231G>A, and (F) c.2455A>G.

(B–F) Sequencing of both gDNA and cDNA is shown for all variants. Variant and missense predictions are shown in the gDNA box. Boxes around the cDNA are colored to indicate the location of the sequence within its respective exon and correspond to the schematic below the chromatogram. Variant and protein consequences confirmed through RNA analysis are depicted in the protein schematic. GenBank: NM_000214.3.

Discussion

Splicing alterations are an important class of disease-causing variants in many human genetic disorders. While alterations in canonical splice sites are readily recognized and predicted to cause abnormal splicing, other alterations, such as the formation of cryptic splice sites generated by nucleotide changes resulting in apparent missense variants, or intronic nucleotide changes outside of the canonical positions, can be harder to interpret. The validation of a splicing effect for non-canonical and missense variants is important for the clinical diagnostics of ALGS since haploinsufficiency is the defined pathomechanism. Missense variants, which do not result in early protein

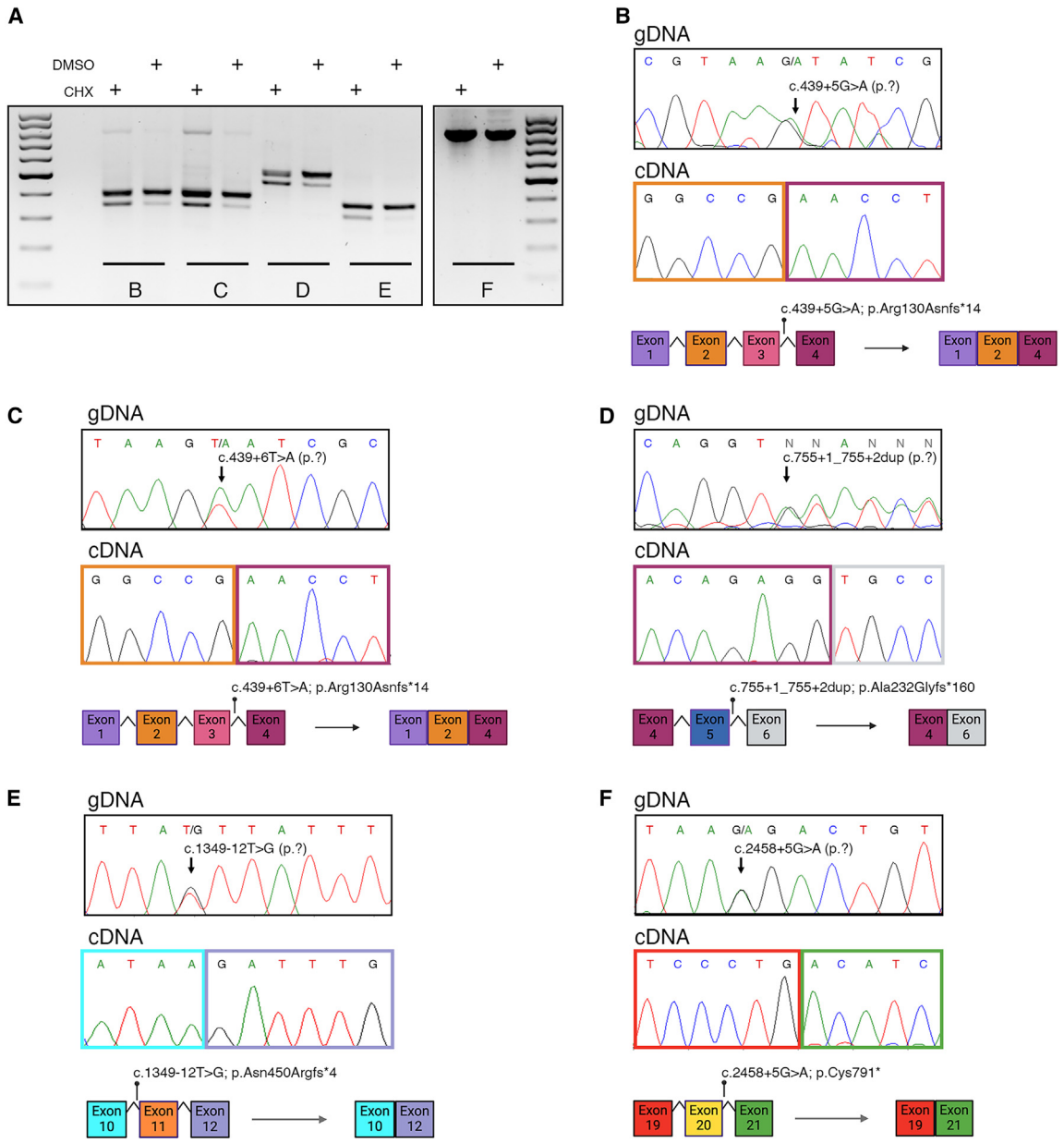

Figure 3. cDNA analysis of putative non-canonical splice variants

(A) Gel of cDNA from cells treated with DMSO or cycloheximide. Bands are labeled according to the panel showing their respective gDNA and cDNA chromatograms, along with a schematic of the splicing effect for (B) c.439+5G>A, (C) c.439+6T>A, (D) c.755+1_755+2dup, (E) c.1349-12T>G, and (F) c.2458+5G>A.

(B–F) Sequencing of both gDNA and cDNA is shown for all variants. Variants are shown in the gDNA box. Boxes around the cDNA are colored to indicate the location of the sequence within its respective exon and correspond to the schematic below the chromatogram. Variant and protein consequences confirmed through RNA analysis are depicted in the protein schematic. GenBank: NM_000214.3.

truncation, and non-canonical splice variants, which may or may not interfere with normal splicing, both lead to uncertain associations with disease and often require functional analysis of the protein product or RNA to confirm pathogenicity. Review of *JAG1* missense variants detected in our cohort of individuals with ALGS identified 43 variants, which we evaluated using splice prediction tools, and identified 7/43 with evidence for a splicing effect. We also identified 7 variants within ~100 bases of an intron/exon boundary, but not within the canonical ±1 or 2 positions. Of these combined 14 unique variants, there were 10 individuals for whom we had biospecimens, making it possible to study RNA for splicing alterations. Eight of 10 putative splicing variants studied were found to result in aberrant splice products, which confirms their pathogenic role in causing ALGS, while two missense variants did not affect splicing and likely result in functionally abnormal or improperly trafficked proteins that would require further validation (c.686G>A [p.Cys229Tyr] and c.2231G>A [p.Arg744Gln]; GenBank: NM_000214.3).

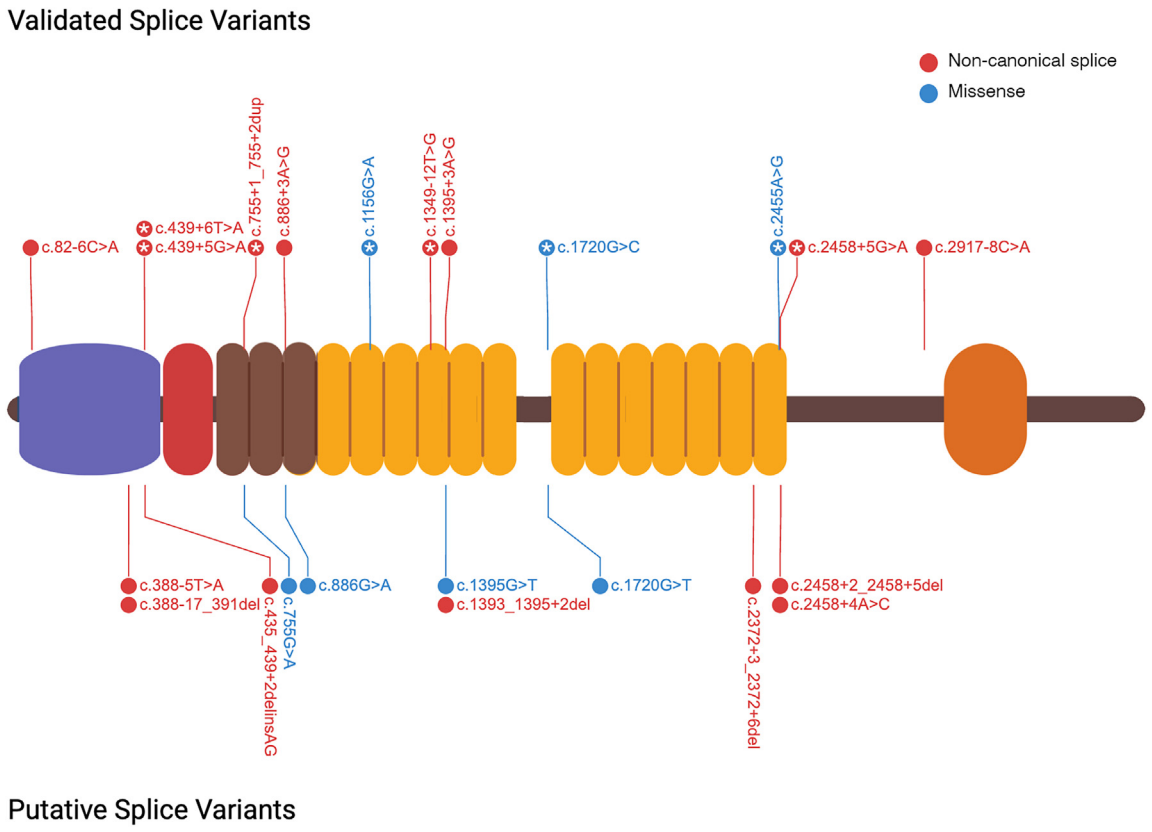

Figure 4. Schematic of all reported putative and confirmed splice variants
Variants are color-coded to indicate whether they were identified as missense (blue) or non-canonical intronic (red) splice variants. Lollipops with an asterisk indicate variants that were included in this study. JAG1 protein domains are color-coded for the C2-like domain (purple), the delta-serate-LAG2 domain (red), NOTCH2-interacting epidermal growth factor (EGF)-like domains (brown), EGF-like domains (yellow), and transmembrane domain (orange).

A review of the literature identified an additional 4 cryptic splice variants with RNA evidence to support aberrant splicing[10,13,14] and 18 putative splice variants without RNA evidence (Table S6). Of these 18 putative splice variants, 7 were included in our study, and we now have data to confirm their effect on splicing.[7,25,30] One additional variant from the literature with prior evidence to support an effect on splicing was also included in our cohort, where we confirmed the reported effect (NM_000214.3:c.1156G>A; p.Asn374Argfs*26).[27] Thus, there are now 12 (52.2%) validated and 11 putative *JAG1* splice variants (Figure 4).

To identify putative *JAG1* cryptic splice variants, we used *in silico* splice prediction tools, including Pangolin and SpliceAI.[3,4] Both tools were created using deep neural networking to predict splicing variants based on near and distant sequence composition, with SpliceAI predictions based on a single cell type in humans and Pangolin based on four cell types across four species. The probability that a variant is more likely to result in a splice alteration over the wild-type sequence is expressed as a Δ score, with higher Δ scores (>0.2 on a 0–1 scale) positively correlated with aberrant splice events. Use of these splice predictor tools for other disease genes has identified instances of both false positives, where variants above the recommended threshold are found not to affect splicing, and false negatives, where variants below this threshold are identified to affect splicing.[19,32] Thus, while these tools represent a notable and valuable development for splice variant identification, confirmatory assays are necessary to establish an RNA effect.

Our data suggest that SpliceAI and Pangolin have reduced sensitivity for variants with low Δ scores (<0.2), highlighting a limitation of these tools that results in the misidentification of true splice variants. Of the 10 variants we studied and 4 additional variants from the literature with RNA evidence, one (NM_000214.3:c.886+3A>G)[13] was validated to be a false negative using RNA evidence (maximum Δ score = 0.06). A second variant from our literature search (NM_000214.3:c.886G>A)[17] also had a maximum Δ score below 0.2 (0.11), but without confirmatory RNA evidence supporting aberrant splicing. Interestingly, both variants are located at the exon/intron 6 junction, which posits that this region's sequence composition may reduce the specificity of both splice predictor tools. Given that we used a Δ score threshold of >0.2 to identify variants in our cohort for this study, we are unable to confirm whether other true splice variants exist that were missed by SpliceAI and Pangolin; however, the finding that the c.886+3A>G variant affects splicing supports this possibility.

The specificity of SpliceAI at a Δ score >0.2 has been reported at ~90%[4,5,33,34]; thus, there is an expected incidence of false positives when using this cutoff. Of the 10 variants we studied and the 4 additional variants from the literature with RNA evidence, only one had a Δ score >0.2, but this did not affect splicing (NM_000214.3:c.2231G>A [p.Arg744Gln]). Although this variant was the second-lowest-scoring variant within the cohort, its maximum Δ score of 0.48 was well above the cutoff of 0.2 used for SpliceAI and Pangolin.[19] The lowest-scoring variant, NM_000214.3:c.686G>A (p.Cys229Tyr), had a Δ score of 0 by both Pangolin and SpliceAI, and it was only retained for study due to its location near the exon 4 terminus and the Alamut prediction that it might affect splicing. A closer review of Alamut indicated that only one of the four splice prediction tools incorporated within the Alamut software projected a moderate effect on splicing. Indeed, of the 10 variants in our study, Alamut incorrectly predicted the splicing effects of 6, indicating that Alamut does not provide the best predictive modeling of splice variants. Although increasing the Δ score threshold for SpliceAI and Pangolin would improve specificity and reduce the rate of false positives for *JAG1* variants, the identification of the two variants described above with Δ scores below 0.2 (c.886G>A and c.886+3A>G; GenBank: NM_000214.3), but with confirmatory RNA evidence supporting aberrant splicing of the c.886+3A>G variant,[13,16] highlights the limitations of balancing sensitivity with specificity. The size of our cohort and the ascertainment bias of excluding variants from study with a Δ score <0.2 inhibits our ability to calculate accurate sensitivity and specificity estimates for SpliceAI and Pangolin, but our finding of both false positives and false negatives when applying this cutoff suggests similarities with other published studies.[4,5,19,33,34]

The lowest maximum Δ score of all predicted splice variants with confirmatory RNA evidence (our study combined with those from the literature, $n = 12$) was 0.64 (both c.2917-8C>A and c.2458+5G>A; GenBank: NM_000214.3), suggesting that Δ scores above this cutoff have a high likelihood of causing aberrant splicing. Sensitivity and specificity should be considered thoroughly when using tools such as SpliceAI and Pangolin as the predicted molecular consequence, location of the variant, and tissue of study may require different thresholds.[35] Although our combined cohort of 14 variants with RNA evidence ($n = 12$ confirmed splice variant and $n = 2$ likely missense variants) presented here or previously described is small, it suggests that a Δ score cutoff >0.6 is recommended for high sensitivity and that there is a greater importance in validating the splicing effects of variants with Δ scores below this using RNA assays. This small cohort size is a limitation of our study and is due in part to both the rare nature of ALGS and the low availability of biospecimens for functional study. Increasing the cohort size through future studies could better resolve cutoff recommendations.

We noted that SpliceAI and Pangolin Δ scores occasionally differed in their strength of confidence. Two variants from the literature at the exon/intron 11 junction, NM_000214.3:c.1395+3A>G and NM_000214.3:c.1395G>T (p.Arg465=), had maximum SpliceAI Δ scores just above the 0.2 cutoff (0.26 and 0.27, respectively), but higher Pangolin Δ scores (0.64 and 0.66, respectively), highlighting differences between the SpliceAI and Pangolin algorithms in the interpretation of sequence composition and suggesting that both tools should be reviewed when evaluating a putative splice variant. Of note, the NM_000214.3:c.1395+3A>G variant has been confirmed to affect splicing based on RNA evidence.[26] The NM_000214.3:c.1395G>T (p.Arg465=) variant occurs at the last nucleotide of exon 11 and encodes a synonymous change; however, predictions from these tools, identification of the variant in affected individuals,[18] and absence of the variant in population databases (gnomAD versions 2 and 3) support a possible effect on splicing that should be confirmed using RNA.

Utilization of cycloheximide in our cDNA assays was essential to inhibit nonsense-mediated decay and allow for sequencing of the splice products. Surprisingly, two variants (c.755+2G>T and c.2458+5G>A; GenBank: NM_000214.3) resulted in aberrantly spliced products that appeared to partially escape nonsense-mediated decay, as is evident by their presence in the DMSO control. The prevailing pathomechanism of ALGS is haploinsufficiency whereby Notch signaling is disrupted due to a depleted dosage of JAG1 protein.[3] A dominant-negative effect has been proposed, with evidence suggesting that soluble JAG1 can bind NOTCH2, but studies have been unable to confirm that mutant JAG1 is released from the endoplasmic reticulum to exert this effect under physiological conditions.[28,36–38] Thus, the presence of these bands in our samples does not indicate whether these products are sequestered intracellularly or secreted, and more research would be needed to discern those effects.

In conclusion, we show that three *JAG1* missense variants and five intronic variants outside of the canonical splice site positions identified in individuals with ALGS cause abnormal splicing and are likely pathogenic. Review of an additional 15 putative or confirmed splice variants from the literature allowed for a comparison of RNA analyses with the *in silico* splice prediction tools SpliceAI and Pangolin. Review of true splice variants with confirmatory RNA evidence indicates that Δ scores >0.6 show high predictive sensitivity. Usage of the recommended cutoff of 0.2 results in instances of both false positives and false negatives, highlighting the limitations of these tools and underscoring the need to integrate RNA evidence with computational predictions for variants with low Δ scores. We expect that this will provide guidance for the incorporation of these tools during clinical variant classification of *JAG1* and potentially other disease genes.

Data and code availability

The published article includes all datasets generated or analyzed during this study.

Acknowledgments

This project was internally funded by the Fred and Suzanne Biesecker Pediatric Liver Center (N.B.S.) and R01-DK134585-01A1 (N.B.S.).

Author contributions

A.M.J., K.M.L., N.B.S., and M.A.G. designed the research plan. E.K.-J., N.V., A.M.J., G.M., A.R., A.T., G.M., K.L.M., N.B.S., and M.A.G. performed the research and analyzed the data. E.K.-J., N.V., and M.A.G. wrote the manuscript. All authors read, edited, and approved the final version of the manuscript.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.xhgg.2024.100351.

Web resources

SpliceAI and Pangolin: https://spliceailookup.broadinstitute.org/.
MaxEntScan: http://hollywood.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html.
GeneSplicer: https://www.cbcb.umd.edu/software/GeneSplicer/gene_spl.shtml.
NNSplice 0.9: https://www.fruitfly.org/seq_tools/splice.html.
Primer3Plus: https://www.primer3plus.com/index.html.

Received: July 2, 2024
Accepted: September 4, 2024

References

1. Faustino, N.A., and Cooper, T.A. (2003). Pre-mRNA splicing and human disease. Genes Dev. *17*, 419–437.
2. Dvinge, H. (2018). Regulation of alternative mRNA splicing: old players and new perspectives. FEBS Lett. *592*, 2987–3006.
3. Zeng, T., and Li, Y.I. (2022). Predicting RNA splicing from DNA sequence using Pangolin. Genome Biol. *23*, 103.
4. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. Cell *176*, 535–548.e24.
5. Kurosawa, R., Iida, K., Ajiro, M., Awaya, T., Yamada, M., Kosaki, K., and Hagiwara, M. (2023). PDIVAS: Pathogenicity predictor for Deep-Intronic Variants causing Aberrant Splicing. BMC Genom. *24*, 601.
6. Smith, C., and Kitzman, J.O. (2023). Benchmarking splice variant prediction algorithms using massively parallel splicing assays. Genome Biol. *24*, 294.
7. Gilbert, M.A., Bauer, R.C., Rajagopalan, R., Grochowski, C.M., Chao, G., McEldrew, D., Nassur, J.A., Rand, E.B., Krock, B.L., Kamath, B.M., et al. (2019). Alagille syndrome mutation update: Comprehensive overview of JAG1 and NOTCH2 mutation frequencies and insight into missense variant classification. Hum. Mutat. *40*, 2197–2220.
8. Mitchell, E., Gilbert, M., and Loomes, K.M. (2018). Alagille Syndrome. Clin. Liver Dis. *22*, 625–641.
9. Gilbert, M.A., and Spinner, N.B. (2018). Genetics of Alagille Syndrome. In Alagille Syndrome: Pathogenesis and Clinical Management, B.M. Kamath and K.M. Loomes, eds. (Springer), pp. 33–48.
10. Vazquez-Martinez, E.R., Varela-Fascinetto, G., Garcia-Delgado, C., Rodríguez-Espino, B.A., Sánchez-Boiso, A., Valencia-Mayoral, P., Heller-Rosseau, S., Pelcastre-Luna, E.L., Zenteno, J.C., Cerbón, M., et al. (2014). Polymorphism analysis and new JAG1 gene mutations of Alagille syndrome in Mexican population. Meta Gene *2*, 32–40.
11. Crosnier, C., Driancourt, C., Raynaud, N., Dhorne-Pollet, S., Pollet, N., Bernard, O., Hadchouel, M., and Meunier-Rotival, M. (1999). Mutations in JAGGED1 gene are predominantly sporadic in Alagille syndrome. Gastroenterology *116*, 1141–1148.
12. Krantz, I.D., Colliton, R.P., Genin, A., Rand, E.B., Li, L., Piccoli, D.A., and Spinner, N.B. (1998). Spectrum and frequency of jagged1 (JAG1) mutations in Alagille syndrome patients and their families. Am. J. Hum. Genet. *62*, 1361–1369.
13. Heritage, M.L., MacMillan, J.C., and Anderson, G.J. (2002). DHPLC mutation analysis of Jagged1 (JAG1) reveals six novel mutations in Australian alagille syndrome patients. Hum. Mutat. *20*, 481.
14. Chen, Y., Liu, X., Chen, S., Zhang, J., and Xu, C. (2019). Targeted Sequencing and RNA Assay Reveal a Noncanonical JAG1 Splicing Variant Causing Alagille Syndrome. Front. Genet. *10*, 1363.
15. Stenson, P.D., Mort, M., Ball, E.V., Evans, K., Hayden, M., Heywood, S., Hussain, M., Phillips, A.D., and Cooper, D.N. (2017). The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum. Genet. *136*, 665–677.
16. Onouchi, Y., Kurahashi, H., Tajiri, H., Ida, S., Okada, S., and Nakamura, Y. (1999). Genetic alterations in the JAG1 gene in Japanese patients with Alagille syndrome. J. Hum. Genet. *44*, 235–239.
17. Guegan, K., Stals, K., Day, M., Turnpenny, P., and Ellard, S. (2012). JAG1 mutations are found in approximately one third of patients presenting with only one or two clinical features of Alagille syndrome. Clin. Genet. *82*, 33–40.
18. Oda, T., Elkahloun, A.G., Pike, B.L., Okajima, K., Krantz, I.D., Genin, A., Piccoli, D.A., Meltzer, P.S., Spinner, N.B., Collins, F.S., and Chandrasekharappa, S.C. (1997). Mutations in the human Jagged1 gene are responsible for Alagille syndrome. Nat. Genet. *16*, 235–242.
19. de Sainte Agathe, J.M., Filser, M., Isidor, B., Besnard, T., Gueguen, P., Perrin, A., Van Goethem, C., Verebi, C., Masingue, M., Rendu, J., et al. (2023). SpliceAI-visual: a free online tool to improve SpliceAI splicing variant interpretation. Hum. Genomics *17*, 7.
20. Shapiro, M.B., and Senapathy, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. *15*, 7155–7174.

21. Yeo, G., and Burge, C.B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J. Comput. Biol. *11*, 377–394.
22. Pertea, M., Lin, X., and Salzberg, S.L. (2001). GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res. *29*, 1185–1190.
23. Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. J. Comput. Biol. *4*, 311–323.
24. Tarpey, P.S., Raymond, F.L., Nguyen, L.S., Rodriguez, J., Hackett, A., Vandeleur, L., Smith, R., Shoubridge, C., Edkins, S., Stevens, C., et al. (2007). Mutations in UPF3B, a member of the nonsense-mediated mRNA decay complex, cause syndromic and nonsyndromic mental retardation. Nat. Genet. *39*, 1127–1133.
25. Warthen, D.M., Moore, E.C., Kamath, B.M., Morrissette, J.J.D., Sanchez-Lara, P.A., Piccoli, D.A., Krantz, I.D., and Spinner, N.B. (2006). Jagged1 (JAG1) mutations in Alagille syndrome: increasing the mutation detection rate. Hum. Mutat. *27*, 436–443.
26. Heritage, M.L., MacMillan, J.C., Colliton, R.P., Genin, A., Spinner, N.B., and Anderson, G.J. (2000). Jagged1 (JAG1) mutation detection in an Australian Alagille syndrome population. Hum. Mutat. *16*, 408–416.
27. Lin, H.C., Le Hoang, P., Hutchinson, A., Chao, G., Gerfen, J., Loomes, K.M., Krantz, I., Kamath, B.M., and Spinner, N.B. (2012). Alagille syndrome in a Vietnamese cohort: mutation analysis and assessment of facial features. Am. J. Med. Genet. *158A*, 1005–1013.
28. Tada, M., Itoh, S., Ishii-Watabe, A., Suzuki, T., and Kawasaki, N. (2012). Functional analysis of the Notch ligand Jagged1 missense mutant proteins underlying Alagille syndrome. FEBS J. *279*, 2096–2107.
29. Wang, H., Wang, X., Li, Q., Chen, S., Liu, L., Wei, Z., Wang, L., Liu, Y., Zhao, X., He, L., et al. (2012). Analysis of JAG1 gene variant in Chinese patients with Alagille syndrome. Gene *499*, 191–193.
30. Samejima, H., Torii, C., Kosaki, R., Kurosawa, K., Yoshihashi, H., Muroya, K., Okamoto, N., Watanabe, Y., Kosho, T., Kubota, M., et al. (2007). Screening for Alagille syndrome mutations in the JAG1 and NOTCH2 genes using denaturing high-performance liquid chromatography. Genet. Test. *11*, 216–227.
31. de Filippis, T., Marelli, F., Nebbia, G., Porazzi, P., Corbetta, S., Fugazzola, L., Gastaldi, R., Vigone, M.C., Biffanti, R., Frizziero, D., et al. (2016). JAG1 Loss-Of-Function Variations as a Novel Predisposing Event in the Pathogenesis of Congenital Thyroid Defects. J. Clin. Endocrinol. Metab. *101*, 861–870.
32. Strauch, Y., Lord, J., Niranjan, M., and Baralle, D. (2022). CI-SpliceAI-Improving machine learning predictions of disease causing splicing variants using curated alternative splice sites. PLoS One *17*, e0269159.
33. Ha, C., Kim, J.W., and Jang, J.H. (2021). Performance Evaluation of SpliceAI for the Prediction of Splicing of NF1 Variants. Genes *12*, 1308.
34. Wai, H.A., Lord, J., Lyon, M., Gunning, A., Kelly, H., Cibin, P., Seaby, E.G., Spiers-Fitzgerald, K., Lye, J., Ellard, S., et al. (2020). Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. Genet. Med. *22*, 1005–1014.
35. Rowlands, C., Thomas, H.B., Lord, J., Wai, H.A., Arno, G., Beaman, G., Sergouniotis, P., Gomes-Silva, B., Campbell, C., Gossan, N., et al. (2021). Comparison of in silico strategies to prioritize rare genomic variants impacting RNA splicing for the diagnosis of genomic disorders. Sci. Rep. *11*, 20607.
36. Boyer-Di Ponio, J., Wright-Crosnier, C., Groyer-Picard, M.T., Driancourt, C., Beau, I., Hadchouel, M., and Meunier-Rotival, M. (2007). Biological function of mutant forms of JAGGED1 proteins in Alagille syndrome: inhibitory effect on Notch signaling. Hum. Mol. Genet. *16*, 2683–2692.
37. Guan, Y., Xu, D., Garfin, P.M., Ehmer, U., Hurwitz, M., Enns, G., Michie, S., Wu, M., Zheng, M., Nishimura, T., et al. (2017). Human hepatic organoids for the analysis of human genetic diseases. JCI Insight *2*, e94954.
38. Xiao, Y., Gong, D., and Wang, W. (2013). Soluble JAGGED1 inhibits pulmonary hypertension by attenuating notch signaling. Arterioscler. Thromb. Vasc. Biol. *33*, 2733–2739.