Genetics in Medicine (2025) 27, 101477

ELSEVIER

Genetics in Medicine
An Official Journal of the ACMG

www.journals.elsevier.com/genetics-in-medicine

ARTICLE

Utility of genome sequencing and group-enrichment to support splice variant interpretation in Marfan syndrome

Susan Walker[1], David J. Bunyan[2], Huw B. Thomas[3], Yesim Kesim[4], Christopher J. Kershaw[5], John Holloway[6], Htoo Wai[6], Michael Day[7], Cassandra L. Smith[1], Gareth Hawkes[7], Andrew R. Wood[7], Michael N. Weedon[7], Ed Blair[8], Stephanie L. Curtis[9], Catherine Fielden[10], Julie Evans[10], Rebecca Whittington[10], Sarah F. Smithson[11], Helen Cox[12], Paul Clift[13], Meriel McEntagart[14], Matina Prapa[14], Suzanne Alsters[14], Deborah Morris-Rosendahl[15,16], John Dean[17], Patrick J. Morrison[18], Abhijit Dixit[19], Ajoy Sarkar[19], Katrina Prescott[20], Leila Amel Riazat Kesh[20], Riya Tharakan[5], Claire Turner[21], Sian Ellard[21,22], Charles Shaw-Smith[21], James Fasham[21], Virginia Clowes[23], Simon Holden[24], Suresh Somarathi[12], Catherine Mercer[25], Ian Berry[10], Raymond T. O'Keefe[3], Siddharth Banka[5,26], Diana Baralle[6], N. Simon Thomas[2], Emma L. Baple[7,21], Jenny C. Taylor[4], Alistair T. Pagnamenta[4,7,*]

ARTICLE INFO

Article history:
Received 27 January 2025
Received in revised form
19 May 2025
Accepted 23 May 2025
Available online 2 June 2025

Keywords:
Aortic aneurysm
FBN1
Marfan syndrome
Pseudoexon
Splicing

ABSTRACT

Purpose: To quantify the impact of noncanonical *FBN1* splice site variants in undiagnosed Marfan syndrome (MFS), a connective tissue disorder associated with skeletal abnormalities and familial thoracic aortic aneurysm disease (FTAAD).
Methods: A systematic analysis of ultrarare *FBN1* variants was performed using genome sequencing data from the 100,000 Genomes Project. Variants were annotated with SpliceAI and the significance of enrichment among individuals with FTAAD was assessed using Fisher's exact test. Experimental validation used RNA sequencing, reverse transcriptase polymerase chain reaction, minigene constructs, and replication analysis was with data from UK Biobank.
Results: Using aggregate data for 78,195 individuals, we identified 13,864 singleton single-nucleotide variants in *FBN1* of which 21 were predicted to affect splicing (SpliceAI > 0.5). Incidence of candidate splice variants in individuals recruited with FTAAD (9/703) was significantly elevated compared with that seen in non-FTAAD participants (12/77,492; odds ratio = 84, $P = 9.7 \times 10^{-14}$). Additional analysis uncovered a further 14 families harboring 11 different *FBN1* splice variants. A total of 20 candidate splice variants in 23 families were identified, of which 70% lay beyond the ±8 splice regions. RNA testing confirmed the

The Article Publishing Charge (APC) for this article was covered by an "Institutional Agreement Discount" between Elsevier and the University of Exeter.
Susan Walker and David J. Bunyan contributed equally to this work.
*Correspondence and requests for materials should be addressed to Dr Alistair T. Pagnamenta, RILD Building, Royal Devon and Exeter Hospital, Barrack Road, Exeter, EX2 5DW, UK. *Email address:* a.pagnamenta@exeter.ac.uk
Affiliations are at the end of the document.

doi: https://doi.org/10.1016/j.gim.2025.101477

predicted splice aberration in 16 of 20 and for 9 of 20, pseudoexonization was the likely splicing anomaly.

Conclusion: Our findings indicate that noncanonical splice variants may account for approximately 3% of families with undiagnosed FTAAD, highlighting the importance of incorporating analysis of introns and confirmatory RNA testing into genetic testing for Marfan syndrome.

Introduction

Marfan syndrome (MFS; OMIM 154700), a connective tissue disorder first described in 1896, has an estimated incidence of 2 to 3 in 10,000 individuals and an average age at diagnosis in the early 20s.[1-3] In addition to a number of typical skeletal features, aneurysm of the thoracic aorta is a cardinal feature, which can, in the absence of surgery, progress to sudden aortic dissection and death. The "Ghent nosology," which primarily comprises a set of cardiovascular, ocular, and skeletal manifestations was developed to facilitate clinical diagnosis of this syndrome and improve patient management/counseling.[4] Although penetrance is typically high and correlated with age, MFS has highly variable expressivity, including absence of aortic pathology in some individuals/families.

Although the first pathogenic variants in *FBN1* (HGNC:3603) described in MFS patients >30 years ago were de novo in origin,[5] inherited pathogenic variants have also been found cosegregating across large multigenerational pedigrees.[6,7] *FBN1* is a large gene spread over 237 kb, with 66 exons of which 65 are coding. The gene encodes fibrillin 1, a glycoprotein that is localized to the extracellular matrix. Pathogenic variants reported to date frequently include those resulting in premature termination codons (PTCs; including those in the last exon), in-frame indels and missense variants, particularly those disrupting cysteine residues that affect the formation of structurally important disulphide bonds.[8-10] Geleophysic and acromicric dysplasia (OMIM 614185 and 102370), conditions characterized by severe short stature, short extremities, and stiff joints, can also be caused by pathogenic *FBN1* variants in exons 41 and 42.[11]

Using conventional methods, rare pathogenic or likely pathogenic *FBN1* variants can be identified in >90% of individuals with clinically diagnosed MFS.[12,13] Families without a molecular diagnosis after clinical testing may be explained by poor phenotypic characterization, the presence of variants in other gene(s) linked to related conditions,[13] or cryptic variants in *FBN1*. Recently, the identification of structural variants, including inversions,[14,15] balanced translocations,[16] and complex interchromosomal insertions,[17] has highlighted the benefits of using genome sequencing data and analyzing intronic sequence of *FBN1* for genetically undiagnosed cases of MFS.

Genome sequencing also enables detection of cryptic splice variants that lie beyond ±8 bp into intronic sequence, which often remain undetected by current clinical testing procedures. Although many pathogenic variants in the literature result in aberrant splicing, reports of noncanonical splice variants in *FBN1* have mainly been limited to case studies (Supplemental Table 1), and many of these were detected using RNA-based testing approaches.

In this study, we performed a systematic assessment of genome sequence data from the 100,000 Genomes Project (100kGP) which included >700 individuals recruited with a diagnosis of familial thoracic aortic aneurysm disease (FTAAD). Although there are over 30 genes linked to this condition, the diagnostic yield for this set of families (for whom nondiagnostic standard-of-care National Health Service (NHS) testing results were obtained previously) is below 10%. We aimed to identify new molecular diagnoses for families with suspected MFS and, in parallel, assess the utility of genome sequencing in combination with in silico splice prediction for the detection of cryptic variants in *FBN1*.

Materials and Methods

Bioinformatic extraction/annotation of singleton variants

A systematic analysis of ultrarare *FBN1* variants was performed using an aggregated data set produced from the v10 release of the 100kGP, a nationwide study that aimed to uncover the genetic basis of disease for individuals in whom a diagnosis had not been obtained by standard-of-care testing.[18] The aggregation included germline data for 78,195 participants from both the rare disease and cancer program of the 100kGP, with FTAAD being the 11th most common disease indication (Supplemental Table 2). This data set (the "AggV2" file) includes >722 million single-nucleotide variants (SNVs) and small indels (≤50 bp) and is split into 1371 genome chunks of approximately equal size. DNA samples were quality controlled, and genome sequencing was conducted using a HiSeqX instrument (Illumina) generating 150-bp paired-end reads. Alignment was to the GRCh38 assembly. Further information about data processing and the construction of these aggregate data

is available in the Supplemental Methods and at https://re-docs.genomicsengland.co.uk/aggv2.

Singleton variants were extracted from the AggV2 file using BCFtools for the entire *FBN1* gene region (chr15:48,408,313-48,645,709, GRCh38). Variant annotation was performed using the Variant Effect Predictor (VEP) and Human Genome Variation Society annotation of variants and was based on the MANE select transcript (NM_000138.5). ORs and 2-tailed significance levels based on Fisher's exact test were calculated using RStudio. Additional scrutiny of in silico splice predictions and the surrounding landscape of absolute (raw) SpliceAI scores was performed using SpliceAI-visual,[19] with the resulting bedGraph files downloaded and imported as a custom track into the UCSC genome browser. The SpliceAI lookup tool (https://spliceailookup.broadinstitute.org) was also used with the max region of effect increased to 500 bp. The option to output absolute scores for REF/ALT alleles was utilized. SpliceAI scores are linked to 1 of 4 possible consequences: acceptor gain (AG), acceptor loss (AL), donor gain (DG), and donor loss (DL). Unless otherwise stated, just the highest of the 4 delta scores (range 0-1) is reported, alongside the delta position.[20]

Secondary analysis involving FTAAD-only data

The October 2024 data release (v19) of the 100kGP includes 72,884 rare disease program participants. Included in this data set were 672 families recruited with FTAAD. The FTAAD category was created in the 100kGP to encompass patients with syndromic and nonsyndromic forms of aortic dilatation (Supplemental Methods). A number of these families were not available on the GRCh38 genome build in data release v10, ie, at the time the AggV2 file was created. Thus, these individuals would not have been assessed in our primary analysis. A secondary analysis was therefore performed, which included all families recruited to the 100kGP with FTAAD, irrespective of which build had been used for read mapping. In this analysis, we did not require rare variants to be unique in the 100kGP, and the population allele frequency threshold was ≤0.001 (based on data from gnomAD v3 and the 1000 Genomes Project). We also used a more sensitive SpliceAI threshold of 0.2. Data were analyzed using a range of custom scripts. We also included a family sequenced as part of NHS England's Genomic Medicine Service using similar procedures to those described above but for which the variant had been picked up by the local Genomic Laboratory Hub using a commercial analysis platform (Congenica) to identify intronic substitutions with a SpliceAI score of >0.1.

RNA sequencing

RNA sequencing (RNAseq) data are available for 5546 probands from the 100kGP, of whom 180 were recruited with FTAAD as the primary diagnosis. Further technical details and information about sample prioritization are available in the Supplemental Methods and elsewhere.[21] In brief, total RNA was extracted from peripheral whole blood using the PAXgene Blood RNA Kit (Qiagen). Libraries were constructed following the Illumina Stranded Total RNA Prep kit, and ligation was with the Ribo-Zero Plus protocol. Sequencing was done on the NovaSeq 6000 using 2 × 100-bp paired-end reads.

Targeted RNA analysis

For targeted RNA testing, blood-derived RNA samples were collected into PAXgene RNA collection tubes. Complementary DNA (cDNA) preparation, reverse transcriptase polymerase chain reaction (RT-PCR) and Sanger sequencing was then carried out as described in the Supplemental Methods and using the primers listed in Supplemental Table 3. The one exception to this was family 20, for which fibroblasts were cultured in the presence/absence of cycloheximide for up to 6 hours to monitor levels of nonsense-mediated decay (NMD).

Minigene constructs

Minigene vectors for both wild-type and variant sequences of *FBN1* were assembled by Gibson cloning using the SK3 minigene vector (a derivative of the pSpliceExpress minigene splice reporter vector, gifted by S. Stamm; Addgene, 32485, RRID:Addgene_32485), as previously described.[22] Human embryonic kidney 293 cells were cultured to 60% to 75% confluency in 2 mL of high-glucose Dulbecco's modified Eagle's medium (Gibco), supplemented with 10% fetal bovine serum (Gibco) in a 6-well culture plate at 37 °C with 5% CO_2. Cells were transiently transfected with 2 µg of either wild-type or variant minigene constructs using Lipofectamine 2000 (Thermo Fisher Scientific) and the manufacturer's standard protocol. After 18 to 21 hours of incubation, RNA was extracted using 1 mL of TRI Reagent (Invitrogen) per well and further purified using the RNAeasy cleanup kit (QIAGEN), which included a DNase digestion step. cDNA was synthesized from up to 4 µg RNA (using an equal amount of RNA for each sample set) by GoScript first strand synthesis using the manufacturer's recommended protocol (Promega). Resulting cDNA was amplified using Q5 Polymerase (New England Biolabs) with the primers listed in Supplemental Table 3. Finally, polymerase chain reaction (PCR) products were run on an agarose gel (1%-4%) supplemented with SYBR Safe (Thermo Fisher Scientific) for visualization. PCR products were purified using a PureLink gel extraction kit (Invitrogen) and sequenced by Sanger sequencing (Eurofins Genomics) to confirm splicing products.

Analysis of *FBN1* splice variants using UK Biobank

UK Biobank (UKB) is a population-based study involving half a million participants from the UK.[23] Recruitment was between 2006 and 2010, and participants were aged 40 to 69

years at the time of recruitment. In addition to the comprehensive demographic and health-related measures, short-read genome sequencing was recently performed, as described (Li S, Carss KJ, Halldorsson BV, Cortes A, UK Biobank Whole-Genome Sequencing Consortium. Whole-genome sequencing of half-a-million UK Biobank participants. medRxiv. 2023.2012.2006.23299426; 2023. https://doi.org/10.1101/2023.12.06.23299426). The R package ukbrapR (https://lcpilling.github.io/ukbrapR) for working in the UKB Research Analysis Platform was used to identify UKB participants with the International Classification of Diseases, 10th revision (ICD10) codes I71 (Aortic aneurysm and dissection) and Q87.4 (MFS). Variants were annotated using Ensemble VEP v110, as described (https://github.com/drarwood/vep_ukb_aou_docker) and SNVs were extracted from the chromosome 15 file (version of March 10, 2024) in the MANE select transcript for *FBN1* (ENST00000316623.10 / NM_000138.5). To align with the systematic analysis performed on 100kGP data, we filtered for variants with a maximum SpliceAI score of 0.5 or more using precomputed scores and the 50-bp analysis window. The analysis was also repeated using a 0.2 SpliceAI threshold. Ultrarare variants were identified by filtering using total population allele frequencies (ie, across all ancestry groups) from the gnomAD v4.1.0 genomes, such that only variants that are singleton variants or else absent in this data set were retained. Variants also were required to have an allele count of 5 or less in UKB. Data access was under the approved UKB project "Understanding the role of rare and common genetic variation in human phenotypes."

Results

Identification of putative splice variants in *FBN1*

To assess the potential contribution of intronic *FBN1* variants in the etiology of FTADD for families recruited to the 100kGP, we performed a systematic analysis of rare variants with SpliceAI scores indicative of a possible splicing alteration. Initially, the analysis included 78,195 participants of the 100kGP, 703 of whom were recruited with FTAAD, ie, 0.90% of the total cohort (Supplemental Table 2). We considered singleton variants (allele count of 1 in the AggV2 data set) in *FBN1* across all phenotype categories and identified 13,864 SNVs. After annotation with SpliceAI, only 21 were seen to have a SpliceAI score of 0.5 or greater (Figure 1A). Incidence of this class of variant in individuals recruited with FTAAD (9/703; 1.3%) was significantly elevated compared with that seen in non-FTAAD participants (12/77,492; 0.015%) and based on Fisher's exact test, this enrichment was highly significant (odds ratio [OR] = 84, $P = 9.7 \times 10^{-14}$). Although 0.5 is the intermediate cutoff recommended for SpliceAI,[20] we explored the effects of varying this threshold for variant inclusion. Although the highest proportion of cases recruited under FTAAD was seen at the 0.8 cutoff (7/10, OR = 260, $P = 5.4 \times 10^{-13}$), the most significant P value (11/34, OR = 53, $P = 6.8 \times 10^{-15}$) was observed at the 0.3 cutoff (Figure 1B). We also explored the effect of relaxing the allele frequency cutoff and repeated this analysis with variants that have AC = 1-5 in the AggV2 file. Although enrichment in the FTAAD cohort was more significant than with the singleton variants at the 0.5 SpliceAI threshold, it came with a lower odds ratio ($P < 2.2 \times 10^{-16}$, OR = 35.2, Supplemental Note 1).

The 703 individuals with FTAAD included in the primary analysis were from a total of 605 families, of which 523 (86%) were recruited with just a single affected individual. However, there were also 70 families with 2 affected individuals, 10 affected trios, 1 affected quad, and 1 affected sextet. Therefore, we performed a secondary analysis including all families recruited to the 100kGP with FTAAD. In this analysis, we no longer required variants to be unique in the data set and reduced the SpliceAI threshold to 0.2. We also included families that had previously been available only on GRCh37 (at the time of data aggregation) and 1 family sequenced as part of the Genomic Medicine Service program. This FTAAD-only analysis led to the identification of 11 noncanonical *FBN1* splice variants in 23 more affected individuals from 14 unrelated families.

In combination with the initial 9 variants, a total of 20 unique variants were identified in 32 individuals from 23 families (Table 1[7,24-28]). These variants were spread throughout the gene from intron 1 to intron 63 (Figure 2A). Although there was no significant clustering, introns 12, 54, and 56 each harbored 2 variants. Overall, 14 of 20 variants lay beyond the ±8 splice regions used for prioritization in the primary analysis performed by Genomics England, and for 9 of these, inclusion of a pseudoexon (PE) was the predicted consequence. Exon extension was also the predicted effect for 9 of 20 variants, in 1 case involving a 5′-untranslated region (5′-UTR) exon (Figure 2B). Predictions of exon skipping and exon contraction were each observed just once. No coding variants predicted to result in anomalous splicing were identified.

The total population allele frequencies for this set of variants in gnomAD v4.1.0 ranged from 0 up to 7 in 1,613,362 for the NM_000138.5:c.1468+24C>T variant in family 2 (Table 1[7,24-28]). The c.1468+24C>T variant is also 1 of 4 for which RNA testing was not possible; therefore, this remains a variant of uncertain significance (VUS). Delta SpliceAI scores ranged from 0.20 in families 12 to 15 (NM_000138.5:c.1589-1217G>T) and family 23 (NM_000138.5:c.4942+19A>T), up to 1.00 in family 1, where NM_000138.5:c.6872-11A>G predicts a change in acceptor site usage (Supplemental Figure 1). An interactive UCSC session showing the position of all 20 variants, SpliceAI-visual tracks showing absolute scores in comparison with the reference and additional tracks showing the

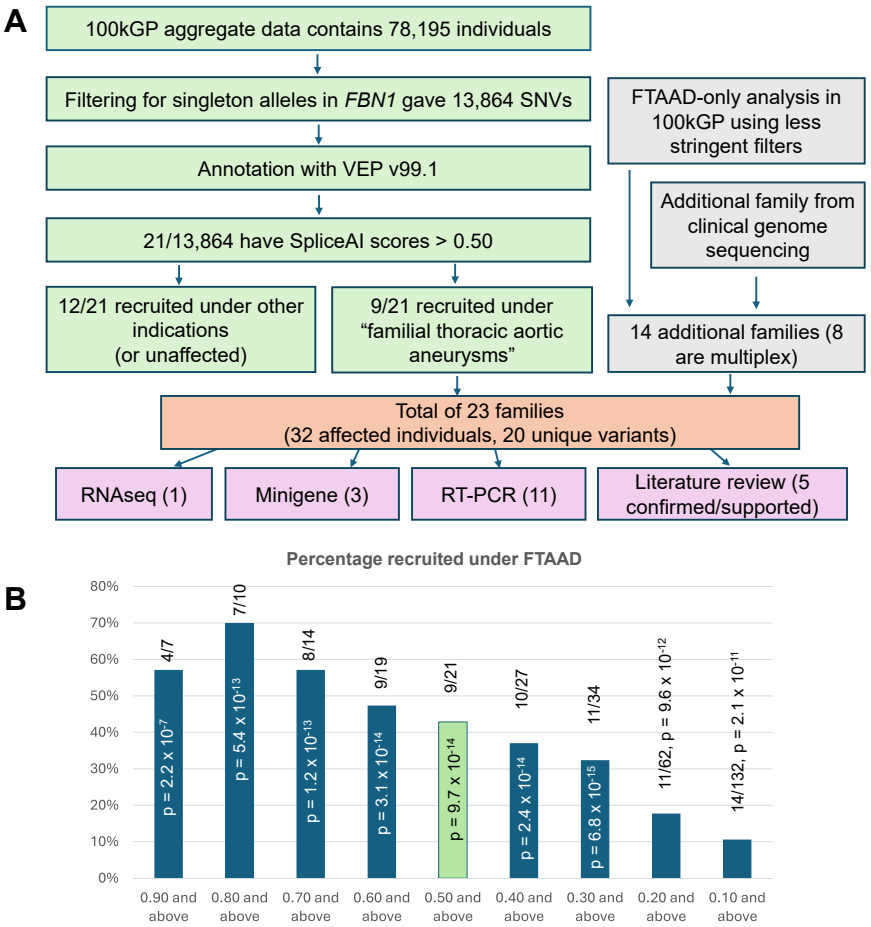

Figure 1 Analytical pipeline and phenotypic enrichment of rare splicing variants in *FBN1*. A. Green boxes denote steps taken in the initial analysis of data from the 100k Genomes Project (100kGP). Singleton SNVs were annotated with SpliceAI and filtered at 0.50 (AG, AL, DG, and DL) because that is the recommended intermediate stringency cutoff. Gray boxes show subsequent steps taken to identify 14 additional families. The salmon box summarizes the combined results from both arms of the study, whereas lilac boxes outline the different methods used to validate RNA effect and the numbers supported by each method. B. Chart showing percentage of individuals with a rare heterozygous splice variant who were recruited under FTAAD at each level of SpliceAI cutoff. As the splice cutoff decreases additional FTAAD families were identified. The green box highlights the initial cutoff used as shown in (A). The highest proportion of *FBN1* variant heterozygotes being from the FTAAD group was seen at a 0.80 cutoff, but this would have resulted in families 2, 7, 17, and 20 being overlooked. *P* values corresponding to Fisher's exact test results of FTAAD enrichment for *FBN1* variant heterozygotes show that the most significant enrichment was seen at the 0.3 cutoff. FTAAD, familial thoracic aortic aneurysm disease; SNV, single-nucleotide variant.

predicted consequences of each variant, is available at https://genome.ucsc.edu/s/AlistairP/FBN1_pseudoexons.

RNA testing for confirmation of aberrant splicing

Short-read RNAseq using blood-derived RNA had already been performed for 10 of 23 families with candidate splice variants in *FBN1*. Coverage at the *FBN1* locus was extremely sparse (Supplemental Figure 2). However, closer scrutiny of the loci of interest in the 10 data sets identified 1 read in 1 individual that supported the predicted splice event, with no coverage for the relevant splice junction in the other 9 data sets. The only confirmed event was in the proband from family 21, with a single read spanning NM_000138.5:c.6872-955C>G and the 3′ end of the predicted 96-bp PE in intron 56 (Supplemental Figure 3). The read alignment supported 1 of the 2 predicted splice junctions, with the soft-clipped bases matching the sequence from the start of exon 57. Similar junctions were not observed in >400 control RNAseq data sets sequenced using the same pipeline (Figure 3A). The variant appears to activate the same cryptic acceptor site as described in a previous study[28] (Figure 3B) and the PE was confirmed by RT-PCR using an independent blood sample (Figure 3C).

For 11 families, RT-PCR was used as a standalone test for assessing the effect of the variant on the *FBN1* transcript (Figure 1A, Table 1[7,24-28]). For 10 of 11 families, blood-derived RNA was used as a template, and despite low expression in this tissue, informative results were obtained in every case. For family 7, multiple rounds of primer design were required, but otherwise, this approach proved to be highly efficient. RT-PCR requires fresh patient-derived

Table 1 Summary of *FBN1* splicing variants in 23 families recruited with familial thoracic aortic aneurysm identified from the 100k Genomes Project or the Genome Medicine Service

Family ID	GRCh38 Coordinates (NC_000015.10)	HGVS Based on NM_000138.5 (SpliceAI Delta Score, Position)	AF in AggV2 (N = 78,195)	AF in UK Biobank (N = 490,640)	RNA Testing Results	Family Structure, Cosegregation and Family History	Notable Variants at Same Locus	ACMG Classification	ACMG Criteria
F1	15:g.48428482T>C	c.6872-11A>G (AG = 1.00, −1; AL = 0.99, −11)	1/156,390	Absent	RT-PCR shows 10-bp extension of exon 57, r.6871_6872insaaucaaacag, p.(Asp2291GlufsTer5)	Singleton in 100kGP. 2 paternal uncles suffering aortic dissection/aneurysm but no other living affected family members	NA	P (SCV005888566)	PVS1_RNA_Very_Strong, PM2_Moderate, PP4_Supporting
F2	15:g.48515363G>A	c.1468+24C>T (DG = 0.63, +6)	1/156,390	1/981,088	NA (patient deceased)	Singleton-no other affected family.	NA	VUS (SCV005888577)	PP3_Supporting, PM2_Supporting, PP4_Supporting
F3	15:g.48441860A>C	c.6038-14T>G (AG = 0.97, −1)	1/156,376	Absent	Minigene assay confirms the introduction of an earlier acceptor splice site that extends the exon 50 by 13 bp, r.6037_6038insgugauucuuuuag, p.(Asp2013GlyfsTer7).	Variant not in unaffected mother.	NA	LP (SCV005888579)	PS3_Strong, PM2_Moderate
F4	15:g.48447995G>C	c.5671+773C>G (DG = 0.80, +1 but also AG = 0.67 +59 if 500 bp distance used)	1/156,390	Absent	RT-PCR and Sanger from blood confirms 59 bp pseudoexon prediction. Pseudoexon is c.5671+714_5671+772 or r.5671_5672ins59, p.(Asp1891GlufsTer2).	Singleton	NA	P (SCV005888580)	PVS1_RNA_Very_Strong, PM2_Moderate, PP4_Supporting
F5	15:g.48644510G>A	c.164+96C>T (DG = 0.96, +2)	1/156,390	Absent	Minigene assay confirms the utilization of a new donor splice site which extends exon 2 by 94 bp and introduces a stop codon shortly into the extended sequence r.164_165ins94, p.(Pro56Ter).	Singleton (father and brother are affected but not in 100kGP).	NA	P (SCV005888581)	PVS1_RNA_Very_Strong, PM2_Moderate, PP4_Supporting
F6	15:g.48432996A>T	c.6617-8T>A (AG = 0.98, −2; AL = 0.63, −8)	1/156,390	1/981,056	RT-PCR from blood shows creation of a novel splice acceptor with in-frame insertion of 2 amino acids, r.6616_6617inscugcag p.(Glu2205_Asp2206insAlaAla).	Variant not passed on to unaffected daughter and no other living family members.	NM_000138.5:c.6617-9_6617-8inv listed in ClinVar as LP (VCV000042404.4), also reported as c.6617-9_6617-8delCTinsAG[25] but that would be 7 bp exon extension.	VUS (SCV005888582)	PM2 and PM4
F7	15:g.48503942A>C	c.1961-3T>G (AL = 0.76, −3 but also AG = 0.17, 130 if 500 distance used).	1/156,390	Absent	RT-PCR from blood showed no difference initially, but no allele imbalance experiments were done. Redone with intronic primers 133-bp extension detected, r.1960_1961ins133 p.(Asp654AlafsTer17).	Singleton and no family history	Variant previously reported in Ogawa et al[26] (2011), described as c.IVS15-3T>G. NM_000138.5:c.1961-3T>C is in ClinVar as VUS (VCV002195588.2).	P (SCV005888583)	PVS1_RNA_Very_Strong, PM2_Moderate, PP4_Supporting
F8	15:g.48499044A>C	c.2114-6T>G (AL = 0.82, −6)	1/156,390	Absent	Prediction is exon 18 skipping (in-frame), but no samples available	Singleton and no family history	NA	VUS (SCV005888584)	PP3_Supporting, PP4_Supporting, PM2_Moderate

(continued)

Table 1 Continued

Family ID	GRCh38 Coordinates (NC_000015.10)	HGVS Based on NM_000138.5 (SpliceAI Delta Score, Position)	AF in AggV2 (N = 78,195)	AF in UK Biobank (N = 490,640)	RNA Testing Results	Family Structure, Cosegregation and Family History	Notable Variants at Same Locus	ACMG Classification	ACMG Criteria
F9	15:g.48645574C>T	c.-182+1G>A (DL = 0.80, +1; DG = 0.30, −3)	1/156,390	Absent	RT-PCR from blood shows the first exon is extended by 4 bp consistent with the in silico prediction. This introduces a new ATG start codon so could affect translation due to uORF creation. The *FBN1* 5′-UTR does not have any existing uAUGs; therefore, it is anticipated that creating one would be deleterious.	Singleton (mother with aortopathy not in 100kGP).	NA	VUS (SCV005888585)	PP3_Supporting, PP4_Supporting, PM2_Moderate
F10	15:g.48417801T>C	c.7820-2034A>G (DG = 0.59, +5 and AG = 0.71, +189 if 500 bp window used)	3/156,390	Absent	Failed to receive a sample from patient. The issue for minigene testing is the size of the fragment needed to incorporate both the variant and the nearest exon. Pseudoexon of 185 bp is predicted.	Proband, sister, and mother are all affected. All 3 individuals have the variant; therefore, there is support from cosegregation.	NA	VUS (SCV005888567)	PP3_Supporting, PP4_Supporting, PM2_Moderate
F11	15:g.48475731T>C	c.3965-1081A>G (DG = 0.95, +1)	2/156,390	Absent	RT-PCR from blood showed pseudoexon of 105 bp, consistent with in silico prediction, r.3964_3965ins105, p.(Asp1322delins36). The inserted amino acid sequence is EVMKKPKIPGPFQSFRWN TWQIACLQTLSNHLIGLH.	Affected father and proband	NA	LP (SCV005888568)	PVS1_Strong, PM2_Moderate, PP4_Supporting
F12-15	15:g.48511386C>A	c.1589-1217G>T (AG = 0.20, −3)	3/156,390[a]	3/979,474	Previously reported to create 202 bp pseudo-exon r.1588_1589ins202 and p.Asp530ValfsTer8.[7]	7 affected individuals from 4 families (3 parent-child and 1 single affected).	Reported in large family with significant cosegregation.[7]	P (SCV005888569)	PVS1_RNA_Very_Strong, PS4_Strong, PM2_Moderate, PP4_Supporting
F16	15:g.48456185G>A	c.5422+452C>T (DG = 0.94, +2 and AG = 0.83, +75 if 500 bp window used)	Absent[b]	Absent	Minigene assay confirms 74 bp pseudoexon r.5422_5423ins74, p.(Ile1809ArgfsTer8), which matches prediction.	Affected father and daughter. Father was tall, given parental heights.	NA	P (SCV005888570)	PVS1_RNA_Very_Strong, PM2_Moderate, PP4_Supporting
F17	15:g.48508577C>G	c.1837+5G>C (DL = 0.38, +5 but also DG = 0.96, −61 if 500-bp window used).	1/156,390	Absent	Proband in 100kGP is deceased; therefore, RNA analysis done on son who is also affected. RT-PCR from blood shows a 66-bp extension of exon, r.1837_1838ins66, p.(Lys612_Asp613ins22). The inserted amino acid sequence is GSCYKTLDHVLHVPFLLLMLFR.	Singleton in 100kGP. Son also affected and confirmed to have variant.	c.1837+5G>A, also DG = 0.83, −61 if 500 distance setting used. 9 submissions in ClinVar VCV000527158.37 as P or LP. Described as p.Lys612_Ile614ins22 after cDNA analysis[27] and therefore similar effect to c.1837+5G>C.	LP (SCV005888571)	PVS1_Strong, PM2, PP4
F18	15:g.48428341C>G	c.6997+5G>C (DL = 0.97, +5)	2/156,390	Absent	51-bp contraction predicted	Variant inherited from affected mother	c.6997+5G>A (DL = 0.87, +5)[26]; Invitae and in ClinVar VCV001437124.7 and c.6997+5G>T (VUS) VCV000954654.7	LP (SCV005888572)	PM2, PP3, PP4, PS1_Moderate

(continued)

Table 1 Continued

Family ID	GRCh38 Coordinates (NC_000015.10)	HGVS Based on NM_000138.5 (SpliceAI Delta Score, Position)	AF in AggV2 (N = 78,195)	AF in UK Biobank (N = 490,640)	RNA Testing Results	Family Structure, Cosegregation and Family History	Notable Variants at Same Locus	ACMG Classification	ACMG Criteria
F19	15:g.48438838CA>C[d]	c.6164-922del (DG = 0.89, +4)	1/156,390	Absent	RT-PCR confirms inclusion of an 81 bp pseudoexon and 27 new amino acids predicted, r.6163_6164ins81, p.(D2055_L2056ins27).	Variant not maternal (father not available but may be affected)	NA	VUS (SCV005888573)	PM2, PM4, PP4
F20	15:g.48513912A>C	c.1469-244T>G (DG = 0.47, +1 and AG = 0.44, +171 if 500 distance setting used)	1/156,390	Absent	RT-PCR and Sanger sequencing confirms 171-bp pseudoexon, r.1468_1469ins171 and predicts p.(Asp490ValfsTer35). Fibroblast cells (± cycloheximide) used for RT-PCR.	Singleton in 100kGP.	NA	P (SCV005888574)	PVS1_RNA_Very_Strong, PM2_Moderate, PP4_Supporting
F21	15:g.48429426G>C	c.6872-955C>G (DG = 0.72, +1)	Absent[c]	Absent	Predicts a 96-bp pseudoexon that contains stop codon, and this is supported by single read in RNAseq data. PAXgene blood RNA from affected mother used for RT-PCR and Sanger analysis confirms r.6871_6872ins96 p.(Asp2291ValfsTer9).	Variant inherited from affected mother. Proband has 2 healthy older siblings. At least 4 further affected maternal relatives.	NM_000138.5:c.6872-961A>G has been shown to create 90-bp pseudoexon that utilizes same cryptic acceptor and causing NMD,[28] VCV000643480.2.	LP (SCV005888575)	PVS1_str, PM2 and PP4
F22	15:g.48433135C>T	c.6617-147G>A (AG = 0.44, −2; DG = 0.57, −59 if 500 bp window used)	Absent[c]	2/981,008	RT-PCR and Sanger sequencing from blood RNA shows 58-bp pseudoexon, r.6616_6617ins6617-145_6617-88, which predicts p.(Asp2206GlyfsTer20) and is consistent with in silico prediction.	Variant transmitted from affected father (proband) to affected son. Cosegregation confirmed in 1 affected relative.	NA	P (SCV005888576)	PVS1_RNA_Very_Strong, PM2
F23	15:g.48465549T>A	c.4942+19A>T (DL = 29, +29; DG = 0.20, +6 but absolute score for donor gain is 0.99)	Absent[c]	Absent	RT-PCR and Sanger sequencing from blood PaxGene sample shows 13-bp extension of exon 40, r.4942_4943insguaaaugguccug, which predicts p.(Asp1648GlyfsTer5).	Variant not in mother, DNA from deceased affected father is unavailable	In ClinVar as likely benign (VCV001564510.6) but out of date splice predictors used.	P (SCV005888578)	PVS1_Very Strong, PM2_Moderate

ACMG classifications are using the Richards et al[24] ACMG framework and accessions for ClinVar submissions are shown.

ACMG, American College of Medical Genetics and Genomics; *AF*, allele frequency; *AG*, acceptor gain; *AL*, acceptor loss; *DG*, donor gain; *DL*, donor loss; *LP*, likely pathogenic; *NA*, not applicable; *P*, pathogenic; *RT-PCR*, reverse transcriptase polymerase chain reaction; *VUS*, variant of uncertain significance.

[a]Only 3 of 7 individuals available on GRCh38 when AggV2 file produced.

[b]Participants from the Genome Medicine Service program are not included in AggV2.

[c]Data analyzed on build GRCh37 and therefore not in AggV2.

[d]Can also be described as NC_000015.10:g.48438840del.

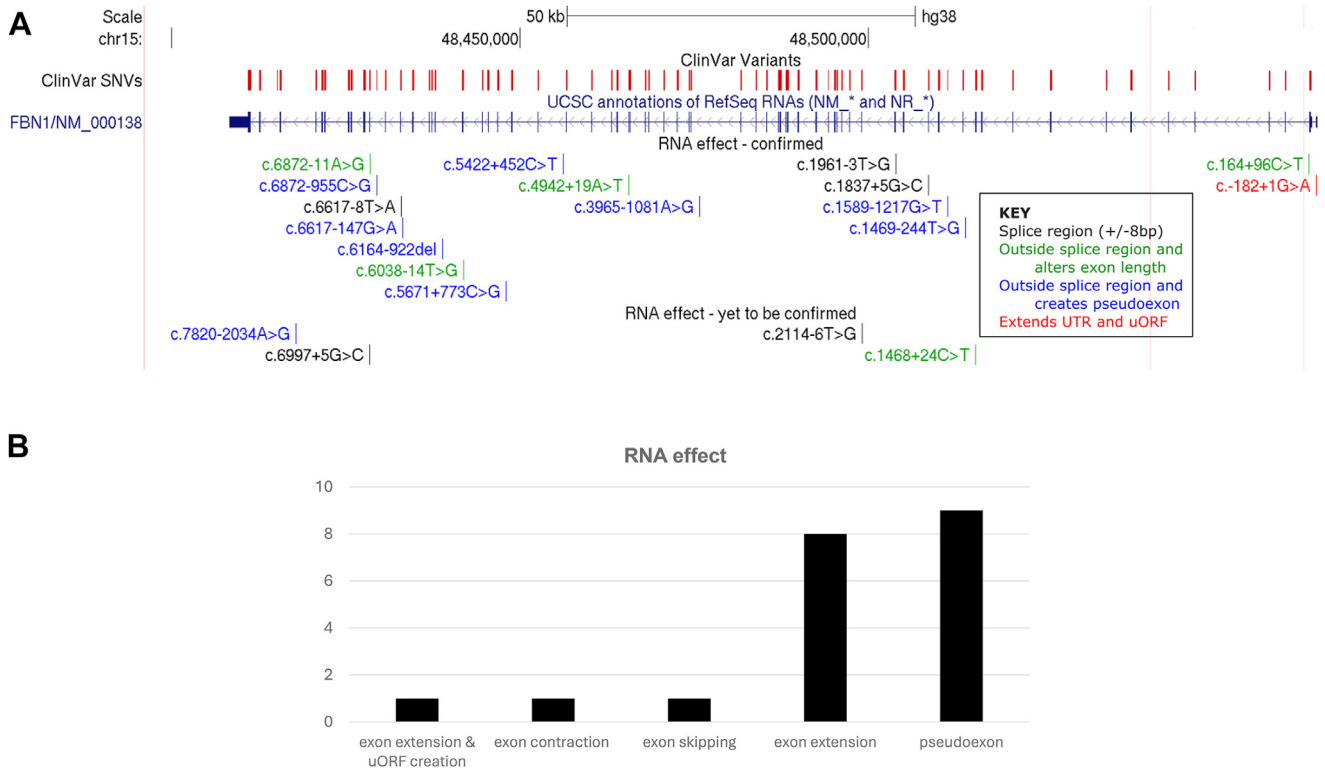

Figure 2 Distribution of noncanonical *FBN1* splicing variants. A. UCSC genome browser graphic showing that variants identified were spread throughout the gene. For viewing purposes, the multiregion view option (15:48396313-48540476, 15:48593651-48615606, and 15:48643624-48648334; GRCh38) was used to artificially shorten introns 2 and 6. An interactive UCSC custom session is available at https://genome.ucsc.edu/s/AlistairP/FBN1_distribution_v3. The ClinVar track highlights that the vast majority of pathogenic/likely pathogenic variants lie in or close to exons, with the only deep intronic variants shown are c.8051+375G>T, c.6872-961A>G, and c.1589-1217G>T. Of the variants described here, only c.1589-1217G>T and c.1961-3T>G were reported previously.[7,26] B. Bar chart summarizing the numbers of variant falling into each category of splice effect, with the most common consequence being creation of a pseudoexon and exon extension. HGVS annotations are based on transcript NM_000138.5. HGVS, Human Genome Variation Society.

RNA samples, and it was not possible to obtain these for all participants. Therefore, to test additional variants and explore the feasibility for alternative approaches, minigene assays were considered. Minigene assays were performed for 3 families and in all cases confirmed the predicted splice events. These included variants in families 3 and 5 resulting in exon extension of 13 and 94 bp, respectively with both predicting the introduction of PTCs (Figure 4A and B). For family 3, blood samples later became available, but instead of confirming the 13-bp extension, RT-PCR results suggested the occurrence of NMD (Supplemental Figure 4). Minigene testing for the heterozygous NM_000138.5:c.5422+452C>T variant found in family 16 confirmed the prediction of a 74-bp PE (Supplemental Figure 5) and the insertion, r.5422_5423ins74, would create a PTC, p.(Ile1809ArgfsTer8).

Overall, experimental confirmation of the putative splice variants has now been performed for 16 of 20 variants, 15 as part of this study and 1 variant (families 12-15) for which RNAseq data had already been reported in the literature.[7] In all cases, aberrant splicing was detected. Overall, we estimated that minigene assays required approximately twice the resources in terms of costs and hands-on time, compared with the simpler RT-PCR approach (Supplemental Methods).

Interpretation of noncanonical splice variants based on in silico approaches

The ability to detect variants likely to affect splicing is essential, but accurate prediction of the alternative splice event is highly beneficial for variant interpretation and for designing appropriate confirmatory RNA testing assays. Typically, the SpliceAI scores used are the delta scores, representing the difference score between reference and variant alleles.[19] In this analysis, we observed the importance of more refined interpretation of candidate splice variants.

Although precalculated SpliceAI scores using the 50-bp window are the scores typically used by researchers and used in our initial prioritization steps here (Figure 1A), for further interpretation, the window size was expanded to 500 bp. Strikingly, for 3 families, the more distant SpliceAI prediction increased the maximum delta SpliceAI score (Table 1[7,24-28]). The clearest example of an increased SpliceAI score using a larger analysis window is for family 17. The proband and her affected son harbored the

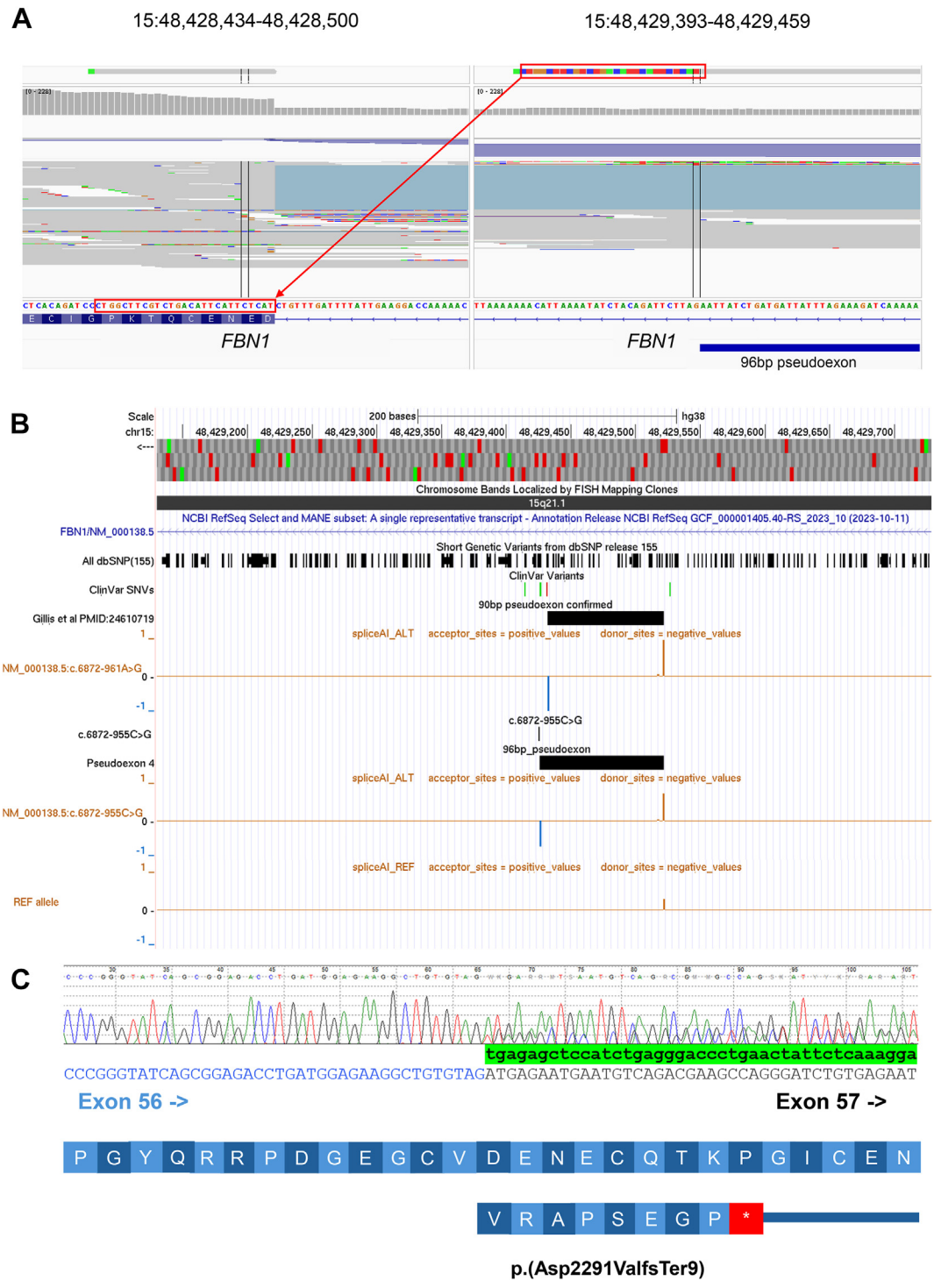

Figure 3 **Deep intronic variant in family 21 leads to activation of a dormant splice acceptor site and a 96-bp pseudoexon.** A. IGV screenshot showing single RNAseq read at the informative locus in intron 56. The red boxes highlight soft-clipped sequence, which match the start of exon 57 and are consistent with the SpliceAI prediction. The lower track shows squished view of reads from 402 merged RNAseq data sets run using a similar pipeline during the same calendar month. No similar junctions are observed. B. UCSC browser image showing positions of c.6872-955C>G and the absolute SpliceAI scores in comparison with the reference, as generated by SpliceAI-visual. The predicted pseudoexon activates the same dormant splice acceptor site that was upregulated by the c.6872-961A>G variant (red) shown in the ClinVar variant track. The positions of the 2 overlapping 96- and 90-bp pseudoexons are shown by black boxes. The dormant splice acceptor site has an absolute score of 0.29. C. PCR-Sanger sequencing of cDNA further supports the in silico prediction, with the sequence highlighted in green corresponding to the 5′ end of the 96-bp pseudoexon. The inserted sequence is predicted to lead to a frameshift and termination after the insertion of 8 novel amino acids. cDNA, complementary DNA; IGV, Integrative Genomics Viewer; RNAseq, RNA sequencing.

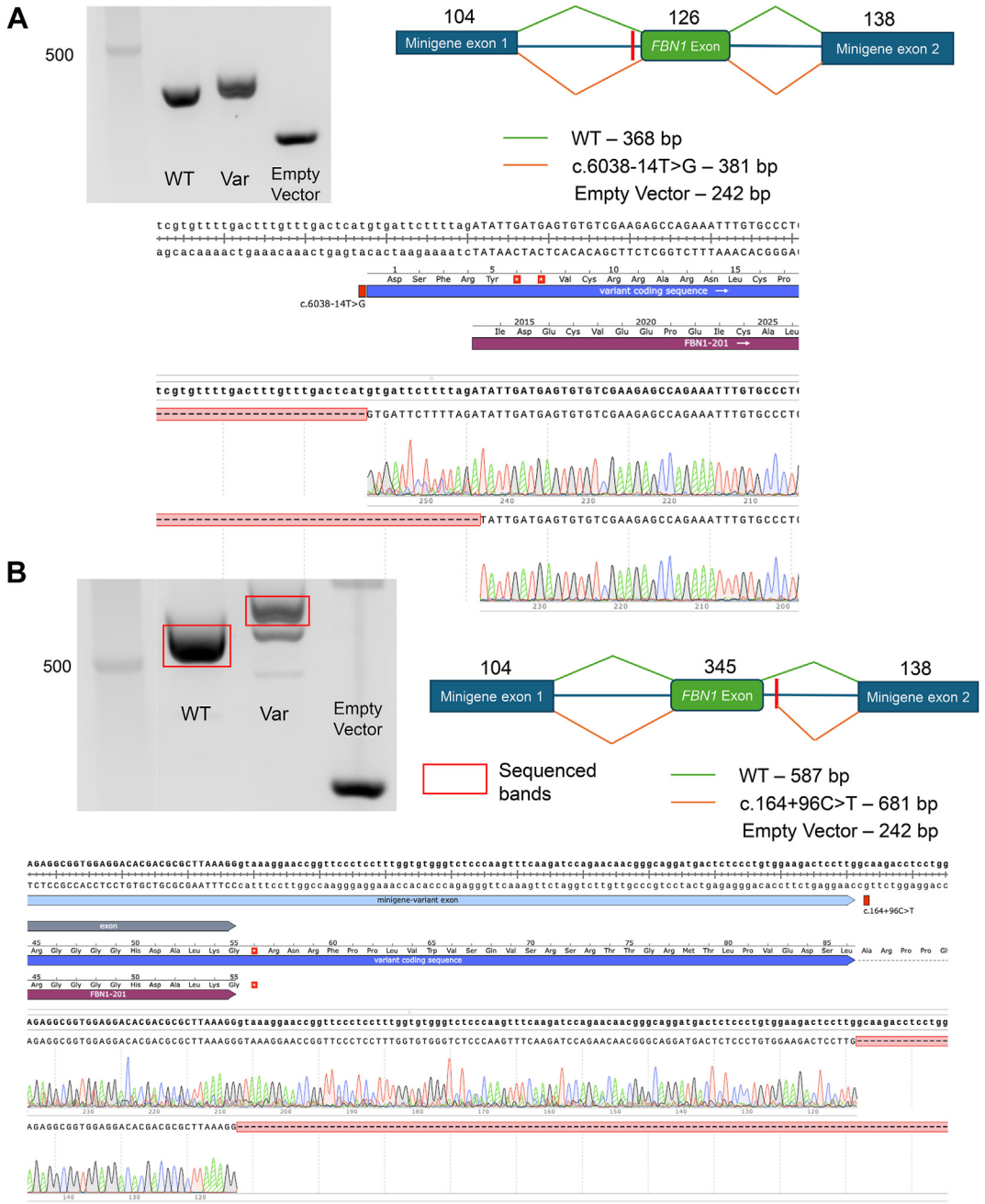

Figure 4 Minigene assay results confirm in silico predictions of exon extension for families 3 and 5. A. Agarose gel electrophoresis showing aberrant RT-PCR product associated with the c.6038-14T>G variant in family 3 because of the introduction of an earlier acceptor splice site consistent with the SpliceAI prediction for a novel acceptor site (AG = 0.97, −1 and AL = 0.20, −14). This extends exon 50 by 13 bp, r.6037_6038insgugauucuuuuag and predicts a frameshift p.(Asp2013GlyfsTer7). B. Gel image and Sanger sequence data showing that the c.164+96C>T variant in family 5 results in the utilization of a new donor site which extends exon 2 by 94 bp and is consistent with the SpliceAI prediction of the introduction of a novel deep intronic donor splice site (DG = 0.96, +2 and DL = 0.45, +96). This introduces a stop codon shortly into the extended sequence, r.164_165ins94, p.(Pro56*). AG, acceptor gain; AL, acceptor loss; DG, donor gain; DL, donor loss; RT-PCR, reverse transcriptase polymerase chain reaction.

NM_000138.5:c.1837+5G>C variant, which has a moderate score for a donor loss (DL = 0.38, +5), but the score increases to DG = 0.96 (−61) if the distance setting is increased. RT-PCR using blood-derived RNA from the son confirmed an in-frame 66-bp extension of exon 15, r.1837_1838ins66, p.(Lys612_Asp613ins22) (Supplemental Figure 6). Similarly, in family 22, NM_000138.5:c.6617-147G>A has a moderate prediction for an AG = 0.44, −2). However, with the larger window, there is a stronger prediction for a donor gain (DG = 0.57, −59). Together, this would result in a PE of 58bp. RT-PCR and Sanger sequencing from blood-derived RNA confirmed this PE,

r.6616_6617ins6617-145_6617-88, p.(Asp2206GlyfsTer20) (Supplemental Figure 7). A third example of this occurrence was for family 10 for which in silico analysis with the larger analysis window predicted a 185-bp PE (Table 1[7,24-28]).

In 1 case, a more accurate splicing prediction was able to overcome some of the difficulty with confirmatory RNA testing, although the maximum SpliceAI score was unchanged. The NM_000138.5:c.1961-3T>G variant identified was described in an independent study, but RNA testing had not been undertaken.[26] For family 7, prior genetic testing using a clinical exome had detected the same variant. However, RNA testing was not indicative of aberrant splicing, but it had been noted that a limitation of the RT-PCR experiment was that the possibility of a large insertion and PCR-bias or transcript degradation due to NMD could not be excluded. Thus, the variant had remained a VUS. Closer scrutiny of the SpliceAI scores using a 500-bp analysis window showed that, although the main effect was likely loss of an acceptor site (AL = 0.76, −3), there was also a weaker cryptic acceptor site predicted 130 bp away (AG = 0.17, 130). Although the delta SpliceAI score was <0.2, the raw reference allele has an absolute score of 0.06 at this position; therefore, the absolute SpliceAI score for the variant was 0.23. Repeating the RT-PCR assay using customized primers confirmed a 133-bp extension of exon 17 in the 5′ direction, r.1960_1961ins133, consistent with the prediction (Supplemental Figure 8). This predicted a frameshift leading to a PTC, p.(Asp654AlafsTer17).

Using the delta scores alone can also limit variant interpretation,[19] especially when there is potentially a high propensity for splicing with the reference allele (and therefore a high raw score). The clearest example of the importance of assessing absolute SpliceAI scores is seen in family 23. In this family, the c.4942+19A>T variant, which is listed as likely benign in ClinVar (VCV001564510.6), was detected in the proband. Although the delta score is low (DG = 0.20, +6), the absolute SpliceAI splice donor scores increase from 0.79 to 0.99 with the variant (Supplemental Figure 9A), resulting in a higher score than the canonical donor site for which the absolute score decreases from 0.99 to 0.84. If assessing just the delta score, this variant would be easy to overlook. RT-PCR and Sanger sequencing confirmed a 13-bp extension of exon 40, r.4942_4943insguaaaugguccug, which predicts p.(Asp1648GlyfsTer5) (Supplemental Figure 9B).

Examples of variants identified

We observed 13 putative loss of function alleles resulting from both inappropriate extension of canonical exons and through the creation of novel PEs. For 7 of 20 variants, a splicing anomaly was observed/predicted that is not expected to result in a PTC. These variants include in-frame indels (2 deletions and 4 insertions, Supplemental Table 4) and 1 variant expected to affect the 5′-UTR. Subsequent interpretation of these variants requires in-depth knowledge of the protein structure and functional domains and/or further functional assessment. In family 6, the c.6617-8T>A variant was shown to extend exon 55 by 6 bp, and this predicts the insertion of 2 alanine resides, the structural consequences of which are hard to determine. In family 8, in-frame exon skipping is predicted, whereas in families 11 and 19, the PEs predict in-frame insertions of 35 and 27 novel amino acids, respectively. The 66-bp exon extension in family 17 and 51 bp contraction in family 18 are also in-frame.

Another notable example is the NM_000138.5:c.-182+1G>A variant detected in family 9. This variant involves a noncoding exon in the 5′-UTR and is predicted to alter splicing of this exon (DL = 0.80, +1; DG = 0.30, −3). RT-PCR was able to confirm the extension of exon 1 by 4 bp (Supplemental Figure 10), and it is notable that this same variant also creates an AUG codon and an upstream open reading frame (uORF) that would encode a short peptide of 27 amino acids. The 5′-UTR of *FBN1* does not contain any other uORFs; therefore, we hypothesize that this variant could alter the efficiency of translational initiation from the canonical start site. The role of uORFs in translational regulation of protein expression among humans is becoming increasingly well understood,[29] but confirmation of such a mechanism would require further functional analysis, such as with luciferase reporter assays. This individual had an MFS phenotype, including nonprogressive dilated aortic root (4.0 cm), pectus, and arachnodactyly, and his mother is similarly affected. Previous testing of *FBN1* by Sanger sequencing and Multiplex Ligation-dependent Probe Amplification (MLPA) had returned negative results, but unfortunately, segregation in the family has not been possible. Examples of SNVs that simultaneously alter both transcription and translation are rare in the literature, with perhaps the best example being 5′-UTR variants in *PAX6*, which alter the reading frame of an existing uORF.[30]

Of the 20 variants identified, only 1 (NM_000138.5:c.1589-1217G>T) was recurrent and identified in 7 individuals from 4 independent families (families 12-15), each recruited from a different geographical region of the United Kingdom. Despite the relatively low SpliceAI score (AG = 0.20; rising to DG = 0.32 if absolute scores and larger analysis window is used), the variant was prioritized on account of the 7 individuals harboring this variant being recruited under FTAAD and the variant not being present in any other individuals recruited to the 100kGP. In parallel with this observation, c.1589-1217G>T was detected by Guo et al[7] after linkage analysis performed in a large MFS pedigree and was shown to result in a novel 202 bp PE by RNAseq.[7] The variant is also listed in ClinVar under accession VCV002664316.1. Because this variant has now been identified in at least 5 families, we performed haplotype studies as described previously[21] to assess whether the variant lay in a shared segment that is identical by descent. This included comparison of data for the 100kGP families with a set of 16 rare SNVs found on the disease haplotype in the original TAA758 family (pers. comm. Dong-Chuan Guo).[7] Altogether, we found no strong evidence supporting a single founder origin. Furthermore, we identified a nearby SNV (256 bp away) that could be phased with

read-level data. In families 12, 14, and 15 c.1589-1217G>T was in cis with the reference allele at rs9806163, whereas in family 13, it was in cis with the nonreference allele (Supplemental Figure 11). Further details of haplotype studies performed, and gnomAD findings are mentioned in Supplemental Note 2. Together, these data suggest independent mutational recurrence for c.1589-1217G>T. It is therefore important to screen for this cryptic variant in MFS cohorts from all ethnicities.

Clinical impact of noncoding variants in *FBN1*

We estimate the diagnostic yield based on the set of noncanonical *FBN1* splicing variants described here in the FTAAD cohort from the 100kGP to be 2.8% in the primary analysis (ie, 17 unrelated probands from 605 FTAAD families included in AggV2) or 3.3% (22/672, based on latest 100kGP release) in the FTAAD-only analysis. In all families for which data are available, *FBN1* testing had been undertaken before recruitment to the 100kGP. Many families had undergone a long diagnostic odyssey as evidenced by the wide range of methods used for prior testing that included denaturing high-performance liquid chromatography, PCR-Sanger, next-generation sequencing panel testing, MLPA, and clinical exome analysis. Aortic root dilatation ranged from being in the normal range for several individuals, up to 9 cm in diameter (before surgery) for the proband in family 3. Cardiac surgery had been performed in at least 11 of 23 families. Chest abnormalities were described in 15 of 23 families, whereas *ectopia lentis* was reported only in 5 of 23. Striae were noted in 8 of 23 and standing heights of 190 cm and above were reported for 9 of 23 families. Further clinical details are available in Supplemental Table 4.

Although the genome sequencing used in this study revealed the likely genetic diagnosis for a total of 32 individuals from 23 families, it is anticipated that cascade testing will be ongoing for several years and will identify many other at-risk individuals who may benefit from regular cardiac surveillance. The best example is family 22, for which the pedigree stretches over 4 generations, and there are 10 to 20 individuals for whom cascade testing for the c.6617-147G>A variant is anticipated (Supplemental Figure 7C). Similarly for family 21, although the full pedigree was not available for review, this nuclear family is part of another large MFS kindred, with at least 4 further affected maternal relatives reported; hence, ongoing cascade testing for the c.6872-955C>G variant will affect cardiac surveillance directly.

The 20 variants described here were assessed using the 2015 American College of Medical Genetics and Genomics (ACMG) framework[24] and with supplemental guidance from the ClinGen SVI Splicing Subgroup.[31] In total, there were 9 of 20 assessed as pathogenic, 5 of 20 as likely pathogenic, and 6 of 20 as a VUS (Supplemental Table 4). Although assessments for this set of variants have been submitted to ClinVar, we note that for some variants, classifications may change over time depending on results from ongoing segregation testing, RNA studies, other related functional experiments, or by expansion of gnomAD and other population databases to include larger and more ethnically diverse genome data sets.

Replication using UKB

Just under 1% (5000/502,284) of UKB participants had the ICD10 code I71 "Aortic aneurysm and dissection" listed either in their hospital episode statistics ($n = 4736$) or in the death record ($n = 264$). The incidence of aortic aneurysm in the general population is thought to be 1% to 2% above the age of 65[32] and therefore broadly in line with data from the UKB. There were also 112 in 502,284 participants with the code of Q87.4 (MFS) listed in the hospital episode statistics data. The overlap of these codes was surprisingly low, with only 40 of 112 of the individuals with Q87.4 also having the I71 code. There were 81 ultrarare SNVs identified with SpliceAI delta scores of 0.50 or more (Supplemental Table 5) found in a total of 115 UKB participants (all heterozygotes). Rare *FBN1* splice variants were not significantly enriched in the I71 cohort (OR = 1.76, $P = .32$), whereas for the more specific Q87.4 code there was a highly significant enrichment (OR = 167, $P = 1.59 \times 10^{-8}$). One of the 4 contributing variants was NM_000138.5:c.863-3641A>G (DG = 0.79, 1) in intron 8 and predicted to create a 485-bp PE. The individual harboring this deep intronic variant did not carry any other likely pathogenic variants in *FBN1*. Because of the nature of UKB consent, we were unable to obtain any further details or fresh biological samples to perform RNA confirmatory studies. Repeating this analysis with a more permissive SpliceAI filter threshold (delta score of 0.2) identified an additional 183 qualifying rare variants in 307 individuals. Enrichment analysis (Supplemental Note 3) showed increased significance (OR = 81, $P = 9.44 \times 10^{-12}$) driven by a single SNV, NM_000138.5:c.5788+5G>A (DG = 0.47, −28).

Discussion

Several case reports have indicated that intronic variants in *FBN1* can result in aberrant splicing and provide a molecular diagnosis for families with MFS.[7,28,33,34] However, the overall contribution of such variants to the etiology of MFS is unclear. In this study, the availability of genome sequencing data across a large clinical cohort that included a large number of participants with a diagnosis of FTAAD allowed us to explore the incidence of splice variants in genetically unsolved cases. Among >600 FTAAD families in the 100kGP without a previously identified molecular diagnosis, ~3% were found to have a candidate intronic small variant, and for 9 of 20 variants, the creation of a PE was predicted. Extrapolation of this yield into other clinical cohorts is difficult. First, MFS was not a specific diagnostic category used for recruitment to the 100kGP; therefore, it is hard to estimate how many of the 672 families recruited

with FTAAD also had a clinical suspicion specifically for MFS. Second, prior genetic testing was a requirement for recruitment to the 100kGP, and because many families were recruited through clinical genetics, it is possible that this collection of unsolved families is biased toward those with a strong family history and/or clinical suspicion of a monogenic disorder. As a result, noncanonical splice variants are likely better represented here than would be the case in MFS cohorts (or unselected population groups such as UKB) naïve to previous genetic testing. We also note that because a single splice prediction tool was used for variant prioritization, there could be additional splice variants in *FBN1* among the families with FTAAD in the 100kGP that were not considered. The fact that no exonic splice variants were identified may be the result of poorer predictive capabilities for splice variants in exonic regions.[35]

The Clinical Genome Resource *FBN1* Variant Curation Expert Panel have recently published consensus recommendations guiding the classification of variants in *FBN1*,[36] following the coding structure of the original ACMG/Association for Molecular Pathology guidance.[24] Minor modifications to the PVS1 criterion were proposed, including for canonical donor/acceptor splice dinucleotide sites, but the coding of deeper intronic splicing variants was not considered. The Variant Curation Expert Panel guidance also more clearly delineates the particular regions of functional relevance in fibrillin via the PM1 criterion, namely the repeated calcium binding epidermal growth factor domains, and regions of the gene with known cysteine-mediated dinucleotide-bonded secondary structure (which include the TB domains and organized N-terminal region of the protein). As our work has shown, noncanonical splicing variants are a recurrent and likely underreported cause of MFS. We would recommend clarifying the PVS1 criterion such that these splice-affecting variants, in cases in which they are demonstrated by cDNA analysis to introduce in-frame pseudoexonic cryptic sequence or delete multiple in-frame residues within these cysteine-rich domains, are subject to the interpretation "disrupts reading frame and is NOT predicted to undergo nonsense-mediated decay" and "truncated/altered region is critical to protein function"; therefore PVS1_strong is applied. In these circumstances, PVS1_strong would supplant the PP3 (in silico evidence of splicing abnormality) and PS3 (functional studies) criteria.

The data for all participants recruited to the 100kGP are available for research in the National Genomic Research Library. There have been a number of new tools published in recent years to enable prediction of splice variants, and this is an active area of research interest. Examples of systematic analyses of splicing variants in the 100kGP have focused for instance on "near-exonic" regions[37] or searching for elusive splice variants in trans with a single heterozygous pathogenic allele.[38,39] For participants with data available in the National Genomic Research Library, primary clinical data are provided as Human Phenotype Ontology terms, which were provided by referring clinicians at the time of recruitment to the 100kGP. Secondary clinical data derived from NHS hospital records are also available, which can provide additional phenotypic information. We acknowledge that this level of phenotype information in large collections of individuals is not uniform and missingness can be limiting for clinical correlation of candidate variants. It is thus not always clear if a participant was considered to have FTAAD or MFS, and in younger patients especially, the clinical presentation is often incomplete. This is why RNA studies are critical for confirming the genetic testing of MFS and appropriate screening to be implemented for cardiac and ocular complications.

To date, SpliceAI and other similar splice prediction tools have most frequently been used for variant prioritization in research analyses and/or for investigation of candidate variants for which there is a high clinical suspicion of a molecular diagnosis in a single or small number of candidate genes. Systematic implementation of such tools in clinical pipelines requires further work to assess an appropriate approach, balancing sensitivity with specificity, particularly considering the need for clinical scientists' time to assess variants. For the majority of novel candidate splice variants, further RNA testing would be required to support variant interpretation.

We observed a strong enrichment (OR = 84, $P = 9.7 \times 10^{-14}$) of singleton splice SNVs in the FTAAD cohort of the 100kGP. However, we failed to replicate this observation among individuals with "Aortic aneurysm and dissection" in the UKB. This lack of replication may be due to recruitment bias in the 100kGP, or a lack of phenotypic specificity in the UKB cohort, in which the ICD10 code, "Aortic aneurysm and dissection," is less precise than the clinical indication in the 100kGP data set (in which families were recruited with FTAAD, ie, "Familial Thoracic"). Abdominal aortic aneurysms are more common and usually related to age, male sex, and atherosclerosis and are thought not to have as significant a genetic contribution. The aortic aneurysm and dissection ICD10 code we used does not differentiate between the thoracic and abdominal aorta. UKB also contained 112 individuals with the more explicit ICD10 code of Q87.4, and for this phenotype, the enrichment of *FBN1* splice variants was highly significant (OR = 167, $P = 1.59 \times 10^{-8}$). Unfortunately, it is unclear whether Q87.4 is typically assigned on a clinical or genetic basis. If the latter, then this supporting evidence is circular and should be treated with caution.

For 16 of 20 candidate variants identified, an impact on mRNA splicing has now been confirmed using at least 1 RNA testing method. A limitation remains in this study that for the remaining 4 variants, RNA testing was not possible for a variety of reasons, including patients not being available to provide samples for testing. However, this limitation highlights challenges with confirmatory RNA testing, especially for genes with low expression in blood, for which bespoke testing methods or invasive sample collection approaches may be required. Even when a result is obtained, PCRs that use low amounts of starting template are in danger of exhibiting stochastic allelic dropout, as was seen in experiments on family 9 (Supplemental Figure 10). In this

study, we made use of a number of different approaches for RNA analysis, and each method comes with its own set of advantages and disadvantages. RT-PCR using RNA extracted from blood was the approach that requires the least resources and was the most commonly used approach used in this study. This technique was performed for 10 families as a standalone test and in 2 further families for confirmation of minigene/RNAseq results. Quantitative RT-PCR using RNA derived from patient fibroblasts was used in a further case, and this approach allows the introduction of NMD inhibitors, such as cycloheximide, to the culture media to help monitor the impact of the c.1469-244T>G variant on NMD (Supplemental Figure 12) and potentially enhance any signal from aberrant transcripts. Although fibroblasts have traditionally been used for RNA analysis of *FBN1*, a recent study proposed urine analysis as a viable noninvasive solution.[40] Minigene assays require more resources but are helpful for patients for whom collection of fresh samples is not feasible. However, effective design of minigene constructs can be problematic for variants that lie too far into an intron. Testing splice donor variants involving exon 1 are not straightforward to design because these exons do not harbor natural splice acceptor sites. Although these can be engineered,[41] this procedure is not trivial. Based on the results presented here, RNAseq from blood is not recommended for analysis of splicing variants in *FBN1* and the single read in data from family 21 capturing one of the predicted PE junctions (Figure 3A) was fortuitous.

Pathogenic variants involving *FBN1* are highly actionable in terms of surveillance/management, and the gene is included in the ACMG list recommended for reporting of secondary findings.[42] It is therefore critically important that RNA validation studies are performed for confirmation of pathogenicity. Currently, provision of follow up RNA testing is variable in existing clinical testing pathways; therefore, many families had experienced long diagnostic odysseys. For 2 families presented here, the variant of interest was within 15 bp of the nearest exon and had been previously identified using a clinical panel test. However, in the absence of confirmatory RNA analysis, the variants were considered as VUS, and for 1 family, additional genetic testing was pursued even after genome sequencing. Increased availability of RNA testing would help resolve VUS and, in some cases, negate repeated DNA testing. Widespread use of RNA testing is hindered by low *FBN1* expression in blood for which the median transcript per million value is just 0.2158 (803 data sets from GTEx; https://gtexportal.org/home/gene/FBN1). In this study, most RNA experiments were performed using RT-PCR by a single clinically accredited genomics laboratory that has significant expertise in working with low abundance transcripts.

Over the last decade, there have been vast improvements in the accuracy of splice prediction algorithms. SpliceAI, which uses a deep neural network to facilitate highly accurate prediction of splice junctions from primary sequence,[20] is one of the most widely used approaches. Our results demonstrate that it is now feasible to scan large intronic regions across thousands of individuals in the search for cryptic splice variants. For the set of variants presented here, in silico predictions all proved to be correct. In contrast, a study from 2012 performed RNA analysis for 36 *FBN1* variants and identified only 2 that caused an abnormality of splicing, leading to the conclusion that, at that time, in silico analysis did not accurately predict the splicing seen experimentally.[43] Further improvements in splice prediction tools may enable the strength of evidence attributed to computational evidence and variant colocalization with other previously reported variants to be increased and potentially reduce the need for complex RNA testing. An example is family 18; although RNA studies were unable to be performed, support for pathogenicity was obtained from variant colocalization at the same +5 position. The proband and affected mother both harbored a NM_000138.5:c.6997+5G>C variant that predicted a strong donor loss (DL = 0.97, +5) and a donor gain (DG = 0.57, +56) that would result in contraction of the exon by 51 bp. Database searches indicated that another variant at the same base c.6997+5G>A with strikingly similar in silico prediction (DL = 0.87, +5, DG = 0.58, +56) has been described previously.[26] The variant is also listed in ClinVar with 1 likely pathogenic assessment (VCV001437124.8). For some genes, deep intronic splice variants can show significant clustering due to the intronic sequence already having exon-like properties and thus only requiring a small increase in splice efficiency to result in PE inclusion.[44,45] Such information should be incorporated into the analysis of candidate splice variants.

Our experience demonstrates that adapting the use of SpliceAI predictions from the default parameters facilitates a more refined analysis of candidate splice variants that increases diagnostic yield and enables improved experimental design of confirmatory RNA analysis. In particular, we note the benefits of increasing the SpliceAI window size from 50 to 500 bp and an analytical window of even 5000 bp has been proposed elsewhere.[46] Here, the increased window size was important for determining the formation of PEs because all 9 PEs detected were over 50 bp in size (58 to 202 bp, Table 1[7,24-28]). However, when designing analytical pipelines, it is important to balance gains in analytical accuracy with the bioinformatic resources required and the interpretation burden. Highly customized modification of parameters is not feasible at scale in high-throughput analysis pipelines, highlighting that further improvements in splice prediction tools would be beneficial.

In the case of *FBN1*, analysis of candidate splice variants can be facilitated by the phenotypic specificity of MFS as the associated condition. For other disorders with a broader phenotypic range and higher genetic heterogeneity, bespoke adaptation of splice parameters with the aim of achieving high sensitivity would be substantially more challenging, likely resulting in a greater analytical challenge and potentially more false positives.

In conclusion, our systematic evaluation of candidate splice variants in *FBN1* in the 100kGP highlights the

benefits of considering rare noncoding variants, prioritized by computational predictions. Meticulous variant analysis led to in silico predictions proving to be correct in all 16 cases in which RNA testing was possible. Our findings indicate that ultrarare *FBN1* variants with supportive SpliceAI scores should be considered as strong candidates in individuals with a FTAAD phenotype and/or suspicion for MFS, indicating that genome sequencing would be an appropriate test method for families without a candidate diagnosis identified. However, for any increased yield to be realized, such changes must also be accompanied by an increased capacity for splicing analysis of low abundance genes being available in clinical laboratories.

Data Availability

Data from the 100kGP and the Genomic Medicine Service are held in the National Genomic Research Library, Genomics England, https://doi.org/10.6084/m9.figshare.4530893.v7. Data from UK Biobank cannot be shared publicly because of data availability and data return policies. Data are available from the UK Biobank for researchers who meet the criteria for access to data sets to UK Biobank (www.ukbiobank.ac.uk).

Acknowledgments

The authors thank all families from the 100kGP and Genomic Medicine Service for their involvement in this study and the late Bruce Castle for clinical input. This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. This research has been conducted using the UK Biobank Resource under Application Number 103356 and uses data provided by patients and collected by the NHS as part of their care and support. UK Biobank protocols were approved by the National Research Ethics Service Committee. Some of this material was previously presented at the 57th European Society of Human Genetics Conference.

Funding

This study was supported by the National Institute for Health and Care Research Oxford and Exeter Biomedical Research Centers. The views expressed are those of the author(s) and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care. Additional funding was from the Medical Research Council (MR/W01761X/1) and the National Institute for Health and Care Research Manchester Biomedical Research Centre (NIHR203308).

Author Contributions

Conceptualization: A.T.P., S.W., J.C.T; Data Curation: E.B., S.L.C., C.F., J.E., R.W., S.F.S., H.C., P.C., M.M., M.P., S.A., D.M.-R., J.D., P.J.M., A.D., A.S., K.P., L.A.R.K., R.T., C.T., S.E., C.S.-S., J.F., V.C., S.H., S.S., C.M.; Formal Analysis: A.T.P., I.B., G.H., S.W.; Validation: D.J.B., H.B.T, C.J.K, J.H., H.W., C.K., Y.K., M.D.; Methodology: A.R.W., C.L.S.; Project Administration: R.T.O'K., D.B., M.N.W., N.S.T., E.L.B., J.C.T.; Supervision: R.T.O'K., S.B., D.B., N.S.T., E.L.B., J.C.T.; Writing-original draft: A.T.P., S.W.; Writing-review and editing: A.T.P., Y.K., S.W., R.T.O'K., S.B., I.B., H.B.T., D.J.B., N.S.T., E.L.B., J.C.T.

Ethics Declaration

Ethics approval was from Cambridge South REC (14/EE/1112), and informed consent was obtained from all participants.

Conflict of Interest

The authors declare no conflicts of interest.

Additional Information

The online version of this article (https://doi.org/10.1016/j.gim.2025.101477) contains supplemental material, which is available to authorized users.

ORCIDs

Susan Walker: https://orcid.org/0000-0002-5016-6426
David Bunyan: https://orcid.org/0000-0001-8686-558X
Deborah Morris-Rosendahl: https://orcid.org/0000-0002-7780-4707
Patrick J. Morrison: https://orcid.org/0000-0002-2823-1762
Raymond T. O'Keefe: https://orcid.org/0000-0001-8764-1289
Emma Baple: https://orcid.org/0000-0002-6637-3411
Jenny C. Taylor: https://orcid.org/0000-0003-3602-5704
Alistair T. Pagnamenta: https://orcid.org/0000-0001-7334-0602

Affiliations

[1]Genomics England Limited, London, United Kingdom; [2]Wessex Genomics Laboratory Service, Salisbury District Hospital, Salisbury, United Kingdom; [3]Division of Evolution, Infection and Genomics, School of Biological Sciences, University of Manchester, Manchester, United Kingdom; [4]Oxford NIHR Biomedical Research Centre, Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; [5]Manchester University Hospitals NHS Foundation Trust, Manchester Centre for Genomic Medicine, Manchester, United Kingdom; [6]University of Southampton, Human Development and Health, Faculty of Medicine, Southampton, United Kingdom; [7]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Royal Devon University Healthcare NHS Foundation Trust, Exeter, United Kingdom; [8]Oxford University Hospitals NHS Foundation Trust, Oxford Centre for Genomic Medicine, Oxford, United Kingdom; [9]University Hospitals Bristol and Weston NHS Foundation Trust, Bristol Heart Institute, Bristol, United Kingdom; [10]North Bristol NHS Trust, Bristol Genetics Laboratory, Bristol, United Kingdom; [11]University Hospitals Bristol NHS Foundation Trust, Department of Clinical Genetics, Bristol, United Kingdom; [12]Birmingham Women's and Children's Hospital, West Midlands Clinical Genetics Service, Birmingham, United Kingdom; [13]Queen Elizabeth Hospital, Grown Up Congenital Heart Disease Unit, Birmingham, United Kingdom; [14]St George's University Hospitals NHS Foundation Trust, SW Thames Centre for Genomic Medicine, London, United Kingdom; [15]Clinical Genetics and Genomics Laboratory, Royal Brompton Hospital, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom; [16]NHLI, Imperial College London, London, United Kingdom; [17]Clinical Medicine, University of Aberdeen, Aberdeen, United Kingdom; [18]Belfast Health and Social Care Trust, Regional Genetics Service, Belfast, United Kingdom; [19]Nottingham University Hospitals NHS Trust, Department of Clinical Genetics, Nottingham, United Kingdom; [20]Clinical Genetics Department, Leeds Teaching Hospitals NHS Trust, United Kingdom; [21]Department of Clinical Genetics, Royal Devon University Healthcare NHS Trust, Exeter, United Kingdom; [22]Exeter Genomics Laboratory, Royal Devon University Healthcare NHS Foundation Trust, United Kingdom; [23]North West Thames Regional Genetics Service, Northwick Park Hospital, United Kingdom; [24]Department of Clinical Genetics, Cambridge University Hospital NHS Foundation Trust, Cambridge, United Kingdom; [25]Wessex Clinical Genetics Service, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom; [26]Division of Evolution, Infection and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom

References

1. Marfan AB. Un cas de déformation congénitale des quatre membres, plus prononcée aux extrémités, caractérisée par l'allongement des os avec un certain degré d'amincissement. *Bull Mem Soc Med Hop Paris*. 1896;13:220-226.
2. Cui RZ, Hodge DO, Mohney BG. Incidence and de novo mutation rate of Marfan syndrome and risk of ectopia lentis. *J AAPOS*. 2023;27(5):273.e1-273.e4. http://doi.org/10.1016/j.jaapos.2023.07.006
3. Groth KA, Hove H, Kyhl K, et al. Prevalence, incidence, and age at diagnosis in Marfan syndrome. *Orphanet J Rare Dis*. 2015;10:153. http://doi.org/10.1186/s13023-015-0369-8
4. Loeys BL, Dietz HC, Braverman AC, et al. The revised Ghent nosology for the Marfan syndrome. *J Med Genet*. 2010;47(7):476-485. http://doi.org/10.1136/jmg.2009.072785
5. Dietz HC, Cutting GR, Pyeritz RE, et al. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature*. 1991;352(6333):337-339. http://doi.org/10.1038/352337a0
6. Dietz HC, Pyeritz RE, Puffenberger EG, et al. Marfan phenotype variability in a family segregating a missense mutation in the epidermal growth factor-like motif of the fibrillin gene. *J Clin Invest*. 1992;89(5):1674-1680. http://doi.org/10.1172/JCI115766
7. Guo DC, Duan X, Mimnagh K, et al. An FBN1 deep intronic variant is associated with pseudoexon formation and a variable Marfan phenotype in a five generation family. *Clin Genet*. 2023;103(6):704-708. http://doi.org/10.1111/cge.14322
8. Dietz HC, Saraiva JM, Pyeritz RE, Cutting GR, Francomano CA. Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome patients at cysteine residues in EGF-like domains. *Hum Mutat*. 1992;1(5):366-374. http://doi.org/10.1002/humu.1380010504
9. Muiño-Mosquera L, Steijns F, Audenaert T, et al. Tailoring the American College of Medical Genetics and Genomics and the association for molecular pathology guidelines for the interpretation of sequenced variants in the FBN1 gene for Marfan syndrome: proposal for a disease- and gene-specific guideline. *Circ Genom Precis Med*. 2018;11(6):e002039. http://doi.org/10.1161/CIRCGEN.117.002039
10. Yoon E, Lee JK, Park TK, et al. Experience of reassessing FBN1 variants of uncertain significance by gene-specific guidelines. *J Med Genet*. 2023;61(1):57-60. http://doi.org/10.1136/jmg-2023-109433
11. Le Goff C, Mahaut C, Wang LW, et al. Mutations in the TGFbeta binding-protein-like domain 5 of FBN1 are responsible for acromicric and geleophysic dysplasias. *Am J Hum Genet*. 2011;89(1):7-14. http://doi.org/10.1016/j.ajhg.2011.05.012
12. Loeys B, De Backer J, Van Acker P, et al. Comprehensive molecular screening of the FBN1 gene favors locus homogeneity of classical Marfan syndrome. *Hum Mutat*. 2004;24(2):140-146. http://doi.org/10.1002/humu.20070
13. Meester JAN, Peeters S, Van Den Heuvel L, et al. Molecular characterization and investigation of the role of genetic variation in phenotypic variability and response to treatment in a large pediatric Marfan syndrome cohort. *Genet Med*. 2022;24(5):1045-1053. http://doi.org/10.1016/j.gim.2021.12.015
14. Clarissa's story. Marfan syndrome. Genomics England. Accessed June 18, 2025. https://www.genomicsengland.co.uk/patients-participants/stories/clarissa
15. Pagnamenta AT, Yu J, Evans J, et al. Conclusion of diagnostic odysseys due to inversions disrupting GLI3 and FBN1. *J Med Genet*. 2023;60(5):505-510. http://doi.org/10.1136/jmg-2022-108753
16. Racine C, Callier P, Touraine R, et al. De novo balanced translocations disrupting the FBN1 gene diagnosed by genome sequencing: an uncommon cause of Marfan syndrome modifying genetic counseling. *Am J Med Genet A*. 2025;197(4):e63923. http://doi.org/10.1002/ajmg.a.63923

17. Bonaglia MC, Salvo E, Sironi M, et al. Case Report: decrypting an interchromosomal insertion associated with Marfan's syndrome: how optical genome mapping emphasizes the morbid burden of copy-neutral variants. *Front Genet*. 2023;14:1244983. http://doi.org/10.3389/fgene.2023.1244983
18. Turnbull C, Scott RH, Thomas E, et al. The 100 000 Genomes Project: bringing whole genome sequencing to the NHS. *BMJ*. 2018;361:k1687. http://doi.org/10.1136/bmj.k1687
19. de Sainte Agathe JM, Filser M, Isidor B, et al. SpliceAI-visual: a free online tool to improve SpliceAI splicing variant interpretation. *Hum Genomics*. 2023;17(1):7. http://doi.org/10.1186/s40246-023-00451-1
20. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535-548.e24. http://doi.org/10.1016/j.cell.2018.12.015
21. Pagnamenta AT, Yu J, Walker S, et al. The impact of inversions across 33,924 families with rare disease from a national genome sequencing project. *Am J Hum Genet*. 2024;111(6):1140-1164. http://doi.org/10.1016/j.ajhg.2024.04.018
22. Thomas HB, Wood KA, Buczek WA, et al. EFTUD2 missense variants disrupt protein function and splicing in mandibulofacial dysostosis Guion-Almeida type. *Hum Mutat*. 2020;41(8):1372-1382. http://doi.org/10.1002/humu.24027
23. Bycroft C, Freeman C, Petkova D, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209. http://doi.org/10.1038/s41586-018-0579-z
24. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. http://doi.org/10.1038/gim.2015.30
25. Lerner-Ellis JP, Aldubayan SH, Hernandez AL, et al. The spectrum of FBN1, TGFβR1, TGFβR2 and ACTA2 variants in 594 individuals with suspected Marfan syndrome, Loeys-Dietz syndrome or Thoracic Aortic Aneurysms and Dissections (TAAD). *Mol Genet Metab*. 2014;112(2):171-176. http://doi.org/10.1016/j.ymgme.2014.03.011
26. Ogawa N, Imai Y, Takahashi Y, et al. Evaluating Japanese patients with the Marfan syndrome using high-throughput microarray-based mutational analysis of fibrillin-1 gene. *Am J Cardiol*. 2011;108(12):1801-1807. http://doi.org/10.1016/j.amjcard.2011.07.053
27. Overwater E, Floor K, van Beek D, et al. NGS panel analysis in 24 ectopia lentis patients; a clinically relevant test with a high diagnostic yield. *Eur J Med Genet*. 2017;60(9):465-473. http://doi.org/10.1016/j.ejmg.2017.06.005
28. Gillis E, Kempers M, Salemink S, et al. An FBN1 deep intronic mutation in a familial case of Marfan syndrome: an explanation for genetically unsolved cases? *Hum Mutat*. 2014;35(5):571-574. http://doi.org/10.1002/humu.22540
29. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci U S A*. 2009;106(18):7507-7512. http://doi.org/10.1073/pnas.0810916106
30. Filatova AY, Vasilyeva TA, Marakhonov AV, et al. Upstream ORF frameshift variants in the PAX6 5′UTR cause congenital aniridia. *Hum Mutat*. 2021;42(8):1053-1065. http://doi.org/10.1002/humu.24248
31. Walker LC, Hoya Wiggins GAR, et al. Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: recommendations from the ClinGen SVI Splicing Subgroup. *Am J Hum Genet*. 2023;110(7):1046-1067. http://doi.org/10.1016/j.ajhg.2023.06.002
32. Golledge J. Abdominal aortic aneurysm: update on pathogenesis and medical treatments. *Nat Rev Cardiol*. 2019;16(4):225-242. http://doi.org/10.1038/s41569-018-0114-9
33. Guo DC, Gupta P, Tran-Fadulu V, et al. An FBN1 pseudoexon mutation in a patient with Marfan syndrome: confirmation of cryptic mutations leading to disease. *J Hum Genet*. 2008;53(11-12):1007-1011. http://doi.org/10.1007/s10038-008-0334-7
34. Kim JA, Jang MA, Jang SY, et al. Overcoming challenges associated with identifying FBN1 deep intronic variants through whole-genome sequencing. *J Clin Lab Anal*. 2024;38(1-2):e25009. http://doi.org/10.1002/jcla.25009
35. Smith C, Kitzman JO. Benchmarking splice variant prediction algorithms using massively parallel splicing assays. *Genome Biol*. 2023;24(1):294. http://doi.org/10.1186/s13059-023-03144-z
36. Drackley A, Somerville C, Arnaud P, et al. Interpretation and classification of FBN1 variants associated with Marfan syndrome: consensus recommendations from the Clinical Genome Resource's FBN1 variant curation expert panel. *Genome Med*. 2024;16(1):154. http://doi.org/10.1186/s13073-024-01423-3
37. Blakes AJM, Wai HA, Davies I, et al. A systematic analysis of splicing variants identifies new diagnoses in the 100,000 Genomes Project. *Genome Med*. 2022;14(1):79. http://doi.org/10.1186/s13073-022-01087-x
38. Lord J, Oquendo CJ, Wai HA, et al. Noncoding variants are a rare cause of recessive developmental disorders in trans with coding variants. *Genet Med*. 2024;26(12):101249. http://doi.org/10.1016/j.gim.2024.101249
39. Moore AR, Yu J, Pei Y, et al. Use of genome sequencing to hunt for cryptic second-hit variants: analysis of 31 cases recruited to the 100 000 Genomes Project. *J Med Genet*. 2023;60(12):1235-1244. http://doi.org/10.1136/jmg-2023-109362
40. Hiraide T, Shimizu K, Miyamoto S, et al. Genome sequencing and RNA sequencing of urinary cells reveal an intronic FBN1 variant causing aberrant splicing. *J Hum Genet*. 2022;67(7):387-392. http://doi.org/10.1038/s10038-022-01016-1
41. Rodriguez-Muñoz A, Liquori A, García-Bohorquez B, et al. Functional assays of non-canonical splice-site variants in inherited retinal dystrophies genes. *Sci Rep*. 2022;12(1):68. http://doi.org/10.1038/s41598-021-03925-1
42. Miller DT, Lee K, Abul-Husn NS, et al. ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: A policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2023;25(8):100866. http://doi.org/10.1016/j.gim.2023.100866
43. Robinson DO, Lin F, Lyon M, et al. Systematic screening of FBN1 gene unclassified missense variants for splice abnormalities. *Clin Genet*. 2012;82(3):223-231. http://doi.org/10.1111/j.1399-0004.2011.01781.x
44. De Angeli P, Flores-Tufiño A, Stingl K, et al. Splicing defects and CRISPR-Cas9 correction in isogenic homozygous photoreceptor precursors harboring clustered deep-intronic ABCA4 variants. *Mol Ther Nucleic Acids*. 2024;35(1):102113. http://doi.org/10.1016/j.omtn.2023.102113
45. Luo X, Wang R, Sun Y, et al. Deep intronic PAH variants explain missing heritability in hyperphenylalaninemia. *J Mol Diagn*. 2023;25(5):284-294. http://doi.org/10.1016/j.jmoldx.2023.02.001
46. Oh RY, AlMail A, Cheerie D, et al. A systematic assessment of the impact of rare canonical splice site variants on splicing using functional and in silico methods. *HGG Adv*. 2024;5(3):100299. http://doi.org/10.1016/j.xhgg.2024.100299