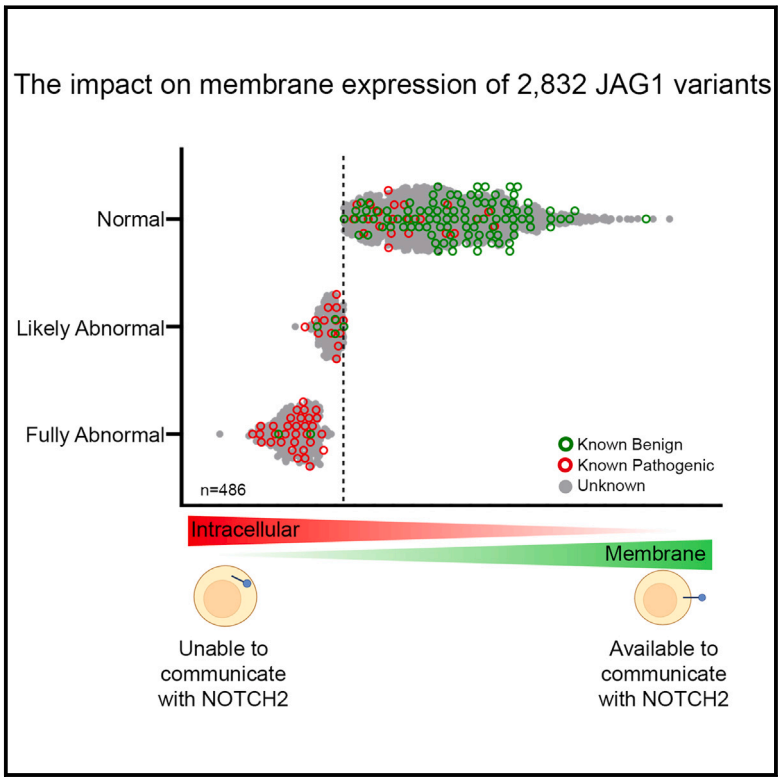
Functional characterization of 2,832 JAG1 variants supports reclassification for Alagille syndrome and improves guidance for clinical variant interpretation

Graphical abstract



Authors

Melissa A. Gilbert, Ernest Keefer-Jacques, Tanaya Jadhav, ..., Ramakrishnan Rajagopalan, Tristan J. Hayeck, Nancy B. Spinner

Correspondence

gilbertma@chop.edu

JAG1 haploinsufficiency is the most common cause of the multi-system disorder, Alagille syndrome. We analyzed cellular localization of 2,832 missense and synonymous variants. We show that 486 tested variants are not appropriately expressed on the membrane and that integration of this data during clinical variant classification reduces uncertainty.

Gilbert et al., 2024, The American Journal of Human Genetics *111*, 1656–1672
August 8, 2024
https://doi.org/10.1016/j.ajhg.2024.06.011

CellPress

ARTICLE

Functional characterization of 2,832 JAG1 variants supports reclassification for Alagille syndrome and improves guidance for clinical variant interpretation

Melissa A. Gilbert,[1,2,3,*] Ernest Keefer-Jacques,[1] Tanaya Jadhav,[1] Daniel Antfolk,[5] Qianqian Ming,[5] Nicolette Valente,[1] Grace Tzun-Wen Shaw,[1] Christopher J. Sottolano,[1,2] Grace Matwijec,[1] Vincent C. Luca,[5] Kathleen M. Loomes,[3,4] Ramakrishnan Rajagopalan,[1,2] Tristan J. Hayeck,[1,2] and Nancy B. Spinner[1,2]

Summary

Pathogenic variants in the *JAG1* gene are a primary cause of the multi-system disorder Alagille syndrome. Although variant detection rates are high for this disease, there is uncertainty associated with the classification of missense variants that leads to reduced diagnostic yield. Consequently, up to 85% of reported *JAG1* missense variants have uncertain or conflicting classifications. We generated a library of 2,832 *JAG1* nucleotide variants within exons 1–7, a region with a high number of reported missense variants, and designed a high-throughput assay to measure JAG1 membrane expression, a requirement for normal function. After calibration using a set of 175 known or predicted pathogenic and benign variants included within the variant library, 486 variants were characterized as functionally abnormal ($n = 277$ abnormal and $n = 209$ likely abnormal), of which 439 (90.3%) were missense. We identified divergent membrane expression occurring at specific residues, indicating that loss of the wild-type residue itself does not drive pathogenicity, a finding supported by structural modeling data and with broad implications for clinical variant classification both for Alagille syndrome and globally across other disease genes. Of 144 uncertain variants reported in patients undergoing clinical or research testing, 27 had functionally abnormal membrane expression, and inclusion of our data resulted in the reclassification of 26 to likely pathogenic. Functional evidence augments the classification of genomic variants, reducing uncertainty and improving diagnostics. Inclusion of this repository of functional evidence during *JAG1* variant reclassification will significantly affect resolution of variant pathogenicity, making a critical impact on the molecular diagnosis of Alagille syndrome.

Introduction

Alagille syndrome (ALGS [MIM: 118450]) is an autosomal dominant disorder caused by defective Notch signaling that leads to a range of clinical features including hepatic, cardiac, vertebral, ocular, renal, and facial phenotypes.[1–6] ALGS is characterized by a high degree of variable expressivity in disease severity with no identified genotype-phenotype correlations.[3,5,7–14] Variants in the Notch signaling ligand jagged1 (*JAG1* [MIM: 601920]) are identified in most patients meeting strict clinical criteria for ALGS (94.3%), while variants in the Notch receptor, *NOTCH2* (MIM: 600275) are identified in an additional 2.5% of patients.[15] Variant identification rates, however, do not directly translate to diagnostic yield, as a subset of variants are classified as variants of uncertain significance (VOUSs).

JAG1 VOUSs are overwhelmingly missense, as protein-truncating variants, including full and partial gene deletions, are expected to lead to disease through an established pathomechanism of haploinsufficiency.[15–20] However, roughly 15% of ALGS-associated *JAG1* variants are missense,[15] which are difficult to classify, often requiring functional characterization to determine whether the protein is defective.[21] Of the over 100 missense variants reported in the Human Gene Mutation Database (HGMD)[22] for *JAG1*, only 15 have been studied (13.3%; 15 out of 113), with results showing impairments to cellular trafficking (intracellular retention), glycosylation, and/or a defect in their ability to activate NOTCH2.[15,17,18,20,23] Uncertainty in the functional effects of the remaining missense variants leads to their distinction as VOUSs, driving the diagnostic yield for identified *JAG1* variants down to just over 85%.

The number of missense variants identified in *JAG1*, however, far exceeds the ALGS disease-associated variants cataloged in HGMD. Given the hepatic disease features of ALGS, *JAG1* is included on many commercially available next-generation sequencing-based panel tests for cholestatic and related diseases, leading to a reported VOUS rate of 51.4% for *JAG1* variants in phenotypically diverse populations.[24] Consequently, there are 575 *JAG1* missense variants classified as uncertain or with conflicting interpretations in ClinVar,[25] a public repository of DNA variants

[1]Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [2]Department of Pathology and Laboratory Medicine, The Perelman School of Medicine at The University of Pennsylvania, Philadelphia, PA 19104, USA; [3]Division of Pediatric Gastroenterology, Hepatology, and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [4]Department of Pediatrics, The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA 19104, USA; [5]Department of Immunology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33612, USA
*Correspondence: gilbertma@chop.edu
https://doi.org/10.1016/j.ajhg.2024.06.011.

and their associated phenotypes, accounting for 86% (n = 575 out of 667) of all classified *JAG1* missense variants. Further obscuring our understanding of the biological significance of *JAG1* missense variants is data from reportedly healthy individuals extracted from the Genome Aggregation Database (gnomAD).[26] Of the 489 *JAG1* missense variants recorded in gnomAD (v2.1.1), 427 (87.3%) are reported at a minor allele frequency (MAF) that is lower than the estimated incidence rate for ALGS (1:30,000; MAF < 3.33E-05).[4,27,28] Given the variable expressivity of ALGS, these low-frequency variants in gnomAD call into question the rate of missed diagnosis of mild disease. Thus, the burden of missense VOUSs is high for *JAG1*, and resolution of this uncertainty by functional characterization of individual variants is necessary to improve diagnostics for cholestatic liver diseases, including ALGS.

Although *JAG1* missense variants are identified across the entire length of the gene, a significantly higher proportion of ALGS disease-associated variants are identified in exons 1–7 compared to exons 8–26.[5,15,29] This region houses important functional domains, including the signal peptide (SP) and C2-like domain, both of which facilitate intracellular and membrane trafficking,[30,31] and the Delta/Serate/Lag-2 (DSL) domain and first three epidermal growth factor-like (EGF-like) repeats, which are involved with physical binding to NOTCH2.[32,33] Of the 113 missense variants reported in HGMD for ALGS, 81% (n = 92) are reported to occur in this region,[22] with functional characterization supporting a pathogenic effect for only 5% (n = 5 out of 92).

We designed a multiplexed assay of variant effects (MAVE) to facilitate high-throughput functional characterization of a *JAG1* mutagenesis library of 2,832 variants encompassing nearly every possible nucleotide permutation in *JAG1* exons 1–7 (97.4%). We challenged our variant library with a flow cytometry assay designed to distinguish cells with JAG1 membrane expression from those without and used long-read sequencing to determine the identity of variants present in each of these two populations. We validated our assay through structural modeling and by confirming the appropriate distribution of a variant control set containing known or predicted benign and pathogenic variants into normal and abnormal membrane expression, respectively. Using this variant control set, we translated our membrane expression data into qualitative evidence strength for use during clinical variant classification and applied this to the reclassification of *JAG1* variants previously associated with ALGS.

Material and methods

JAG1 variant library

The *JAG1* variant library was synthesized (Twist Biosciences, San Francisco, CA) to introduce every possible nucleotide permutation across exons 1–7 (1,008 nucleotides; 336 amino acids [aa]). All synonymous changes were retained. Of 2,907 possible variants, 2,868 variants (98.7%) passed quality control and were included in the linear library.

Stable cell line generation

Wild-type (WT) *JAG1* (GenBank: NM_000214.3) and all mutant variants, including the site saturation variant library (SSVL), were cloned into a plasmid containing an IRES (internal ribosome entry site)-eGFP consensus sequence, allowing for eGFP co-expression (Twist Biosciences, San Francisco, CA) (Figure S1). Lentiviral particles containing plasmid DNA were used to transduce NIH3T3 cells (VectorBuilder, Chicago, IL). NIH3T3 cells express low endogenous levels of JAG1 and have previously been used to overexpress JAG1 mutants.[15,18,23] For generation of the SSVL, cells were transduced at a multiplicity of infection (MOI) of 1, achieving a transduction rate of ~30% to ensure that transduced cells received a single plasmid. Drug resistance (puromycin) and GFP expression were used to confirm and maintain stable cell expression.

Amplicon-based long-read sequencing (Pacific Biosciences, Menlo Park, CA) performed after creation of the *JAG1* SSVL indicated the presence of 2,832 total variants (97.4% of targeted variants), including seven mutant variants that failed quality control for the linear library build but that were generated separately as pure cell lines and spiked into the final cell library (c.41T>C [p.Leu14Pro], c.41T>G [p.Leu14Arg], c.45C>G [p.Ser15Arg], c.47T>C [p.Leu16Pro], c.47T>G [p.Leu16Arg, c.814G>A [p.Val272Ile], and c.814G>T [p.Val272Phe] [GenBank: NM_000214.3]).

Immunofluorescence

Cells overexpressing WT *JAG1* or various mutant constructs were seeded onto glass chamber slides and fixed with 4% formaldehyde in phosphate-buffered saline (PBS). After fixation, slides were placed in a steamer with citrate buffer for 30 min and allowed to cool for 30 min at room temperature for antigen retrieval. Cells were permeabilized for 10 min in 0.25% Triton X-100 followed by blocking in 4% bovine serum albumin (BSA). Primary antibodies against JAG1 (Santa Cruz, sc-390177; 1:30) and GFP (Abcam, ab13970; 1:500) were added overnight. Secondary antibodies for JAG1 (Jackson ImmunoResearch, 111-2995-003; 1:200) and GFP (Abcam, ab225314; 1:1000) were added for 1 h, and cells were visualized using a Leica DMi8 inverted microscope.

Flow cytometry

Cells were lifted from plates using Accutase to preserve membrane protein integrity. Pelleted cells were resuspended in PBS with 25 mM EDTA containing a 1:1,000 dilution of a JAG1 primary antibody (Abcam, ab273571) for 20 min at room temperature and were washed in cell staining solution before incubation in a secondary antibody conjugated to the fluorophore allophycocyanin (APC) (Jackson ImmunoResearch, 115-605-003) for 10 min. After a second wash in cell staining solution, cells were filtered through a 35 μm nylon mesh and propidium iodide (PI) was added to identify live cells.

Cells were sorted on either a FACSMelody cell sorter (BD Biosciences, Franklin Lakes, NJ) or a MoFlo Astrios cell sorter (Beckman Coulter, Brea, CA). Single, live cells were selected using forward and side scatter and by absence of PI fluorescence, and this population was further sorted based on APC and GFP fluorescence. Two populations of cells were collected: (1) APC high; GFP+ (membrane expression of JAG1) and (2) APC low; GFP+ (non-membrane expression of JAG1). Untransfected NIH3T3 cells were

used to define the lower boundary of the GFP+ gate, and WT and mutant JAG1 cell lines were used to define the lower boundary of the APC+ gate (Figures S2–S4). Gating was done conservatively such that the lower boundary of the APC-high population and the upper boundary of the APC-low population were not immediately adjacent to reduce the collection of cells with intermediate APC expression. Seven replicate sorts of the SSVL were collected, and for each sort, roughly one million cells were collected for each population (Table S1). FlowJo (Ashland, OR) was used to generate histograms.

Long-read sequencing

Genomic DNA was extracted from sorted cells using the Nanobind CBB Big DNA Kit (Pacific Biosciences, Menlo Park, CA) and cells were sequenced using amplicon-based long-read PacBio sequencing. A two-step PCR strategy was employed to first amplify *JAG1* exons 1–7 from gDNA and then to add barcodes to amplified products to enable sample multiplexing. Barcodes were added such that all variants from each population (APC high; GFP+ or APC low; GFP+) in a given experimental replicate received the same barcode. For each replicate, 200 ng of gDNA from both the membrane (APC high; GFP+) and non-membrane (APC low; GFP+) populations was amplified in two or three separate reactions, which were later combined to ensure sufficient input before proceeding to library generation. Primers for first-round amplification were blocked at the 5′ end by 5AmMC6 to prevent unbarcoded amplicons from ligating to the SMRTbell adapters during library prep and increasing yield. The following primers were used to amplify *JAG1* (1,488 bp amplicon size):

Forward: 5′ 5AmMC6-gcagtcgaacatgtagctgactcaggtcac**GCAACCAGGTGTGGAAAGTC**.

Reverse: 5′ 5AmMC6-tggatcacttgtgcaagcatcacatcgtag**GGCACACACACTTAAATCCGTTA**.

Following both rounds of PCR, samples from each experiment (including both populations collected from a single sort) were combined at equimolar ratios and used for library generation using the SMRTbell prep kit 3.0 (Pacific Biosciences, Menlo Park, CA). A total of 30 ng of library DNA was used to load a single SMRTcell, and sequencing was performed on a Sequel IIe instrument (Pacific Biosciences, Menlo Park, CA).

Bioinformatic analysis

Following sequencing, Lima (Pacific Biosciences, Menlo Park, CA) was used to demultiplex barcoded samples on a single SMRTcell into the two collected populations (APC high; GFP+ and APC low; GFP+). Sequencing reads were aligned to *JAG1* (GenBank: NM_000214.3) in the Hg38/GRCh38 reference using pbmm2 (Pacific Bioscience, Menlo Park, CA). Overestimation of reads within ~1 codon of the exon boundaries due to non-uniform alignment and sequence similarities in these regions was normalized for during downstream analysis by controlling for depth of coverage and random distribution across these variants in these positions. The depth of coverage and allele counts at each position in *JAG1* exons 1–7 were obtained using the GATK (v4.3) DepthOfCoverage tool[34] and the genomic positions with variants were retained for analysis. Ensembl Variant Effect Predictor (VEP)[35] was used to annotate the variants, including the aa changes and CADD scores. Variants observed at very low abundance that could be attributed to randomly distributed sequencing errors were removed. Variants observed at a frequency of less than one standard deviation in both populations (APC high and APC low) were filtered out of the dataset. Reads covering each position ("anchor" position) were grouped together (read group) and the frequencies of all variants were determined in each bin. JVarkit (https://lindenb.github.io/jvarkit/JvarkitCentral.html) was used to filter out reads with a second variant if the frequency of the secondary variant was higher than a threshold (mean: one standard deviation of all the variants observed in the read group). Table S2 includes the median, minimum, maximum, and standard deviation of reads containing each variant observed across all seven experimental replicates after application of all filtering steps.

Curation of benign and pathogenic JAG1 variant control set

A list of 111 benign and 70 pathogenic variants were ascertained from publicly available databases. SpliceAI[36] and Alamut (SOPHiA Genetics) were used to remove any variants that were predicted to affect splicing. For benign variants, first ClinVar was queried for variants with a benign or likely benign classification ($n = 84$). An additional 14 variants that were listed as benign or likely benign for ALGS specifically were included from variants classified as conflicting. Using a disease incidence of 1:30,000 for ALGS, 11 variants were identified in gnomAD (v2.1.1and v3.1.2) with an MAF > 3.33E-05 that had not previously been identified in ClinVar. Finally, an additional two variants were identified as benign or likely benign in the Leiden Open Variation Database (LOVD[37]) (Table S3).

Pathogenic variants were first ascertained from HGMD (v.2023.4), ClinVar, and LOVD and subsequently filtered to include only those with an MAF <3.33E-05 in gnomAD (in both v2.1.1and v3.1.2). A total of 54 variants were identified in HGMD with a disease-causing (DM) classification. An additional 10 variants were included from ClinVar with a pathogenic or likely pathogenic classification, with one variant listed as likely pathogenic for ALGS specifically among variants classified as conflicting. Finally, six variants were identified as pathogenic or likely pathogenic in LOVD (Table S3).

Statistical framework for variant scoring and classification

Results from sequencing were calibrated and annotated to translate quantitative scores into normal or abnormal function predictions as previously described.[38] For each variant, a weighted and normalized membrane expression score was calculated based on the raw count of how many times a variant was observed in each bin following FACS. First, a variant frequency within bin ($F_{b,v}$) was calculated for each variant as the raw count of each variant for each bin divided by the sum of all counts recorded in that bin, where M was the number of variants with index v for a given bin, b, with $C_{b,v}$ counts of a specific variant in a bin.

$$F_{b,v} = \frac{C_{b,v}}{\sum_{v}^{M} C_{b,v}}$$

Second, the $F_{b,v}$ for each variant was weighted (W_v). The APC low bin was assigned the value 0.1, and the APC high bin was assigned the value 0.9 ($w_1 = 0.9$ and $w_2 = 0.1$).

$$W_v = \frac{\sum_{b}^{B} w_b F_{b,v}}{\sum_{b}^{B} F_{b,v}}$$

Third, the W_v was normalized using experimental data from positive and negative controls to obtain a within replicate variant score (S_v). The W_{WT} calibration was the mean W_v over three calibration variants (chr20:10652540-C-A, chr20:10673486-G-C, and chr20:10673484-A-G [Hg38/GRCh38]), and the W_{mutant} was the mean W_v over the calibration variant (chr20:10673484-A-C [Hg38/GRCh38]). Calibration variants were selected based on assay performance of pure cell lines (Figures S4 and S5).

$$S_v = \frac{W_v - W_{mutant}}{W_{WT} - W_{mutant}}$$

S_v values from replicate runs were averaged to calculate a final membrane expression score (S_α). A minimum of two replicates were required, and confidence intervals (CIs) were calculated using the equation $S_{\alpha,CI} = S_\alpha \pm Z_{97.5} se_v$ where standard error (se_v) is the standard deviation (σ_α) of membrane expression scores divided by the square root of the number of replicates: $se_v = \sigma_\alpha / \sqrt{R_v}$.

A set of known/predicted benign ($n = 111$) and pathogenic ($n = 70$) variants (variant control set) were used to establish thresholds for abnormal and normal function (Table S3). The lowest scoring 5th percentile benign variants defined the upper bound of the functionally abnormal threshold. Variants with both a mean and an upper CI below the threshold were classified as functionally abnormal, and variants with a mean below the threshold and an upper CI crossing the threshold were classified as likely to be functionally abnormal.

Odds of pathogenicity (OddsPath) calculations have been previously described.[39,40] Briefly, OddsPath were calculated where P_1 was the proportion of pathogenic variants in the variant control set, $\frac{total\ variant\ control\ set\ pathogenic}{total\ variant\ control\ set\ pathogenic + total\ variant\ control\ set\ benign}$, and P_2 was the proportion of pathogenic variants in the variant control set with abnormal membrane expression scores out of all abnormal variants, $\frac{total\ variant\ control\ set\ pathogenic\ abnormal}{total\ variant\ control\ set\ abnormal}$. Conversion of OddsPath calculations to American College of Medical Genetics and Genomics (ACMG)/American Association of Molecular Pathology (AMP) evidence weight was previously described.[39]

$$OddsPath = \frac{[P_2 * (1 - P_1)]}{[(1 - P_2) * P_1]}$$

Structural modeling

The structure of human JAG1 (C2-EGF3) in complex with human NOTCH2 (EGF8-12) was reconstituted in AlphaFold2 (DeepMind) using MMseqs2 (ColabFold v1.5.5) and aligned to the crystal structure of rat JAG1 in complex with NOTCH1 (PDB: 5UK5). All structural figures were generated from PyMOL (PyMOL version 2.5.4, Schrödinger). Protein sequences were retrieved from UniProt.[41] SignalP 5.0 (Department of Health Technology, Denmark) was used to predict SP probability using aa 1–70 of human JAG1 (P78504) comparing WT (GenBank: NM_000214.3), c.58C>A (GenBank: NM_000214.3) (p.Leu20Met), and c.59T>G (GenBank: NM_000214.3) (p.Leu20Arg) variants.[42] SP prediction data was visualized with GraphPad Prism 9.5.1 (GraphPad Software). AlphaFold2 (DeepMind) using MMseqs2 (ColabFold v1.5.5) was used to generate predicted template modeling (pTM) scores for human JAG1 with various cysteines mutated. Secondary structures of JAG1 were retrieved from PDBsum (EMBL-EBI, United Kingdom) based on human JAG1 (PDB: 4CBZ). Buried and exposed residues were analyzed using the apo structure of human JAG1 with PDBePISA (EMBL-EBI, United Kingdom).

AlphaMissense and CADD analysis

AlphaMissense (DeepMind) scores[43] for JAG1 were downloaded from the Hege Lab web resource.[44] AlphaMissense scores were plotted for all missense variants with abnormal or likely abnormal membrane expression ($n = 435$) and for all missense variants ($n = 2194$). CADD scores from GRCh38 v1.6 were downloaded and plotted similarly. Data were analyzed using an unpaired t test for statistical significance.

Curation of variants from ClinVar and literature for reclassification

Disease-associated *JAG1* missense variants were identified from ClinVar, with a last check date of March 1, 2024, and from HGMD v.2023.4 (DM characterized as pathogenic and DM? characterized as likely pathogenic). Previously classified variants were also identified from Gilbert et al.[15]

Results

Establishing a high-throughput JAG1 cellular localization assay

To test the impact of *JAG1* nucleotide substitutions on protein localization, we designed an assay to discriminate between membrane and intracellular expression of JAG1. Membrane expression of JAG1 is required for its interaction with the NOTCH2 receptor, and intracellular sequestration of missense variants has been identified as a mechanism of defective JAG1 function.[15,16,18,20,23] Immunofluorescence of stable cell lines confirmed the previously observed intracellular retention of three JAG1 pathogenic variants (c.110T>C [p.Leu37Ser], c.2078G>A [p.Cys693Tyr], and c.2732G>A [p.Cys911Tyr] [GenBank: NM_000214.3])[15,18,23] compared with WT JAG1 (Figures 1A–1D). To design a high-throughput screen, we sorted these cells by flow cytometry using an antibody against the extracellular domain of JAG1 and an APC-conjugated secondary antibody, with increased fluorescence intensity indicating JAG1 membrane expression. All *JAG1* constructs included an IRES-eGFP construct to ensure selection of positively transformed cells (Figure S1). Nearly 80% of cells overexpressing WT JAG1 were GFP+ (Figure S2), indicating successful incorporation of the overexpression construct. Cells overexpressing WT JAG1 were clearly distinguishable from all three pathogenic variants (96.8% of WT cells compared to 6.7% of c.2078G>A [p.Cys693Tyr], 0.14% of c.110T>C p.Leu37Ser], and 0.68% of c.2732G>A [p.Cys911Tyr]) by flow cytometry (Figure 1E). Of note, the c.2078G>A (p.Cys693Tyr) mutant was previously identified to retain partial membrane expression.[15] With our assay, we observe an increase in the percentage of JAG1 present on the membrane of c.2078G>A (p.Cys693Tyr) cells compared to the c.110T>C (p.Leu37Ser) and c.2732G>A (p.Cys911Tyr) lines (6.7% compared to 0.14% and 0.68%, respectively), supporting this prior finding. When these four cell lines are combined in an equal ratio, a bimodal distribution of fluorescence corresponding to

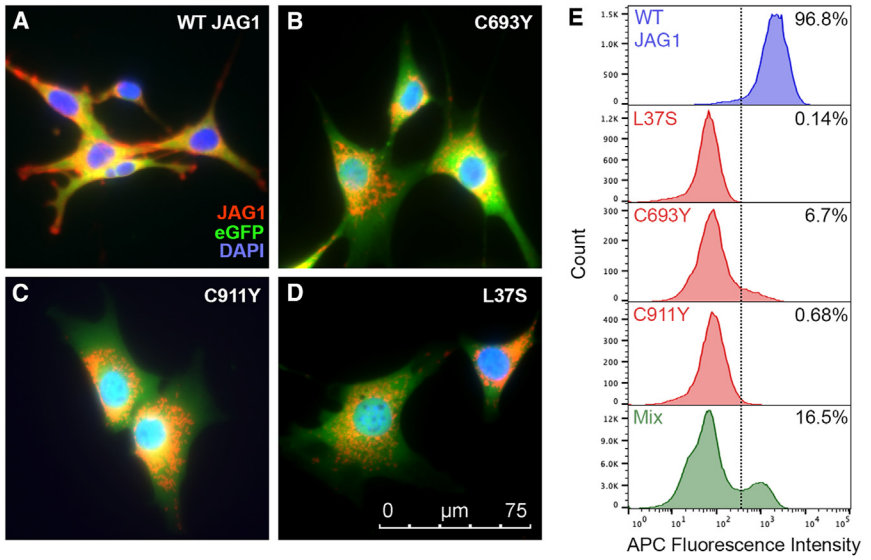


Figure 1. Development of an assay for JAG1 membrane expression

(A–D) WT JAG1 overexpressed from a construct containing an IRES-eGFP cassette (A) shows membrane expression while the JAG1 pathogenic variants (B) Cys693Tyr (C693Y), (C) Cys911Tyr (C911Y), and (D) Leu37Ser (L37S), show intracellular expression of JAG1.

(E) WT and mutant JAG1 cell lines were sorted by flow cytometry using an antibody against the extracellular region of JAG1. WT JAG1 has higher APC fluorescence, indicating membrane expression of JAG1, compared to all three mutants. A mixture of all four cell lines (Mix) shows a bimodal distribution of fluorescence intensity with peaks representing JAG1 non-membrane (APC low) and membrane (APC high) expression. Dotted line represents the lower threshold for the APC high population.

WT JAG1 (high APC) and mutant JAG1 (low APC) is observed, supporting the utility of this assay to clearly distinguish between membrane and non-membrane JAG1 expression (Figure 1E).

Generation and functional characterization of a *JAG1* variant library

To comprehensively characterize the function of all disease-relevant variants in the N terminus of JAG1, we designed an SSVL targeted to include every possible nucleotide permutation in *JAG1* exons 1–7 ($n = 2{,}907$; nucleotides 1–1,008; aa 1–336) (Figure 2A). Since missense variants in ALGS are caused by single-nucleotide substitutions, aa variants produced by the change of two nucleotides within a codon were excluded. All synonymous changes were retained. Of a possible 2,907 variants, 2,868 (98.7%) passed quality control metrics for the linear library build. This SSVL was expressed in NIH3T3 cells from a plasmid containing an IRES-eGFP cassette for the selection of positively transformed cells (Figure S1). Sequencing of the final cell library indicated the presence of 97.2% of targeted nucleotide permutations ($n = 2{,}825$ variants; Table S2).

Five variants falling within a leucine-rich region failed quality control during the linear library build (c.41T>C [p.Leu14Pro], c.41T>G [p.Leu14Arg], c.47T>C [p.Leu16Pro], c.47T>G [p.Leu16Arg], and c.45C>G [p.Ser15Arg] [GenBank: NM_000214.3]). However, their location within the SP and the association of variation at nearby residues (Ser10, Leu17, Leu18, Leu20, Leu21) with ALGS suggested this region was important to study. Along with two other failed variants at an aa residue with a contradictory role in disease (c.814G>A [p.Val272Ile] and c.814G>T [p.Val272Phe] [GenBank: NM_000214.3]),[26,45] these seven variants were generated separately as pure stable cell lines. Stable cell lines were spiked into the *JAG1* library at a ratio to allow for equal representation of all variants, resulting in a final library of 2,832 variants (97.4% of targeted nucleotide permutations), including 2,194 missense and 638 synonymous variants (Table S2).

The JAG1 SSVL was sorted using WT JAG1 and c.47T>G (p.Leu16Arg) pure cell lines as gating controls for membrane (APC high) and non-membrane (APC low) expression, respectively. The c.47T>G (p.Leu16Arg) cell line shows markedly reduced membrane expression compared to the c.110T>C (p.Leu37Ser) cell line with little overlap with WT JAG1 cells and was therefore selected for use as a non-membrane control (Figure S3). The JAG1 variant library exhibited more diffuse expression of GFP than either pure cell line (WT JAG and c.47T>G [p.Leu16Arg]), therefore, untransfected NIH3T3 cells were used to define the lower boundary of the GFP+ gate for selection of positively transfected cells (Figure S4A). Gating for APC fluorescence was done conservatively using the WT JAG1 and c.47T>G (p.Leu16Arg) controls by omitting a small population of cells with intermediate APC fluorescence when creating the lower boundary of the APC high gate and the upper boundary of the APC low gate (Figures 2B and S4B–S4E). Using this gating strategy, two populations of cells were collected corresponding to low and high APC fluorescence. Of WT JAG1 GFP+ cells, nearly 90% were included in the APC high population compared to <1% of c.47T>G (p.Leu16Arg) GFP+ cells and ~50% of the JAG1 library, highlighting two clear populations of cells with (APC high) and without (APC low) JAG1 membrane expression (Figures 2B and S4B–S4E).

Amplicon-based long-read sequencing was used to quantify the frequency of each variant in both the high and low APC populations. A weighted and normalized membrane expression score was calculated for each variant, which ranged from roughly 0 to 1, with lower scores indicating non-membrane expression and higher scores indicating membrane expression (Figures 2C–2E).

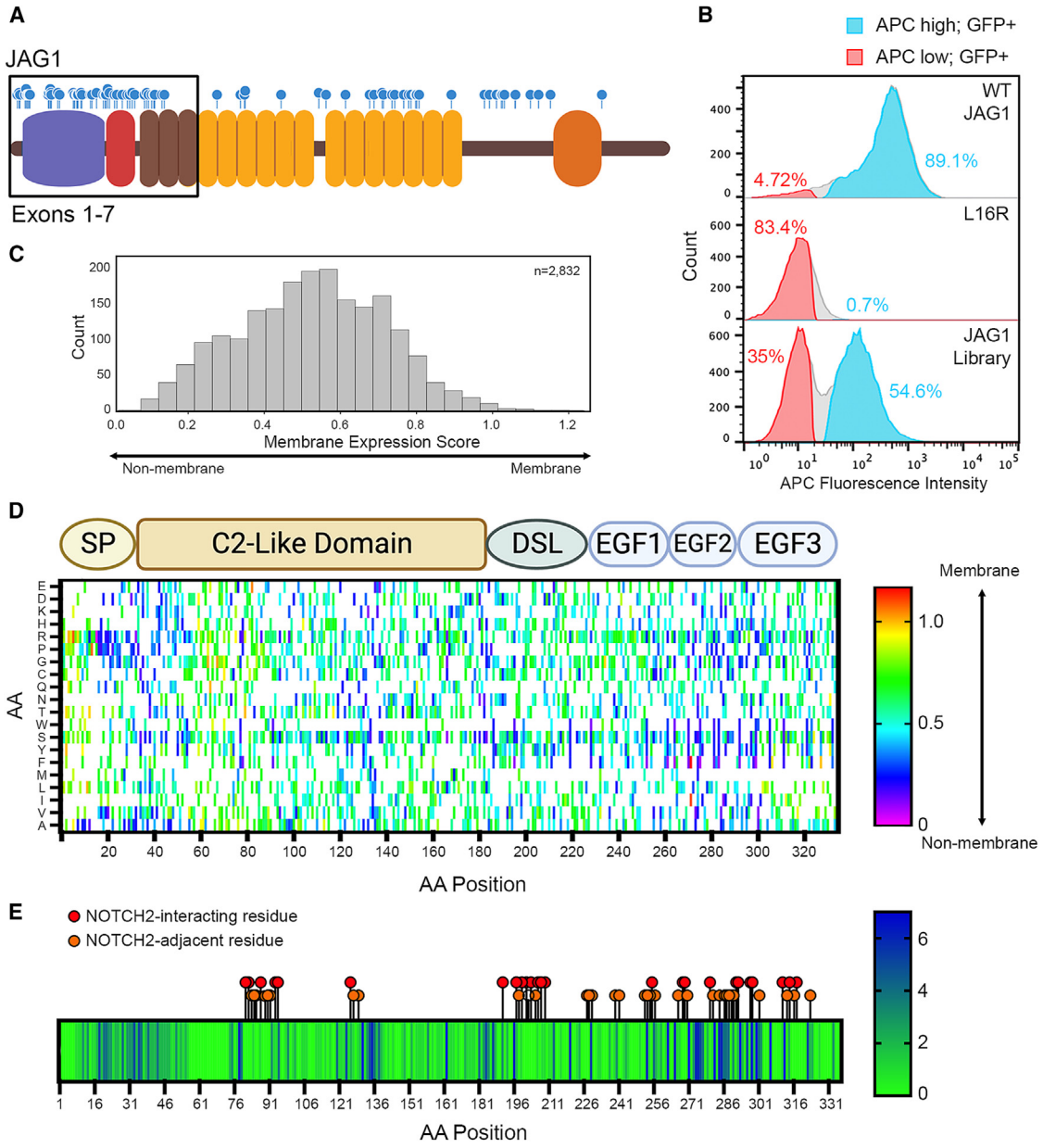


Figure 2. High-throughput functional characterization of a *JAG1* variant library

(A) Schematic of the JAG1 protein highlighting the SSVL targeted region. Blue lollipops indicate previously reported missense variants. Protein domains are color-coded: purple: C2-like; red: Delta/Serate/Lag2 (DSL); brown: NOTCH2-interacting epidermal growth factor-like (EGF-like) repeats; yellow: EGF-like repeats; orange: transmembrane.

(B) Sorting strategy for membrane expression assay. Membrane (APC high; GFP+; blue) and non-membrane (APC low; GFP+; red) populations were gated using WT JAG1 and p.Leu16Arg (L16R) as controls. Red and blue coloration indicates the collected populations. Cells with intermediate APC fluorescence (light gray) were not collected.

(C) Distribution of membrane expression scores across the entire SSVL.

(D) Heatmap showing tolerance (yellow/green) or intolerance (blue/purple) of each JAG1 aa substitution. White indicates variants that were not tested in the assay.

(E) Heatmap showing the tolerance of each aa residue to change. Up to seven alternate variants were assayed per residue, and the number of alternate residues with abnormal membrane expression is plotted. Dark blue indicates sites that are intolerant to variation, green indicates sites that are fully tolerant to variation, and intermediate shades indicate sites with some, but not all, changes disrupting membrane expression. NOTCH2-interacting residues are indicated by red (directly interacting) and orange (likely interacting) lollipops.

Final scores were obtained by combining and analyzing results from seven replicate experiments with over 2 million cells collected from each experiment, ensuring that each variant was represented ~700 times on average (Table S1).

Structural modeling to validate variant functional effects

Analysis of secondary structure shows an enrichment for missense variants with low membrane expression within the α-helix domain compared to the loop and β-strand

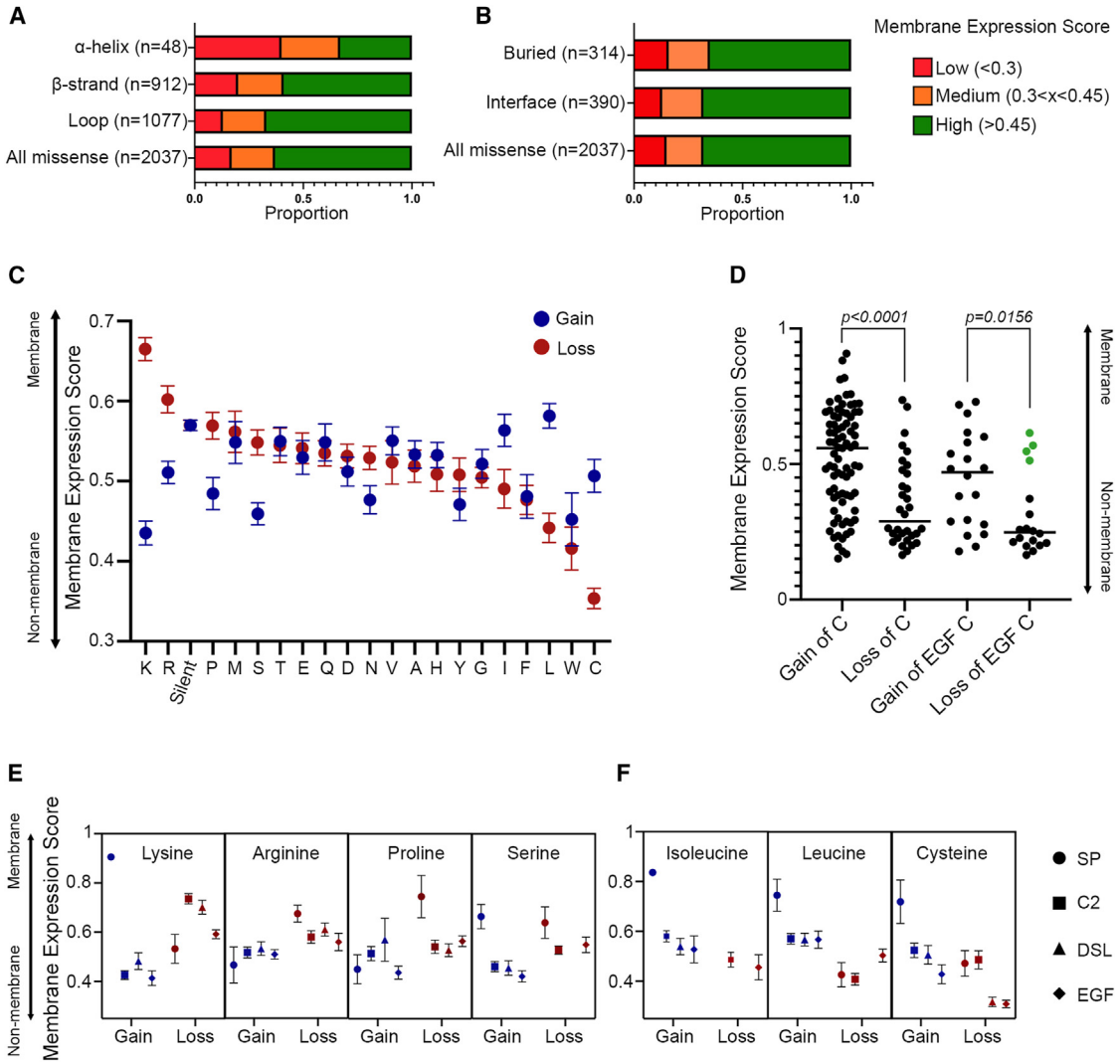


Figure 3. Specific aa residues show sensitivity to either gain or loss

(A and B) The proportion of variants with low, medium, or high membrane expression scores was plotted for (A) secondary structure and (B) buried and interface residues.

(C) Membrane expression scores for aa gain (blue) or loss (red) are plotted. Scores are averaged for each aa across the region.

(D) Membrane expression scores plotted for cysteine gain versus cysteine loss across the entire region assayed and just the EGF-like domains. Data points for cysteine loss represent the average score for all variants at that residue. Green data points indicate four cysteines in the EGF domain that are tolerant to variation. Data were analyzed using an unpaired t test.

(E) Membrane expression scores for aa that are most intolerant to gain are plotted across four regions.

(F) Membrane expression scores for aa that are most intolerant to loss are plotted across four regions.

All error bars represent SEM.

domains (Figure 3A). No differences in membrane expression scores were identified for buried or interface variants compared to all missense variants (Figure 3B).

We analyzed the global effects of single aa loss or gain across the entire region and found that gain of a lysine, arginine, proline, or serine was associated with lower membrane scores (failure to reach the membrane) compared to loss of these aa from WT sequence (Figure 3C). Conversely, loss of an isoleucine, leucine, or cysteine was associated with lower membrane scores than gain of these aa through missense substitution (Figures 3C and 3D). Prior data support a greater intolerance for cysteine loss compared to cysteine gain, particularly within the EGF-like domains.[15,23,46–49] There are three EGF-like domains within our region of study, and each domain contains six cysteines. Surprisingly, the first four cysteines of EGF-like domain 1 were highly tolerant to change in our assay, whereas the remaining 14 cysteines were expectedly intolerant to variation (Figure 3D; Table S2). To better resolve the differences between these cysteines, we modeled the effects of variation at these residues using AlphaFold.[50] When each of these cysteines is changed to a tryptophan, the pTM score, a scoring metric used to evaluate confidence in the predicted protein fold, is similar to WT JAG1 for three of the first four EGF-like cysteines (Cys234, Cys238, and Cys245), supporting our membrane

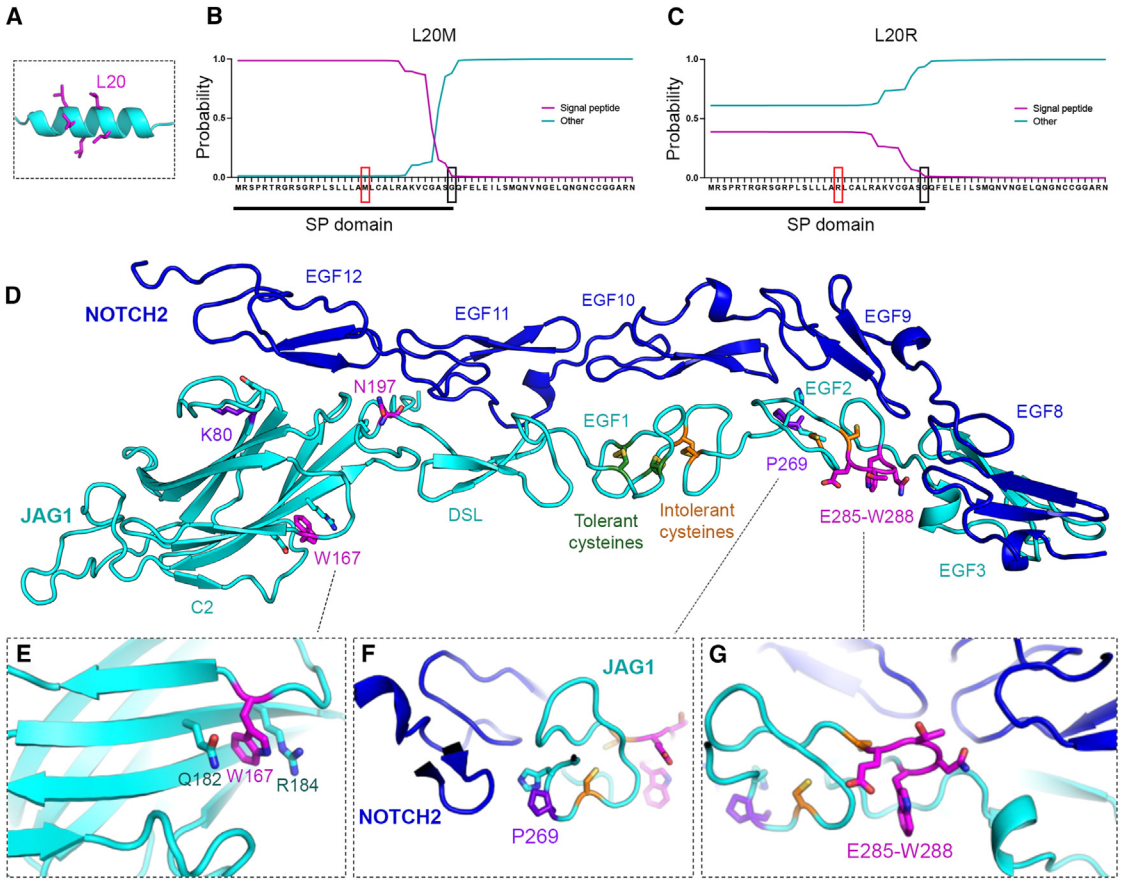


Figure 4. JAG1 structural modeling agrees with membrane expression scores

(A) Structural modeling of Leu16–18 (L16–18) and Leu20–21 (L20–21) within the SP helix.

(B and C) Phobius SP prediction of the (B) Leu20Met (L20M) and (C) Leu20Arg (L20R) variants show a decrease in the confidence of the SP signature for Leu20Arg (L20R) but not Leu20Met (L20M). Red box indicates location of variant. Black box indicates the last residue of the SP domain.

(D) Structural modeling of the JAG1 region of study in contact with NOTCH2. Residues of interest are highlighted, including the four tolerant and 14 intolerant EGF-like cysteines.

(E–G) Structural analysis of (E) Trp167 (W167) packed tightly between a glutamine and an arginine, (F) Pro269 (P269), and (G) hotspot residues Glu285 (E285), Thr286 (T286), Asn287 (N287), and Trp288 (W288) that are highly sensitive to variation.

expression data and indicating that changes at these cysteines may not be critical for protein folding (Table S4).

To test whether intolerance to aa gain is specific for functional domains, we plotted membrane expression scores for lysine, arginine, proline, and serine substitutions occurring in each of the four JAG1 domains. Membrane expression scores for arginine and proline gains are low across all regions (Figure 3E). Low scores for lysine and serine substitutions are driven by a greater intolerance for variation in the C2-like, DSL, and EGF domains, whereas substitution in the SP is more tolerated (Figure 3E). Similarly, we plotted membrane expression scores for isoleucine, leucine, and cysteine substitutions across all four functional domains. Here, leucine loss was more tolerated in the EGF domain compared to the C2-like and DSL domains, whereas cysteine loss was most intolerant in the DSL and EGF domains (Figure 3F).

We used structural modeling to further investigate specific sites that were highly intolerant to variation in our membrane expression assay. Within the SP, there are a series of leucines with abnormal membrane expression scores (Leu18, Leu20, Leu21). We predicted the effects of SP cleavage resulting from mutations at Leu20 using Phobius, a computational tool that predicts SP cleavage sites and probability.[51] This analysis determined that mutation of Leu20 to an arginine, but not a methionine, reduces the probability of SP cleavage, which would be predicted to interfere with trafficking of the protein to the cell surface (confidence scores: c.59T>G [p.Leu20Arg]: 0.3876; c.58C>A [p.Leu20Met]: 0.987; WT JAG1: 0.99) (Figures 4A–4C). This finding is in striking agreement with our membrane expression data where a change to an arginine yields a low membrane expression score (0.201) while a change to methionine is well tolerated with a score (0.812) similar to WT. Concurrently, c.59T>G (GenBank: NM_000214.3) (p.Leu20Arg) has been reported as an ALGS-associated pathogenic variant in the mutation database HGMD, while c.58C>A (GenBank: NM_000214.3) (p.Leu20Met) is absent from both HGMD and ClinVar. Importantly, of the 177 aa residues

that are sensitive to variation in our assay, only eight (4.5%) are fully intolerant to membrane expression, with every tested substitution resulting in intracellular sequestration. The remaining 169 residues display divergent membrane expression, where only specific substitutions result in abnormal membrane expression while the remainder are tolerated (Figure S6; Table S2).

To analyze the effects of additional variants, we generated a structural model of JAG1 bound to NOTCH2 using AlphaFold (Figure 4D). The model had a pTM score of 0.777 and an interface pTM score of 0.795, indicating that the structure was predicted with a high degree of confidence. Aligning the JAG1-NOTCH2 model to the published crystal structure of rJAG1-NOTCH2 (PDB: 5UK5) showed that the binding interface was fully conserved between our model and the experimentally determined complex structure.[33] This JAG1-NOTCH2 model was used to visualize the position of Trp167, a residue located within the DSL domain that is intolerant to all variation in our membrane expression assay. The c.499T>C (GenBank: NM_000214.3) (p.Trp167Arg) variant has been classified as a VOUS in ClinVar, and our data suggest this should be changed to likely pathogenic. Tryptophan is very hydrophobic and is tightly packed between a glutamine and an arginine (Gln182 and Arg184), making it critical for the structural integrity of the DSL domain (Figures 4D and 4E). A change from tryptophan to any other variant tested in our assay (Cys, Leu, Arg, Gly, and Ser), which are all small or hydrophilic residues, would disrupt this region, likely resulting in lack of membrane expression due to domain misfolding or protein aggregation. Similarly, we assessed the structural effects of changes to aa 285–288, a region within the EGF-like domain where variation results in abnormal membrane expression. Structural modeling of this region, consisting of Glu285, Thr286, Asn287, and Trp288, indicates strong intradomain interactions between these aa, with variation likely to negatively affect protein stability (Figures 4D and 4G). Based on the structure of the NOTCH2-JAG1 complex, these residues are predicted to be in proximity to key NOTCH2 interface residues, and it is likely that changes here could affect NOTCH2 binding as well.[33]

We hypothesized that the functional effects of variants located within known NOTCH2-interacting regions could be missed in our assay if variation affects NOTCH2-binding but not protein folding/cellular trafficking. To test this, we visualized the positions of two residues, Pro269 and Lys80, where missense variants have been reported in patients with ALGS but that had high membrane expression scores in our assay, indicating tolerance to variation. Structural analysis revealed that these two residues directly contact NOTCH2 in the model, indicating that variation at these positions is likely to affect binding, and supporting their clinical relevance despite high membrane expression scores in our assay (Figures 4F and S7A). A third variant, c.590A>G (GenBank: NM_000214.3) (p.Asn197Ser), which affects a residue located in the NOTCH2-interacting DSL domain (Figure 2E) and is classified as a VOUS in ClinVar, is predicted to disrupt both NOTCH2-binding and membrane expression (score 0.262) (Figure S7B). Utilization of NOTCH2-binding assays to test these variants may help identify defective binding.

Although we observed changes that were intolerant to membrane expression throughout the entirety of JAG1 exons 1–7, the region from aa 57–73 was highly tolerant to change. This region is located between the SP and the first NOTCH2-interacting residues and thus is likely not critical to NOTCH2-binding nor to protein-folding (Figures 2D and 2E). Correspondingly, there are no pathogenic variants listed in this region in HGMD or ClinVar.

Establishment of a functionally abnormal variant score threshold

To identify a threshold score below which membrane expression could be defined as functionally abnormal, we examined how well our assay results agreed with clinical annotations and variant population frequencies. We defined a variant control set consisting of 111 known or predicted benign and 70 pathogenic variants of which 106 benign and 69 pathogenic variants were scored in our assay (Table S3). Membrane expression scores from benign and pathogenic variants were inversely distributed, indicating that sorting by membrane expression is a good predictor of variant pathogenicity (Figure 5A, upper). Using this variant control set, we established an abnormal score threshold that separates the upper 95% of top-scoring benign variants from the lower 5%. Variants with a mean score and an upper CI below this threshold were categorized as functionally abnormal. Variants with a mean score below this threshold but an upper CI crossing this threshold were defined as likely abnormal (Figure S8). We categorized 486 variants as functionally abnormal (n = 277 abnormal and n = 209 likely abnormal) of which 439 (90.3%) were missense (Figure 5A, lower; Table S2).

Functionally abnormal missense residues correlate well with CADD and AlphaMissense scores, showing significant enrichment for variants that are predicted to be pathogenic by these computational tools (Figures 5B and 5C). Thermodynamic predictions of RNA secondary structure for the lowest-scoring (abnormal) synonymous variant c.624A>T (p.Gly208=) (codon GGA>GGT) indicates a higher base pair probability for the GGT variant compared to the WT variant (GGA), suggesting that the alternate codon results in a more stable 5′ structure (Figure S9). Analysis of a second synonymous variant at the same position, c.624A>G (p.Gly208=) (codon GGA>GGG), with a high membrane expression score (normal/membrane), has a similar base pair probability as the WT codon (GGA) (Figure S9). Alteration to thermodynamic properties of the mature RNA could provide a rationale for the identification of abnormal synonymous variants with our assay via modulation of translation initiation rates.

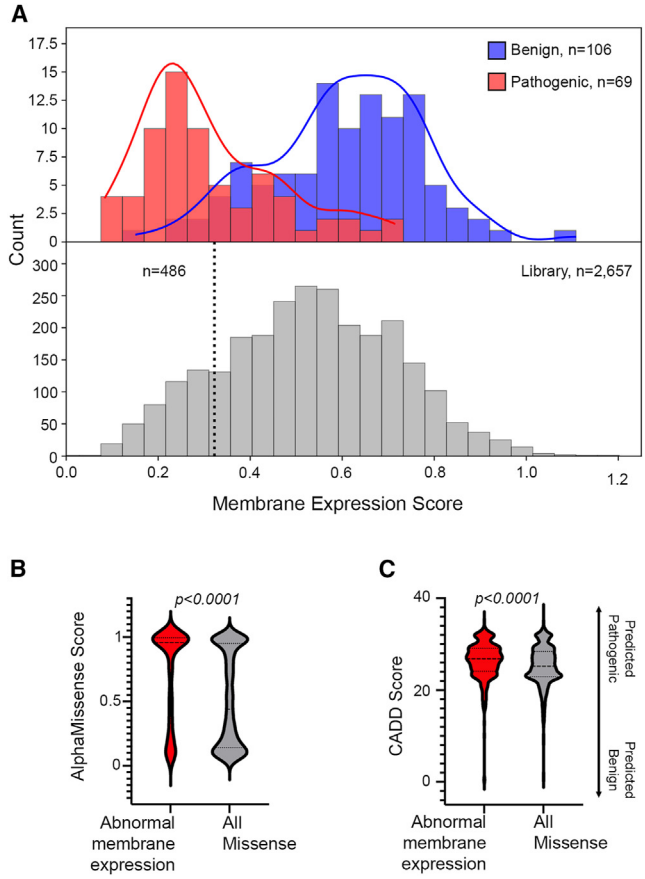


Figure 5. JAG1 membrane expression correlates with pathogenic and benign variant classifications
(A) Histogram of membrane expression scores for 106 benign and 69 pathogenic variants. Black dotted line indicates the threshold for the lower 5% of benign variants, below which variant function is classified as abnormal. Forty-five out of 69 (65.2%) pathogenic variants within the variant control set lie below this line. The distribution of membrane expression scores for the remainder of the library is shown below, with 486 variants scoring below the abnormal threshold.
(B and C) Correlation of (B) AlphaMissense and (C) CADD scores for variants with abnormal membrane expression. Data were analyzed using an unpaired t test.

Defining functional evidence strength to support clinical variant classification

The clear correlation between membrane expression scores and clinical classifications (variant control set) supports the validity of our assay in measuring a disease-relevant functional feature of JAG1. The inclusion of a large set of positive and negative controls that cover the full dynamic range of the assay allows for statistical calculation of assay strength. To empirically determine how well our assay differentiates benign (n = 106) from pathogenic (n = 69) variants, we calculated two odds of pathogenicity (OddsPath) scores, one for variants scoring as abnormal in our assay and a second for those scoring as likely abnormal. OddsPath calculations allow for analysis of the statistical power of a dataset in accurately predicting function for abnormal variants, and scores have been translated to evidence strength levels to allow for application of MAVEs data during clinical variant interpretation.[39] The increased separation between pathogenic and benign controls in the abnormal group resulted in an OddsPath of 24.6, with an OddsPath of 11.5 calculated for variants categorized as having a likely abnormal function (Table S5). Following guidance from the ClinGen Sequence Variant Interpretation (SVI) Working Group, an OddsPath score of 24.6 corresponds to utilization of PS3, a clinical classification criterion for evidence supporting a damaging effect on protein function at strong weight while an OddsPath score of 11.5 corresponds to application of PS3 at moderate weight (PS3_moderate).[39,52] An OddsPath score for benignity (BS3) was not calculated since successful Notch signaling requires both cell surface expression of JAG1 and NOTCH2 binding/signaling initiation, therefore the presence of JAG1 on the membrane is required, but not sufficient, for JAG1 function.

Clinical impact of membrane expression functional data on *JAG1* VOUS reclassification in ALGS patients

Using ACMG clinical variant classification guidelines, we previously classified 38 missense variants identified in an IRB-approved ALGS research study.[15] Of these, 28 variants were included in our MAVEs. We assessed how well

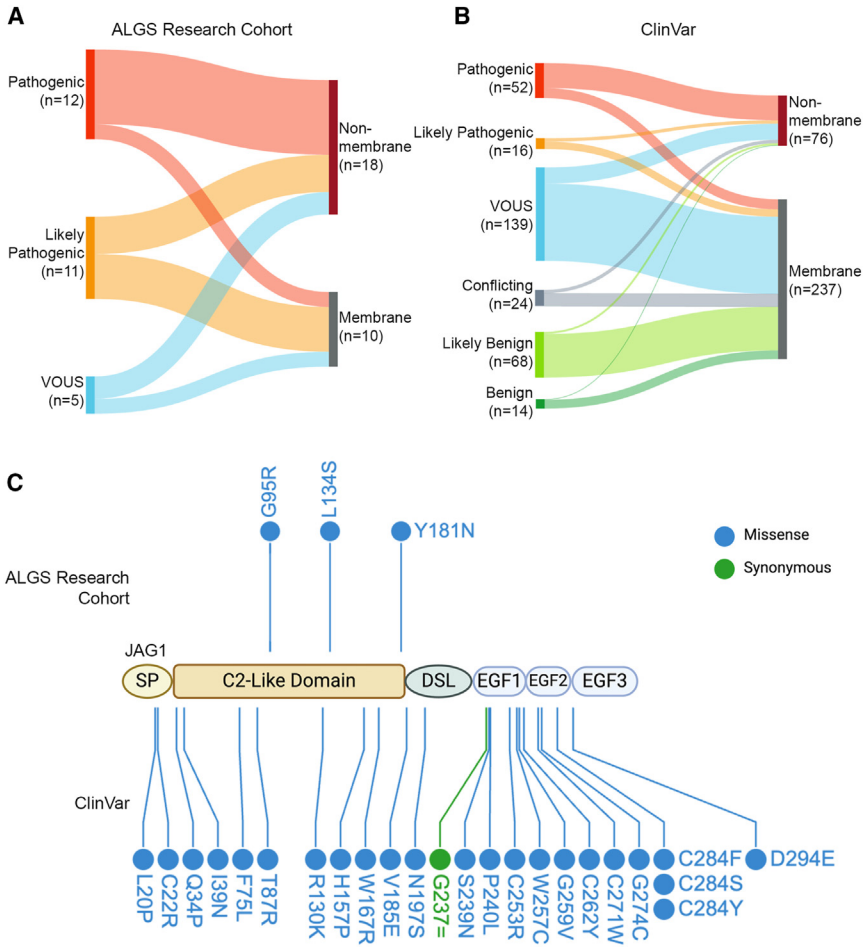


Figure 6. Functional evidence supports *JAG1* VOUS reclassification

(A) Sankey plot showing the percent of previously classified *JAG1* variants from an internal ALGS research cohort with abnormal membrane expression data.

(B) Sankey plot showing the distribution of previously classified variants from ClinVar (all classification categories) and HGMD (DM included with pathogenic classifications, DM? included with likely pathogenic classifications) with membrane expression data.

(C) Analysis of VOUSs with abnormal membrane expression data resulted in reclassification of three VOUSs from the ALGS research cohort to likely pathogenic (upper) and evidence to support the reclassification of 24 VOUSs from ClinVar.

membrane expression scores correlated with our prior classifications and found that the majority of pathogenic (n = 10/12; 83.3%) variants had evidence of non-membrane JAG1 compared to half of likely pathogenic variants (n = 5/11; 45.5%) (Figure 6A). Importantly, a majority of VOUSs also had evidence supporting abnormal membrane expression (n = 3/5; 60%) (Figure 6A). Re-analysis of these three VOUSs (c.283G>C [p.Gly95Arg], c.401T>C [p.Leu134Ser], and c.541T>A [p.Tyr181Asn] (GenBank: NM_000214.3]) with the addition of this functional data resulted in reclassification of all three variants to likely pathogenic, providing diagnostic resolution (Figure 6C; Table S6).

To assess the impact of our functional data on a larger dataset, we reviewed variant classifications from ClinVar. Functional data from our MAVEs was available for 139 missense and synonymous VOUS (Table S7). Upon review, 24 (17.3%) had membrane expression scores indicating they are fully or likely abnormal (Figure 6B). Using ACMG classification criteria, we reanalyzed all 24 ClinVar variants with abnormal membrane expression, resulting in the reclassification of 23 as likely pathogenic (Figure 6C; Table S7). The 115 variants with higher membrane expression scores could not be reclassified with the present data, as lack of membrane expression is only one way to render JAG1 ineffective. Combined with data from our ALGS research study, 26 out 144 VOUSs were reclassified as likely pathogenic (18.1%) (Figure 6C).

Analysis of *JAG1* variants that are associated with other diseases

There are 26 *JAG1* missense variants located in exons 1–7 that have been reported in HGMD to be associated with other diseases, including biliary atresia, miscellaneous cardiac manifestations, congenital anomalies of the kidney and urinary tract, ocular anomalies, pregnancy loss, hypogonadism, hearing loss, and hypothyroidism, that are unrelated to ALGS[22] (Table S8). We hypothesized that the disease pathomechanism for these variants would be different from ALGS-causing variants. As expected, none of the 26 variants had abnormal membrane expression scores indicating that all of them are appropriately expressed on the membrane. These data suggest that these variants are either polymorphisms or act by an alternate disease pathomechanism.

Discussion

Uncertainty in variant classification is a critical problem in genomic diagnostics. Of the over two million variants deposited in ClinVar, 41% are classified as either VOUSs or conflicting, with the VOUS classification undergoing a ~5-fold increase since 2020.[53] Scalable functional assays offer a methodology to interpret the effects of thousands of variants simultaneously, with the advantage of including numerous positive and negative controls (variant control set) to strengthen assay calibration.[54,55] The utility of scalable assays to measure membrane expression is supported by the estimation that up to 30% of all proteins are transmembrane, and generation of assays

such as ours will support improved understanding of this class of proteins.[56]

We describe the development and application of a high-throughput membrane expression assay allowing for measurement of the functional effects of 2,832 nucleotide variants in *JAG1*, a primary cause of ALGS. JAG1 is a transmembrane ligand that interacts with Notch receptors to influence gene expression in a variety of tissues. Therefore, functional JAG1 requires both membrane localization and successful contact with the receptor.[57] With this assay, we identified 486 JAG1 variants with abnormal membrane expression, a known mechanism causing ALGS.[15,17,18,20,23] Functional data are valuable during clinical variant classification, and recent recommendations published by the ClinGen SVI Working Group include guidelines for how to assess the strength of data obtained through MAVEs using OddsPath calculations.[39] Through a Bayesian adaptation of the ACMG variant classification guidelines, numerical OddsPath scores are readily linked to evidence weight (supporting, moderate, and strong) for use during clinical variant classification.[39,40] A rigorous assessment of the utility of OddsPath for clinical use was recently performed on three separate MAVE datasets for BRCA1, TP53, and PTEN, which showed a high predictive value for disease causality for datasets that include many control variants and are able to clearly distinguish abnormal function of pathogenic from benign variants, corresponding to usage of functional data as strong evidence toward pathogenicity.[58] Moreover, a recent workshop to develop consensus recommendations regarding the "Clinical Application of MAVE Data" unanimously agreed that assays should differentiate moderate from strong (or supporting) support for variant classification between variants with different measured effects, supporting the utility of applying different evidence strength for variants based on assay readout.[59] In our MAVEs, we tested a critical functional feature of JAG1, membrane expression, and obtained clear separation of a large set of benign and pathogenic variants, resulting in an OddsPath score of 24.6 for variants with abnormal membrane expression and 11.5 for variants with likely abnormal membrane expression, corresponding to the utilization of this data as strong or moderate evidence, respectively, during clinical variant classification. Review of 144 previously classified VOUSs from ClinVar and published datasets identified 27 with abnormal membrane expression and indicate that inclusion of this evidence during variant reclassification provides significant diagnostic resolution with 26 (18.1%) upgraded to likely pathogenic, demonstrating a high rate of reclassification for variants with abnormal functional data (Figure 6C; Tables S6 and S7).

Our results have multiple implications for improved variant classification for *JAG1*, with potential applicability to other genes. We found that the identification of abnormal function for one aa change at a given residue did not extend to all other substitutions at that location. Rather, there were some aa that were well-tolerated at specific sites while others were not ($n = 169$ of 335 tested residues, 50%), suggesting that loss of the WT aa itself does not drive pathogenicity but rather that pathogenicity is dependent on specific substitutions. We tested this finding by utilizing a software (SignalP 5.0) that predicts the location of SP cleavage sites, which showed that the c.59T>G (GenBank: NM_000214.3) (p.Leu20Arg) variant, which was previously identified in an individual with ALGS and displays abnormal membrane expression in our assay, is very likely to disrupt SP cleavage, while the c.58C>A (GenBank: NM_000214.3) (p.Leu20Met) variant, which has not been previously reported in ALGS and displays normal membrane expression, is not predicted to disrupt SP processing. Observations of discrete functional differences occurring from different aa substitutions is important, particularly since novel missense changes at aa residues where a different missense change has been previously determined to be pathogenic is an ACMG criterion that is given moderate weight during variant classification (criterion PM5).[52] Results from our study suggest that caution should be taken when using this criterion during the classification of *JAG1* variants for ALGS, and possibly for other disease genes, perhaps by modifying the evidence weight from moderate to supporting. To support this using data from our pathogenic variant control set, we identified 19 residues with a single reported pathogenic variant with an abnormal membrane score of which only one occurred at a residue where all other changes were abnormal while two occurred at residues where all other changes were tolerated. For the remaining 16, only 49% of alternative aa changes were abnormal (Table S2). Overall, these analyses highlight a heterogeneity of JAG1 variant effects and suggest that the structure/function relationship in JAG1 is important in understanding the biological significance and diagnostic impact of missense variants.

To further improve clinical variant interpretation for *JAG1*, we also assessed broadly whether certain aa were more likely to result in disease when removed or inserted. We found that gain of a lysine, arginine, proline, or serine was associated with lower membrane expression scores (failure to reach the membrane). Lysine and arginine both contain positively charged side chains that are exposed on the protein surface, allowing for the formation of ionic interactions and hydrogen bonds within the protein, as well as with water molecules, that are important in protein stability.[60] Proline, with its rigid, cyclic structure, is likely to disrupt beta-sheet topology, a common protein secondary structure found throughout JAG1 exons 1–7.[61] A study of the most common disease-causing aa substitutions identified gain of an arginine or proline as the most damaging, which is consistent with our observations.[62] The negative effect of serine gain is less clear, although serine is found at the consensus site for the addition of *O*-linked glycans, and alterations of this conserved signature could have effects on sugar addition and, consequently, protein folding and cellular trafficking.[63]

We found that loss of isoleucine, leucine, or cysteine was associated with lower membrane expression scores (intracellular). Both isoleucine and leucine are branched chain aa, and it is possible that their alteration could disrupt hydrophobic packing, which is crucial to protein stability.[64] Cysteine residues are critical for protein folding, particularly in EGF-like domains, and cysteine loss has been associated with both ALGS and other diseases.[23,46–49] Concurrently, we found that loss of nearly all EGF-like cysteines led to abnormal membrane expression. Analysis of protein structure predictions for three cysteines that were tolerant to variation showed that alterations at these residues yielded models with confidence scores similar to WT, suggesting that these substitutions may not affect cellular localization and supporting our membrane expression findings. These findings reinforce that a biochemical understanding of the JAG1 protein is essential for translating aa variation to functional effect and highlights specific aa alterations that may be of greater significance during clinical variant classification.

In addition to identifying aa alterations with a detrimental effect on membrane localization, we identified a region of 17 aa (aa 57–73) in the C2-like domain that were highly tolerant to all aa substitutions. Consistent with this finding, there have been no reported ALGS-associated variants in this region (HGMD, ClinVar), substantiating the biological significance of this assay. Although we propose that variation in this region is unlikely to cause ALGS, we note that membrane expression of JAG1 is necessary but not sufficient for JAG1 function. A subset of variants with normal membrane expression might have NOTCH2-binding defects that impair their function. Our region of study encompasses the entire NOTCH2-binding domain of JAG1, with interacting residues identified within the C2-like and DSL domains, and EGF-like domains 2 and 3.[33] The identification of two reportedly pathogenic variants from our variant control set, c.238A>G (GenBank: NM_000214.3) (p.Lys80Glu) and c.806C>T (GenBank: NM_000214.3) (p.Pro269Leu), which directly contact NOTCH2, suggests that these variants likely affect NOTCH2-interaction despite not impacting membrane localization. Structural modeling of a third variant, c.590A>G (GenBank: NM_000214.3) (p.Asn197Ser), with a low membrane expression score (intracellular) also showed a likely effect on NOTCH2-interaction, highlighting heterogeneous effects of JAG1 variants. Notably, c.590A>G (GenBank: NM_000214.3) (p.Asn197Ser) is classified as a VOUS in ClinVar, and the data here are expected to upgrade its classification to likely pathogenic.

Previous studies have alluded to mechanistic heterogeneity in JAG1 variant function, with four of 15 (27%) studied variants showing defects in either JAG1 membrane expression or NOTCH2 binding while the remaining 11 (73%) were defective in both functional properties.[15,17,18,20,23] Of the four variants with divergent effects on JAG1 function, two (c.2078G>A [p.Cys693Tyr] and c.2141G>A [p.Cys714-Tyr] [GenBank: NM_000214.3]) have defective membrane expression with normal NOTCH2-binding ability while a third (c.1991G>C [GenBank: NM_000214.3] [p.Cys664Ser]) shows reduced, but not incomplete, membrane expression with normal NOTCH2-binding ability.[15,20,23] The fourth variant, c.821G>A (GenBank: NM_000214.3) (p.Gly274Asp), which is included in our SSVL, also has reduced membrane expression but with defective NOTCH2-binding.[15,17] In our assay, c.821G>A (p.Gly274Asp) has a low membrane expression score; however, it is above the threshold set for abnormal, which likely indicates reduced membrane expression.[17] These discrete functional differences give insight into the complex biology underlying the structure/function relationship for JAG1. Moreover, this highlights a limitation of our assay (measurement of only one of two critical functions of JAG1), providing rationale for the presence of pathogenic variants with membrane expression scores that are higher than our abnormal threshold. Additional limitations, including an inability of our assay to identify effects from cryptic splice variants due to its utilization of cDNA overexpression and an inability to identify variants that are only defective in NOTCH2-binding, oppose utilization of ACMG evidence toward a benign effect on protein function (BS3) during clinical classification for variants that do not score as abnormal. Additional MAVEs designed to test the NOTCH2-binding ability of JAG1 variants and/or expansion of the membrane expression assay to include additional regions of JAG1, thereby incorporating more controls, will increase the functional resolution of variants.

We were surprised by the number of synonymous changes that resulted in abnormal ($n = 16$; 2.5%) or likely abnormal ($n = 31$; 4.9%) membrane expression ($n = 638$ total). Rates of 12%–15% of synonymous variants with abnormal function from other MAVEs have been reported,[65,66] but these are from assays using minigenes that were able to detect splicing effects. Our lower rate of 7.4% abnormal synonymous variants likely reflects the fact that our cDNA assay is not able to capture abnormal splicing events and suggests that effects beyond splicing could impact protein function for synonymous variants. RNA secondary structure can impact protein function through various capacities, including effects on splicing, translation efficiency, mRNA abundance, and protein folding.[67,68] Utilization of a non-optimal codon ("codon bias") for a given aa can alter the speed of translation, which is often coupled to protein folding.[69] Similarly, the folding energy of RNA can impact translation speed, with more stable RNAs requiring more energy to unfold before translation initiation.[69,70] Thermodynamic modeling of the lowest scoring synonymous variant in our assay (c.624A>G [p.Gly208=]; codon GGA>GGT) suggested there was a difference in the base pair probability of the 5′ terminus compared to utilization of the WT codon, suggesting that codon optimality may impact downstream JAG1 protein folding and could be a pathomechanism for some synonymous variants. A recent report identified

over 50 human diseases that are associated with synonymous variants.[71] To date, there has been only one reported synonymous *JAG1* variant associated with ALGS that is predicted to affect splicing,[19] although nine synonymous VOUSs have been reported in ClinVar. Of these nine, only one had likely abnormal expression in our assay (c.711C>T [GenBank: NM_000214.3] [p.Gly237=]). Given the paucity of ALGS disease-causing synonymous variants, we elected to keep this variant as a VOUS rather than upgrade it to likely pathogenic. These data suggest that additional studies should be performed to resolve the effect of synonymous change on JAG1 function and disease.

The diagnostic impact of our variant effect data is evident by its consequence on the reclassification of 26 out of 144 VOUSs, identified through our phenotypically curated ALGS research cohort and ClinVar, to likely pathogenic. The high incidence of the remaining ClinVar VOUSs with normal membrane expression data (n = 115; 82.7%) likely highlights both the limitations of our assay as well as the presence of benign variants within this classification due to the phenotypic diversity and frequent misclassification of this population. Indeed, reanalysis of 26 *JAG1* variants implicated in other diseases, which included biliary atresia (n = 4, 15.4%), the most common cause of liver disease in children, showed that all had normal membrane expression scores, supporting the utility of strict clinical phenotyping when interpreting *JAG1* variants.

Conclusions

The variant effect data for JAG1 membrane expression presented here will improve ALGS diagnostics. Extension of our membrane expression workflow will have general applicability to better understand the ~5,000 transmembrane proteins that are associated with disease.[56] The translation of our data into useable functional evidence and our analysis on the structure/function differences associated with aa gain versus loss provides critical guidance for clinical variant interpretation. Moreover, we have proposed recommendations for both reconsidering the weight of the ACMG classification criterion applied for novel missense changes at aa residues where a different pathogenic missense variant has previously been seen (criterion PM5) and for investigating a possible role for synonymous variants in disease pathogenesis. Although we highlight an immediate effect on 26 previously classified ALGS-associated VOUSs, novel missense variants are continuously identified, and typically classified as uncertain, thus we expect this to be a gross underestimate in the clinical actionability of our dataset. Utilization of this dataset as a prospective repository of functional evidence will enhance clinical variant classification of novel missense variants when they are first detected in children with ALGS, improving clinical care and reducing the distress that accompanies an uncertain diagnosis.

Data and code availability

The dataset supporting the conclusions of this article is included within the supplemental information of the article (Table S2). Additionally, data were uploaded to MaveDB: https://www.mavedb.org/#/, a public repository for datasets from MAVEs[72] (MaveDB: urn:mavedb:00001198-a). Code for the project is available at https://github.com/tris-10/JAG1-Membrane-Expression-Assay.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2024.06.011.

Acknowledgments

The authors wish to thank Avery Zucco in the Division of Genomic Diagnostics at CHOP and the Flow Cytometry Core at CHOP for technical assistance. We also thank members of the Atlas of Variant Effects (AVE) Alliance for useful discussion on data analysis. This project was funded by internal funding from the CHOP Research Institute (M.A.G.), the Fred and Suzanne Biesecker Pediatric Liver Center (N.B.S.), the Sigrid Jusélius Foundation (D.A.), R01-DK134585-01A1 (N.B.S.), and R35GM133482 (V.C.L).

Author contributions

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

Declaration of interests

The authors declare no competing interests.

Received: April 10, 2024
Accepted: June 24, 2024
Published: July 22, 2024

Web resources

AlphaMissense, https://alphamissense.hegelab.org/
ColabFold v1.5.5, https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb
Ensembl VEP, https://useast.ensembl.org/info/docs/tools/vep/index.html
GATK (v4.3) DepthofCoverage, https://gatk.broadinstitute.org/hc/en-us/articles/9570475259291-DepthOfCoverage-BETA
gnomAD, https://gnomad.broadinstitute.org/
JVarkit, https://lindenb.github.io/jvarkit/JvarkitCentral.html.
Lima, https://lima.how/
pbmm2, https://github.com/PacificBiosciences/pbmm2
PDBePISA (EMBL-EBI), https://www.ebi.ac.uk/pdbe/pisa/
PDBsum (EMBL-EBI), https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/
PyMOL (v.2.5.5), https://storage.googleapis.com/pymol-storage/installers/index.html

SignalP 5.0, https://services.healthtech.dtu.dk/services/SignalP-5.0/
SpliceAI, https://github.com/Illumina/SpliceAI
UniProt, https://www.uniprot.org/

References

1. Alagille, D., Odièvre, M., Gautier, M., and Dommergues, J.P. (1975). Hepatic ductular hypoplasia associated with characteristic facies, vertebral malformations, retarded physical, mental, and sexual development, and cardiac murmur. J. Pediatr. *86*, 63–71.
2. Crosnier, C., Lykavieris, P., Meunier-Rotival, M., and Hadchouel, M. (2000). Alagille syndrome. The widening spectrum of arteriohepatic dysplasia. Clin. Liver Dis. *4*, 765–778.
3. Emerick, K.M., Rand, E.B., Goldmuntz, E., Krantz, I.D., Spinner, N.B., and Piccoli, D.A. (1999). Features of Alagille syndrome in 92 patients: frequency and relation to prognosis. Hepatology *29*, 822–829. https://doi.org/10.1002/hep.510290331.
4. Saleh, M., Kamath, B.M., and Chitayat, D. (2016). Alagille syndrome: clinical perspectives. Appl. Clin. Genet. *9*, 75–82. https://doi.org/10.2147/TACG.S86420.
5. Spinner, N.B., Colliton, R.P., Crosnier, C., Krantz, I.D., Hadchouel, M., and Meunier-Rotival, M. (2001). Jagged1 mutations in alagille syndrome. Hum. Mutat. *17*, 18–33. https://doi.org/10.1002/1098-1004(2001)17:1<18::AID-HUMU3>3.0.CO;2-T.
6. Watson, G.H., and Miller, V. (1973). Arteriohepatic dysplasia: familial pulmonary arterial stenosis with neonatal liver disease. Arch. Dis. Child. *48*, 459–466.
7. 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.
8. Dhorne-Pollet, S., Deleuze, J.F., Hadchouel, M., and Bonaïti-Pellié, C. (1994). Segregation analysis of Alagille syndrome. J. Med. Genet. *31*, 453–457.
9. Elmslie, F.V., Vivian, A.J., Gardiner, H., Hall, C., Mowat, A.P., and Winter, R.M. (1995). Alagille syndrome: family studies. J. Med. Genet. *32*, 264–268.
10. Izumi, K., Hayashi, D., Grochowski, C.M., Kubota, N., Nishi, E., Arakawa, M., Hiroma, T., Hatata, T., Ogiso, Y., Nakamura, T., et al. (2016). Discordant clinical phenotype in monozygotic twins with Alagille syndrome: Possible influence of non-genetic factors. Am. J. Med. Genet. *170A*, 471–475. https://doi.org/10.1002/ajmg.a.37429.
11. Kamath, B.M., Bason, L., Piccoli, D.A., Krantz, I.D., and Spinner, N.B. (2003). Consequences of JAG1 mutations. J. Med. Genet. *40*, 891–895.
12. Kamath, B.M., Krantz, I.D., Spinner, N.B., Heubi, J.E., and Piccoli, D.A. (2002). Monozygotic twins with a severe form of Alagille syndrome and phenotypic discordance. Am. J. Med. Genet. *112*, 194–197. https://doi.org/10.1002/ajmg.10610.
13. 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. https://doi.org/10.1086/301875.
14. Shulman, S.A., Hyams, J.S., Gunta, R., Greenstein, R.M., and Cassidy, S.B. (1984). Arteriohepatic dysplasia (Alagille syndrome): extreme variability among affected family members. Am. J. Med. Genet. *19*, 325–332. https://doi.org/10.1002/ajmg.1320190215.
15. 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. https://doi.org/10.1002/humu.23879.
16. Li, L., Krantz, I.D., Deng, Y., Genin, A., Banta, A.B., Collins, C.C., Qi, M., Trask, B.J., Kuo, W.L., Cochran, J., et al. (1997). Alagille syndrome is caused by mutations in human Jagged1, which encodes a ligand for Notch1. Nat. Genet. *16*, 243–251. https://doi.org/10.1038/ng0797-243.
17. Lu, F., Morrissette, J.J.D., and Spinner, N.B. (2003). Conditional JAG1 mutation shows the developing heart is more sensitive than developing liver to JAG1 dosage. Am. J. Hum. Genet. *72*, 1065–1070.
18. Morrissette, J.D., Colliton, R.P., and Spinner, N.B. (2001). Defective intracellular transport and processing of JAG1 missense mutations in Alagille syndrome. Hum. Mol. Genet. *10*, 405–413.
19. 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. https://doi.org/10.1038/ng0797-235.
20. 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. https://doi.org/10.1111/j.1742-4658.2012.08595.x.
21. 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. https://doi.org/10.1007/978-3-319-94571-2_3.
22. 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. https://doi.org/10.1007/s00439-017-1779-6.
23. Bauer, R.C., Laney, A.O., Smith, R., Gerfen, J., Morrissette, J.J.D., Woyciechowski, S., Garbarini, J., Loomes, K.M., Krantz, I.D., Urban, Z., et al. (2010). Jagged1 (JAG1) mutations in patients with tetralogy of Fallot or pulmonic stenosis. Hum. Mutat. *31*, 594–601. https://doi.org/10.1002/humu.21231.
24. Karpen, S.J., Kamath, B.M., Alexander, J.J., Ichetovkin, I., Rosenthal, P., Sokol, R.J., Dunn, S., Thompson, R.J., and Heubi, J.E. (2021). Use of a Comprehensive 66-Gene Cholestasis Sequencing Panel in 2171 Cholestatic Infants, Children, and Young Adults. J. Pediatr. Gastroenterol. Nutr. *72*, 654–660. https://doi.org/10.1097/MPG.0000000000003094.
25. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. *46*, D1062–D1067. https://doi.org/10.1093/nar/gkx1153.
26. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint

spectrum quantified from variation in 141,456 humans. Nature *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
27. Leonard, L.D., Chao, G., Baker, A., Loomes, K., and Spinner, N.B. (2014). Clinical utility gene card for: Alagille Syndrome (ALGS). Eur. J. Hum. Genet. *22*, e1–e4. https://doi.org/10.1038/ejhg.2013.140.
28. Vandriel, S.M., Li, L.T., She, H., Wang, J.S., Gilbert, M.A., Jankowska, I., Czubkowski, P., Gliwicz-Miedzińska, D., Gonzales, E.M., Jacquemin, E., et al. (2023). Natural history of liver disease in a large international cohort of children with Alagille syndrome: Results from the GALA study. Hepatology *77*, 512–529. https://doi.org/10.1002/hep.32761.
29. Masek, J., and Andersson, E.R. (2017). The developmental biology of genetic Notch disorders. Development *144*, 1743–1763. https://doi.org/10.1242/dev.148007.
30. Chillakuri, C.R., Sheppard, D., Ilagan, M.X.G., Holt, L.R., Abbott, F., Liang, S., Kopan, R., Handford, P.A., and Lea, S.M. (2013). Structural analysis uncovers lipid-binding properties of Notch ligands. Cell Rep. *5*, 861–867. https://doi.org/10.1016/j.celrep.2013.10.029.
31. Kapp, K., Schrempf, S., Lemberg, M.K., and Dobberstein, B. (2009). Post-Targeting Functions of Signal Peptides. In Protein Transport into the Endoplasmic Reticulum, R. Zimmerman, ed. (Landes Biosciences). https://www.ncbi.nlm.nih.gov/books/NBK6322/.
32. Pintar, A., Guarnaccia, C., Dhir, S., and Pongor, S. (2009). Exon 6 of human JAG1 encodes a conserved structural unit. BMC Struct. Biol. *9*, 43. https://doi.org/10.1186/1472-6807-9-43.
33. Luca, V.C., Kim, B.C., Ge, C., Kakuda, S., Wu, D., Roein-Peikar, M., Haltiwanger, R.S., Zhu, C., Ha, T., and Garcia, K.C. (2017). Notch-Jagged complex structure implicates a catch bond in tuning ligand sensitivity. Science *355*, 1320–1324. https://doi.org/10.1126/science.aaf9739.
34. Van der Auwera, G.A., and O'Connor, B.D. (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (O'Reilly Media, Incorporated).
35. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.
36. 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. https://doi.org/10.1371/journal.pone.0269159.
37. Fokkema, I.F.A.C., Taschner, P.E.M., Schaafsma, G.C.P., Celli, J., Laros, J.F.J., and den Dunnen, J.T. (2011). LOVD v.2.0: the next generation in gene variant databases. Hum. Mutat. *32*, 557–563. https://doi.org/10.1002/humu.21438.
38. Matreyek, K.A., Stephany, J.J., Ahler, E., and Fowler, D.M. (2021). Integrating thousands of PTEN variant activity and abundance measurements reveals variant subgroups and new dominant negatives in cancers. Genome Med. *13*, 165. https://doi.org/10.1186/s13073-021-00984-x.
39. Brnich, S.E., Abou Tayoun, A.N., Couch, F.J., Cutting, G.R., Greenblatt, M.S., Heinen, C.D., Kanavy, D.M., Luo, X., McNulty, S.M., Starita, L.M., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. *12*, 3. https://doi.org/10.1186/s13073-019-0690-2.
40. Tavtigian, S.V., Greenblatt, M.S., Harrison, S.M., Nussbaum, R.L., Prabhu, S.A., Boucher, K.M., Biesecker, L.G.; and ClinGen Sequence Variant Interpretation Working Group ClinGen SVI (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet. Med. *20*, 1054–1060. https://doi.org/10.1038/gim.2017.210.
41. UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. *49*, D480–D489. https://doi.org/10.1093/nar/gkaa1100.
42. Nielsen, H., Tsirigos, K.D., Brunak, S., and von Heijne, G. (2019). A Brief History of Protein Sorting Prediction. Protein J. *38*, 200–216. https://doi.org/10.1007/s10930-019-09838-3.
43. Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L.H., Zielinski, M., Sargeant, T., et al. (2023). Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science *381*, eadg7492. https://doi.org/10.1126/science.adg7492.
44. Tordai, H., Torres, O., Csepi, M., Padanyi, R., Lukacs, G.L., and Hegedus, T. (2023). Lightway access to AlphaMissense data that demostrates a balanced performance of this missense mutation predictor. Preprint at bioRxiv. https://doi.org/10.1101/2023.10.30.564807.
45. Kola, S., Koneti, N.R., Golla, J.P., Akka, J., Gundimeda, S.D., and Mundluru, H.P. (2011). Mutational analysis of JAG1 gene in non-syndromic tetralogy of Fallot children. Clin. Chim. Acta *412*, 2232–2236. https://doi.org/10.1016/j.cca.2011.08.017.
46. Haritunians, T., Chow, T., De Lange, R.P.J., Nichols, J.T., Ghavimi, D., Dorrani, N., St Clair, D.M., Weinmaster, G., and Schanen, C. (2005). Functional analysis of a recurrent missense mutation in Notch3 in CADASIL. J. Neurol. Neurosurg. Psychiatry *76*, 1242–1248. https://doi.org/10.1136/jnnp.2004.051854.
47. Le Caignec, C., Lefevre, M., Schott, J.J., Chaventre, A., Gayet, M., Calais, C., and Moisan, J.P. (2002). Familial deafness, congenital heart defects, and posterior embryotoxon caused by cysteine substitution in the first epidermal-growth-factor-like domain of jagged 1. Am. J. Hum. Genet. *71*, 180–186. https://doi.org/10.1086/341327.
48. Schrijver, I., Liu, W., Brenn, T., Furthmayr, H., and Francke, U. (1999). Cysteine substitutions in epidermal growth factor-like domains of fibrillin-1: distinct effects on biochemical and clinical phenotypes. Am. J. Hum. Genet. *65*, 1007–1020. https://doi.org/10.1086/302582.
49. Whiteman, P., Willis, A.C., Warner, A., Brown, J., Redfield, C., and Handford, P.A. (2007). Cellular and molecular studies of Marfan syndrome mutations identify co-operative protein folding in the cbEGF12-13 region of fibrillin-1. Hum. Mol. Genet. *16*, 907–918. https://doi.org/10.1093/hmg/ddm035.
50. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
51. Kall, L., Krogh, A., and Sonnhammer, E.L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. *338*, 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.
52. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (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. Genet. Med. *17*, 405–424. https://doi.org/10.1038/gim.2015.30.

53. Fowler, D.M., and Rehm, H.L. (2024). Will variants of uncertain significance still exist in 2030? Am. J. Hum. Genet. *111*, 5–10. https://doi.org/10.1016/j.ajhg.2023.11.005.
54. Gasperini, M., Starita, L., and Shendure, J. (2016). The power of multiplexed functional analysis of genetic variants. Nat. Protoc. *11*, 1782–1787. https://doi.org/10.1038/nprot.2016.135.
55. Gelman, H., Dines, J.N., Berg, J., Berger, A.H., Brnich, S., Hisama, F.M., James, R.G., Rubin, A.F., Shendure, J., Shirts, B., et al. (2019). Recommendations for the collection and use of multiplexed functional data for clinical variant interpretation. Genome Med. *11*, 85. https://doi.org/10.1186/s13073-019-0698-7.
56. Liu, Y., Engelman, D.M., and Gerstein, M. (2002). Genomic analysis of membrane protein families: abundance and conserved motifs. Genome Biol. *3*, research0054. https://doi.org/10.1186/gb-2002-3-10-research0054.
57. Grochowski, C.M., Loomes, K.M., and Spinner, N.B. (2016). Jagged1 (JAG1): Structure, expression, and disease associations. Gene *576*, 381–384. https://doi.org/10.1016/j.gene.2015.10.065.
58. Fayer, S., Horton, C., Dines, J.N., Rubin, A.F., Richardson, M.E., McGoldrick, K., Hernandez, F., Pesaran, T., Karam, R., Shirts, B.H., et al. (2021). Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN. Am. J. Hum. Genet. *108*, 2248–2258. https://doi.org/10.1016/j.ajhg.2021.11.001.
59. Allen, S., Garrett, A., Muffley, L., Fayer, S., Foreman, J., Adams, D.J., Hurles, M., Rubin, A.F., Roth, F.P., Starita, L.M., et al. (2024). Workshop report: the clinical application of data from multiplex assays of variant effect (MAVEs), 12 July 2023. Eur. J. Hum. Genet. *32*, 593–600. https://doi.org/10.1038/s41431-024-01566-2.
60. Li, L., Vorobyov, I., and Allen, T.W. (2013). The different interactions of lysine and arginine side chains with lipid membranes. J. Phys. Chem. B *117*, 11906–11920. https://doi.org/10.1021/jp405418y.
61. Morgan, A.A., and Rubenstein, E. (2013). Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. PLoS One *8*, e53785. https://doi.org/10.1371/journal.pone.0053785.
62. Petukh, M., Kucukkal, T.G., and Alexov, E. (2015). On human disease-causing amino acid variants: statistical study of sequence and structural patterns. Hum. Mutat. *36*, 524–534. https://doi.org/10.1002/humu.22770.
63. Pandey, A., Niknejad, N., and Jafar-Nejad, H. (2021). Multifaceted regulation of Notch signaling by glycosylation. Glycobiology *31*, 8–28. https://doi.org/10.1093/glycob/cwaa049.
64. Zhu, B.Y., Zhou, N.E., Kay, C.M., and Hodges, R.S. (1993). Packing and hydrophobicity effects on protein folding and stability: effects of beta-branched amino acids, valine and isoleucine, on the formation and stability of two-stranded alpha-helical coiled coils/leucine zippers. Protein Sci. *2*, 383–394. https://doi.org/10.1002/pro.5560020310.
65. Gergics, P., Smith, C., Bando, H., Jorge, A.A.L., Rockstroh-Lippold, D., Vishnopolska, S.A., Castinetti, F., Maksutova, M., Carvalho, L.R.S., Hoppmann, J., et al. (2021). High-throughput splicing assays identify missense and silent splice-disruptive POU1F1 variants underlying pituitary hormone deficiency. Am. J. Hum. Genet. *108*, 1526–1539. https://doi.org/10.1016/j.ajhg.2021.06.013.
66. Bhagavatula, G., Rich, M.S., Young, D.L., Marin, M., and Fields, S. (2017). A Massively Parallel Fluorescence Assay to Characterize the Effects of Synonymous Mutations on TP53 Expression. Mol. Cancer Res. *15*, 1301–1307. https://doi.org/10.1158/1541-7786.MCR-17-0245.
67. Faure, G., Ogurtsov, A.Y., Shabalina, S.A., and Koonin, E.V. (2017). Adaptation of mRNA structure to control protein folding. RNA Biol. *14*, 1649–1654. https://doi.org/10.1080/15476286.2017.1349047.
68. Hunt, R.C., Simhadri, V.L., Iandoli, M., Sauna, Z.E., and Kimchi-Sarfaty, C. (2014). Exposing synonymous mutations. Trends Genet. *30*, 308–321. https://doi.org/10.1016/j.tig.2014.04.006.
69. Hanson, G., and Coller, J. (2018). Codon optimality, bias and usage in translation and mRNA decay. Nat. Rev. Mol. Cell Biol. *19*, 20–30. https://doi.org/10.1038/nrm.2017.91.
70. Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science *324*, 255–258. https://doi.org/10.1126/science.1170160.
71. Sauna, Z.E., and Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. Nat. Rev. Genet. *12*, 683–691. https://doi.org/10.1038/nrg3051.
72. Esposito, D., Weile, J., Shendure, J., Starita, L.M., Papenfuss, A.T., Roth, F.P., Fowler, D.M., and Rubin, A.F. (2019). MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. Genome Biol. *20*, 223. https://doi.org/10.1186/s13059-019-1845-6.