Transcript analysis for variant classification resolution in a child with primary ciliary dyskinesia

Alexander Ing,[1,2,3,4] Alissa Wlodaver,[1] Dawn Kirschmann,[1] Erica Toledo,[1] Christopher McCabe,[1] Sabah Kadri,[1,3] Mary Kate McIntyre,[2] Joanne Salazar,[2] Kristina Firestein,[2] Joel Charrow,[1,3,4] Victoria Sanders,[3,4] Theresa Laguna,[2,3] and Kai Lee Yap[1,3]

[1]Center for Genomics, Department of Pathology and Laboratory Medicine, [2]Division of Pulmonary and Sleep Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois 60611, USA; [3]Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA; [4]Division of Genetics, Birth Defects and Metabolism, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois 60611, USA

Abstract Transcriptional analysis can be utilized to reconcile variants of uncertain significance, particularly those predicted to impact splicing. Laboratory analysis of the predicted mRNA transcript may allow inference of the in vivo impact of the variant and aid prediction of its clinical significance. We present a patient with classical features of primary ciliary dyskinesia (PCD) who was identified to have compound heterozygous variants in the *DNAH11* gene (c.10691 + 2T > C, c.13523_13543dup21) via trio whole-exome sequencing in 2013. These variants were originally classified as Mutation and Likely Mutation. However, these variants were downgraded to variants of uncertain significance (VUSs) during reanalysis in 2016 because of uncertainty that they caused a loss of function of the gene. c.10691 + 2T > C is predicted to abrogate the canonical splice site and lead to the skipping of exon 65, but the adjoining of exon 64 and exon 66 in the *DNAH11* transcript preserves the reading frame of the resultant protein. c.13523_13543dup21 is located in the last exon of the *DNAH11* coding sequence, upstream of the canonical stop codon, which suggests a reduced likelihood to trigger nonsense-mediated decay (NMD). Transcriptional analysis was performed to characterize the impact of the variants, resulting in reclassification of c.10691 + 2T > C to Likely Pathogenic by providing evidence that it results in a deleterious effect and subsequent downstream reclassification of c.13523_13543dup21 to Likely Pathogenic as well. Our case illustrates the potential impact of transcriptional analysis on variant resolution, supporting its usage on variants that exert an unpredictable effect on splicing.

Corresponding author:
klyap@luriechildrens.org

INTRODUCTION

Primary ciliary dyskinesia (PCD) is a disorder characterized by abnormal ciliary structure and function that results in chronic respiratory disease, situs abnormalities, and infertility. The incidence of this condition is estimated at approximately 1 in 20,000 individuals (Loges et al. 2018). However, given the variable expressivity and nonspecific initial presentation of this condition (chronic sinusitis, nasal congestion, cough, and otitis media are common presentations in pediatric populations and could be misattributed as "daycare-itis"), PCD may be underreported and the true incidence may be higher (Leigh et al. 2019).

Diagnosis of this condition is typically accomplished via a combination of clinical evaluation, nasal nitric oxide measurement (NNO), electron microscopy of ciliary ultrastructure and genetic testing as noted in a recent practice guideline publication by the American Thoracic Society (ATS) (Shapiro et al. 2018). Genetic testing is typically performed as a next-generation sequencing (NGS) panel of up to 39 genes, with variable reported diagnostic yields because of limited sample sizes (Kim et al. 2014; Marshall et al. 2015; Boaretto et al. 2016; Djakow et al. 2016; Leigh et al. 2016).

Each methodology of diagnosis for this condition has limitations. The clinical presentation of PCD can be nonspecific; year-round daily cough or congestion beginning before the age of 6 months represent two out of four essential and characteristic clinical criteria described in the ATS guidelines. Thirty percent of individuals with PCD can have normal ciliary ultrastructure detected on electron microscopy. NNO testing requires specialized instruments and trained respiratory therapists. It also cannot be reliably performed in individuals under 5 years old or in individuals with acute respiratory illness or sinusitis (Shapiro et al. 2018). Considering these limitations, genetic testing may represent the most accessible tool in the diagnosis of PCD in a pediatric population, and a positive result yields a definitive diagnosis. However, genetic testing is not without disadvantages; the knowledge regarding the genetics of this condition is still incomplete as 30% of patients with a clinical diagnosis of PCD do not have a molecular diagnosis (Leigh et al. 2019) and additional genes are continuously being discovered to be associated with disease in recent publications (Bonnefoy et al. 2018; Fassad et al. 2018; Loges et al. 2018). Because of the novelty of these gene–disease relationships, inconclusive results are common and there are limited tools to clarify results.

NGS has enabled the diagnostic usage of multigene panels in human diseases and the comprehensive identification of genomic variants. Because of widespread usage of such multigene testing methodologies, the identification of variants of uncertain significance (VUSs) poses its own challenge in clinical interpretation. The American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG/AMP) guidelines on interpretation of sequence variants presents a rubric for interpretation that mainly examines factors such as population frequencies, in silico predictions, and a comprehensive literature review (Richards et al. 2015). Overall, the interpretation of protein truncating variants is relatively straightforward because of their expected deleterious effects on the gene function. However, the interpretation of base substitution DNA variants are more challenging, and mostly rely on the scant availability of functional studies to determine the effects of an amino acid substitution. Of note, some DNA variants can also exert an impact even if they do not directly cause amino acid substitutions, and these variants can occur in the exons or introns. These variants may disrupt existing splice sites or splicing regulatory sequences, in addition to either creating or activating cryptic sites. Such "spliceogenic" effects create alternative splicing events in the transcribed mRNA transcript of the gene and may lead to disruption of the open reading frame of the translated protein. In silico prediction tools can help predict the splicing impact of specific DNA variants but are not always accurate reflections of what occurs in vivo (Anna and Monika 2019). Analysis of the mRNA transcript can serve to clarify the functional effects of the variants by providing an expected prediction of the downstream protein and represents an important tool for the assessment of pathogenicity of the variants. In oncological interpretation, transcriptional analyses are common and have been widely cited for resolving the clinical significance of cancer mutations (Jayasinghe et al. 2018). Transcriptional analysis has also been reported in research studies for PCD including for the *DNAH11* gene (Knowles et al. 2012; Andjelkovic et al. 2018). However, the application and usage in clinical testing has been limited thus far. Techniques for analyzing the cDNA reverse-transcribed from the mRNA sequences can include Sanger sequencing of the polymerase chain reaction (PCR)-amplified products to delineate transcript splice junctions, or a more comprehensive whole-transcriptome characterization using NGS-based RNA-seq.

In this case report, we illustrate the usage of transcriptional analysis with PCR amplification and Sanger sequencing to clarify the effects of two variants of uncertain clinical significance in the gene *DNAH11*, in a patient with a clinical diagnosis of PCD. This analysis resulted in a change in classification for the variants from VUS to Likely Pathogenic. We also attempted to predict the effects of the splice variant in *DNAH11* using the MMSplice tool and present data on this analysis as well.

RESULTS

Clinical Presentation and Family History

Our patient was identified to have situs inversus totalis and mirror image dextrocardia by prenatal ultrasound at 20 weeks gestation and subsequently confirmed by prenatal ECHO (Fig. 1A depicts a computed tomography [CT] scan taken at 5 years of age). Patient was born full-term and without complication. She was seen by Genetics at our institution at 4 weeks of age; reported family history was noncontributory to presentation, and chromosomal microarray was ordered, which was nonrevealing. Trio whole-exome sequencing (WES) was performed, and bialleleic heterozygous variants in *DNAH11* were identified (c.10691 + 2T > C and c.13523_13543dup21) (Table 1) by a commercial reference laboratory in 2013, confirmed in *trans*, and initially classified as Mutation and Likely Mutation, supporting a molecular diagnosis of PCD (see Fig. 1B for family pedigree). Given the initial diagnosis by WES, ciliary biopsy was not performed. In light of these findings, the patient was referred to Pulmonology at our institution. Patient was consistently followed beginning at 8 months of age with regular follow-up for chest physiotherapy, airway clearance, and management monitoring. At 6 years of age, patient was seen in the Lurie Children's PCD specialty clinic to establish comprehensive care and follow-up.

Throughout early childhood our patient reported features consistent with PCD. Beginning at 6 months of age, parents noted that the patient had consistent rhinitis, congestion, and cough. At 10 months of age, the patient was referred to Otolaryngology for persistent middle ear effusion, resulting in the placement of ear tubes at 1 year of age. NNO
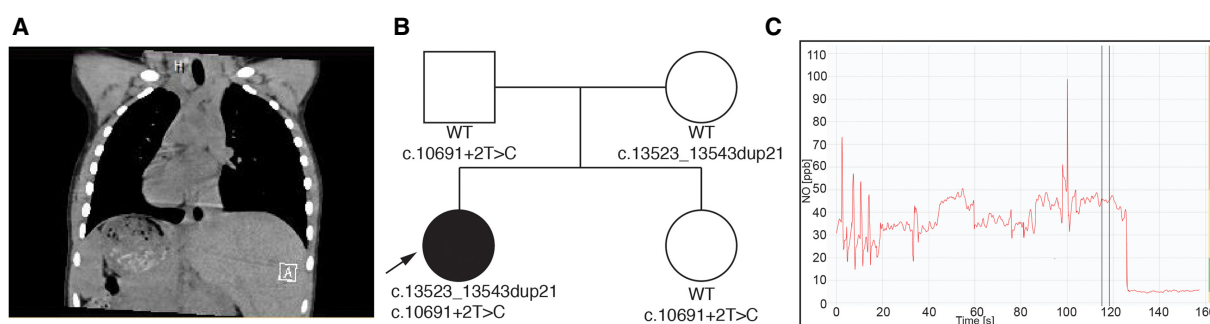

Figure 1. (*A*) Computed tomography (CT) scan taken at 5 years of age demonstrating mirror image dextrocardia of the proband. (*B*) Pedigree of proband's family illustrating the inheritance pattern of the two variants (c.10691 + 2T > C and c.13523_13543dup21). The father is a carrier of the c.10691 + 2T > C variant and the mother is a carrier of the c.13523_13543dup21 variant. The unaffected sibling is a carrier of only the c.10691 + 2T > C variant, and the proband is compound heterozygous for the c.10691 + 2T > C and c.13523_13543dup21 variants. (*C*) Graph of nasal nitric oxide (NNO) levels (*y*-axis) versus time (*x*-axis) illustrating the average reading of NNO levels in the proband that is consistently below the normal reference of 77 nL/min.

Table 1. Variants

Gene	Chromosome	HGVS DNA reference	HGVS protein reference	Variant type	Predicted effect	dbSNP/ dbVar ID	Genotype
DNAH11	Chr 7(GRCh37): g.21857959T > C	c.10691 + 2T > C	p.?	Substitution	Altered splicing	rs886039341	Heterozygous
DNAH11	Chr 7(GRCh37): g.21940844_21940864dup	c.13523_13543dup	p.Ala4508_ Leu4514dup	Duplication	In-frame duplication	n/a	Heterozygous

measurement was indicative of PCD and reproducible on separate visits; 15.2 nL/min and 16.3 nL/min (<77 nL/min diagnostic for PCD) (Fig. 1C).

Although our patient carried a clinical diagnosis of PCD, at 4 years of age, the family was contacted by Genetics that the laboratory that performed the whole-exome sequencing had reclassified both identified variants as VUSs.

Variant Interpretation

The *DNAH11* gene is a member of the dynein heavy chain family and contains 82 exons; there is one described transcript in Refseq (NM_001277115.1) in the GRCh37 genome assembly. Pathogenic variants in this gene cause PCD in an autosomal recessive inheritance pattern. Loss of function is a known mechanism of disease for this gene; of note, patients with pathogenic variants in this gene were identified to have normal ciliary ultrastructure (Knowles et al. 2012).

When this patient presented to our specialty clinic, we recontacted the testing laboratory to verify the classification; the performing laboratory confirmed that these variants had been downgraded to VUS and that this classification was current. These variants were also independently interpreted by our molecular diagnostic laboratory in 2019 using ACMG/AMP variant interpretation guidelines (Richards et al. 2015). We present our laboratory interpretations below.

c.10691 + 2T > C

This variant represents a single-nucleotide substitution at an intronic region of a canonical splice site. At the time of review, this variant was only identified in four Ashkenazi Jewish alleles within the gnomAD database (https://gnomad.broadinstitute.org/variant/7-21857959-T-C) and no homozygotes. Computational splicing predictions (Alamut, SpliceSiteFinder-like, MaxEntScan, NNSPLICE, GeneSplicer) (Shapiro and Senapathy 1987; Reese et al. 1997; Pertea et al. 2001; Yeo and Burge 2004) support a loss of the 5′ splice site. This variant is present in the literature (Retterer et al. 2016) and in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/variation/265100/#summary-evidence), but likely represents the same case as our report. The initial criteria used to assess this variant by our laboratory was PM2 ("Absent from controls or at an extremely low frequency if recessive"). Because this variant occurs at a canonical splice location, based on the nucleotide sequence, this variant is predicted to generate a skip of exon 65 (Fig. 2A, Isoform 1), but the joining of exon 64 and 66 preserves the open reading frame of the protein, reducing confidence that PVS1 ("Null variant in a gene where loss of function (LOF) is a known mechanism of disease") could be applied to this variant.

c.13523_13543dup21 (p.Ala4508_Leu4514dup)

This variant represents the duplication of 21 nt, resulting in an in-frame duplication of seven amino acids (Fig. 2A, right panel). This variant occurs in the last exon of the gene (82/82) and
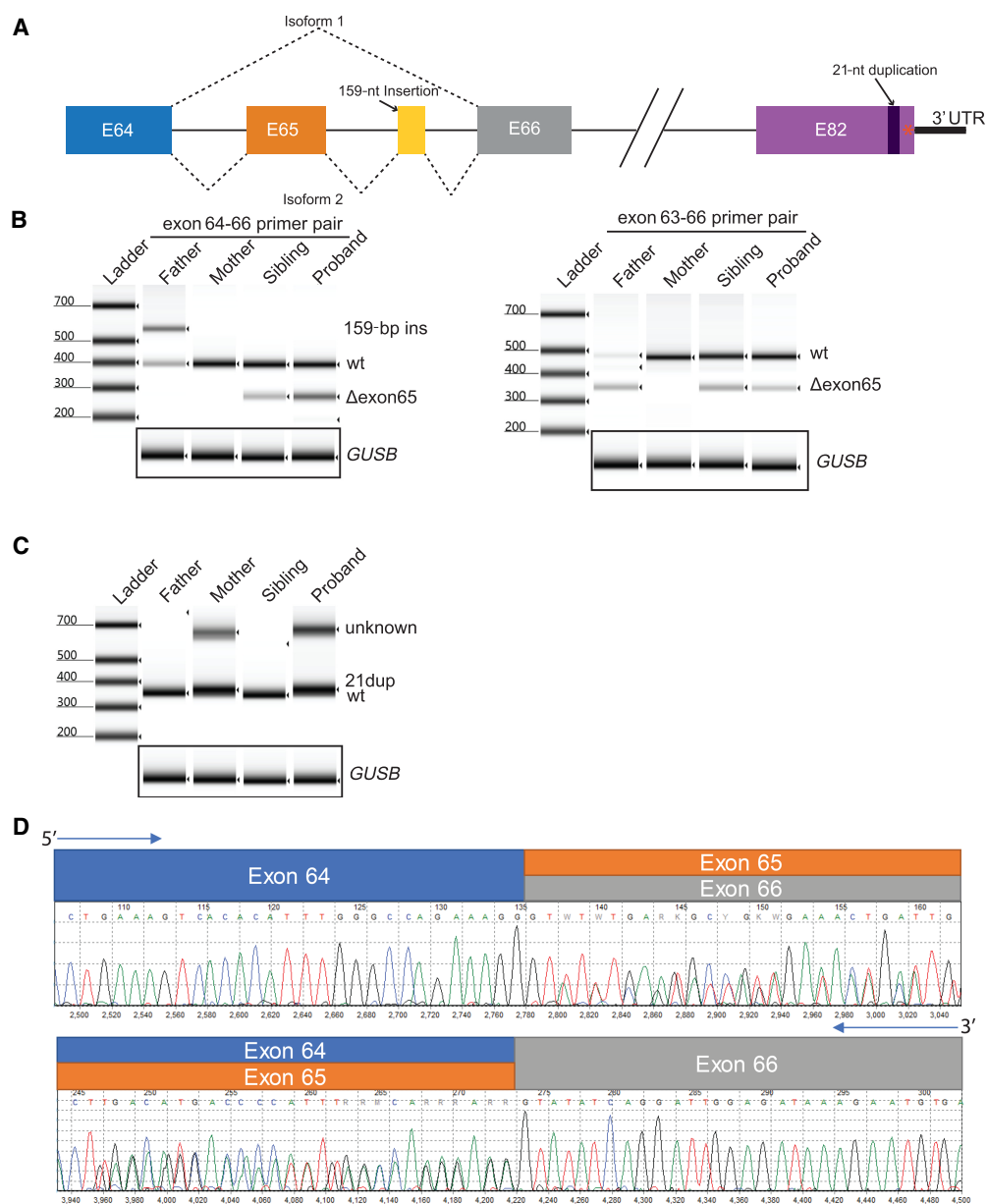
Figure 2. (*A*) Schematic of the potential effects of the two *DNAH11* variants. The *left* panel describes the effects of the c.10691 + 2T > C variant. Isoform 1 with the predicted skipping of exon 65 and the splicing of exon 64 and exon 66 together was the in silico predicted outcome of the variant. Isoform 2 is the alternative transcript that is observed in the father of the proband and does not encompass the skipping of exon 65, but instead utilizes the cryptic splice sites for the additional incorporation of 159 bp in between the junction of exon 65 and 66. This is expected to cause a shift in the open reading frame of the protein. The *right* panel illustrates the location of the c.13523_13543dup21 variant in the terminal exon 82, and proximal to the canonical stop codon (red asterisk). It is unclear if there are any additional splice effects due to this duplication. (*B*) Gel images depicting the PCR products that are obtained when amplification was conducted with a forward primer on exon 64 and the reverse primer on exon 66 of the *DNAH11* coding sequences. In the proband and her sibling, who are both carriers of the c.10691 + 2T > C variant, two PCR products are evident. A wild-type (wt) product of ~400 bp, and a ~280-bp product that is due to a skipping of exon 65 (Δexon65). The PCR amplification product derived from the father of the proband contains the wild-type band but has an additional band that is ~550 bp in size. (*Legend continues on next page.*)

is predicted to extend the protein but preserve the wild-type stop codon. The variant was absent from controls (gnomAD; https://gnomad.broadinstitute.org/) and was not identified in the literature (PubMed; https://www.ncbi.nlm.nih.gov/pubmed/), Human Genome Mutation Database (HGMD; https://portal.biobase-international.com/hgmd/pro/) or ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/). The initial criteria used to assess this variant were PM2 ("Absent from controls or at an extremely low frequency if recessive") and PM4 ("Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop loss variants").

Given the critical position of the canonical splice variant as well as the identification of an additional atypical variant in *trans* in this gene, consistent with the patient's phenotype, additional transcriptional analysis was undertaken to help clarify the impact of these two variants.

Transcript Analysis

Peripheral blood specimens from the proband, mother, father and unaffected sibling were processed for RNA extraction. The RNA specimens were reverse transcribed to cDNA. Transcription analysis utilizing PCR amplification and Sanger sequencing was performed to clarify the potential splicing effect of the variants.

c.10691 + 2T > C

After PCR amplification using primers flanking exons 64 and 66, three different possible PCR products were identified in the family members; wild-type (~400 bp), presumed deletion of exon 65 (~280 bp), and an unexpected larger product (~550 bp) (Fig. 2B). Sequencing analysis identified the ~280-bp product with a deleted exon 65 (Fig. 2D), and the larger product of ~550 bp as a 159-bp intronic sequence insertion from intron 65 (Fig. 3A,B). The 159-bp sequence (*DNAH11*: c.10691 + 8703_10691 + 8861) inserted into the alternate transcript, located at Chr 7[GRCh37]:21866660–21866818, is flanked by cryptic splice sites in the genomic DNA. Immediately upstream of the 159-bp intronic insertion is a 3′ splice site sequence (5′-YAG-3′, SpliceSiteFinder-like score = 81.27, MaxEntScan score = 9.96, NNSPLICE score = 0.04), and immediately downstream from the intronic insertion is a predicted 5′ splice

Figure 2. (*Continued.*) This was determined to be due to a 159-bp insertion between exon 65 and 66 of the *DNAH11* transcript by Sanger sequencing (159 bp ins). The housekeeping gene *GUSB* was amplified as a control to illustrate good quality of the starting cDNA specimens. The gel image on the right depicts the PCR products that are obtained when amplification was conducted with a forward primer on exon 63 and the reverse primer on exon 66 of the *DNAH11* coding sequences. Using this primer set, the products derived from carriers of the c.10691 + 2T > C variant are similar. The larger band that was present in the father was not detected by this primer pair, potentially because of preferential amplification of the smaller products. The wild-type band in the father is also lighter as compared to his offspring. (*C*) Gel images depicting the PCR products that are obtained when amplification was conducted with a forward primer on exon 82 and the reverse primer on 3′-UTR region of *DNAH11*. In the proband and her mother who are both carriers of the c.13523_13543dup21 variant, two main PCR products are evident. A wild-type product of ~350 bp, and a slightly bigger product that is most likely due to the additional 21-bp duplication. Unexpectedly, there is a ~700 bp product that is of unknown origin. Sanger sequencing did not capture any sequences additional other than the wild-type allele and the allele containing the 21-bp duplication. It is unclear if the larger product is due to an alternative transcript or a secondary structure derived from the duplicated allele. The housekeeping gene *GUSB* was amplified as a control to illustrate good quality of the starting cDNA. (*D*) Sanger electropherograms derived from sequencing the PCR product derived from an amplification reaction using the *DNAH11* exon 64 forward primer and exon 66 reverse primer on the proband specimen. The two electropherograms derived via bidirectional Sanger sequencing depict joining of the coding sequences between exon 64 and exon 66 on one allele of the coding sequences of *DNAH11*.
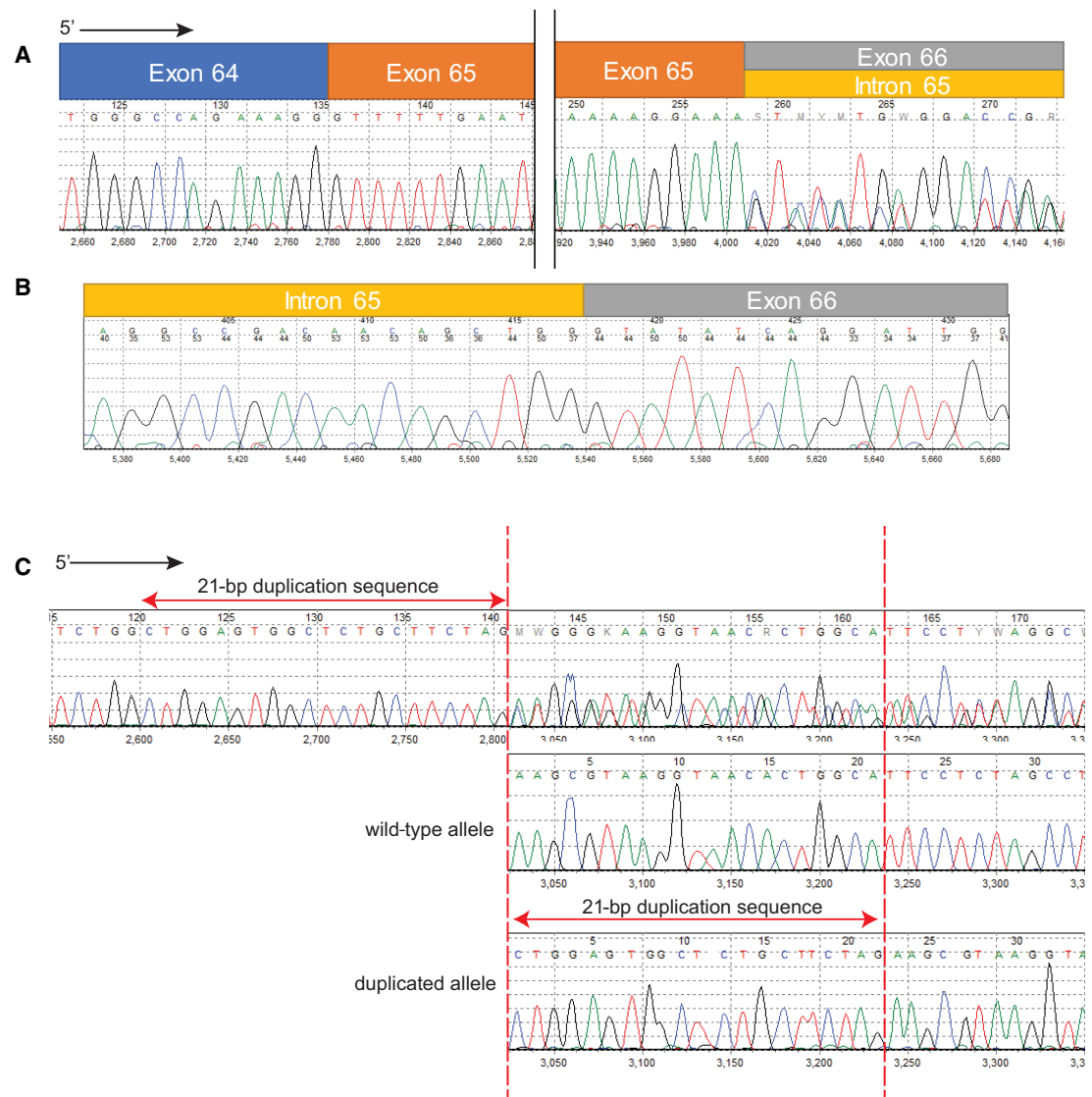
Figure 3. (*A*) Sanger electropherograms derived from sequencing the PCR product derived from an amplification reaction using the *DNAH11* exon 64 forward primer and exon 66 reverse primer on the father's specimen. The electropherogram depict joining of the coding sequences between exon 64 and 65 (*left* panel) and exon 65 with exon 66 on one allele and exon 65 with a 159-bp stretch of sequences from intron 65 on the other allele. The start of the 159-bp intronic sequence is at c.10691 + 8703. (*B*) Sanger electropherograms derived from forward strand sequencing the PCR product derived from an amplification reaction using the *DNAH11* intron 65 forward primer, and exon 66 reverse primer on the father's specimen. The electropherogram depict joining of the coding sequences between the end of the 159-bp insertion at c.10691 + 8861 from intron 65 and exon 66. (*C*) Sanger sequencing electropherogram of the proband, demonstrating the 21-bp duplication in the region of the c.13523_13543dup21 variant. The *top* panel depicts the sequence obtained when a forward primer was used in the sequencing reaction. The *middle* and *bottom* panels show that the sequences obtained by the overlapping peaks were deconvoluted into the two alleles. The heterozygous indel peaks were due to the duplication of the preceding 21 bp in the region (denoted by red arrows).

site sequence (5′-GUAAUUAC-3′, SpliceSiteFinder-like score = 74.52, MaxEntScan score = 6.19, NNSPLICE score = 0.98). Within the Sanger sequence data, exon 66 sequence resumes after the 159-bp intronic insertion (Fig. 3A). These data suggest that both of the cryptic splice sites that flank the inserted sequence were used during splicing of this cryptic exon: the cryptic 3′ splice site upstream of the cryptic exon was used as a splice acceptor with the genuine 5′ splice site in intron 65, and the cryptic 5′ splice site downstream from the cryptic exon was used as a splice donor with the genuine 3′ splice site in intron 65. The consequence of inclusion of these cryptic intronic sequences into the open reading frame is a frameshift mutation and a premature termination codon after 17 codons (NM_001277115.1(*DNAH11*): c.10691_10691 + 1ins159, p.Gly3563_Lys3564ins17*).

Of note, this larger product was only visualized in the father, who is a carrier of this variant. This product was not visualized in the proband or the unaffected sibling, who are both carriers of this variant. Also of note, this product was absent from all family members on repeat assay with different primers. As mRNA levels may physiologically fluctuate over time, it is possible that this transcript is produced at a lower level not detectable by our PCR reaction,or is subjected to nonsense-mediated decay (NMD), preventing its detection in the proband and her sibling. However, the presence and detection of this alternate transcript indicates that the effect of this variant is not only the in-frame skipping of a singular exon. It is possible that both children also generate this larger transcript that could have been visualized if there is a mechanism to stall the process of NMD and enable detection of the aberrant transcript.

c.13523_13543dup21 (p.Ala4508_Leu4514dup)

After PCR amplification using primers flanking exon 82 and the 3′-UTR, two variable products (∼350 bp) comprising the wild-type sequences and sequences with the 21-bp duplication were identified in the mother and proband who are carriers of this variant. Sanger sequencing verified the presence of the 21-bp duplication in the heterozygous state in the mother and the proband (proband data shown in Fig. 3C). However, an unexpected third PCR product band of ∼700 bp in size was also detected in both individuals (labeled as "unknown") (Fig. 2C). Sanger sequencing failed to resolve the sequences in the 700-bp product even after gel separation; therefore, it is possible that the product is due to the formation of a secondary nucleic acid structure caused by the duplicated sequences. Additional in silico prediction models were performed for this variant, which suggested a deleterious impact of the duplication (SIFT: damaging, FATTHM: pathogenic [score = 0.853], Provean: deleterious [score = −5.11]) (Hu and Ng 2013; Choi and Chan 2015; Ferlaino et al. 2017).

MMSplice Analysis of the c.10691 + 2T > C Variant

To further our attempts at quantifying the potential effects of this splice variant, we completed the MMSplice prediction tool that serves to predict quantitatively the physical measures of splicing (Cheng et al. 2019). Using this tool, a prediction score (delta_logit_psi) is generated that quantifies the likelihood of a splicing event that either includes or excludes the exon in question. Delta_logit_psi is the main score predicted by MMSplice, which shows the effect of the variant on the inclusion level (PSI percent spliced in) of the exon. Because the score is on a logit scale, if the score is positive, it shows that variant leads to higher inclusion rate for the exon. If the score is negative, it shows that variant leads to higher exclusion rate for the exon. If delta_logit_psi is >2 or <−2, the effect of the variant can be considered strong. For this particular variant, the calculated delta_logit_psi score was −1.73, predicting that the variant leads to a higher exclusion rate for exon 65. But because the value is <−2, the effect of the variant is not considered strong and could be supportive of the notion that the skipping of exon 65 is not definitive. This leaves room for in vitro transcript analysis that will

provide confirmatory results of the variant transcripts. Overall this variant was given a pathogenicity score of 0.995 by MMSplice.

MMSplice Analysis for the c.13523_13543dup21 Variant

For this particular 21-bp duplication variant, the calculated delta_logit_psi score was 0.0457, which does not indicate any significant exon inclusion/exclusion events. The variant was given a pathogenicity score of 0.954 by MMSplice.

Outcomes

c.10691 + 2T > C

Data from these studies confirmed that the canonical splice variant causes an in-frame skip of one exon; however, our analysis also detected an alternate transcript with a 159-bp intronic insertion due to utilization of two cryptic splice sites. These results suggested a nonsingular outcome for the canonical splicing variant, including the generation of a frameshift and premature stop codon that causes loss of function of the protein. Given the data supporting a premature termination codon outcome, PVS1 (loss of function) criteria was more confidently applied to the interpretation. In addition, PS3 ("Well established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product") was also applied to this variant, reflecting the usage of the transcriptional analysis. Of note, to avoid "double counting" the same data point generated by these studies, the strength of PVS1 was reduced to Moderate and the strength of PS3 was also reduced to Moderate. The utilization of these criteria enabled the upgrading of the classification from VUS to Likely Pathogenic. (PVS1_Moderate + PS3_Moderate + PM2 = Likely Pathogenic).

c.13523_13543dup21 (p.Ala4508_Leu4514dup)

Although data from transcriptional analysis was obtained for this variant that is suggestive of additional transcriptional or structural impact due to the presence of the 21-bp duplication, additional ACMG criteria could not be applied. However, because of the change in classification for the canonical splicing variant to Likely Pathogenic, professional judgment was used to apply PM3 ("For recessive disorders, detected in trans with a pathogenic variant") and was added to the interpretation. The utilization of this criteria enabled the upgrading of the classification from VUS to Likely Pathogenic (PM2 + PM3 + PM4 = Likely Pathogenic).

DISCUSSION

The publication of guidelines by the ACMG to help interpret sequence variants has aided the practice of variant interpretation greatly. However, VUS results are still common, particularly for genes and conditions of which characterization is still emerging such as PCD. A significant barrier to conclusive variant interpretation centers around insufficient data to determine pathogenicity. Splicing variants can be especially difficult to reconcile and rely heavily on computational prediction tools to interpret the potential impact to the splicing machinery. Although some variant positions are intuitive to assess for splicing impacts, it is important to complete an in silico assessment for any potential splicing effects for all detected DNA variants (including variants that cause in-frame duplications and deletions) and not only for variants that more obviously impact a canonical splice site at the ±1 or 2 intronic position. Transcriptional analysis represents a powerful tool to assist in variant classification. By giving additional predictive context to the downstream protein product beyond the DNA level, the

true nature of a variant can be more accurately assessed. As evidenced by our case, the position of the variant at a canonical splice site (c.10691 + 2T > C) was recognized as highly suspicious, but analysis predicting only an in-frame loss of a single exon resulted in a downgraded classification. Only by additional transcriptional analysis conducted in the laboratory were we able to verify the identification of a premature stop codon in another alternative transcript. We propose that transcriptional analysis be more routinely conducted in the clinical laboratories involved in genetic testing, especially on variants with a predicted splicing impact, as additional information obtained can be extremely valuable for variant interpretation.

Although transcriptional analysis has clear benefits, it is not without limitations and is useful for some, but not all, situations and variants and may present some variable data because the analyte of choice is the relatively labile RNA species. During conversion of RNA into cDNA for the purpose of transcript analysis, there is some potential for biases (Minshall and Git 2020). Using a nonquantitative methodology like PCR amplification and Sanger sequencing may also miss the detection of transcripts that are lowly expressed or degraded because of nonsense-mediated decay. A more comprehensive methodology like whole-transcriptome profiling using RNA-seq or a molecular biology approach of utilizing agents such as cycloheximide that may inhibit mRNA degradation before transcriptional analysis may provide the improved sensitivity in detecting the variant transcripts. However, these approaches would be cost-prohibitive or likely out of the scope of practice for most clinical laboratories to undertake. Therefore it is important to be aware of the limitations in our methodology of choice, as this may have caused the unexpected results to be observed in this family (only the father's specimen exhibited the 159-bp insertion). It is possible that the same product is present in the proband and the carrier sibling but not observed because of degradation by NMD. It is also of interest that in the characterized transcript with the insertion of the 159 bp, exon 65 was not noted to be skipped, as exon 65 would have been expected to be skipped if the variant caused significant weakening of the canonical splice donor site. Moreover, based on the predictions by the MMSplice software using the delta_-logit_psi output, the predicted exclusion of exon 65 is not defined as "strong." A previous publication (Anna and Monika 2019) provided evidence of the potential array of variable transcripts that can be generated, even though suggested mutations at canonical splice site sequences are usually only predicted to lead to single exon skipping. Overall this data underscores that splicing predictors are less than perfect, and unexpected outcomes can only be visualized with additional studies like transcript analysis. In silico predictive tools like MMSplice may also be utilized to provide a quantitative assessment of the splicing impact of the variants before transcript analysis is undertaken.

For the second 21-bp duplication, transcriptional analysis allowed visualization of an additional large band. Interestingly, we have been unable to conduct Sanger sequencing of the larger 700-bp PCR product for this variant, even after gel separation to resolve the 350- and 700-bp product. It is possible that the effects of the 21-bp duplication are purely structural and may be affecting protein interaction and subsequent transport of the protein in its cellular context. In support of this hypothesis, missense variants in this carboxy-terminal motor domain of the protein have been reported in the literature, suggesting nonframeshift sequence changes in this region as a potential disease mechanism in addition to the well-established NMD and subsequent loss of function pathogenesis (Knowles et al. 2012). Furthermore, loss of function variants proximal to the termination codon have also been found to be disease-causing even though they are less likely to trigger NMD, suggesting that preservation of wild-type sequences in the carboxy-terminus domain may be important for normal gene function. Therefore, functional studies of a different nature may be needed to help clarify the significance of this variant further. In addition to variant type considerations, it is critical to note that RNA studies represent a "snapshot" in time and are

representative of the transcript expression from the sample tissue source. As such, data from RNA transcriptional analysis may vary in the same patient from time to time, may not detect transcripts that are subjected to NMD, and require critical judgment in its interpretation. Finally, although transcriptional analysis for detection of causative variants has been demonstrated in nonaffected tissue types, for certain diseases a peripheral blood sample may not provide the right specimen type for analysis if the gene is not well-expressed in hematopoietic lines, or this process may be occurring at a lower level relative to wild-type splicing (Kremer et al. 2017). For our case, the presence of the isoform resulting in a premature stop in the unaffected father, but not the affected proband or the unaffected sibling, may represent a selection of these confounding factors complicating the interpretation of the result and supporting the potential necessity for additional samples and/or of a different type to clarify the effect.

Barriers to implementation and broad usage of transcriptional analysis in a clinical laboratory also exist at a practical level. As yield for transcriptional analysis is dependent on variant type, both ordering clinicians and laboratory personnel must collaborate and exercise discretion to determine for which variants this may be useful. Variants that would benefit from transcriptional analysis are only identified once sequencing is complete. RNA is inherently less stable than DNA and an additional subsequent draw from patients and family members would likely be required. As a high number of genetic tests are not performed within the ordering institution, obtaining additional samples to perform transcriptional analysis presents additional logistical challenges.

An additional difficulty with variant interpretation lies within the variance and best usage application of the guideline criteria. Our case report provides an excellent example: The application of transcriptional analysis established the potential usage of PVS1 (loss of function) for the c.10691 + 2T > C variant, but the question was introduced as to the inclusion of PS3 ("Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product"). Although transcriptional analysis can be considered a functional study, it is unclear whether the application of this analysis to prove usage of PVS1 would be "double counting" criteria from the same data point and if the additional utilization of PS3 in the interpretation is appropriate. The published guidelines do not provide detailed insight beyond "expert judgment" on moderating the impact of a given criteria. Additional recommendations providing further granularity for specific criteria have been published, but these publications also acknowledge that further work must be done, including how criteria may be used in relation to each other with PVS1 and PS3 named as a specific example (Abou Tayoun et al. 2018; Brnich et al. 2019). At default strength, application of PVS1 and PS3 to this variant would achieve a classification of Pathogenic. However, given the nature of the transcriptional analysis assay in comparison to a model organism experiment and being cognizant of overcrediting the same criteria, a reduction in strength of the criteria to Moderate was applied (PVS1_Moderate and PS3_Moderate) and resulted in a classification of Likely Pathogenic.

Furthermore, determination on usage of these criteria have downstream impacts for the c.13523_13543dup21 variant. If the c.10691 + 2T > C variant had achieved a classification of Pathogenic, this would have allowed default usage of PM3 ("For recessive disorders, detected in trans with a pathogenic variant") for the c.13523_13543dup21 variant. With default application of PVS1 and PS3, the classification of the c.10691 + 2T > C variant and the c.13523_13543dup21 variant achieve a Pathogenic and Likely Pathogenic classification, respectively; with reduced strength, the variants are default classified as Likely Pathogenic and VUS. Given that the ACMG–AMP guidelines suggest that "expert judgment" should be utilized when applying these criteria and that a classification of Likely Pathogenic represents a >90% certainty of pathogenicity, the decision was made to utilize PM3 for the c.13523_13543dup21 variant to reflect the likely impact of this variant with regards to

disease. Therefore, the decision on application of criteria, even at reduced strength, represents an enormous impact in the classification of both these variants.

VUS results represent a significant source of frustration for patients and clinicians as well. In addition to the uncertainty of a VUS and the subtleties of accurate interpretation of its meaning, to further the complexity, classifications may undergo reassessment and can be changed as more information is published about the variant or gene. Thus, it is imperative that patients are counseled that the classification of variants are only accurate with the current knowledge of the genetics for the disease. Most often in light of additional data, classification can be upgraded from VUS to Pathogenic or Likely Pathogenic, but can also be downgraded from Pathogenic down to VUS/Likely Benign/Benign. This may occur after a diagnosis has been given, medical management has been established, and patients have reconciled the impact of a positive diagnosis on their life.

The current rate of novel genetic discovery within PCD is accelerated and gene-specific characterization (e.g., mechanism of disease, critical domains) is ongoing. Given the scarcity of institutions equipped with access to all diagnostic tools for PCD, genetic testing in diagnosis of PCD is of high importance. In light of this, it is crucial to understand the nuances and the potential unique impacts that variants may have within the gene and how it relates to the phenotype. Illustrating the dichotomous nature of a clinical vs. molecular diagnosis, our case initially presented with classical features of PCD but an incomplete molecular diagnosis. As evidenced by our case report, transcriptional analysis may aid in VUS resolution but is not an all-encompassing solution. Although both variants were reclassified by our work, transcriptional analysis was only able to directly reclassify the c.10691 + 2T > C variant to Likely Pathogenic based on the available criteria. Data sharing, such as within the ClinVar database or collaborative expert review groups such as the ClinGen consortia (https://clinicalgenome.org/) also represent a key direction toward better understanding of these genes and variant interpretation.

In summary, this case report highlights the utility of transcriptional analysis to assist in variant classification by giving insight on the downstream impact of a variant on transcription. Given the high rate of VUS results in the genetic testing of rare pediatric conditions, particularly in conditions like PCD with continual genetic discoveries, transcriptional analysis adds significant value toward variant resolution and should be utilized when available.

METHODS

Total RNA was extracted from 1–3 mL EDTA-anticoagulated whole blood using QIAGEN QIAamp RNA Blood Mini Kits according to manufacturer instructions. The extracted total RNA was reverse-transcribed to cDNA using Invitrogen Superscript Enzyme mix according manufacturer instructions. M13-tailed PCR primers flanking exon 65 (DNAH11-e64a-F1, 5′-TGTAAAACGACGGCCAGT AGAATGTCCACCGAAAATGC-3′; DNAH11-e64b-F2, 5′-TG TAAAACGACGGCCAGT TCCATTCCACTAACCGAAGG-3′; DNAH11-e66-R1/2, 5′-CAGGA AACAGCTATGACC TGCCAATTTTGTGTGAAGGA-3′) and primers flanking the termination codon (DNAH11-e82-F1, 5′-TGTAAAACGACGGCCAGTCCGTGGACAGACAAG AAACC-3′ and DNAH11-3′UTR-R1, 5′-CAGGAAACAGCTATGACCGACTTTCACTCTAAG GATAGCG TTG-3′) were used to PCR-amplify and Sanger-sequence the cDNA. cDNA product was visualized using the Agilent 4200 Tapestation system. Soft Genetics Mutation Surveyor (v5.1.2) was used to analyze Sanger sequence data, which was aligned against the *DNAH11* NCBI Reference Sequence NM_001277115. Alamut Visual v2.11 was used to assess the transcript variants, including splicing effect prediction with SpliceSiteFinder-like, MaxEntScan, NNSPLICE, and GeneSplicer (Shapiro and Senapathy 1987; Reese et al. 1997; Pertea et al. 2001; Yeo and Burge 2004).

ADDITIONAL INFORMATION

Data Deposition and Access

Both variants reported in this case report have been submitted to ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) under accession numbers SCV001164095.1 and SCV001164096.1.

Ethics Statement

Genetic analysis for this family was performed as clinical testing; transcriptional analysis is performed as part of an overall assessment for variant interpretation and/or resolution. Verbal consent for publication was obtained from the family. This study was undertaken as part of the clinical work-up and clinical variant resolution workflow of the Lurie Molecular Diagnostics Laboratory.

Acknowledgments

The authors thank the family for their enthusiasm in this work and in furthering the knowledge of PCD and genetic testing. The authors thank Dr. Larry Jennings and David Dittmann for their technical expertise and assistance.

Author Contributions

A.I., T.L., J.C., V.S., M.K.M., J.S., and K.F. are part of this patient's clinical team and contributed to the management of the patient's medical care. A.I., A.W., C.M., D.K., E.T., S.K., and K.L.Y. contributed to the laboratory analysis of this family's variants. A.I., A.W., K.L.Y., S.K., and T.L. generated the manuscript. All authors reviewed the manuscript prior to submission.

REFERENCES

Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG, Harrison SM, ClinGen Sequence Variant Interpretation Working Group (ClinGen SVI). 2018. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* **39:** 1517–1524. doi:10.1002/humu.23626

Andjelkovic M, Minic P, Vreca M, Stojiljkovic M, Skakic A, Sovtic A, Rodic M, Skodric-Trifunovic V, Maric N, Visekruna J, et al. 2018. Genomic profiling supports the diagnosis of primary ciliary dyskinesia and reveals novel candidate genes and genetic variants. *PLoS ONE* **13:** e0205422. doi:10.1371/journal.pone.0205422

Anna A, Monika G. 2019. Splicing mutations in human genetic disorders: examples, detection, and confirmation [published correction appears in 2019 *J Appl Genet* 2019 **60:** 231]. *J Appl Genet* **59:** 253–268. doi:10.1007/s13353-018-0444-7

Boaretto F, Snijders D, Salvoro C, Spalletta A, Mostacciuolo ML, Collura M, Cazzato S, Girosi D, Silvestri M, Rossi GA, et al. 2016. Diagnosis of primary ciliary dyskinesia by a targeted next-generation sequencing panel: molecular and clinical findings in Italian patients. *J Mol Diagn* **18:** 912–922. doi:10.1016/j.jmoldx.2016.07.002

Bonnefoy S, Watson CM, Kernohan KD, Lemos M, Hutchinson S, Poulter JA, Crinnion LA, Berry I, Simmonds J, Vasudevan P, et al. 2018. Biallelic mutations in LRRC56, encoding a protein associated with intraflagellar transport, cause mucociliary clearance and laterality defects. *Am J Hum Genet* **103:** 727–739. doi:10.1016/j.ajhg.2018.10.003

Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, Kanavy DM, Luo X, McNulty SM, Starita LM, 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. doi:10.1186/s13073-019-0690-2

Cheng J, Nguyen T, Cygan KJ, Çelik MH, Fairbrother WG, Avsec Ž, Gagneur J. 2019. MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol* **20:** 48. doi:10.1186/s13059-019-1653-z

Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31:** 2745–2747. doi:10.1093/bioinformatics/btv195

Djakow J, Kramná L, Dušátková L, Uhlík J, Pursiheimo JP, Svobodová T, Pohunek P, Cinek O. 2016. An effective combination of Sanger and next generation sequencing in diagnostics of primary ciliary dyskinesia. *Pediatr Pulmonol* **51:** 498–509. doi:10.1002/ppul.23261

Fassad MR, Shoemark A, Legendre M, Hirst RA, Koll F, le Borgne P, Louis B, Daudvohra F, Patel MP, Thomas L, et al. 2018. Mutations in outer dynein arm heavy chain *DNAH9* cause motile cilia defects and situs inversus. *Am J Hum Genet* **103:** 984–994. doi:10.1016/j.ajhg.2018.10.016

Ferlaino M, Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. 2017. An integrative approach to predicting the functional effects of small indels in non-coding regions of the human genome. *BMC Bioinformatics* **18:** 442. doi:10.1186/s12859-017-1862-y

Hu J, Ng PC. 2013. SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. *PLoS ONE* **8:** e77940. doi:10.1371/journal.pone.0077940

Jayasinghe RG, Cao S, Gao Q, Wendl MC, Vo NS, Reynolds SM, Zhao Y, Climente-González H, Chai S, Wang F, et al. 2018. Systematic analysis of splice-site-creating mutations in cancer. *Cell Rep* **23:** 270–281.e3. doi:10.1016/j.celrep.2018.03.052

Kim RH, Hall D A, Cutz E, Knowles MR, Nelligan KA, Nykamp K, Zariwala MA, Dell SD. 2014. The role of molecular genetic analysis in the diagnosis of primary ciliary dyskinesia. *Ann Am Thorac Soc* **11:** 351–359. doi:10.1513/AnnalsATS.201306-194OC

Knowles MR, Leigh MW, Carson JL, Davis SD, Dell SD, Ferkol TW, Olivier KN, Sagel SD, Rosenfeld M, Burns KA, et al. 2012. Mutations of *DNAH11* in patients with primary ciliary dyskinesia with normal ciliary ultrastructure. *Thorax* **67:** 433–441. doi:10.1136/thoraxjnl-2011-200301

Kremer LS, Bader DM, Mertes C, Kopajtich R, Pichler G, Iuso A, Haack TB, Graf E, Schwarzmayr T, Terrile C, et al. 2017. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun* **8:** 15824. doi:10.1038/ncomms15824

Leigh MW, Ferkol TW, Davis SD, Lee HS, Rosenfeld M, Dell SD, Sagel SD, Milla C, Olivier KN, Sullivan KM, et al. 2016. Clinical features and associated likelihood of primary ciliary dyskinesia in children and adolescents. *Ann Am Thorac Soc* **13:** 1305–1313. doi:10.1513/AnnalsATS.201511-748OC

Leigh MW, Horani A, Kinghorn B, O'Connor MG, Zariwala MA, Knowles MR. 2019. Primary ciliary dyskinesia (PCD): a genetic disorder of motile cilia. *Transl Sci Rare Dis* **4:** 51–75. doi: 10.3233/TRD-190036

Loges NT, Antony D, Maver A, Deardorff MA, Güleç EY, Gezdirici A, Nöthe-Menchen T, Höben IM, Jelten L, Frank D, et al. 2018. Recessive *DNAH9* loss-of-function mutations cause laterality defects and subtle respiratory ciliary-beating defects. *Am J Hum Genet* **103:** 995–1008. doi:10.1016/j.ajhg.2018.10.020

Marshall CR, Scherer SW, Zariwala MA, Lau L, Paton TA, Stockley T, Jobling RK, Ray PN, Knowles MR, Canada Consortium FORGE, et al. 2015. Whole-exome sequencing and targeted copy number analysis in primary ciliary dyskinesia. *G3 (Bethesda)* **5:** 1775–1781. doi:10.1534/g3.115.019851

Minshall N, Git A. 2020. Enzyme- and gene-specific biases in reverse transcription of RNA raise concerns for evaluating gene expression. *Sci Rep* **10:** 8151. doi:10.1038/s41598-020-65005-0

Pertea M, Lin X, Salzberg SL. 2001. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* **29:** 1185–1190. doi:10.1093/nar/29.5.1185

Reese MG, Eeckman FH, Kulp D, Haussler D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4:** 311–323. doi: 10.1089/cmb.1997.4.311

Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG, et al. 2016. Clinical application of whole-exome sequencing across clinical indications. *Genet Med* **18:** 696–704. doi:10.1038/gim.2015.148

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, 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. doi:10.1038/gim.2015.30

Shapiro MB, Senapathy P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15:** 7155–7174. doi:10.1093/nar/15.17.7155

Shapiro AJ, Davis SD, Polineni D, Manion M, Rosenfeld M, Dell SD, Chilvers MA, Ferkol TW, Zariwala MA, Sagel SD, et al. 2018. Diagnosis of primary ciliary dyskinesia. An official American Thoracic Society clinical practice guideline. *Am J Respir Crit Care Med* **197:** e24–e39. doi:10.1164/rccm.201805-0819ST

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11:** 377–394. doi:10.1089/1066527041410418