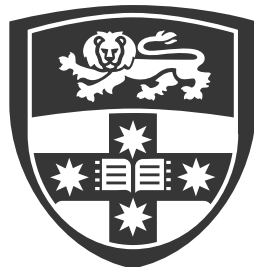# Optimising RNA diagnostics for implementation into clinical practice.

Adam M. Bournazos

Primary supervisor: Professor Sandra T. Cooper

Auxiliary supervisors: Clinical Professor Bruce Bennetts and Clinical Associate Professor Kristi J. Jones

A thesis submitted to fulfil requirements for the degree of Doctor of Philosophy

THE UNIVERSITY OF SYDNEY

Kids Neuroscience Centre, Kids Research

The Children's Hospital at Westmead, Westmead, NSW, Australia

Sydney Medical School, Faculty of Medicine and Health

The University of Sydney, Sydney, NSW, Australia

July 2022

Statement of originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Adam M. Bournazos

July 2022

Authorship attribution statement

Section 2.2 of this thesis was published as:

Akesson LS., **<u>Bournazos A.</u>**, Fennell A, et al. Rapid exome sequencing and adjunct RNA studies confirm the pathogenicity of a novel homozygous *ASNS* splicing variant in a critically ill neonate. *Hum Mutat*. 2020;41(11):1884-1891.

I performed the *in silico* and RNA analysis (RT-PCR and RNA sequencing) and wrote the research testing report to experimentally confirm pathogenicity of a variant identified by rapid exome sequencing. I co-wrote the RNA analysis methods and results, and prepared figure 1d-e, figure 2, figure S2, figure S3.

This paper was a collaboration, procuring a genetic diagnosis for a newborn with cerebellar hypoplasia who underwent rapid exome sequencing as part of a rapid genomic testing program. Lauren S. Akesson, Andrew Fennell and I contributed equally as joint first authors.

Section 3.2 of this thesis was published as:

<u>Bournazos AM</u>, Riley LG, Bommireddipalli S, et al. Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genet Med*. 2021;24(1):130-145.

This paper was the result of a large collaboration where I performed the *in silico* and RNA analysis (RT-PCR and RNA sequencing) to inform variant classification for 67 of the 74 families with genetic variants predicted to impact pre-mRNA splicing. I prepared and co-wrote the manuscript and prepared all the figures and tables.

Section 4.2 of this thesis was published as:

Dawes R, **Bournazos AM**, Bryen SJ et al. SpliceVault: predicting the precise nature of variant-associated mis-splicing. *Nat Genet (NG-AN59054R1)*. Under Revision.

Ruebena Dawes and I contributed equally as joint first authors. I performed experimental validation for the majority of splicing variant cohort and RNA re-analysis, Figure 1A,B, Figure 3 and Figure 4, writing and editing manuscript.

The appendix also includes the following publications:

Jones HF, Bryen SJ, Waddell LB, et al. Importance of muscle biopsy to establish pathogenicity of DMD missense and splice variants. *Neuromuscul Disord*. 2019;29(12):913-919.

I performed the western blots and standard curve for dystrophin quantitation and prepared figure 1C.

Bryen SJ, Ewans LJ, Pinner J, et al. Recurrent TTN metatranscript-only c.39974-11T>G splice variant associated with autosomal recessive arthrogryposis multiplex congenita and myopathy. *Hum Mutat*. 2020;41(2):403-411.

I performed the RNA extraction from a patient muscle biopsy, *in silico* analysis and primer design for RT-PCR.

Waddell LB, Bryen SJ, Cummings BB, et al. WGS and RNA studies diagnose noncoding DMD variants in males with high creatine kinase. *Neurol Genet*. 2021;7(1):e554.

I performed the western blots and standard curve for dystrophin quantitation used in figure 4.

Katiyar D, Anderson N, Bommireddipalli S, et al. Two novel *B9D1* variants causing Joubert syndrome: Utility of mRNA and splicing studies. *Eur J Med Genet*. 2020;63(9):104000.

I performed the *in silico* and RNA analysis and prepared figure 2.

Huq AJ, Thompson BA, Bennett MF, et al. Clinical Impact of Whole Genome Sequencing in Patients with Early Onset Dementia. *J Neurol Neurosurg Psychiatry*. 2022 In Press.

I performed the in silico and RNA analysis for a patient with compound heterozygous variants in SPG21 and prepared figure 4C.

Adam M. Bournazos

July 2022

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Professor Sandra T. Cooper

July 2022

Competing Interests

Professor Sandra Cooper is director of Frontier Genomics Pty Ltd (Australia). Professor Cooper receives no remuneration (salary or consultancy fees) for this role. Frontier Genomics Pty Ltd (Australia) has no existing financial relationships that will benefit from publication of these data.

Acknowledgements

A special thank you to my wife Dr Melissa Gardiner for your endless support, particularly through the pandemic with our beautiful children Nathan and Lachlan.

Abstract

Background

Genetic variants that elicit aberrant splicing of pre-messenger RNA (pre-mRNA) are recognised as causative variants in ~30-50% of genetic disorders. However, it is still not possible to predict reliably if and how a variant will impact splicing, limiting the application of *in silico* splice prediction tools in variant interpretation. Most splicing variants fall outside the essential splice site and, in the absence of RNA testing, remain classed variants of uncertain significance (VUS) according to ACMG-AMP (American College of Medical Genetics and Genomics and Association for Molecular Pathology) guidelines. Sequence analysis of spliced messenger RNA (mRNA) is the only definitive means to determine the precise nature of variant associated mis-splicing. Tissues with limited accessibility, for instance vital organs, present a challenge for RNA testing of genes with tissue specific expression. Fortunately, clinically accessible tissues such as blood and fibroblasts can be used to infer variant associated mis-splicing outcomes in the manifesting tissue. A further challenge arises from the lack of guidance on how functional evidence (PS3/BS3 criteria) from RNA studies should be applied to variant interpretation within the current ACMG-AMP framework. There is an urgent need to establish ACMG-AMP aligned quality standards and guidelines for complex RNA assay data for accurate and consistent variant interpretation between clinical laboratories.

Methods

Families were recruited from local area health districts across Australia and New Zealand using inclusion criteria to ascertain putative splicing variants with high clinical suspicion of causality. More than 120 families with diverse monogenic conditions were triaged into PCR-based RNA testing, with comparative RNA-sequencing for 38 cases. Consensus ascertainment criteria, standard practices for PCR-based RNA testing, and RNA assay interpretation rubric were devised through consultation with the clinical and molecular genetics community via surveys, live polls and SpliceACORD consortium (Australasian Consortium for RNA Diagnostics) meetings.

Results

Informative RNA assay data was obtained for 96% cases, enabling variant re-classification for 75% of variants. RNA testing reports were used to guide clinical care and genetic counselling, and 75% of diagnosis were clinician-reported to have a positive impact for the family. PCR-based RNA diagnostics has the capacity to analyse 81.3% of clinically significant genes and to allow phasing of RNA splicing events. Variant associated mis-splicing was highly reproducible between affected individuals and heterozygotes, and between different biospecimens.

Discussion

We provide a standardised protocol for PCR-based RNA testing and ACMG-AMP aligned recommendations for the interpretation of RNA assay data. Our study

demonstrates the significant diagnostic and health benefits of RNA analysis as adjunct testing to extend diagnostic yield from genomic testing.

Conference and other proceedings

<u>Bournazos AM</u> and Hunter MF. PCR tricks to exploit heterozygous coding SNVs to discern partial from complete mis-splicing. Oral presentation: HGSA 44th ASM. SpliceACORD RNA Workshop: Toward a consensus for pathology interpretation of RNA Assay data for variant re-classification 2021 August 14-17, Adelaide, SA.

<u>Bournazos AM</u>, Bommireddipalli S, Edwards C, et al. Abnormal initiation of transcription: an important consideration for RNA analysis of intron-1 donor splice site variants. Oral presentation: Human Genetics Society of Australasia 44th Annual Scientific Meeting 2021 August 14-17, Adelaide, SA.

<u>Bournazos AM</u>, Bommireddipalli S, Edwards C, et al. Abnormal initiation of transcription: an important consideration for RNA analysis of intron-1 donor splice site variants. Poster presentation: Precision Medicine: A Revolution in Patient Care Conference 2021 May 20-21, Sydney, NSW.

<u>Bournazos A</u>, Riley LG, Bommireddipalli S, et al. Rapid Implementation of rapid splicing studies using blood or skin informs diagnosis and management in clinical and acute care genomics. Poster session: Australian Functional Genomics Conference, 2019 November 22-23, Sydney, NSW.

Bournazos A, Riley LG, Bommireddipalli S, et al. Rapid Implementation of rapid splicing studies using blood or skin informs diagnosis and management in clinical and acute care genomics. Oral presentation: American Society of Human Genetics Annual Meeting, 2019 October 15-19, Houston, Texas.

Bournazos A, Riley LG, Bommireddipalli S, et al. Rapid RT-PCR studies using blood or fibroblast mRNA improves diagnostic yield for putative splice variants in tissue-specific genes. Poster presentation: Westmead Research Showcase, 2019 September 11, Sydney, NSW. Awarded Runner up poster prize.

Bournazos A, Riley LG, Bommireddipalli S, et al. Improving diagnostic yield in clinical and acute care genomics: Rapid PCR studies of splicing variants in tissue-specific genes using blood or fibroblast mRNA. Oral presentation: Australasian Neuromuscular Network Annual Scientific Meeting, 2019 September 3, Sydney, NSW.

Bournazos A, Riley LG, Bommireddipalli S, et al. Rapid RT-PCR analysis of putative splice variants enables molecular diagnosis of rare genetic diseases. Oral presentation: The University of Sydney Children's Hospital Westmead Clinical School 2019 HDR Student Conference, 2019 August 16, Sydney, NSW.

Abbreviations

ACMG   American College of Medical Genetics and Genomics

ACORD   Australian Consortium for RNA Diagnostics

AMP   Association for Molecular Pathology

Bp   Base pair

ClinGen   The National Institutes of Health Clinical Genome Resource

CNV   Copy number variation

DNA   Deoxyribonucleic acid

DROP   Detection of RNA Outlier Pipeline

EBV   Epstein-Barr virus

ES   Exome sequencing

ESE   Exonic splicing enhancer

ESS   Exonic splicing silencer

GS   Genome sequencing

ILS   Intron lariat spliceosome

ISE   Intronic splicing enhancer

ISS   Intronic splicing silencer

Kb   Kilobase

LCL	Lymphoblastoid cell line
LPCWG	Laboratory Practice Committee Working Group
MBS	Medicare benefits schedule
MPS	Massively parallel sequencing
mRNA	Messenger ribonucleic acid
OMIM	Online Mendelian Inheritance in Man
PCR	Polymerase chain reaction
Pre-mRNA	Precursor messenger ribonucleic acid
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
RT-PCR	Reverse transcription polymerase chain reaction
snRNA	Small nuclear ribonucleic acid
snRNP	Small nuclear ribonucleoprotein
SNV	Single nucleotide variation
TSS	Transcription start site
UTR	Untranslated region
VECP	Variant expert curation panel
VUS	Variant of uncertain significance

Table of contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Genomics for rare disease

Rare diseases collectively affect 263-446 million individuals worldwide[1] with ~7000 phenotypes that have a known genetic etiology[2]. Provision of a genetic diagnosis for patients with rare disease can be extremely difficult and prolonged over many years, due to the numerosity and heterogeneity of rare diseases, which for many the causal pathogenetic mechanisms are unknown[3,4]. Over 80% of genetic disorders affect children, most are chronic and can lead to early death[4,5]. Moreover, rare disorders can have significant social, psychological, and economic impacts to patients and their families[6,7]. Advances in DNA sequencing technologies have revolutionised the diagnostic rate of rare disease, enabling a precise molecular diagnosis, aid of clinical management and reproductive counselling, reduction in medical costs and invasive procedures, and for many ends a lengthy diagnostic pursuit[8,9]. Despite these advances, the overall diagnostic rate of clinical genomic sequencing leaves > 50% of those affected by rare disease without a genetic diagnosis[10,11].

Individuals suspected of having a rare genetic disorder have several options for diagnostic testing depending on clinical context. Single gene tests may be offered for

disorders with distinctive clinical features typical for a specific disorder. For more heterogeneous indications, gene panel testing may be used to target gene sets associated with disorders that have overlapping phenotypes. When a phenotype is indistinct with multiple nonspecific features, a less restricted approach such as exome sequencing (ES) or genome sequencing (GS) can have more diagnostic utility. These broader approaches enable the identification of variants in genes not yet associated with disease, with the option for future reanalysis to yield diagnoses considering novel gene discoveries and expanded disease phenotypes.

ES reduces sequencing costs and facilitates variant interpretation by only sequencing the protein coding regions of the human genome (~1.5% of genome), where ~85% of all variants currently recognised to cause disease are located[12,13]. Cost effectiveness and higher overall diagnostic yield (25-58% depending on clinical indication) relative to gene panels has led to the widespread adoption of ES in clinical practice for patients with suspected rare genetic disorders[14–26]. However, hybridisation-based enrichment of the exome leads to incomplete and uneven sequencing coverage, limiting its utility for some classes of variation[13,27].

GS covers the whole genome and unlike ES, GS does not involve an enrichment process, the data is quicker to produce, provides more even coverage, and as a result requires lower average coverage to obtain the same accuracy in variant calling[27,28]. GS extends diagnostic yield to ~40-60% in a range of disorders[11,29–37] primarily due to its ability to detect copy number variants (CNVs), structural rearrangements and repeat expansions[38–40]. This comes at greater sequencing and data storage costs and still leaves approximately half of affected individuals without a diagnosis[11]. More than 99% of an individual's ~4.5 million single nucleotide variants (SNVs), small insertions and deletions detected by GS are in the non-coding regions

of the genome where the functional impact of variation is largely unknown[41,42]. A significant proportion of pathogenic non-coding variants disrupt RNA processes such as pre-mRNA splicing and transcription but require additional RNA studies to validate a functional impact to the gene product[43,44]. Furthermore, the abundance of non-coding variants identified by GS and poor understanding of non-coding sequences makes effectively prioritising candidate variants extremely difficult[45].

1.2 Pre-mRNA splicing

At least 95% of human genes are alternatively spliced to generate distinct mRNA transcript isoforms, expanding the functional proteome able to be translated from a set number of protein coding genes[46,47]. The precise excision of introns and ligation of exons from pre-messenger RNA (pre-mRNAs) to produce mature messenger RNAs (mRNAs) is a fundamental step in gene expression. Regulation of splicing is necessary for tissue-specific and developmental isoform expression, and translation of protein isoforms with specialised function[48,49].

Exon-intron boundaries are demarked by conserved RNA motifs that constitute three essential splice sites; 1) the 'GU' donor splice site (5' splice site) at the 5' end of the intron, 2) the 'AG' acceptor splice site (3' splice site) at the 3' end of the intron, and 3) the branchpoint 'A' sequence ~18 – 44 bp upstream of the acceptor[50–52]. Whilst the essential splice sites are stringently conserved, present in >98% of human introns[53], the extended splice site motifs are more degenerate[54] (Figure 1.1A).

The spliceosome is a large multisubunit ribonucleoprotein complex responsible for pre-mRNA splicing, comprising of five core small nuclear ribonucleoprotein (snRNPs) (U1, U2, U4, U5, U6) and known to associate with hundreds of other

1.2 Pre-mRNA splicing

proteins to execute the splicing reaction[46,50,51,55]. Splice site motifs are recognised by base-pairing interactions with the small nuclear ribonuclear RNAs (snRNAs) within the five snRNPs for precise intron excision and spliceosome remodelling during the splicing cycle[50] (Figure 1.1B).

During each round of splicing, the spliceosome assembles and disassembles to form at least ten distinct spliceosome complexes throughout the splicing cycle[46,50,51,55] (Figure 1.1C). This extensive remodelling into catalytically active confirmations enables two transesterification reactions termed "branching" (lariat formation) and "exon ligation" to produce mature mRNA[46,50,51,55]. Splicing occurs co-transcriptionally and is intimately coupled, in time and space, to RNA polymerase II to facilitate transcription initiation, elongation and termination[56,57].

Splicing is regulated by numerous exonic and intronic RNA regulatory motifs recognised by splicing factors that interact with the spliceosome complex to enhance or silence different phases of the splicing cycle[58]. To date over 1,000 RNAs and proteins associated with the human spliceosome complex have been deposited in the spliceosome database[59]. Figure 1.1 illustrates a simplified overview of the essential splice sites and co-transcriptional splicing reaction.

1.2 Pre-mRNA splicing
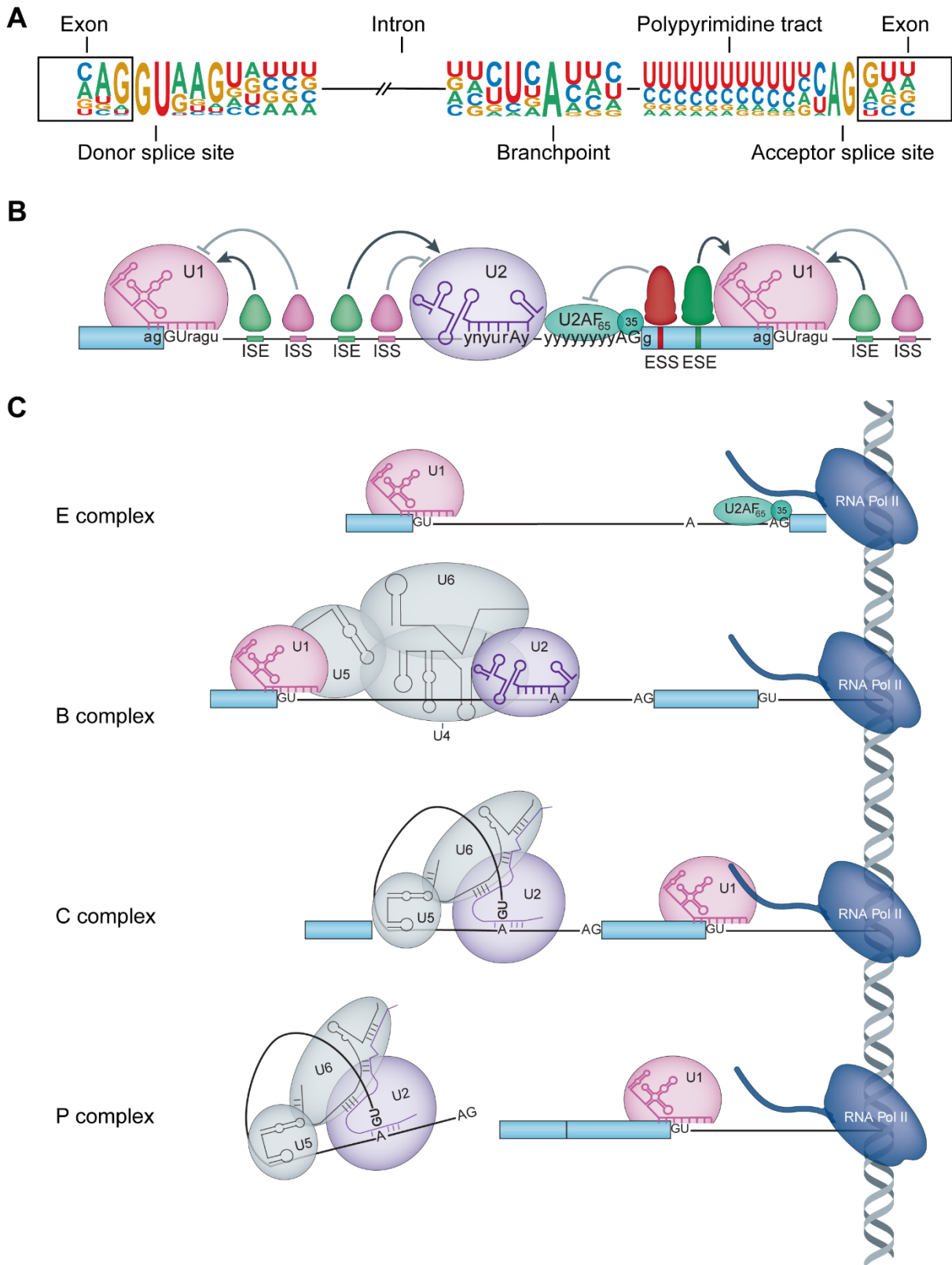
Figure 1.1: Simplified overview of co-transcriptional splicing adapted from Scotti and Swanson, 2016[54]. A) Consensus splice site motifs at the exon intron junctions. **B)** snRNPs U1, U2 and U2AF$_{35}$ base-pair with the consensus splice sites and U2AF$_{65}$ assembles on the polypyrimidine tract. Splice enhancer motifs called intronic and exonic splicing silencers and

enhancers (ISE, ISS, ESE, ESS) are recognised by auxiliary splicing factors to silence or enhance different phases of spliceosome assembly. **C)** The U1 snRNPs binds to the donor and U2AF assembles on the polypyrimidine tract and acceptor to form the E complex. The U2 snRNP binds to the BP and the U4/U6.U5 tri-snRNP is recruited to form the pre-catalytic B complex. U1 and U4 are released to facilitate remodelling into the catalytic C complex to perform the branching reaction, followed by exon ligation to produce the P complex containing the released intron lariat spliceosome (ILS).

1.3 Variant associated mis-splicing in disease

Sequence variants that cause aberrant splicing may represent up to half of all human genetic disease variation[60,61]. Singleton ES alone identifies over 500 splicing VUS within the extended splice site motifs[62]. Splicing variants alter the ability of the core spliceosome machinery and/or auxiliary splicing factors to recognise splice site and RNA regulatory motifs. This can preclude use of canonical splice sites and/or activate a splice site that is preferentially utilised by the spliceosome. Splicing variants are deleterious through removing part of an mRNA transcript or inclusion of ectopic sequence into the mRNA transcript, to encode a truncated protein or frameshift and premature termination codon. Splice-altering variants typically result in exon skipping, intron retention, cryptic splice site activation, and/or ectopic inclusion of a pseudoexon[54].

Splicing disrupting variants can occur almost anywhere within a gene. However, most confirmed pathogenic splicing variants are identified in the essential splice sites and extended splice sites covered by ES, and where clinical laboratories are most cognisant of a potential impact splicing[63]. Beyond the extended splice sites, exonic and deep intronic variants may result in cryptic splice site or pseudoexon activation

1.3 Variant associated mis-splicing in disease

respectively[54,64,65]. Figure 1.2 provides an overview of the different ways in which sequence variants can induce different forms of pre-mRNA mis-splicing.

A) Normal splicing

B) Exon skipping

C) Intron retention

D) Cryptic splice site activation

E) Pseudoexon activation

▼ Variant ablates/weakens consensus splice site
▲ Variant creates/strengthens consensus splice site
▲ Variant creates/strengthens or ▼ ablates/weakens splicing enhancer motif
▲ Variant creates/strengthens or ▲ ablates/weakens splicing silencer motif
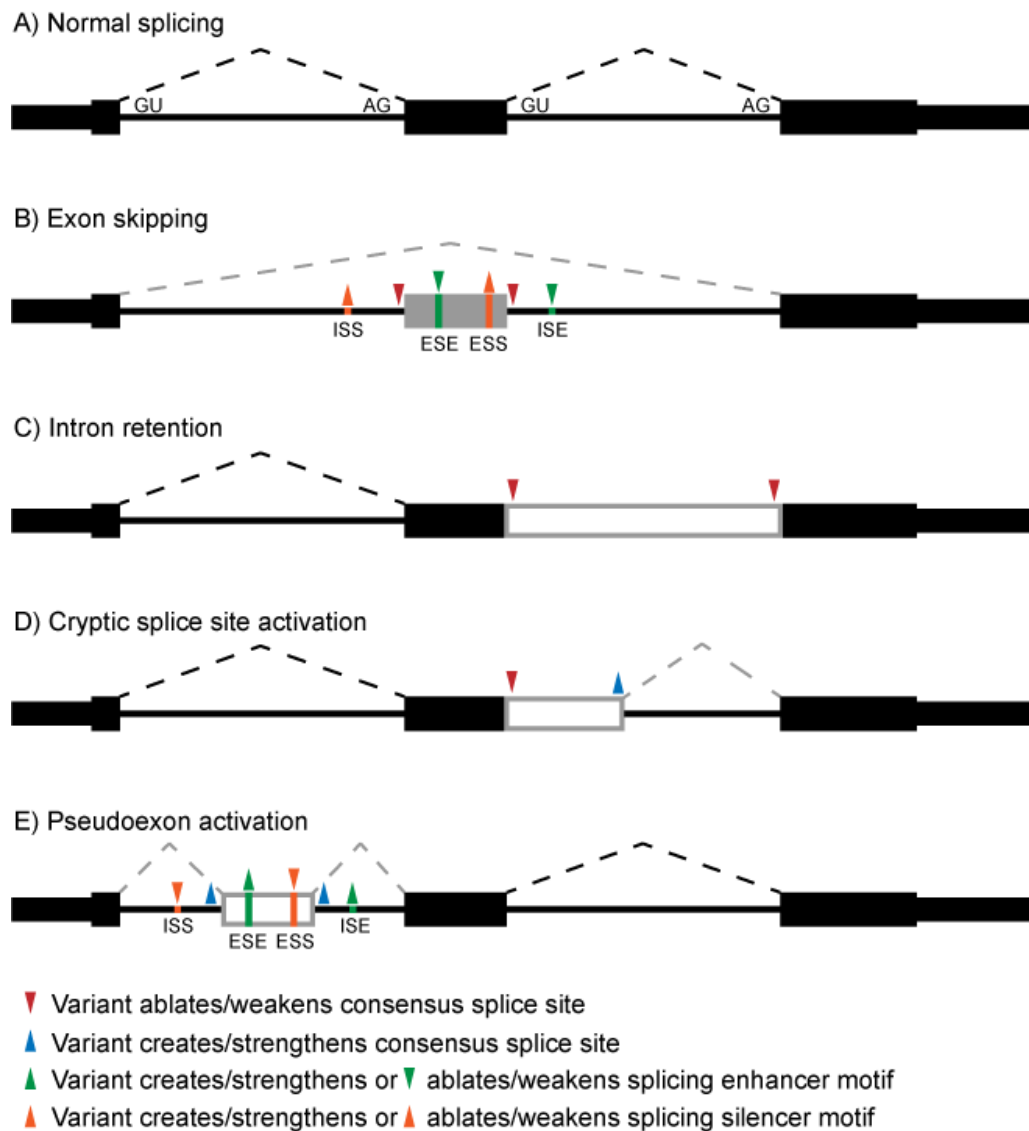
Figure 1.2: Overview of potential variant associated mis-splicing outcomes. Sequence variants (coloured triangles) can impede or preclude use of canonical splice sites. **A)** canonical splicing. **B)** exon skipping due to variant ablated/weakened consensus motif (red) or splice enhancer (green). Alternatively, variant creates/strengthens silencer motif (orange). **C)** intron retention due to ablated/weakened of the consensus splice site (red). **D)** cryptic splice site activation due to ablated/weakened consensus splice site (red) or by variant created/modified donor splice site (blue). **E)** Activation of a pseudoexon due to variant

1.3 Variant associated mis-splicing in disease

created/modified cryptic splice site (blue) or splice enhancer motif (green). Alternatively, variant ablates/weakens silencer motif (orange) to induce activation of a pseudoexon.

Numerous *in silico* tools have been developed to predict whether a sequence variant will disrupt splicing. Whilst essential splice sites variants are almost guaranteed to disrupt splicing, predicting the impact of variants in the extended splice sites and deep intronic regions is far more difficult. Original tools used statistical modelling to predict mis-splicing due to variants affecting the various elements of splicing[66]. Most tools focussed on either donor and acceptor motifs[67–69], the branchpoint and polypyrimidine tract[70,71], or splice enhancer motifs[72,73]. Indeed, early tools capable of analysing multiple elements of splicing[74,75] did so in isolation rather than providing a holistic analysis.

In silico prediction tools now provide important computational evidence for diagnostic interpretation of putative splice variants[76,77]. Pathogenicity of a splicing variant depends on the variant associated splicing outcomes and consequences for the encoded protein function[76,77]. Hence, accurate interpretation is predicated on not only *if* a variant will disrupt splicing, but *how?*

Development of machine learning tools have significantly increased sensitivity and specificity of splice predictions in recent years[66,78]. Machine learning models can provide predictions for variants anywhere within annotated transcripts and consider greater genomic context in their analysis than previous tools[66,78]. However, predicting specific splicing outcomes such as cryptic splice site activation and exon skipping remains erroneous[66,78–80]. Furthermore, there is no consensus on how best to use them as supporting evidence for variant interpretation[66,78]. Difficulty interpreting output, setting optimal threshold values, and conflicting predictions

between tools is a major impediment to their application in clinical practice[66,78,80].
Whilst these tools have been validated on previously analysed sets of variants, their
reported utility in a real-time clinical setting has been underwhelming[80].

1.4 RNA analysis of putative splicing variants

Variant associated mis-splicing must be experimentally validated by RNA analysis
for accurate diagnostic interpretation. Determining for instance, loss-of-function or
gain-of-function effect, can confirm whether mis-splicing outcomes are consistent
with the pathogenetic mechanism of disease and patient phenotype[81–85]. Without
experimental validation of the precise splicing outcomes and effect for the encoded
protein, putative splicing variants often remain classed as VUS which cannot be
used for clinical decision making.

Determining the most appropriate source of RNA for clinical variant interpretation is a
key consideration for RNA analysis. A major limitation of splicing studies stems from
the tissue specific nature of alternative splicing. The affected isoform is often
expressed within tissues with limited accessibility, such as the brain or heart[86].

Genetic constructs called minigenes have been widely used for splicing assays[87–91]
and are not restricted by tissue specific expression or availability of patient-derived
samples. Typically, the affected splice junction and flanking regions are inserted into
an expression plasmid which is transfected into a human cell line for RNA analysis[92].
The splicing consequences of a sequence variant is then assessed, relative to a
reference construct[92].

1.4 RNA analysis of putative splicing variants

A major caveat of minigene assays is the limited genomic sequence that can fit into a minigene construct. Size constraint of the plasmid insert may limit the genomic context to only 3 exons and 2 introns, or to only the flanking intronic regions bearing the consensus motifs[92]. The broader genomic context such as intronic RNA regulatory motifs or genetic modifiers may be relevant to the splicing reaction and penetrance[93,94]. Consequently, splicing in the reference or variant construct may not accurately reflect splicing in the manifesting tissue, and can splice differently depending on which cell line is used to express the minigene[92,95].

Fortunately, readily available biospecimens such as blood and skin fibroblasts can be used to infer splicing outcomes in the predominant isoform expressed in the manifesting tissue[86,96–104]. Blood-derived (whole blood, peripheral blood mononuclear cells, Epstein-Barr virus transformed lymphocytes) and skin fibroblasts are routinely collected in clinical practice and express the majority of OMIM genes at sufficient levels for RNA analysis[86,101,105,106]. If the pattern of splicing is consistent with the manifesting tissue, patient derived samples that more accurately reflect biological conditions are preferred over minigenes for variant interpretation[93]. To overcome strict tissue specific expression of some genes, induced pluripotent stems cells[107] and trans-differentiation of patient cells[98] can been used for RNA analysis, though at the cost of further experiments and time.

Once the source of RNA has been determined, the two main methods currently used for splicing analysis are reverse transcription polymerase chain reaction (RT-PCR) followed by Sanger sequencing or short read RNA sequencing (RNA-seq) by massively parallel sequencing (MPS). RT-PCR and RNA-seq are complimentary approaches that offer different advantages depending on the genomic context and gene expression levels in available tissues (summarised in Figure 1.3).

1.4 RNA analysis of putative splicing variants

Two strategic considerations when choosing the most appropriate method are gene expression levels in the available tissue and whether a candidate variant has been identified. RT-PCR has a lower limit of detection than RNA-seq and primers can be designed to target specific splicing events. The reads are longer, spanning multiple splice junctions to provide a better picture of isoforms by identifying multiple splice junctions within a single amplicon. RT-PCR has been used extensively for splicing analysis[101,103] and the limitations and biases introduced through amplification are well known. Diagnostic sensitivity is heavily influenced by the laboratories' knowledge of splicing as primer design and thermocycler conditions employed will limit detection to anticipated aberrant splicing outcomes[108]. Due to the targeted nature of RT-PCR, multiple experiments and cloning of amplicons may be required to resolve all splicing events. Hence, RT-PCR is best suited to candidate variants or candidate genes with short coding sequences that can be easily amplified.

1.4 RNA analysis of putative splicing variants

RNA source

Minigene

Advantages
• Unrestricted by tissue specific expression or tissue availability

Disadvantages
• Artificial system with limited genomic context may not recapitulate splicing *in situ*
• Splicing outcomes can differ between cell lines used for transfection

Clinically accessible tissue

Advantages
• Blood-derived and skin fibroblasts routing collect in clinical practice
• Complete genomic context
• Can be used for cell transdifferentiation or to generate iPSCs

Disadvantages
• Variant affect on splicing and natural alternative splicing may be different in the manifesting tissue

Biopsy of manifesting tissue

Advantages
• Can assess splicing in the relevant manifesting tissue

Disadvantages
• Limited accessibility to biopsies
• Limited access to age and sex matched controls

RNA analysis

RT-PCR and Sanger sequencing

Advantages
• Lower limit of detection enables analysis of genes with very low expression
• Read length provides greater isoform information and allows phasing of transcripts

Disadvantages
• Only suitable for candidate variants or candidate genes with short coding sequences
• Sensitivity limited by primer design and amplification conditions
• May require multiple experiements and amplicon cloning to resolve all isoforms
• Artefacts due to multi-template PCRs

Short read RNA sequencing (massively parallel sequencing)

Advantages
• Agnostic approach
• Analysis of whole transcriptome
• Allows for allele specific expression analysis
• Does not require a candidate variant

Disadvantages
• Greater lower limit of detection
• Reads more difficult to map unambiguously
• Short reads provide limited isoform information
• Reads too short to phase transcripts

Long read RNA sequencing (Nanopore/PacBio)

Advantages
• Targeted or agnostic approach
• Analysis of whole transcriptome
• Allows for allele specific expression analysis
• Does not require a candidate variant
• Long reads produce more complete isoform information and allow for phasing of transcripts

Disadvantages
• High sequencing error rate
• Technology still emerging
• Immature processing pipelines

Figure 1.3: Relative benefits of the main technical approaches to RNA analysis of putative splicing variants.

1.4 RNA analysis of putative splicing variants

RNA-seq by MPS is an agnostic approach allowing for whole transcriptome analysis, investigation of multiple variants and allele specific expression. Several research-led cohort studies have demonstrated the utility of RNA-seq to increase diagnostic yield over DNA sequencing, even without a candidate variant[98,99,101,102,104,109,110]. In absence of candidate variants, there is an abundance of computational tools to detect expression outliers and aberrant splicing events but significant discordance between them[111–117]. Recently published approaches, OURIDER[118] and FRASER[119], employ machine learning for more accurate gene expression analysis and splicing outlier detection, respectively, than statistical models alone[117].

Numerous RNA-seq workflows exist depending on biological sample and type of analysis required. For example, poly(A) enrichment of mRNA is not recommended for RNA-seq of degraded RNA from formalin-fixed paraffin-embedded tissues as this produces strong 3' bias in transcript coverage[112]. For RNA-seq of whole blood, depletion of highly abundant human beta globin transcripts is recommended to increase read counts for non-globin genes[120].

Although many computational approaches to RNA-seq analysis are available, best practices have not been established for alignment of sequencing reads to the reference genome, filtering and normalisation, and sequencing depth required for splicing analysis[121–123]. Reads generated by RNA-seq typically do not span multiple splice junctions, limiting isoform information and confidence mapping reads that span exon-exon junctions (split reads) to the reference genome[112,124]. This can lead to alignment errors or potentially filtering of diagnostically important sequencing reads[124,125].

1.5 RNA assay data warrants bespoke interpretation guidelines

Now third generation long read sequencing technologies have the potential to overcome caveats of both Sanger sequencing and MPS. Long read sequencing approaches from Oxford Nanopore and Pacific Biosciences can produce average read lengths >10 kb for complete transcript isoform information and phasing[126–128]. Long read sequencing can be applied in a targeted manner to RT-PCR amplicons or to the whole transcriptome[127,128]. However, high cost/throughput ratio, sequencing error rate and immature analysis pipelines relative to MPS currently limit the application of long read sequencing for splicing analysis and RNA diagnostics[126,127].

Due to the complexity of splicing analysis and emergent stage of RNA diagnostics, standardised technical platforms and determination of best practices are needed to integrate RNA analysis into variant interpretation guidelines.

1.5 RNA assay data warrants bespoke interpretation guidelines

In 2000 the ACMG Laboratory Practice Committee Working Group (LPCW) published the first recommended standards for the interpretation of sequence variants as an education resource to aid medical geneticists in clinical reporting[129]. The ACMG LPCW proposed guidelines for report content including interpretation, methodology, limitations, follow-up studies such as segregation, and limiting interpretation and reporting to qualified scientists such as those certified by the American Board of Medical Genetics, and laboratories with Clinical Laboratory Improvement Amendments certification.

It was recommended that sequence variants fall under five categories:

1.5 RNA assay data warrants bespoke interpretation guidelines

1. Sequence variation is previously reported and is a recognized cause of the disorder.

2. Sequence variation is previously unreported and is of the type which is expected to cause the disorder.

3. Sequence variation is previously unreported and is of the type which may or may not be causative of the disorder.

4. Sequence variation is previously unreported and is probably not causative of disease

5. Sequence variation is previously reported and is a recognized neutral variant.

The guidelines referred to splice junction variants predicted to alter reading frame, delete one or more exons, or likely to produce a cryptic splice site, but offered no recommendation of splicing prediction tools or methods of analysis. The authors noted functional studies had not yet been utilised by diagnostics laboratories as of May 2000.

ACMG standards and guidelines have since published numerous iterative developments for specific disorders and testing methodologies[130–155], and major revisions in 2007[156] and 2015[76]. The 2007 revision prescribes the use of Human Genome Variation Society nomenclature[157], National Center for Biotechnology Information reference sequences[158], and a growing number of variant curation databases[159–161]. *In silico* splicing prediction tools[69,72] and RNA analysis were listed as follow-up studies that could be employed to assist in variant classification, and a decision tree was formulated to systematise interpretation. However, the five broad categories of sequence variations remained largely unchanged.

1.5 RNA assay data warrants bespoke interpretation guidelines

In 2013 the ACMG, Association for Molecular Pathology (AMP) and College of American Pathologists formed an expert workgroup to provide significantly more comprehensive guidelines in the 2015 revision for interpretation of sequence variants, warranted by the increased number and complexity of genetic tests, and number of identified variants enabled by advances in sequencing technologies[76]. Over 100 laboratories were involved in revision of the guidelines through surveys and workshops, focusing on interpretation of Mendelian disease variants. Further recommendations included use of the term "variant" instead of "mutation" and "polymorphism", and that pathogenic variants be reported with respect to condition and inheritance pattern. Further utilisation of sequence and disease specific databases[162–167] and now population databases[168–170] for assessing variant frequencies in the general population were recommended when classifying a variant.

The five sequence variation categories were replaced with a standardised terminology applicable to variants in Mendelian genes, with the term "likely" used when there is greater than 90% certainty a variant is pathogenic or benign:

1. Pathogenic
2. Likely pathogenic
3. Uncertain significance
4. Likely benign
5. Benign

Separate weighted criteria were provided for likely/pathogenic and likely/benign variants. Conflicting or insufficient evidence to meet either criteria results in a classification of uncertain significance.

1.5 RNA assay data warrants bespoke interpretation guidelines

Constructing an evidence framework adaptable to any Mendelian variant, whilst allowing for flexibility in evidence weighting, has led to a subjective classification process open to interpretation. Indeed, low concordance in application of ACMG-AMP criteria was reported between diagnostic laboratories[171] and laboratories have subsequently published refined criteria to address issues such as double counting of evidence and lack of specificity[172–174], and The National Institutes of Health Clinical Genome Resource (ClinGen)[175] formed variant curation expert panels (VECP) to assess and refine ACMG-AMP guidelines with respect to specific genes and disorders[176], and specific evidence criteria[77,93,177–179].

Whilst the 2015 revision of the interpretation guidelines established a solid framework for Mendelian variants, it did not cater well for splicing variants which could fall under several evidence criteria with different evidence weighting. Further confounding classification, the "Very Strong PVS1 null variant" criterion could be used for essential splice site variants where loss of function was the *predicted* outcome in the absence of RNA analysis. However, the "PS3 well-established functional studies" criterion could be used for splicing variants where loss of function was *validated* by RNA analysis and weighted as supportive, moderate, or strong evidence depending on the interpretation of "well-established" functional studies. In 2019, updated recommendations for the PS3 criterion cautioned the use of minigene assays with artificial promoters and overexpression that may not accurately reflect biological conditions and suggested further recommendations would be needed to appropriately assign PS3 or PVS1 to splicing assays[93].

Revised recommendations for the PVS1 criterion proposed moderate, strong and very strong weightings depending on effect on reading frame, transcripts targeted by nonsense mediated decay and biological relevance of the effected transcript[77].

However, this was still predicated on the ability to *predict* mis-splicing outcomes associated with essential splice site variants.

The current AMCG-AMP framework is inadequate for the interpretation of putative splicing variants and complex RNA assay data, which can result in multiple in-frame and out-of-frame events, partial mis-splicing, or disruption of transcription. RNA assay data is far more complex than DNA and requires a bespoke interpretation framework for use in clinical variant curation. ClinGen VCEPs have implemented specifications to the ACMG-AMP variant interpretation guidelines tailored to genes in their respective panels for RNA assay data, as have the Cancer Variant interpretation Group UK for hereditary cancer syndromes[180]. However, quality standards for RNA diagnostics and ACMG-AMP aligned interpretation guidelines for *rare disease* are needed urgently. Optimising RNA analysis for implementation into clinical laboratories will increase the sensitivity of genomic testing, enabling definitive genetic diagnosis and improve clinical management for affected individuals.

1.6 Project Aims

Our aim was to optimise the RNA analysis pipeline from *in silico* analysis, to determining optimal source of RNA, technical approaches to RNA analysis and informative diagnostic reporting. This thesis embodies everything our expert team has learnt during our research-led RNA analysis for a cohort of ~150 variants, to enable translation of RNA diagnostics into routine clinical testing.

Specific project aims:

1.6 Project Aims

1. To determine the technical utility of different clinically accessible specimens for RNA diagnostics.

    1.1. Clinically accessible tissues: Whole blood, PBMCs, EBV-transformed lymphocytes, primary fibroblasts, urothelial.

    1.2. Determine diagnostic utility of cycloheximide to rescue transcripts from surveillance by nonsense-mediated decay.

    1.3. Correlate RNA-seq read depth from clinically accessible tissues to determine the median transcripts per million required for mRNA analysis by RT-PCR.

2. Develop ACMG-AMP aligned variant interpretation guidelines for functional mRNA testing of splicing variants.

    2.1. Devise and clinically validate standard operational guidelines for RT-PCR and Sanger sequencing functional studies of pre-mRNA splicing.

    2.2. Comparatively evaluate RT-PCR and short read RNA-seq for diagnostic mRNA testing.

3. To use experimentally determined outcomes from splicing diagnostics to inform iterative development of a novel algorithm to predict splicing abnormalities.

    3.1. Curate splicing outcomes from RNA diagnostics cohort to generate training and test dataset for a machine learning splice prediction algorithm.

Chapter 2 illustrates the diagnostic utility of rapid RNA analysis from clinically accessible tissues in a critically ill neonate and deceased sibling. RT-PCR analysis procured a definitive molecular diagnosis to inform patient management and health economic analysis showed early diagnosis reduced hospitalisation costs.

1.6 Project Aims

Chapter 3 describes standardised practices for PCR-based RNA diagnostics and ACMG-AMP aligned interpretation guidelines for RNA assay data. This study highlights the utility of clinically accessible tissues at scale and the clinical impact on patient management and reproductive counselling for 74 families.

Chapter 4 investigates the capability of unannotated splicing events observed in a set of over 300,000 reference RNA-seq datasets to predict the nature of variant associated mis-splicing at a given splice junction. This splice junction dataset will facilitate variant curation of experimental design of RNA assays.

Chapter 5 outlines all the contributions that I made to variant reclassification or diagnoses throughout my candidature. I performed RNA analysis for 107 variants that was utilised for genetic diagnosis and/or variant interpretation.

Chapter 2

Rapid RNA analysis for a critically ill neonate

2.1 Overview

A novel homozygous essential splice site variant (NM_133436.3:c.1476+1G>A) was identified in a newborn with cerebellar hypoplasia who underwent rapid exome sequencing as part of a rapid genomic diagnosis program[177]. Due to inconclusive results from biochemical analysis of asparagine synthetase levels, this variant remained classified as a variant of uncertain significance.

Subsequently this patient was triaged into our RNA diagnostics program to validate the impact on splicing of *ASNS* transcripts using patient and parental blood samples. I devised the RT-PCR strategy and performed the experiments and analysis to produce an RNA diagnostics report within 10 days of sample receipt. The splicing studies enabled reclassification of the variant from VUS to pathogenic and the patient was subsequently transition from intensive care to palliative care.

A health economic analysis was performed for the patient and deceased sibling who had a similar presentation though remained undiagnosed. Early diagnosis and transition to palliation reduced stay in intensive care and costs by AUD $100,828. DNA extracted from the sibling's archived fibroblast cell line identified presence of

2.1 Overview

the homozygous essential splice site variant and I repeated the RT-PCR analysis in the fibroblasts to confirm an identical pattern of mis-splicing.

This investigation demonstrates the utility and cost effectiveness of rapid exome sequencing and RNA analysis in clinical genomics. Using clinically accessible tissues, blood and fibroblasts, to infer mis-splicing in the brain informed variant classification when the manifesting tissue was unavailable for RNA testing. This chapter was published as a brief report for which I was joint first author:

Akesson LS., **Bournazos A.**, Fennell A, et al. Rapid exome sequencing and adjunct RNA studies confirm the pathogenicity of a novel homozygous *ASNS* splicing variant in a critically ill neonate. *Hum Mutat*. 2020;41(11):1884-1891.

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Rapid exome sequencing and adjunct RNA studies confirm the pathogenicity of a novel homozygous *ASNS* splicing variant in a critically ill neonate

Lauren S. Akesson[1,2,3] | Adam Bournazos[4,5] | Andrew Fennell[3,6] | Emma I. Krzesinski[3,6] | Kenneth Tan[6,7] | Amanda Springer[3,6] | Katherine Rose[3,6] | Ilias Goranitis[8,9] | David Francis[1] | Crystle Lee[1] | Fathimath Faiz[10] | Mark R. Davis[10] | John Christodoulou[1,2,9,11] | Sebastian Lunke[1,9,12] | Zornitza Stark[1,2,9] | Matthew F. Hunter[3,6] | Sandra T. Cooper[4,5,13]

[1]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Parkville, Victoria, Australia

[2]Department of Paediatrics, University of Melbourne, Melbourne, Australia

[3]Genetics Clinic, Monash Health, Monash Medical Centre, Clayton, Victoria, Australia

[4]Kids Neuroscience Centre, Children's Hospital at Westmead, Sydney, Australia

[5]Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[6]Department of Paediatrics, Monash University, Melbourne, Australia

[7]Monash Newborn, Monash Health, Monash Children's Hospital, Clayton, Victoria, Australia

[8]Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia

[9]Australian Genomics Health Alliance, Parkville, Victoria, Australia

[10]Department of Diagnostic Genomics, PathWest Laboratory Medicine, Perth, Australia

[11]Brain and Mitochondrial Research Group, Murdoch Children's Research Institute, Melbourne, Australia

[12]Department of Clinical Pathology, University of Melbourne, Melbourne, Australia

[13]Children's Medical Research Institute, Sydney, Australia

Correspondence
Zornitza Stark, Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Flemington Rd, Parkville, VIC 3052, Australia.
Email: Zornitza.Stark@vcgs.org.au

Abstract

Rapid genomic diagnosis programs are transforming rare disease diagnosis in acute pediatrics. A ventilated newborn with cerebellar hypoplasia underwent rapid exome sequencing (75 h), identifying a novel homozygous *ASNS* splice-site variant (NM_133436.3:c.1476+1G>A) of uncertain significance. Rapid *ASNS* splicing studies using blood-derived messenger RNA from the family trio confirmed a consistent pattern of abnormal splicing induced by the variant (cryptic 5′ splice-site or exon 12 skipping) with absence of normal *ASNS* splicing in the proband. Splicing studies reported within 10 days led to reclassification of c.1476+1G>A as pathogenic at age 27 days. Intensive care was redirected toward palliation. Cost analyses for the neonate and his undiagnosed, similarly affected deceased sibling, demonstrate that early diagnosis reduced hospitalization costs by AU$100,828. We highlight the diagnostic benefits of adjunct RNA testing to confirm the pathogenicity of splicing

variants identified via rapid genomic testing pipelines for precision and preventative medicine.

KEYWORDS

ASNS, asparagine synthetase deficiency, exome sequencing, mRNA splicing analysis, rapid genomic diagnosis program

Rapid genomic diagnosis programs in neonatal and pediatric intensive care are transforming clinical practice by diagnosing monogenic disorders in less than 20 h (Clark et al., 2019). However, assessment of variants of uncertain significance identified in phenotypically concordant genes remains challenging, as supportive functional validation studies such as messenger RNA (mRNA) and functional protein analyses are not typically available within clinically meaningful timeframes. We present a seriously ill newborn who was diagnosed with asparagine synthetase deficiency as part of a rapid genomic diagnosis program, where the pathogenicity of a splicing variant of uncertain significance was confirmed via an integrated pipeline of rapid mRNA analyses. This case demonstrates the clinical and cost benefits of rapid mRNA analyses as part of a rapid genomic diagnosis program in pediatric acute care. A timeline for the diagnostic investigations undertaken in this case, including research-based mRNA splicing studies, is provided in Figure 1c. Additional information on the methodologies is given in the Supplementary Information, available online.

The proband (II:6; Figure 1a,b) was a male neonate born at 36 weeks' gestation to non-consanguineous parents (I:1 and I:2) of South Sudanese ethnicity, following a pregnancy complicated by antenatal detection of microcephaly and suspected pontocerebellar hypoplasia. The family had previously received genetic counseling in 2011 following the death of a male sibling (II:2) at age 8 weeks. The sibling had a similar antenatal course to the proband and was ventilator-dependent from birth. He had significant microcephaly with a head circumference Z-score of −4.2 with otherwise normal growth parameters. In the absence of clinical improvement, the family elected to redirect the sibling's care toward palliation and symptomatic management, and the infant died at age 8 weeks. Postmortem examination was declined. A suspected diagnosis of pontocerebellar hypoplasia type 4 was based on brain magnetic resonance imaging (MRI) findings (Figure 1a). Chromosomal microarray was normal and no further genetic testing was offered. The family received genetic counseling for a presumed autosomal recessive monogenic disorder. A further four siblings are healthy.

The parents declined invasive prenatal genetic testing following identification of microcephaly and suspected pontocerebellar hypoplasia in the present case. Following an emergency Cesarean section for suspected fetal distress, the proband was intubated in the delivery room for poor respiratory effort and transferred to the neonatal intensive care unit. He remained ventilator-dependent with minimal respiratory effort. The proband had microcephaly with a head circumference Z score of −2.5 with otherwise normal growth

parameters. He had abnormal movements with fisting, back-arching, and posturing. There were no electrographic seizures. Neuroimaging with cranial ultrasound and later MRI demonstrated microcephaly with cerebellar hypoplasia (Figure 1b).

The proband was referred for clinical genetics assessment at age 9 h and approved for inclusion in a rapid genomic diagnosis program, the Australian Genomics Acute Care study, at age 12 h. Following genetic counseling, the parents provided written consent for trio exome sequencing. Chromosomal microarray was requested and performed in tandem with rapid exome sequencing (see Figure 1c).

Multiple homozygous variants were observed during variant prioritization, suggesting identity by descent despite absence of known parental consanguinity. Trio exome sequencing identified a homozygous splicing variant in *ASNS* (Chr7(GRCh37):g.97482371C>T; NM_133436.3(ASNS):c.1476+1G>A). The essential splice-site variant was very highly conserved (PhyloP UCSC), absent from population databases (gnomAD, dbSNP, 1000G), and not previously reported in the ClinVar or HGMD databases or the medical literature (all databases accessed January 2019). Segregation analysis confirmed biparental inheritance. Biallelic mutations in *ASNS* cause asparagine synthetase deficiency (MIM# 615574; Alfadhel & El-Hattab, 2018), a diagnosis consistent with the clinical features of the proband and his deceased sibling. The variant was classified as a variant of uncertain significance. Turnaround time from receipt of clinical specimens to issue of report was 75.5 h.

The trio exome sequencing results were disclosed to the parents, and written consent was obtained for supportive analyses to determine the pathogenicity of the variant, including biochemical measurement of asparagine in blood and cerebrospinal fluid (CSF), segregation of the variant in the deceased sibling from DNA extracted from a cryopreserved fibroblast cell line, and mRNA splicing analyses of blood obtained from the proband and both parents, and later from a cryopreserved fibroblast cell line from the deceased sibling.

Blood and CSF asparagine levels were 27 μmol/L (reference range 29–202) and 4 μmol/L (reference range 0–20), respectively. These values were considered potentially consistent with, but not diagnostic of, asparagine synthetase deficiency.

Chromosomal microarray results for the proband became available on Day 9, demonstrating no clinically significant genomic imbalance. Two regions of homozygosity were detected on chromosomes 7 and 8. *ASNS*, located at chromosomal location 7q21.3, was within the chromosome 7 region of homozygosity. Retrospective review of the deceased sibling's microarray performed in
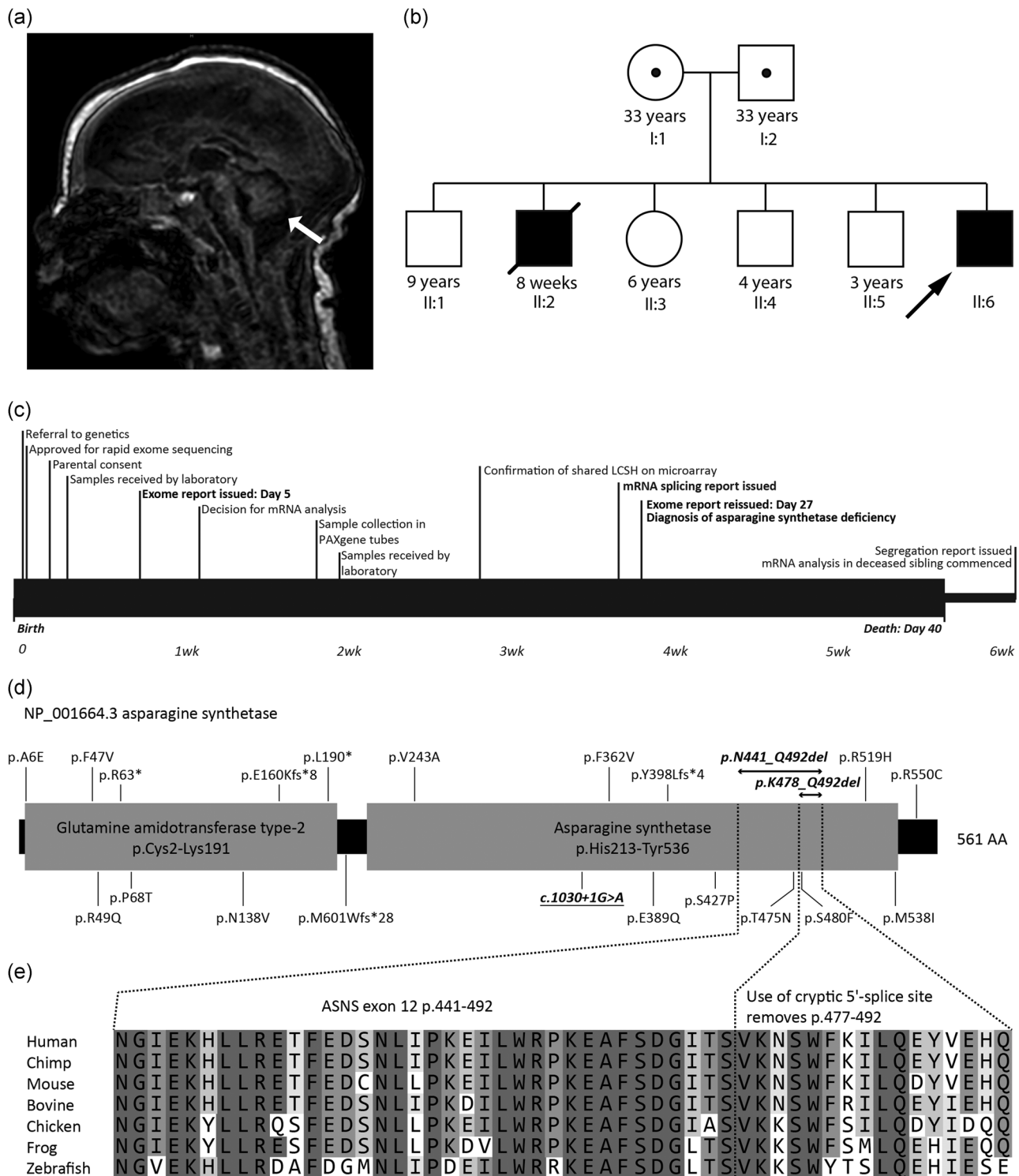
FIGURE 1 Family pedigree, brain imaging, clinical timeline, and schematic overview of pathogenic *ASNS* variants. (a) Magnetic resonance imaging (MRI) of postnatal brain for the proband showing cerebellar hypoplasia. (b) Pedigree showing the proband (arrow) and his deceased affected sibling. (c) Timeline of clinical and genetic investigations, noting sequential reclassification of the homozygous *ASNS* c.1476+1G>A variant from a variant of uncertain significance to a pathogenic variant, and subsequent changes in clinical management. Wks, weeks of age. (d) A schematic of the encoded asparagine synthetase protein showing the position of residues deleted through abnormal splicing induced by the c.1476+1G>A variant, as well as positions of previously reported *ASNS* variants classified as likely pathogenic or pathogenic taken from ClinVar. Residues annotated as comprising the functional domains are taken from Uniprot (P08243). (e) Evolutionary alignment of the amino acid residues encoded by *ASNS* exon 12 (NM_001673.4). Skipping of exon 12, or use of the upstream 5′ cryptic splice site in exon 12, abnormally removes numerous highly conserved amino acids, many of which are invariant throughout vertebrate evolution

2011 demonstrated a similar region of homozygosity on chromosome 7 that also included the *ASNS* gene (Supplementary Information, Figure S1).

Sanger sequencing of DNA extracted from a cryopreserved fibroblast cell line from a skin biopsy obtained from the deceased sibling identified the *ASNS* splicing variant in homozygous form.

Data mining of RNA-seq data obtained from the GTEx Project (Londale et al., 2013) showed that *ASNS* exon 12 is a canonical exon included in predominant *ASNS* isoforms expressed in the brain, blood, and skin (Figure S2). Therefore, splicing outcomes observed in blood and fibroblast RNA maintain relevance to the predominant *ASNS* isoform(s) in the brain.

Reverse transcription polymerase chain reaction (RT-PCR) splicing studies of *ASNS* complementary DNA (cDNA) synthesized from mRNA isolated from whole blood (proband and both parents) and fibroblasts (deceased sibling) confirmed abnormal splicing induced by the homozygous *ASNS* c.1476+1G>A variant (Figure 2). Abnormal splicing events included (a) exon 12 skipping which removes 156 nucleotides from the *ASNS* mRNA. This event is in-frame, removing 52 amino acids from the encoded asparagine synthetase enzymatic domain (p.(Asn441_Gln492del)); (b) use of a cryptic 5′ splice-site 48 nucleotides upstream of the native 5′ splice-site which removes 48 nucleotides from exon 12. This event is in-frame, removing 16 amino acids from the asparagine synthetase enzymatic domain (p.(Val477_Gln492del));(c) retention of intron 12, or both introns 11 and 12, which encode a premature termination codon and may be targeted by nonsense-mediated decay. Transcripts that may escape nonsense-mediated decay encode asparagine synthetase proteins that lack a conserved region within the asparagine synthetase enzymatic domain and are likely to be dysfunctional or nonfunctional (Figure 1d,e). A primer bridging exons 12 and 13, specific for normally spliced *ASNS* transcripts, failed to amplify a product from cDNA for both affected siblings, though robustly amplified correctly spliced *ASNS* cDNA from the parent carriers and unrelated disease controls (Figure 2iii). These data infer there are no, or undetectably low levels of, correctly spliced *ASNS* transcripts in both affected individuals. Amplification of glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) at sub-saturating cycle lengths of 25 and 30 cycles confirmed similar cDNA loading for all samples.

The results of biochemical evaluation, chromosomal microarray, and mRNA splicing analyses of the proband and both parents (reported within 10 days) were available on Day 27 of life. Based on the additional evidence, the homozygous *ASNS* splicing variant was reclassified as pathogenic according to ACMG guidelines (Richards et al., 2015), establishing a diagnosis of asparagine synthetase deficiency (MIM# 615574; Alfadhel & El-Hattab, 2018).

Genetic counseling was provided to the family and after a further 13 days with their child, the parents elected to redirect care toward palliation and symptomatic management, and the proband died at age 40 days (Figure 1c).

Following receipt of segregation results in the deceased sibling, which became available subsequent to the proband's death, confirmatory mRNA splicing analyses were performed on cryopreserved fibroblasts from the deceased sibling (see Figure 2).

Total hospitalization length of stay for the proband and the deceased sibling was 40 and 59 days, respectively, with a difference of 19 days. Based on average daily hospitalization costs in a pediatric intensive care unit of AU$6200 (Schlapbach et al., 2017), less the cost of ultra-rapid trio exome sequencing ($12,000) and research mRNA splicing analysis ($2486; reagents $433, Sanger sequencing $1363, time $690 (in silico analyses, laboratory work, reporting)), this difference in length of stay equates to an estimated cost-saving following early diagnosis of AU$100,828. Potential additional cost savings related to reproductive testing to avoid another recurrence have not been estimated.

Rapid genomic diagnosis programs have the potential to target therapy, clarify prognosis, avoid unnecessary or invasive investigations and interventions, and reduce health costs (Farnaes et al., 2018; French et al., 2019; Meng et al., 2017; Mestek-Boukhibar et al., 2018; Petrikin et al., 2018; Stark et al., 2018; van Diemen et al., 2017; Willig et al., 2015). Variants of uncertain significance are frequently identified by genomic testing (Grody, Thompson, & Hudgins, 2013). In the acute care setting, time taken to achieve diagnostic certainty may have a significant impact on clinical care. Despite recognition of the challenges associated with variants of uncertain significance (Petrikin, Willig, Smith, & Kingsmore, 2015), supportive functional genomics analyses have not been described to date as part of a rapid genomic diagnosis program (Farnaes et al., 2018; French et al., 2019; Meng et al., 2017; Mestek-Boukhibar et al., 2018; Petrikin et al., 2018; Stark et al., 2018; van Diemen et al., 2017; Willig et al., 2015). With a molecular diagnosis of a rare monogenic disorder achieved following mRNA analyses within the first 4 weeks of life, this case demonstrates the clinical and cost benefits of incorporating supportive testing as an adjunct to rapid genomic diagnosis programs. Results facilitated clinical decision making within a clinically appropriate timeframe, resulting in redirection of care toward palliation and symptomatic management, with projected cost savings by reducing hospitalization length of stay and diagnostic investigations following early diagnosis. Importantly, a confirmed diagnosis of asparagine synthetase deficiency now enables reproductive genetic testing.

Asparagine synthetase deficiency is a rare monogenic condition characterized by microcephaly with progressive encephalopathy, severely delayed neurodevelopment, and early death (Alfadhel & El-Hattab, 2018). The spectrum of clinical severity varies between reported cases (Abhyankar et al., 2018; Alfadhel et al., 2015; Alrifai & Alfadhel, 2016; Ben-Salem et al., 2015; Galada et al., 2018; Gupta et al., 2017; Palmer et al., 2015; Sacharow et al., 2018; Seidahmed et al., 2016; Sun et al., 2017; Yamamoto et al., 2017), with no reported genotype–phenotype correlations (Alfadhel & El-Hattab, 2018). It has been suggested that asparagine synthetase deficiency should be considered in any neonate with microcephaly and epileptic encephalopathy, which is the most common clinical presentation for this condition (Radha Rama Devi & Naushad, 2019). The two siblings in this report presented with a severe phenotype, with cerebellar hypoplasia and ventilator dependence from birth. mRNA analyses suggested absence, or extremely low levels, of normal *ASNS* splicing,

(a) (i) Exon10-F + Exon13-R
(ii) Exon12-F + 3'UTR-R
(iii) Exon12/13-F + 3'UTR-R
(iv) Exon10-F + Intron12-R
(v) GAPDH — 25 cycles, 30 cycles

(b) (i) Amplicon 2 - Cryptic 5' splice site
(ii) Amplicon 3 - Exon 12 skipping
(iii) Amplicon 4 - Intron 11 retention
(iv) Amplicon 5 - Intron 12 retention

(c) c.1476+1G>A
Normal splicing
Exon 12 skipping
Cryptic 5' splice-site
Intron 11 retention
Intron 12 retention
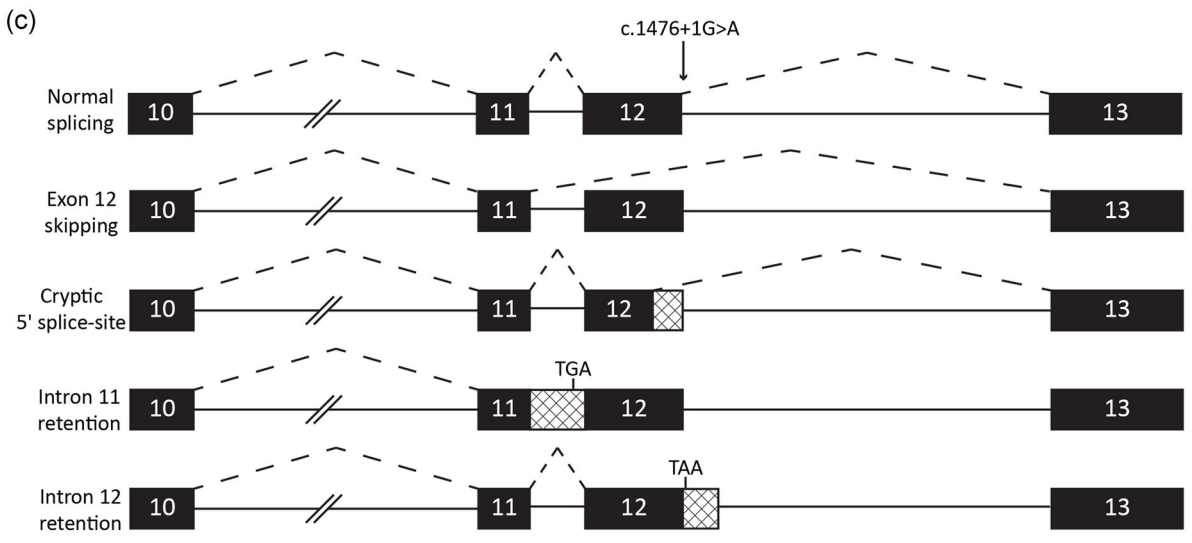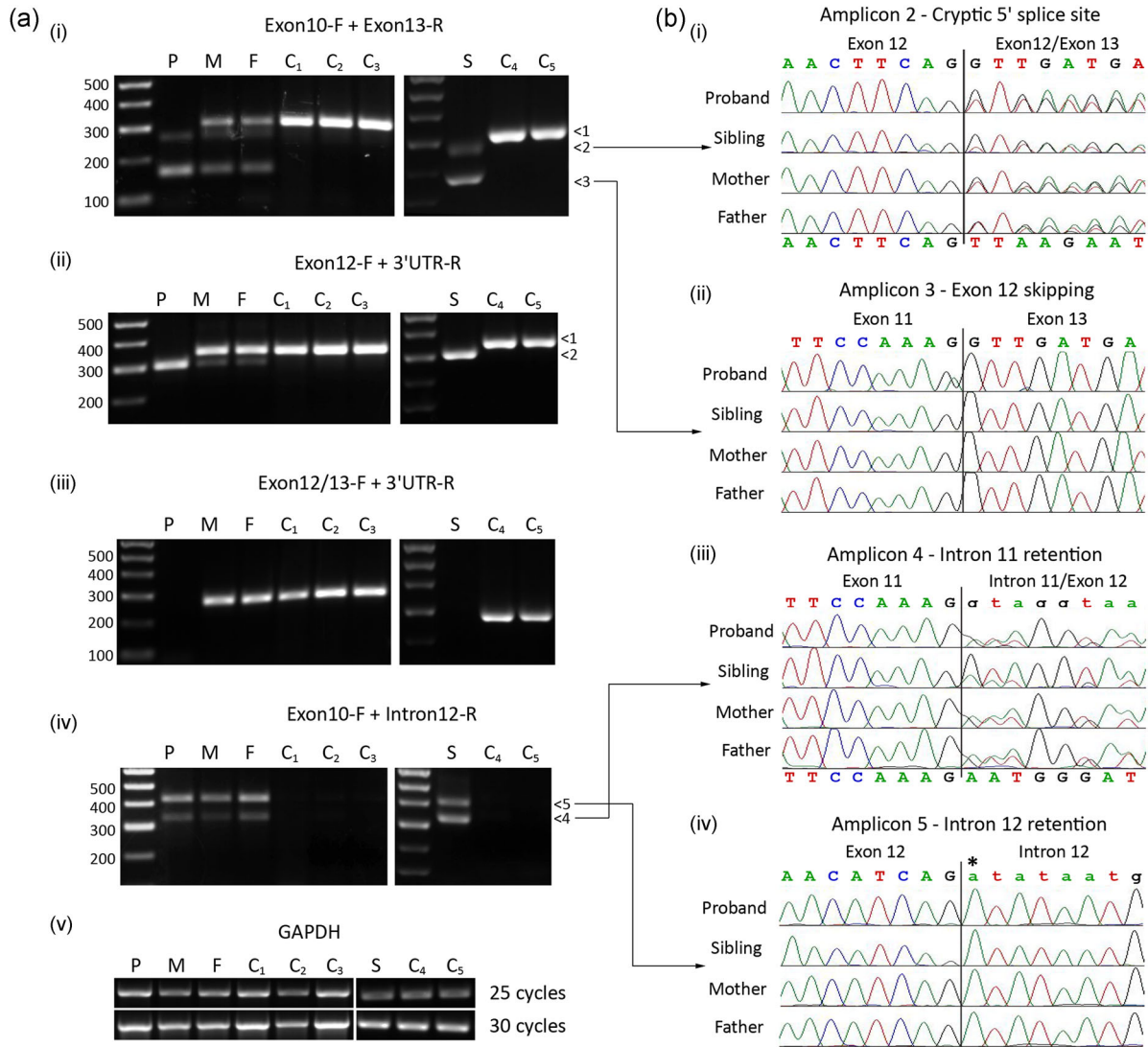
undetectable by polymerase chain reaction (PCR) amplification. Two of the three abnormal splicing events maintained the asparagine synthetase open reading frame, though deleted multiple sequential evolutionarily conserved amino acids that comprise the enzymatic domain (deleted residues are conserved to yeast/zebrafish). The severe neurological phenotype observed in both affected siblings is consistent with the high likelihood of enzymatic dysfunction or protein deficiency resulting from abnormal splicing of *ASNS* transcripts, which affects the conserved asparagine synthetase enzymatic domain. Although emerging technologies in clinical genetics raise the exciting potential for targeted treatments, therapies with asparagine have had mixed results (Alrifai & Alfadhel, 2016; Sprute et al., 2019) and the degree of brain malformation makes the success of postnatal intervention less likely.

Subsequent to diagnosis, RNA derived from blood (proband) and fibroblasts (deceased sibling) was subject to short-read RNA sequencing (150 nt paired-end reads). The blood specimen failed. RNA sequencing of the fibroblasts showed clear evidence for mis-splicing (exon 12 skipping, intron-12 retention, 5′ cryptic splice site use). However, 15% of reads with cryptic splice site use were mis-mapped (100/667) and incorrectly aligned to the authentic 5′ splice site junction with mis-matching (Figure S3). Short-read RNA sequencing is prone to mis-mapping, and particularly for mis-spliced reads that do not match the reference transcriptome. Short-read RNA sequencing is powerful though holds inherent limitations in that reads do not bridge multiple exons, in some cases presenting diagnostic uncertainty related to which isoform(s) and which allele(s) is/are affected by mis-splicing, especially in cases with alternative splicing of exons adjacent to a putative splicing variant. Due to these considerations, we recommend abnormal findings observed by short-read RNA sequencing be confirmed by RT-PCR before being used clinically. Long-read sequencing is entering the diagnostic horizon and may hold improved time and cost-efficiencies for diagnostic use as an alternative to multiple, bespoke PCRs, gel extraction, and Sanger sequencing of amplicons.

In conclusion, we present the neonatal diagnosis of asparagine synthetase deficiency facilitated by rapid genomic testing (exome sequencing and adjunct mRNA analyses) of a critically unwell neonate. Although not all variants of uncertain significance are amenable to clinical functional genomics, establishment of a rapid pipeline for mRNA analyses has the potential to increase the number of definitive diagnoses, with significant clinical and health economic benefits.

CONFLICT OF INTEREST

Professor Sandra Cooper is the director of Frontier Genomics Pty Ltd (Australia). Frontier Genomics has not traded (as of October 2019). Frontier Genomics Pty Ltd (Australia) has no existing financial relationships that will benefit from publication of these data. The remaining co-authors do not have any relationships, financial or otherwise, that may result in a perceived conflict of interest.

DATA AVAILABILITY STATEMENT

The data reported in this case report have not been submitted to a publicly available database to respect patient confidentiality.

FIGURE 2 Adjunct RT-PCR studies of *ASNS* pre-mRNA splicing. (a) Gel electrophoresis of RT-PCR reactions. RT-PCR was performed using mRNA isolated from whole blood from the family trio and three controls (C_1 = disease control 1, male, 7 months; C_2 = disease control 2, male, 5 years; C_3 = control 3, female, 43 years), and subsequently, using cDNA from primary fibroblasts from the affected sibling and two controls (C_4 = disease control 4, male, 8 months; C_5 = control 5, male fetus, 31/40 weeks). P, affected proband; M, unaffected mother; F, unaffected father; S, affected sibling. (i) Forward primer in exon 10 and reverse primer in exon 13. (ii) Forward primer in exon 12 and reverse primer in the 3′UTR. (iii) Forward primer bridging the junction of exon 12/exon 13 to selectively amplify transcripts with normal splicing of exons 12–13. We could not find evidence for normal exons 12–13 splicing in the proband and his sibling. (iv) Forward primer in exon 10 and reverse primer in intron 12 to selectively amplify transcripts with intron 12 retention. Retention of intron 12, and or introns 11 and 12, was not detected in controls and observed only in the affected individuals and parent carriers of the c.1476+1G>A variant. (v) Forward and reverse primers in *GAPDH* used as a cDNA loading control. Replicate samples were subject to PCR for 25 or 30 cycles to confirm sub-saturating PCR conditions and demonstrate loading and quality of cDNA. (b) Sanger sequencing of gel-purified bands. Amplicon 1 was confirmed to correspond to normal splicing of exons 10–11–12–13 (444 bp). Amplicon 2 corresponds to use of an exon 12 cryptic 5′ splice site (396 bp), 48 nucleotides upstream from the authentic 5′ splice site. Amplicon 3 corresponds to exon 12 skipping (288 bp). Amplicon 4 corresponds to retention of intron 12 (322 bp). Amplicon 5 corresponds to retention of both introns 11 and 12 (411 bp). (c) Schematic of the abnormal splicing events induced by the *ASNS* c.1476+1G>A variant, showing the positions of the exon 12 5′ cryptic splice site and encoded stop codons resulting from retention of intron 11 or intron 12. cDNA, complementary DNA; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; mRNA, messenger RNA; PCR, polymerase chain reaction; RT-PCR, reverse transcription polymerase chain reaction

ORCID

Lauren S. Akesson (iD) https://orcid.org/0000-0002-7439-612X

Adam Bournazos (iD) https://orcid.org/0000-0002-6464-4548

Zornitza Stark (iD) https://orcid.org/0000-0001-8640-1371

Sandra T. Cooper (iD) https://orcid.org/0000-0002-7314-5097

REFERENCES

Abhyankar, A., Lamendola-Essel, M., Brennan, K., Giordano, J. L., Esteves, C., Felice, V., ... Jobanputra, V. (2018). Clinical whole exome sequencing from dried blood spot identifies novel genetic defect underlying asparagine synthetase deficiency. *Clinical Case Reports*, 6(1), 200–205. https://doi.org/10.1002/ccr3.1284

Alfadhel, M., Alrifai, M. T., Trujillano, D., Alshaalan, H., Al Othaim, A., Al Rasheed, S., ... Eyaid, W. (2015). Asparagine synthetase deficiency: New inborn errors of metabolism. *JIMD Reports*, 22, 11–16. https://doi.org/10.1007/8904_2014_405

Alfadhel, M., & El-Hattab, A. W. (2018). Asparagine synthetase deficiency. In M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, & A. Amemiya (Eds.), *GeneReviews*. Seattle, WA: University of Washington, Seattle.

Alrifai, M. T., & Alfadhel, M. (2016). Worsening of seizures after asparagine supplementation in a child with asparagine synthetase deficiency. *Pediatric Neurology*, 58, 98–100. https://doi.org/10.1016/j.pediatrneurol.2016.01.024

Ben-Salem, S., Gleeson, J. G., Al-Shamsi, A. M., Islam, B., Hertecant, J., Ali, B. R., & Al-Gazali, L. (2015). Asparagine synthetase deficiency detected by whole exome sequencing causes congenital microcephaly, epileptic encephalopathy and psychomotor delay. *Metabolic Brain Disease*, 30(3), 687–694. https://doi.org/10.1007/s11011-014-9618-0

Clark, M. M., Hildreth, A., Batalov, S., Ding, Y., Chowdhury, S., Watkins, K., ... Kingsmore, S. F. (2019). Diagnosis of genetic diseases in seriously ill children by rapid whole-genome sequencing and automated phenotyping and interpretation. *Science Translational Medicine*, 11(489). https://doi.org/10.1126/scitranslmed.aat6177

Farnaes, L., Hildreth, A., Sweeney, N. M., Clark, M. M., Chowdhury, S., Nahas, S., ... Kingsmore, S. F. (2018). Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. *NPJ Genomic Medicine*, 3, 10. https://doi.org/10.1038/s41525-018-0049-4

French, C. E., Delon, I., Dolling, H., Sanchis-Juan, A., Shamardina, O., Megy, K., ... Raymond, F. L. (2019). Whole genome sequencing reveals that genetic conditions are frequent in intensively ill children. *Intensive Care Medicine*, 45(5), 627–636. https://doi.org/10.1007/s00134-019-05552-x

Galada, C., Hebbar, M., Lewis, L., Soans, S., Kadavigere, R., Srivastava, A., ... Shukla, A. (2018). Report of four novel variants in ASNS causing asparagine synthetase deficiency and review of literature. *Congenit Anom (Kyoto)*, 58(5), 181–182. https://doi.org/10.1111/cga.12275

Grody, W. W., Thompson, B. H., & Hudgins, L. (2013). Whole-exome/genome sequencing and genomics. *Pediatrics*, 132(Suppl 3), S211–S215. https://doi.org/10.1542/peds.2013-1032E

Gupta, N., Tewari, V. V., Kumar, M., Langeh, N., Gupta, A., Mishra, P., ... Kabra, M. (2017). Asparagine synthetase deficiency-report of a novel mutation and review of literature. *Metabolic Brain Disease*, 32(6), 1889–1900. https://doi.org/10.1007/s11011-017-0073-6

Lonsdale, J., Thomas, J., & Salvatore, M., et al (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, 45(6), 580–585. https://doi.org/10.1038/ng.2653

Meng, L., Pammi, M., Saronwala, A., Magoulas, P., Ghazi, A. R., Vetrini, F., ... Lalani, S. R. (2017). Use of exome sequencing for infants in intensive care units: Ascertainment of severe single-gene disorders and effect on medical management. *JAMA Pediatr*, 171(12), e173438. https://doi.org/10.1001/jamapediatrics.2017.3438

Mestek-Boukhibar, L., Clement, E., Jones, W. D., Drury, S., Ocaka, L., Gagunashvili, A., ... Williams, H. J. (2018). Rapid Paediatric Sequencing (RaPS): Comprehensive real-life workflow for rapid diagnosis of critically ill children. *Journal of Medical Genetics*, 55(11), 721–728. https://doi.org/10.1136/jmedgenet-2018-105396

Palmer, E. E., Hayner, J., Sachdev, R., Cardamone, M., Kandula, T., Morris, P., ... Kirk, E. P. (2015). Asparagine synthetase deficiency causes reduced proliferation of cells under conditions of limited asparagine. *Molecular Genetics and Metabolism*, 116(3), 178–186. https://doi.org/10.1016/j.ymgme.2015.08.007

Petrikin, J. E., Cakici, J. A., Clark, M. M., Willig, L. K., Sweeney, N. M., Farrow, E. G., ... Kingsmore, S. F. (2018). The NSIGHT1-randomized controlled trial: Rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. *NPJ Genomic Medicine*, 3, 6. https://doi.org/10.1038/s41525-018-0045-8

Petrikin, J. E., Willig, L. K., Smith, L. D., & Kingsmore, S. F. (2015). Rapid whole genome sequencing and precision neonatology. *Seminars in Perinatology*, 39(8), 623–631. https://doi.org/10.1053/j.semperi.2015.09.009

Radha Rama Devi, A., & Naushad, S. M. (2019). Molecular diagnosis of asparagine synthetase (ASNS) deficiency in two Indian families and literature review of 29 ASNS deficient cases. *Gene*, 704, 97–102. https://doi.org/10.1016/j.gene.2019.04.024

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... ACMG Laboratory Quality Assurance Committee (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. *Genetics in Medicine*, 17(5), 405–424. https://doi.org/10.1038/gim.2015.30

Sacharow, S. J., Dudenhausen, E. E., Lomelino, C. L., Rodan, L., El Achkar, C. M., Olson, H. E., ... Kilberg, M. S. (2018). Characterization of a novel variant in siblings with asparagine synthetase deficiency. *Molecular Genetics and Metabolism*, 123(3), 317–325. https://doi.org/10.1016/j.ymgme.2017.12.433

Schlapbach, L. J., Straney, L., Gelbart, B., Alexander, J., Franklin, D., Beca, J., ... The Australian & New Zealand Intensive Care Society (ANZICS) Paediatric Study Group. (2017). Burden of disease and change in practice in critically ill infants with bronchiolitis. *European Respiratory Journal*, 49, 1601648. https://doi.org/10.1183/13993003.01648-2016

Seidahmed, M. Z., Salih, M. A., Abdulbasit, O. B., Samadi, A., Al Hussien, K., Miqdad, A. M., ... Alkuraya, F. S. (2016). Hyperekplexia, microcephaly and simplified gyral pattern caused by novel ASNS mutations, case report. *BMC Neurology*, 16, 105. https://doi.org/10.1186/s12883-016-0633-0

Sprute, R., Ardicli, D., Oguz, K. K., Malenica-Mandel, A., Daimaguler, H. S., Koy, A., ... Cirak, S. (2019). Clinical outcomes of two patients with a novel pathogenic variant in ASNS: Response to asparagine supplementation and review of the literature. *Human Genome Variation*, 6, 24. https://doi.org/10.1038/s41439-019-0055-9

Stark, Z., Lunke, S., & Brett, G. R., et al (2018). Meeting the challenges of implementing rapid genomic testing in acute pediatric care. *Genet Med*, 20(12), 1554–1563. https://doi.org/10.1038/gim.2018.37

Sun, J., McGillivray, A. J., Pinner, J., Yan, Z., Liu, F., Bratkovic, D., ... Chopra, M. (2017). Diaphragmatic eventration in sisters with asparagine synthetase deficiency: A novel homozygous ASNS mutation and expanded phenotype. *JIMD Reports*, 34, 1–9. https://doi.org/10.1007/8904_2016_3

van Diemen, C. C., Kerstjens-Frederikse, W. S., Bergman, K. A., de Koning, T. J., Sikkema-Raddatz, B., van der Velde, J. K, ... Wijmenga, C. (2017). Rapid targeted genomics in critically ill newborns. *Pediatrics*, 140(4), e20162854. https://doi.org/10.1542/peds.2016-2854

Willig, L. K., Petrikin, J. E., Smith, L. D., Saunders, C. J., Thiffault, I., Miller, N. A., ... Kingsmore, S. F. (2015). Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: A retrospective analysis of diagnostic and clinical findings. *Lancet Respir Med*, 3(5), 377–387. https://doi.org/10.1016/S2213-2600(15)00139-3

Yamamoto, T., Endo, W., Ohnishi, H., Kubota, K., Kawamoto, N., Inui, T., … Fukao, T. (2017). The first report of Japanese patients with asparagine synthetase deficiency. *Brain and Development*, *39*(3), 236–242. https://doi.org/10.1016/j.braindev.2016.09.010

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Rapid exome sequencing and adjunct RNA studies confirm pathogenicity of a homozygous ASNS

splicing variant in a critically ill neonate

Akesson LS, Bournazos A, Fennell A *et al*

SUPPLEMENTARY INFORMATION
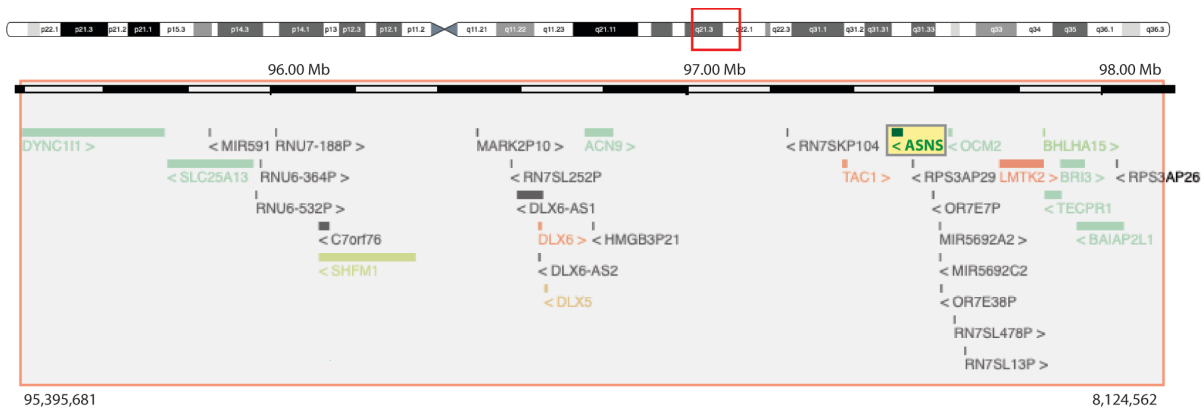
Figure S1. DECIPHER genome browser (Firth *et al*, 2009) showing the shared region of homozygosity

on chromosome 7, including *ASNS*.
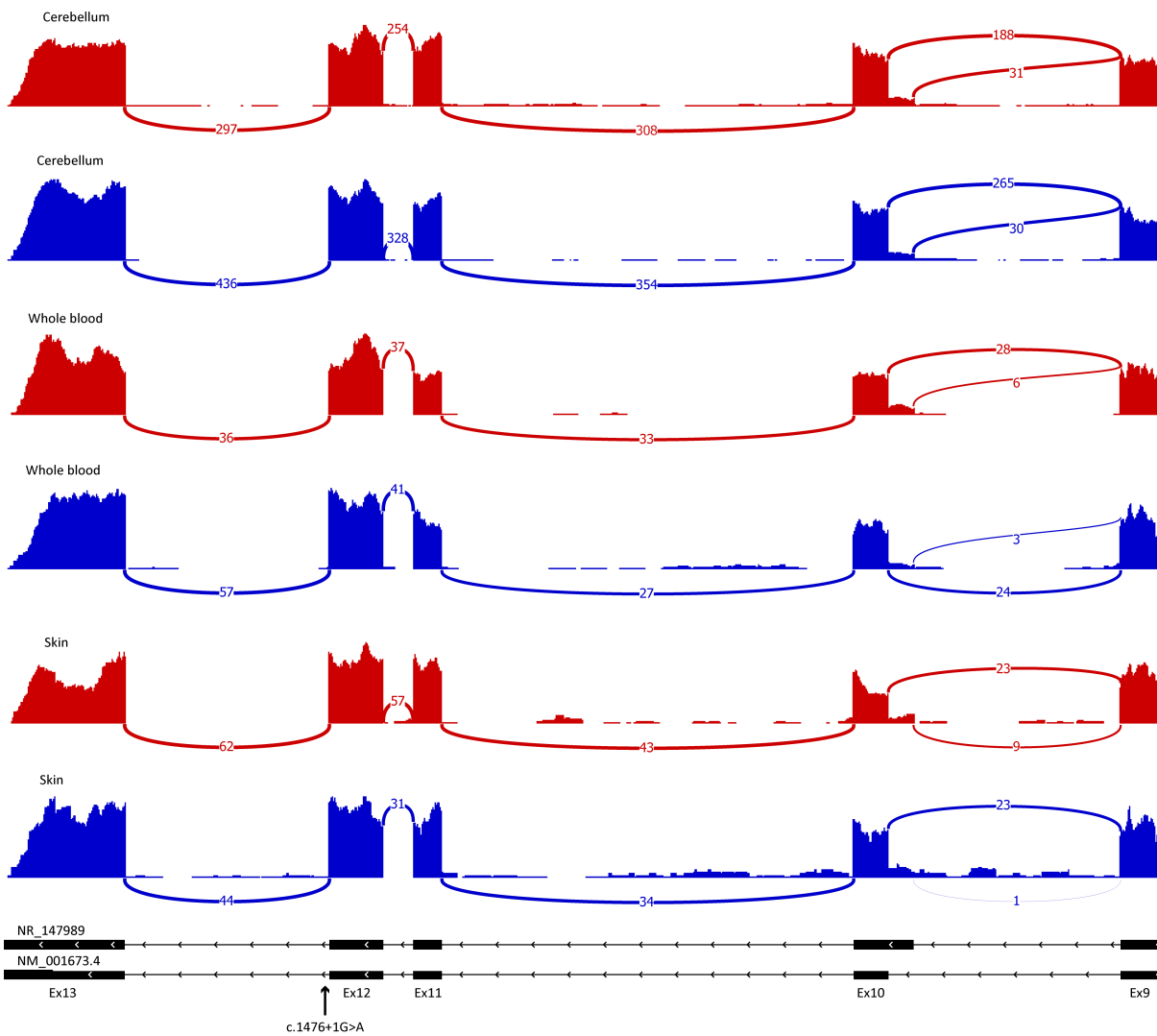
Figure S2. Sashimi plots of RNA sequencing BAM files for *ASNS* (NM_001673.4) showing exons 9-13 are included in all predominant isoforms of *ASNS* found in mRNA derived from the cerebellum, whole blood and skin of two individuals from the GTEx Consortium. *NOTE: ASNS is on the minus strand and therefore sequential exons go from right to left).* *ASNS* exon-12 is a canonical exon included in all predominant *ASNS* isoforms expressed in brain, blood and skin. Red, male, 25 years; Blue, female, 37 years.
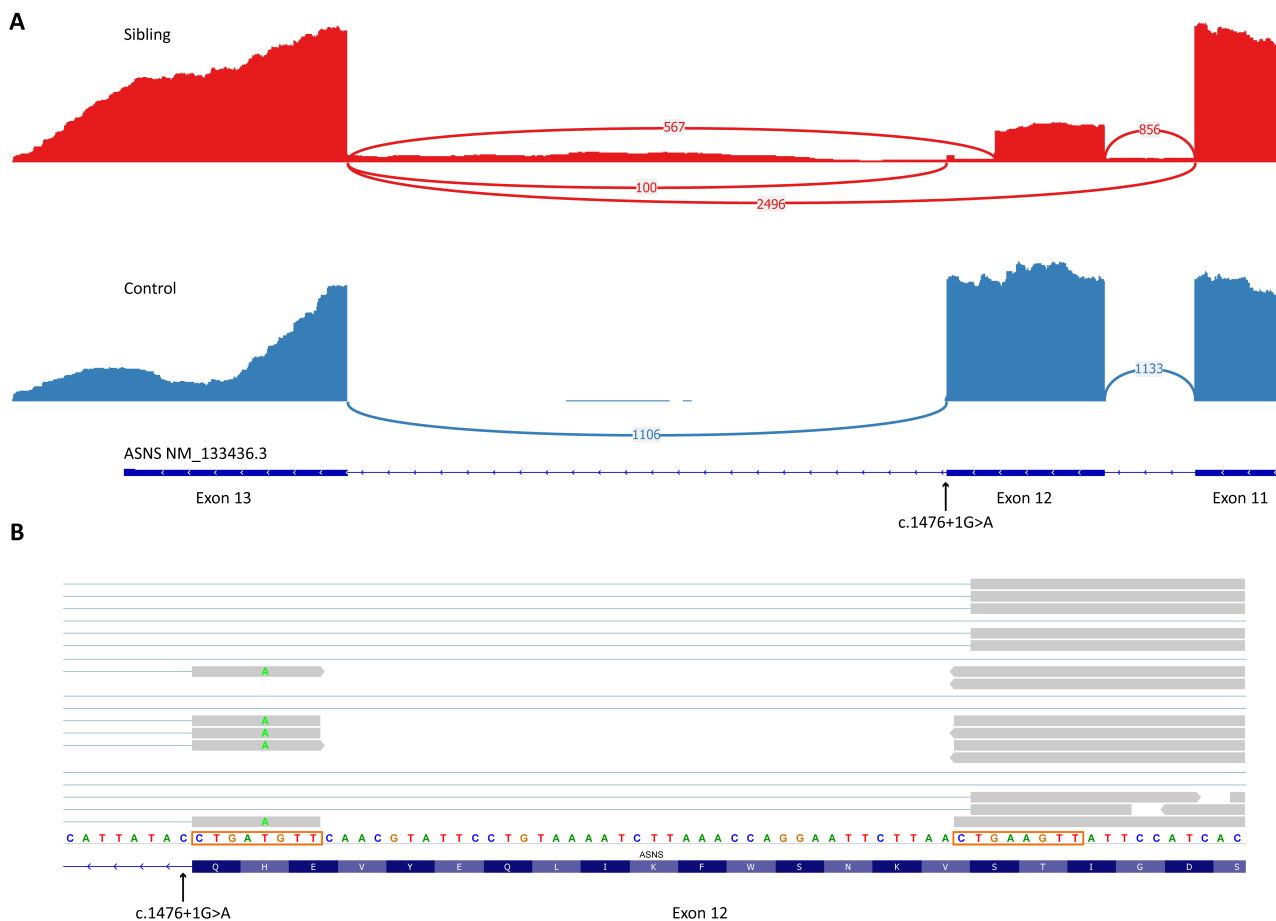
Figure S3. RNA sequencing identifies mis-splicing of ASNS pre-mRNA derived from skin fibroblasts from the affected deceased sibling. (A) RNA-sequencing clearly identifies exon 12 skipping, increased levels of retention of introns 11 and 12 retention, and use of a cryptic 5' splice site deleting 48 nucleotides of exon 12. (B) The orange boxed regions highlight an 8 nucleotide region of homology and basis for mis-alignment of 100 reads corresponding to cryptic 5' splice site use that are incorrectly mapped to the authentic 5'splice site and falsely report a c.1472A>T variant (Chr7:g.97482376T>A). This highlights a current major limitation of short read RNA sequencing, whereby the position of the variant within the read influences mapping accuracy. In this case, 567 junctional reads map correctly to the exon 12 5'cryptic splice site, yet 100 reads report a c.1472A>T variant in order to align the read to the reference splice junction. Despite this, high read count for ASNS enables strong diagnostic certainty of mis-splicing. However, one can appreciate grave diagnostic concerns related to potential mis-mapping with lower read depths of 10 – 50, particularly given mis-spliced reads are those prone to mis-mapping.

MATERIALS AND METHODS

Editorial policies and ethical considerations

The Australian Genomics Acute Care study has Human Research Ethics Committee approval (HREC/16/MH251). All other research described herein was approved by the relevant local ethics committees of the participating institutions (Monash Genetics, Victorian Clinical Genetics Services, Sydney Children's Hospitals Network). Written, informed consent was obtained from the parents of the proband and his deceased sibling. All research was conducted in accordance with the Declaration of Helsinki.

Rapid genomic diagnosis: trio exome sequencing

The Australian Genomics Acute Care study is a rapid genomic diagnosis program for seriously ill neonates and children with suspected monogenic disorders from neonatal and pediatric intensive care units (NICUs/PICUs). The program aims to issue diagnostic reports within 5 days from receipt of clinical samples.

DNA was extracted from peripheral blood specimens using the QIAamp DNA blood mini kit (Qiagen, Düsseldorf, Germany). DNA concentration was measured using the Invitrogen Qubit dsDNA HS Assay kit (Thermo Fisher Scientific, Waltham, MA, USA), and quality metrics for all samples were assessed using the Agilent Tape Station Genomic DNA ScreenTape (Agilent Technologies, Santa Clara, CA, USA).

Trio exome sequencing was performed at the Victorian Clinical Genetics Services (Melbourne, Australia) by massively parallel sequencing using SureSelect QXT CRE exome capture (Agilent Technologies) on a NextSeq 500 sequencer (Illumina, San Diego, CA, USA), with a targeted mean coverage of 100× with a minimum of 90% of bases sequenced to at least 15×. Variants were

characterized using the DRAGEN™ (Dynamic Read Analysis for GENomics) Bio-IT Platform (Illumina) to generate annotated variant calls within the target region (coding exons ± 2 base pairs), via alignment to the reference genome (GRCh37).

Phenotype-driven variant prioritisation was performed by a multidisciplinary team as previously described (Stark *et al.*, 2017; Stark *et al.*, 2016). Variants were assessed in accordance with the American College of Medical Genetics and Genomics (ACMG) variant classification criteria (Richards *et al.*, 2015).

Chromosomal microarray

Chromosomal microarray was performed on the Illumina Infinium GSA-24 v2.0 (Illumina) (proband) and the Illumina HumanCytoSNP-12 v2.1 (Illumina) (deceased sibling) platforms with a copy number change resolution of 0.20 Mb. Interpretation was based on the UCSC hg19 (proband) or hg18 (deceased sibling) human reference sequence (International Human Genome Sequencing Consortium, 2001). Microarray data were analyzed using KaryoStudio (Illumina). Regions of long continuous stretches of homozygosity (LCSH) were detected using KaryoStudio above a genomic size of 2 Mb. Identification of genes within regions of homozygosity was performed using the UCSC genome browser (Kent *et al.*, 2002) and interpreted for clinical relevance using the DECIPHER database (Firth *et al.*, 2009) and Online Mendelian Inheritance in Man (OMIM) (Amberger, Bocchini, Schiettecatte, Scott, & Hamosh, 2015).

Biochemical neurometabolic evaluation

Biochemical measurement of blood and cerebrospinal fluid (CSF) amino acids was performed on ACQUITY UPLC amino acid analyser according to the manufacturer's instructions (Waters

Corporation, Milford, MA, USA). Data analysis was performed using Empower software (Waters Corporation).

ASNS rapid splicing studies

mRNA was extracted from 2.5 ml of whole blood collected in a PAXgene blood RNA tube (PreAnalytiX, Qiagen) according to the manufacturer's instructions. mRNA was also extracted from a cryopreserved fibroblast cell line obtained in 2011 from a skin biopsy from the deceased sibling. Primary skin fibroblasts were cultured in a 6-well plate containing high glucose DMEM (Gibco™, Thermo Fisher Scientific), 10% fetal bovine serum (GE Healthcare, Chicago, IL, USA) and Gentamicin (50 µg/ml) (Gibco™). To inhibit nonsense-mediate decay of abnormally-spliced transcripts, primary fibroblasts were treated with 100 µg/ml cycloheximide (Sigma-Aldrich, St Louis, MO, USA), or dimethyl sulfoxide carrier control, for 6 h before harvesting in TRIzol™ reagent (Invitrogen™, Thermo Fisher Scientific). RNA was isolated using the standard TRIzol™ procedure followed by the RNase-free DNase set (Qiagen) and RNeasy mini kit cleanup protocol (Qiagen). cDNA was synthesized using SuperScript™ IV first-strand synthesis system (Invitrogen™) from 500 ng of RNA according to kit instructions. Recombinant Taq DNA polymerase (Invitrogen™) was used for PCRs; 95°C 3 min; 35 cycles 95°C 30 s, 58°C 30 s, 72°C 30-60 s depending on amplicon length; 72°C 5 min. RT-PCR products were analysed on a 1.2% agarose gel followed by Sanger sequencing of purified bands (GeneJET gel extraction kit, Thermo Fisher Scientific). Primers: a) ASNS forward primers; Ex10-F 5'-CGCAGATCGAACTACTGCTG-3', Ex12-F 5'-CGACCAAAAGAAGCCTTCAG-3', Ex12/13-F 5'-GGAATACGTTGAACATCAGGTT-3'; a) ASNS reverse primers In12-R 5'-TGACAGCTCTGCATCCAAAC-3', Ex13-R 5'-AAATTTCTGGGCTGCATTTG-3', 3'UTR-R 5'-CCCATCCAACACGAAGAAAT-3'; c) GAPDH forward primer GAPDH-F 5'-TGAAGGTCGGAGTCAACGGATTTGGT-3'; GAPDH reverse primer GAPDH-R 5'-CATGTGGGCCATGAGGTCCACCAC-3'

Segregation analysis

DNA was extracted from a cryopreserved fibroblast cell line from the deceased sibling using the QIAamp DNA mini kit (Qiagen). DNA concentration was measured using the NanoDrop® ND (Thermo Fisher Scientific). The (NM_133436.3(ASNS):c.1476+1G>A) variant was detected by Sanger sequencing using custom primers (Integrated DNA Technologies, Coralville, IA, USA) designed using Primer3 algorithm (Untergasser *et al.*, 2012). DNA amplification was performed on the proband sample as a positive control and DNA from the deceased sibling using HotStart Taq (Qiagen). PCR products were purified using Agencourt AMPure beads (Beckman Coulter, Indianapolis, IN, USA). The purified products were sequenced using the Big Dye Terminator Cycle v3.1 Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific) and loaded on the ABI 3730/3730XL Genetic Analyzer (Applied Biosystems). Data analysis was performed using MutationSurveyorV4.09 (SoftGenetics, State College, PA, USA).

Cost analysis

A conservative estimate of the cost savings of earlier diagnosis in the proband compared to the deceased sibling was made by multiplying the difference in hospitalization length of stay between the proband and the deceased sibling by the average daily cost of hospitalization in a pediatric intensive care unit (AU$6,200).(Schlapbach *et al.*, 2017) The cost of trio exome sequencing and research testing of the variant of uncertain significance was subtracted from the cost saving related to days of hospitalization.

REFERENCES (for supplementary information only)

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., & Hamosh, A. (2015). OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res, 43*(Database issue), D789-798. doi:10.1093/nar/gku1205

Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., . . . Carter, N. P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet, 84*(4), 524-533. doi:10.1016/j.ajhg.2009.03.010

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature, 409*(6822), 860-921. doi:10.1038/35057062

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Res, 12*(6), 996-1006. doi:10.1101/gr.229102

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Committee, A. L. Q. A. (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*(5), 405-424. doi:10.1038/gim.2015.30

Schlapbach, L. J., Straney, L., Gelbart, B., Alexander, J., Franklin, D., Beca, J., . . . New Zealand Intensive Care Society Paediatric Study, G. (2017). Burden of disease and change in practice in critically ill infants with bronchiolitis. *Eur Respir J, 49*(6). doi:10.1183/13993003.01648-2016

Stark, Z., Dashnow, H., Lunke, S., Tan, T. Y., Yeung, A., Sadedin, S., . . . James, P. A. (2017). A clinically driven variant prioritization framework outperforms purely computational approaches for the diagnostic analysis of singleton WES data. *Eur J Hum Genet, 25*(11), 1268-1272. doi:10.1038/ejhg.2017.123

Stark, Z., Tan, T. Y., Chong, B., Brett, G. R., Yap, P., Walsh, M., . . . White, S. M. (2016). A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet Med, 18*(11), 1090-1096. doi:10.1038/gim.2016.1

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res, 40*(15), e115. doi:10.1093/nar/gks596

Chapter 3

Standardised practices and interpretation guidelines for RNA diagnostics

3.1 Overview

This chapter contains the most significant contribution to this thesis. I reviewed >450 variants submitted to our splice variant submission portal to assess splicing predictions, loss-of-function constraint, allele frequency, splicing patterns between affected tissues and clinically accessible tissues, and isoform expression to inform experimental design and evaluate in-house *in silico* splice prediction tools. I had a leading role in coordinating this project liaising with clinicians and diagnostic laboratories to arrange biospecimens for testing, research testing consent, performed functional RNA testing (I also performed Western Blot analysis for 5 families) to inform ACMG-AMP variant classification, wrote diagnostic reports for >100 families with genetic variants predicted to impact pre-mRNA splicing. We also evaluated the comparative diagnostic utility of RNA-seq for 38 cases. Here, the utility of clinically accessible tissues was applied at scale as the manifesting tissue was unavailable for testing in the majority cases, which would have otherwise remained undiagnosed without functional testing.

3.1 Overview

Over the analysis of >120 putative splicing variants we acquired extensive knowledge of the potential impacts to pre-mRNA splicing and transcription regulation to devise standardised practices for RNA analysis of splicing variants by RT-PCR and Sanger sequencing. A survey of variant classifiers highlighted the inadequacy of current ACMG-AMP criteria to accurately classify splicing variants prompting the proposal our own recommendations for interpretation of RNA assay data for ACMG-AMP-aligned variant classification.

Upon receipt of RNA diagnostic reports clinicians and diagnostic scientists were surveyed and I collated the diagnostic and clinical impact of this study to publish a manuscript with 108 collaborating authors. The reclassification of 75% of variants of uncertain significance to likely pathogenic or pathogenic had greatest impact to family planning and reproductive counselling. Clinicians reported RNA testing had a positive impact for the family in 75% of cases and that families were relieved to have an established diagnosis with the option for prenatal testing in future pregnancies. This study demonstrates the significant diagnostic and health benefits of RNA diagnostics as adjunct testing to extend diagnostic yield from genomic testing.

I was first author on this publication which included 74 families from this cohort:

Bournazos AM, Riley LG, Bommireddipalli S, et al. Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genet Med.* 2021;24(1):130-145.

3 cases from this cohort were published separately as case/brief reports:

Akesson LS., **Bournazos A.**, Fennell A, et al. Rapid exome sequencing and adjunct RNA studies confirm the pathogenicity of a novel homozygous *ASNS* splicing variant in a critically ill neonate. *Hum Mutat*. 2020;41(11):1884-1891.

3.1 Overview

Katiyar D, Anderson N, Bommireddipalli S, et al. Two novel *B9D1* variants causing Joubert syndrome: Utility of mRNA and splicing studies. *Eur J Med Genet*. 2020;63(9):104000.

Huq AJ, Thompson BA, Bennett MF, et al. Clinical Impact of Whole Genome Sequencing in Patients with Early Onset Dementia. J Neurol Neurosurg Psychiatry. 2022 In Press.

ELSEVIER

Genetics in Medicine

ARTICLE

Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants

ABSTRACT

Purpose: Genetic variants causing aberrant premessenger RNA splicing are increasingly being recognized as causal variants in genetic disorders. In this study, we devise standardized practices for polymerase chain reaction (PCR)-based RNA diagnostics using clinically accessible specimens (blood, fibroblasts, urothelia, biopsy).
Methods: A total of 74 families with diverse monogenic conditions (31% prenatal-congenital onset, 47% early childhood, and 22% teenage-adult onset) were triaged into PCR-based RNA testing, with comparative RNA sequencing for 19 cases.
Results: Informative RNA assay data were obtained for 96% of cases, enabling variant reclassification for 75% variants that can be used for genetic counseling (71%), to inform clinical care (32%) and prenatal counseling (41%). Variant-associated mis-splicing was highly reproducible for 28 cases with samples from ≥2 affected individuals or heterozygotes and 10 cases with ≥2 biospecimens. PCR amplicons encompassing another segregated heterozygous variant was vital for clinical interpretation of 22 of 79 variants to phase RNA splicing events and discern complete from partial mis-splicing.
Conclusion: RNA diagnostics enabled provision of a genetic diagnosis for 64% of recruited cases. PCR-based RNA diagnostics has capacity to analyze 81.3% of clinically significant genes, with long amplicons providing an advantage over RNA sequencing to phase RNA splicing events. The Australasian Consortium for RNA Diagnostics (SpliceACORD) provide clinically-endorsed, standardized protocols and recommendations for interpreting RNA assay data.

© 2021 Published by Elsevier Inc. on behalf of American College of Medical Genetics and Genomics.

Introduction

Genetic variants causing abnormal splicing of premessenger RNA (pre-mRNA) may represent up to half of all disease-causing variations.[1] However, the vast majority of splicing variants outside the conserved GT-AG essential splice site will be classified as variants of unknown significance (VUS) according to the existing American College of Medical Genetics and Genomics and the Association of Molecular Pathology (ACMG-AMP) guidelines.[2] It is often not possible to confidently predict if and how a genetic variant will disrupt splicing. The only way to know with certainty is through functional testing of the spliced mRNA to define consequences for the encoded protein, enabling ACMG-AMP–guided variant reclassification for a definitive molecular diagnosis.[3-6]

*Correspondence and requests for materials should be addressed to Sandra T. Cooper, Kids Neuroscience Centre, Kids Research, The Children's Hospital at Westmead; Children's Medical Research Institute; Discipline of Child and Adolescent Health, Sydney Medical School, The University of Sydney. Locked Bag 4001, Westmead, New South Wales 2145, Australia. *E-mail address:* sandra.cooper@sydney.edu.au
A full list of authors and affiliations appears at the end of the paper.

Technical platforms devised to sequence DNA are imminently transferable to RNA. However, RNA is vastly more complicated than DNA. The central difference is that genomic DNA has 1 reference sequence and all sequencing reads are aligned back to this reference sequence. The challenge with RNA arises from alternative splicing, which leads to multiple reference mRNA isoforms for each gene.

The use of short-read RNA sequencing (RNA-seq) was previously investigated using muscle specimens, diagnosing 35% of 50 exome-negative families with neuromuscular disorders.[3] Despite these successes, short-read RNA-seq showed significant diagnostic limitations. Short reads of ≤150 nucleotides (nt) regularly do not span multiple exons to unambiguously identify which isoform is affected by any identified aberrant splicing. Furthermore, mis-spliced reads often do not match the reference genome and can be filtered out, mis-aligned, and/or present at comparatively low levels because of nonsense-mediated decay (NMD),[3,7,8] an innate surveillance pathway targeting transcripts with a premature termination codon.[9] A further challenge arises for heterozygous variants for which normally spliced mRNA is transcribed from the allele *in trans*. Unless a single read contains a segregated heterozygous variant to phase splicing events (ie, discern from which allele observed mis-splicing and normal splicing is arising), it is impossible to know whether a variant induces complete or partial mis-splicing.

For many genetic disorders, tissues from affected vital organs are rarely available for RNA studies. RNA-seq using RNA from whole blood, fibroblasts, or Epstein-Barr virus transformed lymphocytes (EBV-LCLs) is increasingly being used to improve diagnostic yield.[4-6] However, many OMIM genes are expressed at too low levels in blood, EBV-LCLs, or fibroblasts for diagnostic confidence of splicing outcomes via RNA-seq.[10,11] Using more sensitive reverse transcription polymerase chain reaction (RT-PCR), we have subsequently diagnosed 11 Australian families,[8,12-17] 4 of which were unsolved by muscle RNA-seq.[3] Furthermore, our informatics analyses established that blood, EBV-LCLs, skin fibroblasts, and urothelial cells collectively express 81.3% of clinically significant OMIM genes[18] (Supplemental Table 1) at levels our research indicates is sufficient for diagnostically informative results by RT-PCR (>0.5 transcripts per million [TPM]) (Figure 1A).

In this study, the Australasian Consortium for RNA Diagnostics (SpliceACORD) devise and evaluate standardized practices for PCR-based RNA diagnostics using RNA from clinically accessible specimens. Our specific goals were to (1) establish diagnostic criteria for clinical recommendation of RNA testing with high diagnostic return, (2) triage families in real time to undergo RNA testing and determine reproducibility of variant-associated (mis)splicing between multiple affected individuals or heterozygotes and multiple biospecimens (blood, skin fibroblasts, urothelial cells, available biopsies), (3) devise standard operational procedures and provide evidence deemed to be of sufficient rigor by pathologists and diagnostic scientists for variant classification, (4) perform microcosting analyses, (5) develop recommendations for clinical interpretation of splicing assay data, and (6) collate the diagnostic and health impacts of RNA testing.

Materials and Methods

Patient recruitment

A total of 74 families were recruited from local area health districts across Australia and New Zealand, fulfilling the devised inclusion criteria that are described in the Results section, because devising consensus ascertainment criteria was part of the consultation process for clinical codesign of standard operating procedures.

Sample collection and RNA extraction

Whole blood was collected in a PAXgene (PreAnalytiX) blood RNA tube and RNA was isolated using the PAXgene blood RNA kit according to kit instructions. Peripheral blood mononuclear cells (PBMCs) were isolated using SepMate-15 tubes (StemCell Technologies) and Ficoll Paque Plus (GE Healthcare). Urothelial cell culture was performed as described in steps 1 to 16 of Zhou et al.[19] RNA from PBMCs, transformed lymphocytes, urothelial, and cultured skin-derived fibroblasts was isolated using the standard TRIzol (Invitrogen) procedure followed by the RNase-free DNase set (Qiagen) and RNeasy (Qiagen) mini kit cleanup protocols. PBMCs, transformed lymphocytes, primary fibroblasts, and urothelia were treated with dimethyl sulfoxide (Sigma-Aldrich) or 100 μg/mL cycloheximide (Sigma-Aldrich) for 6 hours before harvesting in TRIzol reagent. Detailed protocols can be found in the Supplemental Methods.

RT-PCR and Sanger sequencing

SuperScript IV (Invitrogen) first-strand synthesis system was used to make complementary DNA (cDNA) from 500 ng of RNA according to kit instructions. Recombinant Taq DNA polymerase (Invitrogen) and MasterAmp 2X PCR PreMix D (Epicentre Biotechnologies) or LongAmp Taq DNA Polymerase (New England BioLabs) were used for the PCRs. Control cDNA was obtained from healthy individuals (when available), family members, or affected individuals from cases with genetic variants in an unrelated gene. All PCR products were analyzed on a 1.2% agarose gel. Bands were manually excised from an agarose gel with a scalpel and the cDNA was purified using the GeneJET gel extraction kit (Thermo Scientific) according to the manufacturer's instructions. Purified cDNA was subject to Sanger sequencing at the Australian Genomics Research Facility.

For detailed protocols, see the Supplemental Methods primers have been provided in Supplemental Table 2.
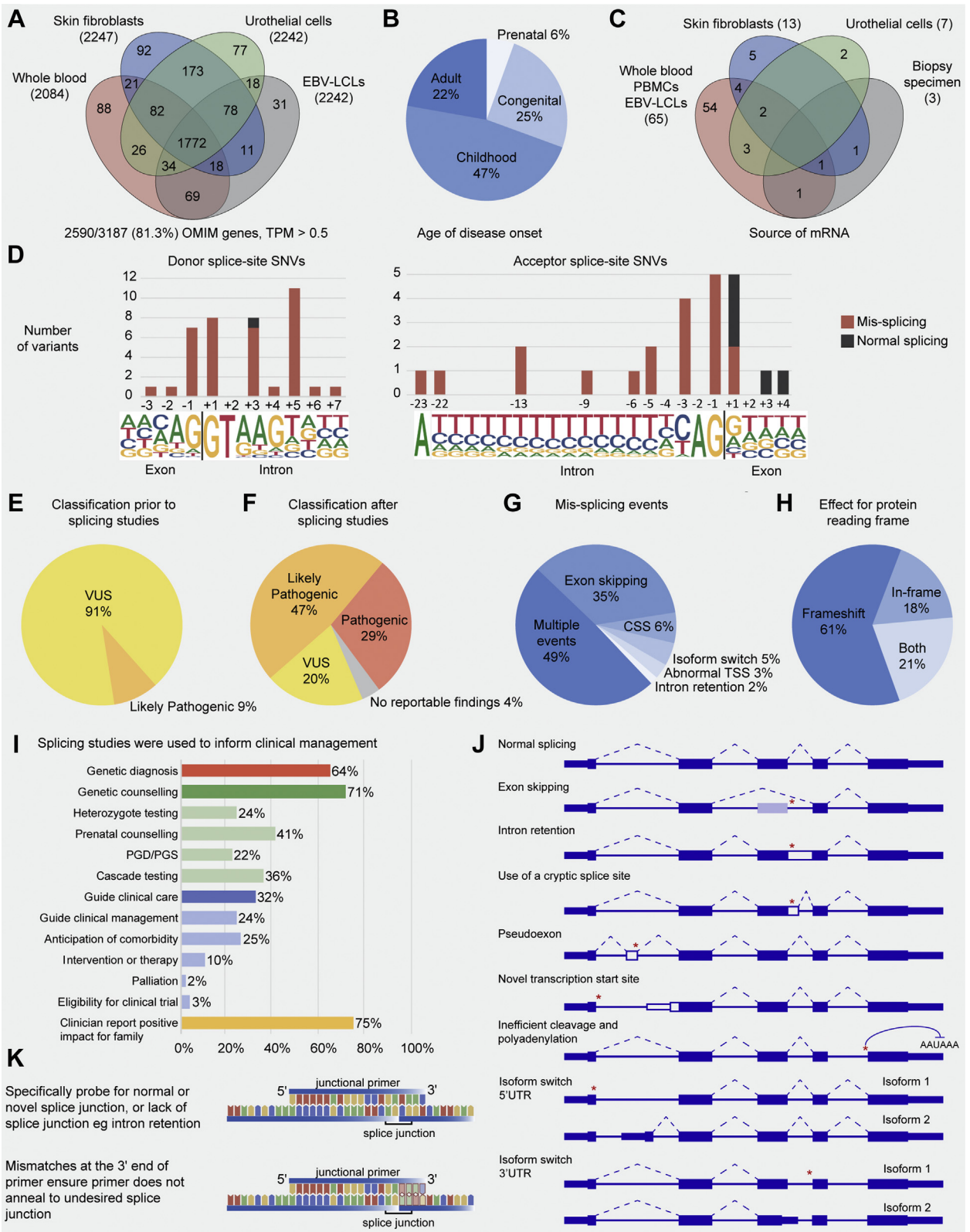
Figure 1 **Overview of the cohort of 74 families triaged in real time from clinical genomics into RNA diagnostics.** A. Informatics analyses of RNA sequencing data from whole blood, EBV-LCLs, skin fibroblasts, or urothelial cells shows that 81.3% of OMIM genes linked to clinically relevant Mendelian disorders are expressed at TPM >0.5 levels feasible for analysis by RT-PCR. B. Pie chart depicting age of disease onset for affected individuals from 74 families subject to RNA diagnostics. C. Venn diagram summarizing the biospecimens used as a source of RNA for this study. D. Position of putative splicing variants analyzed in this study relative to the donor splice site or acceptor splice site. Red: variants shown to induce mis-splicing. Black: variants maintaining normal splicing. Variant classification of

RNA-seq

Replicate Libraries were prepared from 1 µg of RNA using Illumina TruSeq Stranded mRNA (using poly-A selection) or Illumina TruSeq Stranded Total RNA Kit (using RiboZero beads to deplete ribosomal RNA) (Illumina Inc) per the manufacturer's instructions. Sequencing was performed on the NextSeq 500/550 sequencing platform using the High Output Reagent Cartridge v2, 300 cycles, 150 bp paired-end reads (Illumina Inc). Reads were aligned to GRCh37/hg19 reference genome using STAR aligner[20] (detailed information is provided in Supplemental Table 3) and visualized on Integrative Genomics Viewer.[21] Median TPM[22] values for whole blood, fibroblasts, urothelial, and EBV-LCLs are shown in Figure 1A.

In silico analysis and splicing predictions

Prediction algorithms available in Alamut Visual v2.10 (Interactive Biosoftware) and SpliceAI[23] were utilized for splicing predictions, using default thresholds (Alamut) or a Δ score threshold of 0.2 (SpliceAI).

Alternative splicing of the gene was scrutinized using in-house RNA-seq data or RNA-seq data derived from organs and tissues (fetal, child, and adult) obtained from the Genotype Tissue Expression (GTEx) project[24] (http://gtexportal.org/home/) or the Encyclopedia of DNA Elements (ENCODE)[25-27] (https://www.encodeproject.org/).

Variant submission portal and surveys

The Research Electronic Data Capture system (Vanderbilt University)[28] was used to manage case ascertainment and data capture via online surveys (Supplemental Figures 1-3).

Results

Our devised inclusion criteria to ascertain variants with high clinical suspicion of causality were (1) a high likelihood of a monogenic Mendelian disorder, (2) variant allele frequency consistent with disease incidence, (3) putative splicing variant in a clinician-defined, phenotypically concordant gene, and (4) preferably the variant segregates with disease (some cases studied in parallel with segregation because of clinical urgency). A total of 74 families with diverse Mendelian conditions were recruited with 6% with prenatal, 25% with severe-congenital, 47% with early childhood, and 22% with teenage-adult onset (Figure 1B; Supplemental Table 4). RNA was derived from blood (65/74), skin fibroblasts (13/74), urothelial cells (7/74), and/or an available biopsy specimen (3/74) (Figure 1C). RT-PCR and Sanger sequencing were performed for 129 individuals (2 quads, 18 trios, 13 duos, and 41 singletons) encompassing 79 splicing variants (Figure 1D and E). Of note, 19% of the variants affect the essential GT-AG splice sites, 71% affect the extended donor (5′) splice site or acceptor (3′) splice-site regions, 27% were exonic variants (Figure 1D), and 2 were structural/copy number variants.

Informative PCR data establishing normal splicing or mis-splicing of the target mRNA were obtained for 96% of the cases (71/74), enabling variant reclassification for 75% of variants (58 cases) (Figure 1F and H). We were unable to study cases A005-*CFTR* and A103-*PLP1* using blood RNA (<0.1 TPM) and these analyses are being repeated using urothelial cells (*CFTR*) or skin fibroblasts (*PLP1*). Case A158-*EDN3* could not be studied using RNA from blood or urothelia. Splicing studies that confidently established no evidence for variant-induced mis-splicing enabled classification of an alternative putative causal variant in 2 additional cases (3%). Importantly, 87% of the results were reported within a clinically relevant timeframe, including rapid turnaround of 10 to 21 days for 4 neonates in intensive care who had undergone ultrarapid genomic sequencing.[29] Of note, 71% of the diagnoses were used for genetic counseling, including additional diagnoses of 9 similarly affected family members; 41% of diagnoses were used for prenatal counseling with half of these cases with intended use for preimplantation genetic diagnosis or screening; 32% of diagnoses informed clinical care; 75% of diagnoses were clinician-reported to have a positive impact for the family; and for 24% of cases, the classification remain unchanged, having no or neutral impact (Figure 1I). For 2 cases, provision of a molecular diagnosis enabled eligibility for a clinical trial.

Importantly, an identical pattern of variant-induced mis-splicing was observed for all (28) cases with samples from at least 2 affected individuals or heterozygotes and all (10) cases with RNA studies of 2 or more biospecimens (including a manifesting tissue for 3 cases). Critical review of the available RNA-seq data (see Materials and Methods) established that there was no significant alternative splicing in the region of the gene containing the variant between the manifesting tissues and clinically accessible specimens, with case A134-*CDH23* being a notable exception (Supplemental Figure 4). Although our findings provide important validation for the reproducibility of variant-associated

---

putative splicing variants (E) before and (F) after splicing studies. G, H. Summary of mis-splicing events detected and their effect on protein reading frame. I. Summary of clinical impact metrics returned from referring clinicians by survey. J. Schematic illustrating the range of theoretical mis-splicing outcomes that may arise from a splicing variant. K. Schematic design of a junctional PCR primer bridging 2 exon junctions. CSS, cryptic splice site; EBV, Epstein-Barr virus; LCLs, lymphoblastoid cell lines; PGD, preimplantation genetic diagnosis; PGS, preimplantation genetic screening; RT-PCR, reverse transcription polymerase chain reaction; TPM, transcripts per million; TSS, transcription start site.
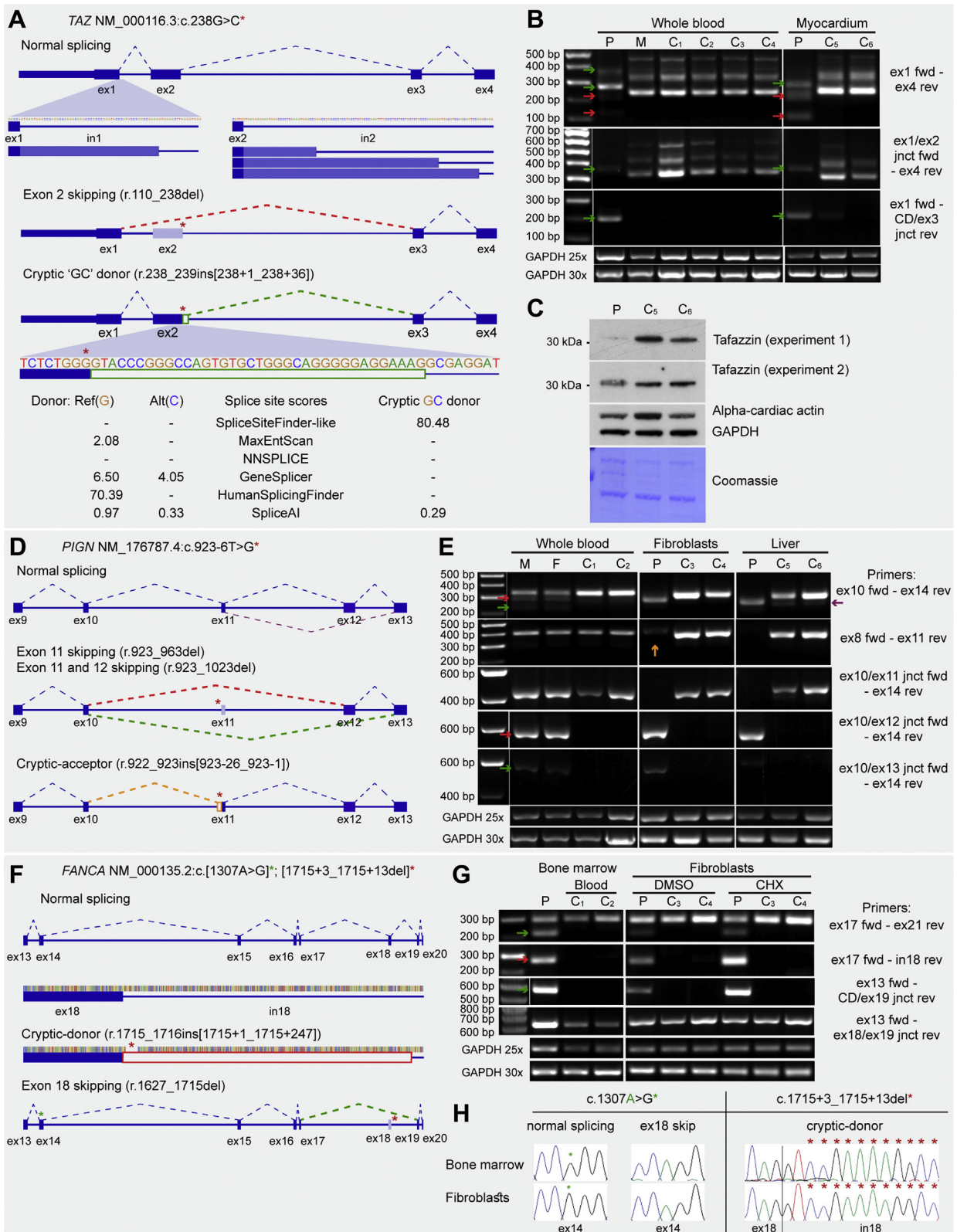
Figure 2 RT-PCR to interrogate for multiple mis-splicing events. A. Schematic of detected mis-splicing for case A022-*TAZ*. The NM_000116.3:c.238G>C variant (red asterisk) led to exon 2 skipping (red splice junction and arrows) and use of a cryptic donor (green splice junction and arrows). B. Left: RT-PCR of blood-derived cDNA from the proband (P: male, 7 months), his mother (M: female, 35 years), and controls (C1: male, 1 year; C2: male, 2 years; C3: female, 31 years; C4 female, 36 years). Right: cDNA derived from myocardium from the proband (P: male, 10 months old) and available myocardium controls (C5: female, 8 months; C6: female, 10 years). C. Western blot of 10 g of myocardial protein lysate shows marked reduction of encoded tafazzin protein. Detected tafazzin protein showed

mis-splicing between biospecimens, it is not possible to be certain of the pattern of splicing in any tissue without testing the RNA.

Standard operating procedures for informed design of PCR-based RNA diagnostics

Procedural guidelines for RT-PCR of cDNA are presented in Supplemental Figure 5. Technical aspects deemed essential by >90% of clinical variant curator (CVC) respondents (N = 18) were (1) critical scrutiny of tissue-specific or developmentally regulated alternative splicing of the target gene between manifesting tissue(s) and clinically accessible biospecimens through mining of RNA-seq data; (2) strategic design of primers to specifically interrogate all theoretically possible (mis)splicing outcomes (Figure 1J and K) and to mitigate technical caveats of PCR; (3) 2 methods confirming each splicing outcome (see Figures 2-4), ideally via 2 separate primer pairs or, if this is not possible, a repeat experiment with the same primer pair; (4) use of at least 2, and preferably 3, age-, sex- and tissue-matched controls; (5) gel extraction and Sanger sequencing of each PCR amplicon from the 2 methods (or experimental repeat). The following were deemed nonessential but highly desirable: (6) when possible for heterozygous splicing variants, use of a long amplicon encompassing a segregated heterozygous variant to phase splicing events from each allele to discern whether a variant induces complete or partial mis-splicing; and (7) test multiple affected individuals or heterozygotes to increase confidence in the reproducibility of variant-induced mis-splicing, although actionable results are possible with testing of a single proband.

Cost of testing

Our rigorous RT-PCR approach required an average of 8 primer pairs per case and an average costs of A$1823 per singleton and A$2563 per trio per research diagnostic report (Supplemental Table 5), which required subsequent evaluation by a genetic pathologist. Cost of testing of A$2500 per trio was deemed acceptable by 56% of survey respondents (clinicians, pathologists, and scientists), unacceptable by 3%, and 41% were unsure and made optional comments declaring that (1) costs of $2500 are insignificant relative to other clinical expenditure and (2) unanimous agreement that despite clear clinical utility of RNA diagnostics, a funding model to support RNA testing does not yet exist within the Australian public health system.

Reproducibility of variant-induced mis-splicing in multiple affected individuals or heterozygotes and different biospecimens

A022-TAZ is a male neonate admitted to intensive care with suspected Barth syndrome (MIM#302060) who presented with cardiac and metabolic abnormalities. Genomic sequencing[29] identified a TAZ missense variant affecting the last nucleotide of exon 2, which was classified as a VUS (ChrX[GRCh37]:g.153640551G>C; NM_000116.3:c.238G>C; p.[Gly80Arg]). RNA studies using blood-derived cDNA from the proband and his mother (segregation not available at time of this testing) established that the hemizygous TAZ c.238G>C variant induces 2 in-frame splicing defects (Figure 2A), namely (1) use of a cryptic-donor splice site in intron 2 (r.238_239ins[238+1_238+36]; p.Trp79_Gly80insArgThrArgAlaSerValLeuGlyArgGlyArg Lys) and (2) exon 2 skipping (r.110_238del; p.Lys37_ Gly80delinsArg). Strategic use of primers bridging splice junctions confirmed undetectable levels of normally spliced TAZ mRNA in the hemizygous proband (exon 1/2 junctional primer) (Figure 2B) and, conversely, that abnormal use of the cryptic-donor is detected only in the proband and is absent in controls (cryptic-donor/exon 3 junctional primer) (Figure 2B). Subsequent to reclassification of c.238G>C as likely pathogenic, the proband required surgical intervention for heart complications and a cardiac specimen was available. RNA studies confirmed identical mis-splicing events in the cardiac tissue with a western blot showing marked reduction of tafazzin protein (Figure 2C), enabling reclassification of c.238G>C as pathogenic. An affected younger

similar molecular weight and may represent the insertion of 12 amino acids, which leads to only a subtle increase in size. Alpha-cardiac actin, GAPDH, and Coomassie stained membrane are shown as loading controls. D. Schematic of detected mis-splicing for case A040-PIGN. The NM_176787.4:c.923-6T>G variant (red asterisk) led to exon 11 skipping (red splice junction and arrows), exon 11 and 12 skipping (green splice junction and arrows), and use of a cryptic acceptor (orange splice junction and arrow). E. RT-PCR of blood-derived cDNA from the parents (M: female, 36 years; F: male, 39 years) and controls (left) or cDNA from fibroblasts (middle) or liver (right) from the affected proband and disease controls. C1 blood cDNA: female, 35 years; C2 blood cDNA: male, 38 years; C3 fibroblast cDNA: male, 2 months; C4 fibroblast cDNA male, 8 months; C5 liver cDNA: male, 5 months; C6 liver cDNA female, 2 months. Diagnostic utility of heterozygous coding SNVs to discern complete from partial mis-splicing. F. Schematic of detected mis-splicing for case A206-FANCA. The NM_000135.2:c.1715+3_1715+13del variant (red asterisk) with NM_000135.2:c.1307A>G (green asterisk) in trans. G. FANCA mRNA studies using fibroblasts and bone marrow showed 2 abnormal splicing events in the proband (P: male, 28 years) that were absent in controls (C1: male, 28 years; C2: male, 48 years), namely exon 18 skipping (green splice junction and arrows) and use of a cryptic donor (red box and arrow). H. Sanger sequencing of cDNA with normal splicing (forward primer in exon 13 and reverse primer annealing to the exon 18/19 splice junction) shows apparent hemizygosity of the missense variant c.1307A>G in trans, establishing undetectable levels of normal splicing arising from the c.1715+3_1715+13del allele. Conversely, Sanger sequencing of smaller band corresponding to exon 18 skipping shows absence of the missense variant c.1307A>G in trans (thus confirming that detected transcripts with exon 18 skipping are arising from the c.1715+3_1715+13del allele). CD, cryptic donor; cDNA, complementary DNA; RT-PCR, reverse transcription polymerase chain reaction.

male sibling established the mother to be germline mosaic. Case A079-*LAMP2* also involved maternal germline mosaicism (Supplemental Figure 6).

Family A040-*PIGN* underwent termination of pregnancy because of multiple congenital anomalies (diaphragmatic defect, pulmonary hypoplasia, cardiovascular malformation, genital malformation, absent olfactory bulbs, and absent 12th ribs) suggestive of Fryn's syndrome[30] (MIM#614080). Genomic testing in the affected fetus identified a homozygous variant in *PIGN* (Chr18[GRCh37]:g.59810585A>C; NM_176787.4:c.923-6T>G) classified as a VUS. RT-PCR using blood-derived mRNA from the proband's heterozygous parents showed the c.923-6T>G variant-induced exon 11 skipping (r.c.923_963del; p.Glu308Glyfs*2) or skipping of exons 11 and 12 (r.923_1023del; p.Glu308Glyfs*9) (Figure 2D). Studies of mRNA derived from skin fibroblasts and liver specimen from the proband confirmed c.923-6T>G–induced complete mis-splicing with no detectable normal splicing of *PIGN* (Figure 2E).

Use of heterozygous variants to phase events and discern complete from partial mis-splicing

Heterozygous coding variants were crucial for clinical interpretation of splicing assay data for 22 of 66 variants with autosomal dominant or compound heterozygous recessive disease to distinguish complete from partial mis-splicing. For example, A206-*FANCA* is a male proband with acute myeloid leukemia and suspected Fanconi anemia (MIM#227650) with a paternal pathogenic missense variant (Chr16[GRCh37]:g.89857863T>C; NM_000135.2:c.1307A>G; p.[Gln436Arg]) and maternal VUS in trans (Chr16[GRCh37]:g.89846264_89846274del; NM_000135.2:c.1715+3_1715+13del) (Figure 2F). *FANCA* mRNA studies using fibroblasts and bone marrow showed 2 mis-splicing events in the proband that were absent in controls (Figure 2G), namely exon 18 skipping (r.1627_1715del; p.Pro543Hisfs*26) and use of a cryptic-donor (r.1715_1716ins[1715+1_1715+258]; p.Ser572Argfs*73). Use of a forward primer upstream of the missense c.1307A>G variant (Figure 2F and H) and a reverse primer annealing to the exon 18/19 splice junction enabled specific amplification of transcripts with normal splicing. This PCR data established that all correctly spliced *FANCA* transcripts (exons 13-18) bear the paternal missense variant c.1307A>G. These data infer that c.1715+3_1715+13del induces (near) complete mis-splicing of all detected *FANCA* transcripts with both mis-splicing outcomes encoding a premature termination codon, enabling variant reclassification as likely pathogenic.

Cycloheximide treatment recommended as a second investigation

We explored the diagnostic utility conferred by an additional RNA sample preparation step of cycloheximide (CHX) inhibition of NMD for 25 cases (Figure 3A-F, Supplemental Figure 7). Of 23 cases, 15 cases were subsequently shown to produce at least 1 NMD-compliant outcome[9] that is an encoded premature termination codon >50 nt upstream of the last exon-exon junction (see Supplemental Table 6, CHX sensitivity). CHX rescue of NMD-compliant events was evident with primers in flanking exons for 10 of 15 cases (eg, case A054-*UBE3A* in Figure 3A-C). However, CHX effects were not readily apparent for 5 of 15 cases using flanking primers that amplify multiple events, with rescue by CHX evident only when using a primer pair specific for the NMD-targeted event (eg, case A064-*GLI3* in Figure 3D-F) likely because of competition inherent with multitemplate PCRs. CHX treatment strengthened evidence in several cases by showing mis-splicing was not a rare event and rather that NMD was effective.

Because cycloheximide treatment doubles the costs and time required for RNA testing, >90% of CVC respondents (*n* = 18) agreed that CHX treatment should be used as a second investigation for cases in which there is clear diagnostic utility for protecting an NMD-compliant mis-splicing outcome. The educational needs of genetic pathologists and diagnostic scientists were reflected by our respondents who declared that they were not aware (45%) or only somewhat aware (30%) that (1) only spliced transcripts successfully transported out of the nucleus for a pilot round of translation in the cytoplasm can activate NMD and (2) there are innate protective mechanisms to prevent mis-spliced mRNA with atypical features (eg, retained intron) from exiting the nucleus. Thus, a proportion of mis-spliced transcripts are retained in the nucleus and are incapable of activating NMD but are also unable to be translated.[31]

Importance of strategic consideration of abnormal initiation or termination of transcription

Although transcription by RNA polymerase and pre-mRNA splicing by the spliceosome are separate processes, there is complex interplay between these 2 key events underpinning gene regulation. Variants affecting promoter regions or untranslated regions or splice-site motifs of intron 1 or a terminal intron can lead to abnormal initiation or termination of transcription.[32] Our study identified 4 cases with an intron 1 splice-site variant shown to induce abnormal initiation of transcription (A001-*CLN5*, A088-*PRPH2*, A094-*NF1*, A100-*TUBA1A*) and 1 case with pathogenic abnormal termination of transcription (A113-*KCNH2*). Because of the complexity of the mechanisms involved, deeper mechanistic investigations of cases A001-*CLN5*, A100-*TUBA1A*, and A113-*KCNH2* will be submitted for publication separately. However, evidence for the activation of an alternative transcription start site for variants affecting the donor splice site of intron 1 in cases A088-*PRPH2* and A094-*NF1* are shown in Figure 3G-L. In both cases, PCR encompassing a distal benign single-nucleotide variation
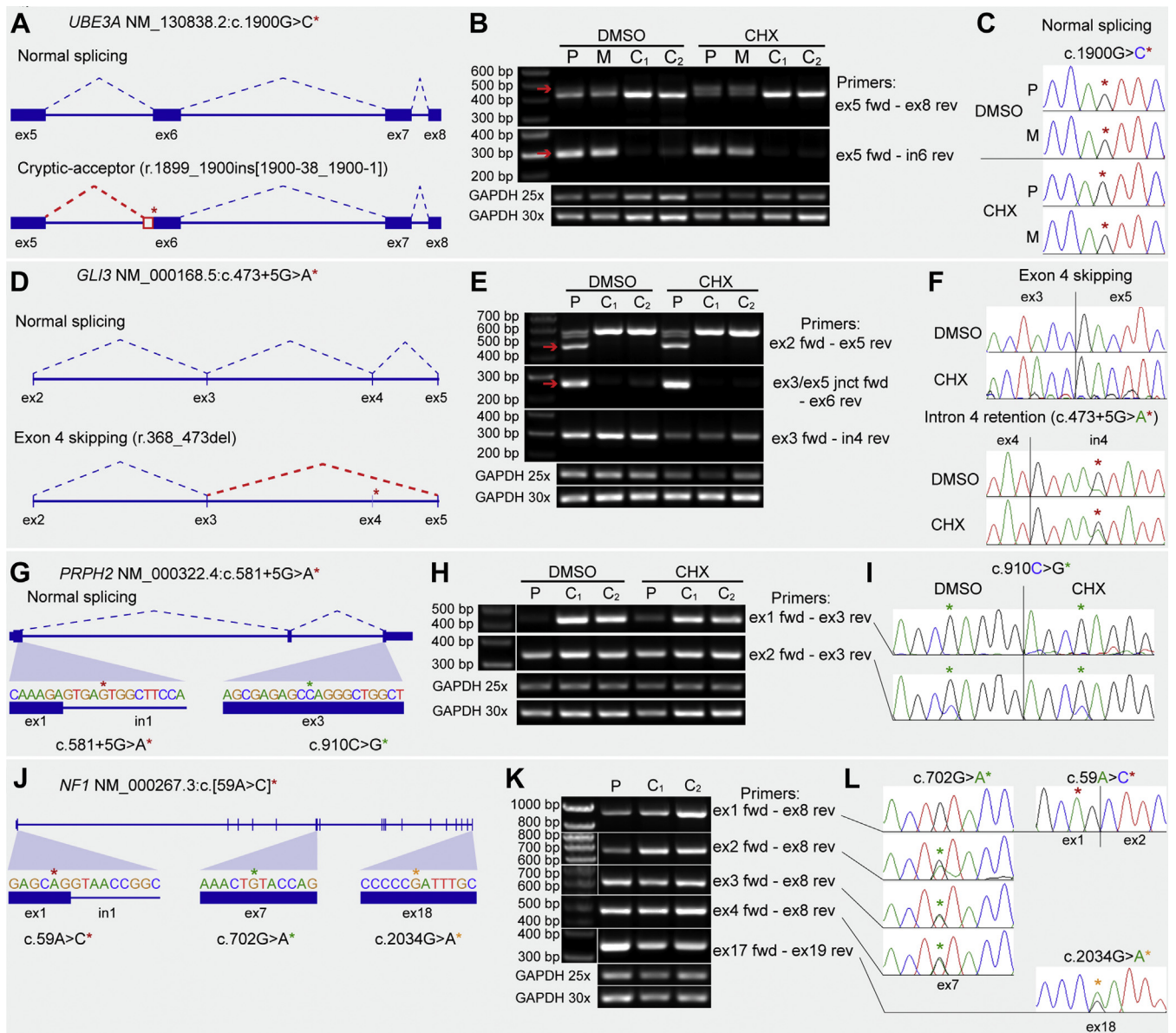
Figure 3   Diagnostic utility of nonsense-mediated decay (NMD) inhibition (A-F): intron 1 variants causing abnormal initiation of transcription (G-O). Isolated peripheral blood mononuclear cells and fibroblasts were treated with CHX or DMSO before RNA extraction. A. Schematic of detected mis-splicing for case A054-*UBE3A*. Maternal NM_130838.2:c.1900G>C variant (red asterisk) in a male proband with Angelman syndrome (MIM#105830) led to use of a cryptic acceptor (red splice junction and arrow). B. CHX treatment increased relative abundance of the higher band (top red arrow) corresponding to use of the cryptic acceptor, inducing a frameshift. Proband (P: male, 7 years); mother (M: female, 31 years); controls (C1: male, 5 years, C2: male, 39 years). C. All detected *UBE3A* transcripts with normal splicing arise from the c.1900G paternal allele. D. Schematic of detected mis-splicing for case A064-*GLI3*. The NM_000168.5:c.473+5G>A variant (red asterisk) identified in a male proband with polydactyly (MIM#174700) led to exon 4 skipping (red splice junction and arrow), inducing a frameshift. E, F. There was little difference in the intensity of amplicons with and without CHX treatment using flanking primers in exon 2 and exon 5. Proband (P: male, 28 years); controls (C1: male, 40 years, C2: male, 48 years). G. A088-*PRPH2* is a female proband with macular dystrophy (MIM#169150) associated with NM_000322.4:c.581+5G>A (red asterisk). H, I. Sanger sequencing trace files of an exon 1-2 amplicon (not shown) or exon 1-3 cDNA amplicon shows apparent hemizygosity of benign variant NM_000322.4:c.910C>G (green asterisk) in trans, whereas an exon 2 to 3 amplicon shows heterozygosity of c.910C>G variant (green asterisk). Acknowledging that PCR is not quantitative, peak height of the c.910C allele in cis with c.581+5G>A was consistently lower—suggestive either of NMD degradation of transcripts from this allele and/or inefficient transcription initiation at an ectopic start site. Proband (P: female, 64 years); controls (C1: female, 43 years, C2: female, 71 years). J. A094-*NF1* is a male proband with neurofibromatosis (MIM#162200) associated with a missense variant in exon 1 NM_000267.3:c.59A>C (red asterisk). K, L. Sanger sequencing showed absence of the c.59A>C allele in amplicons spanning exons 1-8 (red asterisk) (or exons 1-3 or exons 1-4, not shown). Concordantly, NM_000267.3:c.702G>A in cis with c.59A>C was absent in an exon 1-8 amplicon (green asterisk) although it appeared heterozygous in amplicons spanning exons 2 to 8, 3 to 8, or 4 to 8 (green asterisks). Evidence therefore suggests that *NF1* transcription initiates after exon 1. Proband (P: male, 34 years); controls (C1: male 31 years, C2: male 35 years). CHX, cycloheximide; DMSO, dimethyl sulfoxide.
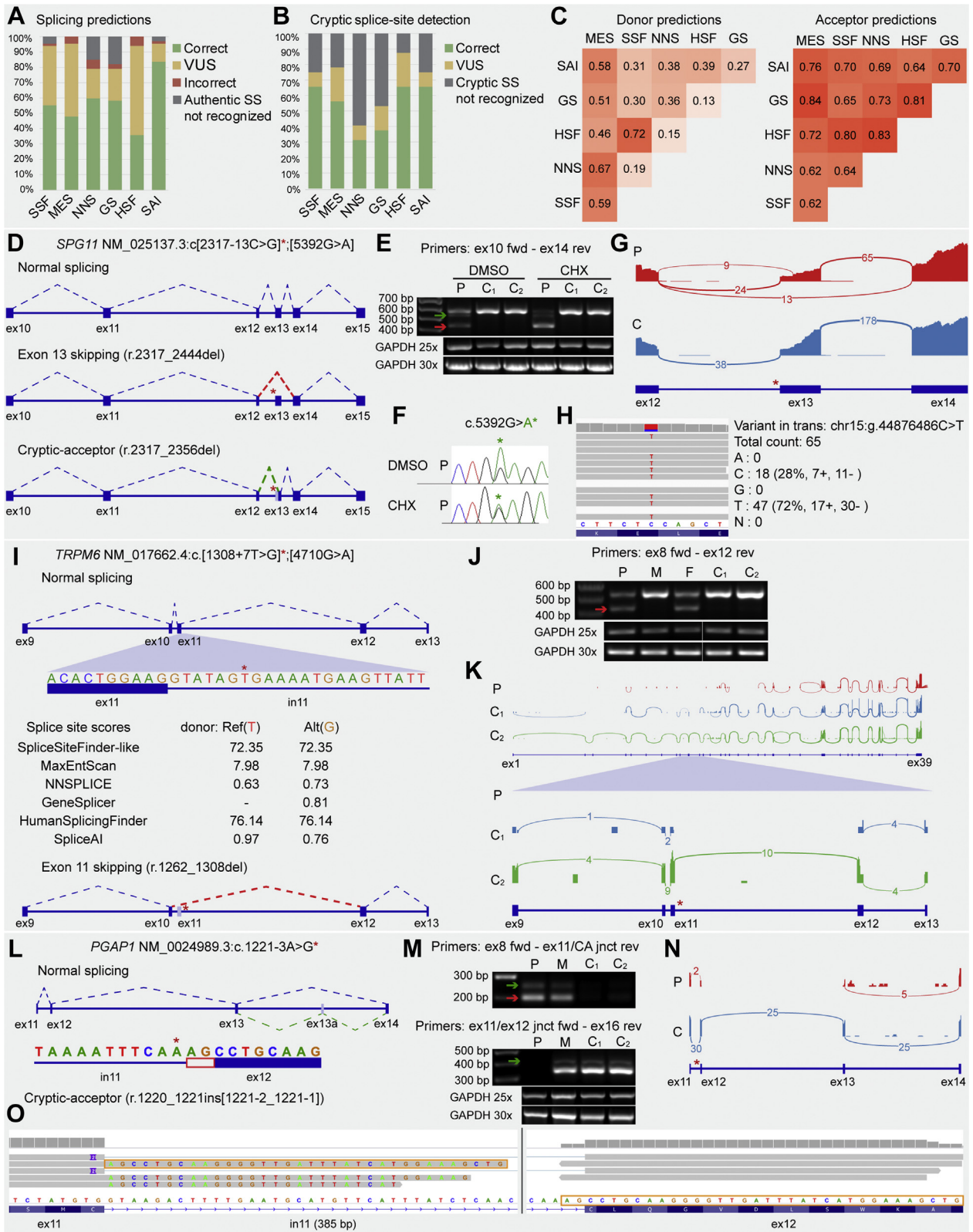
Figure 4   **Utility of in silico predictive algorithms and comparative diagnostic utility of RNA sequencing (RNA-seq).** A, B. Histograms showing the predictive accuracy of splicing prediction algorithms. A. Author-defined confidence thresholds (see Materials and Methods) were used to assign predictions for normal splicing or mis-splicing, or VUS if predictive scores fell outside of confidence thresholds. Green = correct prediction. Red = incorrect prediction. Yellow = VUS. Gray = could not identify the authentic splice site. B. Detection of cryptic splice sites. Gray = cryptic splice site used by the spliceosome not recognized by the algorithm. Green = cryptic splice site score higher than the resultant variant splice site. Yellow = VUS; cryptic splice-site score lower than resultant variant splice site. C. Concordance matrix of splicing prediction algorithms for donor splice-site variants (left) and acceptor splice-site variants (right). D-O. Diagnostic utility of reverse transcription polymerase chain reaction (RT-PCR) vs RNA-seq. D-F. Overview showing detected mis-splicing

(SNV) was crucial to show loss of correctly spliced transcripts containing exon 1 (containing the start AUG). This was achieved by showing that a heterozygous coding SNV appears hemizygous (if in trans) or absent (if in cis) by Sanger sequencing of amplicons derived using an exon 1 forward primer, but heterozygous using a forward primer in exon 2, 3, 4, and so on. These data infer abnormal initiation of transcription downstream of exon 1 and upstream of exon 2. Both *PRPH2* and *NF1* have the start AUG encoded by exon 1; when transcription does not initiate at exon 1, it is impossible to predict the translational start site and any AUG codon within a reasonable Kozak sequence may be used,[33] in frame or out of frame.

Assessment of the reliability of splicing prediction algorithms

We retrospectively evaluated the predictive accuracy and concordance of splicing algorithms within Alamut Visual Biosoftware[34-38] and the deep-learning algorithm SpliceAI[23] (Figure 4A-C). We used thresholds defined by the algorithm's authors to assign predictions of normal splicing or mis-splicing or, if predictive scores fell within an author-defined "grey zone" (outside of confidence thresholds for normal splicing or mis-splicing), we assigned VUS (Figure 4A, yellow segment). SpliceAI was the most accurate (84%) followed by NNSplice (60%). NNSplice and GeneSplicer were often unable to recognize the authentic human splice site to offer a prediction of 15% (NNSplice) or 18% (GeneSplicer) variants (Figure 4A, gray segment). A majority of this difficult cohort of splicing variants were "VUS" (yellow segment), reaching 58% VUS for Human Splicing Finder and 48% VUS for MaxEntScan. Many activated cryptic splice sites were not recognized by the prediction algorithms (Figure 4B, gray segment). Overall, there was significant discordance and inaccuracy in the predicted outcomes among algorithms, especially for donor splice-site variants (Figure 4C), highlighting the significant challenge with clinical interpretation of in silico splicing predictive tools.

Comparative evaluation of diagnostic utility of RNA-seq

RNA-seq (150 bp paired-end reads) was performed subsequently for 19 cases studied by RT-PCR. Diagnostically informative RNA-seq data were obtained for 40% (6/19) of the cases, identifying all splicing events detected by RT-PCR and detectable allele bias of coding SNVs reflecting active NMD (eg, case A014-*SPG11* in Figure 4D-H). RNA-seq was nondiagnostic for 60% (12/19) of cases because of low read depth (eg, case A089-*TRPM6* in Figure 4I-K) and exacerbated by NMD (eg, case A031-*PGAP1* in Figure 4L-O and case A066-*VPS13D* in Supplemental Figure 8). Notably, 4 samples failed library preparation primarily because of low RNA concentration (RT-PCR informed diagnoses were secured before sending residual RNA for RNA-seq). Retrospective analyses indicate a read depth of approximately 5 TPM is required for diagnostically informative RNA-seq, whereas genes in this cohort with TPM values >0.5 could be reliably studied by RT-PCR (35 cycles) (Supplemental Figure 9). Extrapolating these thresholds to our entire cohort infers 42% had a read depth too low for diagnostically informative results by transcriptomic RNA-seq (50M paired-end reads).

Discussion

Our study demonstrates the significant diagnostic and health benefits of RNA diagnostics as adjunct testing to extend diagnostic yield from genomic testing. Although blood, skin fibroblasts, and EBV-LCLs are being used widely for RNA studies,[4-6,10] we demonstrate the unrealized diagnostic utility of urothelial cells. Urine collection is an attractive biospecimen because of ease of sampling, particularly for

---

for case A014-*SPG11*. The NM_025137:c.2317-13C>G variant (red asterisk) identified in a female proband with spastic paraplegia (MIM#604360) was shown by RT-PCR to cause exon 13 skipping (red splice junction and arrow) and use of a cryptic acceptor (green splice junction and arrow). Proband (P: female 43); controls (C1: female, 36 years; C2: female, 37 years). G, H. RNA-seq (CHX untreated sample) confidently identifies exon 13 skipping and cryptic-acceptor use, as well as allele bias over the NM_025137.3:c.5392G>A missense variant (green asterisk) in trans. I, J. Overview showing detected mis-splicing for case A089-*TRPM6*, a male with hypomagnesaemia, seizures, and developmental delay (MIM#602014). The NM_017662.4:c.1308+7T>G variant (red asterisk), inherited in trans with NM_017662.4:c.4710G>A (green asterisk), was shown by RT-PCR to induce exon 11 skipping (red splice junction and arrow). Proband (P: male, 8 months); mother (M: female, 24 years); father (F: male, 31 years); controls (C1: male, 2 years; C2: male, 3 years). Splicing algorithms predict negligible impact of c.1308+7T>G variant. K. RNA-seq sashimi plots of the entire *TRPM6* gene showing 3′ bias because of polyA capture and/or 5′ decay of transcripts with zoom-up showing absence of reads mapping to the exon skipping event. L, M. Overview showing detected mis-splicing for A031-*PGAP1*, a male with mental retardation (MIM#615802). RT-PCR showed the homozygous NM_024989.3:c.1221-3A>G variant (red asterisk) caused use of a cryptic acceptor (red box and arrow). N, O. RNA-seq showed low read depth for *PGAP1* relative to disease controls analyzed in the same run and failed to correctly align reads corresponding to use of the cryptic acceptor with soft clipped reads (orange box) revealed to highlight the alignment error. CA, cryptic-acceptor; CHX, cycloheximide; GS, GeneSplicer; HSF, Human Splicing Finder; MES, MaxEntScan; NNS, NNSplice; SAI, SpliceAI; SSF, SpliceSiteFinder-like; VUS, variants of unknown significance.

young children. Importantly, urothelial cells express a different repertoire of genes than skin fibroblasts or blood cells, increasing the breadth of genes able to be studied via RNA diagnostics.

Health economics analyses performed for A058-*ASNS* demonstrated that an early diagnosis, enabled through rapid RNA diagnostics, reduced hospitalization costs by A\$117,800.[7] Although not yet measured formally, there are significant cost benefits of this study related to additional diagnoses of 9 similarly affected family members and facilitation of preventative medicine. Of note, 41% of diagnoses were used for prenatal counseling with half of those cases intending to use the molecular diagnosis for preimplantation genetic diagnosis or screening, reducing both the significant emotional anxiety for parents related to recurrence risk and the lifetime financial cost of health services caring for children with severe genetic disorders.

Approximately 39% of our CVC respondents (*n* = 18) agree on a requirement for pre-mRNA testing to be performed within an accredited laboratory, whereas 39% of respondents were satisfied with clinical actioning of splicing data from a reputable research laboratory with expertise in splicing studies (with an appropriate ethical and governance framework). The vast majority of SpliceACORD members support a hybrid model of research-pathology laboratory collaboration during this transitional period of RNA diagnostics. All CVC respondents endorsed the following consensus:

(1) An accredited regulatory framework is favored although it does not yet exist.
(2) PCR studies of cDNA are necessarily bespoke. Although protocols used for confirmatory Sanger sequencing of gDNA amplicons are readily adaptable to cDNA, it would be inefficient to perform multiple PCRs in this manner; however, it could be conducted as an accredited test to confirm 1 or more key (mis) splicing outcomes.
(3) The most important factor is the expertise of the testing center in the complexities of pre-mRNA splicing and rigor of the scientific methods used.
(4) There is an urgent need to establish quality standards for RNA diagnostics and ACMG-AMP–aligned interpretation rubrics for complex mRNA assay data (in frame, out of frame, multiple events, abnormal initiation, or termination of transcription).

Therefore, informed by study outcomes, Figure 5 details SpliceACORD consensus recommendations for ACMG-AMP–aligned interpretation of mRNA assay data for variant classification. We consider the PS3/BS3 criterion most appropriate for pre-mRNA splicing assay data using RNA isolated from patient biospecimens. We define recommendations for very strong, strong, and moderate evidence levels—to provide, when appropriate, the same level of evidence afforded by PVS1.[39] We emphasize that

PS3 should not be used in combination with the PVS1 predicted loss-of-function criterion to avoid double counting of evidence. We recommend application of BS3 with robust evidence supporting maintenance of a normal splicing pattern (Figure 5) only when an effect on transcription or pre-mRNA splicing is the sole concerning possible effect of a variant. Protein biochemistry assays must be used to accurately determine the consequences of a synonymous or noncoding variant upon translation. Overall, complex RNA assay data do not retrofit well into existing ACMG-AMP guidelines. Adaptation of the functional evidence criterion to a points system may provide a more viable solution in the longer term to enable collective weighting of functional genomics evidence derived from patient specimens or in vitro assays (eg, RNA, protein, epigenetics).

RT-PCR and RNA-seq have different strengths and weaknesses. Although high sensitivity of PCR is a diagnostic strength, a strong caveat is reliance upon the expertise of the testing center and their strategic positioning of primers—you detect only what your primers are capable of amplifying under the PCR conditions used. RNA-seq is the test of choice for cases with multiple putative splicing variants or for exome-negative cases with strong phenotypic concordance to known associated genes, if there is sufficient read depth for diagnostic confidence. Our experience with short-read RNA-seq highlights several caveats of diagnostic importance, namely (1) mis-spliced reads are regularly mis-aligned or filtered out (eg, A031-*PGAP1*, A058-*ASNS*[7]); (2) short reads are regularly mis-mapped between homologous gene paralogues in contig (eg, hemoglobin, tubulin, myosin heavy chain gene clusters); (3) pathogenic mis-splicing events confirmed by RT-PCR can be missed by RNA-seq because of low read depth as a result of low gene expression in the available biospecimen, NMD, 3′ bias because of polyA selection, or natural RNA decay of the 5′ end of long transcripts (A089-*TRPM6*; A066-*VPS13D*); (4) reads are regularly too short to encompass a heterozygous SNV to phase mis-splicing events and discern complete from partial mis-splicing. SpliceACORD's recommendation therefore is to perform RT-PCR of cDNA to confirm pathogenic mis-splicing detected by RNA-seq until we have greater depth of experience to establish standard operating procedures that minimize the risk of false negatives and false positives.

In the future, clinical RNA diagnostics for Mendelian conditions is likely to require multiple technical platforms tailored to the genomic context and expression levels of the target gene in available specimens, including genome-wide transcriptomics, targeted transcriptomics of genes or gene panels, and RT-PCR for genes with low expression in available specimens. Service delivery of RNA diagnostics must consider the requirement for a biobank and ethical framework of informed consent for diagnostic use of previously tested age- and gender-matched biospecimens as controls for prospective cases. In conclusion, SpliceACORD leverages the collective expertise of approximately 80

	PS3_ Splice Altering_Very strong	PS3_Splice Altering_Strong	PS3_Splice Altering_Moderate
Functional evidence for mis-splicing	**1.** Very strong evidence for variant-associated (near) **complete mis-splicing**, with evidence confirming undetectable levels of normal splicing arising from the splicing variant allele.	**1.** Strong evidence for variant-associated **mis-splicing of > 95 % of transcripts** arising from the variant allele, compared with < 5% mis-splicing from the reference allele in controls.	**1.** Strong evidence for variant-associated mis-splicing, with or without evidence for remnant normal splicing arising from the splicing variant allele (**complete mis-splicing or partial mis-splicing**).

2. Technical approach controls for:
a) **mRNA quality and quantity;**
b) comparatively assesses at least two, and preferably three, **age-, gender- and tissue-matched control mRNA** specimens in the same experiment;
c) **interrogates for all plausible consequences for pre-mRNA splicing;** i) normal splicing, ii) exon-skipping, iii) intron retention, iv) cryptic splice-site activation, v) consideration of possible abnormal initiation of transcription for variants affecting the promoter region, 5'untranslated region or consensus splicing motifs of intron-1 for any transcript, vi) consideration of possible abnormal termination of transcription for variants affecting the consensus splicing motifs of the last intron of any transcript, the 3'untranslated region or polyadenylation signal; and
d) demonstrated reproducibility of detected variant-associated (mis)splicing events.

3. RT-PCR technical requirements: strength of experimental evidence to confidently exclude likelihood of Type I (failed detection of normal splicing) or Type II error (failed detection of mis-splicing).
NOTES: 1) RNA/cDNA method is not selective only for polyadenylated mRNA.

	a) hemizygous or homozygous variant: Two independent approaches confirm **undetectable levels of normal splicing** from the splicing variant allele(s), relative to the reference allele in controls. b) heterozygous variant (confounded by normal splicing from the allele in trans): Evidence for **undetectable levels of normal splicing** from the splicing variant allele. Phasing of normal splicing may be achieved using a PCR primer bridging the correctly spliced exon junction under scrutiny, paired with a second primer that includes a segregated heterozygous coding SNV. (Near) Complete mis-splicing is evidenced by reproducible, apparent hemizygosity of the SNV if in trans or absence if in cis.	a) hemizygous or homozygous variant: Two independent approaches confirm **negligible normal splicing** from the splicing variant allele(s), relative to the reference allele in controls. b) heterozygous variant (confounded by normal splicing from the allele in trans). Evidence for **negligible levels of normal splicing** from the splicing variant allele. Phasing of normal splicing may be achieved using a PCR primer bridging the correctly spliced exon junction under scrutiny, paired with a second primer that includes a segregated heterozygous coding SNV. Variant-associated mis-splicing of the vast majority of transcripts is evidenced by near hemizygosity (> 95%) of the SNV if in trans or near absence (< 5%) if in cis.	a) hemizygous or homozygous variant: two independent approaches confirm a **clinically meaningful reduction** of correctly spliced mRNA arising from the splicing variant allele(s), relative to the reference allele in controls. b) heterozygous variant (confounded by normal splicing from the allele in trans). i) A heterozygous coding SNV is available that establishes a **clinically meaningful reduction** of correctly spliced mRNA arising from the splicing variant allele. ii) A heterozygous coding SNV is not available to phase from which allele normal splicing arises. However, alternative evidence clearly demonstrates variant-associated mis-splicing resulting in a **clinically meaningful reduction** of correctly spliced mRNA arising from the splicing variant allele. c) For any inheritance pattern with normal splicing and NMD-compliant mis-splicing, where a heterozygous coding SNV is not available to phase events, perform repeat testing plus/minus cycloheximide to improve determination of relative levels of variant-associated mis-splicing.

4. RNA-seq technical requirements: Strength of evidence based on read depth and available SNVs (to inform allele bias) to confidently exclude likelihood of Type I (failed detection of normal splicing) or Type II error (failed detection of mis-splicing).
NOTES: 1) RNA/cDNA method is not selective only for polyadenylated mRNA. 2) PCR and Sanger sequencing of cDNA amplicons spanning multiple exons (when possible) to confirm pathogenic mis-splicing of the disease-relevant transcript(s).

	a) homozygous or hemizygous variant: at least 15 reads confirming pathogenic mis-splicing, with **undetectable** levels of normal splicing. b) heterozygous variant (confounded by normal splicing from the allele in trans): at least 15 reads confirming pathogenic mis-splicing. See **3b** for diagnostic requirement of an informative SNV to confirm **undetectable levels of normal splicing** arising from the splicing variant allele by RNA-seq or RT-PCR of cDNA.	a) homozygous or hemizygous variant: at least 15 reads confirming mis-splicing in > **95%** of transcripts arising from the splicing variant allele, compared with **< 5%** mis-splicing in controls. b) heterozygous variant (confounded by normal splicing from the allele in trans): at least 15 reads confirming pathogenic mis-splicing. See **3b** for diagnostic requirement of an informative SNV to confirm **negligible levels of normal splicing** arising from the splicing variant allele by RNA-seq or RT-PCR of cDNA	a) homozygous or hemizygous variant: at least 15 reads confirming a **clinically meaningful level** of variant-associated mis-splicing, compared with controls. b) heterozygous variant (confounded by normal splicing from the allele in trans): at least 15 reads confirming a **clinically meaningful level** of variant-associated mis-splicing of the relevant disease-associated transcript(s), compared with controls. i) A heterozygous coding SNV is available that establishes a **clinically meaningful reduction** of correctly spliced mRNA arising from the splicing variant allele. ii) A heterozygous coding SNV is not available to phase from which allele normal splicing arises. However, alternative evidence clearly demonstrates variant-associated mis-splicing resulting in a **clinically meaningful reduction** of correctly spliced mRNA. c) For any inheritance pattern with normal splicing and NMD-compliant mis-splicing, where a heterozygous coding SNV is not available to phase events, perform repeat testing plus/minus cycloheximide to improve determination of relative levels of variant-associated mis-splicing.

Transcript disease assoc.	**5.** Variant affects splicing of one or more constitutive exon(s) present in the predominant isoform(s) expressed by the manifesting tissue(s); **OR** Variant activates inclusion of ectopic/intronic sequences into the predominant isoform(s) expressed by the manifesting tissue(s); **OR**		
	6. Variant affects an alternatively spliced exon but detected mis-splicing event(s) involve a region of the gene with **multiple variants classified likely/pathogenic** establishing its functional and clinical importance	**6.** Variant affects an alternatively spliced exon but detected mis-splicing event(s) involve a region of the gene with **one or more variants classified likely/pathogenic** establishing its functional or clinical importance	**6.** Variant affects an alternatively spliced exon but detected mis-splicing event(s) involve a **conserved region of the gene** consistent with functional and/or structural importance.

Pathogenetic mechanism	**7.** All detected variant-associated abnormal splicing (or abnormal transcription initiation or termination) results in loss-of-function (LoF) consequences for the encoded protein and LoF variants are a known causal basis for disease; **OR**		
	8. All detected variant-associated abnormal splicing (or abnormal transcription initiation or termination) is in-frame and missense variants and in-frame indels are a known causal basis for disease; **OR**		
	9. Multiple abnormal splicing (or abnormal transcription initiation or termination) events, both in-frame and out-of-frame, and LoF, missense variants or in-frame indels are a known causal basis for disease; **AND;**		
	10 In-frame mis-splicing altering length of the encoded protein affects a variant hotspot or domain that contains **multiple variants classified likely/pathogenic** establishing its functional and clinical relevance	**10.** In-frame mis-splicing altering length of the encoded protein affects a region of the gene with **one or more variants classified likely/pathogenic** establishing its functional or clinical relevance, or there is strong evidence from **evolutionary conservation and/or data from model systems** supporting deleterious consequences for protein function.	**10.** In-frame mis-splicing altering length of the encoded protein is **considered** to exert a deleterious effect on protein function based on presence of causal variants, evolutionary conservation, or data from model systems.

	BS3_Splice Neutral_Strong	BS3_Splice Neutral_Moderate
Functional evidence for normal splicing	**1.** Clear evidence for maintenance of a normal splicing pattern of disease-relevant transcript(s) in the affected individual versus age-, gender- and tissue-matched mRNA a) no evidence for increased levels of natural mis-splicing events b) experimental evidence showing similar levels of correctly spliced mRNA in the proband versus controls c) heterozygous variants: **direct evidence from two or more coding SNVs** showing absence of allele bias and confirming roughly equal levels of correctly spliced mRNA arising from both alleles.	**1.** Clear evidence for maintenance of a normal splicing pattern of disease-relevant transcript(s) in the affected individual versus age-, gender- and tissue-matched mRNA a) no evidence for increased levels of natural mis-splicing events b) experimental evidence showing similar levels of correctly spliced mRNA in the proband versus controls c) Heterozygous variants: **direct evidence from at least two coding SNVs** showing absence of allele bias and confirming roughly equal levels of correctly spliced mRNA arising from both alleles.
	2. Technical approach controls for mRNA quality and quantity, comparatively assesses at least two and preferably three age-, gender- and tissue-matched control mRNA specimens, interrogates for all plausible mis-splicing outcomes and demonstrates reproducibility of undetectable levels of variant-associated mis-splicing a) exon-skipping, intron retention, cryptic splice activation b) promoter region or first exon/intron variant: abnormal initiation of transcription c) last exon/intron variant: abnormal termination or polyadenylation of mRNA	**2.** Technical approach controls for mRNA quality and quantity, comparatively assesses at least two and preferably three age-, gender- and tissue-matched control mRNA specimens, interrogates for all plausible mis-splicing outcomes and demonstrates reproducibility of undetectable levels of variant-associated mis-splicing a) exon-skipping, intron retention, cryptic splice activation b) promoter region or first exon/intron variant: abnormal initiation of transcription c) last exon/intron variant: abnormal termination or polyadenylation of mRNA
	3. Technical approach uses an RNA/cDNA preparation method that is not selective only for polyadenylated mRNA.	**3.** Technical approach uses an RNA/cDNA preparation method that is not selective only for polyadenylated mRNA.
	4. Technical approach deploys **cycloheximide-mediated inhibition of nonsense-mediated decay (NMD)**.	
	5. For RNA-seq evidence, ensure soft-clipped reads are revealed, there are >25 reads for adjacent exon-exon junctions, and similar junctional reads for the exon junction under scrutiny as compared with splice junction(s) immediately upstream and downstream. 25 reads is the minimum read depth to avoid Type II error and failed detection of a mis-splicing event being targeted with 90% efficiency by NMD (3 reads). If <25 reads, confirm normal splicing by RT-PCR of cDNA or repeat RNA-seq with higher read depth interrogating for all plausible mis-splicing events.	**4.** For RNA-seq evidence, ensure soft-clipped reads are revealed, there are >25 reads for adjacent exon-exon junctions, and similar junctional reads for the exon junction under scrutiny as compared with splice junction(s) immediately upstream and downstream. 25 reads is the minimum read depth to avoid type II error and failed detection of a mis-splicing event being targeted with 90% efficiency by NMD (3 reads). If <25 reads, confirm normal splicing by RT-PCR of cDNA or repeat RNA-seq with higher read depth interrogating for all plausible mis-splicing events.
Codon usage and translation	**6.** Synonymous variant does not result in substitution of a rare codon for a common codon, or conversely, a common codon for a rare codon for a gene product known to have dose-dependency. **7.** Non-coding or synonymous variant does not affect the KOZAK sequence of the initiation AUG. NOTES: 1) Apply BS3_Splice Neutral only when an effect on transcription or pre-mRNA splicing is considered the only concerning possible effect of the variant. 2) Protein biochemistry assays must be used to accurately determine consequences of a synonymous or non-coding variant upon translation.	**5.** Synonymous variant does not result in substitution of a rare codon for a common codon, or conversely, a common codon for a rare codon for a gene product known to have dose-dependency. **6.** Non-coding or synonymous variant does not affect the KOZAK sequence of the initiation AUG. NOTES: 1) Apply BS3_Splice Neutral only when an effect on transcription or pre-mRNA splicing is considered the only concerning possible effect of the variant. 2) Protein biochemistry assays must be used to accurately determine consequences of a synonymous or non-coding variant upon translation.

Figure 5   SpliceACORD recommendations for interpretation of RNA functional testing data aligning with ACMG/AMP evidence criteria of PS3 and BS3. PS3 (experimental evidence of mis-splicing outcomes) may not be used together with PVS1 (null variant). BS3 can be applied with robust evidence supporting maintenance of a normal splicing pattern only when an effect on transcription or pre-mRNA splicing is the sole, concerning possible effect of a variant. Protein biochemistry assays must be used to establish the potential impact of a synonymous or noncoding variant upon translation. We concur that a testing laboratory should establish the reproducibility and reliability of their RNA testing assays for at least 11 validation controls including a mix of positive and negative controls.[40] However, it is not possible in the RNA Diagnostics standard operating procedures we propose herein to include a mix of 11 benign and pathogenic variant controls, especially for novel variants, within a single experiment that deploys multiple PCRs to interrogate for each conceivable mis-splicing event. Therefore, our recommendation for each individual tested is comparative analysis with 2 to 4 age-, gender- and specimen-matched controls or disease controls (ie, individuals with a genetic variant in a different gene in an unrelated pathway) and wherever possible, testing of multiple affected individuals or heterozygotes to confirm reproducibility of variant-associated (mis)splicing outcomes. cDNA, complementary DNA; NMD, nonsense-mediated decay; PCR, polymerase chain reaction; pre-mRNA, premessenger RNA; RNA-seq, RNA sequencing; RT-PCR, reverse transcription polymerase chain reaction; SNV, single-nucleotide variation.

multidisciplinary members with diverse expertise across all stages of a patient's journey through genomic diagnosis to propose recommended triage criteria, standard operating procedures, and interpretation rubrics for PCR-based RNA diagnostics using clinically accessible specimens.

Data Availability

Access to data not provided herein may be requested via the corresponding author.

Author Information

Conceptualization: A.M.B., L.G.R., S.T.C.; Data Curation and Formal Analysis: A.M.B., R.D., H.J.; Investigation: all authors; Resources: all authors; Validation: all authors; Visualization: A.M.B., L.G.R., S.T.C.; Funding Acquisition: S.T.C.; Methodology: A.M.B., L.G.R., S.B., A.N.; Project Administration: S.T.C.; Writing-original draft: A.M.B., L.G.R., S.T.C.; Writing-review and editing: all authors; Supervision: K.J.J., B.B., S.T.C.

Ethics Declaration

Consent for diagnostic genomic testing was supported by governance infrastructure of the relevant local ethics committees of the participating Australian Public Health Local Area Health Districts. Kids Neuroscience Centre's biobanking and functional genomics human ethics protocol was approved by the Sydney Children's Hospitals Network Human Research Ethics Committee (protocol 10/CHW/45 renewed with protocol 2019/ETH11736 [July 2019-2024]) with informed, written consent for all participants.

Conflict of Interest

Sandra T. Cooper is director of Frontier Genomics Pty Ltd (Australia). Sandra T. Cooper currently receives no consultancy fees or other remuneration for this role. Frontier Genomics Pty Ltd (Australia) has no existing financial relationships that will benefit from publication of these data. Samuel P. Strom is an employee and shareholder of Fulgent Genetics. Michael F. Buckley is an employee and shareholder of Invitae. The remaining coauthors declare no conflicts of interest.

Australasian Consortium for RNA Diagnostics (SpliceACORD)

Ghusoon Abdulrasool, Lauren S. Akesson, Ghamdan Al Eryani, Mohammad Al-Shinnag, Peer Arts, Richard Bagnall, Naomi L. Baker, Christopher Barnett, Sarah Beecroft, Bruce Bennetts, Marina Berbic, Victoria Beshay, Michael Black, Jim Blackburn, Piers Blombery, Kirsten Boggs, Adam M. Bournazos, Susan Branford, Jimmy Breen, Natasha J. Brown, Samantha J. Bryen, Leslie Burnett, Daffodil Canson, Pak Cheong, Edward Chew, Belinda Chong, John Christodoulou, Seo-Kyung Chung, Mike Clark, Corrina Cliffe, Melissa Cole, Felicity Collins, Alison Compton, Antony Cooper, Sandra T. Cooper, Mark Corbett, Mark Cowley, Mark R. Davis, Martin Delatycki, Tracy Dudding, Matthew Edwards, Stefanie Eggers, Lisa J. Ewans, Eduardo Eyras, Fathimath Faiz, Miriam Fanjul Fernandez, Andrew Fellowes, Andrew Fennell, Michael

Field, Ron Fleischer, Chiara Folland, Lucy Fox, Mary-Louise Freckmann, Clara Gaff, Melanie Galea, Roula Ghaoui, Himanshu Goel, Ilias Gornanitis, Thuong Ha, Bernadette Hanna, James Harraway, Rippei Hayashi, Ian Hayes, Alex Henderson, Luke Hesson, Erin Heyer, Michael Hildebrand, Michael Hipwell, Gladys Ho, Ari E. Horton, Cass Hoskins, Matthew F. Hunter, Matilda Jackson, Paul James, Kristi J. Jones, Justin Jong-Leong Wong, Sarah Josephi-Taylor, Himanshu Joshi, Karin Kassahn, Peter Kaub, Lucy Kevin, Edwin Kirk, Emma Krzesinski, Smitha Kumble, Sarah Kummerfeld, Nigel Laing, Chiyan Lau, Eric Lee, Sarah Leighton, Ben Lundie, Sebastian Lunke, Amali Mallawaarachchi, Chelsea Mayoh, Julie McGaughran, Alison McLean, Mary McPhillips, Cliff Meldrum, Edwina Middleton, Di Milnes, Kym Mina, David Mowat, Amy Nisselle, Emily Oates, Alicia Oshlack, Elizabeth E. Palmer, Gayathri Parasivam, Michael Parsons, Chirag Patel, Jason R. Pinner, Michael Quinn, John Rasko, Gina Ravenscroft, Anja Ravine, Krista Recsei, Matthew Regan, Jacqueline Rehn, Lisa G. Riley, Stephen Robertson, Anne Ronan, Tony Roscioli, Georgina Ryland, Simon Sadedin, Sarah A. Sandaradura, Andreas Schreiber, Hamish Scott, Rodney Scott, Christopher Semsarian, Cas Simons, Emma Singer, Janine M. Smith, Renee Smyth, Amanda Spurdle, Zornitza Stark, Patricia Sullivan, Samantha Sundercombe, Tiong Y. Tan, Michel C. Tchan, Bryony A. Thompson, David Thorburn, John Toubia, Ronald Trent, Emma Tudini, Irina Voneague, Leigh Waddell, Logan Walker, Mathew Wallis, Nick Warnock, Robert Weatheritt, Deborah White, Susan M. White, Mark G. Williams, Meredith J. Wilson, Ingrid Winship, Lisa Worgan, Dale C. Wright, Kathy Wu, Alison Yeung, Andrew Ziolowski.

Additional Information

Authors

Adam M. Bournazos[1,2], Lisa G. Riley[2,3], Shobhana Bommireddipalli[1], Lesley Ades[2,4], Lauren S. Akesson[5,6,7,8,9,10], Mohammad Al-Shinnag[11,12], Stephen I. Alexander[2,13], Alison D. Archibald[5,9], Shanti Balasubramaniam[14,15,16], Yemima Berman[17,18], Victoria Beshay[10], Kirsten Boggs[4,19,20], Jasmina Bojadzieva[21], Natasha J. Brown[5,9], Samantha J. Bryen[1,2], Michael F. Buckley[22], Belinda Chong[9], Mark R. Davis[23], Ruebena Dawes[1,2], Martin Delatycki[5,9], Liz Donaldson[24], Lilian Downie[5,24,25], Caitlin Edwards[23], Matthew Edwards[26], Amanda Engel[27], Lisa J. Ewans[28,29], Fathimath Faiz[23], Andrew Fennell[5,30], Michael Field[31], Mary-Louise Freckmann[32], Lyndon Gallacher[5,9], Russell Gear[9], Himanshu Goel[31,33],

Shuxiang Goh[34], Linda Goodwin[35], Bernadette Hanna[36], James Harraway[37], Megan Higgins[11], Gladys Ho[2,38], Bruce K. Hopper[39], Ari E. Horton[5,30,40,41], Matthew F. Hunter[30,42], Aamira J. Huq[6,24], Sarah Josephi-Taylor[4,36], Himanshu Joshi[1], Edwin Kirk[22,43], Emma Krzesinski[30,42], Kishore R. Kumar[44,45], Frances Lemckert[1,2], Richard J. Leventer[5,25,46], Suzanna E. Lindsey-Temple[34,47], Sebastian Lunke[7,9], Alan Ma[2,4], Steven Macaskill[10], Amali Mallawaarachchi[28,48], Melanie Marty[9], Justine E. Marum[9], Hugh J. McCarthy[2,13], Manoj P. Menezes[2,49], Alison McLean[34], Di Milnes[11], Shekeeb Mohammad[2,49], David Mowat[43,47], Aram Niaz[3], Elizabeth E. Palmer[43,47], Chirag Patel[11], Shilpan G. Patel[50], Dean Phelan[9], Jason R. Pinner[43,47], Sulekha Rajagopalan[34], Matthew Regan[30,42], Jonathan Rodgers[11], Miriam Rodrigues[51], Richard H. Roxburgh[51], Rani Sachdev[43], Tony Roscioli[22,43,52], Ruvishani Samarasekera[4], Sarah A. Sandaradura[1,2,4], Elena Savva[9], Tim Schindler[47,53], Margit Shah[4], Ingrid B. Sinnerbrink[16,35], Janine M. Smith[4,16], Richard J. Smith[54], Amanda Springer[42], Zornitza Stark[5,9], Samuel P. Strom[55], Carolyn M. Sue[44], Kenneth Tan[42,56], Tiong Y. Tan[5,9], Esther Tantsis[2,49], Michel C. Tchan[16,36], Bryony A. Thompson[57,58], Alison H. Trainer[6,8], Karin van Spaendonck-Zwarts[59], Rebecca Walsh[22], Linda Warwick[27], Stephanie White[17], Susan M. White[5,9], Mark G. Williams[60], Meredith J. Wilson[4,16], Wui Kwan Wong[1,2,49], Dale C. Wright[2,16,61], Patrick Yap[62], Alison Yeung[5,9], Helen Young[63], Kristi J. Jones[1,2,4], Bruce Bennetts[2,38], Sandra T. Cooper[1,2,64,*]; on behalf of the Australasian Consortium for RNA Diagnostics

Affiliations

[1]Kids Neuroscience Centre, Kids Research, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; [2]Department of Child and Adolescent Health, Faculty of Medicine and Health, The University of Sydney, Westmead, New South Wales, Australia; [3]Rare Diseases Functional Genomics, Kids Research, Sydney Children's Hospital Network and Children's Medical Research Institute, Westmead, New South Wales, Australia; [4]Department of Clinical Genetics, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; [5]Department of Paediatrics, University of Melbourne, Parkville, Victoria, Australia; [6]Department of Medicine, University of Melbourne, Parkville, Victoria, Australia; [7]Department of Pathology, University of Melbourne, Parkville, Victoria, Australia; [8]Department of Genomic Medicine, The Royal Melbourne Hospital, Parkville, Victoria, Australia; [9]Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Parkville, Victoria, Australia; [10]Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia; [11]Genetic Health Queensland, Royal Brisbane and

Women's Hospital, Herston, Queensland, Australia; [12]The University of Queensland, Herston, Queensland, Australia; [13]Department of Pediatric Nephrology, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; [14]Genetic Metabolic Disorders Service, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; [15]Western Sydney Genetics Program, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; [16]Specialty of Genomic Medicine, Faculty of Medicine and Health, The University of Sydney, Westmead, New South Wales, Australia; [17]Department of Clinical Genetics, Royal North Shore Hospital, St Leonards, New South Wales, Australia; [18]Northern Clinical School, Royal North Shore Hospital, St Leonards, New South Wales, Australia; [19]Australian Genomics Health Alliance, Parkville, Victoria, Australia; [20]Centre for Clinical Genetics, Sydney Children's Hospital Randwick, Randwick, New South Wales, Australia; [21]Department of Clinical Genetics, Austin Health, Heidelberg, Victoria, Australia; [22]NSW Health Pathology, Randwick, New South Wales, Australia; [23]Department of Diagnostic Genomics, PathWest Laboratory Medicine, QEII Medical Centre, Nedlands, Western Australia, Australia; [24]The Royal Melbourne Hospital, Parkville, Victoria, Australia; [25]Murdoch Children's Research Institute, Parkville, Victoria, Australia; [26]Department of Paediatrics, School of Medicine, Western Sydney University, Penrith South, New South Wales, Australia; [27]ACT Genetic Service, ACT Health, The Canberra Hospital, Garran, ACT, Australia; [28]Department of Medical Genomics, Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia; [29]Central Clinical School, The University of Sydney, Camperdown, New South Wales, Australia; [30]Monash Genetics, Monash Health, Clayton, Victoria, Australia; [31]Genetics of Learning Disability Service, Hunter Genetics, Waratah, New South Wales, Australia; [32]Department of Clinical Genetics, The Canberra Hospital, Garran, ACT, Australia; [33]The University of Newcastle, Callaghan, New South Wales, Australia; [34]Department of Clinical Genetics, Liverpool Hospital, Liverpool, New South Wales, Australia; [35]Department of Clinical Genetics, Nepean Hospital, Kingswood, New South Wales, Australia; [36]Department of Genomic Medicine, Westmead Hospital, Westmead, New South Wales, Australia; [37]Sullivan Nicolaides Pathology, Bowen Hills, Queensland, Australia; [38]Department of Molecular Genetics, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; [39]Forster Genetics, Forster, New South Wales, Australia; [40]Monash Heart and Monash Children's Hospital, Monash Health, Clayton, Victoria, Australia; [41]Monash Cardiovascular Research Centre, Clayton, Victoria, Australia; [42]Department of Paediatrics, Monash University, Clayton, Victoria, Australia; [43]Center for Clinical Genetics, Sydney Children's Hospital, Randwick, New South Wales, Australia; [44]Department of Neurogenetics, Kolling Institute, Faculty of Medicine and Health, University of Sydney, Royal North Shore Hospital, St Leonards, New South Wales, Australia; [45]Translational Genomics, Kinghorn Centre for Clinical Genomics, Garvan Institute for Medical Research, Darlinghurst, New South Wales, Australia; [46]Department of Neurology, The Royal Children's Hospital, Parkville, Victoria, Australia; [47]School of Women's and Children's Health, Faculty of Medicine and Health, University of New South Wales, Kensington, NSW, Australia; [48]Division of Genomics and Epigenetics, Garvan Institute of Medical Research, Darlinghurst, New South Wales, Australia; [49]The TY Nelson Department of Neurology and Neurosurgery, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; [50]School of Medicine, The University of Auckland, Auckland, New Zealand; [51]Department of Neurology, Auckland City Hospital, Auckland, New Zealand; [52]Neuroscience Research Australia, University of New South Wales, Randwick, New South Wales, Australia; [53]Newborn Care, Royal Hospital for Women, Randwick, New South Wales, Australia; [54]Molecular Otolaryngology and Renal Research Laboratories, Carver College of Medicine, University of Iowa, Iowa City, IA, United States; [55]Fulgent Genetics, Temple City, CA; [56]Monash Newborn, Monash Children's Hospital, Clayton, Victoria, Australia; [57]Department of Pathology, The Royal Melbourne Hospital, Parkville, Victoria, Australia; [58]Department of Clinical Pathology, University of Melbourne, Parkville, Victoria, Australia; [59]Genetic Health Queensland, Royal Brisbane and Women's Hospital, Herston, Queensland, Australia; [60]Mater Research Institute, The University of Queensland, South Brisbane, Queensland, Australia; [61]Department of Cytogenetics, The Children's Hospital at Westmead, Westmead, New South Wales, Australia; [62]Northern Hub, Genetic Health Service NZ, Auckland, New Zealand; [63]Department of Intensive Care, Austin Hospital, Heidelberg, Victoria, Australia; [64]The Children's Medical Research Institute, Westmead, New South Wales, Australia

References

1. Baralle D, Buratti E. RNA splicing in human disease and in the clinic. *Clin Sci (Lond)*. 2017;131(5):355–368. http://doi.org/10.1042/CS20160211.

2. 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.

3. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*. 2017;9(386):eaal5209. http://doi.org/10.1126/scitranslmed.aal5209.

4. Frésard L, Smail C, Ferraro NM, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med*. 2019;25(6):911–919. http://doi.org/10.1038/s41591-019-0457-8.

5. Rentas S, Rathi KS, Kaur M, et al. Diagnosing Cornelia de Lange syndrome and related neurodevelopmental disorders using RNA sequencing. *Genet Med*. 2020;22(5):927–936. http://doi.org/10.1038/s41436-019-0741-5.

6. Lee H, Huang AY, Wang LK, et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med*. 2020;22(3):490–499. http://doi.org/10.1038/s41436-019-0672-1.

7. Akesson LS, Bournazos A, Fennell A, et al. Rapid exome sequencing and adjunct RNA studies confirm the pathogenicity of a novel homozygous ASNS splicing variant in a critically ill neonate. *Hum Mutat*. 2020;41(11):1884–1891. http://doi.org/10.1002/humu.24101.

8. Bryen SJ, Ewans LJ, Pinner J, et al. Recurrent TTN metatranscript-only c.39974-11T>G splice variant associated with autosomal recessive arthrogryposis multiplex congenita and myopathy. *Hum Mutat*. 2020;41(2):403–411. http://doi.org/10.1002/humu.23938.

9. Lykke-Andersen S, Jensen TH. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat Rev Mol Cell Biol*. 2015;16(11):665–677. http://doi.org/10.1038/nrm4063.

10. Aicher JK, Jewell P, Vaquero-Garcia J, Barash Y, Bhoj EJ. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet Med*. 2020;22(7):1181–1190. http://doi.org/10.1038/s41436-020-0780-y.

11. Wai HA, Lord J, Lyon M, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med*. 2020;22(6):1005–1014. Published correction appears in *Genet Med*. 2020;22(6):1129. https://doi.org/10.1038/s41436-020-0766-9.

12. Bryen SJ, Joshi H, Evesson FJ, et al. Pathogenic abnormal splicing due to intronic deletions that induce biophysical space constraint for spliceosome assembly. *Am J Hum Genet*. 2019;105(3):573–587. http://doi.org/10.1016/j.ajhg.2019.07.013.

13. Bryen SJ, Oates EC, Evesson FJ, et al. Pathogenic deep intronic MTM1 variant activates a pseudo-exon encoding a nonsense codon resulting in severe X-linked myotubular myopathy. *Eur J Hum Genet*. 2021;29(1):61–66. http://doi.org/10.1038/s41431-020-00715-7.

14. Jones HF, Bryen SJ, Waddell LB, et al. Importance of muscle biopsy to establish pathogenicity of DMD missense and splice variants. *Neuromuscul Disord*. 2019;29(12):913–919. http://doi.org/10.1016/j.nmd.2019.09.013.

15. Riley LG, Waddell LB, Ghaoui R, et al. Recessive DES cardio/myopathy without myofibrillar aggregates: intronic splice variant silences one allele leaving only missense L190P-desmin. *Eur J Hum Genet*. 2019;27(8):1267–1273. http://doi.org/10.1038/s41431-019-0393-6.

16. Sandaradura SA, Bournazos A, Mallawaarachchi A, et al. Nemaline myopathy and distal arthrogryposis associated with an autosomal recessive TNNT3 splice variant. *Hum Mutat*. 2018;39(3):383–388. http://doi.org/10.1002/humu.23385.

17. Waddell LB, Bryen SJ, Cummings BB, et al. WGS and RNA studies diagnose noncoding DMD variants in males with high creatine kinase. *Neurol Genet*. 2021;7(1):e554. http://doi.org/10.1212/NXG.0000000000000554.

18. Dawes R, Lek M, Cooper ST. Gene discovery informatics toolkit defines candidate genes for unexplained infertility and prenatal or infantile mortality. *NPJ Genom Med*. 2019;4(1):8. http://doi.org/10.1038/s41525-019-0081-z.

19. Zhou T, Benda C, Dunzinger S, et al. Generation of human induced pluripotent stem cells from urine samples. *Nat Protoc*. 2012;7(12):2080–2089. http://doi.org/10.1038/nprot.2012.115.

20. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21. http://doi.org/10.1093/bioinformatics/bts635.

21. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24–26. http://doi.org/10.1038/nbt.1754.

22. Li B, Ruotti V, Stewart RM, Thomson JA, Dewey CN. RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*. 2010;26(4):493–500. http://doi.org/10.1093/bioinformatics/btp692.

23. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell*. 2019;176(3):535–548.e24.

24. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580–585. http://doi.org/10.1038/ng.2653.

25. Sloan CA, Chan ET, Davidson JM, et al. ENCODE data at the ENCODE portal. *Nucleic Acids Res*. 2016;44(D1):D726–D732. http://doi.org/10.1093/nar/gkv1160.

26. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74. http://doi.org/10.1038/nature11247.

27. Davis CA, Hitz BC, Sloan CA, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res*. 2018;46(D1):D794–D801. http://doi.org/10.1093/nar/gkx1081.

28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–381. http://doi.org/10.1016/j.jbi.2008.08.010.

29. Australian Genomics Health Alliance Acute Care Flagship, Lunke S, Eggers S, et al. Feasibility of ultra-rapid exome sequencing in critically ill infants and children with suspected monogenic conditions in the Australian public health care system. *JAMA*. 2020;323(24):2503–2511. http://doi.org/10.1001/jama.2020.7671.

30. McInerney-Leo AM, Harris JE, Gattas M, et al. Fryns syndrome associated with recessive mutations in PIGN in two separate families. *Hum Mutat*. 2016;37(7):695–702. http://doi.org/10.1002/humu.22994.

31. Kurosaki T, Popp MW, Maquat LE. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol*. 2019;20(7):406–420. Published correction appears in *Nat Rev Mol Cell Biol*. 2019;20(6):384. https://doi.org/10.1038/s41580-019-0126-2.

32. Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol*. 2017;18(10):637–650. http://doi.org/10.1038/s41580-019-0126-2.

33. Acevedo JM, Hoermann B, Schlimbach T, Teleman AA. Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. *Sci Rep*. 2018;8(1):4018. http://doi.org/10.1038/s41598-018-22330-9.

34. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67. http://doi.org/10.1093/nar/gkp215.

35. Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res*. 2001;29(5):1185–1190. http://doi.org/10.1093/nar/29.5.1185.

36. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997;4(3):311–323. http://doi.org/10.1089/cmb.1997.4.311.

37. Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*. 1987;15(17):7155–7174. http://doi.org/10.1093/nar/15.17.7155.

38. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol*. 2004;11(2-3):377–394. http://doi.org/10.1089/1066527041410418.

39. Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat*. 2018;39(11):1517–1524. http://doi.org/10.1002/humu.23626.

40. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med*. 2019;12(1):3. http://doi.org/10.1186/s13073-019-0690-2.

Supplementary Information

Supplementary methods

Sample collection and RNA extraction

Whole blood: 1 - 3 millilitres of whole blood was collected in a PAXgene® blood RNA tube (PreAnalytiX) and RNA was isolated using the PAXgene® blood RNA kit according to kit instructions.

Peripheral blood mononuclear cells (PBMCs): 6 - 8 millilitres of patient whole blood was collected in 2 lithium heparin tubes (Becton Dickinson) or 2 ACD-B tubes (Greiner Bio One). SepMate-15 tubes (StemCell Technologies) and Ficoll® Paque Plus (GE Healthcare) were used to isolate PBMCs. Two wells of a 12-well plate were plated with 1-3 million cells, depending on PBMC yield. PBMCs were cultured in RPMI 1640 Medium (Gibco), 10% fetal bovine serum (GE Healthcare) and penicillin-streptomycin (100 U-µg/mL) (Gibco).

Skin biopsy: Primary skin fibroblasts were cultured to 90% confluency in a 6-well plate containing high glucose DMEM (Gibco), 10% fetal bovine serum (GE Healthcare) and Gentamicin (50 µg/ml) (Gibco).

Urine sample: Urothelial cell culture was performed as described in steps 1-16 of Zhou et al., 2012[19].

Cycloheximide treatment: PBMCs, transformed lymphocytes, primary fibroblasts and urothelia were treated with dimethyl sulfoxide (DMSO) (Sigma-Aldrich) or 100 µg/ml cycloheximide (Sigma-Aldrich) for 6 h before harvesting in TRIzol reagent (Invitrogen).

RNA Extraction: RNA was isolated using the standard TRIzol procedure followed by the RNase-free DNase set (Qiagen) and RNeasy mini kit cleanup protocols (Qiagen).

RT-PCR and Sanger sequencing

cDNA synthesis: SuperScript IV first-strand synthesis system (Invitrogen) was used to make cDNA from 500 ng of RNA according to kit instructions.

PCR: Recombinant Taq DNA polymerase (Invitrogen) and MasterAmp 2X PCR PreMix D (Epicentre Biotechnologies) or LongAmp® Taq DNA Polymerase (New England BioLabs) was used for PCRs. Thermocycling conditions were 94°C for 3 min, 35 cycles 94°C 30 s, 58°C 30 s, 72°C 90 s/kb, then 72°C 10 min for Recombinant Taq and 94°C for 3 min, 35 cycles 94°C 30 s, 58°C 30 s, 65°C 60 s/kb, then 65°C 10 min for LongAmp® Taq. Control cDNA: healthy individuals (where available), family members or affected individuals from cases with genetic variants in an unrelated gene. All PCR products were analysed on a 1.2% agarose gel.

Gel Extraction of PCR Amplicons: Bands were manually excised from an agarose gel with a scalpel and cDNA purified using GeneJET gel extraction kit (Thermo Scientific) according to the Manufacturer's instructions.

Sanger sequencing: 8-75 ng of purified cDNA and 1 pmol of sequencing primer were subject to Sanger sequencing at the Australian Genomics Research Facility. Sanger sequencing chromatograms were analysed using Sequencher® DNA sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA.

Primers: See Table S4.

Western Blot

Tafazzin was analysed by SDS-PAGE using Bolt 10% Bis-Tris polyacrylamide gels (Invitrogen) and 3-(N-Morpholino)propane sulfonic acid running buffer, with PageRuler Plus (Thermo Scientific) size marker. Loading was determined by Pierce BCA assay kit (Thermo Scientific). Proteins were transferred onto methanol-activated

Immobilon®-P PVDF membranes (Millipore). Membranes were blocked using 5% skim milk in PBS containing 0.1% Tween-20 (Amresco). Blots were probed with primary antibodies incubated overnight at 4°C, membranes were washed with PBS containing 0.1% Tween-20 and incubated with horseradish peroxidase (HRP)–conjugated secondary antibodies (Thermo Scientific) for 2 h at room temperature. Membranes were washed in PBS containing 0.1% Tween-20 and developed using ECL reagent and Hyperfilm (Cytiva). The membranes were then stained by coomassie. Primary antibodies: Anti-Tafazzin (PA2135, Boster Biological Technology), Anti-Cardiac Actin (clone Ac1-20.4.2, 03-61075, American Research Products, Inc.), Anti-Glyceraldehyde-3-Phosphate Dehydrogenase (clone 6C5, MAB374, Merck). Secondary antibodies: Goat Anti-Mouse light chain Antibody, HRP conjugate (AP200P, Merck), Goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP conjugate (65-6120, Invitrogen).

Consultation process for recommended PS3/BS3 interpretation rubric

The consultation process had four stages: 1**)** REDCap Surveys distributed to SpliceACORD members. For example, for development of SOPs, surveys were returned from 18 CVCs comprised of 2 genetic pathologists, 1 genetic pathology trainee, 12 clinical scientists and 3 principal scientists (as defined by National Pathology Accreditation Advisory Council of Australia) across 11 laboratories. **2)** Opinion was sought on draft consensus protocols devised, or interpretation rubric, via live Zoom polls at SpliceACORD meetings. Polls were anonymised to provide a safe environment for SpliceACORD members to freely express contrasting opinions and to see in real-time the perspectives of others. Attendance at SpliceACORD meetings where polling took place ranged from 40-60 attendees (primarily diagnostic genomic pathology and clinical genetics members, with translational researcher members ~20%). **3)** Consensus protocols and interpretation rubric were iteratively refined by written feedback by SpliceACORD members during manuscript preparation. **4)** Near-final consensus protocols and interpretation rubric were forwarded/presented to the ClinGen

LGMD expert panel (S.Cooper is member), the ClinGen Splicing SVI specialist subgroup (A.Spurdle is member), and to the ACMG - to garner comments and feedback of contrasting viewpoints, prior to manuscript submission.

Definition of a 'putative splicing variant' used inclusion criteria for ascertainment of variants.

A variant that affects:

1) the extended splice-site region (*Donor*: from third to last nucleotide of exon to +8; *Acceptor*: three nucleotides upstream of branchpoint to third nucleotide of exon),

2) a cryptic splice site (exonic or intronic)

3) a putative splice enhancer or repressor (for example, a deep intronic variant within an intronic region flanked by extant donor and acceptor cryptic essential splice-sites, consistent with possible activation of a pseudoexon).

Table S1: List of Mendelian disease genes with clinically relevant phenotypes[2].

Table S2: List of primers used in this study.

Table S3: RNA-seq sample information metrics.

Table S4: Phenotypic summary of RNA diagnostics cohort.

Table S5: Cost of testing.

Table S6: List of genetic variants, splicing outcomes, and changes in classification[3,4].

Figure S1: Splice variant submission portal form.

Figure S2: Clinical impact survey.

Figure S3: Clinical variant curator survey.

Table S1: List of Mendelian disease genes with clinically relevant phenotypes (Dawes et al. 2019[1])

A4GALT	AAAS	AAGAB	AARS	AARS2	AASS	ABAT	ABCA1	ABCA3	ABCA4	ABCA12	ABCB4	ABCB6	ABCB7	ABCB11
ABCC2	ABCC6	ABCC8	ABCC9	ABCD1	ABCD4	ABCG5	ABCG8	ABHD5	ABHD12	ABL1	ACAD8	ACAD9	ACADM	ACADS
ACADSB	ACADVL	ACAN	ACAT1	ACE	ACO2	ACOX1	ACOX2	ACP4	ACP5	ACSF3	ACSL4	ACSL6	ACTA1	ACTA2
ACTB	ACTC1	ACTG1	ACTG2	ACTN1	ACTN2	ACTN4	ACVR1	ACVR2B	ACVRL1	ACY1	ADA2	ADAM9	ADAM10	ADAMTS2
ADAMTS10	ADAMTS13	ADAMTS17	ADAMTS18	ADAMTSL2	ADAMTSL4	ADAR	ADAT3	ADCY5	ADD3	ADGRE2	ADGRG1	ADGRG2	ADGRG6	ADGRV1
ADIPOQ	ADK	ADNP	ADRA2B	ADRB2	ADSL	ADSSL1	AFF2	AFF4	AFG3L2	AFP	AGA	AGBL1	AGBL5	AGK
AGL	AGPAT2	AGPS	AGRN	AGT	AGTR1	AGXT	AHCY	AHDC1	AHI1	AICDA	AIFM1	AIMP1	AIP	AIPL1
AIRE	AK1	AK2	AKR1C2	AKR1D1	AKT1	AKT2	AKT3	ALAD	ALAS2	ALB	ALDH1A3	ALDH2	ALDH3A2	ALDH4A1
ALDH5A1	ALDH6A1	ALDH7A1	ALDH18A1	ALDOA	ALDOB	ALG1	ALG2	ALG3	ALG6	ALG8	ALG9	ALG11	ALG12	ALG13
ALMS1	ALOX12B	ALOXE3	ALPL	ALS2	ALX3	ALX4	AMACR	AMBN	AMELX	AMER1	AMH	AMHR2	AMMECR1	AMN
AMPD1	AMPD2	AMT	ANG	ANGPTL3	ANGPTL4	ANK1	ANK2	ANKH	ANKK1	ANKRD11	ANKRD26	ANKS6	ANLN	ANO3
ANO5	ANO6	ANO10	ANOS1	ANTXR1	ANTXR2	ANXA11	AP1S1	AP1S2	AP2S1	AP3B1	AP3B2	AP4B1	AP4E1	AP4M1
AP4S1	AP5Z1	APC	APCDD1	APOA1	APOA2	APOA5	APOB	APOC2	APOC3	APOE	APOPT1	APP	APRT	APTX
AQP2	AQP5	AR	ARCN1	ARFGEF2	ARG1	ARHGAP31	ARHGDIA	ARHGEF6	ARHGEF9	ARHGEF18	ARID1A	ARID1B	ARID2	ARL2BP
ARL6	ARL13B	ARMC4	ARMC9	ARPC1B	ARR3	ARSA	ARSB	ARSE	ARV1	ARX	ASAH1	ASB10	ASCC1	ASCL1
ASH1L	ASL	ASNS	ASPA	ASPH	ASPM	ASPSCR1	ASS1	ASXL1	ASXL2	ASXL3	ATAD3A	ATCAY	ATF6	ATIC
ATL1	ATL3	ATM	ATN1	ATOH7	ATP1A2	ATP1A3	ATP2A1	ATP2A2	ATP2C1	ATP6AP1	ATP6AP2	ATP6V1A	ATP6V1B1	ATP6V1B2
ATP6V1E1	ATP6V0A2	ATP6V0A4	ATP7A	ATP7B	ATP8B1	ATP13A2	ATR	ATRX	ATXN1	ATXN2	ATXN3	ATXN7	ATXN8	ATXN10
AUH	AURKC	AUTS2	AVP	AVPR2	AXIN2	B2M	B3GALNT2	B3GALT6	B3GAT3	B3GLCT	B4GALNT1	B4GALT1	B4GALT7	B4GAT1
B9D1	B9D2	BAAT	BAG3	BANF1	BAP1	BBS1	BBS2	BBS4	BBS5	BBS7	BBS9	BBS10	BBS12	BCAP31
BCHE	BCKDHA	BCKDHB	BCKDK	BCL2	BCL7A	BCL11A	BCOR	BCS1L	BDNF	BEAN1	BEST1	BFSP1	BFSP2	BGN
BHLHA9	BICD2	BIN1	BLK	BLM	BLOC1S3	BLVRA	BMP1	BMP2	BMP4	BMP15	BMPER	BMPR1A	BMPR1B	BMPR2
BOLA3	BPGM	BPTF	BRAF	BRAT1	BRCA2	BRF1	BRIP1	BRPF1	BRWD3	BSCL2	BSND	BTD	BTK	BUB1B
C1QA	C1QB	C1QBP	C1QC	C1QTNF5	C1R	C1S	C2	C3	C4A	C4B	C5	C6	C7	C8A
C8B	C8orf37	C9	C9orf72	C12orf57	C12orf65	C15orf41	C19orf12	C21orf2	CA2	CA4	CA5A	CA8	CA12	CABP2
CABP4	CACNA1A	CACNA1C	CACNA1D	CACNA1F	CACNA1G	CACNA1H	CACNA1S	CACNA2D4	CACNB2	CACNB4	CAD	CALM1	CALM2	CAMK2A
CAMK2B	CAMTA1	CANT1	CAPN1	CAPN3	CAPN5	CARD9	CARD11	CARD14	CARS2	CASK	CASP10	CASP14	CASQ1	CASQ2
CASR	CAST	CAT	CATSPER1	CAV1	CAV3	CAVIN1	CBL	CBS	CC2D1A	CC2D2A	CCBE1	CCDC8	CCDC22	CCDC39
CCDC40	CCDC65	CCDC88C	CCDC103	CCDC114	CCDC115	CCDC151	CCDC174	CCM2	CCND2	CCNO	CCNQ	CCT5	CD2AP	CD3D
CD3E	CD3G	CD4	CD8A	CD19	CD27	CD36	CD40	CD40LG	CD55	CD59	CD79A	CD79B	CD81	CD96
CD151	CD320	CDAN1	CDC14A	CDC42	CDC45	CDC73	CDCA7	CDH1	CDH3	CDH15	CDH23	CDHR1	CDK5RAP2	CDK10
CDK13	CDKL5	CDKN1B	CDKN1C	CDKN2A	CDON	CDSN	CDT1	CEACAM16	CEBPE	CEL	CENPF	CENPJ	CEP19	CEP41
CEP55	CEP57	CEP78	CEP83	CEP104	CEP120	CEP135	CEP152	CEP164	CEP290	CERKL	CERS3	CES1	CETP	CFAP43
CFAP53	CFAP298	CFC1	CFD	CFH	CFHR5	CFI	CFL2	CFP	CFTR	CHAMP1	CHAT	CHCHD2	CHCHD10	CHD1
CHD2	CHD4	CHD7	CHEK2	CHKB	CHM	CHMP1A	CHMP2B	CHMP4B	CHN1	CHRDL1	CHRNA1	CHRNA2	CHRNA4	CHRNB1
CHRNB2	CHRND	CHRNE	CHRNG	CHST3	CHST6	CHST14	CHSY1	CHUK	CIB2	CIC	CIITA	CISD2	CIT	CITED2
CKAP2L	CLCF1	CLCN1	CLCN2	CLCN4	CLCN5	CLCN7	CLCNKA	CLCNKB	CLDN1	CLDN10	CLDN14	CLDN16	CLDN19	CLEC7A
CLMP	CLN3	CLN5	CLN6	CLN8	CLP1	CLPB	CLPP	CLRN1	CLTC	CNBP	CNGA1	CNGA3	CNGB1	CNGB3
CNKSR2	CNNM2	CNNM4	CNTNAP1	CNTNAP2	COA6	COASY	COCH	COG1	COG4	COG5	COG6	COG7	COG8	COL1A1
COL1A2	COL2A1	COL3A1	COL4A1	COL4A2	COL4A3	COL4A3BP	COL4A4	COL4A5	COL5A1	COL5A2	COL6A1	COL6A2	COL6A3	COL7A1
COL8A2	COL9A1	COL9A2	COL9A3	COL10A1	COL11A1	COL11A2	COL12A1	COL13A1	COL17A1	COL18A1	COL25A1	COLEC10	COLEC11	COLQ
COMP	COQ2	COQ4	COQ6	COQ8A	COQ8B	COQ9	CORIN	CORO1A	COX4I2	COX6A1	COX6B1	COX7B	COX10	COX15
COX20	CP	CPA6	CPAMD8	CPLANE1	CPN1	CPOX	CPS1	CPT1A	CPT2	CR2	CRADD	CRB1	CRB2	CRBN
CREBBP	CRELD1	CRIPT	CRLF1	CRTAP	CRTC1	CRX	CRYAA	CRYAB	CRYBA1	CRYBA4	CRYBB1	CRYBB2	CRYBB3	CRYGB

CRYGC	CRYGD	CRYGS	CRYM	CSF1R	CSF2RA	CSF2RB	CSF3R	CSNK1D	CSNK2A1	CSPP1	CSRP3	CST3	CSTA	CSTB
CTC1	CTCF	CTDP1	CTH	CTHRC1	CTLA4	CTNNA1	CTNNA3	CTNNB1	CTNND1	CTNS	CTPS1	CTSA	CTSC	CTSD
CTSF	CTSK	CUBN	CUL3	CUL4B	CUL7	CWC27	CWF19L1	CXCR4	CYB5R3	CYBA	CYBB	CYC1	CYCS	CYLD
CYP1B1	CYP2A6	CYP2B6	CYP2C8	CYP2C9	CYP2C19	CYP2R1	CYP2U1	CYP4F22	CYP4V2	CYP7B1	CYP11A1	CYP11B1	CYP11B2	CYP17A1
CYP19A1	CYP21A2	CYP24A1	CYP26B1	CYP26C1	CYP27A1	CYP27B1	D2HGDH	DAB1	DAG1	DARS	DARS2	DBH	DBT	DCAF17
DCC	DCDC2	DCHS1	DCLRE1C	DCN	DCPS	DCTN1	DCX	DDB2	DDC	DDHD1	DDHD2	DDR2	DDRGK1	DDX3X
DDX11	DDX58	DDX59	DEAF1	DENND5A	DEPDC5	DES	DGKE	DGUOK	DHCR7	DHCR24	DHDDS	DHFR	DHH	DHODH
DHTKD1	DHX30	DIABLO	DIAPH1	DIAPH3	DICER1	DIP2B	DIS3L2	DKC1	DLAT	DLD	DLG3	DLL3	DLL4	DLX3
DMD	DMGDH	DMP1	DMPK	DNA2	DNAAF1	DNAAF2	DNAAF3	DNAAF4	DNAAF5	DNAH1	DNAH5	DNAH11	DNAI1	DNAI2
DNAJB2	DNAJB6	DNAJB13	DNAJC5	DNAJC6	DNAJC12	DNAJC19	DNAJC21	DNAL1	DNASE1L3	DNM1	DNM1L	DNM2	DNMT1	DNMT3A
DNMT3B	DOCK2	DOCK6	DOCK7	DOCK8	DOK7	DOLK	DONSON	DPAGT1	DPH1	DPM1	DPM2	DPM3	DPP6	DPY19L2
DPYD	DPYS	DRAM2	DRC1	DRD4	DRD5	DSC2	DSE	DSG1	DSG2	DSG4	DSP	DSPP	DST	DSTYK
DTNA	DTNBP1	DUOX2	DUOXA2	DUSP6	DVL1	DVL3	DYM	DYNC1H1	DYNC2H1	DYNC2LI1	DYRK1A	DYRK1B	DYSF	DZIP1L
EARS2	EBF3	EBP	ECEL1	ECHS1	ECM1	EDA	EDAR	EDARADD	EDN1	EDN3	EDNRA	EDNRB	EED	EEF1A2
EFEMP1	EFEMP2	EFNB1	EFTUD2	EGF	EGFR	EGLN1	EGR2	EHMT1	EIF2AK3	EIF2AK4	EIF2B1	EIF2B2	EIF2B3	EIF2B4
EIF2B5	EIF2S3	EIF4A3	ELAC2	ELANE	ELMO2	ELN	ELOVL4	ELOVL5	ELP1	ELP2	EMC1	EMD	EMG1	EML1
EMP2	EMX2	ENAM	ENG	ENPP1	ENTPD1	EOGT	EP300	EPAS1	EPB41	EPB42	EPCAM	EPG5	EPHA2	EPHX1
EPM2A	EPS8L2	ERAL1	ERBB3	ERBB4	ERCC1	ERCC2	ERCC3	ERCC4	ERCC5	ERCC6	ERCC6L2	ERCC8	ERF	ERLIN1
ERLIN2	ESCO2	ESPN	ESR1	ESRRB	ETFA	ETFB	ETFDH	ETHE1	ETV6	EVC	EVC2	EWSR1	EXOSC2	EXOSC3
EXOSC8	EXPH5	EXT1	EXT2	EXTL3	EYA1	EYA4	EYS	EZH2	F2	F5	F7	F8	F9	F10
F11	F12	F13A1	F13B	FA2H	FADD	FAH	FAM20A	FAM20C	FAM83H	FAM111A	FAM111B	FAM126A	FAM161A	FAN1
FANCA	FANCB	FANCC	FANCD2	FANCE	FANCF	FANCG	FANCI	FANCL	FAR1	FARS2	FAS	FASLG	FAT2	FAT4
FBLN5	FBN1	FBN2	FBP1	FBXL4	FBXO7	FBXO38	FCGR3A	FCGR3B	FCN3	FDPS	FDXR	FECH	FERMT1	FERMT3
FEZF1	FGA	FGB	FGD1	FGD4	FGF3	FGF5	FGF8	FGF9	FGF10	FGF12	FGF14	FGF16	FGF17	FGF23
FGFR1	FGFR2	FGFR3	FGG	FH	FHL1	FIBP	FIG4	FIGLA	FKBP10	FKBP14	FKRP	FKTN	FLAD1	FLCN
FLG	FLI1	FLNA	FLNB	FLNC	FLRT3	FLT4	FLVCR1	FLVCR2	FMN2	FMO3	FMR1	FN1	FOLR1	FOXC1
FOXC2	FOXE1	FOXE3	FOXF1	FOXG1	FOXI1	FOXL2	FOXN1	FOXO1	FOXP1	FOXP2	FOXP3	FOXRED1	FRAS1	FREM1
FREM2	FRMD7	FRMPD4	FRRS1L	FSCN2	FSHB	FSHR	FTCD	FTL	FTO	FTSJ1	FUCA1	FUS	FUT6	FUZ
FXN	FXYD2	FYB1	FYCO1	FZD4	FZD6	G6PC	G6PC3	G6PD	GAA	GABRA1	GABRB1	GABRB2	GABRB3	GABRG2
GALC	GALE	GALK1	GALNS	GALNT3	GALT	GAMT	GAN	GANAB	GARS	GAS8	GATA1	GATA2	GATA3	GATA4
GATA6	GATAD2B	GATM	GBA	GBA2	GBE1	GCDH	GCH1	GCK	GCLC	GCM2	GCNT2	GCSH	GDAP1	GDF1
GDF2	GDF3	GDF5	GDF6	GDI1	GDNF	GFAP	GFER	GFI1B	GFM1	GFPT1	GGCX	GH1	GHR	GHRHR
GHSR	GIF	GINS1	GIPC3	GJA1	GJA3	GJA5	GJA8	GJB1	GJB2	GJB3	GJB4	GJB6	GJC2	GK
GLA	GLB1	GLDC	GLDN	GLE1	GLI2	GLI3	GLIS2	GLIS3	GLMN	GLRA1	GLRB	GLRX5	GLUD1	GLUL
GLYCTK	GM2A	GMNN	GMPPA	GMPPB	GNA11	GNAI2	GNAI3	GNAL	GNAO1	GNAS	GNAT1	GNAT2	GNB1	GNB3
GNB4	GNB5	GNE	GNMT	GNPAT	GNPTAB	GNPTG	GNRHR	GNS	GORAB	GOSR2	GOT1	GP1BA	GP1BB	GP6
GP9	GPAA1	GPC3	GPC6	GPD1	GPD1L	GPHN	GPI	GPIHBP1	GPR68	GPR101	GPR143	GPR179	GPSM2	GPT2
GPX4	GREB1L	GREM2	GRHL2	GRHL3	GRHPR	GRIA3	GRIA4	GRID2	GRIK2	GRIN1	GRIN2A	GRIN2B	GRIN2D	GRIP1
GRK1	GRM1	GRM6	GRN	GRXCR1	GSC	GSDME	GSN	GSS	GTF2E2	GTF2H5	GTPBP3	GUCA1A	GUCA1B	GUCY1A1
GUCY2C	GUCY2D	GUSB	GYG1	GYS1	GYS2	GZF1	H6PD	HAAO	HACE1	HADH	HADHA	HADHB	HAMP	HARS
HAX1	HBA1	HBA2	HBB	HBD	HBG1	HBG2	HCCS	HCFC1	HCN1	HCN4	HDAC8	HECW2	HELLS	HEPACAM
HERC1	HERC2	HES7	HESX1	HEXA	HEXB	HFE	HFM1	HGD	HGF	HGSNAT	HIBCH	HIKESHI	HINT1	HIST1H1E
HIVEP2	HJV	HK1	HLCS	HMBS	HMGCL	HMGCS2	HMOX1	HMX1	HNF1A	HNF1B	HNF4A	HNMT	HNRNPA1	HNRNPDL
HNRNPH2	HNRNPK	HNRNPU	HOGA1	HOXA1	HOXA11	HOXA13	HOXB1	HOXC13	HOXD10	HOXD13	HPCA	HPD	HPGD	HPRT1
HPS1	HPS3	HPS4	HPS5	HPS6	HPSE2	HR	HRAS	HRG	HSD3B2	HSD3B7	HSD11B1	HSD11B2	HSD17B3	HSD17B4

HSD17B10	HSF4	HSPA9	HSPB1	HSPB8	HSPD1	HSPG2	HTR1A	HTRA1	HTRA2	HTT	HUWE1	HYDIN	HYLS1	IARS
ICK	ICOS	IDH2	IDH3B	IDS	IDUA	IER3IP1	IFIH1	IFITM5	IFNGR1	IFNGR2	IFT43	IFT52	IFT80	IFT81
IFT122	IFT140	IFT172	IGBP1	IGF1	IGF1R	IGFALS	IGFBP7	IGHMBP2	IGLL1	IGSF1	IHH	IKBKB	IKBKG	IKZF1
IL1RAPL1	IL1RN	IL2RA	IL2RG	IL7R	IL10RA	IL10RB	IL11RA	IL12B	IL12RB1	IL17RA	IL17RC	IL17RD	IL21R	IL31RA
IL36RN	ILDR1	IMPA1	IMPAD1	IMPDH1	IMPG1	IMPG2	INF2	INPP5E	INPP5K	INPPL1	INS	INSL3	INSR	INVS
IQCB1	IQSEC2	IRAK4	IRF1	IRF6	IRF8	IRX5	ISCA1	ISCA2	ISCU	ISG15	ISPD	ITCH	ITGA2B	ITGA3
ITGA6	ITGA7	ITGA8	ITGB2	ITGB3	ITGB4	ITGB6	ITK	ITM2B	ITPA	ITPR1	IVD	IYD	JAG1	JAGN1
JAK3	JAM3	JPH2	JUP	KANK1	KANK2	KANSL1	KARS	KAT6A	KAT6B	KATNB1	KBTBD13	KCNA1	KCNA2	KCNA5
KCNB1	KCNC1	KCNC3	KCND3	KCNE1	KCNE2	KCNE3	KCNH1	KCNH2	KCNJ1	KCNJ2	KCNJ5	KCNJ6	KCNJ10	KCNJ11
KCNJ13	KCNK3	KCNK9	KCNMA1	KCNN4	KCNQ1	KCNQ2	KCNQ3	KCNQ4	KCNQ5	KCNT1	KCNV2	KCTD1	KCTD7	KCTD17
KDM1A	KDM5C	KDM6A	KDSR	KERA	KHDC3L	KIAA0556	KIAA0586	KIAA1109	KIDINS220	KIF1A	KIF1B	KIF1BP	KIF1C	KIF2A
KIF5A	KIF5C	KIF7	KIF11	KIF21A	KIF22	KIRREL3	KISS1R	KIT	KITLG	KIZ	KL	KLC2	KLF1	KLF11
KLHL3	KLHL7	KLHL10	KLHL15	KLHL24	KLHL40	KLHL41	KLK4	KLKB1	KLLN	KMT2A	KMT2B	KMT2C	KMT2D	KMT5B
KNL1	KPTN	KRAS	KREMEN1	KRIT1	KRT1	KRT2	KRT3	KRT4	KRT5	KRT6A	KRT6B	KRT6C	KRT8	KRT9
KRT10	KRT12	KRT13	KRT14	KRT16	KRT17	KRT18	KRT25	KRT74	KRT81	KRT83	KRT85	KRT86	KY	KYNU
L1CAM	L2HGDH	LAGE3	LAMA1	LAMA2	LAMA3	LAMA4	LAMB1	LAMB2	LAMB3	LAMC2	LAMC3	LAMP2	LAMTOR2	LARGE1
LARP7	LARS2	LAS1L	LAT	LBR	LCA5	LCAT	LCT	LDB3	LDHA	LDLR	LDLRAP1	LEFTY2	LEMD2	LEMD3
LEP	LEPR	LGI1	LGI4	LHB	LHCGR	LHFPL5	LHX3	LHX4	LIAS	LIFR	LIG1	LIG4	LIM2	LIMS2
LINS1	LIPA	LIPC	LIPE	LIPH	LIPN	LIPT1	LIPT2	LITAF	LMAN1	LMBR1	LMBRD1	LMF1	LMNA	LMNB1
LMOD3	LMX1B	LONP1	LOR	LOX	LOXHD1	LPAR6	LPIN1	LPIN2	LPL	LPP	LRAT	LRBA	LRIG2	LRIT3
LRMDA	LRP2	LRP4	LRP5	LRP6	LRPAP1	LRPPRC	LRRC6	LRSAM1	LRTOMT	LSS	LTBP2	LTBP3	LTBP4	LYRM7
LYST	LYZ	LZTFL1	LZTR1	LZTS1	MAB21L2	MAF	MAFB	MAG	MAGED2	MAGEL2	MAGI2	MAGT1	MAK	MALT1
MAML2	MAMLD1	MAN1B1	MAN2B1	MANBA	MAOA	MAP2K1	MAP2K2	MAP3K1	MAP3K7	MAP3K20	MAPKBP1	MAPRE2	MAPT	MARS
MARS2	MARVELD2	MASP1	MASP2	MAT1A	MATN3	MATR3	MBD5	MBOAT7	MBTPS2	MC2R	MC4R	MCCC1	MCCC2	MCEE
MCFD2	MCM4	MCM6	MCM9	MCOLN1	MCPH1	MDH2	MECOM	MECP2	MECR	MED12	MED13L	MED17	MED23	MED25
MEF2C	MEFV	MEGF8	MEGF10	MEIS2	MEN1	MEOX1	MERTK	MESP2	METTL23	MFAP5	MFF	MFN2	MFRP	MFSD2A
MFSD8	MGAT2	MGME1	MGP	MIB1	MICU1	MID1	MINPP1	MIP	MIPEP	MITF	MKKS	MRTFA	MKRN3	MKS1
MLC1	MLH1	MLH3	MLLT10	MLPH	MLYCD	MMAA	MMAB	MMACHC	MMADHC	MME	MMP1	MMP2	MMP9	MMP13
MMP19	MMP20	MMP21	MN1	MNX1	MOCOS	MOCS1	MOCS2	MOGS	MORC2	MPC1	MPDU1	MPDZ	MPI	MPL
MPLKIP	MPO	MPV17	MPZ	MRAP	MRE11	MRPL3	MRPS16	MRPS22	MRPS34	MS4A1	MSH2	MSH3	MSH6	MSMO1
MSN	MSR1	MSRB3	MSTN	MSTO1	MSX1	MSX2	MTAP	MTFMT	MTHFD1	MTHFR	MTM1	MTMR2	MTO1	MTOR
MTR	MTRR	MTTP	MUC1	MUSK	MUT	MUTYH	MVD	MVK	MXI1	MYBPC1	MYBPC3	MYC	MYCN	MYD88
MYF6	MYH2	MYH3	MYH6	MYH7	MYH8	MYH9	MYH11	MYH14	MYL2	MYL3	MYLK	MYLK2	MYMK	MYO1E
MYO3A	MYO5A	MYO5B	MYO6	MYO7A	MYO15A	MYO18B	MYOC	MYOT	MYOZ2	MYPN	MYT1L	NAA10	NAA15	NACC1
NAGA	NAGLU	NAGS	NALCN	NANOS1	NANS	NARS2	NAXE	NBAS	NBEAL2	NBN	NCF1	NCF2	NCSTN	NDE1
NDN	NDP	NDRG1	NDST1	NDUFA1	NDUFA2	NDUFA9	NDUFA10	NDUFA11	NDUFA12	NDUFAF1	NDUFAF2	NDUFAF3	NDUFAF4	NDUFAF5
NDUFAF6	NDUFB3	NDUFB11	NDUFS1	NDUFS2	NDUFS3	NDUFS4	NDUFS6	NDUFS7	NDUFS8	NDUFV1	NDUFV2	NEB	NECTIN1	NECTIN4
NEDD4L	NEFH	NEFL	NEK1	NEK8	NEK9	NEU1	NEUROD1	NEUROG3	NEXMIF	NEXN	NF1	NF2	NFE2L2	NFIA
NFIX	NFKB1	NFKB2	NFKBIA	NFU1	NGF	NGLY1	NHEJ1	NHLRC1	NHP2	NHS	NIPA1	NIPAL4	NIPBL	NKX2-1
NKX2-5	NKX2-6	NKX3-2	NKX6-2	NLGN4X	NLRC4	NLRP1	NLRP3	NLRP7	NLRP12	NME1	NME8	NMNAT1	NNT	NOBOX
NOD2	NODAL	NOG	NOL3	NONO	NOP10	NOP56	NOTCH1	NOTCH2	NOTCH3	NPC1	NPC2	NPHP1	NPHP3	NPHP4
NPHS1	NPHS2	NPPA	NPR2	NPRL2	NPRL3	NR1H4	NR2E3	NR2F1	NR2F2	NR3C1	NR3C2	NR4A3	NR5A1	NRAS
NR0B1	NR0B2	NRL	NRXN1	NSD1	NSD3	NSDHL	NSMCE2	NSMCE3	NSMF	NSUN2	NT5C2	NT5C3A	NT5E	NTF4
NTHL1	NTRK1	NTRK2	NUBPL	NUP62	NUP93	NUP107	NUS1	NYX	OAT	OBSL1	OCA2	OCLN	OCRL	ODAPH
OFD1	OGT	OPA1	OPA3	OPHN1	OPLAH	OPN1LW	OPN1MW	OPN1SW	OPTN	ORAI1	ORC1	ORC4	ORC6	OSBPL2

OSGEP	OSMR	OSTM1	OTC	OTOA	OTOF	OTOG	OTOGL	OTUD6B	OTULIN	OTX2	OVOL2	OXCT1	P2RX2	P2RY12
P3H1	P3H2	P4HA2	P4HB	PABPN1	PACS1	PADI3	PADI6	PAFAH1B1	PAH	PAK3	PALB2	PAM16	PANK2	PAPSS2
PARK7	PARN	PATL2	PAX2	PAX3	PAX4	PAX6	PAX7	PAX8	PAX9	PBX1	PC	PCARE	PCBD1	PCCA
PCCB	PCDH12	PCDH15	PCDH19	PCNT	PCSK1	PCSK9	PCYT1A	PDCD10	PDE3A	PDE4D	PDE6A	PDE6B	PDE6C	PDE6G
PDE6H	PDE8B	PDE10A	PDE11A	PDGFB	PDGFRB	PDHA1	PDHB	PDHX	PDP1	PDSS1	PDSS2	PDX1	PDYN	PDZD7
PEPD	PER2	PET100	PEX1	PEX2	PEX3	PEX5	PEX6	PEX7	PEX10	PEX12	PEX13	PEX14	PEX16	PEX19
PEX26	PFKM	PFN1	PGAM2	PGAP1	PGAP2	PGAP3	PGK1	PGM1	PGM3	PHEX	PHF6	PHF8	PHGDH	PHKA1
PHKA2	PHKB	PHKG2	PHOX2A	PHOX2B	PHYH	PI4KA	PIBF1	PIEZO1	PIEZO2	PIGA	PIGC	PIGG	PIGL	PIGM
PIGN	PIGO	PIGT	PIGV	PIGY	PIH1D3	PIK3CA	PIK3CD	PIK3R1	PIK3R2	PIK3R5	PIKFYVE	PINK1	PIP5K1C	PITPNM3
PITX1	PITX2	PITX3	PJVK	PKD1	PKD1L1	PKD2	PKHD1	PKLR	PKP1	PKP2	PLA2G4A	PLA2G6	PLA2G7	PLAA
PLAU	PLCB1	PLCB4	PLCD1	PLCE1	PLCG2	PLD1	PLEC	PLEKHG2	PLEKHG5	PLEKHM1	PLG	PLIN1	PLK4	PLN
PLOD1	PLOD2	PLOD3	PLP1	PLPBP	PLPP6	PLS3	PML	PMM2	PMP22	PMPCA	PMS2	PMVK	PNKD	PNKP
PNP	PNPLA1	PNPLA2	PNPLA6	PNPO	PNPT1	POC1A	POC1B	POFUT1	POGLUT1	POGZ	POLA1	POLD1	POLE	POLG
POLG2	POLH	POLR1A	POLR1C	POLR1D	POLR3A	POLR3B	POMC	POMGNT1	POMGNT2	POMK	POMP	POMT1	POMT2	POP1
POR	PORCN	POU1F1	POU3F4	POU4F3	PPA2	PPARG	PPIB	PPM1D	PPOX	PPP1CB	PPP1R3A	PPP1R15B	PPP2R1A	PPP2R1B
PPP2R2B	PPP2R5D	PPP3CA	PPT1	PQBP1	PRCC	PRCD	PRDM5	PRDM6	PRDM12	PRDM16	PRDX1	PRF1	PRG4	PRICKLE1
PRIMPOL	PRKAG2	PRKAR1A	PRKCA	PRKCD	PRKCG	PRKCSH	PRKD1	PRKDC	PRKG1	PRKN	PRKRA	PRLR	PRMT7	PRNP
PROC	PRODH	PROK2	PROKR2	PROM1	PROP1	PROS1	PRPF3	PRPF4	PRPF6	PRPF8	PRPF31	PRPH2	PRPS1	PRRT2
PRRX1	PRSS1	PRSS12	PRSS56	PRUNE1	PRX	PSAP	PSAT1	PSEN1	PSEN2	PSENEN	PSMB8	PSMC3IP	PSMD12	PSPH
PSTPIP1	PTCH1	PTCH2	PTDSS1	PTEN	PTF1A	PTGIS	PTH	PTH1R	PTHLH	PTPN11	PTPRC	PTPRO	PTPRQ	PTRH2
PTS	PUF60	PURA	PUS1	PXDN	PYCR1	PYCR2	PYGL	PYGM	PYROXD1	QARS	QDPR	RAB3GAP1	RAB3GAP2	RAB7A
RAB11B	RAB18	RAB23	RAB27A	RAB28	RAB33B	RAB39B	RAC1	RAC2	RAD21	RAD50	RAD51	RAD51C	RAF1	RAG1
RAG2	RAI1	RAP1GDS1	RAPSN	RARB	RARS	RARS2	RASA1	RAX	RAX2	RBBP8	RBCK1	RBM8A	RBM10	RBM20
RBP4	RBPJ	RCBTB1	RD3	RDH5	RDH12	RDX	RECQL4	REEP1	REEP6	RELN	REN	RERE	REST	RET
RETREG1	RFT1	RFX5	RFX6	RFXANK	RFXAP	RGR	RGS9	RGS9BP	RHAG	RHBDF2	RHCE	RHO	RIMS1	RIN2
RIPK4	RIT1	RLBP1	RLIM	RMND1	RNASEH1	RNASEH2A	RNASEH2B	RNASEH2C	RNASEL	RNASET2	RNF43	RNF125	RNF135	RNF139
RNF168	RNF170	RNF212	RNF216	ROBO2	ROBO3	ROGDI	ROM1	ROR2	RORC	RP1	RP1L1	RP2	RPE65	RPGR
RPGRIP1	RPGRIP1L	RPL5	RPL10	RPL11	RPL21	RPL35A	RPS6KA3	RPS7	RPS10	RPS17	RPS19	RPS23	RPS24	RPS26
RPS28	RPS29	RPSA	RRAS2	RRM2B	RS1	RSPH1	RSPH3	RSPH4A	RSPH9	RSPO1	RSPO4	RSPRY1	RTEL1	RTN2
RTN4IP1	RTTN	RUNX1	RUNX2	RUSC2	RXYLT1	RYR1	RYR2	S1PR2	SACS	SAG	SALL1	SALL4	SAMD9	SAMD9L
SAMHD1	SAR1B	SARS2	SATB2	SBDS	SBF1	SBF2	SC5D	SCARB2	SCARF2	SCN1A	SCN1B	SCN2A	SCN2B	SCN3B
SCN4A	SCN4B	SCN5A	SCN8A	SCN9A	SCN10A	SCN11A	SCNN1A	SCNN1B	SCNN1G	SCO1	SCO2	SCYL1	SDCCAG8	SDHA
SDHAF1	SDHAF2	SDHB	SDHC	SDHD	SDR9C7	SEC23A	SEC23B	SEC24D	SEC61A1	SEC63	SECISBP2	SELENON	SEMA4A	SEPSECS
SEPT9	SEPT12	SERAC1	SERPINA1	SERPINA3	SERPINA6	SERPINB7	SERPINB8	SERPINC1	SERPIND1	SERPINE1	SERPINF1	SERPINF2	SERPING1	SERPINH1
SERPINI1	SETBP1	SETD2	SETD5	SETX	SF3B4	SFRP4	SFTPA2	SFTPB	SFTPC	SFXN4	SGCA	SGCB	SGCD	SGCE
SGCG	SGO1	SGPL1	SGSH	SH2D1A	SH3BP2	SH3PXD2B	SH3TC2	SHANK3	SHH	SHOC2	SHOX	SHROOM4	SI	SIK1
SIL1	SIM1	SIN3A	SIX1	SIX3	SIX5	SIX6	SKI	SKIV2L	SLC1A1	SLC1A2	SLC1A3	SLC1A4	SLC2A1	SLC2A2
SLC2A9	SLC2A10	SLC3A1	SLC4A1	SLC4A4	SLC4A11	SLC5A1	SLC5A2	SLC5A5	SLC5A7	SLC6A1	SLC6A2	SLC6A3	SLC6A5	SLC6A8
SLC6A9	SLC6A17	SLC6A19	SLC6A20	SLC7A7	SLC7A9	SLC7A14	SLC9A3	SLC9A3R1	SLC9A6	SLC10A2	SLC11A2	SLC12A1	SLC12A3	SLC12A5
SLC12A6	SLC13A5	SLC16A1	SLC16A2	SLC16A12	SLC17A5	SLC17A8	SLC17A9	SLC18A3	SLC19A2	SLC19A3	SLC20A2	SLC22A5	SLC22A12	SLC24A1
SLC24A4	SLC24A5	SLC25A1	SLC25A3	SLC25A4	SLC25A12	SLC25A13	SLC25A15	SLC25A19	SLC25A20	SLC25A22	SLC25A24	SLC25A26	SLC25A38	SLC25A46
SLC26A2	SLC26A3	SLC26A4	SLC26A8	SLC27A4	SLC29A3	SLC30A2	SLC30A10	SLC33A1	SLC34A1	SLC34A2	SLC34A3	SLC35A1	SLC35C1	SLC35D1
SLC36A2	SLC37A4	SLC38A8	SLC39A4	SLC39A5	SLC39A8	SLC39A13	SLC39A14	SLC40A1	SLC45A1	SLC45A2	SLC46A1	SLC52A1	SLC52A2	SLC52A3
SLCO1B1	SLCO1B3	SLCO2A1	SLFN14	SLITRK1	SLITRK6	SLURP1	SLX4	SMAD3	SMAD4	SMAD6	SMAD9	SMARCA2	SMARCA4	SMARCAD1
SMARCAL1	SMARCB1	SMARCD2	SMARCE1	SMC1A	SMC3	SMCHD1	SMG9	SMN1	SMOC1	SMOC2	SMPD1	SMPX	SMS	SNAI2

SNAP29	SNCA	SNCB	SNIP1	SNRNP200	SNRPB	SNRPE	SNRPN	SNTA1	SNX10	SNX14	SOBP	SOD1	SOHLH1	SON
SOS1	SOS2	SOST	SOX2	SOX3	SOX5	SOX9	SOX10	SOX11	SOX17	SOX18	SP110	SPAG1	SPARC	SPART
SPAST	SPATA5	SPATA7	SPECC1L	SPEG	SPG7	SPG11	SPG21	SPINK1	SPINK5	SPINT2	SPRED1	SPRTN	SPRY4	SPTA1
SPTAN1	SPTB	SPTBN2	SPTLC1	SPTLC2	SQSTM1	SRCAP	SRD5A2	SRD5A3	SRP72	SRY	SSR4	SSTR5	ST3GAL3	ST3GAL5
ST14	STAC3	STAG1	STAG3	STAMBP	STAR	STAT1	STAT2	STAT3	STAT5B	STIL	STIM1	STK4	STK11	STN1
STOX1	STRA6	STRADA	STRC	STS	STUB1	STX1B	STX11	STX16	STXBP1	STXBP2	SUCLA2	SUCLG1	SUFU	SUGCT
SULT2B1	SUMF1	SUMO1	SUN5	SUOX	SURF1	SYCP3	SYN1	SYNE1	SYNE2	SYNE4	SYNGAP1	SYNJ1	SYP	SYT2
SYT14	SZT2	TAB2	TAC3	TACO1	TACR3	TACSTD2	TAF1	TAF2	TAF6	TAF13	TALDO1	TANGO2	TAP1	TAP2
TAPBP	TAPT1	TARDBP	TAT	TAZ	TBC1D7	TBC1D20	TBC1D23	TBC1D24	TBCD	TBCE	TBCK	TBK1	TBL1XR1	TBP
TBX1	TBX3	TBX4	TBX5	TBX6	TBX15	TBX18	TBX19	TBX20	TBX21	TBX22	TBXAS1	TBXT	TCAP	TCF3
TCF4	TCF12	TCIRG1	TCN2	TCOF1	TCTEX1D2	TCTN1	TCTN2	TCTN3	TDGF1	TDP1	TDP2	TDRD7	TEAD1	TECPR2
TECR	TECRL	TECTA	TEK	TELO2	TENM3	TENM4	TEX11	TF	TFAP2A	TFAP2B	TFE3	TFG	TFR2	TFRC
TG	TGDS	TGFB1	TGFB2	TGFB3	TGFBI	TGFBR1	TGFBR2	TGIF1	TGM1	TGM5	TGM6	TH	THAP1	THBD
THOC2	THOC6	THPO	THRA	THRB	TIA1	TIMM8A	TIMM50	TIMMDC1	TIMP3	TINF2	TJP2	TK2	TKT	TLE6
TLL1	TMC1	TMC6	TMC8	TMCO1	TMEM38B	TMEM43	TMEM67	TMEM70	TMEM98	TMEM107	TMEM126A	TMEM126B	TMEM138	TMEM165
TMEM173	TMEM199	TMEM216	TMEM231	TMEM237	TMEM240	TMEM260	TMIE	TMPRSS3	TMPRSS6	TMPRSS15	TMTC3	TNC	TNFAIP3	TNFRSF1A
TNFRSF10B	TNFRSF11A	TNFRSF11B	TNFRSF13B	TNFRSF13C	TNFSF11	TNIK	TNNC1	TNNI2	TNNI3	TNNT1	TNNT2	TNNT3	TNPO3	TNXB
TOE1	TOP1	TOP2A	TOPORS	TOR1A	TP53	TP53RK	TP63	TPI1	TPK1	TPM1	TPM2	TPM3	TPO	TPP1
TPRKB	TPRN	TRAF3IP1	TRAIP	TRAPPC2	TRAPPC6B	TRAPPC9	TRAPPC11	TRAPPC12	TRDN	TREH	TREM2	TREX1	TRHR	TRIM2
TRIM32	TRIM37	TRIO	TRIOBP	TRIP4	TRIP11	TRIP12	TRIP13	TRIT1	TRMT5	TRMT10A	TRMT10C	TRMU	TRNT1	TRPC6
TRPM1	TRPM4	TRPM6	TRPS1	TRPV3	TRPV4	TSC1	TSC2	TSEN2	TSEN15	TSEN54	TSFM	TSHB	TSHR	TSHZ1
TSPAN7	TSPAN12	TSPEAR	TSPYL1	TTBK2	TTC7A	TTC8	TTC19	TTC21B	TTC25	TTC37	TTI2	TTLL5	TTN	TTPA
TTR	TUBA1A	TUBA4A	TUBA8	TUBB	TUBB1	TUBB2A	TUBB2B	TUBB3	TUBB4A	TUBB4B	TUBB8	TUBG1	TUBGCP4	TUBGCP6
TUFM	TULP1	TUSC3	TWIST1	TWIST2	TWNK	TXNL4A	TYK2	TYMP	TYR	TYROBP	TYRP1	UBA1	UBA5	UBE2A
UBE2T	UBE3A	UBE3B	UBIAD1	UBQLN2	UBR1	UBTF	UCHL1	UFM1	UMOD	UMPS	UNC13D	UNC80	UNG	UPB1
UPF3B	UQCRB	UQCRC2	UQCRQ	UROD	UROS	USB1	USH1C	USH1G	USH2A	USP9X	USP9Y	USP18	USP27X	UVSSA
VAC14	VAMP1	VANGL1	VANGL2	VAPB	VARS	VARS2	VCAN	VCL	VCP	VDR	VEGFC	VHL	VIM	VIPAS39
VKORC1	VLDLR	VMA21	VPS11	VPS13A	VPS13B	VPS13C	VPS33A	VPS33B	VPS37A	VPS45	VPS53	VRK1	VSX1	VSX2
VWF	WAC	WARS	WARS2	WAS	WASHC5	WBP2	WDR11	WDR19	WDR26	WDR34	WDR35	WDR36	WDR45	WDR60
WDR62	WDR72	WDR73	WDR81	WFS1	WHRN	WISP3	WNK1	WNK4	WNT1	WNT4	WNT5A	WNT7A	WNT10A	WNT10B
WRAP53	WRN	WT1	WWOX	XDH	XIAP	XK	XPA	XPC	XPNPEP3	XPR1	XRCC4	XYLT1	XYLT2	YAP1
YARS	YARS2	YWHAG	YY1	YY1AP1	ZAP70	ZBTB16	ZBTB18	ZBTB20	ZBTB24	ZC3H14	ZC4H2	ZDHHC9	ZEB1	ZEB2
ZFP57	ZFPM2	ZFYVE26	ZFYVE27	ZIC1	ZIC2	ZIC3	ZMPSTE24	ZMYND10	ZMYND11	ZNF148	ZNF408	ZNF423	ZNF469	ZNF644
ZNF687	ZNF711	ZNF750	ZNHIT3	ZP1	ZP3	ZSWIM6								

Table S2: List of primers used in this study

Case ID	Gene	Primer	Sequence (5' - 3')
A162	ABCA1	Ex36-F	GCTTTCAATCATCCCCTGAA
		Ex37-F	CCAGCCAGCTTTGTCGTATT
		Ex38-F	CTGCTTCCAGCAGAAGTCCT
		In38-R	CTGCCCTCCTTCTGACACTG
		Ex39-R	GTGAACAGCTCCAGCACAAA
		Ex40-R	ATTGCCTGGTTTTTCACCAT
A078	ABCA12	Ex39-F	TCCCTTCCAGCTTACCTCAA
		Ex41-F	CATCAAACCAAAGCCAAACA
		Ex42-F	CATCAAACCAAAGCCAAACA
		In42-F	TCAGCCTCCCAAATAACCAG
		Ex43-R	CCAGCAAGTACATCCAGGAAA
		Ex43-R2	CCACTGACAGGGAAACAATG
		Ex44-R	AACATGGTGCCCTGAGAAAC
		Ex45-R	GCAGGTATCCCAATGCTGAT
A002	ABCB4	Ex3-F	GAAGACAGTGAAAATGATTGGAGT
		Ex4-F	TGGCCATAGCTCACGGATCA
		In5-R	ACGAGTGCTGTCTGTATGCT
		Ex6-R	CACTGAACTCAATACGCGGC
		Ex7-R	GATGGAAGCTCACCCTTGTGA
		Ex14-F	GAAGCAGAGGATCGCCATTG
		Ex15-F	GTCATCGCTGGGTTTGAGGA
		In16-R	AGGCACAGATGACTGCTACT
		Ex17-R	CGTGGGAACAGTATGTGCCAT
		Ex18-R	GCGATGATGCAGTGAAGCAG
A009	ACE	Ex9-F	TTCCCGCTACAACTTCGACT
		Ex10-F	AACGAAACCCACTTTGATGC
		In11-R	ATCAGAGCAGCCAGACACCT
		Ex13-R	CTCAGCCTCATCAGTCACCA
		Ex14-R	CCGTACTTCAGGGTGTGGTT
		Ex10/11-F	TGTGACACCATACATCAGGTACT
		Ex10/12-F	TGACACCATACATCAGGAAGG
		Ex8-F	CTCGGTCTCCACTCCTGAAC
		In9-R	AGAGGCTGGACCAGAAGCTA
		In9-R2	GTGGGAGTCAGGGAAGATCA
		Ex18c/19-R	CAGCCTGCAGCACCTTCCTT
A018	AHI1	Ex16-F	GCCATATTGGTCCGACAGTT
		Ex17-F	AGGGGTGATTGTTGTTTGGA
		Ex18-F	ATTCATCCCAATGGAAAACG
		In18-R	TCCACCTTTTTCAGGAAGGA
		Ex19-R	TCTCCCGATAATTTGCTGCT
		Ex20-R	CCAAATGCACAGAATGCAAC
		Ex8-F	GCTCAATGGAACAAAGCACA
		Ex10-R	CCTCGAAGCAAATAGGGAAA
		Ex17/18-F	GCACCACTGGACTATAAATAAGGAA
		Ex18/19-F	TGGATCTCCGGATATTAGTAGCA
		Ex17/19-F	AAATAAGATTAGTAGCAAGGAAGTTTG
		Ex18c/19-F	TCACAGAATTAGTAGCAAGGAAGTTT
		Ex21-R	GGGCATCTTGACTTTGGTGT
A006	ANO10	Ex5-F	CCTGGTACACTCGGTTTGCT
		Ex6-F	GACATGGAGGTTTGGGCCTT
		In7-R	AGGTTCATGTCTTCTGGTCCAG
		Ex8-R	GCGCAAAAGCTTCATATCTTTCA
		Ex8-R2	GTGAGGCAAAGCAATTGAGGA
A129	APC	Ex2-F	TGGCTTCTTCTGGACAGATTG
		Ex3-F	TCAAAAATGTCCCTCCGTTC
		Ex4-F	TGCTTCTTGCTGATCTTGACA
		In4-F	TTGGCAGTACAACTTATTTGAAACTT
		Ex5-R	CTCTGATTTGCCTTGCTTCA
		Ex6-R	TGTCCTTTTCGATTTGCTGA
A213	ARL2BP	Ex2-F	TGCAGAATTTGATGCTGTGG
		Ex3-F	CCTGGAGTTTGAAGACACAGAAG
		Ex4-F	AGTTCAACATGGCAGCCTTC
		In4-F	AAGATCAGGCATTGGCTCAC
		In4-R	GAGGGCTTCCTCCATAGGAC
		Ex5-R	GTGAAGGTGAGCAGCATGTC
		Ex6-R	TCATTGGCTGGAGGTAGGAC
		Ex10-F	CGCAGATCGAACTACTGCTG
		Ex11-F	TGTCTCTGCCACCAGAAATG
		Ex12-F	CGACCAAAAGAAGCCTTCAG
		In12-R	TGACAGCTCTGCATCCAAAC
A058	ASNS	Ex13-R	AAATTTCTGGGCTGCATTTG
		Ex13-2R	CCACTTGGGCATCCAGTAAT
		3'UTR-R	CCCATCCAACACGAAGAAAT

Case ID	Gene	Primer	Sequence (5' - 3')
A050	HSD17B4	Ex1-F	ACCAGAAATCGGCAAGTCAC
		Ex2-R	TAACGCTCCTCTTTCTGCAA
		Ex3-R	GAAGTCCCCTCCCAAATCAT
		Ex13-F	ATGGGTGGAGGATTAGCAGA
		Ex14-F	TAAACCACTTCCCAGAGCAG
		Ex15-F	TGTGAAGCAGTTGTTGCTGA
		In15-R	TGTCATACACAAAGCCTATCAATC
		Ex16-R	TGTCTGATGTCCGTTTTCCA
		Ex17-R	CAGCATCAGGAGGGTCTATTAGG
		Ex18-R	GGGATTCCAGTCTCCACTGA
A085	HSPG2	Ex14/16-F	ACCACTTCCCAGAGCAG TCT
		Ex37-F	ACTGCGAGCTATGTGAATGC
		Ex38-F	TTTGTGCCCCTGGCTACTAC
		Ex39-F	GCTACACTGGCCAGTACTGT
		In39-F	GTGGAGGACAAATGGTGGAC
		Ex39/40-F	GGCCAGTACTGTGAGCAGTGT
		Ex40-R	CCCTTGCACACTGGGGTTAC
		Ex41-R	CATCCTCACGGGACCAATAG
		Ex42-R	GCAAATGTAGACCCCAGCAT
		Ex39c-R	TACTGTGTGGCCCAGGTTAC
A077	IBA57	Ex1-F	CTAGGGCTGCTGACCAATGA
		Ex1-2F	ACAGCTCCTGCAGTCCTGGT
		Ex1a-F	CAGAGCTCTTTGCGTTGGTC
		Ex2-F	CACCTCGCGCTATACAGGAT
		In2-R	ATTGAACATCGTCCAGCTCC
		In2/Ex3-F	TCCAATCCCTGCAGGCGTT
		Ex2/3-R	CAAGGCGTTCCTGAGGGGGT
		Ex3-R	GTAGCAGCCTTTGGTGAAGC
		Ex3-2R	CAAGGCTTCGGACTACTTGG
		Ex4-F	ATGAGGGCACCTACATCTGC
		Ex5-F	GTTTACCCACCCGCTACCTC
		Ex/In6-F	GGAGCTGATAGCTAGAGGTAGGAC
		Ex7-F	CTGGAGCCAGTACCGGATTA
		Ex7-R	CTCTGCAAGCTCACATCCAG
A069	IL11RA	Ex8-F	GCCTGCGGGTAGAGTCAGTA
		Ex/In8-F	GTCCACAGTGAGGCCTGGA
		In8-F	GATTTCACGATCCTGGGTGT
		Ex9-R	GGCACTGACTCGTACAGCAT
		Ex10-R	AAGTAGCCGAGGGTGTGGTT
		Ex11-R	CACCAGTCCCAGGAAAGAAA
		Ex11-F	AAATGCACATGCAGCTCTTC
		Ex12-F	TGGGATCAAGTTGCTGTCTG
A180	IQSEC2	Ex13-F	GAAAGGTATGATGCGGCCTA
		Ex13-R	AGAGAGGTCTCGCAGGGAAC
		Ex14-R	AGCGATCCCACGCTGTTAC
		Ex15-R	CCAGGAAGGATGAGGAGTTG
		Ex3-F	AGGTGGTGATGGAGAAGGACAT
		Ex4-F	TGACAGCCATGGACAACCAC
		Ex6-R	CTTGTACTCAGGCAGCACGTC
		Ex7-R	CAGTCTTCAGCAGCCCGAT
A113 & A195	KCNH2	In5-F	GGTGGGAGTGTCTTGGTCTT
		In5-R2	AGGTCTGCCTCCCTCCAAGAG
		Ex6F	CTGCTCATCTTCGGCTCTGG
		Ex7F	CGGAAGCTGGATCGCTACTC
		Ex8F	CCAACGGCATCGACATGAAC
		Ex9F	CGCCCTGTACTTCATCTCCC
		In9F	TGTCCTCTCCATGGCCTCC
		In9R	CACTGCACCCTTATAAGCAATGT
		Ex8R	CTGGTGGAAGCGGATGAACT
		Ex8R2	TCGATGCCGTTGGTGTAGGA
		Ex9aR2	TAAAGGAGCCCAGTGACCCT
		Ex9aR3	CAGAGGGCATTTCCAGTCCA
		Ex10R	GTACATGTCCAGCACCTCCA
		Ex11R	CCCTCTAACTCCGTACTGCC
		Ex12R	GCCCTCATCCTCACTGCTC
A079	LAMP2	Ex4-F	TTGGCCATCAGAATTCCATT
		Ex5-F	GCTGCAGCTGAACATCACTC
		Ex6-F	GCCGTTCTCACACTGCTCTA
		Ex7-R	TGCTGATGTTCACTTCCTTCA
		In7-R	CAGGCCAGTGCTTTGCTAAC
		Ex8-R	TCACATTGAAAGGCTGAACC
		Ex9-R	CCTGAAAGACCAGCACCAAC
		Ex2-F	TGGGTCTCTTTGGATAACTGG

Left column

ID	Gene	Primer	Sequence
		Ex12/13-F	GGAATACGTTGAACATCAGGTT
		Ex11/13-F	CAGAAATGAGAATTCCAAAGGTT
		Ex17-F	CCTACACCCTGACCAGCAAT
		Ex18-F	ATGAAGAGACAGCCCAGGAA
		Ex19-F	TGTGATCCTGGCAGAAAATG
		In19-F	GTGTCTCTGCCCACCGTAAG
		In19-2F	CCACTAACTGACGTCCCGTG
A091	ATP1A3	Ex20-R	CGACAACGATGCTCACAAAG
		Ex20-2R	GGGTCTTGCAGATGATCAGA
		Ex21-R	CTCAAACAGCCCGAAGATCA
		Ex22-R	CAGGATGAGTTTGCGGATTT
		Ex23-R	GTGGGGCTGAGGTCAGTAGT
		Ex19/20c-F	TTACGGGCAGCAGTGGAGGA
		Ex2-F	TTTGTGTACGGCCAGGACT
		Ex3-F	TAAAAGCACCAACCCCTACG
		Ex4-F	CAGCGTGTATGGACCAGATG
		In4-F	TTAGGTTCAGGCTGGGAATG
		Ex5-R	AAACATGGGGATGGTCCTTT
		1UTR-R	TTTTCCTCTGCTCCCTCAGA
A017	B9D1	Ex6-F	TAGTGCTCCTGGTTGTTTCC
		Ex6-R	GGAAACAACCAGGAGCACTA
		Ex7-F	TCCCCCATACAGGTTACAGC
		Ex7-R	GGCAGCCCTTCATTATCAGA
		Ex7-F2	CGAGTACACAGACCCCAAGG
		Ex7-R2	TACACAGACCCCAAGGTGGT
		Ex8-R	ACCACAGGCTAGAGGCTTCA
		Ex9-R	AAAGAAGGGTGTGTGGGATG
		Ex2-F	GATGACAAGACCCCCATGTC
		Ex3-F	CCTGGGGTTCATCTTCCATA
		Ex4-F	CTGCAGGAACCCACTTCAAT
		Ex4-R	TATCCCTGACACGCTGCTTCA
		In4-R	GGATCCAGCTGAGGACACAT
		Ex5-R	AGGATGGCAAAGAAGAGCAG
		Ex6-R	TCAGCACAGCAAAAGGATG
A053 & A155	CACNA1E	Ex31-F	GTGTGGCACTTTGTGGTGTC
		Ex32-F	TCCTGAAGGTCATCGCTTTT
		In32-F	TGGGAAACTGAAGCCAAAGT
		Ex33-F	TCACCGTGATTGGCAGTATC
		Ex33-R	TCTGTGATACTGCCAATCACG
		Ex34-R	GACGCAGGAGCTTTATGAGG
		Ex35-R	TGGCAATTAAAAGGCAGACA
		Ex35a-R	CTTTGTGCTTGGTGCTGTGT
		Ex36-R	CCGGTTGATGTGACTCTCCT
		Ex32/34-F	CGCTTTTGGCTTTTTGCTGG
		Ex33/34-R	CACTGGTGTTCACCAGCTTG
		Ex4-F	GGCTAACCACCTCCCCTAAG
		Ex5-F	GATGGGAAAACGACCAGAGA
		In5-F	TGCTGTCCCAGAGCCTTAGT
A070	CAMTA1	Ex6-F	GGCTGCTATGTCCATTCCTC
		In6-R	TGTGTCAGGAAGGGGAGAAG
		Ex7-R	TTCTTGTCGGTGTTGATGGA
		Ex8-R	GAGAAGCCTGAGCTGCTGTT
		Ex9-R	TGTTTGGCGCTGTTACACTT
		Ex5-F	CAAGGAAATGGTGTCCGAAA
		Ex6-F	ACTTCCAGAACAGGCTTTGC
		Ex6/8-F	ATTGCAAGCAGCGAGGCCAG
A097	CC2D2A	Ex7-F	TGGAGACTGAATTTGGCACA
		In7-F	GCAGAGCTATGCACAGAGGA
		Ex8-R	TTTTCGTGCAGAATCATGGA
		Ex9-R	CTTTCCTCCTCCTTGTGCAG
		Ex2-F	AGATGCAATGGTCCAAGGTC
		Ex3-F	GCAGATGAATGGAACAGCAA
		Ex3/4-R	GTGCTGCTGTTTTCCACC
		Ex3/5-R	GAGTTCCATTATCCTCCACC
A146	CD96	Ex4-F	ACGGATTCTTGGGTCCTTCT
		Ex4-R	AAGGAATCCGTGCTGCTGTTT
		In4-R	CCGTTCTTTAGCTGAATCTGG
		Ex5-R	TGGTGGAGCTCCTCAAGATT
		Ex6-R	ACCAAGACATCCGTGGAGTT
		Ex11-F	CCACTCTTCATCCAGGTGGT
		Ex13-F	CCCACCACTTCATCATCTCC
		In13-F	TAGATCCCACCTTCCACTGC
		Ex13/14-F	CCGCTACGACTTTGATCTCT
		Ex13/14cF	CCGCTACGACTTTGATGTGA
		Ex13/14cF2	CCGCTACGACTTTGATCCAG

Right column

ID	Gene	Primer	Sequence
		Ex3-F	TCTTGAGGGAGTGGATACGG
		Ex4-F	CAGCAGAGCAAAGCTCACAC
		Ex4-F2	TGCCATGTCAACTTTGTTCG
		Ex/In4-F	CGGAAGGTAATGGTCTCTCAA
A239	MANBA	In4-R	AACCCCACTAAGGGATCAGG
		Ex4/5-F	TCAACTTTGTTCGGAAGGAG
		Ex4c-F	ACTTGTGCAGAAGGGAGC
		Ex5-R	GATTCCCTGGGTAGGAAAGG
		Ex6-R	CAATCCTTTTCCCAGGTTGA
		Ex7-R	AAGCCTCCATCCAGTTCAAA
		Ex4-F	CCAAGAAGGGCTGTAAACGA
		Ex5-F	AGGAAGAGGAGCAGGAAA
		Ex6-F	TACGTCCCTGCTGAAGTGTG
		Ex/In6-R	CTGACAATAATGTCACCTGACA
		In6-F	TGTCAGGTTTGTTTCCCACT
		Ex7-R	TTCATCCACAAGCACCAGAG
		Ex8-R	CTATTGCCGTCCCATCAAAT
		Ex9-R	CTATTGCCGTCCCATCAAAT
A087	MSH6	Ex4-F	CCAAGAAGGGCTGTAAACGA
		Ex5-F	AGGAAGAGGAGCAGGAAA
		Ex6-F	TACGTCCCTGCTGAAGTGTG
		Ex/In6-R	CTGACAATAATGTCACCTGACA
		In6-F	TGTCAGGTTTGTTTCCCACT
		In6-R	TTTGCTTCTTGTATATGAGCCTTTT
		Ex7-R	TTCATCCACAAGCACCAGAG
		Ex8-R	CTATTGCCGTCCCATCAAAT
		Ex9-R	CCTCTGGGAGATTAGCAAGC
		Ex5-F	TTGGCTGCTGTTCAAGATCA
		Ex8-R	CTGCAGACCACTGTGCACTT
		Ex14-F	TGGTTTGAACCATCACCTCA
		Ex16-F	TGACTTCCTCCACTGCCTCT
A026	NBAS	Ex41-F	GATTTCACTCAGGCGCAGAT
		Ex43-R	GGTGAAAGGCTTCTGGATCA
		Ex19-R	AGTTCCTTTTGTTCCAGTGTCA
		Ex21-R	TTCGGGCAGCAAAACAGAAT
		Ex26-R	TCCTCTTTCTGGCAGACATTCT
		Ex7-F	CCACTGATCCCCAAGTCTGT
		Ex8-F	AGAGCTGTGGCCAGTGTACC
A085	NDUFV1	Ex9-F	AACAAGGTGATGGCACGTTT
		In9-R	GGGGACCATGTGAAACAGAG
		Ex10-R	ATGGATAGACGCAGGACAGC
		Ex10-R2	GGCAGCACTCGCTTTATTGT
		5'UTR-F	CGTGGAAAGGATCCCACTTC
		Ex1/2-F	CCGCTTCGACGAGCCGCTTC
		Ex15-F	TGGAATCCTGATGCTCCTGT
		Ex16-F	TGGTTGCGGGAAATATTGAT
		Ex17-F	CTCTCTCCGGAAGGGAAAG
		Ex18-R	TTCAGTGTCAGGGTTCCACA
		Ex19R	TCAGCAGTGCCATCACTCTT
		Ex1-F	CTCCACAGACCCTCTCCTTG
		Ex2-R	GTGAGGCCGCTTATAACCAA
		Ex3-R	CCCAGCAAGACATTTTTCCA
		Ex4-R	TTTCTGGCAGCAACTGTTTG
		In1-2F	GGACAGAGTAGGTGAGGGGA
		In16-F	CCTTCAAGTTGGGGCATAGA
		In1-F	GGGGGTGGGGACAGAGTAG
		5UTR-F	CGTGGAAAGGATCCCACTTC
A094 & A251	NF1	Ex1-F	CTCCACAGACCCTCTCCTTG
		Ex1/2-F	CCGCTTCGACGAGCCGCTTC
		In1-F	GGGGGTGGGGACAGAGTAG
		In1-2F	GGACAGAGTAGGTGAGGGGA
		In1-R	CTGGAGGAAAGGAAGGGAAG
		Ex2-F	TTGGTTATAAGCGGCCTCAC
		Ex2-R	GTGAGGCCGCTTATAACCAA
		Ex3-F	TATTTGGAGAAGCTGCTGAA
		Ex3-R	CCCAGCAAGACATTTTTCCA
		Ex4-F	GTCAAACAGTTGCTGCCAGA
		Ex4-R	TTTCTGGCAGCAACTGTTTG
		Ex5-F	TGGATTGTGCAAAATTAAAACG
		Ex8-R	TTGTAGTGGCCAAACTGCTG
		Ex9-F	GCTGACAGAAAGTGCTGCAA
		Ex12-F	CACACCCAGCAATACGAATG
		Ex17-F	CGTACTCCTGGAGCCTCTCT
		Ex19-R	TCAGCAGTGCCATCACTCTT
		Ex20-R	TCTTCCATTTTGGCTTTTGG

ID	Gene	Primer	Sequence
A134	CDH23	Ex14-R	CACATGGTCAGGCACACTCT
		Ex14-2R	TGGCTCCTTTTCCATATTGC
		Ex15-R	TGCCTGCATCATTGTCAGTT
		Ex16-R	ACGTTGTCGTTGACATCCAA
		Ex46-F	CTGTCCTCCAACCAGACCAT
		Ex47-F	ACCCTCACTGTCCATCTGCT
		Ex48-F	CACAGCCATTGTCACCATTC
		Ex49-R	ACCTCATAGGTGGCAACCAG
		Ex50-R	CTTGAACTGGGGCGTATTGT
A005	CFTR	Ex10-F	GGTGATGACAGCCTCTTCTTC
		Ex12-R	CTTGCTCGTTGACCTCCACT
		Ex14-R	TTAGCCATCAGTTTACAGACACAG
		Ex19-F	TCCAAACCTCACAGCAACTC
		Ex20-F	GCTGGTTCCAAATGAGAATAG
		Ex21-F	TTGCAGTGGGCTGTAAACTC
		Ex23-R	TTGAATCCCAAGACACACCA
		Ex24-R	CCACTGTTCATAGGGATCCA
A137	CHD7	Ex9-F	TAACAGCCTTCTGGGAGGAG
		Ex24-F	TGTCCGCATGCTGTACTACC
		Ex25-F	TTGGTGGGATAAGGAAGCAG
		Ex25/26-F	TGGAGTGTTCAAACATGAGC
		In25-F	TGTGGCAGTGCTGTGATTTTG
		In25-R	GCACTGCCACAACACAAGTT
		In26c-F	TGGAGTGTTCAAACATGAGC
		In26c-F2	TGGAGTGTTCAAACATGAGC
		Ex26-R	CGACTCGTTCCAGAAAGCAC
		Ex27-R	TTTGAACGGTGTTCTGGTTG
		Ex28-R	GGCATGTATTTCCATGGATTC
A093	CHD8	Ex14-F	ACCGTCACAATTTCCCTCAG
		Ex15-F	TCGGGCTATTTTGGAGAAGA
		In15-F	TGGTTCACTGGATCTCAACA
		Ex16-R	ATTCCGTGAAGCTTGCCATA
		Ex16-F	GCTTAAAGCTGGTGGCCATA
		Ex17-R	CGAGTTAGAGGCAACCTTCG
		Ex18-R	AAGGTGTACCGCCTCATCAC
		Ex15/In15c-F	CCACCCATATCTCATCAATGTTAAA
		Ex15/16c-F	CCATATCTCATCAATGGCCA
A200	CIC	Ex1-F	ATGTATTCGGCCCACAGG
		Ex2-F	GCACTCCTTAGTCCCCTTCC
		Ex3-F	AGGGAGTGCTGACCCTGAG
		Ex3-R	AGGGTCAGCACTCCCTGTC
		In2-R	GGCAGCAGGGACAGTGAG
		Ex4-R	CAGAAGATGAGTCCCGTTCC
		Ex5-R	GTTGTCCTGGTTGGGATGAC
A001	CLN5	Ex1-F	CGGGGAGGTGTCATGCG
		Ex1-F2	CAGGGCGCCGAGATGC
		Ex1-F3	TGGCTCGCGGTGGTTC
		In1-F	ACCGCGGAACATTTGGTACA
		In1-R	CTTGTACCCAGAAGTCCCTCC
		In1-R2	TGAGGGCAAAACTGGGAAGT
		Ex2-R	TTCCCATACTGGGGCTTGTA
		Ex3-R	TTTCAGGTCGGAGATGGAAG
A141	COL4A1	Ex40-F	CAGCAGGAGAGAAGGGTGAA
		Ex41-F	GAAGAGGATTCCCAGGGTTT
		Ex42-F	TGGATCCAAAGGAGAGCAAG
		In42-R	ATTCTGTCCCAGTCCTCAGC
		Ex43-F	CAGCCTGGATTTCCTTTGTC
		Ex44-F	TCCTGGAGAGCCACCAATAC
		Ex19-F	CAGGAAAGGGGAGATGAAGG
		Ex21-F	TGGTTCACCAGGTCTTCCAG
		Ex22-F	GTGACATCGTTTTTCGCAAG
		In22-R	CTCAGGGGGAAAAATCAGTG
		Ex23-R	ATGTAACCCTGGCAATCCTG
		Ex24-R	TGGAAATCCTGGAAGACCTG
		Ex32-F	GACCACTGGGTCAAAGAGGA
		Ex33-F	ATGATGGGCTTTCCTGGAG
		Ex34-R	CCTCTCTGGCCCTTTACTCC
		Ex34-F	TCCCACGTAATAGGGGACAA
A120	COL4A3	In34-F	TTGCTTGTTTGGAATCAGGA
		In34-R	TCCTGATTCCAAACAAGCAA
		Ex35-F	CCTTTGAGGCCCTTTAAACC
		Ex36-F	ATACCACGCAGTCCTGGTTC
		Ex37-R	CCTGGCCTCTCCCTTATCTCC
		Ex32-R	AAGGTGAAATGGGACCACTG
		Ex33-F	GATTGGGATGATGGGCTTTC
		Ex5-F	TTCCGGAAGGAACAAATAGG
		Ex6-F	ACTGGATCGGCACTTACACC
		Ex7-F	CCAGGAAGTTCAGGAGTCCA
		Ex7/8-F	AGCCTGTCTTGGCCTTTCTT
		In8-R	TTCAATCGGAATGGAAAGAC
		Ex8/9-R	TTCTTGTATTCGTGTAAACTGAGTTGG
		Ex8/9-F	CAACTCAGTTTACAGAATACAAGAAATC
		Ex9-R	GCTGTCCTGGAAGTTTGCAT
		Ex10-F	CCAAAACACTGACCATGACG
A024 & A157	OPHN1	Ex10-R	CGTCATGGTCAGTGTTTTGG
		Ex11-F	GGGCCCTTGGACTTAACACT
		Ex11-R	ACTGGATCGGCACTTACACC
		Ex12/16-R	CTGGCATTTAGGATAGGT
		In14-R	GAGTTGTCATCTCCCCTTGG
		Ex14/15-F	ATTATTGAGACCAAAGGGAT
		Ex14/16-F	ATTATTGAGACCAAAGATCC
		Ex14/15c-F	ATTATTGAGACCAAAGATC
		Ex15-F	GAGTTGTCATCTCCCCTTGG
		Ex15-R	ACAGTGCGGTACAACCCTTC
		Ex17-R	GCAGAGACCAGCTCTTTGTG
		Ex18-R	TTCCAGCATCTCTCGGTTCT
		Ex16-R	GCACTTGCTTACCACGTTGA
		Ex15-R	CACCACTTTCCTTGAAGACAATC
A031	PGAP1	Ex6-F	CAGACCACCTCTCCATTGTG
		Ex8-F	CAAAACATTTTGAGGAAAATCCA
		Ex9-F	TGTCCAAATGGACCTATGTAGC
		Ex11/12-R	ATCAACCCCTTGCAGGCA
		Ex11/In11c-F	TCAACCCCTTGCAGGCTC
		Ex12-F	ATGGAAAGCTGAACTGCTGC
		Ex14-R	AAAAAGATGAGTTACAGGAAGCTG
		Ex11/12-F	AAACAGCACTTCTATGTGCC
		Ex11/11c-R	AACAGCACTTCTATGTGAGC
		In11-R	AAATGAACATGCATTCAAAAGA
A040	PIGN	Ex8-F	TCATCATGCCAGAAACAACC
		Ex9-F	CCTCTGACCATGGAATGACA
		Ex10-F	GGCTGGTCATCCTTCAGAGA
		In10-R	TGAACGATACAATCAGACACAAGAT
		Ex11-F	GCCTCTTCCAATTCTCCAATC
		Ex12-R	GGAAGTCATCAATGGTGCAA
		Ex13-R	AACATGCTCTCTGCTTTGAAGA
		Ex14-R	AACTTCTTTCTTCTGAGTCATTTTCA
		Ex10/11-R	AAGAGTGGAGATTGGAGAATT
		Ex10/12-R	AAGGCTGATATTGCACCATTG
		Ex10/13-R	AAGGGAATCCTTCCTGTGGAT
		Ex10/11-F	TGATGCATTTTTGAAAGAGTGG
		Ex10/12-F	GATGCATTTTTGAAAGGCTGA
		Ex10/13-R	TGATGCATTTTTGAAAGGGAAT
		Ex16-R	TAACAACATTGACGCCCAAA
		In10-F	CCTTCCTCAGCCTCTGATGT
		In11-F	AAGCAATACAACCATGAAAAGC
		Ex10/In10c-F	TGATGATGCATTTTTGAAAGATCT
		Ex11-F	GACGTGGTGGTGGAAAACAG
		Ex12-F	GTCATTTATCAGACGCGGGC
		Ex12/14-F	GGTCTTCAAGCTCTCAGTGGACC
		Ex12/14-R	TCCCCAAAGGTCCACTGAGA
A118	PKD1	Ex14-R	GGGAAGGACTCGTTGTACGG
		Ex15-R	CCCGAAGTCCCACGTGTAAA
		In13-F	TCACGGGGGTTGCTTTTCTGA
		In13-R	CAAACCGGCCCCCCGAG
		Ex13-R	AGGAAGGCCACCTCCAC
A103	PLP1	Ex3-F	GGCCACTGGATTGTGTTTCT
		Ex4-F	GGCGACTACAAGACCACCAT
		Ex5-R	AGGTGGTCCAGGTGTTGAAG
		In5-F	GTGCTTTGGCTCTCCTACCC
		In5/Ex6-R	TCCATGGGAGAACACCCTAA
		Ex6-R	GAAAGCATTCCATGGGAGAA
		Ex7-R	ACAGGTGGAAGGTCATTTGG
		Ex8-R	CAGCATTGTAGGCTGTGTGG
A029	PPP2R5D	Ex4-F	TTTACCCTGAGGCTGTCACC
		Ex5-F	ACCTTCATCGAATCCCACAG
		Ex6-F	CCTGATTTCCAGCCAAACAT
		Ex7-R	CCCCAAAAACTTGCCATAGA
		Ex8-R	TGTGATGCTCCGTCTCGTAG
		Ex9-R	GGGAAGTAGGACACGGATGA
		Ex1-F	ATAGGGATGGGGGTGCTATC

ID	Gene	Primer	Sequence
		Ex34-F	TCCCACGTAATAGGGGACAA
		In34-R	TCCTGATTCCAAACAAGCAA
		Ex36-R	ATACCACGCAGTCCTGGTTC
		Ex37-R	CCTGGCTCTCCCTTATCTCC
		Ex13-F	CAGAACCAGTTCCCGTCATC
		Ex14-F	TCTGTGGCTACCCCTCAGTC
		Ex15-F	CTGACGTACCTGTGCTGGAA
		In15-R	GTTGCGATACGCAGTCAATG
		Ex16-R	TGCGAAGGAGATGTTGACTG
		Ex17-R	TGCCGGAAAGGTAATGACTC
A251	CREBBP	Ex22-F	TTCGTTGATTGCAAGGAGTG
		Ex23-F	ACTGGCAGACCTCGAAAAGA
		In24-F	GCAGTGGCCAGAGTTAGAGG
		In24-R	AAGGCTGTTGCTGACAAGGT
		Ex25-R	GAGCCGTATTCTTGGACGTG
		Ex27-R	TCCTTCACTTGGAGGACAGG
		Ex28-R	CACTGGTGAGCCTGTCTTCA
		Ex2-F	AAAATCCAGCGTGGACAATG
		Ex2a-F	CCTTTTGCTCCATTTTCTGC
		Ex3-F	TGGATACCTCCCAAGTCCTG
A240	CTNNB1	In3-F	CCTTACTGAAAGTCAGAATGCAG
		Ex4-R	AGCCAAACGCTGGACATTAG
		Ex5-R	AACGCATGATAGCGTGTCTG
		Ex6-R	AGCAAGGCAACCATTTTCTG
		Ex8-F	GCCACAACAGACAGCTTTGA
		Ex9-F	CAGCTTACAGGGTGGGTGTC
		In9-R	GCAGCAGCCATTTCTTTCAT
A060	DFNA5 (GSDME)	Ex10-R	TCTGGAGTTTGCAGCAAGTG
		Ex11-R	TGCCTAAAGCACAGAGTCCA
		Ex8/10-F	GGTCCTGGAACCAGTGAAAT
		Ex7-F	AGGTGGCTTCGAGAACAAGA
		In9-F	GCTGCTTTAAAATGCTTTGCCT
		Ex13-2F	CTCACTCACATGGTGGTGGT
		Ex14-F	TCGATGGGCAAACATCTGTA
		Ex14/15-F	GTCTTACTGAAGAACAGTGCC
A208	DMD	Ex15-F	TGCATGGCTTTCAGAAAAAG
		In15-R	TGGGTTTTTATAAGACCATTGAAA
		Ex16-R	CAACACCGGGCAAAGTTATC
		Ex17-R	CCCTTGTGGTCACCGTAGTT
		Ex3-F	CAGAGCTCCCATGGAGACATT
		Ex4-F	CCCTCGTCAGCAAGACAGAA
		Ex5-R	TTGCCGACAATTGCACTTCC
A168	DNAJB11	Ex6-R	CCAGCGTTCGTTCTTCATTCA
		In4-F	TCTTGGGTGGATATGCTTGAGA
		In4-F2	TGGAGGAGACCACCAGGTAG
		In4-R	CTCAAGCATATCCACCCAAGAGA
		In4-R2	CAAGTCAGTTGCATCTTCTGTGA
		Ex53-F	ATTATCGTTCCCTGGGAGGT
		Ex55-F	AGCTGGATGATGCTTTCCAC
		Ex55-R	AAAACAGTGAAGGGCTGGTG
		Ex56-F	GCTGGAAATGACCTTGGAGA
A108	DYSF	In56-F	CATTGGACCTGCTGTTGAGA
		Ex57-F	CTGTTCCTGGCCATCTTCAT
		In57-R	TGAGGATACAGAGCCCAGCT
		Ex58-R	GGACTCTCCTGCCCTGTAGA
		3'UTR-R	TCACCCCACTTCCATCATTT
		Ex4-F	AGGTCCTCCTGGTCCTCAAG
		Ex5-F	AAGCTGGAACTCGAGAAAACC
		Ex6-F	CCAAGGGTCAGCAATTCAAG
		Ex7-F	TTCAGGTGGAGTGCTCAATG
A165	EDA	Ex7/8-R	CAGTGAAGTTGATGTAGTATACTTCT
		Ex7/8c-R	CAGTGAAGTTGATGTAGTATACATGA
		In7-R	CTTTCAACTCCCTCCCAGTG
		Ex8-R	ATCCACCACCACCTCATAGC
		Ex8-R2	ATCTTGACGGCGATCTTCTG
		5UTR-F	CGCGCTCTGAAAGTTTATGAC
		Ex1-F	CTCCTTTTCGGGCTCACAG
A158	EDN3	Ex2-F	CTGAGGGGGACTGTGAAGAG
		In2-R	CAGGCTCTGGGCTAACTGAG
		Ex3-R	GGCAGGCCTTGTCATATCTC
		Ex4-R	TTCCCTTCCTCTTCTTTGTCTG
		Ex1-F	GATACCGAGCTGACCACCTT
		Ex2-F	TCGAGTACGAGACCCAGAGG
		Ex3-F	TTCGACTAGAGGGGATGCAG
		In3-F	CACTCCAGCAGCCTTAGGAG

ID	Gene	Primer	Sequence
		Ex1-2F	TGCAGATCGAGTTCAAATGC
		Ex/In1-F	GAGTGAATGGCTTCCAGTCC
		Ex1a-F	CCAGTCTGAGGGAGAAGCTG
A088	PRPH2	Ex1b-F	CAGCTGTGTCCCTTAGGTG
		In1-F	CAGTCCTGGGGTCAGAGCTA
		Ex2-R	CTCCGTCTGGTGGTCGTAAC
		Ex3-R	CTGGCTCTCGCTCTCAGATT
		3'UTR-R	ATCCACGTTTCTTGGAGTGC
		Ex2-F	AGCTGCTGCAATCCTAGCTC
		Ex2-F2	TGCCCTGCTGAGCTACTACA
		Ex1-F	CTCGTAAAGGACCGCAATGT
		Ex1-F2	CGCAGCAGCACTACTATGAGAAG
		Ex3-F	ATGGAGGAGCTGGAGGAAAT
A085	PYGM	Ex5-F	TCACGCTACCTGTGCACTTC
		Ex5-F2	CTGTGCACTTCTACGGCCAT
		Ex/In5-F	GGTGGACACACAAGTGAGGA
		Ex6-R	TGAAGTCATTGGGAGCCTTG
		Ex6-R2	GCGCATGGTGTTGACAACAT
		Ex7-R	CATTGGGGTACAGGACACGA
		Ex8-R	GACTTGAAGCGACGGATGAT
		Ex8-F	GATGTACAGAACGCGCTTGA
		Ex9-F	GCAAGAGATTTGGCTTTGGA
		Ex9/11-F	GGGCATTTCTTCAAAAAGTGC
A025	SETD5	In9-R	TATTTGGAGTCATGGGGTGG
		Ex10-F	GCCCGTACTTTCGGTAATGA
		Ex10-R	GATGCCCGTACTTTCGGTAA
		Ex11-R	ACACATGATTGCAGATGGGA
		Ex12-R	GAAAAGCACGACGGAAAGAG
NMD inhibition control	SRSF1	Ex1-F	ACATCGACCTCAAGAATCFCCGC
		Ex3-F	CACTGGTGTCGTGGAGTTTGTACGG
		3'UTR-R	GGGCAGGAATCCACTCCTATG
		Ex5-F	CTTTCAGGCCACCATAGAGC
		Ex6-F	TTAGGAATGTGGACAGCAACC
		In8-F	TGAACTAATTTAATATTTGCTCTTGTG
		Ex9-F	AAAGCAGTAGCTGCAGAATCG
A154 & A237	SPAST	Ex9-R	AAGGTTGCATTCGATTCTGC
		Ex10-F	TGGCTCGAGAACTTCAACCT
		Ex12-F	GCTTGATGAGGCTGTTCTCA
		Ex14-F	GGAAGTCCATTGACCCAAAA
		Ex16-R	TGGCAGACATATTCTTCACCTG
		Ex17-R	GCTGACGCTGCGTTTTATTT
		5UTR-R	TGTTTCTGTAGCCGATGACG
		Ex10-F	CCCCAAAGTAAAGGAGAGCA
		Ex11-F	CAGCGCCATTTTAAACAACA
		Ex12-F	AGTTGATCCCACCACAGAGC
A014	SPG11	Ex14-R	AGTTGATCCCACCACAGAGC
		Ex15-R	TGCTGGGTTTGAAATTCTCC
		Ex29-F	AGATGGACAATTCGCTTTGG
		Ex30-R	GCTCTGCAGTGCAATACCAA
		Ex9-F	CCTGTGCTCCATGAATTGAC
		Ex12-F	ACCTGTCCCAGATGCTGAAG
		Ex13-F	ACAAACTCTGCCGAGTGGAG
		Ex13/14-F	AACTCTGCCGAGTGGAGCAG
A164	STXBP1	Ex13/15-R	TTCCTCCGTGATGCCTGCT
		Ex14-F	ATTCTGCTGGATGCCAATGT
		In14-R	CTCAGTTTCCCCAAGACAGG
		Ex15-R	TGAGCCATGTTGGTGATGAT
		Ex16-R	CACCGTGAGAGCTGGTAGGT
		Ex14c/Ex15-F	GCTGGATGCCAATGCATC
		5'UTR-F	CTCCCCAGTGACGAGAGAGC
		Ex1-F	CACCTACAGCTGCTTCTGGA
		Ex1-2F	TGCTTCTGGACCAGTGAGTG
		Ex1/2-F	CAGCTGCTTCTGGACCAAGT
		Ex1/3-F	AGCTGCTTCTGGACCAGGAT
		Ex2-F	CACAACAGGGAGGTGCTGTA
		Ex2-R	TGATTGGGACACGGTGATGAG
A022	TAZ	In2-R	AAGGGGTTTGTTCTGACGAG
		Ex2/3-F	ACCCTCATCTCTGGCGGAT
		Ex3-R	ACGCATCAACTTCAGGTTCC
		Ex4-R	GTGGGAGTGTAGCTCCTTGG
		Ex1/2-F	CAGCTGCTTCTGGACCAAGT
		Ex2/3-F	CTCTGGGGGATCCTGAAACT
		Ex1/3-F	AGCTGCTTCTGGACCAGGAT
		Ex2c/3-F	GGAGGAAAGGGATCCTGAAA
		Ex2/In2c-R	AAGGGATCCTGAAACTCCGC

ID	Gene	Primer	Sequence
A205	EMD	Ex3/4-F	TTACTCTACCAGAGCAAGGGCTA
		Ex4-F	ACTTATGGGGAGCCCGAGTC
		Ex4-R	GATGGAAAGCGTCAGCATCT
		Ex5-R	AGACAAAAGATCGTCATCATGC
		Ex6-R	TGATGCTCTGGTAGGCACTG
		Ex6-2R	GGTGAGCCATGAAGAGGAAG
		Ex13-F	TGATGCTGAGTGCAGAGGAG
		Ex15-R	GGAGACTCAAAAGGCACGAG
		Ex16-F	AAGTACCGCTCCCTCCTCAC
		Ex17-F	TCAGCTGGGGACATTACTGAG
		Ex18-F	CATGGTGTTTGAGCATACGG
		Ex18-R	ATGGCCTTTTCAACATCCTG
A206	FANCA	In18-R	CAGAAATGGGACACACTCCA
		Ex18/19-R	CCAGCATATTCAGGAGGCCT
		In18-2R	TGTTTCACAGCAAATGCACA
		Ex19-R	GGAGGAAGTGGGACACGTAG
		Ex20-R	TATAAACGCCACACGGGAGT
		Ex21-R	CTTCTGGCTTCTCTTCAGCAG
		In18c/Ex19-R	GGGCATATTCAGGAGGCCTT
A236	FMR1	Ex2-F	TTCATGAAGATTCAATAACAGTTGC
		Ex3-F	GCAGCCTGATAGGCAGATTC
		Ex3/4-F	GAAAGTGATGAAGTTGAGGTGT
		Ex4-F	AAAGAGCCTTGCTGTTGGTG
		Ex4-2F	GCCTTGCTGTTGGTGGTTAG
		In4-R	TCCACACGGGAATGTACAAC
		Ex5-R	CTTTTGTGGCAGGTTTGTTG
		Ex6-R	ACAGAAAAGGCACCAACTGC
A219	FXN	Ex1/2-R	CCACGTTGGTTCGAACTTGC
		Ex1/3-R	GTGGTCTCATCTAGAGAGTGC
		Ex1/4-R	GACAGTTAAGACACCACTTGC
		Ex1-F	GGAGCAGCATGTGGGACTCTC
		In1-F	ATGGAGCACTCGGTTACAGG
		Ex2-F	TGAGGAAATCTGGAACTTTGG
		Ex3-R	GGCTTGTCTGCAAGGTCTTC
		Ex4-R	AGCCAGATTTGCTTGTTTGG
A036	GAA	Ex1-F	AGCTGACGGGGAAACTGAG
		Ex2-F	ATCCAGCTAACAGGCGCTAC
		Ex2-R	CAGCTGTAGCTGGGGTAGC
		Ex3-F	ATCCAGCTAACAGGCGCTAC
		Ex4-F	TGCCCTCGCAGTATATCACA
		Ex4-R	ATCCAGCTAACAGGCGCTAC
		Ex5-F	ATCCAGCTAACAGGCGCTAC
		Ex5-R	GCTGTTTAGCAGGAACACCC
		Ex6-R	CCTGTCGACCTCCAGCTAAG
		Ex7-R	GGTCATGTTCTCCACCACCT
		Ex8-R	CCATCCTTGTTGAACGTGAA
		In3-F	AGGCCTGGAAAGGTTCTGTT
		In4-R	CTGGGACATGCAGGAGACG
		In5-R	AGGCATCAGCCCATAGGAC
		WTF	CCACCTCTAGGTTCTCCTCGT
		WTR	GGCCTGGACAGCTCCTACA
		Ex7-F	GTGAGGAGCGATCTCAGCAG
		Ex8-F	CATGGCCCTACCAGACACTG
		Ex8/9-F	GCGTGATGATCTAGCCTCCA
		Ex8/9-R	GGAGGCTAGATCATCACGCT
A220	GALT	In8-F	GTTGCTCCCAGTAGGGTCAG
		In8-R	GGTCCAGATGCTGACCCTAC
		Ex9-F	TTTCCCTACTCCATGGGCTG
		Ex10-R	AGCCTGAGCAAGCATTTCGT
		Ex11-R	TGCCCCAGGTGGTAATGAAC
Loading control	GAPDH	Ex3-F	TCACCAGGGCTGCTTTTAAC
		Ex6-R	GGCAGAGATGATGACCCTTT
A064	GLI3	Ex2-F	CGAACAGATGTGAGCGAGAA
		Ex3-F	CAACATCGAGTGACGAGAGG
		Ex4-F	CCTTCCATCCTCCTGTACCA
		In4-R	CAGCATCTCGTTCCATTTCA
		Ex5-R	TGAGATCATGGAGAGCGATG
		Ex6-R	TATTCTGCTGGGCTGACTCC
A063	GYS1	Ex12-F	GCTCCCTGTGGACTATGAGG
		Ex13-F	AGTGCACGGTTATGGGAATC
		In13-R	CAGAAAACATCCTGGGGAGA
		Ex14-F	CTCTCCGACCTTCTGGACTG
		Ex14-R	CAGTCCAGAAGGTCGGAGAG
		Ex15-R	GGCTCGTAGGTGAAGTGCTC
		Ex16-R	CTCCTCGTCCTCATCGTAGC
A198	TCF12	Ex5-R	CATGTGCCTGCCTGTGTCTA
		Ex13-F	CCAGTTGGATCACCTTCACC
		Ex14-F	CCTTCATCCCCAAGCTATGA
		Ex15-F	TGCAAGATGCAATGTCCTTC
		In15-F	TCAGGTGGTTGCTTTTGAAG
		Ex15/16-F	TAAAGGATGTCTGTGAGCAG
		Ex15/In15c-F	TAAAGGATGTCTGTGAGagC
		Ex15/17-R	TAAAGGATGTCTGTGAGGTT
		Ex16-F	CAGTCTCGAATGGAGGATCG
		Ex17-R	AGGACTGAATGATTGCCATTG
		Ex18-R	TCCAGACTGACTTTGCAAGC
		Ex18-2R	TTTGGGAGGATTCATCATCTG
A089	TRPM6	Ex10-F	CCAACGTCATCCTGTCAGTG
		Ex11-F	AGTCCAAGCACCTTTTCCAA
		Ex12-F	AAGAGCAGCAAGACCTGGAC
		Ex13-F	GGTGGACATTGCCAAGAAAC
		In13-R	TCCGAATAACTTCATTTTCCCTA
		Ex14-R	CTTCCAGTCGAGGGATGGTA
		Ex15-R	TGGACGAGATGATGCAAGAG
		Ex16-R	TGCTGCGATATGCTCTACCA
		Ex28-F	AGATCGGCCCAGAGTAGTGA
		Ex29a-R	GCTGGTTTAAGCCACAGGTC
		Ex30-R	CTGTTTTCTCCTGGCTCTGG
		Ex31-R	TGATTTGTATAGCACATTGTCCA
A100	TUBA1A	5SS1-R	GGAGATGCACTCACGAGCG
		SS1-F2	GAAGCAGCAACCATGCGT
		5SS2-R	TGGAGATGCACTCACGGAAA
		SS2F	CCTTTAAGCTCTTTTCTTTCCGT
		5SS3-R	TGGAGATGCACTCACGCATG
		SS3F	ATTCCCGGGAAAACATGCGT
		5SS4-R	GGAGATGCACTCACGCTGC
		5UTR-F	GGTCTGGACCAACAGGAAAA
		5UTR-F2	GCAACAACCTCTCCTCTTCG
		Ex1-R	CCCCAGGCTCTACTTGGAAA
		Ex1-F2	CGAGGGTGGTGGTATGTTTT
		Ex1a-R	CAGACCACACCCATATGCAG
		Ex2-R	ATTGGGGGAGGAGATGATTC
		Ex3-R	ACACCATTGGCAAGGAGATC
		Ex4-R	GCAAGAAGTCCAAGCTGGAG
		Ex4-R2	GGATGGAGTTGTAGGGCTCA
A188	TUSC3	Ex1-F	TTTCCCTTCCTTCCTGCT
		Ex2-F	ATGGAATGGAGTTCCAGACG
		Ex3-F	CTGGCGCTATTCATCTGCTT
		In3-R	CCCAGGTTCAGACGATGAAT
		Ex4-R	TCCGTTCTGTCAGCAATCC
		Ex5-R	AAACCTCCAACAAGCGACAC
		Ex6-R	GTCCACGGATATGGTTCCAC
		Ex5-F	TGAGGGAGGTGTTTCCAAAG
		Ex6-F	CAGGAAGCTAATGGGGAAAA
		In6-R	CTATGGTGGCCTCAATTTACC
A054	UBE3A	Ex7-F	TCTTTTTGGTAACCCAATGATG
		Ex7-R	TGGAGTATGAAGGGAATGTGG
		Ex8-R	CAGTTCAAGGCTTTTCGGAG
		Ex9-R	ATGACGGTGGCTATACCAGG
A066	VPS13D	Ex7-F	TCGATTGCACTTTACTGGGG
		Ex8-F	GGAGCCTGTGTTTGCATCTG
		Ex9-F	AGGAGAGGCAGGTGAAGTTC
		In9-R	CATGACCACACTGAGAAGCC
		Ex10-R	GCAACGTTTCCTCTGCTCTC
		Ex11-R	TCCTCAAAGCTCTGTTCCTCT
		Ex12-R	TGGAGATACTGCAGCATCCC
		Ex8/9-F	CCTTGAAACTCTCTCAGCTGC
		Ex8/10-R	ACCATTCTCGGCAGCTGA
		Ex50-F	TTCTCCCTGGATGGTGGTAG
		Ex52-R	ACTTCAAAGCCTCCAGACCA
A143	WDR45	Ex8-F	GGGACACAAGTGTGGGAGTC
		Ex9-F	GTCTGCTCCATTCACGATCA
		Ex10-F	CACGACTCCTCCTTCCTCTG
		In10-R	CACTGGCTAAGCCCAGGTAT
		Ex11-R	GAGTCCACGTACTGCCCAAT
		Ex12-R	GCCTCTCTGTTGCAGTTTCC
		3UTR-R	ATTAATGCTTGCTGGCTGGT
		Ex10c/Ex11-F	GCTTCCAGTGATAAGGGCTG

Table S3: RNA-seq sample information metrics

Sample ID	Individual	Affected status	Tissue	RIN	Total number of reads (paired-end)	Alignment Information (% Bases)				
						Aligned Reads (%)	Coding	UTR	Intronic	Intergenic
A009-ACE	Father	Unaffected heterozygote	Urothelia	9.5	54,225,459	97.81	47.56	35.27	14.80	2.37
A014-SPG11	Proband	Affected	Blood	7.6	36,893,595	96.50	38.66	38.07	19.14	4.13
A018-AHI1	Proband	Affected	Blood	8.6	37,113,904	97.41	34.71	39.42	20.76	5.11
A022-TAZ	Proband	Affected	Blood	8.3	Library prep failed					
A031-PGAP1	Proband	Affected	Fibroblasts	9.4	33,992,246	96.48	43.07	40.40	14.10	2.43
A031-PGAP1	Proband	Affected	Blood	8.9	Library prep failed					
A036-GAA	Proband	Affected	Blood	9.2	Library prep failed					
A040-PIGN	Proband	Affected	Liver	6.8	36,055,870	95.20	45.53	33.40	17.67	3.40
A050-HSD17B4	Proband	Affected	EBV-LCL	9.8	30,621,719	92.61	48.93	31.28	16.42	3.38
A058-ASNS	Sibling	Affected	Fibroblasts	9.5	58,039,163	98.04	46.90	36.65	14.25	2.20
A058-ASNS	Proband	Affected	Blood	9	Library prep failed					
A066-VPS13D	Proband	Affected	Blood	8.8	48,668,769	96.82	38.04	34.37	22.68	4.90
A069-IL11RA	Proband	Affected	Blood	8	49,046,071	97.31	44.41	34.42	16.91	4.26
A077-IBA57	Proband	Affected	Blood	8.5	46,482,693	97.81	47.25	33.55	15.74	3.46
A078-ABCA12	Proband	Affected	EBV-LCL	9.5	37,814,562	98.30	50.37	32.68	14.40	2.55
A079-LAMP2	Proband	Affected	Blood	8.4	46,174,801	95.90	39.64	37.08	16.65	6.63
A079-LAMP2	Sibling	Affected	Blood	8.5	45,176,837	92.91	38.94	38.15	18.40	4.50
A081-ARMC4	Proband	Affected	Blood	7.8	51,108,224	97.13	40.35	38.19	17.51	3.94
A089-TRPM6	Proband	Affected	Blood	8.3	45,747,546	96.82	50.57	32.27	13.05	4.12
D19-1532	Control	Unaffected	Urothelia	9.5	57,208,270	97.75	47.72	35.40	14.70	2.17

Illumina TruSeq Stranded mRNA Kit with Poly(A) + enrichment using oligo-dT coated beads
RNA-Seq Alignment, Version 2.0.2

STAR (Aligner)	STAR_2.6.1a
Reference Genome	GRCh37/hg19
Annotation Source	Refseq

Sample ID	Individual	Affected status	Tissue	RIN	Total number of reads (single-end)	Alignment Information (% Bases)				
						Aligned Reads (%)	Coding	UTR	Intronic	Intergenic
A001-CLN5	Proband	Affected	Blood	9.3	53,502,502	98.64	30.43	17.89	35.57	16.11
A002-ABCB4	Proband	Affected	Blood	8.6	52,011,956	98.67	39.37	16.77	29.74	11.12
A006-ANO10	Proband	Affected	Blood	9.7	49,892,314	98.37	29.28	24.55	36.27	9.89

Illumina TruSeq Stranded Total RNA Kit with RiboZero beads to deplete rRNA
RNA-Seq Alignment, Version 1.1.0

STAR (Aligner)	STAR_2.5.0b
Reference Genome	GRCh37/hg19
Annotation Source	Refseq

Table S4: Phenotypic summary of RNA diagnostics cohort

Patient	Year of birth	Age at testing	Gene	Phenotype MIM #	Phenotype	Inheritance	Trio/duo/singleton	Phenotypic concordance
A001	1988	31 y	CLN5	# 256731	Ceroid lipofuscinosis, neuronal 5	AR	Trio	Yes
A002	2007	10 y	ABCB4	# 602347	Cholestasis, progressive familial intrahepatic 3	AR	Quad	Yes
A005	2011	7 y	CFTR	# 219700	Cystic fibrosis	AR	Singleton	Yes
A006	1969	48 y	ANO10	# 613728	Spinocerebellar ataxia, autosomal recessive 10	AR	Singleton	Yes
A009	Fetal death @38 weeks M 29 y, F 40 y		ACE	# 267430	Renal tubular dysgenesis	AR	Duo	Yes
A014	1974	43 y	SPG11	# 604360	Spastic paraplegia 11, autosomal recessive	AR	Singleton	Yes
A018	2014	5 y	AHI1	# 608629	Joubert syndrome 3	AR	Trio	Yes
A022	2017	7 mo	TAZ	# 302060	Barth syndrome	XLR	Duo	Yes
A024	2013	5 y	OPHN1	# 300486	Mental retardation with cerebellar hypoplasia and distinctive facial appearance	XLR	Duo	Yes
A025	2015	4 y	SETD5	# 615761	Mental retardation, autosomal dominant 23	AD	Singleton	Somewhat
A026	2018	8 mo	NBAS		Osteogenesis imperfecta	AR	Singleton	Uncertain
A029	1992	28 y	PPP2R5D	# 616355	Mental retardation, autosomal dominant 35	AD	Trio	Yes
A031	2011	7 y	PGAP1	# 615802	Mental retardation, autosomal recessive 42	AR	Duo	Yes
A036	1989	29 y	GAA	# 232300	Glycogen storage disease II	AR	Singleton	Yes
A040	No live births M 36 y, F 39 y		PIGN	# 614080	Multiple congenital anomalies-hypotonia-seizures syndrome 1	AR	Trio	Yes
A050	2012	6 y	HSD17B4	# 233400	Perrault syndrome 1	AR	Singleton	Yes
A053	2002	16 y	CACNA1E	# 618285	Epileptic encephalopathy, early infantile, 69	AD	Singleton	Yes
A054	2012	7 y	UBE3A	# 105830	Angelman syndrome	AD	Duo	Yes
A058	2019	1 mo	ASNS	# 615574	Asparagine synthetase deficiency	AR	Quad	Yes
A060	1967	52 y	GSDME	# 600994	Deafness, autosomal dominant 5	AD	Singleton	Yes
A063	2014	4 y	GYS1	# 611556	Glycogen storage disease 0, muscle	AR	Trio	Somewhat
A064	1991	28 y	GLI3	# 174700	Polydactyly, preaxial IV	AD	Singleton	Yes
A066	2003	15 y	VPS13D	# 607317	Spinocerebellar ataxia, autosomal recessive 4	AR	Trio	Somewhat
A069	2014	4 y	IL11RA	# 614188	Craniosynostosis and dental anomalies	AR	Duo	Yes
A070	2013	6 y	CAMTA1	# 614746	Cerebellar ataxia, non-progressive, with mental retardation	AD	Singleton	Uncertain
A077	2012	7 y	IBA57	# 616451	Spastic paraplegia 74, autosomal recessive	AR	Trio	Yes
A078	2017	2 y	ABCA12	# 242500	Ichthyosis, congenital, autosomal recessive 4B (harlequin)	AR	Singleton	Yes
A079	2017	2 y	LAMP2	# 300257	Danon disease	XLD	Duo	Yes
			NDUFV1	# 618255	Mitochondrial complex I deficiency, nuclear type 4	AR		
			PYGM	# 232600	McArdle disease	AR		
A085	1977 M 42 y / 1983 F 36 y		HSPG2	# 224410 / # 255800	Dyssegmental dysplasia, Silverman-Handmaker type / Schwartz-Jampel syndrome, type 1	AR	Duo	Carrier screening
A087	1957	62 y	MSH6	# 120435	Lynch Syndrome	AD	Singleton	Yes
A088	1955	64 y	PRPH2	# 169150	Macular dystrophy, patterned 1	AD	Singleton	Yes
A089	2018	1 y	TRPM6	# 602014	Hypomagnesemia 1, intestinal	AR	Trio	Yes
A091	1973	46 y	ATP1A3	# 128235	Dystonia 12	AD	Singleton	Yes
A093	1999	20 y	CHD8	# 615032	Autism, susceptibility to, 18	AD	Singleton	Somewhat
A094	1985	34 y	NF1	# 162200	Neurofibromatosis, type 1	AD	Singleton	Yes
A097	Fetus M 29 y, F 31 y		CC2D2A	# 612285	Joubert syndrome 9	AR	Duo	Somewhat
A100	2009	10 y	TUBA1A	# 611603	Lissencephaly 3	AD	Singleton	Yes
A103	1976 F 43 y		PLP1	# 312080	Pelizaeus-Merzbacher disease	XLR	Singleton	Yes
A107	1994	25 y	B9D1	# 617120	Joubert syndrome 27	AR	Singleton	Yes
A108	1997	21 y	DYSF	# 253601	Muscular dystrophy, linb-girdle, autosomal recessive 2	AR	Trio	Yes
A113	2007	12 y	KCNH2	# 613688	Long QT syndrome 2	AD	Duo	Yes
A118	1952	68 y	PKD1	# 173900	Polycystic kidney disease 1	AD	Singleton	Yes
A120	1954	66 y	COL4A3	# 104200	Alport syndrome 3, autosomal dominant	AD	Singleton	Yes
A129	1973	46 y	APC	# 175100	Adenomatous polyposis coli	AD	Singleton	Yes
A134	2018	1 y	CDH23	# 601386	Deafness, autosomal recessive 12	AR	Trio	Yes
A137	2018	3 y	CHD7	# 214800	CHARGE syndrome	AD	Trio	Yes
A141	2017	3 y	COL4A1	# 175780	Brain small vessel disease with or without ocular anomalies	AD	Trio	Somewhat
A143	2016	2 y	WDR45	# 300894	Neurodegeneration with brain iron accumulation 5	XLD	Singleton	Uncertain
A146	2009	11 y	CD96	# 211750	C syndrome	AD	Duo	Somewhat
A154	1960	60 y	SPAST	# 182601	Spastic paraplegia 4, autosomal dominant	AD	Singleton	Yes
A155	2019	1 y	CACNA1E	# 618285	Epileptic encephalopathy, early infantile, 69	AD	Trio	Yes
A157	2008	12 y	OPHN1	# 300486	Mental retardation with cerebellar hypoplasia and distinctive facial appearance	XLR	Duo	Yes
A158	2017	3 y	EDN3	# 613712	Hirschsprung disease, susceptibility to, 4	AD	Trio	Somewhat
A162	1970	50 y	ABCA1	# 604091	HDL deficiency, familial, 1	AD	Singleton	Yes
A164	2018	2 y	STXBP1	# 612164	Epileptic encephalopathy, early infantile, 4	AD	Singleton	Somewhat

A165	2004 17 y	EDA	# 305100	Ectodermal dysplasia 1, hypohidrotic, X-linked	AD	Singleton	Yes
A168	2005 15 y	DNAJB11		Nephronophthisis	AR	Trio	Yes
A180	2011 9 y	IQSEC2	# 309530	Mental retardation, X-linked 1/78	XLD	Duo	Somewhat
A188	1993 27 y	TUSC3	# 611093	Mental retardation, autosomal recessive 7	AR	Trio	Somewhat
A195	1997 23 y	KCNH2	# 613688	Long QT syndrome 2	AD	Singleton	Yes
A198	2016 4 y	TCF12	# 615314	Craniosynostosis 3	AD	Singleton	Yes
A200	2012 8 y	CIC	# 617600	Mental retardation, autosomal dominant 45	AD	Singleton	Somewhat
A205	1970 49 y	EMD	# 310300	Emery-Dreifuss muscular dystrophy 1, X-linked	XLR	Singleton	Yes
A206	1991 29 y	FANCA	# 227650	Fanconi anemia, complementation group A	AR	Singleton	Yes
A208	2006 14 y	DMD	# 310200	Duchenne muscular dystrophy	XLR	Singleton	Yes
A213	1994 15 y	ARL2BP	# 615434	Retinitis pigmentosa with or without situs inversus	AR	Singleton	Uncertain
A219	1988 32 y	FXN	# 229300	Friedreich ataxia	AR	Singleton	Yes
A220	2014 6 y	GALT	# 230400	Galactosemia	AR	Trio	Yes
A236	2012 8 y	FMR1	# 300624	Fragile X syndrome	XLD	Singleton	Somewhat
A237	2006 14 y	SPAST	# 182601	Spastic paraplegia 4, autosomal dominant	AD	Trio	Somewhat
A239	1985 35 y	MANBA	# 248510	Mannosidosis, beta	AR	Singleton	Carrier screening
A240	2012 9 y	CTNNB1	# 615075	Neurodevelopmental disorder with spastic diplegia and visual defects	AD	Singleton	Somewhat
A251	2020 4 mo	CREBBP	# 180849	Rubinstein-Taybi syndrome 1	AD	Trio	Uncertain
A255	1968 52 y	NF1	# 162200	Neurofibromatosis, type 1	AD	Singleton	Somewhat

Table S5: Cost of testing

Reagent costs of RT-PCR analysis					
Item	**Supplier**	**Cat#**		**Total cost (Singleton)**	**Total cost (Trio)**
PAXgene Blood RNA Tube	BD Biosciences	762165		$16.18	$48.53
PAXgene Blood RNA Kit	Qiagen	762174		$71.10	$71.10
RNase AWAY™ Decontamination Reagent	Invitrogen	10328011		$35.00	$35.00
SuperScript™ IV First-Strand Synthesis System	Invitrogen	18091200		$13.25	$39.75
MasterAmp™ 2X PCR PreMix D 5ml	Astral Scientific	MO7205D		$14.48	$21.72
Taq DNA Polymerase, recombinant	Invitrogen	10342020		$30.40	$45.60
Agarose 250g	Astral Scientific	BIOD0012-250g		$6.97	$9.29
Biotium GelRed Nucleic Acid Gel Stain	Gene Target Solutions	41003		$15.00	$20.00
HyperLadder™ 100bp	Bioline	BIO-33030		$16.70	$22.27
5x DNA Loading Buffer Blue	Bioline	BIO-37045		$2.37	$3.55
GeneJET Gel Extraction Kit	Thermo Scientific	K0691		$50.00	$80.00
1.5ml Graduated Microfuge Tubes	Interpath	121000		$1.28	$2.05
8 Strip PCR Tubes	Interpath	324500		$22.00	$34.00
Neptune Barrier Tips	Pathtech	NEPBT10XLS3/20/200/1250		$55.00	$110.00
Primers	Sigma Aldrich	VC00021		$40.00	$40.00
Sanger sequencing	Australian Genome Research Facility			$398.00	$636.80
Total reagent cost				**$803.73**	**$1,235.67**

Time costs	**Total hours (Singleton)**	**Total hours (Trio)**
In silico analysis and primer design	3	4
RNA extraction/cDNA synthesis	3	3
PCRs/Gels	3	4
Gel extractions/sequencing reactions	3	5
Sequencing analysis	3	5
Report (Hospital Scientist)	8	10
Report (Senior Hospital Scientist)	1	1
Report (Principal Scientist)	1	1
Total hours	25	33
Total time cost	**$1,019.70**	**$1,327.30**
Total RT-PCR analysis cost (AUD)	**$1,823.43**	**$2,562.97**

Assuming number of:	Singleton	Trio
40 µl cDNA synthesis reactions	1	3
PCR reactions	32	48
Agarose gels	6	8
Agraose gel extractions	20	32
Sanger sequencing reactions	40	64
Primers	8	8

Cost of RNA-seq				50M PE reads		100M PE reads	
Item	Supplier	Cat#		Total cost (Singleton)	Total cost (Trio)	Total cost (Singleton	Total cost (Trio)
PAXgene Blood RNA Tube	BD Biosciences	762165		$16.18	$48.53	$16.18	$48.53
PAXgene Blood RNA Kit	Qiagen	762174		$22.38	$67.14	$22.38	$67.14
RNase AWAY™ Decontamination Reagent	Invitrogen	10328011		$35.00	$35.00	$35.00	$35.00
Neptune Barrier Tips	Pathtech	NEPBT10XLS3/20/200/1250		$55.00	$55.00	$55.00	$55.00
Total RNA Library Prep	Australian Genome Research Facility			$243.00	$729.00	$243.00	$729.00
NovaSeq S4 Lane, 300 cycle	Australian Genome Research Facility			$140.22	$420.66	$280.44	$841.32
Total reagent cost				**$511.78**	**$1,355.33**	**$652.00**	**$1,775.99**
Time costs				Singleton analysis time (hours)		Trio analysis time (hours)	
In silico analysis					2		2
RNA extraction					3		3
Adapter trimming					3		3
Alignment/sorting/indexing					4		4
Sequencing analysis					2		3
Report (Hospital Scientist)					6		8
Report (Senior Hospital Scientist)					1		1
Report (Principal Scientist)					1		1
Total					22		25
Total time cost				**$991.65**		**$1,106.67**	
				50M PE reads		100M PE reads	
				Singleton	Trio	Singleton	Trio
Total RNA-seq analysis cost (AUD)				**$1,503.43**	**$2,346.98**	**$1,758.67**	**$2,882.66**

Assuming S4 flow cell	**Number of samples per flow cell**
50M PE reads	50
100M PE reads	25

Table S6: List of genetic variants, splicing outcomes, and changes in classification

Case ID	Gene	Inheritance	Variant(s)	Zygosity	Splicing outcome(s)	Reading frame	CHX sensitivity	Classification Before	After	Tissue(s) tested	Published as case report
A001	CLN5	AR	Chr13(GRCh37):g.77566411G>A NM_006493.2:c.320+5G>A p.?	Hom	Cryptic donor (r.320_321ins[320+1_320+581]; p.Arg108*) Cryptic donor (r.160_320del; p.Val54Alafs*3)	Frameshift/PTC Frameshift/PTC		3	4	Whole blood	
A002	ABCB4	AR	Chr7(GRCh37):g.87083848_87083849insAA NM_000443.3:c.344+2_344+3insTT p.?	Het	Exon skipping (r.287_344del; p.Val96Aspfs*48)	Frameshift/PTC		3	4		
			Chr7(GRCh37):g.87056063T>A NM_000443.3:c.2064+3A>T p.?	Het	Exon skipping (r.1894_2064del; p.Thr632_Leu688del)	In-frame		3	4	Whole blood	
A005	CFTR	AR	Chr7(GRCh37):g.117267573C>A NM_000492.3:c.3469-3C>A p.?	Het	Unable to amplify in blood			3	N/A		
			Chr7(GRCh37):g.117199646_117199648del NM_000492.3:c.1521_1523del p.(Phe508del)	Het						Whole blood	
A006	ANO10	AR	Chr3(GRCh37):g.43616365T>C NM_018075.3:c.1163-9A>G p.?	Hom	Exon skipping (r.1163_1218del; p.Glu388Valfs*69) Cryptic donor (r.1162_1163ins[1163-8_1163-1]; p.Glu388Valfs*3)	Frameshift/PTC Frameshift/PTC		3	5	Whole blood	
A009	ACE	AR	Chr17(GRCh37):g.61561337G>C NM_000789.3:c.1709+5G>C p.?	Hom	Exon skipping (r.1587_1709del; p.Tyr530_Arg570del) Cryptic 'GC' donor (r.1693_1709del; p.Ala565Glufs*64)	In-frame Frameshift/PTC	CHX insensitive CHX sensitive	3	4	Whole blood Fibroblasts Urothelial cells	
A014	SPG11	AR	Chr15(GRCh37):g.44914558G>C NM_025137.3:c.2317-13C>G p.?	Het	Exon skipping (r.2317_2444del; p.Glu774Leufs*21) Cryptic acceptor (r.2317_2356del; p.Val773Argfs*5)	Frameshift/PTC Frameshift/PTC	CHX sensitive CHX sensitive	3	4		
			Chr15(GRCh37):g.44876486C>T NM_025137.3:c.5392G>A p.(Glu1798Lys)	Het						Whole blood PBMCs	
A018	AHI1	AR	Chr6(GRCh37):g.135751015C>T NM_001134831.1:c.2492+5G>A p.?	Het	Exon skipping (r.2374_2492del; p.Glu792Ilefs*18) Cryptic donor (r.2492_2493ins[2492+1_2492+40]; p.Ile831Metfs*2)	Frameshift/PTC Frameshift/PTC		3	4		
			Chr6(GRCh37):g.135778732G>A NM_001134831.1:c.1051C>T p.(Arg351*)	Het						Whole blood	
A022	TAZ	XLR	ChrX(GRCh37):g.153640551G>C NM_000116.3:c.238G>C p.(Gly80Arg)	Hem	Exon skipping (r.110_238del; p.Lys37_Gly80delinsArg) Cryptic 'GC' donor(r.238_239ins[238+1_238+36]; p.Trp79_Gly80insArgThrArgAlaSerValLeuGlyArgGlyArgLys)	In-frame In-frame		3	5	Whole blood Myocardium	
A024	OPHN1	XLR	ChrX(GRCh37):g.67431946T>C NM_002547.2:c.702+4A>G p.?	Hem	Exon skipping (r.598_702del; p.Val200_Asn234del)	In-frame	CHX insensitive	3	4	Whole blood PBMCs	
A025	SETD5	AD	Chr3(GRCh37):g.9483807C>G NM_001080517.1:c.960-5C>G p.?	Het	Exon skipping (r.960_1077del; p.Lys320_Cys396)	Frameshift/PTC	CHX insensitive	3	4	Whole blood PBMCs	
A026	NBAS	AR	Chr2(GRCh37):g.15427178_15427393dup Chr2(GRCh37):g.15614210_15614501dup Chr2(GRCh37):g.15679451G>A NM_015909.3:c.409C>T p.(Arg137Trp)	Het Het	Not determined Not determined	Transcripts degraded Transcripts degraded	CHX sensitive CHX sensitive	3	5	Fibroblasts	
A029	PPP2R5D	AD	Chr6(GRCh37):g.42975694G>A NM_006245.3:c.748G>A p.(Glu250Lys)	Het	Splicing unaltered		CHX insensitive	4	4	Whole blood PBMCs	
A031	PGAP1	AR	Chr2(GRCh37):g.197750202T>C NM_024989.3:c.1221-3A>G p.?	Hom	Cryptic acceptor (r.1220_1221ins[1221-2_1221-1]; p.Cys407*)	Frameshift/PTC	CHX sensitive	3	5	Whole blood Fibroblasts	
A036	GAA	AR	Chr17(GRCh37):g.78081601C>T NM_000152.4:c.861C>T p.(Pro287=)	Het	Splicing unaltered Exon skipping (r.-32_546del; p.?)	Transcripts degraded Frameshift/PTC		3	3		
			Chr17(GRCh37):g.78078341T>G NM_000152.4:c.-32-13T>G p.?	Het	Cryptic acceptor(r.-33_-32ins[-32-154_-32-1]; p.?) Cryptic acceptor (r.-32_486del; p.?)	5'UTR Frameshift/PTC				Whole blood	
A040	PIGN	AR	Chr18(GRCh37):g.59810585A>C NM_176787.4:c.923-6T>G p.?	Hom	Exon skipping (r.c.923_963del; p.Glu308Glyfs*2) Two exons skipped (r.923_1023del; p.Glu308Glyfs*9) Cryptic acceptor (r.922_923ins[923-26_923-1]; p.Glu308Aspfs*19)	Frameshift/PTC Frameshift/PTC Frameshift/PTC		3	4	Whole blood Fibroblasts Liver	
A050	HSD17B4	AR	Chr5(GRCh37):g.118842585G>C NM_000414.3:c.1333+1G>C p.?	Het	Exon skipping (r.1262_1333del; p.Gly421_Asp444del)	In-frame	CHX insensitive	3	4		
			Chr5(GRCh37):g.118788316G>A NM_000414.3:c.46G>A p.(Gly16Ser)	Het						EBV transformed lymphocytes	

ID	Gene	Inh	Variant	Zyg	Splicing outcome	Consequence	CHX			Sample	Reference
A053	CACNA1E	AD	Chr1(GRCh37):g.181547008G>A NM_001205293.1:c.616+3G>A p.?	Het	Splicing unaltered			3	3	Whole blood	
A054	UBE3A	AD	Chr15(GRCh37):g.25601203C>G NM_130838.2:c.1900G>C p.(Val634Leu)	Het	Cryptic acceptor (r.1899_1900ins[1900-38_1900-1]; p.Val634Phefs*19)	Frameshift/PTC	CHX sensitive	3	5	Whole blood PBMCs	
A058	ASNS	AR	Chr7(GRCh37):g.97482371C>T NM_001673.4:c.1476+1G>A p.?	Hom	Exon skipping (r.1321_1476del; p.Asn441_Gln492del) Cryptic donor (r.1429_1476del; p.Lys478_Val493del) Intron retention (r.1476_1477ins[1476+1_1477-1]; p.Val493Ilefs*2)	In-frame In-frame Frameshift/PTC	CHX insensitive CHX insensitive CHX sensitive	3	5	Whole blood Fibroblasts	Akesson et al. 2020[3]
A060	GSDME	AD	Chr7(GRCh37):g.24745798C>T NM_004403.2:c.1183+5G>A p.?	Het	Exon skipping (r.991_1183del; p.Cys331Lysfs*42)	Frameshift/PTC	CHX insensitive	3	4	Fibroblasts	
A063	GYS1	AR	Chr19(GRCh37):g.49473965_49473967del NM_002103.4:c.1646-1_1647del p.?	Hom	Intron retention (r.1645_1646ins[1645+1_1646-1]; p.Ile550Glnfs*6) Cryptic acceptor (r.1646_1718del; p.Gly549Valfs*29)	Frameshift/PTC Frameshift/PTC		3	5	Whole blood	
A064	GLI3	AD	Chr7(GRCh37):g.42116346C>T NM_000168.5:c.473+5G>A p.?	Het	Exon skipping (r.368_473del; p.His123Argfs*58)	Frameshift/PTC	CHX sensitive	3	5	Fibroblasts	
A066	VPS13D	AR	Chr1(GRCh37):g.12317147A>G NM_015378.3:c.941+3A>G p.?	Het	Exon skipping (r.841_941del; p.Gln282Profs*11) Intron retention (r.941_942ins[941+1_942-1]; p.Asn314Lysfs*2)	Frameshift/PTC Frameshift/PTC	CHX sensitive CHX sensitive	3	4	Whole blood Fibroblasts Urothelial cells	
			Chr1(GRCh37):g.12422904C>T NM_015378.3:c.10270C>T p.(Gln3424*)	Het	Exon skipping (r.10142_10272del; p.Ile3382Asnfs*24) Intron retention (r.10272_10273ins[10272+1_10273-1]; p.Gln3424*)	Frameshift/PTC Frameshift/PTC	CHX sensitive CHX sensitive				
A069	Il11RA	AR	Chr9(GRCh37):g.34658680G>A NM_001142784.2:c.810G>A p.(Thr270=)	Het	Exon skipping (r.647_810del; p.Leu216Cysfs*88) Isoform switch, alternative donor (r.810_811ins[810+1_810+12]; p.Val27_Glu272insValArgProGlyVal)	Frameshift/PTC In-frame		4	4	Whole blood	
			Chr9(GRCh37):g.34657328C>T NM_001142784.2:c.475C>T p.(Gln159*)	Het							
A070	CAMTA1	AD	Chr1(GRCh37):g.7527939G>A NM_015215.3:c.488G>A p.(Arg163Gln)	Het	Splicing unaltered			4	4	Whole blood	
A077	IBA57	AR	Chr1(GRCh37):g.228362733A>G NM_001010867.4:c.679+3A>G p.?	Het	Intron retention (r.679_680ins[679+1_680-1]; p.Pro229Glyfs*53)	Frameshift/PTC		3	4	Whole blood	
			Chr1(GRCh37):g.228353779dup NM_001010867.4:c.262dup p.(Ala88Glyfs*22)	Het							
A078	ABCA12	AR	Chr2(GRCh37):g.215820086C>G NM_173076.2:c.6234-1G>C p.?	Hom	Exon skipping (r.6234_6393del; p.Tyr2079Leufs*2)	Frameshift/PTC	CHX sensitive	3	5	EBV transformed lymphocytes	
A079	LAMP2	XLD	ChrX(GRCh37):g.119576451T>A NM_013995.2:c.928+3A>T p.?	Hem	Exon skipping (r.865_928del; p.Lys289Phefs*36)	Frameshift/PTC		3	5	Whole blood	
	NDUFV1	AR	Chr11(GRCh37):g.67379846C>A NM_007103.3:c.1312C>A p.(Leu438Met)	Het	Splicing unaltered			3	3		
	PYGM	AR	Chr11(GRCh37):g.64525251C>T NM_005609.2:c.660G>A p.(Gln220=)	Het	Exon skipping (r.529_660del; p.Met177_Gln220del) Two exons skipped (r.425_660del; p.Ala142Glyfs*32)	In-frame Frameshift/PTC		3	4		
A085	HSPG2	AR	Chr1(GRCh37):g.22188247T>A NM_001291860.1:c.4958+3A>T p.?	Het	Exon skipping (r.4872_4958del; p.Phe1625_Gln1653del) Cryptic donor (r.4952_4958del; p.Glu1652Valfs*35)	In-frame Frameshift/PTC		3	4	Whole blood	
A087	MSH6	AD	Chr2(GRCh37):g.48032171_48032174delinsCATTATTGTCAGG NM_000179.2:c.3556+5_3556+8delinsCATTATTGTCAGG p.?	Het	Exon skipping (r.3439_3556del; p.Ala1147Valfs*9) Intron retention (r.3556_3557ins[3556+1_3557-1]; p.Glu1187Aspfs*17)	Frameshift/PTC Frameshift/PTC		3	3	Whole blood	
A088	PRPH2	AD	Chr6(GRCh37):g.42689487C>T NM_000322.4:c.581+5G>A p.?	Het	Disruption of transcription start site (r.?; p.?)	Uncertain	CHX insensitive	3	4	Fibroblasts	
A089	TRPM6	AR	Chr9(GRCh37):g.77431578A>C NM_017662.4:c.1308+7T>G p.?	Het	Exon skipping (r.1208_1308del; p.Gly403Alafs*2)	Frameshift/PTC		3	5		
			Chr9(GRCh37):g.77376687C>T NM_017662.4:c.4710G>A p.(Tro1570*)	Het						Whole blood	
A091	ATP1A3	AD	Chr19(GRCh37):g.42473068C>A NM_152296.4:c.2689-1G>T p.?	Het	Intron retention (r.2688_2689ins[2688+1_2689-1]; p.Thr897Valfs*119) Cryptic acceptor (r.2689_2700del; p.Thr897_Gln900del)	Frameshift/PTC In-frame		3	4	Whole blood	

ID	Gene	Inheritance	Variant	Zygosity	Consequence	Classification			Tissue	Reference
A093	CHD8	AD	Chr14(GRCh37):g.21871823C>G NM_001170629.1:c.3308-1G>C p.?	Het	Cryptic acceptor (r.3307_3308ins[3308-66_3308-1]; p.Gly1103Valfs*3) Cryptic acceptor (r.3308_3379del; p.Ala1104Hisfs*12)	Frameshift/PTC Frameshift/PTC	3	5	Whole blood	
A094	NF1	AD	Chr17(GRCh37):g.29422386A>C NM_000267.3:c.59A>C p.(Gln20Pro)	Het	Disruption of transcription start site (r.?; p.?)	Uncertain	3	3	Whole blood	
A097	CC2D2A	AR	Chr4(GRCh37):g.15504547G>T NM_001080522.2:c.438+1G>T p.?	Hom	Exon skipping (r.337_438del; p.Ser113_Glu146del)	In-frame	3	3	Whole blood	
A100	TUBA1A	AD	Chr12(GRCh37):g.49582757del NM_006009.3:c.3+3del p.?	Het	Disruption of transcription start site (r.?; p.?)	Uncertain	3	4	Whole blood Fibroblasts	
A103	PLP1	XLR	ChrX(GRCh37):g.103042898G>C NM_001128834.2:c.622+3G>C p.?	Het	Unable to amplify in blood		3	N/A	Whole blood	
A107	B9D1	AR	Chr17(GRCh37):g.19251097C>T NM_015681.3:c.341G>A p.(Arg114Gln) Chr17(GRCh37):g.19246718C>G NM_015681.3:c.529G>C p.(Asp177His)	Het Het	Exon skipping (r.245_341del; p.Trp82Cysfs*45)	Frameshift/PTC	3	4	Whole blood	Katiyar et al. 2020[4]
A108	DYSF	AR	Chr2(GRCh37):g.71909660G>A NM_003494.3:c.6057G>A p.(Arg2019=) Chr2(GRCh37):g.71901357_71901358del NM_003494.3:c.5698_5699del p.(Ser1900Glnfs*14)	Het Het	Splicing unaltered		3	3	Whole blood	
A113	KCNH2	AD	Chr7(GRCh37):g.150646165del NM_000238.3:c.2399-28del p.?	Het	Isoform switch: loss of expression of the long KCNH2 isoforms (NM_172057.2; NM_000238.4) Transcripts terminate at altenative 3'UTR (NM_172056.2; NM_001204798.2)		3	5	Whole blood	
A118	PKD1	AD	Chr16(GRCh37):g.2162784C>T NM_001009944.2:c.3161+5>A p.?	Het	Inconclusive: possible exon skipping (r.2986_3161del; p.Leu996Valfs*46), at limit of detection (~20% mosaic)	Frameshift/PTC	3	3	Whole blood	
A120	COL4A3	AD	Chr2(GRCh37):g.228149062G>A NM_000091.4:c.2881+1G>A p.?	Het	Exon skipping (r.2747_2881del; p.Ser917_Gly961del)	In-frame	3	5	Fibroblasts	
A129	APC	AD	Chr5(GRCh37):g.112111440T>A NM_000038.5:c.531+6T>A p.?	Het	Exon skipping (r.423_531del; p.Arg141Serfs*8)	In-frame	3	3	Whole blood	
A134	CDH23	AR	Chr10(GRCh37):g.73406215G>A NM_022124.5:c.1291-1G>A p.? Chr10(GRCh37):g.73553240G>T NM_022124.5:c.6555G>T p.(Glu2185Asp)	Het Het	Crytpic acceptor (r.1291_1335del; p.Leu431_Lys445del) Cryptic acceptor (r.1291_1382del; p.Leu431Profs*25) Intron retention (r.1290_1291ins[1290+1_1291-1]; p.Lys431Valfs*22)	In-frame Frameshift/PTC Frameshift/PTC	3	5	Whole blood	
A137	CHD7	AD	Chr8(GRCh37):g.61763030T>A NM_017780.3:c.5405-22T>A p.?	Het	Cryptic acceptor (r.5404_5405ins[5405-20_5405-1]); p.Gly1802Alafs*35) Cryptic 'TG' acceptor (r.5404_5405ins[5405-22_5405-1]); p.(Gly1802Glufs*10) Intron retention (r.5404_5405ins[5404+1_5405-1]); p.(Tyr1803Lysfs*3)	Frameshift/PTC Frameshift/PTC Frameshift/PTC	3	5	Whole blood	
A141	COL4A1	AD	Chr13(GRCh37):g.110822893C>T NM_001845.5:c.3742+1G>A p.?	Het	Exon skipping (r.3557_3742del; p.Ser1187_Gly1248del)	In-frame	4	5	Whole blood	
A143	WDR45	XLD	ChrX(GRCh37):g.48933018C>G NM_007075.3:c.830+5G>C p.?	Het	Intron retention (r.830_831ins[830+1_831-1]; p.Leu278*) Cryptic donor (r.779_830del; p.Thr261Trpfs*10)	Frameshift/PTC Frameshift/PTC	3	4	Whole blood	
A146	CD96	AD	Chr3(GRCh37):g.111296397_111296415del NM_198196.2:c.591+1_591+19del p.?	Het	Exon skipping (r.544_591del; p.Asn183_Glu198del)	In-frame	3	3	Whole blood	
A154	SPAST	AD	Chr2(GRCh37):g.(32370077_32372286)_(32372328_32379442)dup NM_014946.3:c.(1687+1_1688-1)_(1728+1_1729-1)dup	Het	Exon duplication (r.1688_1728dup; p.Met577Asnfs*2)	Frameshift/PTC	3	4	Whole blood	
A155	CACNA1E	AD	Chr1(GRCh37):g.181731712C>T NM_001205293.1:c.4608C>T p.(Asn1536=)	Het	Splicing unaltered		3	3	Whole blood	
A157	OPHN1	XLR	ChrX(GRCh37):g.67412836C>T NM_002547.2:c.1202-1G>A p.?	Hem	Exon skipping (r.1202_1276del; p.Gly401_Phe425del) Three exons skipped (r.1105_1276del; p.Tyr370Leufs*23) Cryptic acceptor (r.1202del; p.Ile402Serfs*20)	In-frame Frameshift/PTC Frameshift/PTC	3	4	Whole blood	

ID	Gene	Inh.	Variant	Zyg.	Splicing consequence	Effect	CHX			Sample
A158	EDN3	AD	Chr20(GRCh37):g.57876778G>C NM_207034.2:c.365+1G>C p.?	Het	Unable to amplify in blood or urothelia			3	N/A	Whole blood Urothelial cells
A162	ABCA1	AD	Chr9(GRCh37):g.107558587T>G NM_005502.3:c.5237+3A>C p.?	Het	Exon skipping (r.5122_5237del; p.Cys1708Valfs*34)	Frameshift/PTC		3	4	Whole blood
A164	STXBP1	AD	Chr9(GRCh37):g.130438221G>C NM_003165.3:1249G>C p.(Gly417Arg)	Het	Exon skipping (r.1111_1249del; p.Asp371Alafs*7) Intron retention (r.1249_1250ins[1249+1_1250-1]; p.Ile418Glyfs*80) Cryptic donor (r.1195_1249del; p.Val399Alafs*7)	Frameshift/PTC Frameshift/PTC Frameshift/PTC		3	5	Whole blood
A165	EDA	XLR	ChrX(GRCh37):g.69253383G>A NM_001399.4:c.924+5G>A p.?	Hem	Isoform switch, alternative donor (r.919_924del; p.Glu308_Val309del)	In-frame		3	4	Whole blood
A168	DNAJB11	AR	Chr3(GRCh37):g.186296198T>C NM_016306.5:c.456+609T>C p.?	Hom	Pseudo-exon (r.456_457ins[456+565_456+620]; p.Val154Glnfs*19) Pseudo-exon (r.456_457ins[457-1490_457-1267]; p.Val153Phefs*11) Two pseudo-exons (r.456_457ins[456+185_456+267;456+565_456+620]; p.Val153Leufs*4) Two pseudo-exons (r.456_457ins[456+565_456+620;457-1490_457-1267]; p.Val154Glnfs*53) Pseudo-exon and cryptic acceptor (r.456_457ins[456+565_456+620];457_460del; p.Val154Glnfs*19)	Frameshift/PTC Frameshift/PTC Frameshift/PTC Frameshift/PTC Frameshift/PTC	CHX sensitive CHX sensitive CHX sensitive CHX sensitive CHX sensitive	3	4	Urothelial cells
A180	IQSEC2	XLD	ChrX(GRCh37):g.53265559G>A NM_001111125.2:c.3396C>T p.(Gly1132=)	Hem	Splicing unaltered			3	3	Whole blood
A188	TUSC3	AR	Chr8(GRCh37):g.15508323G>A NM_001356429.1:c.426G>A p.(Gln142=)	Hom	Exon skipping (r.309_426del; p.Arg103Serfs*4)	Frameshift/PTC	CHX insensitive	3	4	Urothelial cells
A195	KCNH2	AD	Chr7(GRCh37):g.150654364_150654374delinsGT NM_000238.3:c.1128+5_1128+15delinsAC p.?	Het	Exon skipping (r.917_1128del; p.Ala307Profs*10)	Framshift/PTC		3	4	Whole blood
A198	TCF12	AD	Chr15(GRCh37):g.57545457C>G NM_001322151.1:c.1261-3C>G p.?	Het	Cryptic acceptor (r.1260_1261ins[1261-2_1261-1]; p.Gln421Serfs*9)			3	5	Whole blood
A200	CIC	AD	Chr19(GRCh37):g.42791393G>T NM_015125.4:c.452+1G>T p.?	Het	Cryptic donor (r.359_452del; p.Gly120Alafs*54)	Frameshift/PTC		4	5	Whole blood
A205	EMD	XLR	ChrX(GRCh37):g.153608591A>G NM_000117.2:c.266-3A>G p.?	Hem	Cryptic acceptor (r.266_308del; p.Tyr90Leufs*18) Intron retention (r.265_266ins[265+1_266-1]; p.Tyr90Lysfs*25)	Frameshift/PTC Frameshift/PTC		4	5	Whole blood Fibroblasts
A206	FANCA	AR	Chr16(GRCh37):g.89846264_89846274del NM_000135.2:c.1715+3_1715+13del p.? Chr16(GRCh37):g.89857863T>C NM_000135.2:c.1307A>G p.(Gln436Arg)	Het Het	Exon skipping (r.1627_1715del; p.Pro543Hisfs*26) Cryptic donor (r.1715_1716ins[1715+1_1715+258]; p.Ser572Argfs*73)	Frameshift/PTC Frameshift/PTC	CHX sensitive CHX sensitive	3	4	Bone marrow Fibroblasts
A208	DMD	XLR	ChrX(GRCh37):g.32591642C>T NM_004006.2:c.1812+5G>A p.?	Hem	Exon skipping (r.1705_1812del; p.Cys569_Ala604del) Intron retention (r.1812_1813ins[1812+1_?]; p.Leu606Tyrfs*4)	In-frame Frameshift/PTC	CHX insensitive CHX sensitive	3	4	Whole blood Urothelial cells
A213	ARL2BP	AR	Chr16(GRCh37):g.57283769G>A NM_012106.3:c.293+5G>A p.?	Hom	Exon skipping (r.208_293del; p.Ile70Alafs*3) Exon skipping and pseudoexon insertion (r.208_293delins[293+63_293+260]; p.Ile70Alafs*18) Intron retention (r.293_294ins[293+1_294-1]; p.His99*)	Frameshift/PTC Frameshift/PTC Frameshift/PTC		3	4	Whole blood
A219	FXN	AR	Chr9(GRCh37):g.71661296T>G NM_000144.5:c.166-5T>G p.? 180x intron 1 GAA repeat expansion	Het Het	Exon skipping (r.166_263del; p.Ser56Leufs*4) Two exons skipped (r.166_384del; p.Ser57_Ser129del)	Frameshift/PTC In-frame	CHX sensitive CHX insensitive	3	4	Whole blood PBMCs
A220	GALT	AR	Chr9(GRCh37):g.34648972T>G NM_000155.3:c.821-23T>G p.?	Hom	Exon skipping (r.821_904del; p.Asp274_His301del) Cryptic donor (r.855-904del; p.(Ala303Serfs*49) Intron retention (r.820_821ins[820+1_821-1]; p.(Arp274Glyfs*6) Intron retention and cryptic donor (r.820_821ins[820+1_821-1];855_904del; p.Arp274Glyfs*6)	In-frame Frameshift/PTC Frameshift/PTC Frameshift/PTC		3	4	Whole blood
A236	FMR1	XLD	ChrX(GRCh37):g.147009911G>A NM_002024.5:c.270G>A p.(Glu90=)	Hem	Exon skipping (r.199_270del; p.Val67_Glu90del)	In-frame		4	5	Whole blood
A237	SPAST	AD	Chr2(GRCh37):g.32353477G>C NM_014946.3:c.1174G>C p.(Ala392Pro)	Het	Splicing unaltered			3	3	Whole blood
A239	MANBA	AR	Chr4(GRCh37):g.103644027C>T NM_005908.3:c.549+1G>A p.?	Het	Exon skipping (r.379_549del; p.Ser127_Lys183del)	In-frame		3	4	Whole blood

A240	CTNNB1	AD	Chr3(GRCh37):g.41266442C>A NM_001904.4:c.242-3C>A p.?	Het	Exon skipping (r.242_495del; p.Asp81Glyfs*4) Cryptic acceptor (r.244_390del; p.Ile82Alafs*10)	Frameshift/PTC Frameshift/PTC		3	4 Whole blood
A251	CREBBP	AD	Chr16(GRCh37):g.3789725C>A NM_004380.3:c.4134G>T p.(Arg1378=)	Het	Exon skipping (r.4134_4280del; p.Phe1379_Arg1427del) Intron retention (r.4133_4134ins[4133+1_4134-1]; p.Phe1379Serfs*27)	In-frame Frameshift/PTC		3	5 Whole blood
A255	NF1	AD	Chr17(GRCh37):g.29541605dup NM_001042492.2:c.1527+2dup p.?	Het	Exon skipping (r.1393_1527del; p.Val67_Glu90del) Intron retention (r.1527_1528ins[1527+1_1528-1]; p.Asn510Valfs*16)	In-frame Frameshift/PTC		3	4 Whole blood

Splice variant submission form

Important: The Kids Neuroscience splicing diagnostics team will hold the final splice site variant review meeting for 2019 on the 9th December. We will not be reviewing any new cases until our first meeting of 2020 on the 13th January.

We will consider any urgent/acute care cases where a diagnosis is required to inform clinical management of the affected individual.

Note: Once you have submitted, you will be able to download a PDF copy of your submission for reference purposes.

Thank you for taking the time in submitting this variant and related details. Please carefully check entered data and ensure it is correct prior to submitting.

Page 1) Submitter details *
Page 2) Affected individual & genotype details
Page 3) Family individuals, genotype & segregation details

* page you are currently on

1	Patient identifier	
		_____
		(i.e. Referring centre patient ID e.g. MRN-XXX-XXX-XXX)

2	Referral Centre	◯ SCHN
		◯ VCGS
		◯ PathWest
		◯ SEALS
		◯ Other

	Referral centre (other)	
		_____

3	Clinician - Name	
		_____

	Clinician - Email	
		_____

4	Genetic Pathologist or Molecular Geneticist (responsible for variant classification) - Name	
		_____

	Genetic Pathologist or Molecular Geneticist (responsible for variant classification) - Email	
		_____

5	Additional recipient - Name	
		_____

	Additional recipient - Email	
		_____

Page 1) Submitter details

Page 2) Affected individual & genotype details *

Page 3) Family individuals, genotype & segregation details

* page you are currently on

6   Patient (proband) - Name

       _____

       Patient (proband) - Date of birth

       _____

       Patient (proband) - Gender   ◯ Male   ◯ Female

       Brief clinical description of proband. OMIM#

       _____
       (e.g. Joubert Syndrome; OMIM# 213300)

       Age of onset   ◯ Fetal death
                            ◯ Congenital
                            ◯ Infantile
                            ◯ Juvenile
                            ◯ Adult
                            ◯ Uncertain

Note:
- In urgent cases, informal results can be provided within 10 days.
- Our goal is to produce a diagnostic report within 4-6 weeks. There can be multiple abnormal splicing outcomes and validation of all events can take time. As the resultant clinical decision-making is often significant (e.g. PGD) we feel this degree of analysis is essential.

       How urgently are results from Splicing Diagnostics   ◯ Urgent
       required?   ◯ Not urgent

       Comments regarding Splicing Diagnostics urgency

       _____

7   Gene name

       _____
       (e.g. AHI1)

       Assembly   ◯ GRCh37 (hg19)   ◯ GRCh38 (hg38)

Chromosome	◯ chr1
	◯ chr2
	◯ chr3
	◯ chr4
	◯ chr5
	◯ chr6
	◯ chr7
	◯ chr8
	◯ chr9
	◯ chr10
	◯ chr11
	◯ chr12
	◯ chr13
	◯ chr14
	◯ chr15
	◯ chr16
	◯ chr17
	◯ chr18
	◯ chr19
	◯ chr20
	◯ chr21
	◯ chr22
	◯ chrX
	◯ chrY

gDNA nomenclature

_____
(e.g. g.XXXXXXXXC>TPlease ensure cDNA and gDNA annotations match & are correct via Mutalyzer or Variant validator)

Transcript

_____
(e.g. ENST00000269305.4, NM_001126113.2)

cDNA nomenclature

_____
(e.g. c.YYYY+1G>A)

8	Is [gene_name] a good phenotypic match for the clinical presentation?	◯ Yes
		◯ No
		◯ Somewhat
		◯ Uncertain

Is [gene_name] a good phenotypic match for the clinical presentation? (comments)

_____

9	Are any of the following available?	☐ Skin fibroblasts
		☐ Frozen biopsies
		☐ PBMCs
		☐ Uncertain
		☐ None available

Are any of the following available? (comments)

_____

| **Page 1) Submitter details** |
| **Page 2) Affected individual & genotype details** |
| **Page 3) Family individuals, genotype & segregation details *** |
| |
| *** page you are currently on** |

| 10 | Please tick which family members you can provide information for | ☐ Father
☐ Mother
☐ Sibling 1
☐ Sibling 2
☐ None available |

| 10a | Father - Name | _____ |

	Father - Date of birth
_____	

| | Father - Status | ○ Affected
○ Unaffected
○ Unknown
○ Other |

| | Father - Comments | _____ |

| 10b | Mother - Name | _____ |

	Mother - Date of birth
_____	

| | Mother - Status | ○ Affected
○ Unaffected
○ Unknown
○ Other |

| | Mother - Comments | _____ |

| 10c | Sibling 1 - Name | _____ |

	Sibling 1 - Date of birth
_____	

| | Sibling 1 - Gender | ○ Male   ○ Female |

Sibling 1 - Status	○ Affected ○ Unaffected ○ Unknown ○ Other

Sibling 1 - Comments	_____

10d Sibling 2 - Name

_____

Sibling 2 - Date of birth	

_____

Sibling 2 - Gender	○ Male ○ Female

Sibling 2 - Status	○ Affected ○ Unaffected ○ Unknown ○ Other

Sibling 2 - Comments	_____

11 Zygosity
○ Heterozygous
○ Homozygous
○ Hemizygous

12 Consanguinous
○ Yes   ○ No   ○ Unknown

13 Suspected inheritance pattern of the disorder
☐ Autosomal dominant
☐ Autosomal recessive
☐ X-linked
☐ Mitochondrial
☐ Uncertain

Suspected inheritance pattern of the disorder
(comments)

_____

14 Segregation of variant
○ Paternal
○ Maternal
○ De novo
○ Not determined

Segregation (comments)

_____

15	What is the current ACMG classification for this variant?	○ Benign ○ Likely Benign ○ Variant of Uncertain Significance ○ Likely Pathogenic ○ Pathogenic ○ Not yet classified
16	For recessive conditions, has a second variant in [gene_name] been identified?	○ Yes ○ No ○ Disorder is not autosomal recessive ○ Not determined

For recessive conditions, has a second variant in [gene_name] been identified? (comments)

_____

Assembly (for second variant)　　　　　　　　○ GRCh37 (hg19)　　○ GRCh38 (hg38)

Chromosome (for second variant)

○ chr1
○ chr2
○ chr3
○ chr4
○ chr5
○ chr6
○ chr7
○ chr8
○ chr9
○ chr10
○ chr11
○ chr12
○ chr13
○ chr14
○ chr15
○ chr16
○ chr17
○ chr18
○ chr19
○ chr20
○ chr21
○ chr22
○ chrX
○ chrY

gDNA nomenclature (for second variant)

_____
(e.g. g.XXXXXXXXC>TPlease ensure cDNA and gDNA annotations match & are correct via Mutalyzer or Variant validator)

Transcript (for second variant)

_____
(e.g. ENST00000269305.4, NM_001126113.2)

cDNA nomenclature (for second variant)

_____
(e.g. c.YYYY+1G>A)

	Second variant (segregation)	◯ Paternal
		◯ Maternal
		◯ de novo
		◯ Not determined

Segregation (comments) (for second variant)

_____

15	What is the current ACMG classification (for second variant)?	◯ Benign
		◯ Likely Benign
		◯ Variant of Uncertain Significance
		◯ Likely Pathogenic
		◯ Pathogenic
		◯ Not yet classified

17	Are there any other genetic abnormalities relevant to interpretation of splicing outcomes for this variant?	◯ CNV detected
		◯ Other
		◯ None yet identified

Are there any other genetic abnormalities relevant to interpretation of splicing outcomes for this variant? (comments)

_____

18	Experimental studies will support decision-making related to clinical classification of this variant?	◯ Agree
		◯ Not sure
		◯ Disagree

Experimental studies will support decision-making related to clinical classification of this variant? (comments)

_____

Assessing the clinical impact of splicing studies

Splice variant, classification, report and impact

1. Report Sample No.

_____
(Note: This refers to "SAMPLE NO:" e.g.
Z77_JoDo_A1)

Report Date

_____
(Note: This refers to "Date of report:")

2. What is your role in caring for the family with this variant?

○ Specialist clinician
○ Genetic pathologist responsible for variant classification
○ Molecular genetic scientist responsible for variant classification
○ Other (enter comments below)

Comments
(for Question 2)

_____

3A. What was the ACMG classification of this variant prior to mRNA studies?

○ Pathogenic
○ Likely pathogenic
○ Benign
○ Likely benign
○ Uncertain significance
○ Not classified

Please elaborate on the evidence used for classifying this variant as [prior_acmg_class]
(Related to 3A)

_____
(ACMG evidence e.g. PM2, etc.)

3B. What was the ACMG classification of this variant after mRNA studies?

○ Pathogenic
○ Likely pathogenic
○ Benign
○ Likely benign
○ Uncertain significance
○ Not classified

Please elaborate on the evidence used for classifying this variant as [after_acmg_class]
(Related to 3B)

_____
(ACMG evidence e.g. PM2, etc.)

4. Was the report received in a clinically relevant timeframe?

○ Yes   ○ No   ○ Somewhat

Comments
(for Question 4)

_____

5. The splicing diagnostics report was easily understood and clearly informed variant classification

○ Strongly agree
○ Agree
○ Neutral
○ Disagree
○ Strongly disagree

Comments
(for Question 5)

_____

---

6. What other information/data would you find useful to be included in this report?

_____

---

7. Did this report inform clinical management of the patient/family?

○ Yes *
○ Not yet, but possible in future *
○ No
○ No contact with family
 (* Please select from the list below)

Clinical management
(Note: Select all relevant options)

☐ Genetic Diagnosis
☐ Genetic counselling *
☐ Guide clinical care *
☐ Prognostic counselling
☐ Eligibility for clinical trial
☐ Other
 (* Further sub-categories will be shown below)

Genetic counselling (sub-categories)
(Note: Select all relevant options)

☐ Carrier testing
☐ Prenatal Counselling
☐ Preimplantation genetic diagnosis/screening
☐ Screening/diagnosis of siblings

Guide clinical care (sub-categories)
(Note: Select all relevant options)

☐ Guide clinical management
☐ Anticipation of co-morbidity or complications linked to specific genetic disorder
☐ Intervention or therapy
☐ Palliation

---

8. In which settings would you use this service when a variant has been identified in an OMIM gene consistent with an affect individual's clinical presentation?

☐ To provide evidence to reclassify a Likely pathogenic variant to a pathogenic variant
☐ To provide evidence to reclassify a VUS to a likely/pathogenic variant
☐ To provide evidence to reclassify a VUS to a likely/benign variant

---

Comments
(for Question 8)

_____

---

9. The cost of this service is $2,500 for testing of 3 family members.

Would you use this testing to reclassify a variant in a gene consistent with an affected individual's clinical presentation?

○ Yes
○ No *
○ Unsure *
 (* Please elaborate why in comments box below)

Comments
(for Question 9)

_____

---

10. From your clinical consultation with the family,
what is your opinion of the impact of this testing
for the family?
(Note: Select all relevant options)

☐ Positive impact
☐ No impact / Neutral impact
☐ Negative impact
☐ No contact with family
☐ Prefer not to comment

Comments
(for Question 10)

_____

---

11. We would like to invite families who have participated in our Splicing Diagnostics research program to anonymously share their views on what it means for them to obtain a precise genetic diagnosis. We welcome any feedback on the impact for the families from participation in this research study (both positive and negative).

Please might you share this link with the participating families when/if you convey the results from the Splicing Diagnostics report to the family (http://kidsneuroscience.org.au/family_feedback).

Feedback on Splicing Diagnostics report design, layout and content

12. The design and layout of the report was clear and
easy to read

○ Strongly agree
○ Agree
○ Neutral
○ Disagree
○ Strongly disagree

Comments
(for Question 12)

_____

---

13. This report contained sufficient scientific
explanation

○ Strongly agree
○ Agree
○ Neutral
○ Disagree *
 (* Additional sub-categories will be displayed
 below to better understand)

You disagreed because there was ...

○ Insufficient detail
○ Excessive detail
○ Other
 (Please elaborate in Comments box below)

Comments
(Please elaborate ...)

_____

---

14. The figures in this report were comprehensible
and well labelled

○ Strongly agree
○ Agree
○ Neutral
○ Disagree *
 (* Additional sub-categories will be displayed
 below to better understand)

REDCap

You disagreed because there was ...	◯ Insufficient detail
	◯ Excessive detail
	◯ Other
	(Please elaborate in Comments box below)

Comments
(Please elaborate ...)

_____

REDCap

Variant Classifiers survey

What constitutes Acceptable Functional Evidence from PCR based mRNA studies for classification of splicing variants?

Part A. Acceptable models for provision of functional testing of splicing variants for clinical consideration. 4 multiple choice questions.

Our goal from this series of questions is to determine the acceptable operational models for provision of functional testing results used for clinical decision-making (what are the minimum acceptable requirements, what are the requirements accepted by most i.e. consensus)

1 Diagnostic investigations of pre-mRNA splicing used to assist variant classification;

a) Must be performed as an accredited test by a NATA accredited laboratory if used to support variant classification.

○ Strongly agree
○ Agree
○ Disagree
○ Strongly disagree
○ Uncertain

Comments

_____

b) May be performed as a non-accredited test by a NATA accredited laboratory if used to support variant classification.

○ Strongly agree
○ Agree
○ Disagree
○ Strongly disagree
○ Uncertain

Comments

_____

2 Do you consider functional studies of pre-mRNA (mis)splicing published in peer-reviewed journals to be valid evidence for variant classification?

○ Yes
○ Sometimes
○ No

Comments

_____

3 Please rank the following criteria governing studies of pre-mRNA splicing performed by a reputable research laboratory in order for these studies to be used to assist variant classification:

	Essential	Desirable	Non essential
a			

REDCap®

	There is Human Ethics governance approval for the study and informed, written consent by the study participants	○	○	○
b	The experimental studies are considered by the variant classifier to have been performed with robust scientific rigour and adequate controls	○	○	○
c	The laboratory is affiliated with an accredited diagnostic laboratory and subject to monitoring of practices to ensure appropriate procedures related to sample providence	○	○	○

Comments

_____

Part B. Technical aspects of Functional Testing results. 9 multiple choice questions. 15 minutes.

The protocols used for this Splicing Diagnostics program1 were devised to ensure reliability of PCR of mRNA (cDNA) derived from 'easily obtained' biospecimens (blood, skin or urine cells) to assist variant classification in Mendelian disorders with tissue-specific presentations. Our study establishes that (mis)splicing events are highly reproducible between biological replicates (different carriers of a variant), multiple biospecimens from individual carriers (including the manifesting tissue when available) and between experimental replicates.

It is important to now consider if and how functional testing of splicing variants may transition into a diagnostic context.

Therefore, our goal from this series of questions is to determine the Variant Classifier's perspective of which steps are essential, and which are non-essential, with respect to:

Specific technical findings that enable satisfactory confidence of normal splicing for variant classification Specific technical findings that enable satisfactory confidence of mis-splicing Illustrative reports which highlight different technical aspects of the mRNA analysis are available to assist in completing this survey.

Report 1 – Use of junctional primer to specifically amplify transcript which utilise a cryptic splice site (Part B Q6)

Report 2 – RT-PCRs demonstrate all abnormal splicing is detected from variant allele and all normal splicing is detected from allele in trans (Part B Q6)

Report 3 – Use of allele bias to infer abnormal initiation of transcription (Part B Q6)

Report 4 – RNA sequencing data showing the comparative isoform expression in the manifesting tissue relative to the biospecimen tested (Part B Q1, Part C Q2)

1Funded through a Sydney Health Partners Medical Research Futures Foundation Rapid Applied Research Translation Grant; MRFF RART.

1   The design of our RT-PCR protocol was based on careful review of RNA sequencing data from controls to scrutinise natural alternative splicing or mis-splicing of the target gene in the manifesting tissue(s), relative to blood, skin or urine cells.

How important is critical analysis of natural, tissue-specific alternative or mis-splicing of the gene with respect to functional testing of splicing variants, to ensure that:

		Essential	Very important	Somewhat important	Not important	Uncertain
a	Functional testing to establish normal or mis-splicing of a gene in e.g. blood can be used to infer splicing outcomes for that gene in the manifesting tissue(s)	○	○	○	○	○
b	RT-PCR primers are positioned in constitutive exons present in all major isoforms of the gene expressed in the biospecimen tested and the manifesting tissue(s)	○	○	○	○	○
c	RT-PCR strategies are informed by natural mis-splicing events in this region of the gene, which are commonly enhanced in the context of a splice site variant	○	○	○	○	○

What is your opinion or understanding of the strengths and weaknesses of use of expression databases, such as GTEx or ENCODE, which present an algorithmic analysis of predicted isoform expression to perform such analyses - over direct review of RNA sequencing data (which can be requested from GTEx and ENCODE)?

_____

2   Our RT-PCR protocol requires strategic design of primers to specifically interrogate for all possible splicing outcomes:

Normal splicing Exon skipping Intron retention Use of cryptic site(s) Abnormal initiation of transcription Abnormal termination of transcription. In order to provide acceptable evidence for variant classification, how important is designing PCR primers for each possible mis-splicing event?

○ Essential
○ Desirable
○ Excessive

Comments

_____

3   Our RT-PCR protocol included an experimental repeat of every observed splicing outcome to confirm reproducibility of the findings (either via use of two separate primer pairs to confirm each observed splicing event, or where this was not possible, a repeat experiment performed with the same primer pair). This requirement resulted in an average of 6 - 8 primer pairs/case and average reagent costs (including Sanger sequencing) of AUS$500 for a singleton and AUS$900 for a trio.

How important is performing experimental repeats for every observed splicing outcome to a) confirm reproducibility of pre-mRNA splicing and b) provide acceptable evidence for variant classification?

○ Essential
○ Desirable
○ Excessive

Comments

_____

---

4   Our RT-PCR protocol performed gel excision and Sanger Sequencing of each PCR amplicon (from the proband, carrier parents if available and controls) from both experimental repeats.

In order to provide acceptable evidence for variant classification, how important is Sanger sequencing of both experimental repeats?

◯ Essential
◯ Desirable
◯ Excessive

Comments

_____

---

5   Experimental repeats of splicing outcomes may be deemed essential by the majority of Variant Classifiers. If so, we propose an experimental protocol that allows time and cost savings (gel extraction and Sanger sequencing not required for Experiment 2):

Experiment 1 uses gel extraction and Sanger sequencing of PCR amplicons (from patient and controls) to reliably determine the DNA sequence of the amplicon and nature of the mis/splicing event and; Experiment 2 provides a technical repeat though uses chromatography to determine the molecular weight of amplicons, and, infers via molecular weight correlations with Experiment 1 that the splicing events are the same. and; Any instance where a discrepancy is observed between Experiment 1 and 2 must be subject to repeat RT-PCR with Sanger sequencing of gel extracted amplicon.

◯ Protocol acceptable as stated
◯ Protocol acceptable with the following changes
◯ Protocol unacceptable
◯ Alternate protocol suggested

Comments

_____

Comments

_____

---

6   For heterozygous variants our RT-PCR protocol used technical strategies (junctional primers, and/or, allele-specific single nucleotide variants) to show: a) all observed mis-splicing was arising from the variant allele and/or b) observed normal splicing was arising from the allele in trans.

---

a   What importance do you place on the use of technical strategies to confirm that mis-splicing events occurr only in carriers of the variant and not in age- and gender-matched controls?

◯ Essential
◯ Very important
◯ Somewhat important
◯ Not important
◯ Uncertain

Comments

---

b   What importance do you place on the use of technical strategies to amplify a long PCR amplicon to interrogate for presence of any other single nucleotide variant (SNV) identified within the coding region of the same gene for which segregation data is available to show:

REDCap

		Essential	Very important	Somewhat important	Not important	Uncertain
a	All abnormal splicing detected was coming from the variant allele	◯	◯	◯	◯	◯
b	All normal splicing detected was coming from the allele without the splicing variant inherited in trans (for heterozygous splicing variants)	◯	◯	◯	◯	◯
c	Abnormal initiation of transcription for variants affecting the 5'splice site (donor) of exon 1 (which can prevent exon recognition by the RNA polymerase)	◯	◯	◯	◯	◯
d	For variants not observed to induce mis-splicing, presence of roughly equal levels of a heterozygous SNV at a region distal to the variant to provide evidence supporting normal transcription and splicing of both alleles	◯	◯	◯	◯	◯

Comments

_____

7 How important is trio testing or testing multiple carriers of the variant (biological replicates), to establish reproducibility of splicing outcomes in more than one carrier of the variant?

◯ Essential
◯ Very important
◯ Somewhat important
◯ Not important
◯ Uncertain

Comments

_____

8 How important is the use of specimen- , age- and gender- matched controls for RT-PCR (requiring a Biobank and informed consent to archive samples previously diagnosed for use as controls)?

◯ Essential
◯ Very important
◯ Somewhat important
◯ Not important
◯ Uncertain

Comments

_____

projectredcap.org

REDCap®

9   In this study, we used cycloheximide (CHX) inhibition of nonsense-mediated decay (NMD) for 22 cases: 9/22 cases were confirmed to have theoretically NMD-compliant mis-splicing outcome(s). NMD-compliant transcripts were rescued by CHX treatment in 6/9 cases; whereas 3/9 cases were unresponsive to CHX treatment. Effective inhibition of NMD was confirmed using a positive control for all cases.

a) Are you aware that only spliced transcripts that are successfully transported out of the nucleus for a pilot round of translation in the cytoplasm can activate nonsense-mediated decay (NMD)?

○ Fully aware
○ Somewhat aware
○ Not aware
○ Uncertain

Comments

_____

b) Are you aware that a proportion of mis-spliced transcripts with a NMD-compliant premature termination codon are retained in the nucleus and are incapable of activating nonsense-mediated decay, though are also unable to be translated?

○ Fully aware
○ Somewhat aware
○ Not aware
○ Uncertain

Comments

_____

c) CHX treatment strengthened evidence for some variants by showing mis-splicing was not a rare event; rather, nonsense-mediated decay was effective. However, CHX treatment adds an additional technical step of cell culture, increases costs of testing by 50 %, doubles the number of specimens for analysis, and requires delivery of specimens to the laboratory within 24 hours. Whereas, PAX RNA tubes are stable for 3 – 5 days at room temperature, 5 – 7 days at 4°C and up to 1 year frozen at -20°C. Based on these pros and cons, do you agree that cycloheximide treatment should be performed as a second investigation for cases where evidence from mRNA studies without CHX treatment is insufficient for variant classification?

○ Yes
○ No
○ Uncertain

Comments

_____

Part C. Provision of technical data in our RT-PCR Splicing Diagnostic reports. 6 multiple choice questions.

1   How useful is provision of the splicing prediction using ALAMUT® visual biosoftware?

○ Essential
○ Very useful
○ Somewhat useful
○ Not useful
○ Uncertain

Comments

_____

2   How useful is provision of sashimi plots from RNA sequencing data showing the comparative isoform expression in the manifesting tissue relative to the biospecimen(s) tested?

○ Essential
○ Very useful
○ Somewhat useful
○ Not useful
○ Uncertain

Comments

_____

3   How useful is presentation of multiple RT-PCR gels that show specific interrogation for all possible mis-splicing outcomes (1. Exon skipping; 2. Cryptic Splice site use; 3. Intron retention), even if some mis-splicing event were not detected?

○ Essential
○ Very useful
○ Somewhat useful
○ Not useful
○ Uncertain

Comments

_____

4   How useful is provision of the Sanger sequencing chromatogram files for expert scrutiny?

○ Essential
○ Very useful
○ Somewhat useful
○ Not useful
○ Uncertain

Comments

_____

REDCap®

5 How useful is provision of the schematic with exons, introns and dashed lines showing the detected splicing abnormalities?

◯ Essential
◯ Very useful
◯ Somewhat useful
◯ Not useful
◯ Uncertain

Comments

_____

6 How useful is detailing within the Diagnostic Reports the consequences of any observed mis-splicing events for the encoded protein?

◯ Essential
◯ Very useful
◯ Somewhat useful
◯ Not useful
◯ Uncertain

Comments

_____

Part D. Further comments or recommendations regarding how the Splicing Diagnostic may be improved.
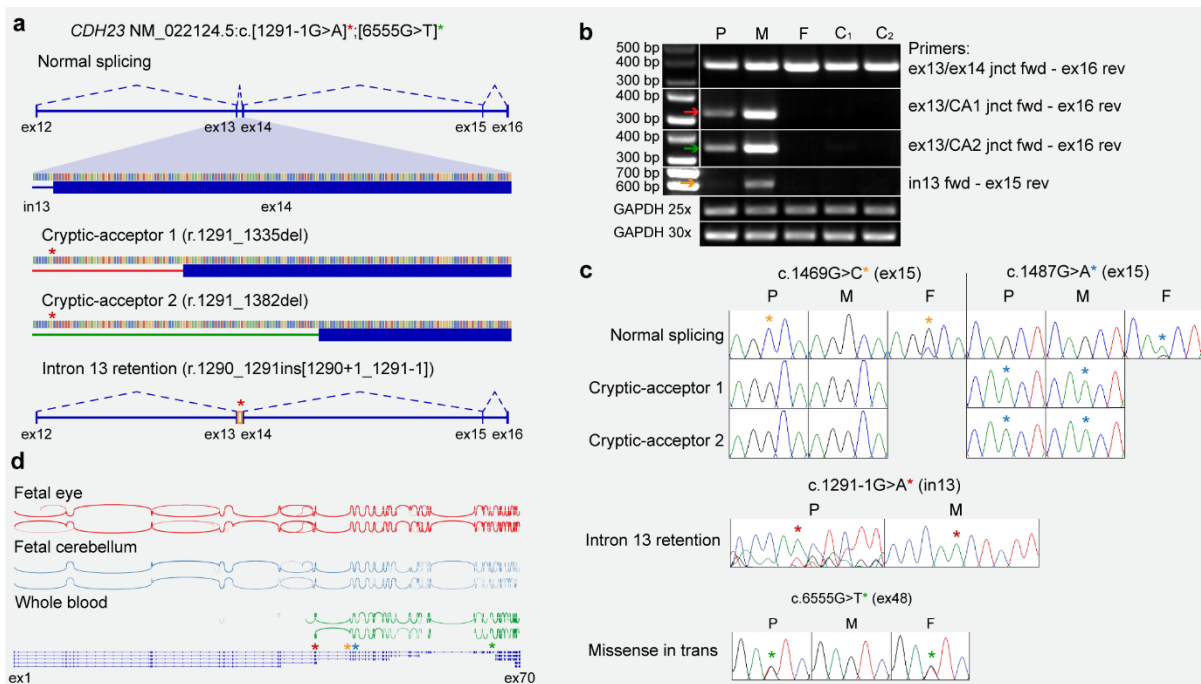
Figure S4: A case of significant differential expression between the manifesting tissue and clinically accessible tissue tested (A134-*CDH23*).

a, Overview showing detected mis-splicing for case A134-*CDH23* with bilateral profound sensorineural hearing loss (MIM#601386). Compound heterozygous variants NM_022124.5:c.1291-1G>A (red asterisk) and NM_022124.5:c.6555G>T (green asterisk) *in trans* were both classified as VUS. **b**, *CDH23* mRNA studies using blood showed three abnormal splicing events in the proband and heterozygote mother, absent in controls; use of two cryptic-acceptors (CA1:cryptic-acceptor 1, red bar and arrow; CA2:cryptic-acceptor 2, green bar and arrow), and intron 13 retention (orange box and arrow). Proband (P: male, 1 year); mother (M: female, 32 years); father (F: male, 32 years); controls (C1: male 2 years; C2: male, 3 years). **c**, Sanger sequencing of cDNA with normal splicing (forward primer annealing to the exon 13/14 splice junction and reverse primer in exon 16) shows apparent hemizygosity of two benign variants (NM_022124.5:c.1469G>C *in trans* (orange asterisk) and NM_022124.5:c.1487G>A *in cis* (blue asterisk))*,* establishing undetectable levels of normal splicing arising from the maternal allele. Also, Sanger sequencing of the band corresponding to intron 13 retention shows all detectable levels of intron 13 retention were expressed from the maternal allele. RNA studies enabled re-classification of the c.1291-

1G>A variant from VUS to pathogenic. PM3 was subsequently used to re-classify the c.6555G>T variant *in trans* likely pathogenic **d**, RNA-seq sashimi plots showing predominant expression of the full length *CDH23* isoforms in fetal eye and fetal cerebellum that are absent in blood.

Recommended ascertainment criteria
a. High likelihood of a monogenic disorder
b. Variant in a phenotypically concordant gene
c. Variant segregates with disease
d. Variant has allele frequency consistent with disease incidence

In silico analyses to define possible effects on pre-mRNA splicing
a. Review if variant weakens a consensus splice site, creates or strengthens a cryptic splice site.
b. Consider variant impact upon splice enhancers or repressors: particularly for exonic variants in alternatively spliced regions.
c. Theorise possible consequences for pre-mRNA splicing (exon-skipping, intron retention, cryptic splice-site use) with respect to an encoded premature termination codon and NMD.
d. For variants affecting the promoter region, 5'UTR or consensus splice-sites of intron 1, be mindful of loss of transcription or activation of an alternative transcription start site.
e. For variants affecting the last intron of any transcript, consider potential disruption of transcription termination or polyadenylation

Mine RNA sequencing (RNA-seq) data from manifesting tissue(s) and clinically accessible biospecimens
a. Determine expression levels of the target gene (Guide: > 0.2 TPM feasible for RT-PCR; > 5.0 TPM feasible for RNA-seq)
b. Establish whether the variant affects a constitutive exon present within the predominant isoform(s) of the target gene expressed in the manifesting tissue(s) versus the biospecimens potentially available for RNA testing (e.g. blood, skin or urine cells).
c. Identify any differences in isoforms expressed by the manifesting tissue(s) and blood, skin or urine cells relevant to design or interpretation of PCR.
d. Use clues from natural mis-splicing/alternative splicing in this region of the gene to inform PCR strategies, as these events are commonly enhanced in the context of a variant affecting the consensus splice site.
   - If two adjacent exons are naturally skipped together, ensure primers flanking variant do not lie within either of these exons.
   - If use of natural cryptic/alternative splice sites is observed, design a junctional primer, or strategically position a primer, to ensure you are able to detect any variant-associated enhancement of this event.
   - If natural intron retention occurs, use a reverse primer at the beginning of the intron paired with a forward primer two exons upstream to detect this event. Reciprocally, a forward primer at the end of the intron, paired with a reverse primer two exons downstream, can inform if the entire intron is retained.

PCR of cDNA
a. PCR detects only what primer design and PCR conditions allow. Ensure PCR strategy **specifically interrogates for all potential mis-splicing events**; 1) normal splicing, 2) exon skipping, 3) cryptic splice site use, 4) intron retention; _for first or last exons of transcripts_; 5) disruption of initiation of transcription, 6) disruption of termination of transcription
b. Position PCR primers in **constitutive exons** of the predominant isoform(s) of the target gene expressed by the manifesting tissue(s) and specimen(s) tested.
c. Position paired **primers more than one splice junction apart** to distinguish cDNA amplification from gDNA amplification.
d. Allow PCR extension times of ~ 1 minute per 1000 nt of cDNA, mindful of **amplification bias** for shorter PCR amplicons and against longer PCR amplicons.
e. Primers bridging exon junctions are useful tools to probe for specific splicing events, though due to exonic conservation at splice-sites can result in non-specific amplification of other cDNA templates (why Sanger sequencing of amplicons is requisite).
f. Consider impact of NMD acting on mis-spliced transcripts, especially for heterozygous variants. For example, when using primers in exons flanking a splicing variant, there will be _enormous_ PCR amplification bias against longer transcripts with intron retention activating NMD.
g. PCRs that detect multiple splicing outcomes (e.g. primers in flanking exons) are multi-template PCRs susceptible to **heteroduplex formation**.
h. Where possible test parent heterozygote(s) and affected proband: **Biological replicates** increase confidence of reproducibility of variant-associated (mis)splicing events.
i. Use **age-, sex- and specimen-matched cDNA from at least two and preferably three controls** to assess natural mis-splicing/alternative splicing.
j. Use amplification of a relevant control gene to show similar quantity and quality of cDNA in each reaction (we recommend use of 25 and 30 cycles to show amplification conditions are sub-saturating).
k. For heterozygous variants, we emphasize diagnostic utility of segregated, heterozygous coding single nucleotide variants (SNV) in the same gene to show:
   - For variants <u>not</u> observed to induce mis-splicing, confirm presence of roughly equal levels of a heterozygous SNV at a region distal to the variant to provide evidence supporting normal transcription and splicing of both alleles.
   - For variants observed to induce mis-splicing, establish whether/if transcripts with normal splicing arise from the allele _in trans_. For example, derive an amplicon corresponding to normal splicing of multiple exons (a bridging primer may be required for cases with activation of a nearby cryptic splice-site) to encompass a segregated heterozygous SNV, which will appear hemizygous if _in trans_ or absent if _in cis_ by Sanger sequencing.
   - Evidence of abnormal initiation of transcription. For example, if there is an additional SNV _in trans_ in exon 4, compare Sanger sequencing chromatograms of an exon 1/exon 5 amplicon versus an exon 2/exon 5 amplicon, scrutinizing whether the exon 4 SNV transitions from hemizygous to heterozygous. If it does, these data support abnormal initiation of transcription downstream of exon 1.

Sanger sequencing and interpretation notes
a. Be mindful of 'messy sequence' from gel extracted amplicons, as this can reflect superimposed sequences due to multi-template PCR and be diagnostic of a mis-splicing event; for example, use of a nearby cryptic splice site that creates a similar sized amplicon. Resolution of closely migrating amplicons may be resolved via diagnostic chromatography.
b. Only mis-spliced transcripts capable of nuclear export and pilot translation activate NMD. For heterozygous variants, a mis-spliced product present at 10% levels of a normally spliced product (of similar size) may be consistent with: _i)_ a complete splicing defect targeted by NMD, or _ii)_ a partial splicing defect not targeted by NMD. See **4k** for diagnostic use of a coding SNV to phase splicing events to discern complete from partial mis-splicing. Consider repeat testing using an additional RNA preparation step of cycloheximide inhibition of NMD.
c. Natural mis-splicing can confound or obscure 'increased mis-splicing'. Quantitative PCR approaches may be necessary for some cases.

Figure S5: Procedural guidelines for RNA Diagnostics via RT-PCR and Sanger sequencing endorsed by Clinical Variant Curators (genetic pathologists and qualified diagnostic scientists).
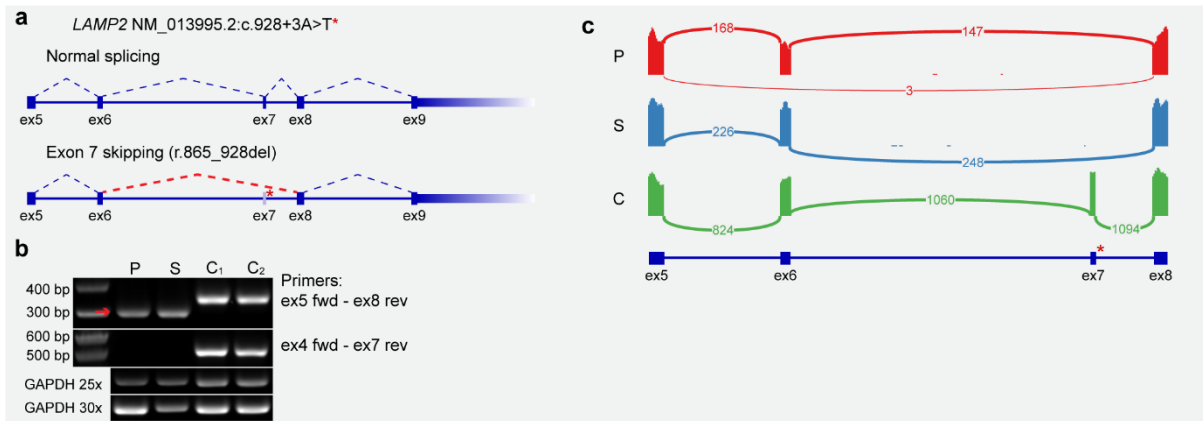
Figure S6: A case of maternal germline mosaicism (A079-*LAMP2*).

a, Schematic of detected mis-splicing for case A079-L*AMP2*: A hemizygous NM_013995.2:c.928+3A>T variant was identified in a proband and sibling with severe concentric hypertrophic cardiomyopathy and proximal muscle weakness (MIM#300257). The mother tested negative for c.928+3A>T using blood gDNA, though subsequently is established to be germline mosaic. **b**, RT-PCR and **c**, RNA-seq identify abnormal exon 7 skipping (red splice junction and arrow). Collective evidence from RNA studies enabling re-classification of c.928+3A>T from VUS to pathogenic. Proband (P: male, 2 years); sibling (S: male, 3 years); controls (C1: male, 7 months; C2: male, 5 years).
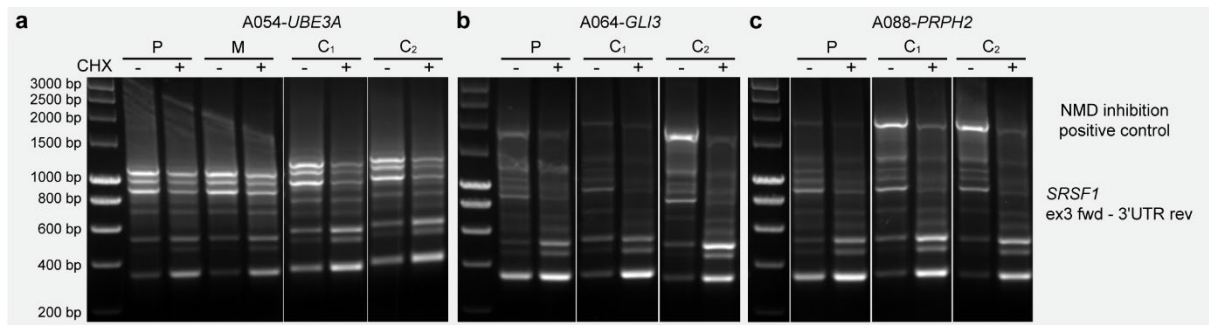
Figure S7: Positive control for cycloheximide inhibition.

a-c, *SRSF1* negatively autoregulates its expression via alternative splicing of transcript isoforms that are targeted by NMD[5]. Thus we used RT-PCR of *SRSF1* as a positive control for CHX treatment (Fig 3b,e,h). Effective inhibition of NMD using CHX results in increased abundance of the short *SRSF1* isoforms.
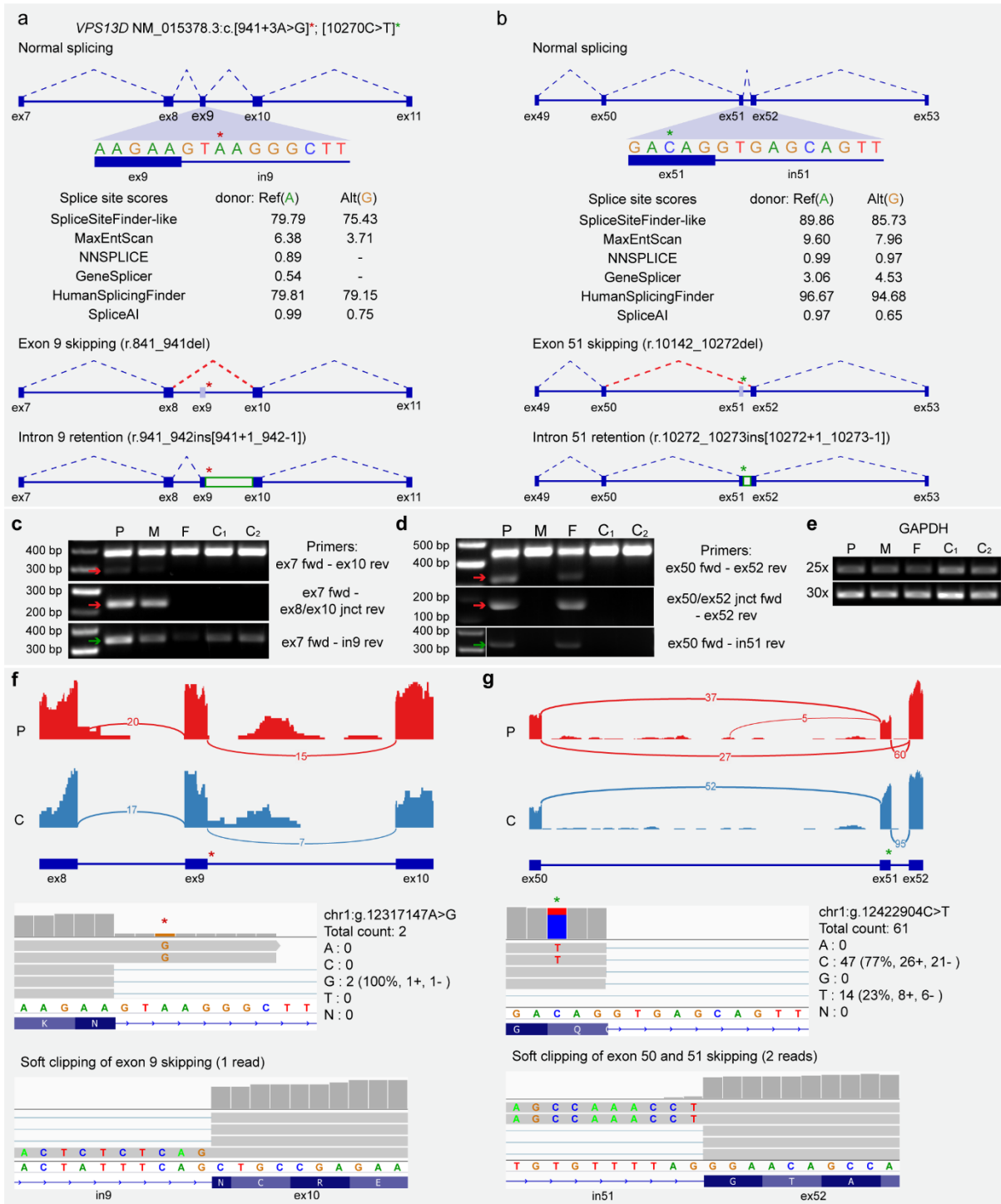
Figure S8: A complex case with pathogenic partial mis-splicing (A066-*VPS13D*).

Overview of case A066-*VPS13D*: a male proband 15 years of age with atypical spinocerebellar ataxia (MIM#607317) associated with compound heterozygous variants in *VPS13D* **a**, NM_015378.3:c.941+3A>G (red asterisk) and **b**, NM_015378.3:c.10270C>T (green asterisk). **c-e**, Though initially referred for the maternal c.941+3A>G splice-site

variant, both variants were anticipated to induce splicing defects and shown to do so by RT-PCR. **c**, The maternal c.941+3A>G variant induces exon 9 skipping (red splice junction and arrow) and intron 9 retention (green box and arrow), with residual normal splicing of exons 8-9-10. **d**, The paternal c.10270C>T induces exon 51 skipping (red splice junction and arrows) or intron 51 retention (green box and arrow) with residual normal splicing of exons 50-51-52. Thus, whether due to a splicing defect or encoded p.(Gln3424*) nonsense variant, all transcripts arising from the paternal *VPS13D* allele are loss-of-function. Proband (P: male, 17 years); mother (M: female, 43 years); father (F: male, 47 years); controls (C1: male, 28 years; C2: male, 28 years). **f**, Exon 9 skipping and elevated levels of intron 9 retention induced by the maternal c.941+3A>G variant is obscured by low read depth at the 5' end of long *VPS13D* transcripts, due to 3' bias due to polyA capture and/or 5' mRNA decay compounded by NMD. Importantly for variant interpretation, complete loss of *VPS13D* is associated with murine early embryonic lethality (MGI:2448530) and cellular non-viability[6-8]. Therefore, residual normal splicing of *VPS13D* from the maternal allele did not refute likely pathogenicity of the c.941+3A>G variant, as two loss-of-function variants is likely to be associated with embryonic lethality. On critical evaluation of collective genetic, phenotypic and functional evidence supporting likely haploinsufficiency of encoded full-length vacuolar protein-sorting 13D, c.941+3A>G was re-classified likely pathogenic. **g**, RNA-seq showed clear evidence for exon 51 skipping arising from the paternal allele at the 3' end of the gene and evidence for allele bias (lower levels of paternal transcripts with c.10270T due to active NMD).
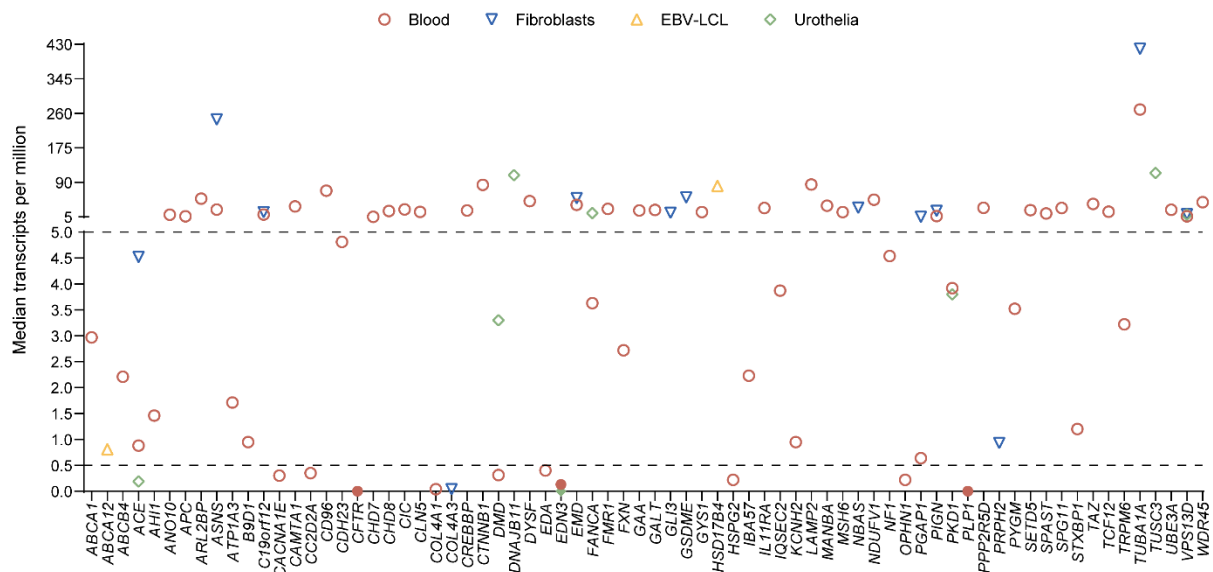
Figure S9: Median transcripts per million (TPM) values for all genes studied in the tissue available for testing.

TPM values for cases studied using blood (red circles), fibroblasts (blue triangles), EBV-LCLs (orange triangles), and urothelia (green diamonds). Shapes with no fill represent cases that were successfully studied by RT-PCR. Filled shapes represent cases that failed RT-PCR amplification due to low expression levels. Genes with TPMs above 0.5 could be reliably studied by 35 cycles of RT-PCR. Retrospective analyses indicate a TPM of ~5 was required for diagnostically informative RNA-seq (50M paired-end reads).

References (for supplementary information)

1. Zhou T, Benda C, Dunzinger S, et al. Generation of human induced pluripotent stem cells from urine samples. *Nat Protoc*. 2012;7(12):2080-2089.
2. Dawes R, Lek M, Cooper ST. Gene discovery informatics toolkit defines candidate genes for unexplained infertility and prenatal or infantile mortality. *NPJ Genom Med*. 2019;4(1):8.
3. Akesson LS, Bournazos A, Fennell A, et al. Rapid exome sequencing and adjunct RNA studies confirm the pathogenicity of a novel homozygous ASNS splicing variant in a critically ill neonate. *Hum Mutat*. 2020;41(11):1884-1891.

4. Katiyar D, Anderson N, Bommireddipalli S, Bournazos A, Cooper S, Goel H. Two novel B9D1 variants causing Joubert syndrome: Utility of mRNA and splicing studies. *Eur J Med Genet*. 2020;63(9):104000.

5. Sun S, Zhang Z, Sinha R, Karni R, Krainer AR. SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. *Nat Struct Mol Biol*. 2010;17(3):306-312.

6. Blomen VA, Májek P, Jae LT, et al. Gene essentiality and synthetic lethality in haploid human cells. *Science*. 2015;350(6264):1092-1096.

7. Hart T, Chandrashekhar M, Aregger M, et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell*. 2015;163(6):1515-1526.

8. Wang T, Birsoy K, Hughes NW, et al. Identification and characterization of essential genes in the human genome. *Science*. 2015;350(6264):1096-1101.

Chapter 4

Predicting variant associated mis-splicing

4.1 Overview

Splice prediction tools used for variant curation have become far more proficient at identifying if a variant will disrupt splicing but fail to predict the precise outcomes and the relative frequencies with which they occur. This chapter describes a dataset of natural unannotated splice junctions from >300,000 publicly available RNA-seq samples termed 300K-RNA. Ranking the most common unannotated splice junctions (Top-4) utilised at each exon-intron junction accurately predicts splicing outcomes when a genetic variant disrupts the splice-site at a given exon-intron junction.

For RNA Diagnostic testing performed in our laboratory (Chapter 3), we routinely interrogated RNA-seq data to assess patterns of alternative splicing of the target gene between the manifesting tissues and clinically accessible specimens. We observed that the predominant, variant-associated mis-splicing events identified in patients often observed as rare splice-junctions in control RNA-seq data.

Using a cohort of experimentally-verified splice-altering variants, predominantly from Chapter 3 RNA diagnostics cohort, 300K-RNA Top-4 substantively outperformed

4.1 Overview

current machine learning splice prediction tools. 300K-RNA has utility in both variant curation and strategic experimental design of RNA assays to specifically target probable mis-splicing events. Indeed, retrospective RNA reanalysis identified additional mis-splicing events missed during initial RNA analysis for 3/4 cases examined. We propose new recommendations for consideration of 300K-RNA Top-4 for application of the PVS1 null variant criterion for classification of essential splice site variants.

This chapter was prepared as an analysis article type manuscript, now under revision at *Nature Genetics*, for which I am joint first author. My contributions were initial observation that mis-splicing events present as rare splice junctions in RNA-seq data during analysis for our RNA diagnostics service, experimental validation for the majority of splicing variant cohort and RNA reanalysis, Figure 1A,B, Figure 3 and Figure 4, writing and editing manuscript.

Dawes R, **Bournazos AM**, Bryen SJ et al. SpliceVault: predicting the precise nature of variant-associated mis-splicing. *Nat Genet (NG-AN59054R1)*. Under Revision.

SpliceVault: predicting the precise nature of variant-associated mis-splicing.

Ruebena Dawes[1,2,a], Adam Bournazos[1,2,a], Samantha J. Bryen[1,2,3], Shobhana Bommireddipalli[1], Himanshu Joshi[1,b] and Sandra T. Cooper[1,2,3,b*]

[1]Kids Neuroscience Centre, Kids Research, Children's Hospital at Westmead, Sydney, NSW 2145, Australia

[2]Discipline of Child and Adolescent Health, Faculty of Health and Medicine, University of Sydney, Sydney, NSW2006, Australia

[3]The Children's Medical Research Institute, 214 Hawkesbury Road, Westmead NSW 2145, Sydney, Australia

a Ruebena Dawes and Adam Bournazos contributed equally as joint first authors

b Himanshu Joshi and Sandra Cooper contributed equally as joint senior authors.

Corresponding author:

Professor Sandra Cooper

Joint Head, Scientific Director, Kids Neuroscience Centre, The Children's Hospital at Westmead, Locked Bag 4001, Sydney, NSW 2145, Australia.

Discipline of Child and Adolescent Health, Faculty of Health and Medicine, University of Sydney, NSW 2006, Australia.

Telephone: (+61) (02) 9845 1455

E-mail: sandra.cooper@sydney.edu.au

4.2 Predicting variant associated mis-splicing

Abstract

Clinical interpretation of splicing variants depends critically upon the nature of variant-associated mis-splicing and consequence(s) for the encoded gene product. Arrestingly, ranking the four most common unannotated splicing events across 335,301 reference RNA-sequencing samples (300K-RNA Top-4), identifies the nature of variant-associated mis-splicing with remarkable prescience. 300K-RNA Top-4 correctly identifies 96% of exon-skipping events and 82% of cryptic splice-sites induced by 86 variants across 72 genes and 139 affected individuals or heterozygotes subject to RNA Diagnostics. In comparison, applying interpretative rules to SpliceAI Δ-scores correctly identifies 55% of exon-skipping events and 67% of cryptic splice-sites. Importantly, RNA re-analyses showed we had missed 300K-RNA Top-4 events for several early cases tested prior to 300K-RNA. In conclusion, 300K-RNA provides an evidence-based method that predicts with 91% sensitivity the nature of variant-associated mis-splicing. The SpliceVault web portal allows users easy access to 300K-RNA, to augment both pathology consideration of PVS1 and RNA Diagnostic investigations.

4.2 Predicting variant associated mis-splicing

Introduction

Genetic variants that induce mis-splicing of precursor messenger RNA (pre-mRNA) are a common cause of inherited disorders[1,2]. Interpreting pathogenicity of a splicing variant depends on the nature of detected mis-splicing, relative to the known pathogenetic mechanism(s) of disease for that gene and disorder (i.e., loss-of-function/gain-of-function)[3–8].

Variants impacting essential splice-sites, the almost invariant GT-AG flanking each intron, are virtually guaranteed to induce mis-splicing. Due to triplet codons, mis-splicing of pre-mRNA commonly induces a frameshift or encodes a premature termination codon (PTC), supporting rationale for consideration of essential splice-site variants under the PVS1 Null Variant (**V**ery **S**trong evidence level) criterion of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG–AMP) guidelines[9]. In 2018, revised PVS1 guidelines recommended application of the PVS1 code for essential splice-site variants, at varying strengths, based upon theoretical consideration of consequences from exon-skipping, intron retention and use of any cryptic splice-site within 20 nucleotides (nt)[10]. While only ~20% of variant-activated cryptic donors are within 20 nt[11], consideration of a larger window is unfeasible in diagnostic genetic pathology, due to the large number of potential cryptic splice-sites present in the genome. In addition, factors that induce multi-exon skipping (or retention of multiple introns) associated with some splice-altering variants are unknown.

For RNA Diagnostic testing performed in our laboratory[3], we routinely interrogate RNA sequencing (RNA-Seq) data from control specimens (in house or from GTEx[12]

4.2 Predicting variant associated mis-splicing

or ENCODE[13]) to assess patterns of alternative splicing of the target gene between the manifesting tissues and clinically accessible specimens. We observed that the predominant, variant-associated mis-spliced transcript(s) identified in specimens from affected individuals and heterozygotes were often observed as rare, stochastic splice-junctions in control RNA-Seq data. Brandão and colleagues detailed a similar finding, with dominant variant-induced mis-spliced *BRCA1* or *BRCA2* transcripts often seen as rare events in disease controls[14].

In Dawes et al., 2021[11], we analysed 5145 variants activating cryptic splice-sites and established that 87% of activated cryptic splice-sites are those detected as rare splice junctions in 40,233 RNA-Seq samples from GTEx[12] and Intropolis[15] (40K-RNA database[11]). The key insight that cryptic donors activated by genetic variants are also seen as rare events in population-based RNA-Seq data, led us to explore whether other forms of variant-associated mis-splicing may be predicted by quantifying the relative prevalence of stochastic, natural, unannotated splicing events.

We therefore created 300K-RNA, an expanded resource detailing the most common unannotated splicing events local to each exon-intron junction of Ensembl[16] and RefSeq[17] transcripts, based on splice-junctions detected across 335,301 publicly available RNA-Seq samples from Genotype Tissue Expression dataset (GTEx)[12] and Sequence Read Archive (SRA)[18] (300K-RNA). 300K-RNA is updated to the GRCh38 genome assembly and is hosted in a web resource called SpliceVault, together with 40K-RNA (GRCh37)[11]. Unannotated splice-junctions in 300K-RNA constitute evidence that a splicing event is biophysically possible and possesses the requisite constellation of features for the splicing reaction to be executed. Our

4.2 Predicting variant associated mis-splicing

central hypothesis is that a genetic variant impeding or precluding spliceosomal use of an annotated splice site is most likely to enhance or activate stochastic mis-splicing events that occur naturally.

Herein we demonstrate that 300K-RNA Top-4 ranked events correctly identifies 96% of exon-skipping events (including multi-exon skipping) and 82% of activated cryptic splice-sites induced by 86 variants in 72 genes for 139 affected individuals or heterozygotes subject to RNA Diagnostics. We provide a comparison with two machine-learning methods of predicting mis-splicing; MMSplice[19] predictions of exon-skipping and SpliceAI[20] predictions of cryptic splice-site activation. We additionally apply custom interpretive rules to SpliceAI Δ-scores +/- 5000 nt of variants to infer predictions of exon skipping and intron retention.

4.2 Predicting variant associated mis-splicing

Results

A set of experimentally-verified splice-altering variants

We performed retrospective analysis of 86 variants across 72 genes that affect an annotated splice-site and are confirmed by RNA diagnostics to disrupt pre-mRNA splicing (see Methods). Reverse transcription PCR (RT-PCR), RNA-Seq, and/or minigene assay were performed for 139 affected individuals or heterozygotes with diverse Mendelian conditions[3,21–24] (Figure 1A). The majority of probands had neurological (*n*=26), skeletal muscle (*n*=20), or malformation syndrome (*n*=9) phenotypes. 32% of variants affect the essential GT (*n*=18) or AG (*n*=10) splice-sites and 68% affect the extended donor or acceptor splice-site regions. The dataset included 77 single nucleotide variants (SNVs), 2 insertions, 5 deletions and 2 deletion-insertion variants (Figure 1B).

Just over half of the variants (44/86) induced two or more mis-splicing events (Figure 1C; 147 events). Variants most frequently caused skipping of a single exon (66/147 events, 45%), followed by cryptic activation (45/147 events, 31%) and intron retention (29/147 events, 20%), and rarely caused multi-exon skipping (7/147 events, 5%) (Figure 1C).

Unannotated splicing events in 300K-RNA

The 300K-RNA database describes natural variation in splicing among 335,301 publicly available RNA-Seq samples from GTEx[12] and SRA[18] (see Methods). For each donor and acceptor in Ensembl[16] and RefSeq[17] transcripts, we collate all unannotated, stochastic splicing events surrounding that splice-site (Figure 1D-E). These splice-junctions provide experimental evidence for an executed splicing

4.2 Predicting variant associated mis-splicing

reaction using: ***a)*** a paired donor and acceptor from different introns, reflecting skipping of one or more consecutive exons normally present in that transcript (Figure 1D, *exon skipping*); or ***b)*** an annotated donor or acceptor, paired with a non-annotated acceptor or donor, respectively, indicating cryptic splicing (Figure 1E, *cryptic splicing*).

We use the four most frequent unannotated events at each exon-intron junction (300K-RNA Top-4) as our prediction of the probable mis-splicing outcomes induced by disruption of an annotated splice-site.

4.2 Predicting variant associated mis-splicing
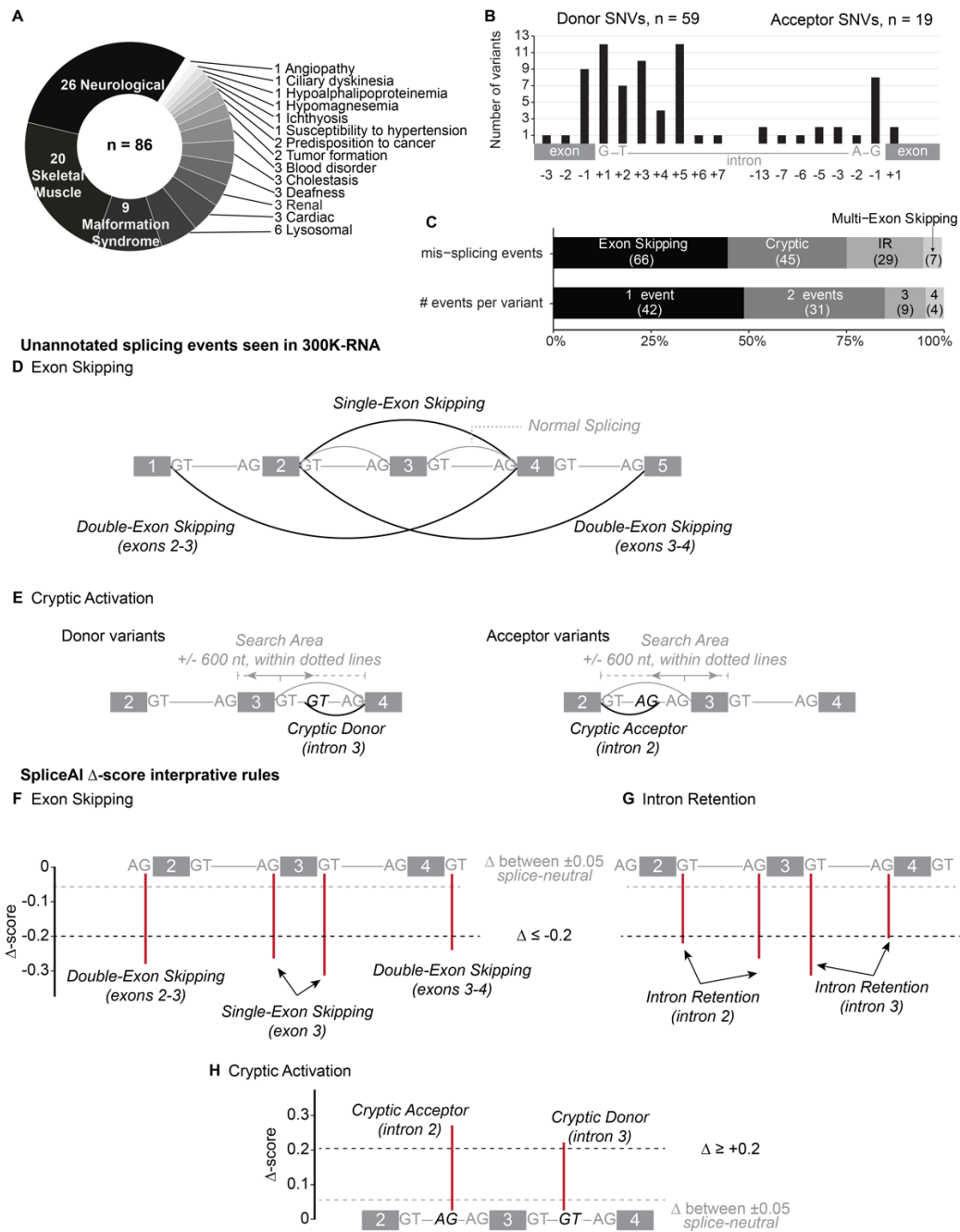
Figure 1. Cohort description and collated mis-splicing events. A) Phenotypes associated with 86 experimentally-verified clinical splicing variants and **B)** Position of the 78/86 variants that are single nucleotide variants (SNVs) relative to the essential splice-sites. **C)** Nature of 147 unannotated splicing events (mis-splicing) induced by the 86 variants. **D-E)** Natural splicing events collated within 300K-RNA, as mined from unannotated splice-junctions in public RNA-Seq data. **D)** Exon Skipping events are evidenced by split-reads spanning non-consecutive exons within the transcript. **E)** Cryptic Activation events are evidenced by split-reads spanning: *i)* an annotated acceptor and an unannotated donor or *ii)* an annotated donor and an unannotated acceptor. **F-H)** Custom interpretive rules applied to

4.2 Predicting variant associated mis-splicing

SpliceAI Δ-scores to predict the nature of mis-splicing. Heights of red lines denote example Δ-scores which would predict mis-splicing events according to our rules. **F)** *Single-Exon Skipping* is predicted if both splice-sites flanking the *exon* have a donor and acceptor loss Δ-scores ≥ 0.20, and *Double-Exon Skipping* was inferred if the splice-site of the upstream or downstream intron also had donor loss or acceptor loss Δ-score ≥ 0.20. **G)** *Intron Retention* was predicted if both splice-sites flanking an *intron* had donor loss and acceptor loss Δ-scores ≥ 0.20. **H)** *Cryptic Activation* was predicted by donor gain or acceptor gain Δ-scores ≥ 0.20 for any unannotated donor or acceptor.

Validation of 300K-RNA Top-4 and comparative analysis with deep learning predictions of variant associated mis-splicing

We compare the accuracy of 300K-RNA Top-4, with machine/deep learning algorithms SpliceAI and MMSplice, to predict the nature of mis-splicing induced by our experimentally-verified set of 86 splice-altering variants.

We adapt SpliceAI to offer a prediction of the nature of mis-splicing, by assessing Δ-scores +/-5000 nt of the variant. As scanning +/- 5000 nt of each annotated splice-site produces 20,000 Δ-scores, we exclude Δ-scores ≤ 0.05 as our assigned SpliceAI threshold for splice-neutral outcomes (grey dashed line, Figure 1F-H and Figures 2A and 2B). We employ the recommended high sensitivity threshold of Δ ≥ 0.2[20] as a prediction of variant associated mis-splicing (black dashed line, Figure 1F-H and Figures 2A-B). *Exon Skipping* is inferred for donor and acceptor variants if both splice-sites flanking an *exon* have a Δ ≥ 0.2 (Figure 1F). *Double-Exon Skipping* is inferred if the relevant splice-site of the upstream or downstream intron also has a Δ ≥ 0.2 (Figure 1F). *Intron Retention* is inferred if both splice-sites flanking an *intron* have a Δ ≥ 0.2 (Figure 1G). *Cryptic Activation* is predicted by Δ ≥ 0.2 for any unannotated donor or acceptor (Figure 1H). SpliceAI predicts at least one mis-

4.2 Predicting variant associated mis-splicing

splicing event for 76/86 variants according to these interpretative rules. For the remaining 10/86 variants, either a $\Delta \geq 0.2$ was returned only for the annotated splice site (6/10, inferred as a prediction of mis-splicing without an inferred prediction of its nature), or the threshold of $\Delta \geq 0.2$ was not reached (4/10). MMSplice predictions of exon skipping were inferred using the recommended threshold of -2[19].

300K-RNA Top-4 correctly identifies 96% (70/73) of variant-activated exon skipping events confirmed by RNA studies – including 7/7 of detected multi-exon skipping events (none of which are predicted by either SpliceAI or MMSplice) (Figure 2A). SpliceAI predicts 40/73 (55%) using the $\Delta \geq 0.2$ threshold and MMSplice predicts 50/73 (68%) of the exon skipping events detected by RNA studies.

82% (37/45) of cryptics confirmed to be activated by RNA studies are in the 300K-RNA Top-4 (Figure 2B), with SpliceAI predicting 67% (30/45) of cryptics activated using the $\Delta \geq 0.2$ threshold (Figure 2B). MMSplice cannot predict cryptic activation. For intron retention events, which cannot be predicted using 300K-RNA or MMSplice, SpliceAI shows a sensitivity of only 7% (2/29) (Figure 2C). While intron retention is readily theorised, with deleterious consequences (frameshift or premature termination codon) apparent in most instances, improved methods to predict likely instances of intron retention are needed.

4.2 Predicting variant associated mis-splicing
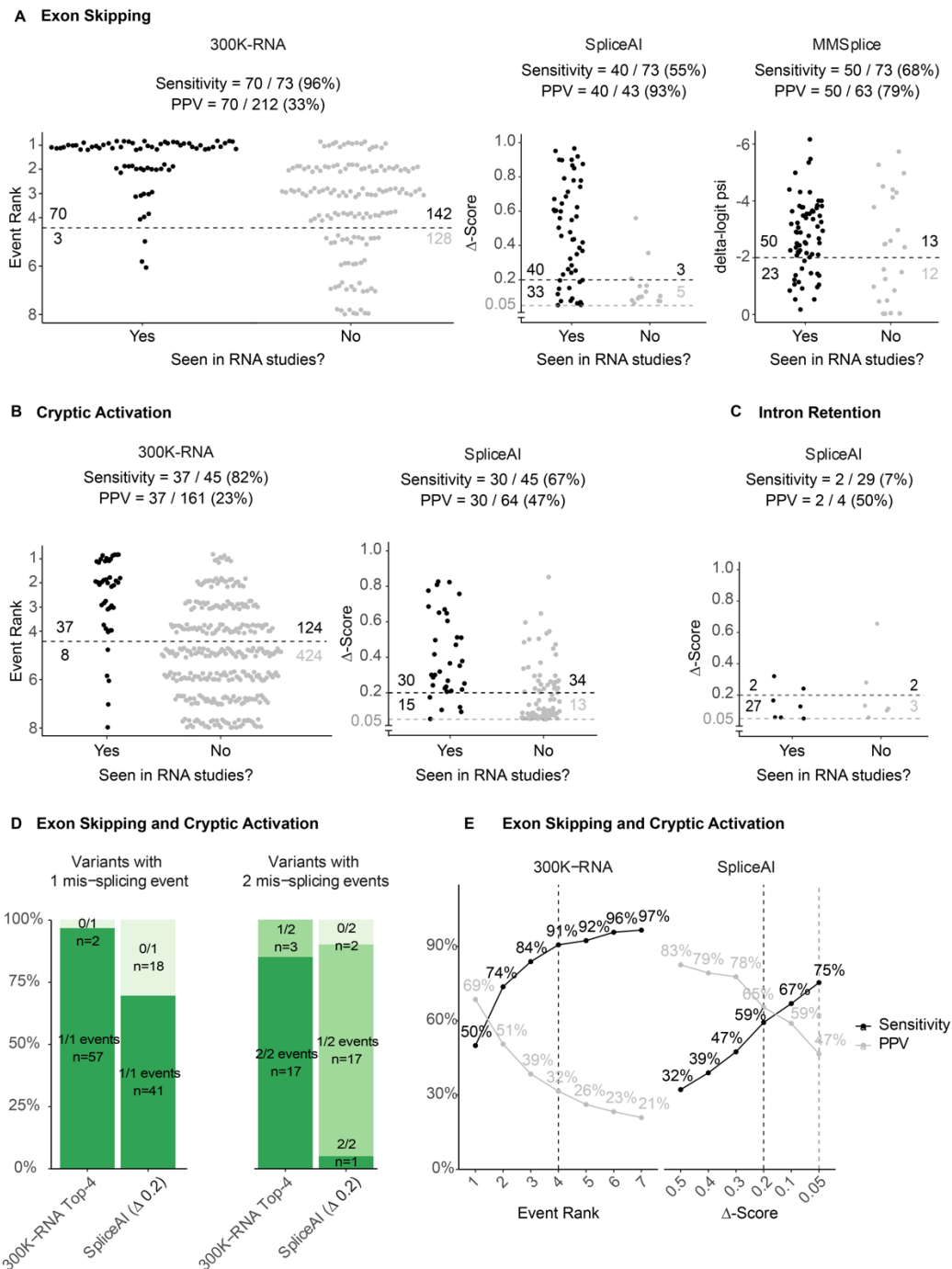
Figure 2. Accuracy of 300K-RNA, SpliceAI and MMSplice to predict the nature of variant-associated mis-splicing. A) Sensitivity and Positive Predictive Value (PPV) of 300K-RNA Top-4, SpliceAI and MMSplice to predict single- and multi- exon skipping. 300K-RNA predictions are inferred as the Top-4 unannotated splicing events proximal to the variant splice-site. Only SpliceAI Δ-scores ≥ 0.05 (grey dashed line) are depicted, with statistical metrics according to application of the high sensitivity threshold of Δ-scores ≥ 0.20 (black dashed line). For MMSplice, the author's recommended score of -2[19] was used as a prediction of exon skipping. **B)** Sensitivity and PPV of 300K-RNA and SpliceAI for predicting cryptic activation events (MMSplice only predicts exon skipping). **C)** Sensitivity and PPV of

126

4.2 Predicting variant associated mis-splicing

SpliceAI for predicting intron retention events (300K-RNA cannot currently predict intron retention). **D)** 300K-RNA shows higher sensitivity than SpliceAI for variants inducing 2 mis-splicing events. For variants with 1 mis-splicing event, 300K-RNA predicted 57/59 events and SpliceAI 41/59. For variants with 2 mis-splicing events, 300K-RNA predicted 37/40 events and SpliceAI predicted 19/40. **E)** Sensitivity and PPV of 300K-RNA and SpliceAI for predicting exon skipping and cryptic activation events at different thresholds.

300K-RNA shows high sensitivity, including for variants inducing multiple mis-splicing events

300K-RNA Top-4 predicted both events for 17/20 variants that induce two mis-splicing events (37/40 events, excluding intron retention). Whereas SpliceAI predicted only a single mis-splicing event for 17/20 variants and neither mis-splicing events for 2/20 variants (19/40 events, Figure 2D).

Remarkably, 50% of all exon-skipping and cryptic mis-splicing events activated by a splicing variant are also the Top-1 ranked event in 300K-RNA, and 91% are in the Top-4 (Figure 2E). In total, 116/118 (98%) of all variant-activated exon skipping and cryptic activation events were present in 300K-RNA (all events). The two mis-splicing events seen in RNA studies but not present in 300K-RNA were cryptic donors activated in the context of one variant. The variant in question (NM_001271208.1 :c.12018+1G>A, individual D8 in Cummings et al.[21]) strengthened a cryptic donor in *NEB* inducing its use (and predicted by SpliceAI with a Δ-score of 0.37), as well as activating another proximal cryptic donor (not predicted by SpliceAI).

RNA re-analysis identifies previously undetected Top-4 events.

4.2 Predicting variant associated mis-splicing

We noted many 300K-RNA Top-4 events not detected in our early RNA diagnostics cases (prior to 40K-RNA or 300K-RNA) involved double-exon skipping events (39%) or cryptic activation events further than 250 nt from the annotated splice site (11%), which may have been missed on initial analysis. Prior to the development of 40K-RNA[11], our laboratory practice included critical review of all cryptic splice-sites within 250 nt of the annotated donor[3]. Therefore, RNA re-analysis was performed for 4 cases (A024-*OPHN1* c.702+4A>G; A060-*GSDME* c.1183+5G>A; A014-*SPG11* c.2317-13C>G; A205-*EMD* c.266-3A>G)[3] to assess 9 undetected 300K-RNA Top-4 events.

We identified or clarified variant-associated enhanced use of 1/4 multi-exon skipping events (Figure 3A,B, *SPG11*, red), 4/4 cryptic splice sites (Figure 3C,D, *GSDME* and *EMD,* red), and 1 single-exon skipping event (Figure 3D, *EMD*, red). Skipping of multiple exons associated with *SPG11* c.2317-13C>G was not detected initially by RT-PCR due to primer placement in exons too proximal to the splice variant, and undetected by RNA-Seq due to low read depth exacerbated by NMD. Activation of two cryptic donors and two cryptic acceptors associated with *GSDME* c.1183+5G>A and *EMD* c.266-3A>G, respectively, were missed initially due to competition inherent with multi-template PCRs, heteroduplex formation and challenges resolving multi-trace chromatograms by Sanger sequencing.

4.2 Predicting variant associated mis-splicing
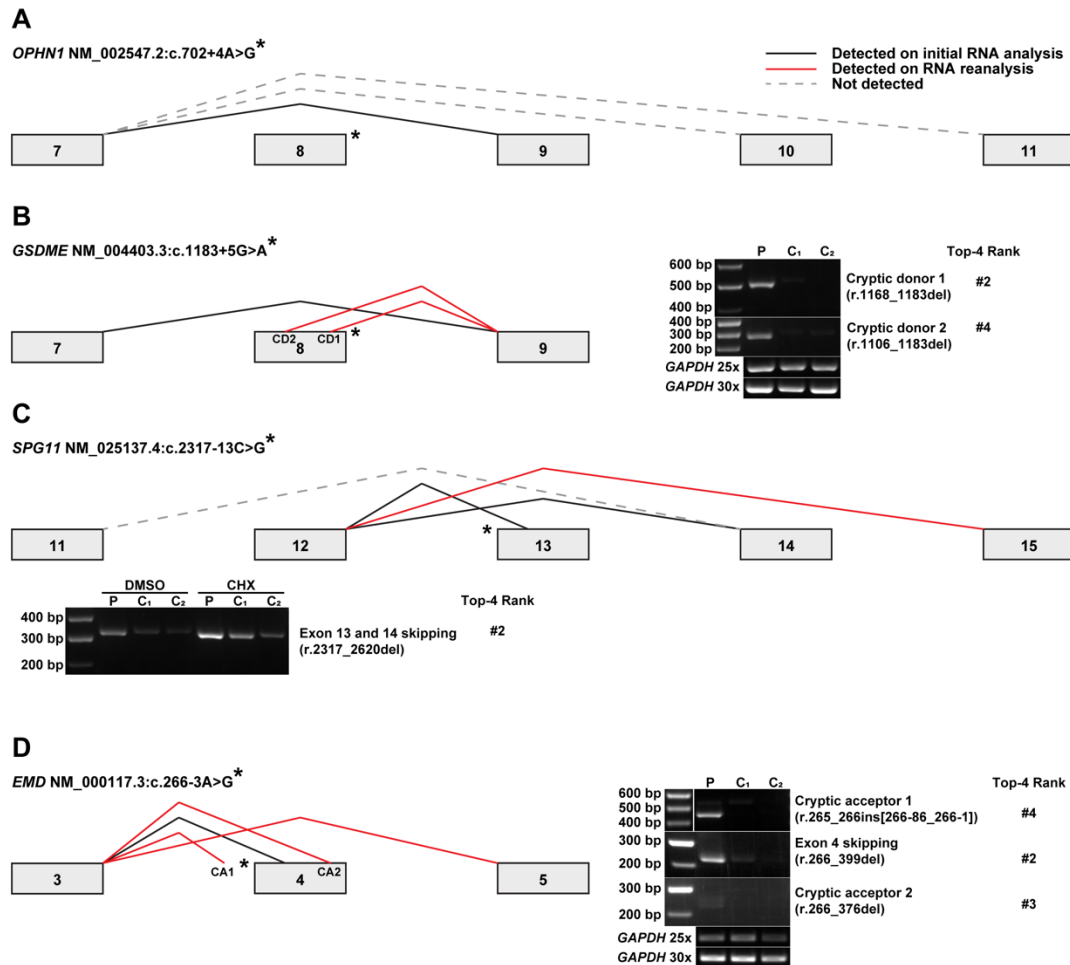
Figure 3. RNA re-analysis to check for undetected 300K-RNA Top-4 mis-splicing events. *Black* lines represent variant-associated mis-splicing identified during initial RNA analysis[3]. *Red* lines show Top-4 events detected upon re-analysis, with *Grey* lines representing Top-4 events undetected upon re-analysis. **A)** No additional Top-4 events were identified for *OPHN1* c.702+4A>G. **B)** Multi-exon skipping event associated with *SPG11* c.2317-13C>G identified upon re-analysis was not detected using initial RT-PCR primer combinations or by RNA-seq due to low read depth and NMD. **C)** RT-PCR using primers specific for two exonic cryptic donors shows their variant-associated increased use for *GSDME* c.1183+5G>A. These rare events were missed during initial RNA analysis due to PCR biases and challenges resolving Sanger sequencing chromatograms due to heteroduplex formation. **D)** RT-PCR identifies rare use of two cryptic acceptors and exon 4 skipping associated with *EMD* c.266-3A>G missed during initial RNA analysis due to PCR biases and heteroduplex formation. P=proband, C1=control 1, C2=control 2, DMSO= dimethyl sulfoxide, CHX=cycloheximide, CD1=cryptic donor 1, CD2=cryptic donor 2, CA1=cryptic acceptor 1, CA2=cryptic acceptor 2.

4.2 Predicting variant associated mis-splicing

Discussion

Clinical variant interpretation relies on predicting, or experimentally verifying, the nature of variant-induced mis-splicing to confirm variant impact on the encoded protein. This is of particular importance when applying the PVS1 (null variant) criterion to essential splice site variants[10]. No previous method can predict the nature of variant-induced mis-splicing. While the impact of exon skipping and intron retention on protein reading frame can be theorized, it has remained difficult to predict whether exon-skipping or cryptic splice-site activation will occur - and if a cryptic splice-site is activated, which one of the many potential sites present in the vicinity will be selected by the spliceosome.

Our empirical method of using 300K-RNA Top-4 accurately predicts the nature of variant associated mis-splicing with 91% sensitivity for 86 variants across a broad range of genes and disorders. Comparative analyses show 300K-RNA Top-4 outperforms SpliceAI (and MMSplice) to correctly identify exon-skipping, double-exon skipping, cryptic splice site activation and multiple mis-splicing events. We emphasize that 300K-RNA Top-4 cannot be used for variants creating or modifying the essential splice site motif of a cryptic splice site and recommend use of SpliceAI for this category of variant[11]. It is important to acknowledge the low positive predictive value of 300K-RNA Top-4 (Figure 2E, 32%) when used as proxy for a prediction of the nature of mis-splicing. However, we feel that prioritising sensitivity is of greatest importance, to avoid false negative predictions, even if this requires pathology consideration of 4 events that may include false positives.

4.2 Predicting variant associated mis-splicing

All approaches, including 300K-RNA Top-4, have limited or no ability to predict intron retention. However, our future research will focus on extending features of 300K-RNA to improve prediction of intron retention, leveraging recent advances in RNA-seq data analysis[25,26]. Remarkably, 300K-RNA Top-4 reliably predicts mis-splicing events in clinical samples (blood and fibroblasts) despite being informed by RNA-Seq splice-junction data from a vast array of tissues. Our current research is exploring whether high read depth RNA-Seq of manifesting tissues, or enrichment approaches, has the potential to further improve the accuracy of our methodology. Consideration of Top-3 events, if shown to maintain > 90% sensitivity, could substantially improve positive predictive value.

RNA re-analysis of four cases with one or more undetected 300K-RNA Top-4 events via our initial RNA Diagnostics testing revealed we had missed 6/9 of these rare events, due to experimental design and/or technical limitations. *A priori* knowledge of 300K-RNA Top-4 mis-splicing events has been transformative for our research-led clinical RNA Diagnostics program, facilitating both variant curation and strategic experimental design of RNA assays to specifically target probable mis-splicing events; expressly important for RT-PCR where primer design and extension times strongly influence which products may be amplified. Further, knowledge of all 300K-RNA events enables improved custom alignment of RNA-Seq data by informing the aligner of unannotated splice junctions.

Our reinterrogation of early cases, showing we had missed several rare events, raises the possibility that 300K-RNA Top-4 positive predictive values could be higher than we currently estimate. It also reinforces clinical benefits of being able to reliably predict probable mis-splicing events to improve completeness and accuracy of

4.2 Predicting variant associated mis-splicing

conclusions drawn from RNA diagnostics. Importantly, we cross-checked all other early cases (before 300K-RNA) to confirm that interpretation of likely pathogenicity would not be impacted by any undetected Top-4 events that may have resulted (neither feasible nor economic to re-test all specimens).

To our knowledge, 300K-RNA Top-4 is the first evidence-based method for predicting the nature of variant-associated mis-splicing and will assist clinical laboratories in application of PVS1 or PP3/BP4, prioritisation of VUS for RNA analysis, as well as guide RNA-diagnostic testing to experimentally determine consequences for pre-mRNA splicing. Figure 4 details our recommendations for application of the PVS1 criterion to essential splice site variants.

Informed by this investigation, our recommendation for application of PVS1 is theorized consideration of intron retention and the 300K-RNA Top-4: a feasible, evidence-based method to reliably assess for exon-skipping and probable cryptic activation within a larger distance window of 600 nt. We prescribe very strong, strong and moderate evidence levels based on the prediction of mis-splicing outcomes, their impact to clinically-relevant transcripts and pathogenetic mechanism of disease (Figure 4).

We provide SpliceVault, a web portal to access 300K-RNA (and 40K-RNA in hg19), which quantifies natural variation in splicing and potently predicts the nature of variant-associated mis-splicing: (https://kidsneuro.shinyapps.io/splicevault/). Users require no bioinformatics expertise and can retrieve stochastic mis-splicing events for any splice-junction annotated in Ensembl or RefSeq. Default settings display 300K-RNA Top-4 output according to the optimised parameters we describe herein,

4.2 Predicting variant associated mis-splicing

with the option to return all events, customise the number of events returned, distance scanned for cryptic splice-sites or maximum number of exons skipped. We hope SpliceVault will improve the ability to classify and study splicing variants with accuracy and completeness, avoiding the non-actionable diagnostic endpoint of a variant of uncertain significance (VUS).

	PVS1_Very strong	PVS1_Strong	PVS1_Moderate
Empirical evidence	Very strong evidence for loss-of-function. **1.** Intron retention (IR) and 300K-RNA Top-4 events all result in a frameshift or encode a premature termination codon (PTC).	Strong evidence for loss-of-function. **1.** IR and 300K-RNA Top-4 events are in-frame, out-of-frame events and/or encode a PTC, **OR** 300K-RNA Top-4 events are all in-frame.	Moderate evidence for loss-of-function. **1.** IR and 300K-RNA Top-4 events are in-frame, out-of-frame events and/or encode a PTC, **OR** 300K-RNA Top-4 events are all in-frame.
Transcript disease association.	**2.** IR and 300K-RNA Top-4 events affect splicing of one or more constitutive exon(s) present in the clinically-relevant isoform(s) expressed by the manifesting tissue(s); **AND/OR** IR and 300K-RNA Top-4 events activate inclusion of ectopic/intronic sequences into the clinically-relevant isoform(s) expressed by the manifesting tissue(s); **AND/OR**		
	3. IR and 300K-RNA Top-4 events affect an alternatively spliced exon involving a region of the gene with **multiple variants classified likely/pathogenic** establishing its functional and clinical importance.	**3.** IR and 300K-RNA Top-4 events affect an alternatively spliced exon involving a region of the gene with **one or more variants classified likely/pathogenic** establishing its functional or clinical importance.	**3.** IR and 300K-RNA Top-4 events affect an alternatively spliced exon involving a **conserved region of the gene** consistent with functional and/or structural importance.
Pathogenetic mechanism	**4.** Out-of-frame events resulting in a frameshift or PTC are predicted to activate nonsense mediated decay, and loss-of-function variants are a known causal basis for disease; **AND/OR** Out-of-frame events resulting in a frameshift or PTC are predicted to evade nonsense mediated decay (NMD), and truncating variants are an established pathogenetic mechanism for disease.	**4.** Out-of-frame events resulting in a frameshift or PTC are predicted to activate nonsense mediated decay, and loss-of-function variants are a known causal basis for disease; **AND/OR** In-frame and/or NMD-evading events altering length of the encoded protein affect a region of the gene that contains **multiple missense variants, in-frame indels or truncating variants classified as likely/pathogenic** establishing its functional and clinical importance. Collective evidence strongly supports loss-of-function outcomes resulting from the detected in-frame or NMD-evading splicing events.	**4.** Out-of-frame events resulting in a frameshift or PTC are predicted to activate nonsense mediated decay, and loss-of-function variants are a known causal basis for disease; **AND/OR** In-frame and/or NMD-evading events altering length of the encoded protein affect a region of the gene with **one or more missense variants, in-frame indels or truncating variants as classified as likely/pathogenic** establishing its functional or clinical importance. Collective evidence supports high likelihood of loss-of-function outcomes resulting from the detected in-frame or NMD-evading splicing events.

Figure 4. Guidelines for use of empirical evidence from 300K-RNA to assist application of the PVS1 criterion for essential splice site variants. We recommend pathology consideration of Intron Retention (IR) and the 300K-RNA Top-4 events for all disease relevant transcript(s). PVS1 levels of evidence are influenced by the collective nature of probable induced mis-splicing, relative to evidence supporting null outcomes for the encoded gene product. For use of PVS1 at a Very Strong evidence level, all Top-4 events should be consistent with null outcomes. PVS1 applied at Strong or Moderate should be considered with one or more in-frame events, adjusting the evidence weighting according to known clinical relevance of the affected region of the disease-relevant transcript(s) (3.) and established pathogenetic mechanism(s) associated with a given gene and disorder (4.).

4.2 Predicting variant associated mis-splicing

Methods

Creating 300K-RNA

The 300K-RNA database collates splice-junctions detected in 335,301 publicly available RNA-Seq samples from GTEx[12] and SRA[18]. 18,858 files provided in GTEx detailing splice-junctions detected in each sample were obtained from GTEx (phs000424.v8.p2). Using Datamash[27], splice-junction read counts were summarised across all samples for each unique splice-junction. Splice-junction read counts derived from 316,443 human RNA-Seq samples from SRA were similarly downloaded from the public resource recount3[18]. We then filtered split-reads to those which span at least one annotated splice site, and for each splice-junction detected we tallied the number of samples it occurred in across the two data sources.

We ranked splice junctions according to the number of samples in which the event was detected. Top-4 events herein were limited to single- or double- exon skipping and cryptic splicing using an unannotated donor or acceptor within 600 nt of the annotated donor or acceptor (distance limit chosen to maximise sensitivity). Code used to create 300K-RNA is available at https://github.com/kidsneuro-lab/300K-RNA.

SpliceAI Δ-score interpretive rules

By default, SpliceAI outputs four delta (Δ) scores: acceptor loss, acceptor gain, donor loss and donor gain; for each the maximum Δ-score within +/- 50 nt of the variant is reported. To adapt SpliceAI to the prediction of mis-splicing, we retrieved all Δ-scores +/- 5000 nt of each variant using code available at

4.2 Predicting variant associated mis-splicing

https://github.com/kidsneuro-lab/SpliceAILookup. Two Δ-scores returned at each base (variant nucleotide versus reference nucleotide) generated 20,000 Δ-scores per variant, of which we excluded all Δ-scores $\leq$ 0.05 as our assigned threshold for a SpliceAI prediction of splice-neutral outcome.

389 scores were returned above the 0.05 threshold. 83 of these were donor loss or acceptor loss scores of the affected annotated splice-site, denoting a prediction of mis-splicing. 87 were donor loss or acceptor loss scores of an unannotated splice-site and were discarded as uninterpretable. The remaining 219 predictions were manually annotated using the rules shown in Figure 1F-H.

MMSplice scores

MMSplice[19] predictions were retrieved using code from the MMSplice GitHub repository (https://github.com/gagneurlab/MMSplice_MTSplice), and a threshold of delta logit-psi $\leq$ -2 was used for a prediction of exon skipping as recommended by the authors.

Ethics declaration

Consent for diagnostic genomic testing was supported by governance infrastructure of the relevant local ethics committees of the participating Australian Public Health Local Area Health Districts. Kids Neuroscience Centre's biobanking and functional genomics human ethics protocol was approved by the Sydney Children's Hospitals Network Human Research Ethics Committee (protocol 10/CHW/45 renewed with

4.2 Predicting variant associated mis-splicing

protocol 2019/ETH11736 (July 2019 – 2024)) with informed, written consent for all participants.

RNA re-analysis

RNA extraction from whole blood, peripheral blood mononuclear cells or fibroblasts, and reverse transcription polymerase chain reaction and Sanger sequencing performed as described in Bournazos et al[3].

Table 1. Primers used for RNA re-analysis.

Case	Event	Forward (5' - 3')	Reverse (5' - 3')
OPHN1	Exons 8-9 skipping	TTCCGGAAGGAACAAATAGG	GCCTGGGCTTTAGGAATATC
	Exons 8-10 skipping	TTCCGGAAGGAACAAATAGG	CCTGGGCCCTTGGACTTA
GSDME	Cryptic donor 1	AGGTGGCTTCGAGAACAAGA	GCGCTATCTGGCATTCAAG
	Cryptic donor 2	GGGTGCAGCTTACAGGAAAT	TGCCTAAAGCACAGAGTCCA
SPG11	Exons 12-13 skipping	TGAACTTTTGAAGAATATGGTA	TGCTGGGTTTGAAATTCTCC
	Exons 13-14 skipping	CCCCAAAGTAAAGGAGAGCA	GGGGAATATGATTTGTATTCAAA
EMD	Exon 4 skipping	TCTACCAGAGCAAGGGTGC	GGTGAGCCATGAAGAGGAAG
	Cryptic acceptor 1	TCTACCAGAGCAAGGGTCC	GGTGAGCCATGAAGAGGAAG
	Cryptic acceptor 2	TCTACCAGAGCAAGGGTCC	GGTGAGCCATGAAGAGGAAG

4.2 Predicting variant associated mis-splicing

Competing Interests

Sandra T. Cooper is director and shareholder of Frontier Genomics Pty Ltd (Australia). Sandra T. Cooper currently receives no consultancy fees or other remuneration for this role. Himanshu Joshi offers Technology advice to Frontier Genomics Pty Ltd (Australia) and receives no remuneration for this role.

4.2 Predicting variant associated mis-splicing

Data Availability

All code required to create 300K-RNA is available at https://github.com/kidsneuro-lab/300K-RNA. 300K-RNA can be easily accessed and queried through SpliceVault (https://kidsneuro.shinyapps.io/splicevault/). Code used to create SpliceVault is available at https://github.com/kidsneuro-lab/SpliceVault/.

Code Availability

All code required to perform analyses and generate Figure 1C and Figure 2 is available at https://github.com/kidsneuro-lab/SpliceVault_figures.

References

1. Baralle, D. & Buratti, E. RNA splicing in human disease and in the clinic. *Clinical Science* **131**, 355–368 (2017).

2. López-Bigas, N., Audit, B., Ouzounis, C., Parra, G. & Guigó, R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* **579**, 1900–1903 (2005).

4.2 Predicting variant associated mis-splicing

3. Bournazos, A. M. *et al.* Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genetics in Medicine* (2021) doi:10.1016/j.gim.2021.09.001.

4. Wai, H. A. *et al.* Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med* **22**, 1005–1014 (2020).

5. Murdock, D. R. *et al.* Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *Journal of Clinical Investigation* (2020) doi:10.1172/JCI141500.

6. Maddirevula, S. *et al.* Analysis of transcript-deleterious variants in Mendelian disorders: implications for RNA-based diagnostics. *Genome Biology* **21**, 145 (2020).

7. Frésard, L. *et al.* Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nature Medicine* **25**, 911–919 (2019).

8. Lee, H. *et al.* Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genetics in Medicine* **22**, 490–499 (2020).

9. Richards, 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* **17**, 405–424 (2015).

10. Abou Tayoun, A. N. *et al.* Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat* **39**, 1517–1524 (2018).

11. Dawes, R., Joshi, H. & Cooper, S. T. 90% of variant-activated cryptic-donors are also (mis)used stochastically in population-based RNA-Seq data. **NCOMMS-21-28137A**, Revision under review (2021).

12.   GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

13.   Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

14.   Brandão, R. D. *et al.* Targeted RNA-seq successfully identifies normal and pathogenic splicing events in breast/ovarian cancer susceptibility and Lynch syndrome genes. *International Journal of Cancer* **145**, 401–414 (2019).

15.   Nellore, A. *et al.* Human splicing diversity and the extent of unannotated splice junctions across human RNA-seq samples on the Sequence Read Archive. *Genome Biology* **17**, 266 (2016).

16.   Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Research* **49**, D884–D891 (2021).

17.   O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745 (2016).

18.   Wilks, C. *et al.* recount3: summaries and queries for large-scale RNA-seq expression and splicing. *bioRxiv* 2021.05.21.445138 (2021) doi:10.1101/2021.05.21.445138.

19.   Cheng, J. *et al.* MMSplice: modular modeling improves the predictions of genetic variant effects on splicing. *Genome Biol* **20**, 48 (2019).

20.   Jaganathan, K. *et al.* Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**, 535-548.e24 (2019).

21.   Cummings, B. B. *et al.* Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Science Translational Medicine* **9**, (2017).

22.   Akesson, L. S. *et al.* Rapid exome sequencing and adjunct RNA studies confirm the pathogenicity of a novel homozygous ASNS splicing variant in a critically ill neonate. *Human Mutation* **41**, 1884–1891 (2020).

23.   Katiyar, D. *et al.* Two novel B9D1 variants causing Joubert syndrome: Utility of mRNA and splicing studies. *European Journal of Medical Genetics* **63**, 104000 (2020).

24.   Jones, H. F. *et al.* Importance of muscle biopsy to establish pathogenicity of DMD missense and splice variants. *Neuromuscular Disorders* **29**, 913–919 (2019).

25.   Broseus, L. & Ritchie, W. Challenges in detecting and quantifying intron retention from next generation sequencing data. *Comput Struct Biotechnol J* **18**, 501–508 (2020).

26.   Zheng, J.-T., Lin, C.-X., Fang, Z.-Y. & Li, H.-D. Intron Retention as a Mode for RNA-Seq Data Analysis. *Front. Genet.* **11**, (2020).

27.   Free Software Foundation, I. *GNU Datamash, Available at: https://www.gnu.org/software/datamash/.* (2014).

Chapter 5

Contributions to genetic diagnoses and variant classification during PhD candidature.

5.1 Overview

Chapter 5 outlines all the contributions that I made to variant reclassification or diagnoses throughout my candidature. 73/94 (78%) variants were reclassified for cases which clinical impact surveys were completed, and informative results were obtained in 91/94 cases to facilitate variant interpretation (Figure 5.1).

Of the 86/94 variants initially classed VUS, 66/86 VUS were upgraded to likely/pathogenic and 2/86 were downgraded to likely/benign after RNA studies. RNA testing was performed for 8 variants already classed likely pathogenic to provide confidence in the diagnosis or to be upgraded to pathogenic for preimplantation genetic testing/screening.

Table 5.1 lists 107 variants and my contributions towards RNA analysis that was utilised for genetic diagnosis and/or variant interpretation.
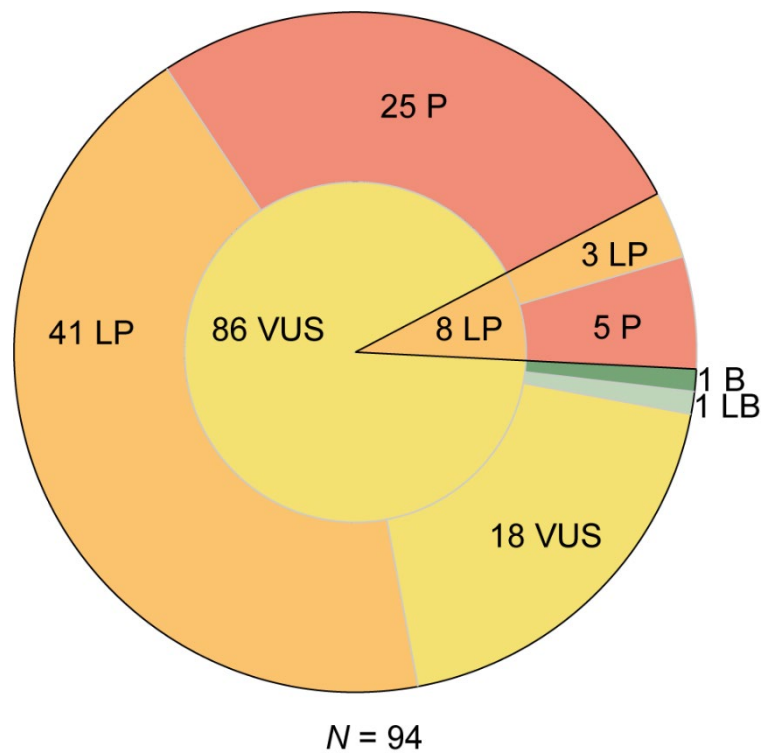
Figure 5.1: Classification prior to (inner pie) and after (outer pie) RNA analysis for 94 variants with completed clinical impact survey data. B, benign; LB, likely benign; VUS, variant of uncertain significance; LP, likely pathogenic' P, pathogenic.

Table 5.1: Phenotypic description, variant details, and contributions to variant classification and/or diagnosis.

Case ID	Section	Phenotype	Gene	No. tested	Inheritance	Variant 1 (GRCh38)	Variant 2 (GRCh38)	Contributions
A009	3.2, 4.2	Renal tubular dysgenesis	ACE	Duo	AR	Chr17:g.63483976G>C NM_000789.3:c.1709+5G>C p.?	Chr17:g.63483976G>C NM_000789.3:c.1709+5G>C p.?	RNA analysis, prepared diagnostics report
A014	3.2, 4.2	Spastic paraplegia 11, autosomal recessive	SPG11	Singleton	AR	Chr15:g.44622360G>C NM_025137.3:c.2317-13C>G p.?	Chr15:g.44584288C>T NM_025137.3:c.5392G>A p.(Glu1798Lys)	RNA analysis, prepared diagnostics report
A018	3.2, 4.2	Joubert syndrome 3	AHI1	Trio	AR	Chr6: g.135429877C>T NM_001134831.1:c.2492+5G>A p.?	Chr6:g.135457594G>A NM_001134831.1:c.1051C>T p.(Arg351Ter)	RNA analysis, prepared diagnostics report
A022	3.2, 4.2	Barth syndrome	TAZ	Duo	XLR	ChrX:g.154412214G>C NM_000116.3:c.238G>C p.(Gly80Arg)	N/A	RNA analysis, Western blot analysis, prepared diagnostics report
A024	3.2, 4.2	Mental retardation with cerebellar hypoplasia and distinctive facial appearance	OPHN1	Duo	XLR	ChrX:g.68212104T>C NM_002547.2:c.702+4A>G p.?	N/A	RNA analysis, prepared diagnostics report
A025	3.2, 4.2	Mental retardation, autosomal dominant 23	SETD5	Singleton	AD	Chr3:g.9442123C>G NM_001080517.1:c.960-5C>G p.?	N/A	RNA analysis, prepared diagnostics report
A026	3.2	Osteogenesis imperfecta	NBAS	Singleton	AR	Chr2:g.15287054_15287269dup Chr2:g.15474086_15474377dup	Chr2:g.15539327G>A NM_015909.3:c.409C>T p.(Arg137Trp)	RNA analysis, prepared diagnostics report
A029	3.2	Mental retardation, autosomal dominant 35	PPP2R5D	Trio	AD	Chr6:g.43007956G>A NM_006245.3:c.748G>A p.(Glu250Lys)	N/A	RNA analysis, prepared diagnostics report
A030	3.2	Neurodegeneration with brain iron accumulation 4	C19orf12	Singleton	AD	Chr19:g.29702836C>T NM_001031726.3:c.335G>A p.(Trp112Ter)	N/A	RNA analysis, Western blot analysis, prepared diagnostics report
A031	3.2	Mental retardation, autosomal recessive 42	PGAP1	Duo	AR	Chr2:g.196885478T>C NM_024989.3:c.1221-3A>G p.?	Chr2:g.196885478T>C NM_024989.3:c.1221-3A>G p.?	RNA analysis, prepared diagnostics report
A036	3.2, 4.2	Glycogen storage disease II	GAA	Singleton	AR	Chr17:g.80107802C>T NM_000152.4:c.861C>T p.(Pro287=)	Chr17:g.80104542T>G NM_000152.4:c.-32-13T>G p.?	RNA analysis, prepared diagnostics report
A040	3.2, 4.2	Multiple congenital anomalies-hypotonia-seizures syndrome 1	PIGN	Trio	AR	Chr18:g.62143352A>C NM_176787.4:c.923-6T>G p.?	Chr18:g.62143352A>C NM_176787.4:c.923-6T>G p.?	RNA analysis, prepared diagnostics report

5.2 Contributions to variant interpretation and diagnoses

A050	3.2, 4.2	Perrault syndrome 1	HSD17B4	Singleton	AR	Chr5:g.119506890G>C NM_000414.3:c.1333+1G>C p.?	Chr5:g.119452621G>A NM_000414.3:c.46G>A p.(Gly16Ser)	RNA analysis, prepared diagnostics report
A053	3.2	Epileptic encephalopathy, early infantile, 69	CACNA1E	Singleton	AD	Chr1:g.181577872G>A NM_001205293.1:c.616+3G>A p.?	N/A	RNA analysis, prepared diagnostics report
A054	3.2, 4.2	Angelman syndrome	UBE3A	Singleton	AD	Chr15:g.25356056C>G NM_130838.2:c.1900G>C p.(Val634Leu)	N/A	RNA analysis, prepared diagnostics report
A058	2.2, 3.2, 4.2	Asparagine synthetase deficiency	ASNS	Quad	AR	Chr7:g.97853059C>T NM_001673.4:c.1476+1G>A p.?	Chr7:g.97853059C>T NM_001673.4:c.1476+1G>A p.?	RNA analysis, prepared diagnostics report
A060	3.2, 4.2	Deafness, autosomal dominant 5	GSDME	Singleton	AD	Chr7:g.24706179C>T NM_004403.2:c.1183+5G>A p.?	N/A	RNA analysis, prepared diagnostics report
A063	3.2, 4.2	Glycogen storage disease 0, muscle	GYS1	Trio	AR	Chr19:g.48970708_48970710del NM_002103.4:c.1646-1_1647del p.?	Chr19:g.48970708_48970710del NM_002103.4:c.1646-1_1647del p.?	RNA analysis, prepared diagnostics report
A064	3.2, 4.2	Polydactyly, preaxial IV	GLI3	Singleton	AD	Chr7:g.42076747C>T NM_000168.5:c.473+5G>A p.?	N/A	RNA analysis, prepared diagnostics report
A066	3.2, 4.2	Spinocerebellar ataxia, autosomal recessive 4	VPS13D	Trio	AR	Chr1:g.12257090A>G NM_015378.3:c.941+3A>G p.?	Chr1:g.12362848C>T NM_015378.3:c.10270C>T p.(Gln3424*)	RNA analysis, prepared diagnostics report
A069	3.2, 4.2	Craniosynostosis and dental anomalies	Il11RA	Duo	AR	Chr9:g.34658683G>A NM_001142784.2:c.810G>A p.(Thr270=)	Chr9:g.34657331C>T NM_001142784.2:c.475C>T p.(Gln159*)	RNA analysis, prepared diagnostics report
A070	3.2	Cerebellar ataxia, non-progessive, with mental retardation	CAMTA1	Singleton	AD	Chr1:g.7467879 G>A NM_015215.3:c.488G>A p.(Arg163Gln)	N/A	RNA analysis, prepared diagnostics report
A077	3.2, 4.2	Spastic paraplegia 74, autosomal recessive	IBA57	Trio	AR	Chr1:g.228175032A>G NM_001010867.4:c.679+3A>G p.?	Chr1:g.228166078dup NM_001010867.4:c.262dup p.(Ala88Glyfs*22)	RNA analysis, prepared diagnostics report
A078	3.2, 4.2	Ichthyosis, congenital, autosomal recessive 4B (harlequin)	ABCA12	Singleton	AR	Chr2:g.214955362C>G NM_173076.2:c.6234-1G>C p.?	Chr2:g.214955362C>G NM_173076.2:c.6234-1G>C p.?	RNA analysis, prepared diagnostics report
A079	3.2, 4.2	Danon disease	LAMP2	Duo	XLD	ChrX:g.120442596T>A NM_013995.2:c.928+3A>T p.?	N/A	RNA analysis, prepared diagnostics report
A081	3.2, 4.2	Ciliary dyskinesia, primary, 23	ARMC4 (ODAD2)	Singleton	AR	Chr10:g.27944217C>G NM_018076.4:c.1743+5G>C p.?	Chr10:g.27860806G>T NM_018076.4:c.2840C>A p.(Ser947Ter)	RNA analysis, prepared diagnostics report

5.2 Contributions to variant interpretation and diagnoses

A085	3.2	Mitochondrial complex I deficiency, nuclear type 4	NDUFV1	Duo	AR	Chr11:g.67612375C>A NM_007103.3:c.1312C>A p.(Leu438Met)	Chr11:g.67612375C>A NM_007103.3:c.1312C>A p.(Leu438Met)	Carrier testing RNA analysis, prepared diagnostics report
A085	3.2, 4.2	McArdle disease	PYGM	Duo	AR	Chr11:g.64757779C>T NM_005609.2:c.660G>A p.(Gln220=)	Chr11:g.64757779C>T NM_005609.2:c.660G>A p.(Gln220=)	Carrier testing RNA analysis, prepared diagnostics report
A085	3.2, 4.2	Dyssegmental dysplasia, Silverman-Handmaker type Schwartz-Jampel syndrome, type 1	HSPG2	Duo	AR	Chr1:g.21861754T>A NM_001291860.1:c.4958+3A>T p.?	Chr1:g.21861754T>A NM_001291860.1:c.4958+3A>T p.?	Carrier testing RNA analysis, prepared diagnostics report
A087	3.2, 4.2	Lynch Syndrome	MSH6	Singleton	AD	Chr2:g.47805032_47805035delins CATTATTGTCAGG NM_000179.2:c.3556+5_3556+8de linsCATTATTGTCAGG p.?	N/A	RNA analysis, prepared diagnostics report
A088	3.2	Macular dystrophy, patterned 1	PRPH2	Singleton	AD	Chr6:g.42721749C>T NM_000322.4:c.581+5G>A p.?	N/A	RNA analysis, prepared diagnostics report
A089	3.2, 4.2	Hypomagnesemia 1, intestinal	TRPM6	Trio	AR	Chr9:g.74816662A>C NM_017662.4:c.1308+7T>G p.?	Chr9:g.74761771C>T NM_017662.4:c.4710G>A p.(Tro1570*)	RNA analysis, prepared diagnostics report
A091	3.2, 4.2	Dystonia 12	ATP1A3	Singleton	AD	Chr19:g.41968916C>A NM_152296.4:c.2689-1G>T p.?	N/A	RNA analysis, prepared diagnostics report
A093	3.2, 4.2	Autism, susceptibility to, 18	CHD8	Singleton	AD	Chr14:g.21403664C>G NM_001170629.1:c.3308-1G>C p.?	N/A	RNA analysis, prepared diagnostics report
A094	3.2	Neurofibromatosis, type 1	NF1	Singleton	AD	Chr17:g.31095368A>C NM_000267.3:c.59A>C p.(Gln20Pro)	N/A	RNA analysis, prepared diagnostics report
A097	3.2, 4.2	Joubert syndrome 9	CC2D2A	Duo	AR	Chr4:g.15502924G>T NM_001080522.2:c.438+1G>T p.?	Chr4:g.15502924G>T NM_001080522.2:c.438+1G>T p.?	RNA analysis, prepared diagnostics report
A098	3.2	Neuromuscular disease, congenital, with uniform type 1 fiber	RYR1	Duo	AR	Chr19:g.38490250G>A NM_000540.2:c.5989G>A (Glu1997Lys)	Chr19:g.38490250G>A NM_000540.2:c.5989G>A (Glu1997Lys)	RNA analysis, prepared diagnostics report
A100	3.2, 4.2	Lissencephaly 3	TUBA1A	Singleton	AD	Chr12:g.49188974del NM_006009.3:c.3+3del p.?	N/A	RNA analysis, Western blot analysis, prepared diagnostics report

5.2 Contributions to variant interpretation and diagnoses

A107	3.2, 4.2	Joubert syndrome 27	B9D1	Singleton	AR	Chr17:g.19347784C>T NM_015681.3:c.341G>A p.(Arg114Gln)	Chr17:g.19343405C>G NM_015681.3:c.529G>C p.(Asp177His)	RNA analysis, prepared diagnostics report
A108	3.2	Muscular dystrophy, linb-girdle, autosomal recessive 2	DYSF	Trio	AR	Chr2:g.71682530G>A NM_003494.3:c.6057G>A p.(Arg2019=)	Chr2:g.71674227_71674228del NM_003494.3:c.5698_5699del p.(Ser1900Glnfs*14)	RNA analysis, Western blot analysis, prepared diagnostics report
A115	3.2	Epileptic encephalopathy	CCDC120	Singleton	XLR	Chr23:g.49067173dup NM_001271835.1:c.957-3dup p.?	N/A	RNA analysis, prepared diagnostics report
A119	3.2	Cornelia de Lange syndrome 5	HDAC8	Singleton	XLD	Chr23:g.72488930T>C NM_018486.2:c.737+3A>G p.?	N/A	RNA analysis, prepared diagnostics report
A120	3.2, 4.2	Alport syndrome 3, autosomal dominant	COL4A3	Singleton	AD	Chr2:g.227284346 G>A NM_000091.4:c.2881+1G>A p.?	N/A	RNA analysis, prepared diagnostics report
A124	3.2, 4.2	Diamond-Blackfan anemia 7	RPL11	Singleton	AD	Chr1:g.23694793T>A NM_000975.3:c.396+2T>A p.?	N/A	RNA analysis, prepared diagnostics report
A126	3.2	Alport syndrome 3, autosomal dominant	COL4A3	Singleton	AD	Chr2:g.227266514G>T NM_000091.4:c.1408+5G>T p.?	N/A	RNA analysis, prepared diagnostics report
A127	3.2	Holoprosencephaly 12, with or without pancreatic agenesis	CNOT1	Singleton	AD	Chr16:g.58543905T>G NM_016284.4:c.4138-2A>C p.?	N/A	RNA analysis, prepared diagnostics report
A129	3.2, 4.2	Adenomatous polyposis coli	APC	Singleton	AD	Chr5:g.112775743T>A NM_000038.5:c.531+6T>A p.?	N/A	RNA analysis, prepared diagnostics report
A134	3.2, 4.2	Deafness, autosomal recessive 12	CDH23	Trio	AR	Chr10:g.71646458G>A NM_022124.5:c.1291-1G>A p.?	Chr10:g.71793483G>T NM_022124.5:c.6555G>T p.(Glu2185Asp)	RNA analysis, prepared diagnostics report
A137	3.2	CHARGE syndrome	CHD7	Trio	AD	Chr8:g.60850471T>A NM_017780.3:c.5405-22T>A p.?	N/A	RNA analysis, prepared diagnostics report
A141	3.2, 4.2	Brain small vessel disease with or without ocular anomalies	COL4A1	Trio	AD	Chr13:g.110170546 C>T NM_001845.5:c.3742+1G>A p.?	N/A	RNA analysis, prepared diagnostics report
A143	3.2, 4.2	Neurodegeneration with brain iron accumulation 5	WDR45	Singleton	XLD	ChrX:g.49075359C>G NM_007075.3:c.830+5G>C p.?	N/A	RNA analysis, prepared diagnostics report

5.2 Contributions to variant interpretation and diagnoses

A146	3.2, 4.2	C syndrome	CD96	Duo	AD	Chr3:g.111577550_111577568del NM_198196.2:c.591+1_591+19del p.?	N/A	RNA analysis, prepared diagnostics report
A154	3.2	Spastic paraplegia 4, autosomal dominant	SPAST	Singleton	AD	Chr2:g.(32145008_32147217)_(32147259_32154373)dup NM_014946.3:c.(1687+1_1688-1)_(1728+1_1729-1)dup	N/A	RNA analysis, prepared diagnostics report
A155	3.2	Epileptic encephalopathy, early infantile, 69	CACNA1E	Trio	AD	Chr1:g.181762576C>T NM_001205293.1:c.4608C>T p.(Asn1536=)	N/A	RNA analysis, prepared diagnostics report
A157	3.2	Mental retardation with cerebellar hypoplasia and distinctive facial appearance	OPHN1	Duo	XLR	ChrX:g.68192994C>T NM_002547.2:c.1202-1G>A p.?	N/A	RNA analysis, prepared diagnostics report
A161	3.2	Mitochondrial complex IV deficiency, nuclear type 5	LRPPRC	Singleton	AR	Chr2:g.43947774T>A NM_133259.3:c.1922A>T p.(Asp641Val)	Chr2:g.43947774T>A NM_133259.3:c.1922A>T p.(Asp641Val)	RNA analysis, prepared diagnostics report
A162	3.2, 4.2	HDL deficiency, familial, 1	ABCA1	Singleton	AD	Chr9:g.104796306T>G NM_005502.3:c.5237+3A>C p.?	N/A	RNA analysis, prepared diagnostics report
A164	3.2, 4.2	Epileptic encephalopathy, early infantile, 4	STXBP1	Singleton	AD	Chr9:g.127675942G>C NM_003165.3:1249G>C p.(Gly417Arg)	N/A	RNA analysis, prepared diagnostics report
A165	3.2	Ectodermal dysplasia 1, hypohidrotic, X-linked	EDA	Singleton	XLR	ChrX:g.70033533G>A NM_001399.4:c.924+5G>A p.?	N/A	RNA analysis, prepared diagnostics report
A170	3.2, 4.2	Vici syndrome	EPG5	Trio	AR	Chr18:g.45876353_45876357del NM_020964.3:c.5943-9_5943-5del p.?	Chr18:g.45876353_45876357del NM_020964.3:c.5943-9_5943-5del p.?	RNA analysis, prepared diagnostics report
A180	3.2	Mental retardation, X-linked 1/78	IQSEC2	Duo	XLD	ChrX:g.53236377G>A NM_001111125.2:c.3396C>T p.(Gly1132=)	N/A	RNA analysis, prepared diagnostics report
A181	3.2	Mental retardation, X-linked 1/78	IQSEC2	Singleton	XLD	ChrX:g.53236494C>T NM_001111125.2:c.3279G>A p.(Ser1093=)	N/A	RNA analysis, prepared diagnostics report
	3.2, 4.2	Mental retardation, autosomal recessive 44	METTL23	Duo	AR	Chr17:g.76733217dup NM_001080510.4:c.322+2dup p.?	Chr17:g.76733062_76733065del NM_001080510.4:c.169_172del p.(His57Valfs*11)	RNA analysis, prepared diagnostics report
A188	3.2, 4.2	Mental retardation, autosomal recessive 7	TUSC3	Trio	AR	Chr8:g.15650814G>A NM_001356429.1:c.426G>A p.(Gln142=)	Chr8:g.15650814G>A NM_001356429.1:c.426G>A p.(Gln142=)	RNA analysis, prepared diagnostics report

A198	3.2	Craniosynostosis 3	TCF12	Singleton	AD	Chr15:g.57253259C>G NM_001322151.1:c.1261-3C>G p.?	N/A	RNA analysis, prepared diagnostics report
A199	3.2, 4.2	Gallbladder disease 1	ABCB4	Singleton	AD	Chr7:g.87449963C>T NM_000443.3:c.833+5G>A p.?	N/A	RNA analysis, prepared diagnostics report
A200	3.2, 4.2	Mental retardation, autosomal dominant 45	CIC	Singleton	AD	Chr19:g.42287241G>T NM_015125.4:c.452+1G>T p.?	N/A	RNA analysis, prepared diagnostics report
A205	3.2, 4.2	Emery-Dreifuss muscular dystrophy 1, X-linked	EMD	Singleton	XLR	ChrX:g.154380231A>G NM_000117.2:c.266-3A>G p.?	N/A	RNA analysis, prepared diagnostics report
A206	3.2, 4.2	Fanconi anemia, complementation group A	FANCA	Singleton	AR	Chr16:g.89779856_89779866del NM_000135.2:c.1715+3_1715+13del p.?	Chr16:g.89791455T>C NM_000135.2:c.1307A>G p.(Gln436Arg)	RNA analysis, prepared diagnostics report
A208	3.2, 4.2	Duchenne muscular dystrophy	DMD	Singleton	XLR	ChrX:g.32573525C>T NM_004006.2:c.1812+5G>A p.?	N/A	RNA analysis, prepared diagnostics report
A213	3.2	Retinitis pigmentosa with or without situs inversus	ARL2BP	Singleton	AR	Chr16:g.57249857G>A NM_012106.3:c.293+5G>A p.?	Chr16:g.57249857G>A NM_012106.3:c.293+5G>A p.?	RNA analysis, prepared diagnostics report
A218	3.2, 4.2	Mast syndrome	SPG21	Singleton	AR	Chr15:g.64976556C>T NM_016630.6:c.226-1G>A p.?	Chr15:g.(64965461_64969254)_(64970223_64974601)del NM_016630.6:c.(452+1_453-1)_(669+1_670-1)del	RNA analysis, prepared diagnostics report
A219	3.2, 4.2	Friedreich ataxia	FXN	Singleton	AR	Chr9:g.69046380T>G NM_000144.5:c.166-5T>G p.?	180x intron 1 GAA repeat expansion	RNA analysis, prepared diagnostics report
A223	3.2	Marfan syndrome	FBN1	Singleton	AD	Chr15:g.48444542T>C NM_000138.4:c.6036A>G p.(Glu2012=)	N/A	RNA analysis, prepared diagnostics report
A225	3.2	Neurofibromatosis, type 1	NF1	Singleton	AD	Chr17:g.31214587dup NM_001042492.2:c.1527+2dup p.?	N/A	RNA analysis, prepared diagnostics report
A233	3.2	Neurofibromatosis, type 1	NF1	Singleton	AD	Chr17:g.31182503 T>G NM_000267.3:c.731-5T>G p.?	N/A	RNA analysis, prepared diagnostics report
A234	3.2	Alport syndrome 2, autosomal recessive	COL4A3	Singleton	AR	Chr2:g.227253638G>A NM_000091.4:c.765G>A p.(Thr255=)	Chr2:g.227307878T>C NM_000091.4:c.4421T>C p.(Leu1474Pro)	RNA analysis, prepared diagnostics report
A236	3.2	Fragile X syndrome	FMR1	Singleton	XLD	ChrX:g.147928393G>A NM_002024.5:c.270G>A p.(Glu90=)	N/A	RNA analysis, prepared diagnostics report

5.2 Contributions to variant interpretation and diagnoses

A237	3.2	Spastic paraplegia 4, autosomal dominant	SPAST	Trio	AD	Chr2:g. 32128408G>C NM_014946.3:c.1174G>C p.(Ala392Pro)	N/A	RNA analysis, prepared diagnostics report
A239	3.2, 4.2	Mannosidosis, beta	MANBA	Singleton	AR	Chr4:g.102722870C>T NM_005908.3:c.549+1G>A p.?	N/A	Carrier testing RNA analysis, prepared diagnostics report
A240	3.2, 4.2	Neurodevelopmental disorder with spastic diplegia and visual defects	CTNNB1	Singleton	AD	Chr3:g.41224951C>A NM_001904.4:c.242-3C>A p.?	N/A	RNA analysis, prepared diagnostics report
A243	3.2, 4.2	Homocystinuria due to MTHFR deficiency	MTHFR	Trio	AR	Chr1:g.11794724C>G NM_005957.5:c.1166+5G>C p.?	Chr1:g.11794466_11794480del NM_005957.5:c.1228_1242del p.(Ser410_Lys414del)	RNA analysis, prepared diagnostics report
A246	3.2	Intellectual developmental disorder, X-linked, Turner type	HUWE1	Trio	XLD	ChrX:g.53617335C>T NM_031407.6:c.1779+5G>A p.?	N/A	RNA analysis, prepared diagnostics report
A251	3.2, 4.2	Rubinstein-Taybi syndrome 1	CREBBP	Trio	AD	Chr16:g.3739724C>A NM_004380.3:c.4134G>T p.(Arg1378=)	N/A	RNA analysis, prepared diagnostics report
A252	3.2, 4.2	Sialidosis, type II	NEU1	Trio	AR	Chr6:g.31861188C>T NM_000434.3:c.615G>A p.(Gln205=)	Chr6:g.31861188C>T NM_000434.3:c.615G>A p.(Gln205=)	RNA analysis, prepared diagnostics report
A255	3.2, 4.2	Neurofibromatosis, type 1	NF1	Singleton	AD	Chr17:g.31214587dup NM_001042492.2:c.1527+2dup p.?	N/A	RNA analysis, prepared diagnostics report
A257	3.2	Intellectual developmental disorder, X-linked 99	USP9X	Duo	XLR	ChrX:g.41170628A>G NM_001039590.2:c.3027+9A>G p.?	N/A	RNA analysis, prepared diagnostics report
A271	3.2	Capillary malformation-arteriovenous malformation 2	EPHB4	Singleton	AD	Chr7:g.100823645T>A NM_004444.5:c.410A>T p.(Lys137Met)	N/A	RNA analysis, prepared diagnostics report
A277	3.2	Microphthalmia, syndromic 2	BCOR	Duo	XLD	Chr4:g.40053931dup NM_017745.5:c.4834dup p.(Leu1612Profs*6)	N/A	RNA analysis, prepared diagnostics report
A279	3.2	Aortic aneurysm, familial thoracic 7	MYLK	Singleton	AD	Chr3:g.123722123C>A NM_053025.3:c.1804+5G>T p.?	N/A	RNA analysis, prepared diagnostics report
A282	3.2, 4.2	Meckel syndrome 6	CC2D2A	Singleton	AR	Chr4:g.15596210A>G NM_001080522.2:c.4437+3A>G p.?	Chr4:g.15580047G>C NM_001080522.2:c.3851G>C p.(Arg1284Pro)	RNA analysis, prepared diagnostics report

5.2 Contributions to variant interpretation and diagnoses

A284	3.2	Aicardi-Goutieres syndrome 2	RNASEH2B	Trio	AR	Chr13:g.50931046C>G NM_024570.3:c.321+287C>G p.?	Chr13:g.50945445G>A NM_024570.3:c.529G>A p.(Ala177Thr)	RNA analysis, prepared diagnostics report
A289	3.2	Capillary malformation-arteriovenous malformation 1	RASA1	Duo	AD	Chr5:g.(87383781_87385300)_(87386904_87389392)dup NM_002890.3:c.(2758+1_2759-1)_(2925+1_2926-1)dup p.?	N/A	RNA analysis, prepared diagnostics report
A291	3.2	Orofaciodigital syndrome I	OFD1	Singleton	XLD	ChrX:g.13749411A>G NM_003611.2:c.829-16A>G p.?	N/A	RNA analysis, prepared diagnostics report
A293	3.2, 4.2	Ciliary dyskinesia, primary, 7, with or without situs inversus	DNAH11	Singleton	AR	Chr7:g.21600934A>C NM_001277115.1:c.3255+4A>C p.?	Chr7:g.21738715C>A NM_001277115.1:c.7660C>A p.(Pro2554Thr)	RNA analysis, prepared diagnostics report
A295	3.2, 4.2	Radioulnar synostosis with amegakaryocytic thrombocytopenia 2	MECOM	Singleton	AD	Chr3:g.169112783T>A NM_004991.3:c.2577+4A>T p.?	N/A	RNA analysis, prepared diagnostics report
A296	3.2	Spherocytosis, type 3	SPTA1	Duo	AR	Chr1:g.158617532C>G NM_003126.3:c.6600+5G>C p.?	Chr1:g.158620191_158620196delinsGCCC NM_003126.3:c.6391_6396delinsGGGC p.(Trp2131Glyfs*5)	RNA analysis, prepared diagnostics report
A299	3.2, 4.2	Tuberous sclerosis-1	TSC1	Singleton	AD	Chr9:g.132921360T>C NM_000368.5:c.737+3A>G p.?	N/A	RNA analysis, prepared diagnostics report
A306	3.2, 4.2	Vertebral, cardiac, renal, and limb defects syndrome 3	NADSYN1	Trio	AR	Chr11:g.71497476T>A NM_018161.5:c.1765-7T>A p.?	Chr11: g.71481963C>T NM_018161.5:c.1088C>T p.(Ala363Val)	RNA analysis, prepared diagnostics report
A311	3.2	Developmental delay with or without dysmorphic facies and autism	TRRAP	Trio	AD	Chr7(GRCh38):g.98978324G>A NM_001375524.1:c.8498+1G>A p.?	N/A	RNA analysis, prepared diagnostics report
A312	3.2	Central core disease	RYR1	Duo	AR	Chr19(GRCh38):g.38580122G>T NM_000540.3:c.14505G>T p.(Gly4835=)	Chr19(GRCh38):g.38443616A>G NM_000540.3:c.329A>G p.(His110Arg)	RNA analysis, prepared diagnostics report
A313	3.2	Encephalopathy, progressive, early-onset, with brain edema and/or leukoencephalopathy, 2	NAXD	Singleton	AR	NM_001242882.1:c.441+3A>G p.?	NM_001242882.1:c.441+3A>G p.?	RNA analysis, prepared diagnostics report

5.2 Contributions to variant interpretation and diagnoses

A314	3.2	Congenital disorder of deglycosylation	NGLY1	Singleton	AR	Chr3(GRCh38):g.25733931T>A NM_018297.4:c.1201A>T p.(Arg401*)	Chr3(GRCh38):g.25739586C>G NM_018297.4:c.872G>C p.(Arg291Pro)	RNA analysis, prepared diagnostics report
A318	3.2	Periodic fever, familial	TNFRSF1A	Singleton	AD	Chr12(GRCh38):g.6333164_6333185dup NM_001065.4:c.473-36_473-15dup p.?	N/A	RNA analysis, prepared diagnostics report
A235	3.2	Ceroid lipofuscinosis, neuronal, 3	CLN3	Trio	AR	Chr16:g.28482323T>C NM_001286104.1:c.890+4A>G p.?	Chr16:g.28482323T>C NM_001286104.1:c.890+4A>G p.?	RNA analysis, prepared diagnostics report
A340	3.2	GM1-gangliosidosis	GLB1	Duo	AR	Chr3:g.33058083A>G NM_001317040.2:c.877+6T>C p.?	Chr3:g.33058083A>G NM_001317040.2:c.877+6T>C p.?	RNA analysis, prepared diagnostics report

Chapter 6

Conclusions and perspectives

6.1 Adjunct RNA analysis for the diagnosis of rare disease

Our study demonstrates RNA analysis using clinically accessible tissues can extend diagnostic yield from genomic testing in rare disease and hereditary cancer syndromes. RT-PCR yields high diagnostic return for candidate variants, reclassifying 78% of putative splicing variants in our cohort. We expand the number of genes available to study through urine-derived epithelial cell culture. Urine is minimally invasive to collect, urothelia are simple to culture, and importantly express a different repertoire of genes to fibroblasts and blood-derived samples. For 3 cases, A234-*COL4A3*, A188-*TUSC3*, A295-*MECOM*, urothelia was the only available sample with sufficient expression for splicing analysis. Leveraging the expertise of SpliceACORD we have devised standardised procedures for RT-PCR analysis of candidate splice-altering variants.

RNA-seq remains the ideal choice for initial testing, assuming sufficient read depth is achieved, with the ability to analyse the whole transcriptome and tools to detect aberrant splicing when no candidate variant has been identified. RNA-seq was especially effective for coding variants such as A066-*VPS13D*

6.1 Adjunct RNA analysis for the diagnosis of rare disease

NM_015378.4:c.10270C>T, A223-*FBN1* NM_000138.5:c.6036A>G and A234-*COL4A3* NM_000091.4:c.765G>A, in regions with little alternative splicing. These coding variants at the end of the exon allowed for phasing of correctly spliced transcripts. Presence of the *VPS13D* c.10270C>T and *FBN1* c.6036A>G variants in correctly spliced transcripts confirmed partial mis-splicing whereas absence of the *COL4A3* c.765G>A variant in normally spliced transcripts indicated complete mis-splicing. Interrogation of allele bias of a distal heterozygous coding variant in DMSO and CHX treated samples showed no allele imbalances consistent with detected in-frame splicing outcomes associated with *COL4A3* c.765G>A. Conversely, in cases A014-*SPG11*, A295-*MECOM*, allele imbalances of a distal heterozygous coding variants that were corrected after CHX treatment provided evidence for effective degradation by NMD.

In case A277-*BCOR*, allele specific expression provided evidence of skewed X chromosome inactivation in the mother of the proband. The maternally inherited NM_017745.5:c.4834dup variant was associated with X-linked dominant inheritance of syndromic microphthalmia in the proband, yet the mother was unaffected. RNA-seq of a maternal blood sample showed apparent hemizygosity of two heterozygous coding variants in *BCOR* and the majority of heterozygous X chromosome variants confirmed by ES. This result indicates *BCOR* transcripts are expressed from a single allele in the maternal sample, likely due to skewed X chromosome inactivation.

Despite the abundance of RNA-seq tools, significant issues with mapping aberrant transcripts and filtering of diagnostically important information persist. Indeed STAR aligner[178] mis-aligned or soft clipped reads corresponding to variant associated mis-splicing in 16/24 RNA-seq samples from our cohort with adequate coverage of the target gene.

6.1 Adjunct RNA analysis for the diagnosis of rare disease

STAR aligner could not align split reads to splice junctions where a variant had created an essential splice site absent from the reference genome. In case A284-*RNASEH2B*, the deep intronic NM_024570.3:c.321+287C>G variant created an essential donor splice site that resulted in ectopic inclusion of a pseudoexon. Split reads to the pseduoexon acceptor could be aligned, however, reads corresponding to splicing from the pseudoexon donor were soft clipped. Incorporation of the pseudoexon was not initially clear due to high levels of intron retention mapping to the same intron. Only after a custom alignment could splicing to and from the pseudoexon be clearly observed (Figure 6.1). In the event diagnostically important reads are mis-aligned and/or soft clipped, we recommend performing a custom alignment to inform the aligner of variant associated aberrant splice junctions, followed by RT-PCR validation of key events.
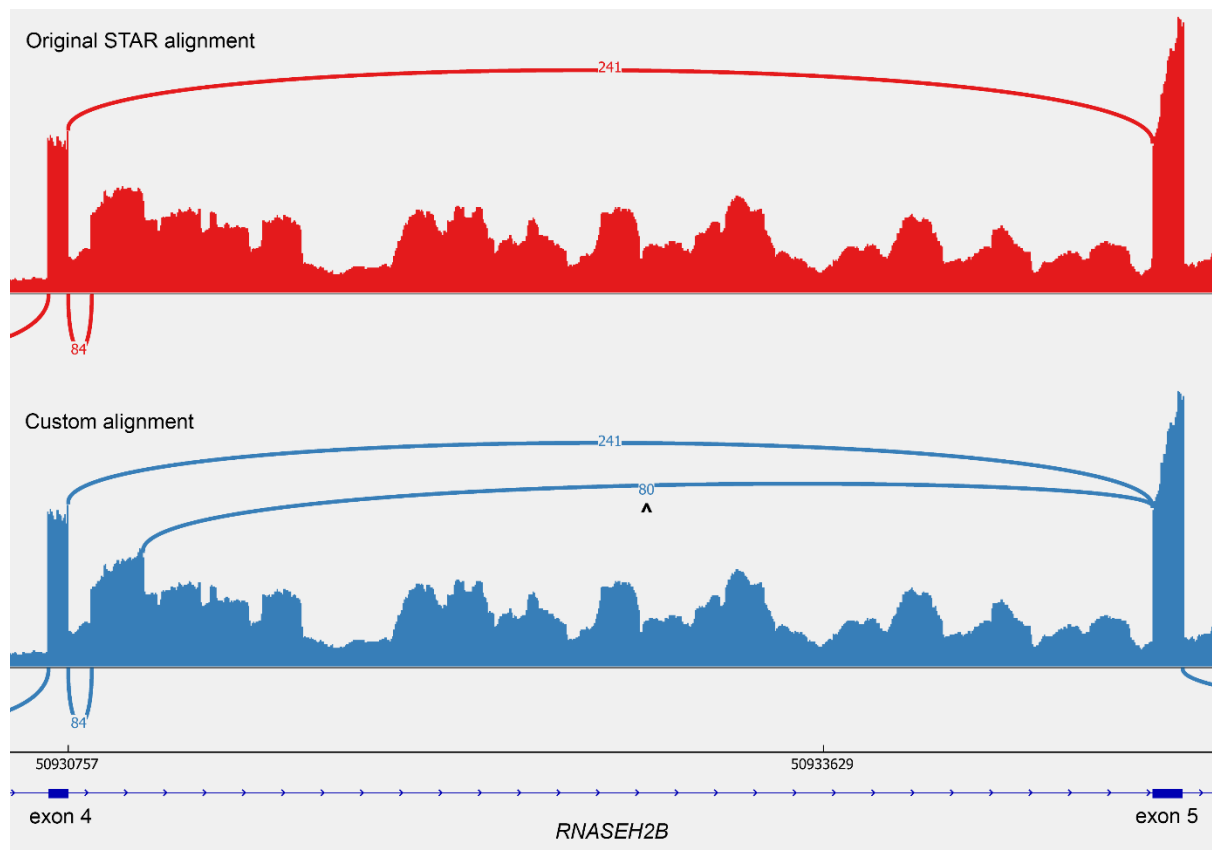
Figure 6.1:Custom alignment of a pseudoexon activated by a deep intronic variant.
A284-*RNASEH2B* NM_024570.3:c.321+287C>G variant created an essential donor splice site that resulted in ectopic inclusion of a pseudoexon into intron 4. Splicing of the pseudoexon was not initially clear due to high levels of intron 4 retention. Only after a custom alignment could 80 reads splicing from the pseudoexon donor be correctly aligned (^).

Clinical laboratories surveyed in our study emphasised the utility of distinguishing complete from partial mis-splicing for accuracy of variant interpretation. Interpretation of heterozygous splicing variants can be confounded by the presence of normal splicing expressed from the allele *in trans*. Apparent levels of aberrant transcripts are often misleading as it is difficult to distinguish partial mis-splicing from transcripts degraded by NMD. For example, in case A306-*NADSYN1* mis-splicing associated with paternal variant NM_018161.5:c.1765-7T>A appears far less abundant than normal exon 18-19 splicing (Figure 6.2). This could have been due to effective NMD or due to partial mis-splicing from a hypomorphic allele. Using a distal coding SNV

6.1 Adjunct RNA analysis for the diagnosis of rare disease

(NM_018161.5:c.1088C>T, maternally inherited) to phase correctly spliced transcripts, we could not detect any normal exon 18-19 splicing from the paternal allele. This result indicated a complete loss of normal exon 18-19 splicing from the paternal allele, providing clarity around variant interpetation.
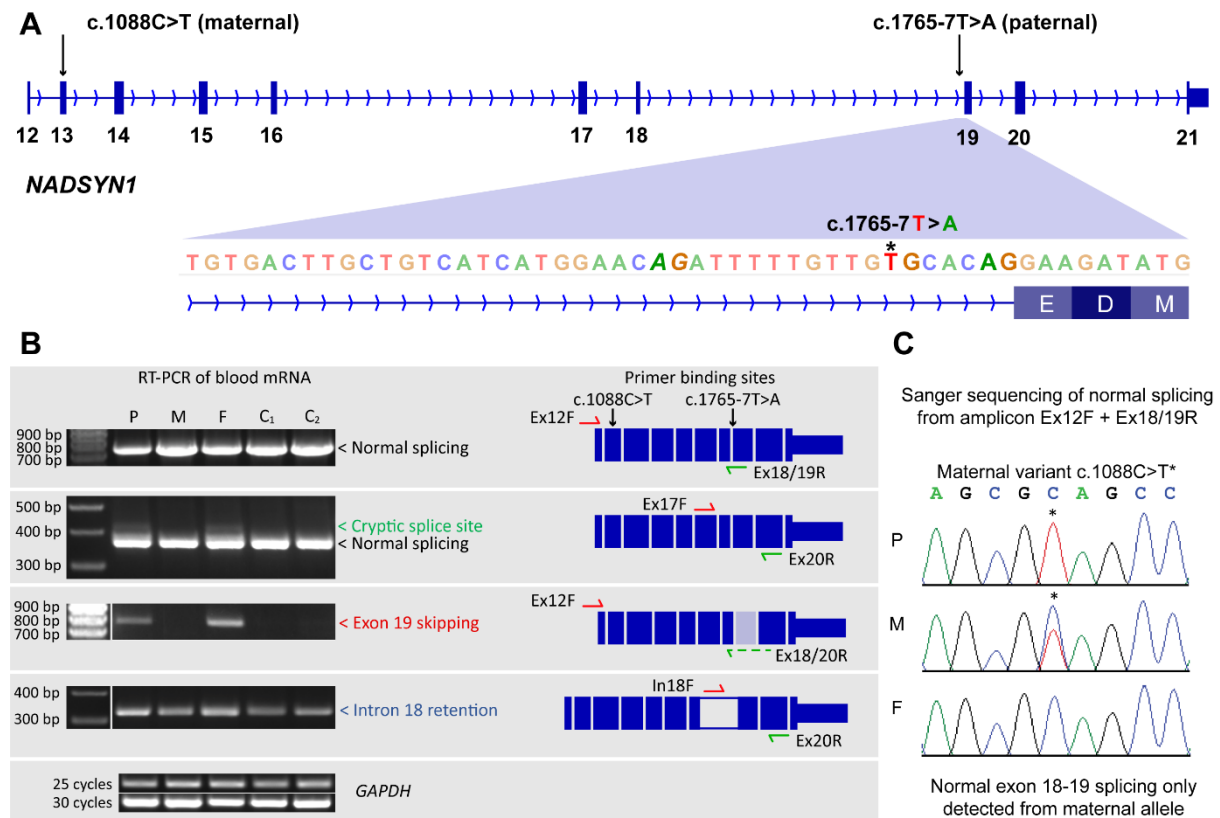
Figure 6.2: Phasing using a distal heterozygous coding variant to discern complete from partial mis-splicing. A) Positions of maternal c.1088C>T and paternal c.1765-7T>A (asterisk) variants in *NADSYN1* associated with vertebral, cardiac, renal, and limb defects syndrome 3 (MIM# 618845) in a male proband. The annotated AG acceptor is in bold whilst use of the bold and italicised AG acceptor is unique to proband. **B)** RT-PCR and Sanger sequencing analysis for case A304-NADSYN1. Using multiple primer pairs we show that the paternal variant results in use of a cryptic acceptor within intron 18, exon 19 skipping and increased levels of intron 18 retention relative to controls. Notably the variant associated mis-splicing outcomes appear less abundant than normal exon 18-19 splicing in the proband. **C)** Sanger sequencing of an amplicon specific for normal exon 18-19 splicing that also encompasses the maternal variant (asterisk) shows all normal exon 18-19 splicing in the proband is detected from the paternal allele. This result indicates the paternal variant results in a complete loss of normal splicing. P = proband, M = mother, F = father, C_1 = control 2, C_2 = control 2.

6.1 Adjunct RNA analysis for the diagnosis of rare disease

Phasing correctly spliced transcripts provided evidence for partial mis-splicing for 7 variants and complete mis-splicing for 23 variants in our cohort. This proved more informative than attempting quantitation of mis-spliced transcripts where NMD may be acting on one or both alleles. In our opinion, phasing normally spliced transcripts using a coding SNV is more diagnostically useful evidence and preferred by laboratories we surveyed. In most cases short read RNA-seq had insufficient read length for phasing and we recommend subsequent RT-PCR to obtain this evidence.

We assessed diagnostic utility of NMD inhibition using CHX treatment to detect aberrant transcripts degraded by NMD. In most cases we did not observe an increase in sensitivity to enable detection of additional mis-spliced transcripts, rather a shift in the relative abundance of transcripts (see A054-*UBE3A* and A014-*SPG11*). Confirming aberrant transcripts were being effectively targeted by NMD was diagnostically useful, however, this additional evidence would not have changed classification of most variants at the expense of doubling experimental work and costs. Therefore, routine CHX treatment is not recommended. CHX treatment has been most informative in instances where only one pathogenic variant has been identified in a candidate gene and assessing allele imbalance with and without CHX shows the allele *in trans* is being degraded, encouraging further investigations to identify the second variant.

Although *in silico* splice prediction tools have become increasingly more accurate, we are still far from being able to curate splicing variants based on predictions alone. Even when clinical laboratories begin to offer RNA analysis of splicing variants, *in silico* tools will be important for curation and identification of candidate variants. In 300K-RNA we have created an accurate empirical method for prediction of variant associated mis-splicing that is simple to use without requirement of bioinformatics

expertise and will only become more powerful with additional RNA-seq data in the future. 300K-RNA can accurately predict multiple mis-splicing and outperforms machine learning tools for many splice-altering variants. 300K-RNA will aid diagnostic scientists with variant curation of essential spice site variants, where mis-splicing is virtually guaranteed, and using our recommendations, assist in the application of the PVS1 criterion.

Using collective knowledge gained through *in silico* analysis, experimental validation of >120 splice altering variants and multidisciplinary expertise of SpliceACORD we have devised the first interpretation guidelines for RNA assay data to inform variant interpretation and diagnosis of rare disease. It is impossible to account for every type of variation and genomic context, but we anticipate these guidelines form a solid foundation for interpretation of most splicing variants. Further iterative development of interpretation guidelines will be needed as experience and understanding increases, with specific recommendations for distinct classes of variants and disorders.

6.2 Current challenges and future directions

Establishing standards for technical and analytical best practices will expand the utility of RNA analysis. The inevitable translation into routine clinical testing will provide increased access to testing for families impacted by rare disease and cancer.

RT-PCR has great diagnostic sensitivity for candidate splicing variants when priming for all possible outcomes. Clinical laboratories have used RT-PCR for years in cancer diagnostics making it readily translatable[179–184]. The quick turnaround from

sample receipt to informative RT-PCR results allowed rapid RNA analysis for 8 urgent acute care cases in our cohort[181]. Conversely, priming for all possible events to overcome the targeted nature of amplification is extremely laborious, making routine RT-PCR testing impracticable for splice variants outside of confirming key mis-splicing events and/or phasing of transcripts.

RNA-seq is seemingly the best approach for initial RNA analysis if sufficient read depth is not a concern. There are several technical stages to RNA-seq and currently an overwhelming number of approaches exist for library preparation, trimming, alignment, counting, normalisation, differential expression and splicing outlier analysis to generate hundreds of combinations processing pipelines[111,189–191]. Improved RNA-seq computational workflows for rare disease and tools for the detection of aberrant splicing are continuing to emerge[110,192], for instance Yepez et al have combined expression outlier, splicing outlier and mono allelic expression detection into a single computational workflow, DROP (Detection of RNA Outlier Pipeline)[192]. Murdock et al utilised DROP to diagnose 17% of an ES/GS negative Mendelian disease cohort testing RNA from whole blood and skin fibroblasts[102]. Yepez et al achieved a diagnosis for 16% of ES negative cases performing routine clinical RNA-seq in parallel for patients with a suspected Mendelian disease[193]. DROP also allows for incorporation of clinically-relevant gene lists and disease phenotypes such as OMIM genes[194] and Human Phenotype Ontology[195] terms, to reduce the number of genes analysed for outlier detection.

Incorporation of phenotypic data to identify strong candidate genes significantly increases likelihood of diagnosis by RNA-seq[109]. Matching unrelated cases with overlapping clinical presentations and either, variants in the same rare disease gene, or the same rare variant, has proven utility in candidate prioritisation[196]. Linking

6.2 Current challenges and future directions

similar splice variants could provide sufficient evidence for transcript-disease association and implicate pathogenicity. Multiple tools exist to connect unrelated cases with matching genotypes and phenotypes, enabling diagnosis and novel disease-gene discovery[196–198]. Matchmaker Exchange, which links eight matchmaking tools, has more than 120,000 case submissons across 13,600 unique genes[196]. The Gene Curation Coalition is an international collaboration that provides expert consensus on Mendelian gene-disease associations to improve diagnostic outcomes of genomic testing[199]. Shariant enables sharing of curated variants between clinical laboratories in Australia[200]. As these data bases grow, linking patient and variant data will continue to benefit rare disease diagnosis and interpretation of genomic testing for patients.

Reporting RNA-seq data in an easily digestible way for clinical laboratories so that scientists can review the data and form their own interpretation has proved difficult for several cases. Aberrant splicing in regions with many alternative splice junctions or very large introns as well as capturing aberrant splice junctions, allele imbalances and explanation of diagnostically important mis-aligned and/or soft clipped reads can be difficult to communicate without taking numerous screenshots of sequencing reads and sashimi plots. For example, intron retention and several large multiexon skipping events across 30kb of sequence were observed in control samples for A314-*NGLY1*. The resulting sashimi plots were extremely complicated and obscured the variant associated increase in relative levels of intron retention, as well as single exon and multiexon skipping events unique to the proband. Similarly for A186-*METTL23*, five alternative exon 1 donor splice sites, and alternative splicing of exons 2 and 3, obscured single and multiexon skipping events associated with NM_001080510.4:c.322+2dup. We are still working on a better solution where

6.2 Current challenges and future directions

impact to splicing can be more clearly visualised in conjunction with tabulated results providing relative quantitation of splice junction usage to aid interpretation.

Emergent long read sequencing technologies have several advantages over RNA-seq including greater mapping accuracy and transcript isoform identification whilst continuing to surmount initial caveats of high costs, sequencing error rate, throughput challenges and immature analysis pipelines[201]. Long read DNA sequencing has performed well for analysis for repetitive regions, pseudogene discrimination, transposable elements, structural variants, and phasing[127,201,202]. Long read transcriptome sequencing and direct RNA sequencing has identified many full length and previously unannotated isoforms, transcription start sites and polyadenylation sites, and is now showing efficacy in rare disease diagnosis and splicing analysis[203,204]. De Roeck et al demonstrated the benefits of long reads to detect intron retention, link multiple aberrant splice junctions to a single variant and phase splicing events associated with *ABCA7* variants[205]. de Jong et al have shown long reads can be used in conjunction with RT-PCR for a targeted approach that overcomes the disadvantages of Sanger sequencing amplicons from multi-template RT-PCRs[206]. As with RNA-seq, users of long read sequencing technologies already have an overwhelming choice of available tools for processing long read data, with 587 tools in the long read tools database as of March 2022[207].

Controlling for tissue specific and developmental regulation of splicing, as well as age and sex, demands a diverse range of control biospecimens. Perhaps in time, with the accumulation of RNA-seq data in clinical laboratories and expansion of RNA-seq data repositories, 'control transcriptomes' could be established removing the need for controls to be tested concurrently with each sequencing run as with massively parallel sequencing for constitutional variants[208]. For now, collaboration is

162

needed between research and clinical laboratories to accumulate this vast set of specimens, with appropriate ethics approval, to properly control RNA analyses.

Access to RNA diagnostic testing is another challenge for families suffering from rare disease. Despite the cost effectiveness of genomic sequencing[15,209,210], in Australia, only a small portion of available genomic testing is covered by Medicare leaving those without insurance or funds to directly cover costs with limited access to testing[211,212]. Testing may be ordered by clinicians if their department has a standalone budget, however, much of genomic testing for rare disease is accessed through research projects and clinical trials[211]. As of May 2020, five years since the first Australian laboratory received accreditation from the National Association of Testing Authorities for ES[213], ES and GS for childhood syndromes and intellectual disability for eligible children 10 years and under was added to the Medicare Benefits Schedule (MBS) in a win for eligible families[214]. In addition to lack of funding support for RNA testing, lack of splicing expertise and staffing resources to offer subsequent resource intensive RNA analysis remain significant challenges to translation into routine clinical practice.

6.3 Clinical impact of RNA diagnostics

In our study, clinicians reported RNA diagnostics had a positive impact for 75% of families recruited to date. 73% of reports were used for genetic counselling and 33% to inform clinical care. The greatest impact was to family planning with 44% were used for prenatal counselling and giving families the option of for preimplantation genetic testing or screening in 25% of cases. This was especially important to many families who had previously suffered one or more miscarriages or termination of

pregnancy and using assisted reproduction services. For instance, RNA analysis for A089-*TRPM6* secured the diagnosis of a treatable disease for which the family had previously lost another child, likely to the same condition. This diagnosis informed reproductive risk and will allow prenatal testing in the future.

A genetic diagnosis provided clarity around the suspected clinical diagnosis for many genes that have few reported cases and variable highly variable phenotypes. It also guided therapies and for four families enabled eligibility for clinical trials. In the case of A063-*GYS1* the patient suffered from a metabolic condition due to the of NM_002103.4:c.1646-1_1647del which was reclassified from VUS to pathogenic. This allowed the clinical team to implement a plan to prevent episodes of further deterioration, in addition to surveillance for known cardiac complications.

Determining the NM_015125.4:c.452+1G>T variant in case A200-*CIC* resulted in a loss of function had significant prognostic, and potentially surveillance recommendation implications as loss of function in *CIC* causes intellectual disability and a neuro-behavioural phenotype, whereas a gain of function causes a neurodegenerative phenotype. In the case of A134-*CDH23* and A306-*NADSYN1* reclassification of splicing VUS to pathogenic enabled use of PM3 to upgrade a missense VUS *in trans* to likely pathogenic to secure a genetic diagnosis.

Determining the *absence* of mis-splicing also provided clarity around diagnosis. Two rare candidate genes with very few or no reported cases were being considered as causal of a child's undiagnosed phenotype in cases A155-*CACNA1E* and A115-*CCDC120* respectively. Here confirming maintenance of normal splicing allowed prioritisation of a second candidate variant in both cases and the putative splicing variants were downgraded in classification.

6.3 Clinical impact of RNA diagnostics

Clinicians also reported benefits for some patients who had already received a clinical diagnosis and treatment, but not having a reason for their presentation had affected the patient's mental health. RNA testing had economic benefits in the case of A236-*FMR1*, where reclassification of NM_002024.5:c.270G>A to pathogenic assisted the patient's access to disability support funding.

Our clinical impact assessment of RNA diagnostic testing showed rare disease diagnoses had significant personal implications particularly around reproductive decision making, diagnosis of additional family members, reduced health services costs and for many provided a long sought after answer to their clinical presentation.

Below are quotes from our family feedback survey from families that were willing to share their story and express the significance of receiving a genetic diagnosis in their own words.

"We were desperately trying to find answers as to what was wrong with our son, why his neurological issues were so complex and why he didn't fit a simple ASD profile.

Having a genetic diagnosis that explains his challenges has been wonderful for our family. We can now stop questioning what has been missed and focus on helping our child.

We found comfort in this new knowledge. Thank you so much."

".. my daughter is suffering for Classical Galactosemia. It has been 6 years since my daughter was born and we were still looking for answers as to what variant had caused disruption in her GALT gene.

Your research has now found now she has a variant that has never been reported in any population databases. The information is pivotal for future genetic counselling for my daughter and our extended family.

Thank you again for your extensive research and answering the unanswered."

"For our family to have a diagnosis was really important, it meant we could be linked in with other families with children with the same genetic condition and receive the support we needed."

6.4 Concluding remarks

Unlike genomic DNA testing where sequencing is performed with respect to a single reference genome, alternative splicing and tissue/age/sex specific expression of RNA makes clinical RNA testing exceedingly more complicated. A range of different clinical and genomic contexts will require RNA for testing to be sourced from a wide range of specimens with appropriate controls. Depending on the variant context, for example whether it is hetero-/homo-/hemi-zygous, coding or noncoding, in the essential or extended splice site or deep intronic, will need different technical approaches to achieve the evidence required for accurate variant interpretation. Likely a combination of RT-PCR, RNA-seq and long read sequencing will be required to adapt to genomic context, gene expression levels and read length required to obtain adequate evidence for clinical interpretation of a give variant. Each

6.4 Concluding remarks

RNA source and technical approach have different strengths and weaknesses. There is an enormous amount of assay optimisation and standardisation required to demonstrate reproducibility and robustness of the various RNA testing pipelines before clinical laboratories can adopt RNA diagnostics for rare disease and hereditary cancer syndromes.

This body of work has laid the collaborative infrastructure in Australia to enable integration of RNA diagnostic into health services. By leveraging the multidisciplinary expertise of SpliceACORD we have established national consensus standard operating procedures for PCR based RNA diagnostics with interpretation guidelines that establish appropriate level of scientific rigour and functional evidence for RNA testing. Australia now poised to embed accredited RNA diagnostics into clinical practice to provide genetic diagnosis for families and empower clinicians managing their care.

6.4 Concluding remarks

References

1. Nguengang Wakap S, Lambert DM, Olry A, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet*. 2019;28(2):165-173. doi:10.1038/s41431-019-0508-0

2. OMIM Gene Map Statistics. Accessed September 30, 2021. https://www.omim.org/statistics/geneMap

3. Boycott KM, Rath A, Chong JX, et al. International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. *Am J Hum Genet*. 2017;100(5):695-705. doi:10.1016/J.AJHG.2017.04.003

4. Valdez R, Ouyang L, Bolen J. Public Health and Rare Diseases: Oxymoron No More. *Prev Chronic Dis*. 2019;13(1). doi:10.5888/PCD13.150491

5. Lee CE, Singleton KS, Wallin M, Faundez V. Rare Genetic Diseases: Nature's Experiments on Human Development. *iScience*. 2020;23(5):101123. doi:10.1016/J.ISCI.2020.101123

6. Ouyang L, Grosse SD, Riley C, et al. A comparison of family financial and employment impacts of fragile X syndrome, autism spectrum disorders, and intellectual disability. *Res Dev Disabil*. 2014;35(7):1518-1527. doi:10.1016/J.RIDD.2014.04.009

7. Rofail D, Maguire L, Kissner M, Colligs A, Abetz-Webb L. A Review of the

6.4 Concluding remarks

Social, Psychological, and Economic Burdens Experienced by People with Spina Bifida and Their Caregivers. *Neurol Ther*. 2013;2(1-2):1-12. doi:10.1007/s40120-013-0007-0

8.   Splinter K, Adams DR, Bacino CA, et al. Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. *N Engl J Med*. 2018;379(22):2131-2139. doi:10.1056/NEJMoa1714458

9.   Esquivel-Sada D, Nguyen MT. Diagnosis of rare diseases under focus: impacts for Canadian patients. *J Community Genet*. 2018;9(1):37. doi:10.1007/S12687-017-0320-X

10.   Wise AL, Manolio TA, Mensah GA, et al. Genomic medicine for undiagnosed diseases. *Lancet*. 2019;394(10197):533-540. doi:10.1016/S0140-6736(19)31274-7

11.   Mattick JS, Dinger M, Schonrock N, Cowley M. Whole genome sequencing provides better diagnostic yield and future value than whole exome sequencing. *Med J Aust*. 2018;209(5):197-199. doi:10.5694/mja17.01176

12.   Taylor A, Alloub Z, Tayoun AA. A Simple Practical Guide to Genomic Diagnostics in a Pediatric Setting. *Genes (Basel)*. 2021;12(6):818. doi:10.3390/GENES12060818

13.   Barbitoff YA, Polev DE, Glotov AS, et al. Systematic dissection of biases in whole-exome and whole-genome sequencing reveals major determinants of coding sequence coverage. *Sci Rep*. 2020;10(1):1-13. doi:10.1038/s41598-020-59026-y

14.   Stark Z, Tan TY, Chong B, et al. A prospective evaluation of whole-exome

6.4 Concluding remarks

sequencing as a first-tier molecular test in infants with suspected monogenic disorders. *Genet Med*. 2016;18(11):1090-1096. doi:10.1038/gim.2016.1

15.   Tan TY, Dillon OJ, Stark Z, et al. Diagnostic Impact and Cost-effectiveness of Whole-Exome Sequencing for Ambulant Children With Suspected Monogenic Conditions. *JAMA Pediatr*. 2017;171(9):855-862. doi:10.1001/jamapediatrics.2017.1755

16.   Dillon OJ, Lunke S, Stark Z, et al. Exome sequencing has higher diagnostic yield compared to simulated disease-specific panels in children with suspected monogenic disorders. *Eur J Hum Genet*. 2018;26(5):644-651. doi:10.1038/s41431-018-0099-1

17.   Seo GH, Kim T, Choi IH, et al. Diagnostic yield and clinical utility of whole exome sequencing using an automated variant prioritization system, EVIDENCE. *Clin Genet*. 2020;98(6):562-570. doi:10.1111/CGE.13848

18.   Retterer K, Juusola J, Cho MT, et al. Clinical application of whole-exome sequencing across clinical indications. *Genet Med*. 2015;18(7):696-704. doi:10.1038/gim.2015.148

19.   Lee H, Deignan JL, Dorrani N, et al. Clinical Exome Sequencing for Genetic Identification of Rare Mendelian Disorders. *JAMA*. 2014;312(18):1880-1887. doi:10.1001/JAMA.2014.14604

20.   Neveling K, Feenstra I, Gilissen C, et al. A Post-Hoc Comparison of the Utility of Sanger Sequencing and Exome Sequencing for the Diagnosis of Heterogeneous Diseases. *Hum Mutat*. 2013;34(12):1721-1726. doi:10.1002/HUMU.22450

21. Yang Y, Muzny DM, Reid JG, et al. Clinical Whole-Exome Sequencing for the Diagnosis of Mendelian Disorders. *N Engl J Med*. 2013;369(16):1502-1511. doi:10.1056/NEJMOA1306555

22. Shashi V, McConkie-Rosell A, Rosell B, et al. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet Med*. 2013;16(2):176-182. doi:10.1038/gim.2013.99

23. Sawyer SL, Hartley T, Dyment DA, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet*. 2016;89(3):275-284. doi:10.1111/CGE.12654

24. Yang Y, Muzny DM, Xia F, et al. Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA*. 2014;312(18):1870-1879. doi:10.1001/JAMA.2014.14601

25. Nambot S, Thevenon J, Kuentz P, et al. Clinical whole-exome sequencing for the diagnosis of rare disorders with congenital anomalies and/or intellectual disability: substantial interest of prospective annual reanalysis. *Genet Med*. 2017;20(6):645-654. doi:10.1038/gim.2017.162

26. Fung JLF, Yu MHC, Huang S, et al. A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. *npj Genomic Med*. 2020;5(1):1-11. doi:10.1038/s41525-020-00144-x

27. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics*. 2014;15(1):1-11. doi:10.1186/1471-2105-15-247

6.4 Concluding remarks

28.   Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of Exome and Genome Sequencing Technologies for the Complete Capture of Protein-Coding Regions. *Hum Mutat*. 2015;36(8):815-822. doi:10.1002/HUMU.22813

29.   Gilissen C, Hehir-Kwa JY, Thung DT, et al. Genome sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511(7509):344-347. doi:10.1038/nature13394

30.   Stavropoulos DJ, Merico D, Jobling R, et al. Whole-genome sequencing expands diagnostic utility and improves clinical management in paediatric medicine. *npj Genomic Med*. 2016;1(1):1-9. doi:10.1038/npjgenmed.2015.12

31.   Petrikin JE, Cakici JA, Clark MM, et al. The NSIGHT1-randomized controlled trial: rapid whole-genome sequencing for accelerated etiologic diagnosis in critically ill infants. *npj Genomic Med*. 2018;3(1):1-11. doi:10.1038/s41525-018-0045-8

32.   Ostrander BEP, Butterfield RJ, Pedersen BS, et al. Whole-genome analysis for effective clinical diagnosis and gene discovery in early infantile epileptic encephalopathy. *npj Genomic Med*. 2018;3(1):1-10. doi:10.1038/s41525-018-0061-8

33.   Clark MM, Stark Z, Farnaes L, et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *npj Genomic Med*. 2018;3(1):1-10. doi:10.1038/s41525-018-0053-8

34.   Soden SE, Saunders CJ, Willig LK, et al. Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental

disorders. *Sci Transl Med*. 2014;6(265):265ra168-265ra168.
doi:10.1126/SCITRANSLMED.3010076

35.   Willig LK, Petrikin JE, Smith LD, et al. Whole-genome sequencing for identifi
cation of Mendelian disorders in critically ill infants: a retrospective analysis of
diagnostic and clinical fi ndings. *Lancet Respir Med*. 2015;3:377-387.
doi:10.1016/S2213-2600(15)00139-3

36.   Palmer EE, Sachdev R, Macintosh R, et al. Diagnostic Yield of Whole Genome
Sequencing After Nondiagnostic Exome Sequencing or Gene Panel in
Developmental and Epileptic Encephalopathies. *Neurology*.
2021;96(13):e1770-e1782. doi:10.1212/WNL.0000000000011655

37.   Bagnall RD, Ingles J, Dinger ME, et al. Whole Genome Sequencing Improves
Outcomes of Genetic Testing in Patients With Hypertrophic Cardiomyopathy. *J
Am Coll Cardiol*. 2018;72(4):419-429. doi:10.1016/J.JACC.2018.04.078

38.   Bick D, Jones M, Taylor SL, Taft RJ, Belmont J. Case for genome sequencing
in infants and children with rare, undiagnosed or genetic diseases. *J Med
Genet*. 2019;56(12):783-791. doi:10.1136/JMEDGENET-2019-106111

39.   Dolzhenko E, Vugt JJFA van, Shaw RJ, et al. Detection of long repeat
expansions from PCR-free whole-genome sequence data. *Genome Res*.
2017;27(11):1895-1903. doi:10.1101/GR.225672.117

40.   Dong Z, Wang H, Chen H, et al. Identification of balanced chromosomal
rearrangements previously unknown among participants in the 1000 Genomes
Project: implications for interpretation of structural variation in genomes and
the future of clinical cytogenetics. *Genet Med*. 2017;20(7):697-707.
doi:10.1038/gim.2017.170

41.   Levy S, Sutton G, Ng PC, et al. The Diploid Genome Sequence of an

        Individual Human. *PLOS Biol*. 2007;5(10):e254.

        doi:10.1371/journal.pbio.0050254

42.   Ng PC, Levy S, Huang J, et al. Genetic Variation in an Individual Human

        Exome. *PLOS Genet*. 2008;4(8):e1000160. doi:10.1371/journal.pgen.1000160

43.   French JD, Edwards SL. The Role of Noncoding Variants in Heritable Disease.

        Trends Genet. 2020;36(11):880-891.

        doi:https://doi.org/10.1016/j.tig.2020.07.004

44.   Gloss BS, Dinger ME. Realizing the significance of noncoding functionality in

        clinical genomics. *Exp Mol Med*. 2018;50(8):1-8. doi:10.1038/s12276-018-

        0087-0

45.   Cooper GM, Shendure J. Needles in stacks of needles: finding disease-causal

        variants in a wealth of genomic data. *Nat Rev Genet*. 2011;12(9):628-640.

        doi:10.1038/nrg3046

46.   Lee Y, Rio DC. Mechanisms and Regulation of Alternative Pre-mRNA Splicing.

        Annu Rev Biochem. 2015;84(1):291-323. doi:10.1146/annurev-biochem-

        060614-034316

47.   Cieply B, Carstens RP. Functional roles of alternative splicing factors in human

        disease. *Wiley Interdiscip Rev RNA*. 2015;6(3):311-326.

        doi:10.1002/wrna.1276

48.   Baralle FE, Giudice J. Alternative splicing as a regulator of development and

        tissue identity. *Nat Rev Mol Cell Biol*. 2017;18(7):437-451.

        doi:10.1038/nrm.2017.27

49. Black DL. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annu Rev Biochem*. 2003;72(1):291-336. doi:10.1146/annurev.biochem.72.121801.161720

50. Wilkinson ME, Charenton C, Nagai K. RNA Splicing by the Spliceosome. *Annu Rev Biochem*. 2020;89(1):359-388. doi:10.1146/annurev-biochem-091719

51. Shi Y. Mechanistic insights into precursor messenger RNA splicing by the spliceosome. *Nat Rev Mol Cell Biol*. 2017;18(11):655-670. doi:10.1038/nrm.2017.86

52. Signal B, Gloss BS, Dinger ME, Mercer TR. Machine learning annotation of human branchpoints. *Bioinformatics*. 2018;34(6):920-927. doi:10.1093/bioinformatics/btx688

53. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. *J Appl Genet*. 2018;59(3):253-268. doi:10.1007/s13353-018-0444-7

54. Scotti MM, Swanson MS. RNA mis-splicing in disease. *Nat Rev Genet*. 2016;17(1):19-32. doi:10.1038/nrg.2015.3

55. Papasaikas P, Valcárcel J. The Spliceosome: The Ultimate RNA Chaperone and Sculptor. *Trends Biochem Sci*. 2016;41(1):33-45. doi:10.1016/J.TIBS.2015.11.003

56. Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol*. 2017;18(10):637-650. doi:10.1038/nrm.2017.63

57. Tellier M, Maudlin I, Murphy S. Transcription and splicing: A two-way street.

6.4 Concluding remarks

WIREs RNA. 2020;11(2):e1593. doi:10.1002/wrna.1593

58.   Matera AG, Wang Z. A day in the life of the spliceosome. *Nat Rev Mol Cell Biol*. 2014;15(2):108-121. doi:10.1038/nrm3742

59.   Cvitkovic I, Jurica MS. Spliceosome Database: a tool for tracking components of the spliceosome. *Nucleic Acids Res*. 2013;41(D1):D132-D141. doi:10.1093/NAR/GKS999

60.   Baralle D, Buratti E. RNA splicing in human disease and in the clinic. *Clin Sci*. 2017;131(5):355-368. doi:10.1042/CS20160211

61.   López-Bigas N, Audit B, Ouzounis C, Parra G, Guigó R. Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett*. 2005;579(9):1900-1903. doi:10.1016/J.FEBSLET.2005.02.047

62.   Jagadeesh KA, Paggi JM, Ye JS, et al. S-CAP extends pathogenicity prediction to genetic variants that affect RNA splicing. *Nat Genet*. 2019;51(4):755-763. doi:10.1038/s41588-019-0348-4

63.   Truty R, Ouyang K, Rojahn S, et al. Spectrum of splicing variants in disease genes and the ability of RNA analysis to reduce uncertainty in clinical interpretation. *Am J Hum Genet*. 2021;108(4):696-708. doi:10.1016/j.ajhg.2021.03.006

64.   Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. *Hum Genet*. 2017;136(9):1093-1111. doi:10.1007/s00439-017-1809-4

65.   Li Q, Wang Y, Pan Y, Wang J, Yu W, Wang X. Unraveling synonymous and deep intronic variants causing aberrant splicing in two genetically undiagnosed

epilepsy families. *BMC Med Genomics*. 2021;14(1):152. doi:10.1186/s12920-021-01008-8

66.   Riolo G, Cantara S, Ricci C. What's Wrong in a Jump? Prediction and Validation of Splice Site Variants. *Methods Protoc*. 2021;4(3):62. doi:10.3390/mps4030062

67.   Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res*. 2001;29(5):1185-1190. doi:10.1093/nar/29.5.1185

68.   Yeo G, Burge CB. Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. *J Comput Biol*. 2004;11(2-3):377-394. doi:10.1089/1066527041410418

69.   Reese MG, Eeckman FH, Kulp D, Haussler2 D. Improved Splice Site Detection in Genie. *J Comput Biol*. 1997;4(3):311-323. doi:10.1089/cmb.1997.4.311

70.   Shibata A, Okuno T, Rahman MA, et al. IntSplice: prediction of the splicing consequences of intronic single-nucleotide variations in the human genome. *J Hum Genet*. 2016;61(7):633-640. doi:10.1038/jhg.2016.23

71.   Corvelo A, Hallegger M, Smith CWJJ, Eyras E. Genome-Wide Association between Branch Point Properties and Alternative Splicing. *PLOS Comput Biol*. 2010;6(11):e1001016. doi:10.1371/JOURNAL.PCBI.1001016

72.   Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res*. 2003;31(13):3568-3571. doi:10.1093/nar/gkg616

6.4 Concluding remarks

73.   Fairbrother WG, Yeo GW, Yeh R, et al. RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res*. 2004;32(suppl_2):W187-W190. doi:10.1093/nar/gkh393

74.   Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 2009;37(9):e67. doi:10.1093/nar/gkp215

75.   Schwartz S, Hall E, Ast G. SROOGLE: webserver for integrative, user-friendly visualization of splicing signals. *Nucleic Acids Res*. 2009;37(suppl_2):W189-W192. doi:10.1093/nar/gkp320

76.   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):405-423. doi:10.1038/gim.2015.30

77.   Abou Tayoun AN, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat*. 2018;39(11):1517-1524. doi:10.1002/humu.23626

78.   Rowlands CF, Baralle D, Ellingford JM. Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. *Cells*. 2019;8(12):1513. doi:10.3390/CELLS8121513

79.   Tubeuf H, Charbonnier C, Soukarieh O, et al. Large-scale comparative evaluation of user-friendly tools for predicting variant-induced alterations of splicing regulatory elements. *Hum Mutat*. 2020;41(10):1811-1829. doi:10.1002/HUMU.24091

80.   Riepe T V., Khan M, Roosing S, Cremers FPM, 't Hoen PAC. Benchmarking deep learning splice prediction tools using functional splice assays. *Hum Mutat*. 2021;42(7):799-810. doi:10.1002/HUMU.24212

81.   Wang H, Guan J, Guan L, et al. Further evidence for "gain-of-function" mechanism of DFNA5 related hearing loss. *Sci Rep*. 2018;8(1):8424. doi:10.1038/s41598-018-26554-7

82.   Carvill GL, Engel KL, Ramamurthy A, et al. Aberrant Inclusion of a Poison Exon Causes Dravet Syndrome and Related SCN1A-Associated Genetic Epilepsies. *Am J Hum Genet*. 2018;103(6):1022-1029. doi:https://doi.org/10.1016/j.ajhg.2018.10.023

83.   Krüger J, Schubert J, Kegele J, et al. Loss-of-function variants in the KCNQ5 gene are associated with genetic generalized epilepsies. *medRxiv*. Published online April 20, 2021:2021.04.20.21255696. doi:10.1101/2021.04.20.21255696

84.   Bruni V, Spoleti CB, La Barbera A, et al. *A Novel Splicing Variant of COL2A1 in a Fetus with Achondrogenesis Type II: Interpretation of Pathogenicity of In-Frame Deletions*. Vol 12. Multidisciplinary Digital Publishing Institute; 2021:1395. doi:10.3390/genes12091395

85.   Nix P, Mundt E, Coffee B, et al. Interpretation of BRCA2 Splicing Variants: A Case Series of Challenging Variant Interpretations and the Importance of Functional RNA Analysis. *Fam Cancer*. 2022;21(1):7-19. doi:10.1007/s10689-020-00224-y

86.   Aicher JK, Jewell P, Vaquero-Garcia J, Barash Y, Bhoj EJ. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet Med*. 2020;22(7):1181-

1190. doi:10.1038/s41436-020-0780-y

87. Morbidoni V, Baschiera E, Forzan M, et al. Hybrid Minigene Assay: An Efficient Tool to Characterize mRNA Splicing Profiles of NF1 Variants. *Cancers (Basel)*. 2021;13(5):999. doi:10.3390/CANCERS13050999

88. Valenzuela-Palomo A, Bueno-Martínez E, Sanoguera-Miralles L, et al. Mis-splicing in breast cancer: identification of pathogenic BRCA2 variants by systematic minigene assays. *J Pathol*. 2019;248(4):409-420. doi:10.1002/PATH.5268

89. Valenzuela-Palomo A, Bueno-Martínez E, Sanoguera-Miralles L, et al. Splicing predictions, minigene analyses, and ACMG-AMP clinical classification of 42 germline PALB2 splice-site variants. *J Pathol*. Published online November 30, 2021. doi:10.1002/path.5839

90. Leman R, Gaildrat P, Gac GL, et al. Novel diagnostic tool for prediction of variant spliceogenicity derived from a set of 395 combined in silico/in vitro studies: an international collaborative effort. *Nucleic Acids Res*. 2018;46(15):7913-7923. doi:10.1093/NAR/GKY372

91. Dionnet E, Defour A, Da Silva N, et al. Splicing impact of deep exonic missense variants in CAPN3 explored systematically by minigene functional assay. *Hum Mutat*. 2020;41(10):1797-1810. doi:10.1002/HUMU.24083

92. Cooper TA. Use of minigene systems to dissect alternative splicing elements. *Methods*. 2005;37(4):331-340. doi:https://doi.org/10.1016/j.ymeth.2005.07.015

93. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP

sequence variant interpretation framework. *Genome Med*. 2019;12(1):3. doi:10.1186/s13073-019-0690-2

94. Bergsma AJ, in 't Groen SLMM, van den Dorpel JJAA, et al. A genetic modifier of symptom onset in Pompe disease. *EBioMedicine*. 2019;43:553-561. doi:https://doi.org/10.1016/j.ebiom.2019.03.048

95. Lin JH, Wu H, Zou W Bin, et al. Splicing Outcomes of 5′ Splice Site GT>GC Variants That Generate Wild-Type Transcripts Differ Significantly Between Full-Length and Minigene Splicing Assays. *Front Genet*. 2021;12:1419. doi:10.3389/FGENE.2021.701652/BIBTEX

96. Miller TE, You L, Myerburg RJ, Benke PJ, Bishopric NH. Whole blood RNA offers a rapid, comprehensive approach to genetic diagnosis of cardiovascular diseases. *Genet Med*. 2007;9(1):23-33. doi:10.1097/GIM.0b013e31802d74de

97. Frésard L, Smail C, Ferraro NM, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med*. 2019;25(6):911-919. doi:10.1038/s41591-019-0457-8

98. Gonorazky HD, Naumenko S, Ramani AK, et al. Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *Am J Hum Genet*. 2019;104(3):466-483. doi:10.1016/j.ajhg.2019.01.012

99. Lee H, Huang AY, Wang L, et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med*. 2020;22(3):490-499. doi:10.1038/s41436-019-0672-1

100. Rentas S, Rathi KS, Kaur M, et al. Diagnosing Cornelia de Lange syndrome and related neurodevelopmental disorders using RNA sequencing. *Genet*

Med. 2020;22(5):927-936. doi:10.1038/s41436-019-0741-5

101. Wai HA, Lord J, Lyon M, et al. Blood RNA analysis can increase clinical diagnostic rate and resolve variants of uncertain significance. *Genet Med*. 2020;22(6):1005-1014. doi:10.1038/s41436-020

102. Murdock DR, Dai H, Burrage LC, et al. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J Clin Invest*. 2021;131(1). doi:10.1172/JCI141500

103. Maddirevula S, Kuwahara H, Ewida N, et al. Analysis of transcript-deleterious variants in Mendelian disorders: implications for RNA-based diagnostics. *Genome Biol*. 2020;21(1):145. doi:10.1186/s13059-020-02053-9

104. Kremer LS, Bader DM, Mertes C, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun*. 2017;8(1):15824. doi:10.1038/ncomms15824

105. Douglas AGL, Baralle D. Translating RNA Splicing Analysis into Diagnosis and Therapy. *OBM Genet*. 2021;5(1):1-23. doi:10.21926/OBM.GENET.2101125

106. Rowlands CF, Taylor A, Rice G, et al. MRSD: A quantitative approach for assessing suitability of RNA-seq in the investigation of mis-splicing in Mendelian disease. *Am J Hum Genet*. Published online January 21, 2022. doi:10.1016/J.AJHG.2021.12.014

107. Holliday M, Singer ES, Ross SB, et al. Transcriptome Sequencing of Patients with Hypertrophic Cardiomyopathy Reveals Novel Splice-Altering Variants in MYBPC3. *Circ Genomic Precis Med*. 2021;14(2):183-191. doi:10.1161/CIRCGEN.120.003202

108. Whiley PJ, consortium on behalf of the E, de la Hoya M, et al. Comparison of mRNA Splicing Assay Protocols across Multiple Laboratories: Recommendations for Best Practice in Standardized Clinical Testing. *Clin Chem*. 2014;60(2):341-352. doi:10.1373/CLINCHEM.2013.210658

109. Cummings BB, Marshall JLJL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*. 2017;9(386):eaal5209. doi:10.1101/074153

110. Frésard L, Smail C, Ferraro NM, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med*. 2019;25(6):911-919. doi:10.1038/s41591-019-0457-8

111. Mehmood A, Laiho A, Venäläinen MS, et al. Systematic evaluation of differential splicing tools for RNA-seq studies. *Brief Bioinform*. 2020;21(6):2052-2065. doi:10.1093/BIB/BBZ126

112. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet*. 2019;20(11):631-656. doi:10.1038/s41576-019-0150-2

113. Van den Berge K, Hembach KM, Soneson C, et al. RNA Sequencing Data: Hitchhiker's Guide to Expression Analysis. *Annu Rev Biomed Data Sci*. 2019;2(1):139-173. doi:10.1146/annurev-biodatasci-072018-021255

114. Lord J, Baralle D. Splicing in the Diagnosis of Rare Disease: Advances and Challenges. *Front Genet*. 2021;0:1146. doi:10.3389/FGENE.2021.689892

115. Jenkinson G, Li YI, Basu S, Cousin MA, Oliver GR, Klee EW. LeafCutterMD: an algorithm for outlier splicing detection in rare diseases. *Bioinformatics*. 2020;36(17):4609-4615. doi:10.1093/BIOINFORMATICS/BTAA259

116. Aguet F, Barbeira AN, Bonazzola R, et al. Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science (80-)*. 2020;369(6509). doi:10.1126/SCIENCE.AAZ5900/SUPPL_FILE/AAZ5900-FERRARO-SM.PDF

117. Schlieben LD, Prokisch H, Yépez VA. How Machine Learning and Statistical Models Advance Molecular Diagnostics of Rare Disorders Via Analysis of RNA Sequencing Data. *Front Mol Biosci*. 2021;8:473. doi:10.3389/FMOLB.2021.647277/BIBTEX

118. Brechtmann F, Mertes C, Matusevičiūtė A, et al. OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data. *Am J Hum Genet*. 2018;103(6):907-917. doi:10.1016/J.AJHG.2018.10.025

119. Mertes C, Scheller IF, Yépez VA, et al. Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun*. 2021;12(1):1-13. doi:10.1038/s41467-020-20573-7

120. Harrington CA, Fei SS, Minnier J, et al. RNA-Seq of human whole blood: Evaluation of globin RNA depletion on Ribo-Zero library method. *Sci Rep*. 2020;10(1):1-12. doi:10.1038/s41598-020-62801-6

121. Conesa A, Madrigal P, Tarazona S, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17(1):1-19. doi:10.1186/S13059-016-0881-8

122. Su Z, Łabaj PP, Li SS, et al. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol*. 2014;32(9):903-914. doi:10.1038/nbt.2957

123. Li S, Tighe SW, Nicolet CM, et al. Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat Biotechnol*. 2014;32(9):915-925. doi:10.1038/nbt.2972

124. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. *Cold Spring Harb Protoc*. 2015;2015(11):pdb.top084970. doi:10.1101/PDB.TOP084970

125. Marco-Puche G, Lois S, Benítez J, Trivino JC. RNA-Seq Perspectives to Improve Clinical Diagnosis. *Front Genet*. 2019;10(November):1-7. doi:10.3389/fgene.2019.01152

126. Pollard MO, Gurdasani D, Mentzer AJ, Porter T, Sandhu MS. Long reads: their purpose and place. *Hum Mol Genet*. 2018;27(R2):R234-R241. doi:10.1093/hmg/ddy177

127. Mantere T, Kersten S, Hoischen A. Long-Read Sequencing Emerging in Medical Genetics. *Front Genet*. 2019;10:426. doi:10.3389/fgene.2019.00426

128. Wai H, Douglas AGLL, Baralle D. RNA splicing analysis in genomic medicine. *Int J Biochem Cell Biol*. 2018;108:61-71. doi:10.1016/j.biocel.2018.12.009

129. Kazazian J, Boehm CD, Seltzer WK. ACMG recommendations for standards for interpretation of sequence variations. *Genet Med*. 2000;2(5):302-303. doi:10.1097/00125817-200009000-00009

130. Bradley LA, Palomaki GE, McDowell GA. Technical standards and guidelines: Prenatal screening for open neural tube defects. *Genet Med*. 2005;7(5):355-369. doi:10.1097/00125817-200505000-00010

131. Palomaki GE, Lee JES, Canick JA, McDowell GA, Donnenfeld AE. Technical standards and guidelines: Prenatal screening for Down syndrome that includes

first-trimester biochemistry and/or ultrasound measurements. *Genet Med*. 2009;11(9):669-681. doi:10.1097/gim.0b013e3181ad5246

132. Wolff DJ, Van Dyke DL, Powell CM. Laboratory guideline for Turner syndrome. *Genet Med*. 2010;12(1):52-55. doi:10.1097/gim.0b013e3181c684b2

133. Cowan TM, Blitzer MG, Wolf B. Technical standards and guidelines for the diagnosis of biotinidase deficiency. *Genet Med*. 2010;12(7):464-470. doi:10.1097/gim.0b013e3181e4cc0f

134. Wang RY, Bodamer OA, Watson MS, Wilcox WR. Lysosomal storage diseases: Diagnostic confirmation and management of presymptomatic individuals. *Genet Med*. 2011;13(5):457-484. doi:10.1097/gim.0b013e318211a7e1

135. Lyon E, Gastier Foster J, Palomaki GE, et al. Laboratory testing of CYP2D6 alleles in relation to tamoxifen therapy. *Genet Med*. 2012;14(12):990-1000. doi:10.1038/gim.2012.108

136. Rehder CW, David KL, Hirsch B, Toriello H V., Wilson CM, Kearney HM. American College of Medical Genetics and Genomics: standards and guidelines for documenting suspected consanguinity as an incidental finding of genomic testing. *Genet Med*. 2013;15(2):150-152. doi:10.1038/gim.2012.169

137. Cooley LD, Lebo M, Li MM, Slovak ML, Wolff DJ. American College of Medical Genetics and Genomics technical standards and guidelines: microarray analysis for chromosome abnormalities in neoplastic disorders. *Genet Med*. 2013;15(6):484-494. doi:10.1038/gim.2013.49

138. Monaghan KG, Lyon E, Spector EB. ACMG Standards and Guidelines for

fragile X testing: a revision to the disease-specific supplements to the Standards and Guidelines for Clinical Genetics Laboratories of the American College of Medical Genetics and Genomics. *Genet Med*. 2013;15(7):575-586. doi:10.1038/gim.2013.61

139. South ST, Lee C, Lamb AN, Higgins AW, Kearney HM. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med*. 2013;15(11):901-909. doi:10.1038/gim.2013.129

140. Hegde M, Ferber M, Mao R, Samowitz W, Ganguly A. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). *Genet Med*. 2013;16(1):101-116. doi:10.1038/gim.2013.166

141. Bean L, Bayrak-Toydemir P. American College of Medical Genetics and Genomics Standards and Guidelines for Clinical Genetics Laboratories, 2014 edition: technical standards and guidelines for Huntington disease. *Genet Med*. 2014;16(12):e2-e2. doi:10.1038/gim.2014.146

142. Spector EB, Grody WW, Matteson CJ, et al. Technical standards and guidelines: Venous thromboembolism (Factor V Leiden and prothrombin 20210G>A testing): A disease-specific supplement to the standards and guidelines for clinical genetics laboratories. *Genet Med*. 2005;7(6):444-453. doi:10.1097/01.gim.0000172641.57755.3a

143. Kishnani PS, Austin SL, Abdenur JE, et al. Diagnosis and management of glycogen storage disease type I: a practice guideline of the American College of Medical Genetics and Genomics. *Genet Med*. 2014;16(11):e1-e1.

6.4 Concluding remarks

doi:10.1038/gim.2014.128

144. Mikhail FM, Heerema NA, Rao KW, Burnside RD, Cherry AM, Cooley LD. Section E6.1–6.4 of the ACMG technical standards and guidelines: chromosome studies of neoplastic blood and bone marrow–acquired chromosomal abnormalities. *Genet Med*. 2016;18(6):635-642. doi:10.1038/gim.2016.50

145. Cooley LD, Morton CC, Sanger WG, Saxe DF, Mikhail FM. Section E6.5–6.8 of the ACMG technical standards and guidelines: chromosome studies of lymph node and solid tumor–acquired chromosomal abnormalities. *Genet Med*. 2016;18(6):643-648. doi:10.1038/gim.2016.51

146. Sharer JD, Bodamer O, Longo N, Tortorelli S, Wamelink MMC, Young S. Laboratory diagnosis of creatine deficiency syndromes: a technical standard and guideline of the American College of Medical Genetics and Genomics. *Genet Med*. 2017;19(2):256-263. doi:10.1038/gim.2016.203

147. Strovel ET, Cowan TM, Scott AI, Wolf B. Laboratory diagnosis of biotinidase deficiency, 2017 update: a technical standard and guideline of the American College of Medical Genetics and Genomics. *Genet Med*. 2017;19(10):1079-1079. doi:10.1038/gim.2017.84

148. Pasquali M, Yu C, Coffee B. Laboratory diagnosis of galactosemia: a technical standard and guideline of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2017;20(1):3-11. doi:10.1038/gim.2017.172

149. Hirsh B, Brothman AR, Jacky PB, Rao KW, Wolff DJ. Section E6 of the ACMG technical standards and guidelines: Chromosome studies for acquired abnormalities. *Genet Med*. 2005;7(7):509-513.

6.4 Concluding remarks

doi:10.1097/01.gim.0000177416.12323.58

150. Maddalena A, Bale S, Das S, Grody W, Richards S. Technical Standards and Guidelines: Molecular Genetic Testing for Ultra-Rare Disorders. *Genet Med*. 2005;7(8):571-583. doi:10.1097/01.gim.0000182738.95726.ca

151. Sherman S, Pletcher BA, Driscoll DA. Fragile X syndrome: Diagnostic and carrier testing. *Genet Med*. 2005;7(8):584-587. doi:10.1097/01.gim.0000182468.22666.dd

152. Shaffer LG, Beaudet AL, Brothman AR, et al. Microarray analysis for constitutional cytogenetic abnormalities. *Genet Med*. 2007;9(9):654-662. doi:10.1097/gim.0b013e31814ce3d9

153. Rinaldo P, Cowan TM, Matern D. Acylcarnitine profile analysis. *Genet Med*. 2008;10(2):151-156. doi:10.1097/gim.0b013e3181614289

154. Prior TW. Technical standards and guidelines for myotonic dystrophy type 1 testing. *Genet Med*. 2009;11(7):552-555. doi:10.1097/gim.0b013e3181abce0f

155. Cooley LD, Mascarello JT, Hirsch B, et al. Section E6.5 of the ACMG technical standards and guidelines: Chromosome studies for solid tumor abnormalities. *Genet Med*. 2009;11(12):890-897. doi:10.1097/gim.0b013e3181bb7808

156. Richards CS, Bale S, Bellissimo DB, et al. ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genet Med*. 2008;10(4):294-300. doi:10.1097/GIM.0b013e31816b5cae

157. den Dunnen J, Antonarakis S. Nomenclature for the description of human sequence variations. *Hum Genet*. 2001;109(1):121-124. doi:10.1007/S004390100505

158. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 2005;33(suppl_1):D501-D504. doi:10.1093/nar/gki025

159. Huijun Ring AZ, Kwok P-Y, Cotton RG, for correspondence A. Human Variome Project: an international collaboration to catalogue human genetic variation. *Pharmacogenomics*. 2006;7(7):969-972. doi:10.2217/14622416.7.7.969

160. Smigielski EM, Sirotkin K, Ward M, Sherry ST. dbSNP: a database of single nucleotide polymorphisms. *Nucleic Acids Res*. 2000;28(1):352-355. doi:10.1093/NAR/28.1.352

161. Cooper DN, Ball E V., Krawczak M. The human gene mutation database. *Nucleic Acids Res*. 1998;26(1):285-287. doi:10.1093/NAR/26.1.285

162. Dalgleish R, Flicek P, Cunningham F, et al. Locus Reference Genomic sequences: an improved basis for describing human DNA variants. *Genome Med*. 2010;2(4):1-7. doi:10.1186/GM145

163. Kogelnik AM, Lott MT, Brown MD, Navathe SB, Wallace DC. MITOMAP: A Human Mitochondrial Genome Database. *Nucleic Acids Res*. 1996;24(1):177-179. doi:10.1093/NAR/24.1.177

164. Firth H V, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*. 2009;84(4):524-533. doi:https://doi.org/10.1016/j.ajhg.2009.03.010

165. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships

among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(D1):D980-D985. doi:10.1093/nar/gkt1113

166. Fokkema IFAC, Dunnen JT den, Taschner PEM. LOVD: Easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. *Hum Mutat*. 2005;26(2):63-68. doi:10.1002/HUMU.20201

167. Hamosh A, Scott AF, Amberger J, Valle D, Mckusick VA. Online Mendelian Inheritance in Man (OMIM). *Hum Mutat*. 2000;15:57-61. doi:10.1002/(SICI)1098-1004(200001)15:1<57::AID-HUMU12>3.0.CO;2-G

168. Altshuler DM, Durbin RM, Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. doi:10.1038/nature11632

169. Lek M, Karczewski KJ, Minikel E V., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291. doi:10.1038/nature19057

170. Auer PL, Reiner AP, Wang G, et al. Guidelines for Large-Scale Sequence-Based Complex Trait Association Studies: Lessons Learned from the NHLBI Exome Sequencing Project. *Am J Hum Genet*. 2016;99(4):791-801. doi:10.1016/J.AJHG.2016.08.012

171. Amendola LM, Jarvik GP, Leo MC, et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among Nine Laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet*. 2016;98(6):1067-1076. doi:10.1016/J.AJHG.2016.03.024

172. Nykamp K, Anderson M, Powers M, et al. Sherloc: a comprehensive

refinement of the ACMG–AMP variant classification criteria. *Genet Med*.
2017;19(10):1105-1117. doi:10.1038/gim.2017.37

173. Niehaus A, Azzariti DR, Harrison SM, et al. A survey assessing adoption of the
ACMG-AMP guidelines for interpreting sequence variants and identification of
areas for continued improvement. *Genet Med*. 2019;21(8):1699-1701.
doi:10.1038/s41436-018-0432-7

174. Strande NT, Brnich SE, Roman TS, Berg JS. Navigating the nuances of
clinical sequence variant interpretation in Mendelian disease. *Genet Med*.
2018;20(9):918-926. doi:10.1038/s41436-018-0100-y

175. Rehm HL, Berg JS, Brooks LD, et al. ClinGen — The Clinical Genome
Resource. *N Engl J Med*. 2015;372(23):2235-2242.
doi:10.1056/NEJMsr1406261

176. Rivera-Muñoz EA, Milko L V., Harrison SM, et al. ClinGen Variant Curation
Expert Panel experiences and standardized processes for disease and gene-
level specification of the ACMG/AMP guidelines for sequence variant
interpretation. *Hum Mutat*. 2018;39(11):1614-1622. doi:10.1002/HUMU.23645

177. Clinical Genome Resource. Accessed August 12, 2021.
https://www.clinicalgenome.org/site/assets/files/5182/pm2_-
_svi_recommendation_-_approved_sept2020.pdf

178. Clinical Genome Resource. Accessed August 12, 2021.
https://www.clinicalgenome.org/site/assets/files/3717/svi_proposal_for_pm3_cr
iterion_-_version_1.pdf

179. Ghosh R, Harrison SM, Rehm HL, Plon SE, Biesecker LG. Updated

recommendation for the benign stand-alone ACMG/AMP criterion. *Hum Mutat*. 2018;39(11):1525-1530. doi:10.1002/HUMU.23642

180. Garrett A, Callaway A, Durkie M, et al. Cancer Variant Interpretation Group UK (CanVIG-UK): An exemplar national subspecialty multidisciplinary network. *J Med Genet*. 2020;57(12):829-834. doi:10.1136/jmedgenet-2019-106759

181. Lunke, S., Eggers, S., Wilson, M., Patel, C., Barnett, C. P., Pinner, J., Sandaradura, S. A., Buckley, M. F., Krzesinski, E. I., de Silva, M. G., Brett, G. R., Boggs, K., Mowat, D., Kirk, E. P., Adès, L. C., Akesson, L. S., Amor, D. J., Ayres, S., Baxenda Z. Feasibility of Ultra-Rapid Exome Sequencing in Critically Ill Infants and Children With Suspected Monogenic Conditions in the Australian Public Health Care System. *JAMA*. 2020;323(24):2503-2511. doi:10.1001/jama.2020.7671

182. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

183. Costa C, Giménez-Capitán A, Karachaliou N, Rosell R. Comprehensive molecular screening: From the RT-PCR to the RNA-seq. *Transl Lung Cancer Res*. 2013;2(2):87-91. doi:10.3978/J.ISSN.2218-6751.2013.02.05

184. Takeuchi K, Choi YL, Soda M, et al. Multiplex Reverse Transcription-PCR Screening for EML4-ALK Fusion Transcripts. *Clin Cancer Res*. 2008;14(20):6618-6624. doi:10.1158/1078-0432.CCR-08-1018

185. Bernard PS, Wittwer CT. Real-Time PCR Technology for Cancer Diagnostics. *Clin Chem*. 2002;48(8):1178-1185. doi:10.1093/CLINCHEM/48.8.1178

186. Tu JJ, Rohan S, Kao J, Kitabayashi N, Mathew S, Chen Y-T. Gene fusions

between TMPRSS2 and ETS family genes in prostate cancer: frequency and transcript variant analysis by RT-PCR and FISH on paraffin-embedded tissues. *Mod Pathol*. 2007;20(9):921-928. doi:10.1038/modpathol.3800903

187. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. *Nat Rev Genet*. 2016;17(5):257-271. doi:10.1038/nrg.2016.10

188. Wangensteen T, Felde CN, Ahmed D, Mæhle L, Ariansen SL. Diagnostic mRNA splicing assay for variants in BRCA1 and BRCA2 identified two novel pathogenic splicing aberrations. *Hered Cancer Clin Pract*. 2019;17(1):14. doi:10.1186/s13053-019-0113-9

189. Corchete LA, Rojas EA, Alonso-López D, De Las Rivas J, Gutiérrez NC, Burguillo FJ. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Sci Rep*. 2020;10(1):19737. doi:10.1038/s41598-020-76881-x

190. Ding L, Rath E, Bai Y. Comparison of Alternative Splicing Junction Detection Tools Using RNASeq Data. *Curr Genomics*. 2017;18(3):268-277. doi:10.2174/1389202918666170215125048

191. Chao HP, Chen Y, Takata Y, et al. Systematic evaluation of RNA-Seq preparation protocol performance. *BMC Genomics*. 2019;20(1):1-20. doi:10.1186/s12864-019-5953-1

192. Yépez VA, Mertes C, Müller MF, et al. Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc*. 2021;16(2):1276-1296. doi:10.1038/s41596-020-00462-5

193. Yépez VA, Gusic M, Kopajtich R, et al. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med 2022 141*. 2022;14(1):1-26. doi:10.1186/S13073-022-01019-9

194. Home - OMIM. Accessed July 5, 2022. https://www.omim.org/

195. Köhler S, Gargano M, Matentzoglu N, et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res*. 2021;49(D1):D1207-D1217. doi:10.1093/NAR/GKAA1043

196. Boycott KM, Azzariti DR, Hamosh A, Rehm HL. Seven years since the launch of the Matchmaker Exchange: The evolution of genomic matchmaking. *Hum Mutat*. 2022;43(6):659-667. doi:10.1002/HUMU.24373

197. Philippakis AA, Azzariti DR, Beltran S, et al. The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. *Hum Mutat*. 2015;36(10):915-921. doi:10.1002/HUMU.22858

198. Wohler E, Martin R, Griffith S, et al. PhenoDB, GeneMatcher and VariantMatcher, tools for analysis and sharing of sequence data. *Orphanet J Rare Dis*. 2021;16(1):1-10. doi:10.1186/S13023-021-01916-Z/FIGURES/4

199. DiStefano MT, Goehringer S, Babb L, et al. The Gene Curation Coalition: A global effort to harmonize gene–disease evidence resources. *Genet Med*. Published online May 4, 2022. doi:10.1016/J.GIM.2022.04.017

200. Shariant. Accessed May 23, 2022. https://shariant.org.au/accounts/login/

201. Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. *Genome Biol*. 2020;21(1):30. doi:10.1186/s13059-020-1935-5

202. Mitsuhashi S, Matsumoto N. Long-read sequencing for rare human genetic diseases. *J Hum Genet*. 2019;65(1):11-19. doi:10.1038/s10038-019-0671-8

203. Clark MB, Wrzesinski T, Garcia AB, et al. Long-read sequencing reveals the complex splicing profile of the psychiatric risk gene CACNA1C in human brain. *Mol Psychiatry*. 2020;25(1):37-47. doi:10.1038/s41380-019-0583-1

204. Anvar SY, Allard G, Tseng E, et al. Full-length mRNA sequencing uncovers a widespread coupling between transcription initiation and mRNA processing. *Genome Biol*. 2018;19(1):46. doi:10.1186/s13059-018-1418-0

205. De Roeck A, Van den Bossche T, van der Zee J, et al. Deleterious ABCA7 mutations and transcript rescue mechanisms in early onset Alzheimer's disease. *Acta Neuropathol*. 2017;134(3):475-487. doi:10.1007/s00401-017-1714-x

206. de Jong LC, Cree S, Lattimore V, et al. Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events. *Breast Cancer Res*. 2017;19(1):127. doi:10.1186/s13058-017-0919-1

207. Amarasinghe SL, Ritchie ME, Gouil Q. long-read-tools.org: an interactive catalogue of analysis methods for long-read sequencing data. *Gigascience*. 2021;10(2):1-7. doi:10.1093/GIGASCIENCE/GIAB003

208. Rehder C, Bean LJH, Bick D, et al. Next-generation sequencing for constitutional variants in the clinical laboratory, 2021 revision: a technical standard of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2021;23(8):1399-1415. doi:10.1038/s41436-021-01139-4

209. Li C, Vandersluis S, Holubowich C, et al. Cost-effectiveness of genome-wide

sequencing for unexplained developmental disabilities and multiple congenital anomalies. *Genet Med*. 2021;23(3):451-460. doi:10.1038/S41436-020-01012-W

210. Aaltio J, Hyttinen V, Kortelainen M, et al. Cost-effectiveness of whole-exome sequencing in progressive neurological disorders of children. *Eur J Paediatr Neurol*. 2022;36:30-36. doi:10.1016/J.EJPN.2021.11.006

211. Burns BL, Bilkey GA, Coles EP, et al. Healthcare system priorities for successful integration of genomics: An Australian focus. *Front Public Heal*. 2019;7:41. doi:10.3389/FPUBH.2019.00041/BIBTEX

212. Stark Z, Boughtwood T, Phillips P, et al. Australian Genomics: A Federated Model for Integrating Genomics into Healthcare. *Am J Hum Genet*. 2019;105(1):7-14. doi:10.1016/J.AJHG.2019.06.003

213. National First for SA Pathology and CCB. Accessed January 18, 2022. https://www.centreforcancerbiology.org.au/news-and-events/archive/National_First/

214. Item 73358 | Medicare Benefits Schedule. Accessed January 17, 2022. http://www9.health.gov.au/mbs/fullDisplay.cfm?type=item&q=73358&qt=item

Appendix

Additional publications

Jones HF, Bryen SJ, Waddell LB, et al. Importance of muscle biopsy to establish pathogenicity of DMD missense and splice variants. *Neuromuscul Disord*. 2019;29(12):913-919.

I performed the western blots and standard curve for dystrophin quantitation and prepared figure 1C.

Bryen SJ, Ewans LJ, Pinner J, et al. Recurrent TTN metatranscript-only c.39974-11T>G splice variant associated with autosomal recessive arthrogryposis multiplex congenita and myopathy. *Hum Mutat*. 2020;41(2):403-411.

I performed the RNA extraction from a patient muscle biopsy, *in silico* analysis and primer design for RT-PCR.

Waddell LB, Bryen SJ, Cummings BB, et al. WGS and RNA studies diagnose noncoding DMD variants in males with high creatine kinase. *Neurol Genet*. 2021;7(1):e554.

Additional publications

I performed the western blots and standard curve for dystrophin quantitation used in figure 4.

Katiyar D, Anderson N, Bommireddipalli S, et al. Two novel *B9D1* variants causing Joubert syndrome: Utility of mRNA and splicing studies. *Eur J Med Genet*. 2020;63(9):104000.

I performed the *in silico* and RNA analysis and prepared figure 2.

Huq AJ, Thompson BA, Bennett MF, et al. Clinical Impact of Whole Genome Sequencing in Patients with Early Onset Dementia. J Neurol Neurosurg Psychiatry. 2022 In Press.

I performed the in silico and RNA analysis for a patient with compound heterozygous variants in SPG21 and prepared figure 4C.

Importance of muscle biopsy to establish pathogenicity of *DMD* missense and splice variants

Hannah F Jones [a,b,c], Samantha J Bryen [a,b], Leigh B Waddell [a,b], Adam Bournazos [a,b],
Mark Davis [d], Michelle A Farrar [e,f], Catriona A McLean [g], David R Mowat [h], Hugo Sampaio [e],
Ian R Woodcock [i,j], Monique M Ryan [i,j], Kristi J Jones [a,b,c,1], Sandra T Cooper [a,b,1,*]

[a] *Kids Neuroscience Centre, The Children's Hospital at Westmead, Sydney, New South Wales 2145, Australia*
[b] *Discipline of Child and Adolescent Health, University of Sydney, Sydney, New South Wales, Australia*
[c] *Department of Clinical Genetics, Children's Hospital at Westmead, Sydney, New South Wales, Australia*
[d] *Department of Diagnostic Genomics, PathWest Laboratory Medicine, QEII Medical Centre, Perth, Western Australia, Australia*
[e] *Department of Neurology, Sydney Children's Hospital, Sydney, New South Wales, Australia*
[f] *Discipline of Paediatrics, School of Women's and Children's Health, UNSW Medicine, UNSW Sydney, New South Wales, Australia*
[g] *Anatomical Pathology and Victorian Neuromuscular Laboratory Service, Alfred Health and Monash University, Australia*
[h] *Centre for Clinical Genetics, Sydney Children's Hospital, Sydney, New South Wales, Australia*
[i] *Department of Neurology Royal Children's Hospital, Murdoch Childrens Research Institute and University of Melbourne, Parkville, Victoria, Australia*
[j] *Murdoch Childrens Research Institute, Melbourne, Victoria, Australia*

Abstract

A precise genetic diagnosis of a dystrophinopathy has far-reaching implications for affected boys and their families. We present three boys with *DMD* single nucleotide variants associated with Becker muscular dystrophy presenting with myalgia, reduced exercise capacity, neurodevelopmental symptoms and elevated creatine kinase. The *DMD* variants were difficult to classify: AIII:1 a synonymous variant in exon 13 c.1602G>A, p.Lys534Lys; BIII:1 an essential splice-site variant in intron 33 c.4674+1G>A, and CII:1 a missense mutation within the cysteine-rich domain, exon 66 c.9619T>C, p.Cys3207Arg. Complementary DNA (cDNA) analysis using muscle-derived mRNA established splice-altering effects of variants for AIII:1 and BIII:1, and normal splicing in CII:1. Western blot analysis demonstrated mildly to moderately reduced dystrophin levels (17.6 – 36.1% the levels of controls), supporting dystrophinopathy as a probable diagnosis. These three cases highlight the diagnostic utility of muscle biopsy for mRNA studies and western blot to investigate *DMD* variants of uncertain pathogenicity, by exploring effects on splicing and dystrophin protein levels.
© 2019 Elsevier B.V. All rights reserved.

Keywords: Becker muscular dystrophy; Duchenne muscular dystrophy; mRNA studies; Splice variants; Muscle biopsy; Missense variants.

1. Introduction

Dystrophinopathies (Duchenne and Becker muscular dystrophies and X-linked dilated cardiomyopathy) are disorders of striated muscle in which dystrophin is absent, reduced or dysfunctional. Dystrophin is encoded by the giant *DMD* gene, spanning two megabases of chromosome Xp21 [1].

The causative genetic variant in *DMD* is found in 96% of Duchenne muscular dystrophy (DMD) cases and 82% percent of Becker muscular dystrophy (BMD) cases [2,3]. Around one third of mutations in *DMD* are *de novo* in the affected male proband [4,5]. The most common genetic variants within *DMD* are large deletions (approximately 70%) or duplications (10 - 14%); often encompassing numerous exons [5,6]. The remaining *DMD* cases involve small deletions or insertions of one or more bases causing a frameshift (3 - 4%), nonsense substitutions creating a premature stop

---

codon (9 - 10%), or splice site mutations (2 - 3%) [5,6]. Pathogenic missense variants in *DMD* are comparatively rare (< 1%) [7,8], but have been identified within key functional domains, such as the N-terminal actin binding domain (ABD1, exons 1–8) where they are most commonly associated with the milder Becker muscular dystrophy phenotype [9,10], and the conserved ZZ *β*-dystroglycan binding domain (aa 3307–3354), which typically cause the more severe Duchenne muscular dystrophy phenotype [11]. Most missense variants are of uncertain significance, and assigning pathogenicity is difficult, especially considering that sarcoglycanopathies and other diseases involving the dystrophin-associated glycoprotein complex can cause secondary abnormalities in dystrophin [12].

Genetic diagnosis of a dystrophinopathy can be determined by multiplex ligation-dependent probe amplification (MLPA) or X-chromosome comparative genomic hybridization array (CGH array) to detect deletions or duplications, or, through parallel or targeted sequencing of *DMD* [13]. Massively parallel sequencing can reveal more difficult to interpret genetic variants, such as missense variants, putative splicing variants or structural rearrangements, which are becoming recognized as important rare causes of dystrophinopathy [5,14,15]. In the subset of individuals for whom a rare, segregating *DMD* variant is identified as a variant of uncertain significance, muscle biopsy remains an important diagnostic investigation to establish abnormal levels or size of dystrophin via immunohistochemistry and western blot, and provide *DMD* mRNA for analysis of abnormal pre-mRNA splicing [2,4,16].

Herein we describe three case reports of single nucleotide variants identified in *DMD* in three families with male probands presenting with myalgia and/or muscle weakness with elevated serum creatine kinase (CK). In two cases aberrant splicing was confirmed to result from the synonymous or splice-site variant, through targeted reverse transcription polymerase chain reaction (RT-PCR) of mRNA isolated from skeletal muscle. mRNA transcripts were normal in the third case, indicating that dystrophin levels were reduced by a different mechanism. These simple mRNA studies extend a growing body of evidence indicating a proportion of synonymous (or missense) variants cause splicing abnormalities and are therefore pathogenic more commonly than is currently recognized. If the splicing pattern is normal, other mechanisms for pathogenesis must also be considered, as illustrated by case 3.

2. Results

2.1. Family A

The male proband AIII:1 from Family A presented at age 10 years with a persistently elevated serum creatine kinase (CK) of 19,372 U/L (normal levels < 200 U/L) and myalgia with exercise, but no associated weakness. He had a history of mild speech delay and was diagnosed with autism spectrum disorder at the age of 14 years and

major depression at 16 years of age. At 18 years of age, he experiences muscle cramps when exercising for longer than one hour. His-power remains normal. He has mild tendoachilles and hamstring contractures. Electrocardiography and echocardiograms have been normal. Family history revealed that the maternal grandfather (AI:1, Fig. 1A) was diagnosed with Becker muscular dystrophy (with superimposed inflammatory myositis) after presenting at the age of 50 with mild limb-girdle muscle weakness and a modestly elevated CK. AI:1 was noted to have large calves and reported difficulty playing sport in childhood. Muscle biopsy from AI:1 showed patchy dystrophin staining and rimmed vacuoles.

MLPA did not identify any deletions or duplications. Sanger sequencing for AIII:1 identified a variant in *DMD* (GRCh37 chrX:32613874C>T, NM_004006.2:c.1602G>A, p.Lys534Lys); the last base of exon 13, initially reported as, 'a variant of uncertain significance'. Segregation analysis identified the same variant in AI:1 and AII:2. Alamut Visual® v2.9.0 splicing prediction programs MaxEntScan and NNSPLICE predict this synonymous variant abolishes the donor splice site; Human Splicing Finder (HSF) and SpliceSiteFinder-like (SSF) predicted weakening of the donor site (−11.5% and −14.4% respectively). Muscle histopathology showed two focal areas of myofibre destruction associated with histiocytic infiltrate and some myofibre size variation. Immunohistochemistry demonstrated mildly reduced antibody staining against DYS1 (rod domain) and DYS3 (C-terminal), normal staining for DYS2 (N-terminal), and reduced and patchy labeling for gamma-sarcoglycan (Fig. 1B). Reduced levels of dystrophin were confirmed by western blot (Fig. 1C, 31.8 ± 5.2% levels observed in controls, of normal molecular weight). RT-PCR studies of extracted mRNA showed clear evidence of *DMD* splicing abnormalities (Fig. 2A). Primers in exons 11 and 15 amplified a smaller cDNA product for AIII:1 compared to an age-matched control, with Sanger sequencing consistent with in-frame skipping of exon 13 (abnormal exon 13 skipping was also confirmed in amplicons from exons 12–14, *not shown*). Exon 13 skipping causes an in-frame deletion within the central rod domain of the encoded DMD protein; p.Val495_Lys534del. No normal-splicing of exons 12–13–14 was observed in muscle mRNA isolated from AIII:1, using a forward primer bridging exons 12 and 13 (Fig. 2A, this primer will anneal only to normally-spliced cDNA). We did not detect evidence for elevated levels of intron-13 retention in AIII:1 (using intron 13 primers, *not shown*). Upon review of the mRNA studies and western blot analysis, c.1602G>A was re-classified using ACMG criteria as a likely pathogenic variant [17].

2.2. Family B

BIII:1 was found incidentally to have a persistently elevated serum CK of 5397 U/L at 2.5 years of age during investigation for restless sleep. There was no family history of weakness or myalgia. At 13 years of age, he can swim
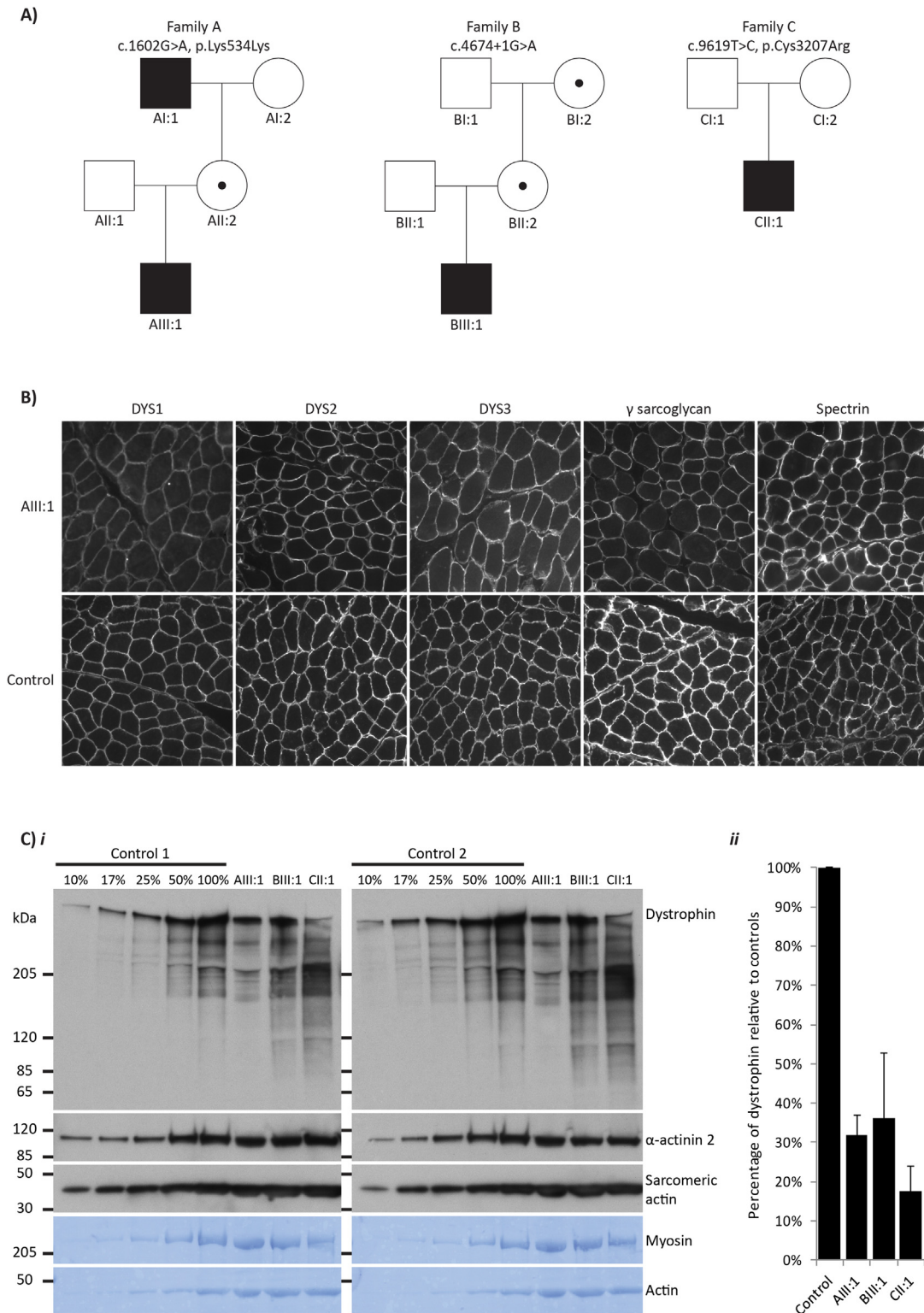
Fig. 1. (**A**) Pedigrees for family A, B and C, with the *DMD* variant numbered as per NM_004006.2. Carriers for the *DMD* variant are denoted with a black dot. Segregation data for family A was unavailable, thus AII:2 is presumed a carrier. (**B**) Immunohistochemical staining with antibodies directed against dystrophin (DYS1, DYS2 and DYS3), γ sarcoglycan and spectrin. Staining with DYS1 and DYS3 was abnormal. Staining with DYS2 was normal. γ sarcoglycan showed secondary patchy staining. (**C**) (**i**) Western blot confirmed a mild-moderate reduction in dystrophin levels in skeletal muscle from the probands (AIII:1, BIII:1, CII:1), consistent with Becker muscular dystrophy. 10 μg total protein was loaded for each proband, alongside a standard curve of 1–10 μg total protein from two skeletal muscle controls (Control 1 - male, 16 years; Control 2 – male, 14 years). Levels of α-actinin-2, sarcomeric actin, and Coomassie staining for myosin and actin, demonstrate protein loading. (**ii**) The levels of dystrophin in patients relative to control standard curves in four replicate gels, with error bars showing the standard deviation between gels. The relative densities of the dystrophin and myosin bands were determined using ImageJ for each gel. The standard curve was used to quantify the levels of dystrophin relative to controls, which were normalised to the myosin loading control.
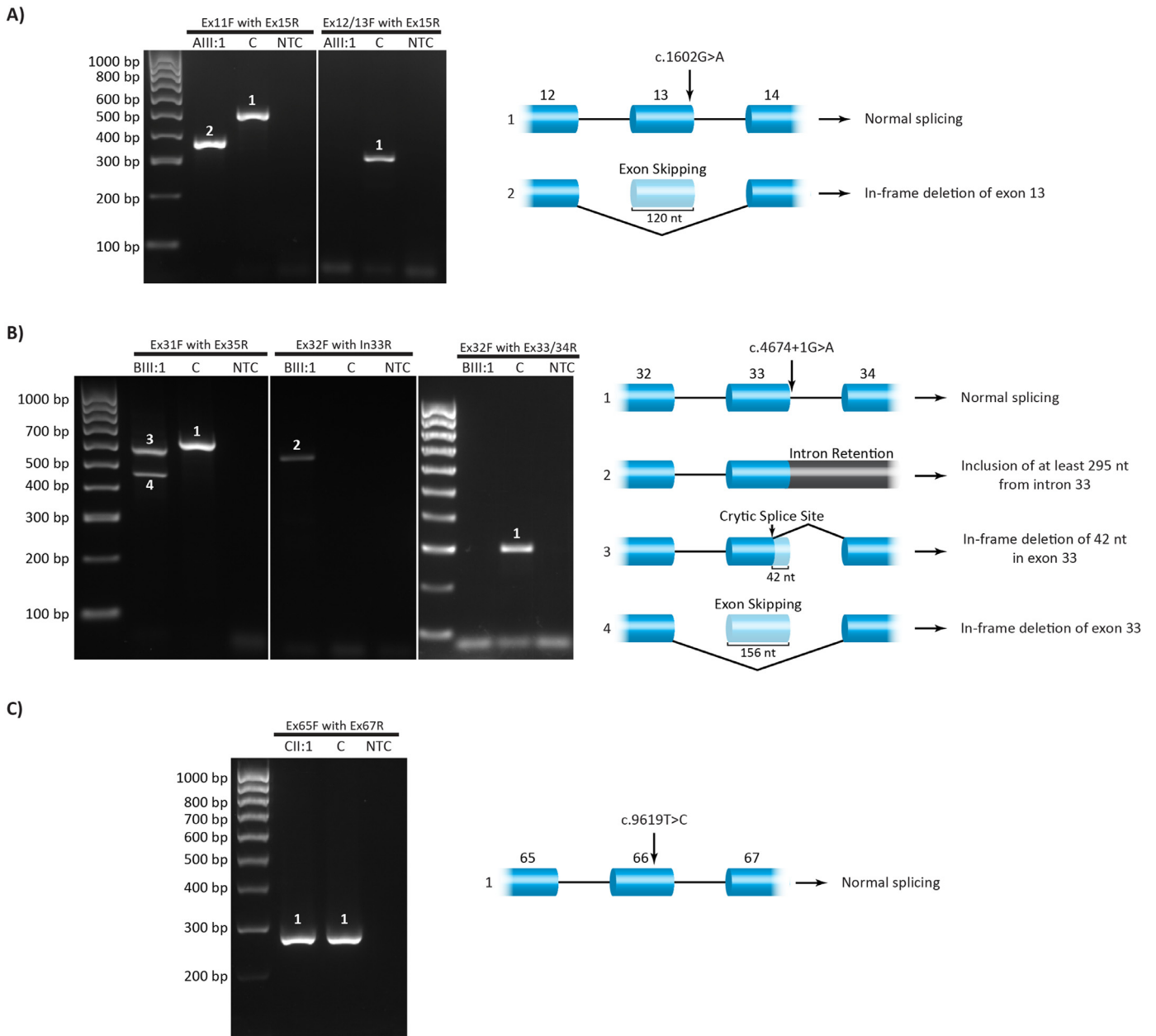
Fig. 2. RT-PCR of mRNA extracted from muscle for probands AIII:1, BIII:1 and CII:1, with *DMD* mRNA consequences illustrated. **A)** AIII:1 - Using primers in exons 11 and 15 of *DMD*, a 482 bp band corresponding to correctly spliced *DMD* mRNA can be seen in the control. In contrast, for AIII:1 the c.1602G>A variant induces exon 13 skipping, resulting in a 120 bp decrease in product size (362 bp). No normal product was detected in AIII:1 using a forward primer bridging exon 12 and exon 13 (to specifically amplify normally-spliced *DMD* mRNA) with a reverse primer in exon 15. **B)** BIII:1 – primers in exons 31 and 35 of *DMD* amplified two products for BIII:1; shorter than the expected 597 bp correctly-spliced *DMD* product seen in the control. Sanger sequencing of the shorter amplified products revealed use of a cryptic splice site in exon 33 resulting in an in-frame deletion of 42 bp (555 bp product), and exon 33 in-frame skipping (441 bp). RT-PCR using primers in exon 32 and intron 33 (to amplify products with intron 33 retention) amplified a 569 bp band for BIII:1, absent from the control, and confirmed by Sanger sequencing to correspond to intron 33 retention which encodes a stop codon. A reverse primer bridging exon 33 and exon 34 (to specifically amplify normally-spliced *DMD* mRNA) amplified a 291 bp product in the control, but not in BIII:I, suggesting negligible levels of normal splicing in BIII:1. (**C**) CII:1 - primers in exons 65 and 67 of *DMD* amplified the same sized products (280 bp) for CII:1 as the control. Sanger sequencing of the amplified products revealed normal splicing of *DMD* pre-mRNA.

for up to 45 min before complaining of myalgia. BIII:1 has some inattentiveness but does not meet diagnostic criteria for attention deficit hyperactivity disorder. Cognition and cardiac investigations are normal. On examination he has

4+/5 power in his proximal upper and lower limbs with a negative Gowers sign.

MLPA did not identify any deletions or duplications. *DMD* Sanger sequencing identified a splice site mutation in *DMD* (GRCh37 chrX:32404426C>T,

NM_004006.2:c.4674+1G>A, intron 33). This variant was initially reported as, 'a splice site mutation, predicted to lead to altered mRNA splicing of dystrophin'. Alamut Visual® v2.9.0 predicted ablation of the 5′ donor splice site of intron 33 using SSF, MaxEntScan, NNSPLICE and HSF. Segregation analysis revealed that his mother (BII:2) and maternal grandmother (BI:2) are both carriers of the same variant. BIII:1 muscle immunocytochemistry reported antibody staining against DYS1, DYS2 and DYS3 as normal. Histopathology was reported to show some variation in fibre size and possible mild increase in internal nuclei (not shown). Western blot demonstrated reduced levels of dystrophin protein to 36.1±16.5% the levels of controls, of normal molecular weight (Fig. 1C). RT-PCR studies showed normal splicing of exons 32–33–34 was not observed in muscle mRNA (Fig. 2B). Three abnormal splicing events were detected: (1) Use of a cryptic splice donor in exon 33, inducing loss of 42 nucleotides from the *DMD* mRNA, and deletion of 13 amino acids from the encoded rod-domain of dystrophin protein; (2) In-frame exon 33 skipping, and deletion of 52 amino acids from the encoded rod-domain of dystrophin; (3) Elevated levels of intron 33 retention within spliced *DMD* mRNA transcripts, resulting in a frameshift p.Thr1560Cysfs*4. Upon review of the mRNA studies and western blot analysis, c.4674+1G>A was re-classified using ACMG (American College of Medical Genetics and Genomics) criteria as a pathogenic variant [17].

2.3. Family C

CII:1 was found to have elevated serum creatine kinase >700 U/L at eight years of age during investigations for learning difficulties and attention deficit hyperactivity disorder. His developmental milestones were mildly delayed. He walked at 18 months of age, with only single words at age two years. At 25 years, he has moderate intellectual disability and some obsessive-compulsive traits. He walks long distances without myalgia. Mild cardiomyopathy was detected in his early 20′s. On examination there was no calf hypertrophy and power was normal. MLPA testing did not detect any deletions or duplications. *DMD* Sanger sequencing identified a *de novo* missense variant in exon 66 (GRCh37 chrX:31224729A>G, NM_004006.2:c.9619T>C, p.Cys3207Arg). This mutation was not previously reported but was interpreted as, 'very likely to be pathogenic' based on its position within the cysteine-rich domain, in which other mutated cysteine residues have been associated with DMD. *In silico* splicing prediction software did not predict aberrant splicing. Histopathology showed mild Type II atrophy (not shown). Immunoperoxidase studies for Dystrophin 1, 2 and 3 were normal (not shown). Western blot (DYS1) demonstrated reduced dystrophin levels to 17.6±6.4% the levels of the controls (Fig. 1C). RT-PCR studies of mRNA extracted from skeletal muscle showed normal splicing of *DMD*, using primers located in exons 65 and 67 (Fig. 2C). This result was confirmed with a second set of primers in exons 64 and 68 (not shown). We did not detect evidence for elevated

levels of intron-65 or intron-66 retention in CII:1 (using intron 65 or 66 primers, *not shown*). Western blot and mRNA studies confirmed the variant c.9619T>C, p.Cys3207Arg was pathogenic through a mechanism other than aberrant splicing, causing reduced dystrophin levels consistent with Becker muscular dystrophy. Upon review of the mRNA studies and western blot analysis, c.9619T>C was re-classified using ACMG criteria as a pathogenic variant [17].

3. Discussion

These three cases illustrate the challenges in diagnosing boys with potential dystrophinopathies due to single nucleotide variants in *DMD* causing missense substitutions or splicing abnormalities, and the importance of muscle biopsy for accurate diagnosis [18]. Interpretation of potential splicing variants is difficult, and affected boys are at risk of remaining undiagnosed [5,6,14]. With the recent explosion in genomic medicine, geneticists commonly turn to *in silico* predictive algorithms, which effectively predict adverse consequences of essential splice site variants (affecting the almost invariant GT and AG at either end of an intron) [19], but have demonstrable weaknesses in their abilities to accurately predict consequences of extended splice site variants and variants creating cryptic splice sites in either exons or introns [20,21]. Existing algorithms such as those offered within Alamut Visual® biosoftware can offer mixed predictions, and it is difficult to derive a clinically meaningful prediction of pathogenicity from a diminution in splice site strength.

While the increased statistical likelihood that a *de novo* missense variant in a phenotypically consistent gene is sufficient in some cases to enable classification as likely pathogenic, missense variants in large muscle genes (*DMD*, *TTN*, *NEB*) are particularly challenging to interpret [22,23]. In the cases described, each at the mild end of the BMD spectrum, evidence from muscle pre-mRNA splicing studies and western blots showing reproducible abnormalities in dystrophin supported their (re)classification as likely/pathogenic variants.

The missense variant p.Cys3207Arg identified in CII:1 lies within the EF hand domain of the cysteine-rich domain, which facilitates interaction between the WW domain of dystrophin and *β*-dystroglycan [11]. The effect of the missense substitution on dystrophin function is uncertain, though reduced dystrophin levels suggest the mutation leads to protein instability. The synonymous splice variant p.Lys534Lys detected in AIII:1 was impossible to interpret without RNA studies, which confirmed exon 13 skipping with no evidence of normal splicing, as demonstrated by Hagiwara for a different substitution at the same nucleotide [24]. For BIII:1, three abnormal splicing events were detected that evoked different in-frame or out-of-frame consequences for the encoded protein.

In all three cases immunohistochemistry failed to provide compelling evidence for dystrophin abnormalities, but quantitative western blot reproducibly demonstrated a mild-moderate reduction in dystrophin levels, with four repeat

western blots (using standard curves from two controls) confirming the subtle reduction in dystrophin levels.

While RNA sequencing is emerging on the diagnostic horizon, we show that clinically meaningful results can be conferred by established RT-PCR approaches. However, it is important to acknowledge the limitations of amplification-based approaches; you detect only what your primers amplify. Primer design must probe specifically for different abnormal splicing events; exonic primers to probe for exon skipping or use of cryptic splice sites, coupled with intronic primers to probe for intron retention. Technical consideration must be applied to minimize caveats associated with nonsense-mediated decay and PCR amplification bias for shorter (exon-skipping) versus longer (intron-retention) amplicons.

A precise genetic diagnosis of dystrophinopathies has far-reaching implications for carrier testing and genetic counseling, cardiac surveillance, and informing prognosis and treatment [20]. Genetic testing has replaced muscle biopsy analysis for diagnosis of many dystrophinopathies. However, in a small but important proportion of cases, current analysis methods may not detect clinically significant splice variants, complex rearrangements, or reliably infer likely pathogenicity of missense variants and a muscle biopsy is needed. We advocate that a high index of suspicion is maintained for any *DMD* variant identified in boys presenting with myalgia and elevated CK, with or without associated weakness, particularly in the presence of neurocognitive disorders. This group is often difficult to diagnose, but establishing a diagnosis has important clinical implications [25].

Beyond the immediate clinical management, an accurate genetic diagnosis will be increasingly valuable in the era of targeted genetic therapies. Boys with *DMD* splice-altering variants could benefit from personalized-medicine in the form of morpholino-based therapies, to mask an out-of-frame cryptic splice site and/or promote an in-frame abnormal splicing event to reinstate or elevate levels of dystrophin [26,27].

In conclusion, we demonstrate the clinical utility of *DMD* mRNA studies and dystrophin western blot analysis to enable a confirmed genetic diagnosis of a pathogenic *DMD* splice or missense variant in three cases with male probands presenting with myalgia, reduced exercise capacity, neurodevelopmental symptoms and an elevated CK.

Acknowledgements

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.nmd.2019.09.013.

References

[1] Ferlini A, Neri M, Gualandi F. The medical genetics of dystrophinopathies: molecular genetic diagnosis and its impact on clinical practice. Neuromuscul Disord. 2013;23(1):4–14.

[2] Taylor PJ, Maroulis S, Mullan GL, Pedersen RL, Baumli A, Elakis G, et al. Measurement of the clinical utility of a combined mutation detection protocol in carriers of Duchenne and Becker muscular dystrophy. J Med Genet. 2007;44(6):368–72.

[3] Santos R, Goncalves A, Oliveira J, Vieira E, Vieira JP, Evangelista T, et al. New variants, challenges and pitfalls in DMD genotyping: implications in diagnosis, prognosis and therapy. J Hum Genet 2014;59(8):454–64.

[4] Aartsma-Rus A, Ginjaar IB, Bushby K. The importance of genetic diagnosis for Duchenne muscular dystrophy. J Med Genet. 2016;53(3):145–51.

[5] Juan-Mateu J, Gonzalez-Quereda L, Rodriguez MJ, Baena M, Verdura E, Nascimento A, et al. DMD mutations in 576 dystrophinopathy families: a step forward in genotype-phenotype correlations. PLoS One. 2015;10(8):e0135189.

[6] Bladen CL, Salgado D, Monges S, Foncuberta ME, Kekou K, Kosma K, et al. The treat-NMD DMD global database: analysis of more than 7000 Duchenne muscular dystrophy mutations. Hum Mutat 2015;36(4):395–402.

[7] Flanigan KM, Dunn DM, von Niederhausern A, Soltanzadeh P, Gappmaier E, Howard MT, et al. Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. Hum Mutat. 2009;30(12):1657–66.

[8] Tuffery-Giraud S, Beroud C, Leturcq F, Yaou RB, Hamroun D, Michel-Calemard L, et al. Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. Hum Mutat 2009;30(6):934–45.

[9] Henderson DM, Lee A, Ervasti JM. Disease-causing missense mutations in actin binding domain 1 of dystrophin induce thermodynamic instability and protein aggregation. Proc Natl Acad Sci USA 2010;107(21):9632–7.

[10] Singh SM, Kongari N, Cabello-Villegas J, Mallela KM. Missense mutations in dystrophin that trigger muscular dystrophy decrease protein stability and lead to cross-beta aggregates. Proc Natl Acad Sci USA 2010;107(34):15069–74.

[11] Vulin A, Wein N, Strandjord DM, Johnson EK, Findlay AR, Maiti B, et al. The ZZ domain of dystrophin in DMD: making sense of missense mutations. Hum Mutat. 2014;35(2):257–64.

[12] Barresi R. From proteins to genes: immunoanalysis in the diagnosis of muscular dystrophies. Skelet Muscle 2011;1(1):24.

[13] Birnkrant DJ, Bushby K, Bann CM, Apkon SD, Blackwell A, Brumbaugh D, et al. Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and neuromuscular, rehabilitation, endocrine, and gastrointestinal and nutritional management. Lancet Neurol 2018;17(3):251–67.

[14] Takeshima Y, Yagi M, Okizuka Y, Awano H, Zhang Z, Yamauchi Y, et al. Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. J Hum Genet. 2010;55(6):379–88.

[15] Gurvich OL, Tuohy TM, Howard MT, Finkel RS, Medne L, Anderson CB, et al. DMD pseudoexon mutations: splicing efficiency, phenotype, and potential therapy. Ann Neurol 2008;63(1):81–9.

[16] Deburgrave N, Daoud F, Llense S, Barbot JC, Recan D, Peccate C, et al. Protein- and mRNA-based phenotype-genotype correlations in dmd/bmd with point mutations and molecular basis for bmd with nonsense and frameshift mutations in the dmd gene. Hum Mutat. 2007;28(2):183–95.

[17] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. Genet Med 2015;17(5):405–24.

[18] Tuffery-Giraud S, Saquet C, Chambert S, Echenne B, Marie Cuisset J, Rivier F, et al. The role of muscle biopsy in analysis of the dystrophin gene in Duchenne muscular dystrophy: experience of a national referral centre. Neuromuscul Disord 2004;14(10):650–8.

[19] Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res 2014;42(22):13534–44.

[20] Soemedi R, Cygan KJ, Rhine CL, Wang J, Bulacan C, Yang J, et al. Pathogenic variants that alter protein code often disrupt splicing. Nat Genet 2017;49(6):848–55.

[21] Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frebourg T, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in Silico tools. PLoS Genet 2016;12(1):e1005756.

[22] Savarese M, Sarparanta J, Vihola A, Udd B, Hackman P. Increasing role of Titin mutations in neuromuscular disorders. J Neuromuscul Dis 2016;3(3):293–308.

[23] Lehtokari VL, Kiiski K, Sandaradura SA, Laporte J, Repo P, Frey JA, et al. Mutation update: the spectra of nebulin variants and associated myopathies. Hum Mutat 2014;35(12):1418–26.

[24] Hagiwara Y, Nishio H, Kitoh Y, Takeshima Y, Narita N, Wada H, et al. A novel point mutation (G-1 to T) in a 5′ splice donor site of intron 13 of the dystrophin gene results in exon skipping and is responsible for Becker muscular dystrophy. Am J Hum Genet 1994;54(1):53–61.

[25] Bushby K, Finkel R, Birnkrant DJ, Case LE, Clemens PR, Cripe L, et al. Diagnosis and management of Duchenne muscular dystrophy, part 2: implementation of multidisciplinary care. Lancet Neurol 2010;9(2):177–89.

[26] Mendell JR, Goemans N, Lowes LP, Alfano LN, Berry K, Shao J, et al. Longitudinal effect of eteplirsen versus historical control on ambulation in Duchenne muscular dystrophy. Ann Neurol 2016;79(2):257–71.

[27] Moulton HM, Moulton JD. Morpholinos and their peptide conjugates: therapeutic promise and challenge for Duchenne muscular dystrophy. Biochim Biophys Acta 2010;1798(12):2296–303.

BRIEF REPORT

Recurrent *TTN* metatranscript-only c.39974–11T>G splice variant associated with autosomal recessive arthrogryposis multiplex congenita and myopathy

Samantha J. Bryen[1,2] | Lisa J. Ewans[3,4] | Jason Pinner[3,26] | Suzanna C. MacLennan[5,6] | Sandra Donkervoort[7] | Diana Castro[8] | Ana Töpf[9] | Gina O'Grady[1,2] | Beryl Cummings[10,11,12] | Katherine R. Chao[10,11,12] | Ben Weisburd[10,11,12] | Laurent Francioli[10,11,12] | Fathimath Faiz[13] | Adam M. Bournazos[1,2] | Ying Hu[7] | Carla Grosmann[23] | Denise M. Malicki[14] | Helen Doyle[15] | Nanna Witting[16] | John Vissing[16] | Kristl G. Claeys[17,18] | Kathryn Urankar[28] | Ana Beleza-Meireles[27] | Julia Baptista[20,21] | Sian Ellard[20,21] | Marco Savarese[24] | Mridul Johari[24] | Anna Vihola[24] | Bjarne Udd[24,25] | Anirban Majumdar[19] | Volker Straub[9] | Carsten G. Bönnemann[7] | Daniel G. MacArthur[10,11,12] | Mark R. Davis[13] | Sandra T. Cooper[1,2,22]

[1]Kids Neuroscience Centre, Kids Research, Children's Hospital at Westmead, Westmead, New South Wales, Australia

[2]Discipline of Child and Adolescent Health, The University of Sydney Children's Hospital Westmead Clinical School, Westmead, New South Wales, Australia

[3]Department of Medical Genomics, Royal Prince Alfred Hospital, Camperdown, New South Wales, Australia

[4]Central Clinical School, University of Sydney, Sydney, New South Wales, Australia

[5]Neurology Department, Women's and Children's Hospital, North Adelaide, South Australia, Australia

[6]School of Paediatrics and Reproductive Health, University of Adelaide, Adelaide, South Australia, Australia

[7]Neurogenetics Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland

[8]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, Texas

[9]John Walton Muscular Dystrophy Research Centre, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom

[10]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts

[11]Center for Mendelian Genomics, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts

[12]Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts

[13]Department of Diagnostic Genomics, PathWest Laboratory Medicine, Nedlands, WA, Australia

[14]Department of Pathology, Rady Children's Hospital University of California San Diego, San Diego, California

[15]Department of Histopathology, The Children's Hospital at Westmead, Sydney Children's Hospital Network, Westmead, NSW, Australia

[16]Copenhagen Neuromuscular Center, Department of Neurology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

[17]Department of Neurology, University Hospitals Leuven, Leuven, Belgium

[18]Laboratory for Muscle Diseases and Neuropathies, Department of Neurosciences, Experimental Neurology, KU Leuven-University of Leuven, Leuven, Belgium

[19]Paediatric Neurology, Bristol Royal Hospital For Children, University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom

[20]Molecular Genetics Department, Royal Devon and Exeter NHS Foundation Trust, Exeter, United Kingdom

[21]Institute of Biomedical and Clinical Science, University of Exeter Medical School University of Exeter, Exeter, United Kingdom

[22]Functional Neuromics, The Children's Medical Research Institute, Westmead, New South Wales, Australia

[23]Department of Neurology, Rady Children's Hospital University of California San Diego, San Diego, California

[24]Folkhälsan Research Center, Medicum, University of Helsinki, Haartmaninkatu 8, Helsinki, 00290, Finland

[25]Tampere Neuromuscular Center, Tampere University Hospital, Teiskontie 35, Tampere, 33520, Finland

[26]Centre for Clinical Genetics, Sydney Children's Hospital, Randwick, NSW, 2031, Australia

[27]Clinical Genetics, Bristol Royal Hospital For Children, University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom

[28]Department of Neuropathology, Southmead Hospital, University Hospitals Bristol NHS Foundation Trust, Bristol, United Kingdom

Correspondence
Sandra T. Cooper, Kids Neuroscience Centre, Kids Research, The Children's Hospital at Westmead and the Children's Medical Research Institute; Discipline of Child and Adolescent Health, Sydney Medical School, The University of Sydney, Locked Bag 4001, Westmead, NSW 2145, Australia.
Email: sandra.cooper@sydney.edu.au

Abstract

We present eight families with arthrogryposis multiplex congenita and myopathy bearing a *TTN* intron 213 extended splice-site variant (NM_001267550.1:c.39974-11T>G), inherited in *trans* with a second pathogenic *TTN* variant. Muscle-derived RNA studies of three individuals confirmed mis-splicing induced by the c.39974-11T>G variant; in-frame exon 214 skipping or use of a cryptic 3′ splice-site effecting a frameshift. Confounding interpretation of pathogenicity is the absence of exons 213-217 within the described skeletal muscle *TTN* N2A isoform. However, RNA-sequencing from 365 adult human gastrocnemius samples revealed that 56% specimens predominantly include exons 213-217 in *TTN* transcripts (inclusion rate ≥66%). Further, RNA-sequencing of five fetal muscle samples confirmed that 4/5 specimens predominantly include exons 213-217 (fifth sample inclusion rate 57%). Contractures improved significantly with age for four individuals, which may be linked to decreased expression of pathogenic fetal transcripts. Our study extends emerging evidence supporting a vital developmental role for *TTN* isoforms containing metatranscript-only exons.

KEYWORDS
alternative splicing, arthrogryposis, congenital titinopathies, intronic splice variant, *TTN* metatranscript-only

1 | INTRODUCTION

Titin is the largest known human protein measuring approximately 1.2 μm in length, and is the third most abundant protein in striated muscle (Chauveau, Rowell, & Ferreiro, 2014). Spanning half the length of the sarcomere, titin is a vital structural scaffold for sarcomere formation during development, and underpins the intrinsic elasticity of striated muscles to enable rapid and repeated lengthening and shortening of the sarcomere during muscle contractions (Chauveau et al., 2014). Titin is encoded by *TTN*, arguably one of the most complex human genes; with 364 exons encoding extensively alternatively spliced transcripts that are approximately 100,000 nucleotides in length (Bang et al., 2001; Freiburg et al., 2000; Labeit & Kolmerer, 1995). Further contributing to the complexity, *TTN* bears a triplicated repeat region that encompasses three, near-identical replicated blocks of nine exons, which are alternatively spliced (Savarese et al., 2018), and technically very challenging to sequence. When a variant is identified within the triplicated repeat region, in many cases it is impossible to be certain in which exon it resides.

Pathogenic variants in *TTN* are associated with a heterogeneous group of cardiac and muscle disorders with varying ages of onset (Savarese, Sarparanta, Vihola, Udd, & Hackman, 2016). Recently, recessive *TTN* variants have been linked to a phenotypic spectrum of severe early-onset disorders collectively termed "congenital titinopathies" (Oates et al., 2018), including centronuclear myopathy, core myopathy with heart disease, early onset myopathy with fatal cardiomyopathy, and arthrogryposis multiplex congenita (Chervinsky et al., 2018; Fernández-Marmiesse et al., 2017; Oates et al., 2018).

Owing to extensive *TTN* alternative splicing, variants are often described in reference to the inferred complete *TTN* metatranscript (NM_001267550.1); a theoretical isoform that includes all putative *TTN* exons. Although, in the context of a myopathy, typically only variants in exons described in the skeletal muscle isoform N2A (NM_133378.4) are considered. However, six pathogenic variants in *TTN* exons <u>not</u> included within the described skeletal muscle isoform N2A (NM_133378.4) are recently reported affecting 10 different families with recessive congenital titinopathies (Chervinsky et al., 2018; Fernández-Marmiesse et al., 2017; Oates et al., 2018). These pathogenic variants in *TTN* metatranscript-only exons support emerging evidence that there are numerous developmental *TTN* transcripts isoforms yet to be formally described (Savarese et al., 2018).

Herein, we describe a recurrent pathogenic *TTN* haplotype that includes a metatranscript-only intron 213 extended splice site variant (Chr2(GRCh37):g.179514069A>C; NM_001267550.1:c.39974-11T>G) identified in eight unrelated families with autosomal recessive arthrogryposis multiplex congenita and myopathy.

The ten affected individuals presented at delivery with arthrogryposis multiplex congenita and globally reduced muscle bulk (Figure 1a,b; Table S1). Contractures were varied, with most

(a)

(b)

i) BII:1 - 11 yrs   *ii*) CII:1 - neonatal   *iii*) CII:1 - 2 mo   *iv*) CII:1 - 9 mo, 13 mo

v) GII:1 - 10 mo   *vi*) HII:1 - 18 mo   *vii*) HII:2 - 2 yrs

(c)

i) AII:1 H&E - 26/40   *ii*) BII:1 H&E - 12 mo   *iii*) BII:1 H&E - 12 yrs   *iv*) DII:1 H&E - 2 yrs   *v*) HII:1 H&E - 18 mo
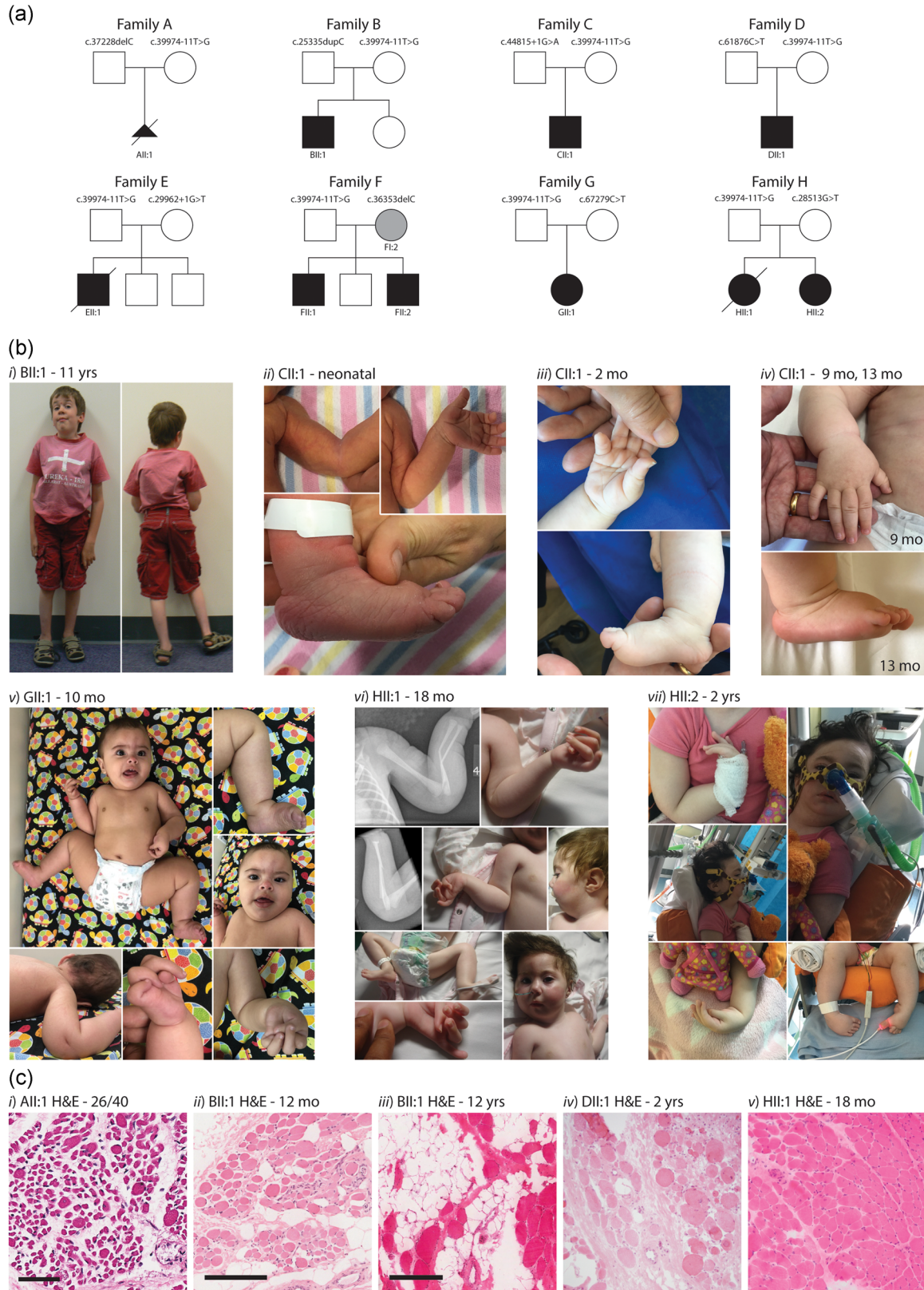
FIGURE 1 Continued.

occurring distally. Congenital fractures were observed in 3/10 cases. Dysmorphic facial features were observed in all affected individuals; with 7/10 noted to have elongated faces and 5/10 noted with micrognathia. Excluding AII:1 (terminated at 26 weeks gestation), all affected individuals presented at birth with; generalized hypotonia that persisted into early childhood and feeding difficulties in the newborn period. Marked axial weakness was noted for 5/9 cases; 7/9 affected individuals were noted to have facial weakness; 8/9 had high arched palates; and 6/9 affected individuals had neonatal respiratory difficulties. GII:1 presented in poor condition at birth, requiring cardiopulmonary resuscitation and epinephrine treatment, and placed on a head cooling protocol for hypoxic-ischemic encephalopathy (see Figure 1bv). Severe restrictive lung function in HII:1 and HII:2 persisted into infancy, with HII:1 succumbing to respiratory failure at age of 2 years.

All liveborn affected individuals had delayed motor milestones, with DII:1 never achieving independent walking. Muscle weakness was observed proximally and distally, with scapular winging in two cases. Neck weakness was pronounced, with all individuals noted to have reduced head control in infancy. Scoliosis developed in 3/9 individuals, with FII:1 presenting with scoliosis at birth. Joint hypermobility was observed in 7/9 cases. Notably, congenital contractures showed improvement with age for 4/9 affected individuals. For example, multiple contractures present in BII:1 at birth resolved throughout childhood, with a resolution of talipes, wrist, and knee contractures on examination at 18 years of age, although finger and elbow contractures persisted (see Figure 1b-i-ii). BII:1 showed delayed motor milestones though progressed to walk independently at 6 years; at 18 years BII:1 can walk short distances (20 m) and is dependent on the use of a wheelchair for longer distances.

Muscle biopsy performed for six probands (Figure 1c; Table S1) showed variation in fiber size and increased internalized nuclei; adipose replacement and fiber splitting were seen in 3/6 cases, and fiber-type disproportion was seen in 5/6 cases although varied between individuals. Serum creatine kinase was within normal limits for all affected individuals. No cardiac abnormalities have thus far been detected in any of the probands. All affected individuals were recorded to have normal intellect, although BII:1 has autism spectrum disorder.

Parallel sequencing (see Supporting Information Materials and Methods) revealed all affected individuals were heterozygous for the metatranscript-only c.39974-11T>G intron 213 extended splice site variant in TTN; present in the gnomAD population database (Lek et al., 2016) at a frequency of 0.000062 (6/96636 alleles) and not previously reported in ClinVar (Landrum et al., 2018). Each affected individual inherited a second pathogenic or likely pathogenic TTN variant in trans with the c.39974-11T>G TTN variant (see Figure 2a-i and Table S2). In each case, the second TTN variant is a truncating or splicing variant, with low allele frequency or is absent from gnomAD and likely to induce severe dysfunction or loss-of-function for the encoded titin protein. Two frameshift variants (Family A c.37228delC and Family F c.36353delC) are also located within metatranscript-only exons (see Figure 2a-i), with c.37228delC lying within the triplicated repeat region. FI:2 is a heterozygous carrier of the TTN c.36353delC variant but does not carry the c.39974-11T>G variant present in her more severely affected children (FII:1, FII:2). FI:2 presented in childhood with mild proximal muscle weakness, slight clinodactyly of toes, and very mild scoliosis. FI:2 has an elongated face with micrognathia and a high arched palate. Scrutiny of parallel sequencing data did not identify an additional likely causative TTN variant in FI:2, and thus remains undiagnosed.

Three heterozygous TTN missense variants were found to co-segregate with the c.39974-11T>G variant in all families for which full exome sequencing data was available (7/8 families); chr2:179585312G>A, c.23177C>T, p.Ser7726Leu; chr2:79486223C>T, c.45328G>A, p.Asp15110Asn; and chr2:179440163C>G, c.70696G>C, p.Gly23566Arg with gnomAD allele frequencies of 0.007 (19 homozygotes), 0.008 (23 homozygotes), and 0.012 (47 homozygotes), respectively. Further, each TTN missense variant is reported multiple times in ClinVar and LOVD as benign (see Table S2). In addition, we confirmed that the six carriers of c.39974-11T>G splice variant in gnomAD also carried the c.23177C>T, c.45328G>A and c.70696G>C variants. Thus, the collective data infer that the c.39974-11T>G splice variant lies within a common TTN haplotype encompassing three missense variants c.23177C>T, c.45328G>A, c.70696G>C (found most commonly in European [Finnish] populations in gnomAD).

Reverse transcription polymerase chain reaction (RT-PCR) performed on mRNA extracted from skeletal muscle biopsies from AII:1, BII:1, and DII:1 showed an identical pattern of abnormal splicing in the three affected individuals, compared with eight controls of different ages, using multiple primer pairs (Figure 2a-ii and -iii; shows data using two different primer pairs, see Supporting Information Materials and Methods). Sanger sequencing of amplicons

FIGURE 1 Eight families presenting with arthrogryposis multiplex congenita and myopathy. (a) Family pedigrees and segregation of the recessive TTN variants, including the common c.39974-11T>G haplotype. NOTE: FI:2 does not carry the c.39974-11T>G variant. (b) Clinical photos; i) BII:1 (11 yrs), showing an elongated face, reduced muscle bulk, and left talipes valgus; ii) CII:1 (neonatal) showing ulnar deviation and elbow contractures; iii) the ulnar deviation improving in CII:1 at 2 months of age and iv) further resolution of wrist and finger contractures at 13 months of age; v) GII:1 (10 months) presenting with hip dysplasia; finger, wrist, ankle, elbow, and knee contractures, and a mild flat nasal bridge; vi) HII:1 (18 months) showing congenital fractures, micrognathia and hip, finger, wrist, ankle, and elbow contractures; vii) HII:2 (2 years) presenting with severe restrictive lung disease, micrognathia and wrist, finger and ankle contractures. (c) Hematoxylin and eosin (H&E) staining of muscle biopsy cryosections show variation in fiber size, internalized nuclei, and areas of fatty/fibrotic replacement of muscle fibers; i) AII:1 psoas from autopsy sample, (scale bar 60 μm); ii) BII:1 quadriceps at 12 months (scale bar 200 μm); iii) BII:1 quadriceps at 12 years of age (scale bar 300 μm); iv) DII:1 vastus lateralis at 2 yrs of age (scale bar unavailable); v) HII:1 vastus lateralis at 18 months of age (scale bar unavailable)
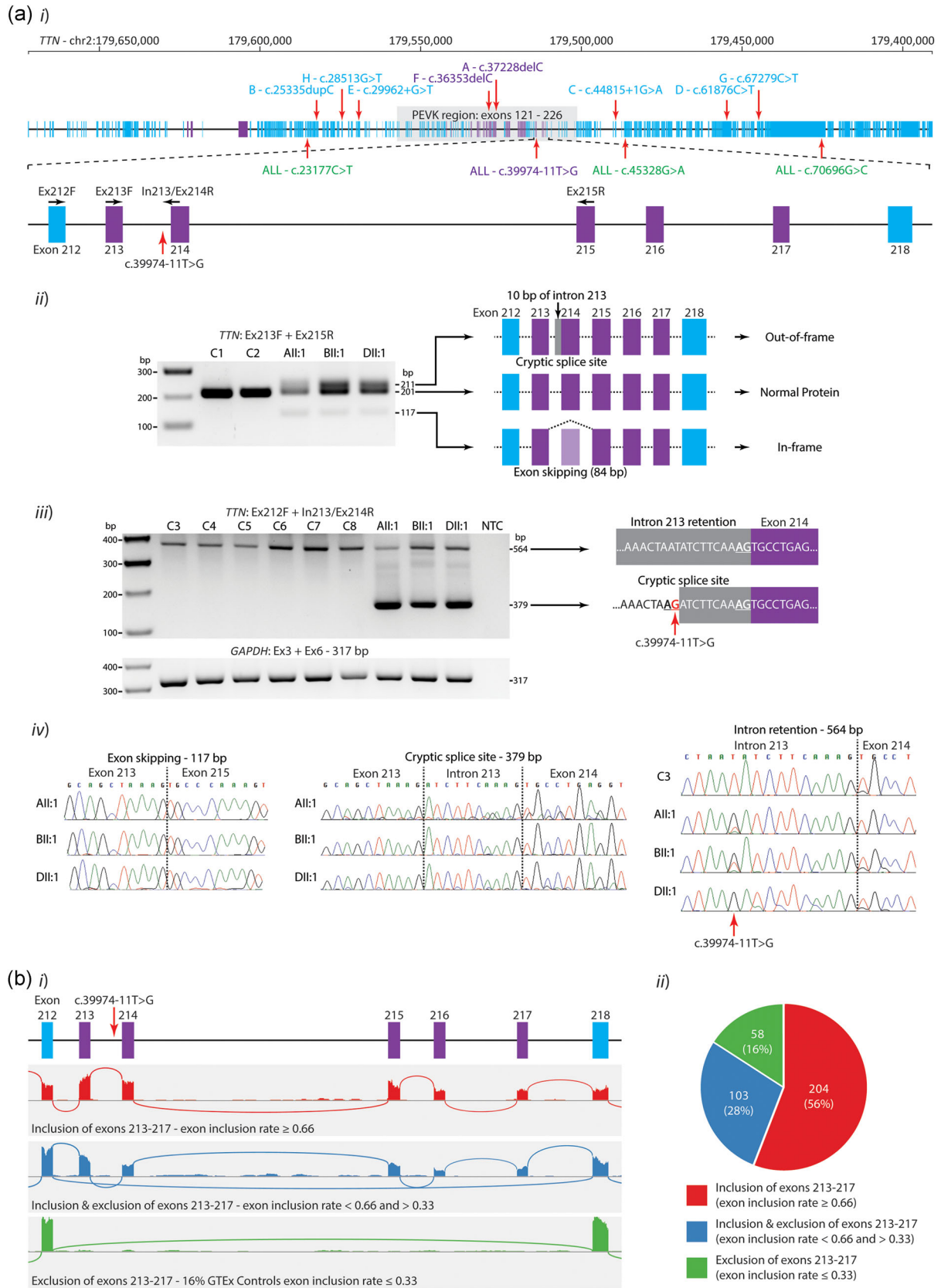
FIGURE 2 Continued.

confirmed two main abnormal splicing events: (a) Use of a cryptic 3′ splice site that abnormally includes 10 nucleotides of intron 213, leading to a frameshift and premature termination codon (r.39973_39974ins39974-10_39974-1, p.Val13325Aspfs*6) or (b) exon 214 skipping, which is in-frame and results in the loss of 28 residues (r.39974_40057del, p.Glu13327_Pro13354del); affecting one of the proline-glutamine–valine–lysine (PEVK) repeat regions (Figure 2a-iv). Multiple primer pairs variably positioned within exons 211-218 reproducibly confirmed the use of the cryptic 3′ splice site and exon 214 skipping induced by the c.39974-11T>G variant, as well as naturally occurring alternative splicing involving exons 213-217 (data not shown). We did not find evidence for increased levels of intron 214 retention resulting from the c.39974-11T>G variant (Figure 2a-iii).

At this point, collective evidence was strongly suggestive of pathogenicity of the *TTN* c.39974-11T>G variant haplotype. However, formal classification of the c.39974-11T>G as a pathogenic variant remained a challenge, given exons 213-217 are metatranscript-only exons and are not described within the skeletal muscle *TTN* N2A isoform (NM_133378.4). We, therefore, performed detailed analyses of RNA-seq from 365 control human muscle samples (gastrocnemius) within the GTEx database (see Supporting Information Materials and Methods). Our analyses confirm that 56% of gastrocnemius specimens predominantly include exons 213-217 in *TTN* transcripts, 16% specimens predominantly skip exons 213-217, and 28% show a mix of both events (Figure 2b). We further performed RNA-seq of five fetal muscle specimens and showed 4/5 fetal muscle RNA samples showed predominant inclusion of exons 213-217 (inclusion rate ≥0.66) with the remaining fetal muscle sample showing an inclusion rate of 0.57. Our data are supported by recent studies, which show that the metatranscript-only exons 213-217 are more highly expressed in fetal muscle than the adult muscle (Savarese et al., 2018).

In addition, RNA-seq data was available for individual BII:1 (paraspinal muscle biopsy was taken at 12 yrs) and HII:A (quadriceps biopsy at age 4 years). RNA-seq for BII:1 showed predominant skipping of exons 213-217 (inclusion rate of 0.1 at 12 yrs). Due to the low number of reads for exons 213-217, abnormal splicing events arising from the c.39974-11T>G variant was unable to be determined with confidence using RNA-seq data (see Figure S1). RNA-seq for HII:A confirmed the same mis-splicing events detected by RT-PCR for AII:1, BII:1 and DII:1, showing abnormal use of the upstream 3′ cryptic splice site (490 junctional reads) and exon 214 skipping (517 junctional reads bridging exons 213-215) (see Figure S1).

In conclusion, we identify eight families with arthrogryposis multiplex congenita and myopathy with a novel *TTN* c.39974-11T>G variant inherited in *trans* with a second pathogenic *TTN* variant. RT-PCR of muscle RNA confirms the c.39974-11T>G variant induces abnormal use of a cryptic 3′ splice site resulting in a frameshift, or exon 214 skipping, which removes 28 amino-acids from the encoded titin protein. While the use of the cryptic 3′ splice-site inducing a frameshift may readily be interpreted as exerting damaging consequences for the encoded titin protein; it remains difficult to interpret the functional implications attributable to the loss of 28 residues within the differentially-spliced PEVK region of titin. Exons within the PEVK region are extensively alternatively spliced, regulating passive tension and muscle elasticity (Freiburg et al., 2000; Ottenheijm et al., 2009; Savarese et al., 2018). However, exon 214 skipping is not observed in control muscle (Savarese et al., 2018), is a very rare/absent event among our muscle RNA-seq data from approximately 50 disease controls (data not shown) and absent from eight control samples by RT-PCR (results for C1 and C2 are shown in Figure 2a-ii).

◄

FIGURE 2 RNA studies of *TTN* transcripts in the muscle (a) i) schematic of *TTN* genomic locus with exons described in the N2A isoform (NM_133378.4) in blue rectangles and metatranscript-only exons in Purple. Green: Missense variants within the shared haplotype, classified as benign in ClinVar. Zoomed region: Exons 212-218 and the location of the c.39974-11T>G variant and primers used for RT-PCR. ii) RT-PCR of complementary DNA extracted from skeletal muscle from a fetal control (C1), an adult control (C2), AII:1 (fetal quadriceps), BII:1 (paraspinal muscle, 12 years), and DII:1 (vastus lateralis, 2 years) using primers in *TTN* exons 213 and 215 (Ex213F + Ex215R, product size 201 bp). Compared with controls, AII:1, BII:1, and DII:1 showed identical additional bands of 117 and 211 bp. Sanger sequencing revealed these three bands corresponded to the use of a cryptic 3′ splice site (inclusion of 10 bp, product size 211 bp), or exon 214 skipping (loss of 84 bp, product size 117 bp) or normal splicing (product size 201 bp). iii) RT-PCR of cDNA extracted from skeletal muscle from AII:1, BII:1 and DII:1 and six controls (two fetal (C3, C4), 10 months of age (C5), 8 years (C6), 18 years (C7, vastus medialis), and 26 years of age (C8, quadriceps)), using a forward primer in exon 212 and a reverse primer spanning the intron 213 and exon 214 junction (Ex212F + In213/Ex214R). A band corresponding to intron retention was observed in all samples (564 bp product) with cryptic 3′ splice site use only present in AII:1, BII:1, and DII:1 (379 bp product). Primers in exons 3 and 6 of *GAPDH* (Ex3F + Ex6) were used as a loading control. iv) Sanger sequencing chromatograms of purified gel products. Sanger sequencing of intron retention confirms the c.39974-11T>G variant is present in AII:1, BII:1, and DII:1 (muscle type unknown for C1-C6). (b) The inclusion rate of exons 213-217 within *TTN* transcripts in RNA-seq data from 365 GTEx skeletal muscle samples (mostly gastrocnemius), calculated using the equation (I/6)/[(I/6) + E] (see Supporting Information Materials and Methods). GTEx samples were divided into three groups based on their exon inclusion rate; 1) ≥0.66, considered to predominantly include exons 213-217, 2) <0.66 and >0.33, considered to have a mix of both events, and 3) ≤0.33, considered to predominantly skip exons 213-217 i) Examples of sashimi plots of RNA-seq data showing the three patterns of exon 213-217 inclusion into *TTN* transcripts. Top: An example of a skeletal muscle biospecimen showing exclusive inclusion of all exons 213-217 (red). Middle: A muscle sample showing a mix of inclusion and skipping of exons 213-217 (blue). Bottom: A muscle specimen showing exclusive skipping of exons 213-217 (green). ii) Pie Chart showing the relative proportion of GTEx muscle samples showing the different patterns of inclusion and skipping of exons 213-217, as defined above

Weighting collective evidence from eight families presenting with an overlapping clinical and histopathological phenotype consistent with congenital titinopathy (Oates et al., 2018), plausible combined deleterious effects evoked by splicing abnormalities (frameshift or deletion of 28 amino acids), and experimental evidence confirming the affected alternatively spliced exons 213-217 are expressed highly in both fetal and adult skeletal muscle specimens; the c.39974-11T>G variant has been classified as a pathogenic variant, when inherited in *trans* with a second, loss-of-function likely/pathogenic *TTN* variant. While the evidence from gnomAD and ClinVar infer the three *TTN* missense variants within the haplotype are benign due to frequent homozygosity; we cannot exclude potential additive pathogenic contributions of these *TTN* missense variants to the manifesting phenotype. However, RNA studies support abnormal splicing induced by the *TTN* c.39974-11T>G variant as the primary pathogenic element within the haplotype.

Genetic diagnosis of a recessive congenital titinopathy was further complicated for Families A and F whose second *TTN* frameshift variants involved metatranscript-only exons (exons 181 and 170, respectively). In support of pathogenicity, there are several recent reports of autosomal recessive congenital titinopathies associated with variants within metatranscript-only exons 163, 172, 181, 201 (Oates et al., 2018), 197 (Fernández-Marmiesse et al., 2017), and 167 (Chervinsky et al., 2018). Junctional reads bridging exons 181 and 170 are detected in adult skeletal muscle (Savarese et al., 2018), inferring the novel frameshift variants found in families A and F affect transcripts expressed in skeletal muscle. Of note, exons 213-217 are not expressed at significant levels in cardiac tissue (Savarese et al., 2018) and reported individuals with a metatranscript-only pathogenic *TTN* variant did not present with a cardiac phenotype (herein and in Chervinsky et al., 2018; Fernández-Marmiesse et al., 2017; Oates et al., 2018).

Careful analyses of developmentally regulated *Ttn* expression in mice and rabbits reveal that titin protein is observably larger in fetal muscle than adult muscle; with clear, age-related decrement in titin size (Ottenheijm et al., 2009). Accompanying transcriptomics infer that the increased molecular weight of titin relates primarily to alternative splicing of the complex PEVK region—and more common inclusion of these exons during development (Ottenheijm et al., 2009). Increased inclusion of the metatranscript-only exons 213-217 in fetal muscle, compared with adult muscle, has been independently confirmed in our study, and in detailed transcriptomic analyses reported in Savarese et al. (2018). Therefore, it is plausible that the improvement of severe contractures present at delivery for 4/9 individuals may be due to decreased reliance on PEVK repeats in mature muscle transcripts; an important finding for prognostic counseling. This hypothesis is supported by individual BII:1, whose contractures had mostly resolved at 18 yrs, and for whom RNA-seq shows skipping of exons 213-217 in 90% of *TTN* transcripts in a muscle biopsy taken at 12 years of age.

TTN is an extraordinarily complex gene, with the full extent of *TTN* alternative splicing only beginning to be elucidated (Guo,

Bharmal, Esbona, & Greaser, 2010; Ottenheijm et al., 2009; Savarese et al., 2018). Interpretation of the functional impact of putative pathogenic variants in *TTN* metatranscript-only exons will benefit greatly from emerging technologies in long-read RNA-seq, potentially from isolated fibers, to better define *TTN* isoforms expressed in developing and adult muscles of different fiber types.

Titin is the cornerstone for sarcomere assembly and is largely responsible for the passive tension and elasticity in the muscle (Chauveau et al., 2014; Ottenheijm et al., 2009). It is conceivable that *TTN* variants leading to abnormal muscle development may yet be associated with a range of developmental phenotypes. Muscles with unique tensile or contractile properties may uniquely depend on a subgroup of *TTN* isoforms, which may be dispensable in other muscles.

In conclusion, we identify a recurrent *TTN* c.39974-11T>G splice variant haplotype as the likely causal basis for arthrogryposis multiplex congenita and myopathy in eight families, when co-inherited with a second, loss-of-function, likely/pathogenic *TTN* variant. The *TTN* c.39974-11T>G variant may be missed by genomics platforms that do not assess or capture all *TTN* metatranscript exons. We advocate screening for this variant in any individual presenting with arthrogryposis who bears one *TTN* likely/pathogenic variant, and is shown to also carry missense variants within the common haplotype (c.23177C>T, c.45328G>A, and c.70696G>C). Our results extend emerging evidence linking recessive metatranscript-only *TTN* variants with severe, arthrogryposis multiplex congenita, and myopathy; due to a crucial role for *TTN* transcripts bearing metatranscript-only exons during development.

CONFLICTS OF INTERESTS

Professor Sandra Cooper is the director of Frontier Genomics Pty Ltd (Australia). Frontier Genomics has not traded (as of October

AUTHOR CONTRIBUTIONS

S. J. B. performed RT-PCR studies and analysis of RNA-seq data with critical analysis and interpretation of results and played a major role in the preparation of manuscript, figures, and collation of clinical data. L. J. E. made clinical review and management Family A; analysis of clinical phenotype and histotype; provision and preparation of data for Figure 1 and Table S1 and manuscript review. J. P. made clinical review and management Family A and C; analysis of clinical phenotype and histotype; provision and preparation of data for Figure 1 and Table S1. S. C. M. made clinical review and management Family B; analysis of clinical phenotype and histotype; provision and preparation of data for Figure 1 and Table S1. S. D. performed genomic analyses, consultation, and research consenting of Family D; provision and preparation of data for Figure 1 and Table S1; co-ordination of shipment of RNA. D. C. made clinical review and management Family G; analysis of clinical phenotype and histotype; provision and preparation of data for Figure 1 and Table S1. A. T. made genomic analyses and co-ordination of data for Families E and F; provision and preparation of data for Figure 1 and Table S1. G. O'G. made a clinical review, management, and genomic testing of Family B; analysis of clinical phenotype and histotype; provision and preparation of data for Figure 1 and Table S1. B. C. extracted and analysis of GTEx RNA-seq data. K. R. C. analyzed sequencing data; identified additional families for this study and initiated collaboration between researchers. B. W. and L. F. analyzed Broad sequencing data; identified common haplotype in gnomAD data. F. F. performed RNA-seq for five fetal muscle samples. A. M. B. extracted RNA from a muscle biopsy of individual AII:1 and designed primers for RT-PCR. Y. H. extracted RNA from a muscle biopsy of individual DII:1 and organized the shipping of the muscle-derived RNA from DII:1. C. G. made clinical review and management Family D. D. M. M. performed histological analysis of individual DII:1. H. D. performed an autopsy and histological analysis of individual AII:1. N. W. and J. V. made clinical review and management Family E; Analysis of clinical phenotype and histotype; Provision and preparation of data for Figure 1 and Table S1. K. G. C. made clinical review and management Family F; analysis of clinical phenotype and histotype; provision and preparation of data for Figure 1 and Table S1. K. U., A. B.-M. and A. M. made clinical review and management Family H; analysis of clinical phenotype and histotype; provision and preparation of data for Figure 1 and Table S1. J. B. and S. E. analysis, curation and interpretation of genomic variants identified by trio exome sequencing for Family H. M. S. and M. J. RNA isolation, cDNA synthesis and RNA sequencing from skeletal muscle for Family H.

A. V. performed titin protein studies for Family H. B. U. study design and oversight for functional investigations for Family H. V. S. made clinical review and management Family E and F; provision and preparation of data for Figure 1 and Table S1. C. B. made a clinical review, management, and genomic testing of Family D; analysis of clinical phenotype and histotype. D. G. M. performed massively parallel sequencing (WES, WGS, and RNA-seq). M. R. D. performed massively parallel sequencing NMD panel; identified plausible pathogenicity of the *TTN* splice variant; recruitment of additional families for the study; clinical classification of variants. S. T. C. led functional genomics study and oversight of experimentation; analyzed and interpreted clinical, pathological, and laboratory results; major role in the preparation of the manuscript; editing of figures.

ORCID

Samantha J. Bryen http://orcid.org/0000-0002-4140-8622
Adam M. Bournazos http://orcid.org/0000-0002-6464-4548
Sandra T. Cooper http://orcid.org/0000-0002-7314-5097

REFERENCES

Bang, M.-L., Centner, T., Fornoff, F., Geach, A. J., Gotthardt, M., McNabb, M., & Labeit, S. (2001). The complete gene sequence of titin, expression of an unusual ≈700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system. *Circulation Research*, 89(11), 1065–1072. https://doi.org/10.1161/hh2301.100981

Chauveau, C., Rowell, J., & Ferreiro, A. (2014). A rising titan: TTN review and mutation update. *Human Mutation*, 35(9), 1046–1059. https://doi.org/10.1002/humu.22611

Chervinsky, E., Khayat, M., Soltsman, S., Habiballa, H., Elpeleg, O., & Shalev, S. (2018). A homozygous TTN gene variant associated with lethal congenital contracture syndrome. *American Journal of Medical Genetics, Part A*, 176(4), 1001–1005. https://doi.org/10.1002/ajmg.a.38639

Fernández-Marmiesse, A., Carrascosa-Romero, M. C., Alfaro Ponce, B., Nascimento, A., Ortez, C., Romero, N., & Couce, M. L. (2017). Homozygous truncating mutation in prenatally expressed skeletal isoform of TTN gene results in arthrogryposis multiplex congenita and myopathy without cardiac involvement. *Neuromuscular Disorders*, 27(2), 188–192. https://doi.org/10.1016/J.NMD.2016.11.002

Freiburg, A., Trombitas, K., Hell, W., Cazorla, O., Fougerousse, F., Centner, T., & Labeit, S. (2000). Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. *Circulation Research*, 86(11), 1114–1121. https://doi.org/10.1161/01.RES.86.11.1114

Guo, W., Bharmal, S. J., Esbona, K., & Greaser, M. L. (2010). Titin diversity-alternative splicing gone wild. *Journal of Biomedicine and Biotechnology*, 2010, 753675. https://doi.org/10.1155/2010/753675

Labeit, S., & Kolmerer, B. (1995). Titins: Giant proteins in charge of muscle ultrastructure and elasticity. *Science*, 270(5234), 293–296. http://www.jstor.org.ezproxy1.library.usyd.edu.au/stable/2888540

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., & Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, 46(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., & Consortium, E. A. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, *536*(7616), 285–291. https://doi.org/10.1038/nature19057

Oates, E. C., Jones, K. J., Donkervoort, S., Charlton, A., Brammah, S., Smith, J. E., & Laing, N. G. (2018). Congenital Titinopathy: Comprehensive characterization and pathogenic insights. *Annals of Neurology*, *83*(6), 1105–1124. https://doi.org/10.1002/ana.25241

Ottenheijm, C. A. C., Knottnerus, A. M., Buck, D., Luo, X., Greer, K., Hoying, A., & Granzier, H. (2009). Tuning Passive mechanics through differential splicing of titin during skeletal muscle development. *Biophysical Journal*, *97*(8), 2277–2286. https://doi.org/10.1016/J.BPJ.2009.07.041

Savarese, M., Sarparanta, J., Vihola, A., Udd, B., & Hackman, P. (2016). Increasing role of titin mutations in neuromuscular disorders. *Journal of Neuromuscular Diseases*, *3*(3), 293–308. https://doi.org/10.3233/JND-160158

Savarese, M., Jonson, P. H., Huovinen, S., Paulin, L., Auvinen, P., Udd, B., & Hackman, P. (2018). The complexity of titin splicing pattern in human adult skeletal muscles. *Skeletal Muscle*, *8*(1), 11. https://doi.org/10.1186/s13395-018-0156-z

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

WGS and RNA Studies Diagnose Noncoding *DMD* Variants in Males With High Creatine Kinase

Leigh B. Waddell, PhD, Samantha J. Bryen, BSc (Hons), Beryl B. Cummings, PhD, Adam Bournazos, BSc (Hons), Frances J. Evesson, PhD, Himanshu Joshi, B Software Engineering, B Business (Finance), Jamie L. Marshall, PhD, Taru Tukiainen, PhD, Elise Valkanas, BA (Biology), Ben Weisburd, BS (Computer Sc), Simon Sadedin, PhD, Mark R. Davis, PhD, Fathimath Faiz, PhD, Rebecca Gooding, PhD, Sarah A. Sandaradura, MBChB, FRACP, PhD, Gina L. O'Grady, MBChB, FRACP, PhD, Michel C. Tchan, MBBS, FRACP, PhD, David R. Mowat, MBBS, FRACP, Emily C. Oates, MBBS, FRACP, PhD, Michelle A. Farrar, MBBS, FRACP, PhD, Hugo Sampaio, MBBCh, FRACP, MPhil, Alan Ma, MBBS, FRACP, Katherine Neas, MBChB, FRACP, Min-Xia Wang, PhD, Amanda Charlton, MBChB, FRCPA, Charles Chan, MBBS (Hons), FRCPA, PhD, Diane N. Kenwright, MBBS, FRCPA, Nicole Graf, MBBS, FRCPA, Susan Arbuckle, MBBS, FRCPA, Nigel F. Clarke, MBChB, FRACP, PhD,* Daniel G. MacArthur, PhD, Kristi J. Jones, MBBS, FRACP, PhD, Monkol Lek, PhD, and Sandra T. Cooper, PhD

Correspondence
Dr. Cooper
Sandra.cooper@sydney.edu.au

Abstract

Objective
To describe the diagnostic utility of whole-genome sequencing and RNA studies in boys with suspected dystrophinopathy, for whom multiplex ligation-dependent probe amplification and exomic parallel sequencing failed to yield a genetic diagnosis, and to use remnant normal *DMD* splicing in 3 families to define critical levels of wild-type dystrophin bridging clinical spectrums of Duchenne to myalgia.

Methods
Exome, genome, and/or muscle RNA sequencing was performed for 7 males with elevated creatine kinase. PCR of muscle-derived complementary DNA (cDNA) studied consequences for *DMD* premessenger RNA (pre-mRNA) splicing. Quantitative Western blot was used to determine levels of dystrophin, relative to control muscle.

Results
Splice-altering intronic single nucleotide variants or structural rearrangements in *DMD* were identified in all 7 families. Four individuals, with abnormal splicing causing a premature stop codon and nonsense-mediated decay, expressed remnant levels of normally spliced *DMD* mRNA. Quantitative Western blot enabled correlation of wild-type dystrophin and clinical severity, with 0%–5% dystrophin conferring a Duchenne phenotype, 10% ± 2% a Becker phenotype, and 15% ± 2% dystrophin associated with myalgia without manifesting weakness.

Glossary

bp = base pair; **CK** = creatine kinase; **DMD** = Duchenne muscular dystrophy; **gnomAD** = Genome Aggregation Database; **GTEx** = Genotype-Tissue Expression; **IGV** = Integrative Genomic Browser; **MLPA** = multiplex ligation-dependent probe amplification; **mRNA** = messenger RNA; **nt** = nucleotide; **RNA-seq** = RNA sequencing; **RT-PCR** = reverse transcription PCR; **SNV** = single nucleotide variant; **WB** = Western blot; **WGA** = wheat germ agglutinin; **WT** = wild type.

Conclusions

Whole-genome sequencing relied heavily on RNA studies to identify *DMD* splice-altering variants. Short-read RNA sequencing was regularly confounded by the effectiveness of nonsense-mediated mRNA decay and low read depth of the giant *DMD* mRNA. PCR of muscle cDNA provided a simple, yet informative approach. Highly relevant to genetic therapies for dystrophinopathies, our data align strongly with previous studies of mutant dystrophin in Becker muscular dystrophy, with the collective conclusion that a fractional increase in levels of normal dystrophin between 5% and 20% is clinically significant.

Duchenne muscular dystrophy (DMD) is a severe X-linked disorder primarily affecting approximately 1 in 5,000 male births.[1–3] DMD shows a relentlessly progressive course, resulting in loss of ambulation in teens, and early mortality due to cardiac or respiratory involvement.[4,5] Dystrophinopathies range clinically from the severe DMD to asymptomatic hyperCKemia.[5–12] DMD is associated with the absence of dystrophin in muscle due to loss-of-function variants in the *DMD* gene encoding dystrophin,[5,6] whereas Becker muscular dystrophy (BMD) is associated with variants in *DMD* that result in reduced levels of (mutated) dystrophin.[5,6]

The *DMD* gene is the largest gene in the human genome, with numerous enormous introns.[13,14] One-third of pathogenic *DMD* variants are de novo,[15,16] with most affected individuals bearing insertions or deletions (indels) of coding exons.[15,17] Pathogenic *DMD* missense variants are rare,[6,15,18] and noncoding variants are emerging as an important rare cause of dystrophinopathy.[15,17,19,20] Approximately 5% of patients clinically diagnosed with DMD do not have a genetic diagnosis after mutational analysis.[5]

Herein, we show the diagnostic application of whole-genome sequencing, transcriptomics, and dystrophin protein

biochemistry to secure a genetic diagnosis for 13 affected males from 7 families with elevated creatine kinase (CK) who remained undiagnosed following multiplex ligation-dependent probe amplification (MLPA) and exomic sequencing. Importantly, we identify 3 families with *DMD* splicing variants who produce varying levels of mis-spliced transcripts that encode a premature stop codon and are targeted by nonsense-mediated decay, though express varying levels of remnant, normally spliced *DMD* mRNA. Therefore, quantitative Western blot (WB) of muscle biopsy specimens from these 3 dystrophin hypomorphs has uniquely enabled specific correlation of levels of wild-type (WT) dystrophin with clinical severity.

Methods

Standard Protocol Approvals, Registrations, and Patient Consents

This study was approved by the Children's Hospital at Westmead Human Research Ethics Committee (Biospecimen Bank_10/CHW/45) with informed, written consent from all participants.

We describe a retrospective cohort of boys diagnosed with *DMD* variants from genomic and RNA studies, who had elevated CK and dystrophic muscle biopsies, and were undiagnosed after MLPA and exomic parallel sequencing.

Immunohistochemistry and Western Blotting

Immunohistochemistry[21] and Western blotting[22] were performed as previously described; WB used NuPAGE 3%–8% Tris-Acetate precast gels (Invitrogen by Thermo Fisher Scientific, NSW, Australia). Antibodies: for immunohistochemistry, muscle fiber membranes were stained with anti-dystrophin DYS1, DYS2, DYS3, and anti-spectrin SPEC1 (Leica Biosystems, VIC, Australia); with anti-mouse Alexa Fluor 555 secondary antibody, membranes were counterstained with wheat germ agglutinin-AF488 (WGA), and nuclei were stained with DAPI (Invitrogen Thermo Fisher Scientific). WBs were probed with DYS1 (Leica Biosystems), rabbit polyclonal dystrophin antibody (Rb-DMD; ab15277; Abcam), α-actinin-2 (4A3, gift from A. Beggs, Children's Hospital Boston, Boston, MA), sarcomeric actin (clone 5C5, A2172; Sigma-Aldrich), and the anti-mouse or anti-rabbit IgG light chain HRP-conjugated secondary antibodies (GE Healthcare, NSW, Australia). The rabbit polyclonal dystrophin antibody (Rb-DMD; ab15277; Abcam) detects a 10-fold serial dilution, whereas DYS2 is less sensitive (detects a 4-fold serial dilution). Therefore, ab15277 was selected due to provision of a more informative standard curve for semiquantification of dystrophin levels in the probands. ImageJ[23] was used to measure the densities of the patient and serially diluted controls bands to create a standard curve, as previously described.[19] Semiquantitation of dystrophin levels was performed by comparing densities of the dystrophin band in patient sample relative to the standard curves of dystrophin in

2 age- and sex-matched controls across 3 experimental replicates.

Massively Parallel Sequencing

Whole-exome sequencing (probands and AI:1, AI:2, and AII:2), PCR-free whole-genome sequencing (probands from families A and B, D–G), and RNA sequencing (RNA-seq; probands from families A, B, D, E, and G) were performed at the Broad Institute of Harvard and MIT as previously described.[20] RNA-seq was performed for CII:2 at PathWest Laboratory Medicine WA as previously described for the fetal samples in reference 24.

Sanger Sequencing and RT-PCR

RNA was extracted, and reverse transcription PCR (RT-PCR) was performed as previously described.[25] Primers used for AII:1 have been previously described.[20] The remaining primer details are as follows: Ex42F 5′-CAATGCTCCTGACCTC TGTG-3′; Ex43/44R 5′-CTGTCAAATCGCCCTTGTCG-3′; *LINC00251*Ex3R 5′-CTGAAATGGGTGGGATGAAG-3′; *LINC00251*Ex2F 5′-GATGCCCCTTAACCAAGGAC-3′; Ex26F 5′-GATGCACGAATGGATGACAC-3′; Ex27R 5′-TGTGCTACAGGTGGAGCTTG-3′; Ex26/27F 5′-GCAGTTGAAGAGATGAAGAGAGC-3′; Ex29R 5′-TGGGTTATCCTCTGAATGTCG-3′; In26PF 5′-AAA-CTTAGTTCGGCCCCATG-3′; Ex48F 5′-GTTAAAT-CATCTGCTGCTGTGG-3′; Ex54R 5′-ACTGGCG-GAGGTCTTTGG-3′; Ex49/52F 5′-ACTCAGCCAGTGA-AGGCAAC-3′; Ex53R 5′-TCCTAAGACCTGCTCAGCT-TC-3′; Ex51F 5′-CGACTGGCTTTCTCTGCTTG-3′; Ex50/52F 5′-CAAATCCTGCATTGTTGCAGG-3′; *GAPDH*Ex3F 5′-TCACCAGGGCTGCTTTTAAC-3′; and *GAPDH*Ex6R 5′-GGCAGAGATGATGACCCTTT-3′. Confirmation and segregation analysis of *DMD* variants was performed by Sanger sequencing,[21] except for family F in which DNA was not available. Primers used for families A, D, E, and G have been previously described.[20] The remaining primer details are as follows: family B—In43F 5′-TTTAGTTTCCAGC-CACTCCTGTC-3′ with chr8R 5′-TAGCAGGGGCAAGG-GTTG-3′ and chr8F 5′-TGCCTCTCCAGAATGAGGAC-3′ with In43R 5′-CGGGGAACATCACACACC-3′ to confirm insertion breakpoints; family C—In26F 5′-CGAAGGAAAC-TGGTATGTAG-3′ with In26R 5′-AAAGCCGTATGACA-GATTCG-3′ to determine causative variant. PCR conditions were 5 minutes 95°C; 35 cycles—30 seconds 95°C, 30 seconds 58°C, and 1 minute 72°C; 8 minutes 72°C; or as described in reference 20.

Whole-Genome Sequencing Analysis

PCR-Free whole-genome sequencing was performed on an Illumina HiSeq X Ten using 2 × 150 paired end reads at 30× mean coverage. The sequencing reads were aligned to the GRCh37 genome reference and single nucleotide variants (SNVs), small insertions and deletions (indels) were detected using methods previously described in reference 20. A reanalysis of rare (Genome Aggregation Database [gnomAD] AF < 0.005) SNVs and indels

Table Clinical Presentation, *DMD* Variants, and Dystrophin Western Blot

	AII:1	BIV:1	CII:2	DII:1	EII:1	FII:6	GII:1
Clinical symptoms	Muscle pain, fatigue, and myoglobinuria with exercise	Proximal weakness and bilateral calf hypertrophy	Progressive limb-girdle weakness and falling regularly	Proximal weakness, calf hypertrophy, and positive Gowers sign	Muscle weakness and calf hypertrophy	Proximal muscle weakness, positive Gowers sign, and calf hypertrophy	Calf hypertrophy and positive Gowers sign; proximal weakness, elbow contractures, and learning difficulties
Onset	15 y	5 y	9 y	3.5 y	6 y	3.5 y	5 y
Family history	2 affected brothers reporting myalgia and serum CK levels of 300–14,700 U/L	Four-generation family segregating with an X-linked muscular dystrophy with cardiomyopathy	Nil	Nil	Has a similarly affected brother	The mother (FI:6) also has muscle pain and elevated serum CK levels of ~500 U/L	Nil
Serum CK, U/L	1,400–7,500	9,964	420	14,500	18,889	24,000	>12,000
Ambulance	Remains ambulant	Remains ambulant	Intermittent use of a wheelchair from 13 y	Wheelchair dependent at 13 y	Wheelchair dependent at 9 y	Remains ambulant and toe walking at 9 y	Wheelchair dependent at 7 y
Cardiac and respiratory involvement	Nil	Normal echocardiogram cardiomyopathy in BIII:2 and BIII:7	Nocturnal bilateral positive airway pressure at 28 y; normal cardiac function	Cardiac-reduced contractility (ejection fraction 30%–35%) with normal left ventricle size	Nil	Nil	Borderline increase in heart size at 9 y; died at age 10 y from cardiac complications
DMD variant	Pseudoexon inclusion in *DMD* intron 43 NM_004006.2:c.6290+30954C>T	~116 kb chr8 insertion in *DMD* intron 43 NM_004006.2:c.6290+28627_6290+28628ins[TGTGGGCAAAGGC; NR_038901.1:-100749_430-3036; NM_004006.2:6290+28628_6290+28751]	Pseudoexon inclusion in *DMD* intron 26 NM_004006.2:c.3603+820G>T	Inversion of *DMD* exon 51 NM_004006.2:c.7310-2629_7542+1338inv	Inversion of *DMD* exons 1–18 NM_004006.2:c.-1950935_2293-1933inv	Inversion of *DMD* exons 1–44 NM_001304548.1:6818-10658_NM_004006.2:6438+112319inv; 4233+10599_5325+387dup; 6117+6701_6438+112319dup]	Inversion of *DMD* exons 1–60 NM_004006.2:c.[-117965533_9085-12259inv;-117965534_-117965551del; 9085-12258_9085-12196del]
Western blot	15% ± 2%	10% ± 2%	0%–5%	0%–5%	0%–5%	0%–5%	0%–5%

Abbreviations: CK = creatine kinase; DMD = Duchenne muscular dystrophy; *DMD* = *DMD* gene or transcript.

revealed no pathogenic *DMD* variants. The Manta tool from Illumina (PMID: 26647377) was used to identify structural variants or split read abnormalities within the *DMD* gene. Putative structural variants were manually inspected within Integrative Genomic Browser (IGV) to validate and resolve exact breakpoints of structural rearrangements.

RNA-seq Analysis

RNA-seq analysis was performed as described in reference 20. Briefly, all samples were jointly processed and aligned with the Genotype-Tissue Expression Consortium (GTEx)[26] to identify spliced reads only seen in patients or groups of patients and missing in controls. In addition, given the nature of the previously suspected diagnosis of a dystrophinopathy, in cases in which this approach did not lead to a diagnosis, exonic read depth was mapped in each patient and compared with controls and sashimi plots of patients were manually inspected using the IGV for the *DMD* gene. In cases, in which RNA-seq identified a mis-splicing event, patient exome and genomes were manually evaluated, depending on availability.

Data Availability

Data not published within this article are available by request from any qualified investigator.

Figure 1 Pedigree of Families A–G
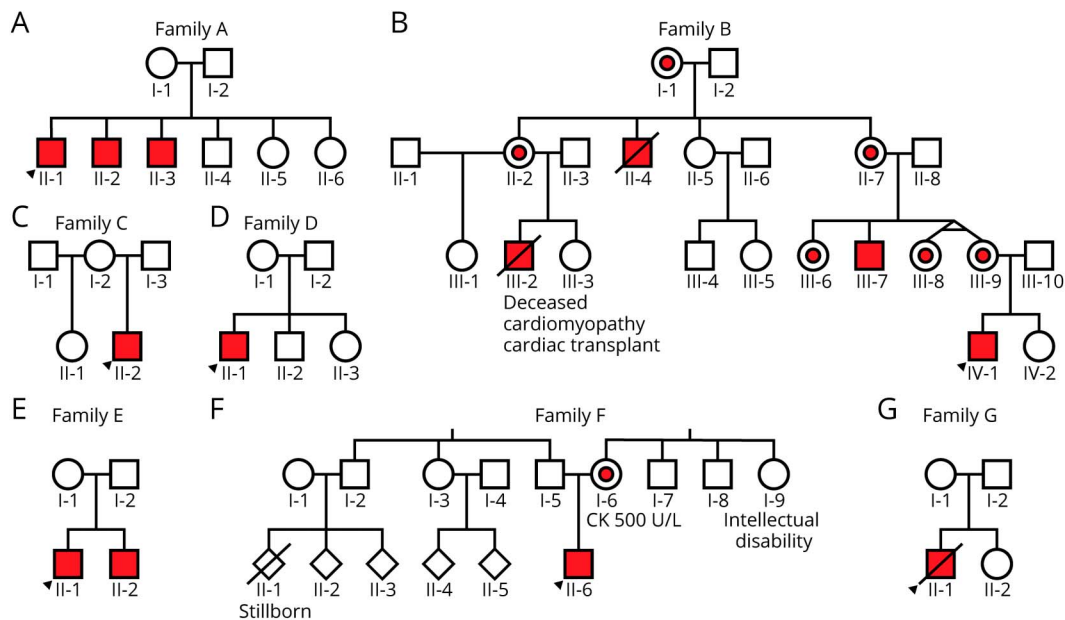
Index patient for each family denoted with black arrow. Affected members colored in red, and carriers part colored in red.

Results

Clinical Presentation

Four families have been described previously in reference 20: AII:1 as N33; DII:1 as C3; EII:1 as C4; and GII:1 as C2. Clinical presentation, *DMD* variants, and dystrophin WB results are summarized in the table. Briefly, AII:1 presented at 15 years with muscle pain, fatigue, and episodes of myoglobinuria with exercise and elevated serum CK (CK 1,400–7,500 U/L, normal range <200 U/L). He has 2 affected brothers with myalgia and elevated serum CK (300–14,700 U/L (figure 1A). Family B is a 4-generation family with an X-linked muscular dystrophy with cardiomyopathy. BIII:2 was diagnosed with dilated cardiomyopathy in his 20s, underwent cardiac transplantation at age 29 years, and died of transplant-related complications at age 31 years. BIII:7 was diagnosed with BMD in his mid-teens. He has no known history of cardiomyopathy and remains ambulant in his 40s (figure 1B). BIV:1 showed elevated serum CK 9,964 U/L at age 6 months. Now age 5 years, he has proximal muscle weakness, bilateral calf hypertrophy, and normal echocardiogram. CII:2 presented at age 9 years with progressive limb-girdle weakness, requiring intermittent use of a wheelchair from age 13 years and nocturnal bilateral positive airway pressure (BiPAP) from age 28 years. He has normal cardiac function with serum CK of 420 U/L at age 31 years (figure 1C). DII:1 presented at age 3.5 years with proximal weakness, calf hypertrophy, positive Gowers sign, and serum CK of 14,500 U/L. He required use of a wheelchair from age 13 years. Echocardiogram at age 17 years showed reduced contractility (ejection fraction 30%–35%) with normal left ventricle size (figure 1D). EII:1 presented at age 6 years with muscle weakness, enlarged calves, and serum CK of 18,889 U/L. He required use of a wheelchair at age 9 years and

has no known cardiac or respiratory involvement. EII:1 has a similarly affected brother (figure 1E). FII:6 presented at age 3.5 years with proximal muscle weakness, positive Gowers sign, prominent calves, and serum CK of 24,000 U/L. He remains ambulant, but is toe walking at age 9 years. He has no known cardiac or respiratory involvement. FII:6's mother (FI:6) reports muscle pain and has elevated serum CK of ~500 U/L (figure 1F). GII:1 presented at age 5 years with waddling gait, calf hypertrophy, positive Gowers sign, and serum CK levels of >12,000 U/L. He required use of a wheelchair from age 7 years. Echocardiogram at age 9 years showed borderline increase in heart size, and he died at age 10 years from cardiac complications (figure 1G).

DMD Diagnostic Genetic Testing

DMD MLPA and Sanger sequencing were performed and reported normal for AII:1, BIV:1, CII:2, DII:1, EII:1, and GII:1. *DMD* MLPA performed for FII:6 revealed duplications of exons 31–37 and 43–44, which were predicted to be in-frame and therefore considered inconsistent with his severe Duchenne-like phenotype, though with high clinical suspicion of causality. A genetic basis could not be identified via whole-exome sequencing (AII:1, BIV:1, DII:1, EII:1, FII:6, and GII:1, with duplications of exons 31–37 and 43–44 confirmed for FII:6) or massively parallel sequencing of a targeted neuromuscular gene panel (CII:2).

Immunohistochemistry Demonstrates Dystrophin Abnormalities in Skeletal Muscle Biopsies

Skeletal muscle immunohistochemistry for AII:1, BIV:1, DII:1, EII:1, FII:6, and GII:1 confirms abnormalities in dystrophin

Figure 2 Muscle RNA Studies of *DMD* in Patients
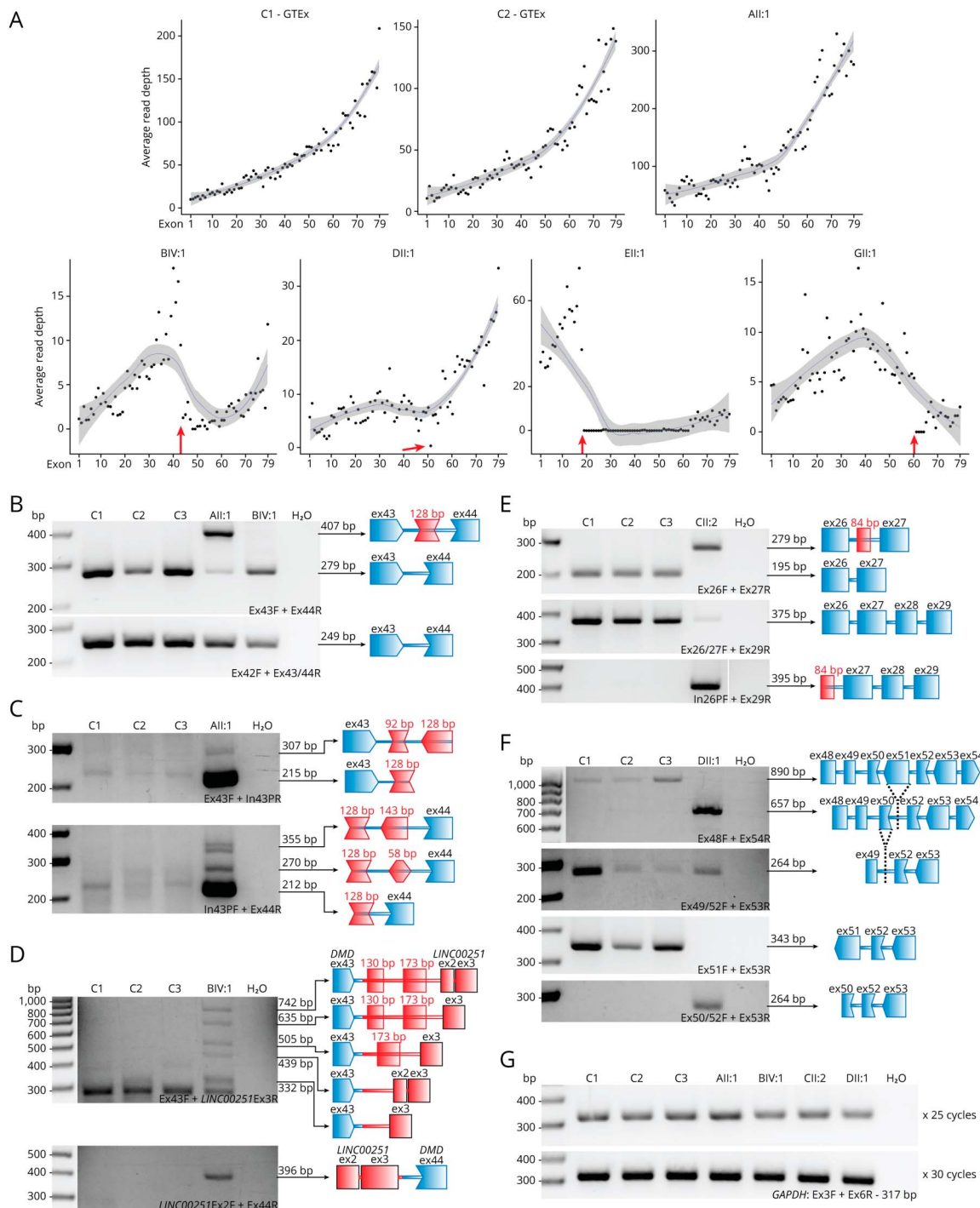
(A) RNA-seq read coverage of *DMD* exons in muscle RNA from AII:1, BIV:1, DII:1, EII:1, and GII:1 and 2 GTEx controls. Red arrows indicate the reduction in read depth, which corresponds with the location of *DMD* structural variants for BIV:1, DII:1, EII:1, and GII:1. (B–G) RT-PCR studies of muscle-derived RNA of patients with splicing abnormalities and 3 male controls (C1, quadriceps, 6.5 years; C2, vastus lateralis, 17 years; C3, unknown, 20 years). Primers used are listed at the bottom right of each gel image and are labeled according to their location (exon; Ex, intron; In, pseudoexon; P) and orientation (forward; F, reverse; R). Bridging primers span a splice junction and are denoted by X/Y, where X and Y are exons the primer spans. All results were confirmed by Sanger sequencing. (B) RT-PCR showing reduced levels of correctly spliced *DMD* transcript (exons 43 and 44) in AII:1 and BIV:1 compared with controls. AII:1 shows the inclusion of a 128-bp pseudoexon. (C) Primers specific to the 128 bp pseudoexon revealed that the inclusion is specific to AII:1 (Sanger sequencing showed that the faint bands in C1 were non-*DMD* sequences). Sequencing reveals that faint bands in AII:1 correspond to multiple pseudoexons in *DMD* incorporated into a minority of *DMD* transcripts. (D) Various chr8 pseudoexons and *LINC00251* exons are included in *DMD* transcripts as a result of the chr8 insertion in BIV:1. The lowest band detected in all samples in the top gel corresponds to non-*DMD* sequences. (E) RT-PCR confirms the inclusion of a 84-bp pseudoexon in CII:2 in the majority of *DMD* transcripts. Normal splicing can only be detected in very low levels in CII:2 by bridging primers. The 92 bp pseudoexon is absent in control samples. (F) RT-PCR of DII:1 confirms that exon 51 is absent from all *DMD* transcripts. A bridging primer indicates that skipping of both exons 50 and 51 is a low frequency event observed in both controls and DII:1. (G) *GAPDH* loading controls to indicate that similar concentrations of complementary DNA were used for both control and patient samples. DMD = Duchenne muscular dystrophy; *DMD* = *DMD* gene or transcript; RT-PCR = reverse transcription PCR.

Figure 3 Schematics of Variants Identified in Families A–G
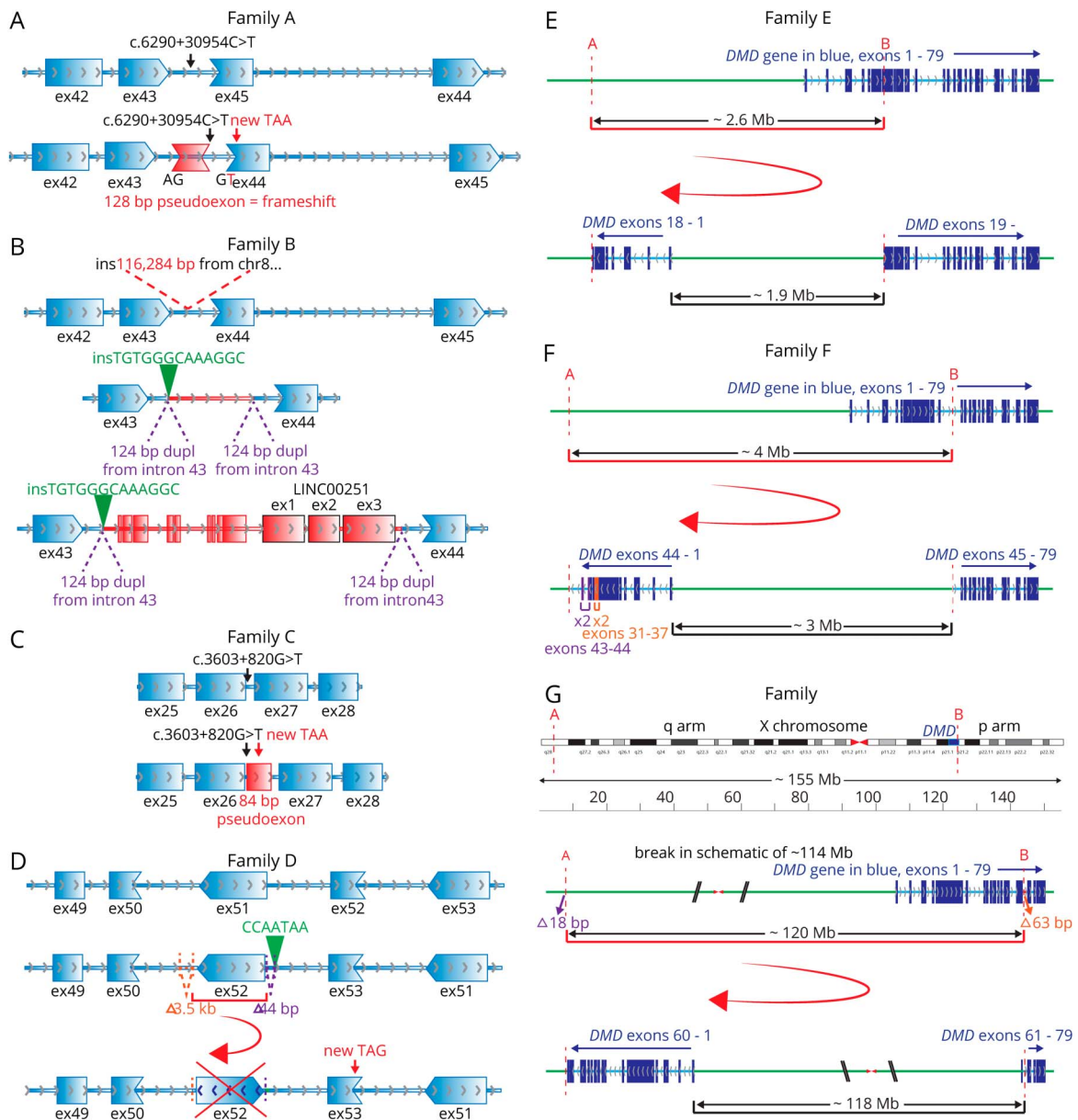
(A) Family A: intronic c.6290+30954C>T (black arrow) creates a cryptic donor splice site, leading to inclusion of a 128-bp pseudoexon (red, within *DMD* intron 43) into the *DMD* mRNA, causing a frameshift and stop codon (red arrow) encoded by exon 44 (ex44). Gene direction is demonstrated by gray arrows. Reading frame between exons is shown by shape complementarity. (B) Family B: insertion of 116,284 bp of chr8 (red sequence) into *DMD* intron 43. The insertion includes LINC00251 exons 1–3 (black outlined exons). A 124-bp sequence of intron 43 of *DMD* (chrX:32,276,895-32,277,018) is duplicated as part of the structural rearrangement and now flanks the chr8 insertion. In addition, there is an insertion of 13 bp (insGCCTTTGCCCACA, shown in green) adjacent to 1 copy of the 124-bp duplication. mRNA studies show evidence for numerous, different abnormal splicing events from *DMD* exon 43 to various pseudoexons (red exons) and LINC00251 exons (red exons with black outlines) within the chr8 insertion. Low levels of normal *DMD* splicing (from exons 43 and 44; blue exons) are also observed. Frame of splicing in pseudoexons and LINC00251 exons not shown. (C) Family C: intronic c.3603+820G>T (black arrow) increases the strength of the polypyrimidine tract leading to use of a cryptic acceptor splice site (3/5 algorithms within Alamut Visual biosoftware predictions; MaxEntScan, NNSPLICE, and GeneSplicer) leading to inclusion of a 84-bp pseudoexon (red, within *DMD* intron 26) into the *DMD* mRNA, encoding a stop codon (red arrow) 39 nucleotides into the pseudoexon. Gene direction is demonstrated by gray arrows. Reading frame between exons is shown by shape complementarity. (D) Family D: inversion of *DMD* exon 51 and flanking adjacent intronic sequence. Flanking the structural rearrangement are 2 intronic deletions (orange 3.5 kb and purple 44 bp) and an insertion of CCAATA (green). mRNA studies show exon 51 skipping, causing a frameshift and a premature stop codon (TAG, encoded by exon 52; red arrow). (E) Family E: A 2.6-Mb inversion on the X chromosome between 2 breakpoints; A in intron 45 of *CFAP47*, 1.9 Mb upstream of exon 1 of *DMD* (GRCh37:chrX:35,180,364) and B in intron 18 of *DMD* (GRCh37:chrX:32,521,892, NM_004006.2). This reverses the orientation of exons 1–18 of *DMD*, which are now joined to *CFAP47* sequences upstream of exon. The *DMD* gene is in blue, exons dark blue, and introns light blue. Intergenic sequence (non-*DMD* ChrX in green). (F) Family F: A 4.1-Mb inversion on the X chromosome between 2 breakpoints; A is 3.8 Mb upstream of exon 1 of *DMD* (GRCh37:chrX:36236087) and B in intron 44 of *DMD* (GRCh37:chrX:32122714). This reverses the orientation of exons 1–44 of *DMD*, which are now joined to intergenic sequence upstream of exon 1. This is accompanied by duplication of exons 31–37 (orange) and exons 43 and 44 (purple) around the breakpoint. (G) Family G: A 119.8-Mb inversion on the X chromosome between 2 breakpoints; A in an intergenic region on the q arm of the X chromosome, 118 Mb upstream of exon 1 of *DMD* (GRCh37:chrX: 151,194,962), and B in intron 60 of *DMD* (GRCh37:chrX:31,379,010, NM_004006.2). This reverses the orientation of exons 1–60 of *DMD*, which are now joined to intergenic sequence upstream of exon 1. In addition, 2 deletions were identified at these breakpoints; an intronic 63 bp deletion (orange, GRCh37:chrX: 31,378,947-31,379,009) and an intergenic 18 bp deletion (purple, GRCh37:chrX:151,194,963-151,194,980). X chromosome displayed in unusual orientation with q arm to the left, so the *DMD* gene is presented with exons in order. *DMD* = *DMD* gene or transcript. mRNA = messenger RNA.

(figure e-1, links.lww.com/NXG/A367). Using 3 anti-dystrophin antibodies, AII:1 and BIV:1 showed reduced dystrophin staining, whereas DII:1, EII:1, FII:6, and GII:1 showed absent staining (figure e-1 absent dystrophin staining shown only for GII:1). WGA outlines the myofibers and labels the endomysium in patient and control skeletal muscle samples.

Correlation of Splicing Analyses With Whole-Genome Sequencing Identifies Pathogenic Intronic and Structural Variants Inducing Abnormal *DMD* Splicing

Six individuals (AII:1, BIV:1, DII:1, EII:1, FII:6, and GII:1) were subject to whole-genome sequencing, 6 individuals were subject to RNA-seq (AII:1, BIV:1, CII:2, DII:1, EII:1, and GII:1), and 4 individuals (AII:1, BIV:1, CII:2, and DII:1) were analyzed by RT-PCR of muscle-derived mRNA. Scrutiny of *DMD* transcripts (NM_004006.2, 11,058 nucleotides [nt] in length) shows typical 3′ bias in read depth (vastly more reads at the 3′ end compared with the 5′ end of *DMD* transcripts). Acknowledging 3′ bias, an abnormal profile of *DMD* transcript read depth was apparent for BIV:1, DII:1, EII:1, and GII:1 (figure 2A), relative to multiple muscle controls from the GTEx consortium.[26]

Standard variant filtering approaches of genomic sequencing failed to identify most causal variants. RNA-seq identified abnormal pseudoexon inclusion into *DMD* transcripts for families A and C. The remaining pathogenic variants were identified only through the combination of whole-genome sequencing, bioinformatics, and RNA analyses.

A genetic diagnosis in AII:1 was identified in a previous study[20] with a deep intronic pathogenic variant GRCh37: ChrX:32274692G>A; c.6290+30954C>T inducing partial mis-splicing of *DMD*. The *DMD* c.6290+30954C>T variant creates a cryptic donor 5′ splice site resulting in inclusion of a variant-activated pseudoexon of 128 nt inserted between exon 43 and exon 44, which encodes 59 missense amino acids and effects a frameshift, resulting in a premature termination codon encoded by exon 44 (figure 3A). RT-PCR confirmed abnormal inclusion of the variant-activated pseudoexon and residual normal splicing of *DMD* exons 42-43-44-45 (figure 2, B and C).

RNA-seq for BIV:1 showed low levels of *DMD* transcripts, with a distinct drop in reads from exon 44 onward (figure 2A, arrow). Bespoke realignment and analyses of WGS data identified insertion of ~118,000 nt of chromosome 8 (chr8) sequences within *DMD* intron 43, encompassing the *LINC00251* gene locus. RT-PCR showed that the chr8 insertion induced abnormal splicing of the *DMD* gene (figure 3B). Multiple adverse events were detected that involved splicing from exon 43 of *DMD* to various pseudoexons and

---

Figure 4 Western Blot Panel for All Patients
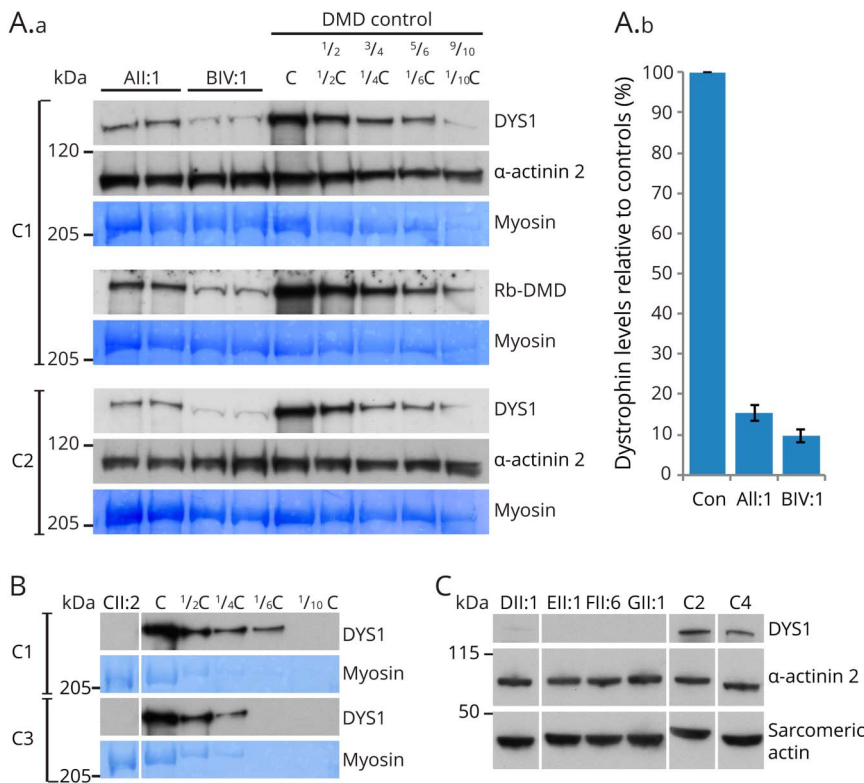
(A.a) Western blot was performed on skeletal muscle from index patients from families A and B (AII:1 and BIV:1) against DYS1 (rod domain epitope) and Rb-DMD (C-terminal epitope) with serial dilutions (1/2, 3/4, 5/6, and 9/10) human control skeletal muscle. Muscle lysate derived from an individual with Duchenne muscular dystrophy and undetectable levels of dystrophin by Western blot (DMD control; deltoid, 14-year-old boy, GRCh37:chrX:32364116G>A, NM_004006.2: c.5530C>T, p.Arg1844*) were added to diluted controls to normalize total protein loading in each lane of the gel. Loading controls: a-actinin-2 and myosin (coomassie). (A.b) Image J[23] was used to measure the densities of the patient and serially diluted controls bands to create a standard curve. Quantification of relative dystrophin levels was performed by comparing patient sample densities to the control standard curves across the 3 gels shown. AII:1 demonstrates 15.5% ± 1.9% levels of dystrophin protein relative to controls. BIV:1 demonstrates 9.6% ± 1.7% levels of dystrophin relative to control. (B) Western blot analysis on skeletal muscle from patient CII:2 against DYS1 shows undetectable levels of dystrophin compared with controls. Loading controls: myosin (coomassie). (C) Western blot analysis on skeletal muscle from patients DII:1, EII:1, FII:6, and GII:1 against DYS1 compared with human control skeletal muscle. DII:1 shows very low levels of dystrophin. EII:1 FII:6, and GII:1 show undetectable levels of dystrophin. Loading controls: a-actinin-2 and sarcomeric actin. Male controls used: C1, tibialis anterior, 16 years; C2, unknown, 5.5 years; C3, unknown, 14 years; C4, quadriceps, 4.5 years. DMD = Duchenne muscular dystrophy.

LINC00251 exons within the chr8 insertion. Sanger sequencing with bespoke PCR over the breakpoints on gDNA confirmed the chr8 inclusion in intron 43 and provided a diagnostic assay that confirmed segregation of the insertion within the family pedigree. Normal splicing of *DMD* exons 42-43-44-45 was observed as a low-frequency event (figure 2, B and D).

For CII:2, manual analysis of RNA-seq data identified abnormal inclusion of 84 nt from intron 26 into a majority of *DMD* transcripts (figure 3C). Sanger sequencing of the genomic region in gDNA from CII:2 identified a deep intronic variant GRCh37: ChrX:32471959C>A, c.3603+820G>T that was absent in gnomAD. The *DMD* c.3603+820G>T variant in intron 26 disrupts an AG, creating an AG-exclusion zone between an available consensus lariat branch point and 3′ splice site.[27] Spliceosomal use of a naturally occurring consensus 5′ splice site sequence and this strengthened 3′ splice site result in the inclusion of a variant-activated pseudoexon into a majority of *DMD* transcripts, encoding 19 missense amino acids followed by a stop codon (figure 3C). RT-PCR confirmed abnormal inclusion of the variant-activated pseudoexon into *DMD* transcripts and residual, low levels of *DMD* transcripts with normal splicing of exons 25-26-27 (figure 2E). Sanger sequencing confirmed that the c.3603+820G>T variant was de novo in CII:2.

For DII:1, RNA-seq in a previous study[20] showed low levels of *DMD* transcripts with exon 51 skipping, inducing a frameshift and premature stop codon encoded by exon 52 (r.7310_7542del, p.Ser2437Cysfs*33, figure 3D). Interrogation of WGS determined presence of a *DMD* structural rearrangement rendering *DMD* exon 51 in the reverse orientation and unable to be spliced into the *DMD* mRNA, confirmed by Sanger sequencing. RT-PCR confirms exon 51 skipping as the predominant mis-splicing event in DII:1, with skipping of exons 50 and 51 a low-frequency, in-frame event observed in both DII:1 and controls (figure 2F). Low levels of exon 50 and 51 skipping are consistent with low levels of dystrophin detected by WB analysis (figure 4C).

RNA-seq showed an abrupt loss of transcripts after exon 18 in EII:1, as previously described in reference 20 (figure 2). WGS showed evidence for an inversion within the *DMD* gene reversing the orientation of exons 1–18 of *DMD*, which are now joined to intergenic sequences upstream of exon 1, explaining the presence of abruptly terminating exon 1–18 transcripts transcribed from the *DMD* promoter (figure 3E). The 1.9 Mb intergenic region included in the inversion contains *FAM47A*, *FAM47B*, and *TMEM47* genes. Sanger sequencing of genomic DNA over the breakpoints confirmed the inversion.

FII:6 with in-frame duplications of exons 31–37 and 43–44 identified on *DMD* MLPA, was shown by WGS to have a larger, more complex structural rearrangement (figure 3F), which reverses the orientation of exons 1–44 of *DMD* which are now joined to intron 45 of *CFAP47*, upstream of exon 1. Expression of *CFAP47* is likely to be disrupted. However, the clinical significance of loss of *CFAP47* expression is unknown.

In GII:1, RNA-seq in a previous study[20] showed low read count for *DMD* transcripts, with evidence for even fewer reads from exon 60. Closer scrutiny of whole-genome sequencing data identified a structural rearrangement reversing the orientation of exons 1–60 of *DMD*, which are now joined to intergenic sequences upstream of exon 1 (figure 3G). Sanger sequencing of genomic DNA over the breakpoints confirmed the inversion.

WB Analyses Define the Threshold of WT Dystrophin Conferring Clinical Phenotypes of Duchenne to Myalgia

Our splicing studies reveal that AII:1, BIV:1, and CII:2 each have residual levels of normally spliced *DMD* transcripts, with abnormal splicing events apparently targeted for degradation by nonsense-mediated decay (figure 2, B–E). Therefore, these individuals uniquely provide an opportunity to quantify levels of WT dystrophin and correlate with clinical phenotype. Quantitative WB (figure 4) using skeletal muscle biospecimens reveals (1) ∼15% ± 2% normal dystrophin levels in AII:1, correlating with a myalgia phenotype without apparent weakness; (2) ∼10% ± 2% levels of dystrophin in BIV:1 (figure 4A) with Becker muscular dystrophy, mild weakness, and cardiac phenotype; and (3); 0%–5% levels of dystrophin in affected individuals who present with a severe Becker (CII:2) or Duchenne phenotype (DII:1, EII:1, FII:6, and GII:1) (figure 4, B and C).

Discussion

Our study further substantiates *DMD* splicing variants as an important causal basis for males presenting with symptoms consistent with a dystrophinopathy, for whom exomic sequencing approaches or MLPA return negative findings. A causal splicing variant in *DMD* was identified in all 7 families within our dystrophinopathy cohort and includes 13 affected males presenting with hyperCKemia with pain and/or muscle weakness and/or cardiac involvement.

Importantly, identification of the causative variant in *DMD* within this hard-to-diagnose cohort required deployment of WGS, RNA-seq and/or bespoke RT-PCR studies of mRNA isolated from skeletal muscle. For example, for CII:2, RNA-seq was crucial to identify the inclusion of an 84 base pair (bp) pseudoexon encoding a frameshift which prompted Sanger sequencing of this region, which lead to the identification of the casual intron 26 c.3603+820G>T variant, which was undetectable by gene panel testing, Sanger sequencing of the individual exons or MLPA. Although multiple genetic investigations are costly and not available currently to many diagnostic laboratories, costs incurred through muscle biopsy, WGS, or RNA studies are insignificant relative to the cost burden to health services for dystrophinopathy cases, for example, the heart transplantation for family B. A precise genetic diagnosis for an X-linked disorder has important and wide-reaching implications for genetic, prenatal, and prognostic counseling across the wider family unit and can inform

reproductive decision making. In addition, a genetic diagnosis could enable future customizable treatments such as splice-modulating antisense oligonucleotide drugs,[28] which would theoretically be applicable to families A, C, and D.

Although WGS and RNA-seq bring powerful adjunct tests to clinical genomics, shortcomings of short read massively parallel sequencing were clearly observed in this study. As human exons are typically 100–150 bp in length, short-read RNA-seq is limited in that a single read does not effectively bridge multiple exons. Most significantly, RNA-seq for BIV:1, DII:1, EII:1, and GII:1 was confounded by the effectiveness of nonsense-mediated decay, an innate surveillance mechanism that degrades mRNA bearing a premature stop codon.[15,17] We suspect that the reason we do not see a profound reduction in read depth for AII:1 (figure 2A) is due in part to higher read depth across the transcriptome, including *DMD*, and in part to the residual normal splicing of a significant proportion of *DMD* transcripts (27%). Nonsense-mediated decay amplifies inherent challenges associated with RNA-seq of very large mRNAs, where mRNA capture and sequencing library construction result in a characteristic bias in read depth, with vastly more reads at the 3′ end than the 5′ end of a very long mRNA. Notably, common disease genes in neuromuscular disorders are among the largest coding mRNAs in humans, with *DMD* mRNA ~14,000 nt, *NEB* mRNA ~50,000 nt, and *TTN* mRNA ~100,000 nt. Therefore, ribosomal RNA depletion and/or long read RNA-seq approaches, which display reduced 3′ bias, may be more effective for diagnosing neuromuscular disorders.

Regular data filtering approaches of genomic sequencing failed to identify most of the causal variants (excluding families A and C found on RNA-seq). This is likely due to the nature of the variants themselves (noncanonical splice affecting variants or structural variants), small read lengths, and mapping restrictions against the reference sequence. The structural rearrangement within *DMD* intron 43 of family B took extensive bioinformatic analysis to delineate, even when our RT-PCR (data not shown) and RNA-seq studies had indicated intron 43 as the likely location of the problem. Although (in retrospect) the copy number variation of the duplicated region of chr8 is evident, informatics approaches to map split reads to precisely define the breakpoints were challenging and ultimately required both informatics and Sanger sequencing of PCR amplicons to fully resolve. Of note, the bespoke PCR uniquely identifying the *DMD* intron 43 structural rearrangement was clinically preferred as the diagnostic test for segregation and carrier testing due to its greater specificity relative to the microarray to detect the chr8 copy number variation. The availability of a validated bespoke PCR also means that carrier females in this family could have prenatal diagnosis of male pregnancies.

Although families B, D, and G have cardiac involvement that is common in dystrophinopathy, families A, C, E, and F do not have reported cardiac symptoms and are being monitored for possible development of cardiac symptoms. The profound cardiac involvement in family B raises suspicion of potential differences in *DMD* pre-mRNA mis-splicing between cardiac and skeletal muscle activated by the insertion of 118 kb of Chr8 sequences containing the LINC00251 gene. It is plausible that the severe cardiac involvement in family B is due to more fully penetrant *DMD* mis-splicing in cardiac tissue compared with skeletal muscle. Unfortunately, no stored cardiac specimens were available for mRNA studies from other affected family members who had undergone transplant surgery. It is also possible that levels of inclusion of the frameshifting pseudoexon in family C may differ between skeletal muscle (and potentially between different skeletal muscles) and cardiac muscle.

In conclusion, we highlight *DMD* splicing variants as an important causal basis in individuals with a suspected dystrophinopathy who remain undiagnosed after exomic sequencing or MLPA approaches. Causative *DMD* variants identified in AII:1, BIV:1, and CII:2 that induce partial mis-splicing of *DMD* mRNA provided us with a unique opportunity; each affected individual produced varying levels of remnant, normally spliced *DMD* mRNA, with all mis-spliced transcripts encoding a premature stop codon and targeted by nonsense-mediated decay. Therefore, we were able to use quantitative WB to correlate levels of WT dystrophin with clinical severity. We establish a steep therapeutic range of WT dystrophin protein levels (figure 4A); with ~15% WT dystrophin associated with myalgia without apparent weakness, ~10% levels of WT dystrophin associated with Becker muscular dystrophy, mild weakness, and cardiac phenotype, and <5% WT dystrophin associated with a severe Becker or Duchenne-like phenotype. Our findings broadly concur with previous studies correlating levels of mutated dystrophin in BMD with clinical severity,[7,29–31] supporting the notion of a functional redundancy within the spectrin-like repeats of the dystrophin rod domain. Of great relevance to international efforts to develop genetic therapies in DMD, our data provide compelling evidence that with early intervention, only fractional increases in levels of dystrophin are likely to result in clinical improvement.

Disclosure

S.T. Cooper is director of Frontier Genomics Pty Ltd. (Australia). Frontier Genomics has not traded (as of September 25, 2020). Frontier Genomics Pty Ltd. (Australia) has no existing financial relationships that will benefit from publication of these data. N.F. Clarke is deceased; disclosures are not included for this author. The remaining coauthors do not have any relationships, financial or otherwise, that may result in a perceived conflict of interest. Go to Neurology.org/NG for full disclosures.

Publication History

Appendix Authors

Name	Location	Contribution
Leigh B. Waddell, PhD	The Children's Hospital at Westmead, New South Wales, Australia	Concept and design of the study, acquisition and analysis of data, and drafting the manuscript/figures for intellectual content
Samantha J. Bryen, BSc (Hons)	The Children's Hospital at Westmead, New South Wales, Australia	Concept and design of the study, acquisition and analysis of data, and drafting the manuscript/figures for intellectual content
Beryl B. Cummings, PhD	Broad Institute of MIT & Harvard, Cambridge, MA	Acquisition and analysis of data and drafting the manuscript/figures for intellectual content
Adam Bournazos, BSc (Hons)	The Children's Hospital at Westmead, New South Wales, Australia	Acquisition and analysis of data and drafting the manuscript/figures for intellectual content

Appendix (continued)

Name	Location	Contribution
Frances J. Evesson, PhD	The Children's Hospital at Westmead, New South Wales, Australia	Acquisition and analysis of data and drafting the manuscript/figures for intellectual content
Himanshu Joshi, B Software Engineering, B Business (Finance)	The Children's Hospital at Westmead, New South Wales, Australia	Acquisition and analysis of data and drafting the manuscript/figures for intellectual content
Jamie L. Marshall, PhD	Broad Institute of MIT & Harvard, Cambridge, MA	Acquisition and analysis of data
Taru Tukiainen, PhD	Broad Institute of MIT & Harvard, Cambridge, MA	Acquisition and analysis of data
Elise Valkanas, BA (Biology)	Broad Institute of MIT & Harvard, Cambridge, MA	Acquisition and analysis of data
Ben Weisburd, BS (Computer Sc)	Broad Institute of MIT & Harvard, Cambridge, MA	Acquisition and analysis of data
Simon Sadedin, PhD	Broad Institute of MIT & Harvard, Cambridge, MA	Acquisition and analysis of data
Mark R. Davis, PhD	PathWest Laboratory Medicine WA, Nedlands, Australia	Acquisition and analysis of data
Fathimath Faiz, PhD	PathWest Laboratory Medicine WA, Nedlands, Australia	Acquisition and analysis of data
Rebecca Gooding, PhD	PathWest Laboratory Medicine WA, Nedlands, Australia	Acquisition and analysis of data
Sarah A. Sandaradura, MBChB, FRACP, PhD	The Children's Hospital at Westmead, New South Wales, Australia	Acquisition and analysis of data and drafting the manuscript/figures for intellectual content
Gina L. O'Grady, MBChB, FRACP, PhD	The Children's Hospital at Westmead, New South Wales, Australia	Acquisition and analysis of data
Michel C. Tchan, MBBS, FRACP, PhD	Westmead Hospital, New South Wales, Australia	Acquisition and analysis of data
David R. Mowat, MBBS, FRACP	Sydney Children's Hospital, Randwick, New South Wales, Australia	Acquisition and analysis of data
Emily C. Oates, MBBS, FRACP, PhD	The Children's Hospital at Westmead, New South Wales, Australia	Acquisition and analysis of data
Michelle A. Farrar, MBBS, FRACP, PhD	Sydney Children's Hospital, Randwick, New South Wales, Australia	Acquisition and analysis of data

Continued

References

1. Moat SJ, Bradley DM, Salmon R, Clarke A, Hartley L. Newborn bloodspot screening for Duchenne muscular dystrophy: 21 years experience in Wales (UK). Eur J Hum Genet 2013;21:1049–1053.
2. Helderman-Van Den Enden ATJM, Madan K, Breuning MH, et al. An urgent need for a change in policy revealed by a study on prenatal testing for Duchenne muscular dystrophy. Eur J Hum Genet 2013;21:21–26.
3. Mendell JR, Shilling C, Leslie ND, et al. Evidence-based path to newborn screening for Duchenne muscular dystrophy. Ann Neurol 2012;71:304–313.
4. Emery AEH. The muscular dystrophies. Lancet 2002;359:687–695.
5. Flanigan KM. Duchenne and Becker muscular dystrophies. Neurol Clin 2014;32: 671–688.
6. Flanigan KM, Dunn DM, Von Niederhausern A, et al. Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. Hum Mutat 2009;30:1657–1666.
7. Bushby KMD, Gardner-Medwin D. The clinical, genetic and dystrophin characteristics of Becker muscular dystrophy: I. Natural history. J Neurol 1993;240:98–104.
8. Yazaki M, Yoshida K, Nakamura A, et al. Clinical characteristics of aged Becker muscular dystrophy patients with onset after 30 years. Eur Neurol 1999;42:145–149.
9. Heald A, Anderson LVB, Bushby KMD, Shaw PJ. Becker muscular dystrophy with onset after 60 years. Neurology 1994;44:2388–2390.
10. Bushby KMD, Cleghorn NJ, Curtis A, et al. Identification of a mutation in the promoter region of the dystrophin gene in a patient with atypical Becker muscular dystrophy. Hum Genet 1991;88:195–199.
11. Minetti C, Tanji K, Chang HW, et al. Dystrophinopathy in two young boys with exercise-induced cramps and myoglobinuria. Eur J Pediatr 1993;152:848–851.
12. Melis MA, Cau M, Muntoni F, et al. Elevation of serum creatine kinase as the only manifestation of an intragenic deletion of the dystrophin gene in three unrelated families. Eur J Paediatr Neurol 1998;2:255–261.
13. Muntoni F, Torelli S, Ferlini A. Dystrophin and mutations: one gene, several proteins, multiple phenotypes. Lancet Neurol 2003;2:731–740.
14. Ferlini A, Neri M, Gualandi F. The medical genetics of dystrophinopathies: molecular genetic diagnosis and its impact on clinical practice. Neuromuscul Disord 2013;23: 4–14.
15. Bladen CL, Salgado D, Monges S, et al. The TREAT-NMD DMD global database: analysis of more than 7,000 Duchenne muscular dystrophy mutations. Hum Mutat 2015;36:395–402.
16. Laing NG. Molecular genetics and genetic counselling for Duchenne/Becker muscular dystrophy. Mol Cel Biol Hum Dis Ser 1993;3:37–84.
17. Juan-Mateu J, Gonzalez-Quereda L, Rodriguez MJ, et al. DMD mutations in 576 dystrophinopathy families: a step forward in genotype-phenotype correlations. PLoS One 2015;10:e0135189.
18. Tuffery-Giraud S, Béroud C, Leturcq F, et al. Genotype-phenotype analysis in 2,405 patients with a dystrophinopathy using the UMD-DMD database: a model of nationwide knowledgebase. Hum Mutat 2009;30:934–945.
19. Jones HF, Bryen SJ, Waddell LB, et al. Importance of muscle biopsy to establish pathogenicity of DMD missense and splice variants. Neuromuscul Disord 2019;29: 913–919.
20. Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. Sci Transl Med 2017;9:eaal5209.
21. Waddell LB, Tran J, Zheng XF, et al. A study of FHL1, BAG3, MATR3, PTRF and TCAP in Australian muscular dystrophy patients. Neuromuscul Disord 2011;21: 776–781.
22. Cooper ST, Lo HP, North KN. Single section Western blot: improving the molecular diagnosis of the muscular dystrophies. Neurology 2003;61:93–97.
23. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 2012;9:671–675.
24. Bryen SJ, Ewans L, Pinner J, et al. Recurrent TTN metatranscript-only c.39974-11T>G splice variant associated with autosomal recessive arthrogryposis multiplex congenita and myopathy. Hum Mutat 2020;41:403–411.
25. Bryen SJ, Joshi H, Evesson FJ, et al. Pathogenic abnormal splicing due to intronic deletions that induce biophysical space constraint for spliceosome assembly. Am J Hum Genet 2019;105:573–587.

26. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580–585.

27. Wimmer K, Schamschula E, Wernstedt A, et al. AG-exclusion zones revisited: lessons to learn from 91 intronic NF1 3′ splice site mutations outside the canonical AG-dinucleotides. Hum Mutat 2020;41:1145–1156.

28. Kim J, Hu C, El Achkar CM, et al. Patient-customized oligonucleotide therapy for a rare genetic disease. N Engl J Med 2019;381:1644–1652.

29. Muntoni F. Is a muscle biopsy in Duchenne dystrophy really necessary? Neurology 2001;57:574–575.

30. Hoffman EP, Kunkel LM, Angelini C, Clarke A, Johnson M, Harris JB. Improved diagnosis of Becker muscular dystrophy by dystrophin testing. Neurology 1989;39:1011–1017.

31. Van Den Bergen JC, Wokke BH, Janson AA, et al. Dystrophin levels and clinical severity in Becker muscular dystrophy patients. J Neurol Neurosurg Psychiatry 2014; 85:747–753.

Two novel B9D1 variants causing Joubert syndrome: Utility of mRNA and splicing studies

Disha Katiyar [a,b], Neil Anderson [c], Shobhana Bommireddipalli [d], Adam Bournazos [d], Sandra Cooper [d], Himanshu Goel [b,e,*]

[a] University of New England, Armidale, NSW, 2351, Australia
[b] University of Newcastle, Callaghan, NSW, 2308, Australia
[c] John Hunter Children's Hospital, Hunter New England Local Health District (HNELHD), New Lambton Heights, NSW, Australia
[d] Kids Neuroscience Centre, The Children's Hospital at Westmead Sydney, Australia
[e] Hunter Genetics, Hunter New England Local Health District (HNELHD), Waratah, NSW, 2298, Australia

ARTICLE INFO

ABSTRACT

The primary cilium is an organelle which plays an important role in the transduction of signals in the Wnt and Sonic hedgehog pathways. Abnormal or absent primary cilia result in various neurodevelopmental, retinal, renal, hepatic and musculoskeletal abnormalities. Joubert syndrome (JS) is a ciliopathy with a prevalence estimated to be between 1:80 000 and 1:100 000. JS occurs due to bi-allelic mutations in one of the 34 identified genes, all of which encode for protein components of the primary cilia. The presentation of JS is highly variable, however a clinical diagnosis can be established by the presence of the molar tooth sign on axial brain MRI, hypotonia in infancy, and developmental delay. JS is less severe than Meckel syndrome (MKS), which is another recessive, and often lethal, ciliopathy. This report outlines an interesting case of JS, in which two novel mutations in *B9D1* were identified. This gene is not commonly associated with JS, and is often implicated in MKS. Functional mRNA study was helpful in delineating the pathogenic role of novel variants in this case.

1. Introduction

Joubert syndrome (JS) is a ciliopathy with variable clinical presentations; due to this phenotypic heterogeneity, the prevalence of JS is difficult to ascertain and has been estimated in literature to range between 1:80 000 and 1:100 000 (Romani et al., 2013). As is the case for other ciliopathies, JS is caused by mutations in genes that are responsible for the development of the primary cilia. Variants in 34 genes have been implicated in JS – 33 are inherited in an autosomal recessive manner and one is X-linked (M. Parisi and Glass, 1993).

The primary cilium is a nonmotile, microtubule-based organelle which protrudes from the surface of most human cells (Bialas et al., 2009). Primary cilia play a role in sensory processes and in the transduction of signals for the Wnt, Sonic hedgehog and other important cellular signalling pathways (Bialas et al., 2009).

A diagnosis of JS can be established by the presence of the following: the molar tooth sign on an axial brain MRI (Fig. 1); intellectual impairment or developmental delay; and hypotonia in infancy (M. A. Parisi, 2009). Other supportive findings include: abnormal eye

movements such as ocular motor apraxia; cerebellar signs such as nystagmus and ataxia; tachypnoea, apnoea or both in infancy (Romani et al., 2013). The phenotypic presentation and organ involvement of JS is highly variable: polydactyly is seen in 10–15%; facial features may be normal or dysmorphic; retinal, renal or hepatic defects may present at any stage (Romani et al., 2013). We present an adult female with mild intellectual disability, abnormal eye movements and molar tooth sign on MRI brain. She had two biallelic novel variants in *B9D1*.

2. Case report

The proband is a 24-year-old female with a clinical diagnosis of JS and a brain MRI demonstrating the molar tooth sign. As a child she displayed global delay involving late walking and a poor sense of balance, and required educational support through primary and high school. On examination, cerebellar signs are present with a head tilt and rotational nystagmus. She also had dysarthria and hypometric saccades.

The proband's siblings and parents do not have any clinical features of JS. In order to establish a molecular diagnosis, clinical exome
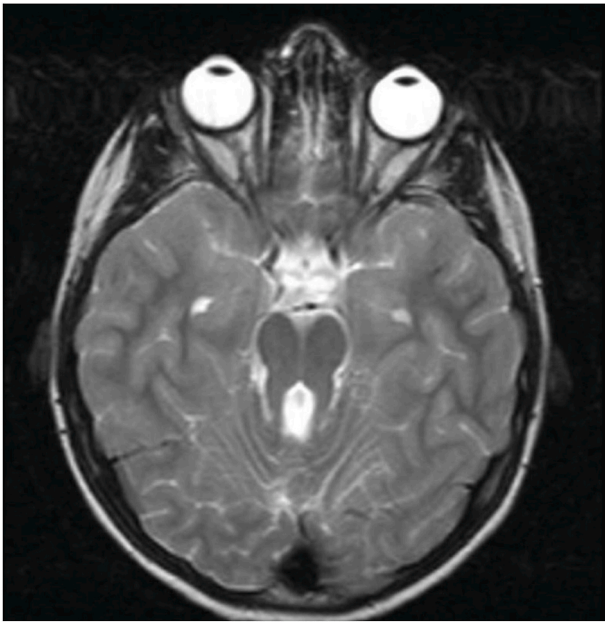
Fig. 1. The pathognomic molar tooth sign of Joubert syndrome, seen on axial brain MRI.

sequencing was performed and revealed bi-allelic variants in *B9D1*: NM_015681.3(B9D1):c.341G > A p.(Arg114Gln) and c.529G > C p.(Asp177His). The exome was negative for pathogenic or likely pathogenic variants in other JS genes, specifically *TMEM231* and *CC2D2A*. Digenic and triallelic inheritance has been suspected in JS; however, it has not been proven without any doubt (Kroes et al., 2016).

There are seven individuals heterozygous for the B9D1 c.341G > A, p.(Arg114Gln) variant in the Genome Aggregation Database (gnomAD, n > 120 000 exomes and >15 000 genomes). No homozygotes were observed in the dataset. PolyPhen, SIFT, and MutationTaster algorithms give conflicting in silico predictions; only MutationTaster predicts the variant as deleterious. Arginine at amino acid position 114 is moderately conserved. However, the nucleotide change affects the last nucleotide of an exon, and all four splicing prediction algorithms of the Alamut Visual software (SSF, MaxEntScan, NNSPLICE and GeneSplicer) predict the variant to weaken or abolish the natural splice donor site and thus likely lead to aberrant splicing, and to an abnormal or absent protein. To our knowledge, the variant has not been published in the relevant medical literature or reported in the disease-related variation databases such as ClinVar or HGMD.

The B9D1 c.529G > C, p.(Asp177His) variant has not been observed in the large reference population cohorts of the Genome Aggregation Database (gnomAD). The variant is predicted damaging by all three in silico tools used (PolyPhen, SIFT, MutationTaster). Aspartic acid at amino acid position 177 was highly conserved among approximately 100 vertebrates. To our knowledge, the variant has not been published in the relevant medical literature or reported in the disease-related variation databases such as ClinVar or HGMD. Both these variants have now been entered in DECIPHER database (HUN413143).

Given the clinical diagnosis of JS, RNA study was performed to confirm that these variants were affecting the functionality of the gene. This showed that these previously unknown variants were responsible for this clinical presentation.

3. Molecular methodology

Splicing studies involved performing RT-PCR on mRNA extracted from whole blood of the proband. Normal mRNA transcripts were not produced by the c.341G > A missense variant. This variant ablated the 5′-splice site of *B9D1* intron four, resulting in out-of-frame skipping of exon four (r.245_341del) (Fig. 2). This produced a frameshift mutation which caused a premature stop codon. Given the important role of exon four of *B9D1* in both brain and blood tissue, it was concluded that findings from the proband's blood sample can be extrapolated to the brain.

Normal splicing of *B9D1* transcripts were still present in the proband's sample. Sanger sequencing of RT-PCR amplicons confirmed that all the normal splicing transcripts were derived from the c.529G > C variant. While the splicing is normal, this deleterious missense mutation in exon seven is possibly damaging; furthermore, the c.529G > C variant was noted to affect an alternatively spliced exon of *B9D1* present in 10–30% of brain transcripts.

4. Discussion

There are 34 genes implicated in the pathogenesis of JS, and they all code for components of the primary cilium; mutations in *B9D1* (Chr17p11.2) are not often associated with JS (Bachmann-Gagescu et al., 2015; Hopp et al., 2011; M. Parisi and Glass, 1993); Table 1 outlines the clinical phenotypes of some *B9D1* variants. The majority of the cases of JS studied by Bachmann-Gagescu et al. (2015) involved mutations in C5ORF42, CC2D2A, CEP290, AHI1 and TMEM67 genes. B9 domain-containing protein 1 is an important protein present in the transition zone of the primary cilium. This is the region where the primary cilium joins the basal body, which anchors it to the plasma membrane of the cell (Chih et al., 2012; Romani et al., 2013). Disruption of this complex, which also involves proteins encoded by the *TMEM231* and *CC2D2A* genes, results in decreased cilia number and loss of signalling receptors (Chih et al., 2012).

The molar tooth sign on axial sections of a brain MRI is pathognomic of JS; it is seen due to cerebellar vermis hypoplasia, rotation and elongation of the cerebellar peduncles, and a deep interpeduncular fossa (Romani et al., 2013). These cerebellar and brainstem abnormalities can be explained by defective signalling pathways, such as the Sonic hedgehog pathway, which are dependent on functional primary cilia; this also explains the malformations seen in another ciliopathy called Meckel syndrome (MKS) (Chih et al., 2012). MKS is a rare and lethal autosomal recessive ciliopathy, characterised by severe renal and central nervous system malformations (Chih et al., 2012; Dowdle et al., 2011). Thirteen genes overlap between JS and MKS, including *B9D1* (Romani et al., 2014). In mice models studied by Dowdle et al. (2011), it was found that "loss of *B9D1* resulted in MKS-like phenotypes". The phenotype seen in MKS has more severe malformations than in JS, such as sinus inversus, cleft lip and palate, and skeletal defects; patients classically present with polydactyly, renal cysts, hepatic malformations and encephaloceles (Romani et al., 2014).

The severity of an autosomal recessive disorder is influenced by the less severe variant in the affected gene (Furu et al., 2003). Autosomal recessive polycystic kidney disease (ARPKD; OMIM 263200) is a severe hereditary form of polycystic disease affecting the kidneys and biliary tract, with a widely variable clinical spectrum. Furu et al. (2003) found the presence of two chain-terminating mutations invariably resulted in perinatal lethality. However, missense hypomorphic variants were more commonly seen in milder presentation of ARPKD. In our patient c.341G > A is a loss of function variant. A loss-of-function variant such as this would cause a more severe phenotype, similar to those seen in MKS (Dowdle et al., 2011); however, c.529G > C is a missense variant and expected to be a milder variant. The exon seven containing the c.529G > C, p.(Asp177His) variant is alternatively spliced (in ~10–30 of transcripts in brain). Exon seven has a likely pathogenic missense variant reported in ClinVar; this is a previously reported 3-bp in-frame deletion (c.520_522delGTG, NG_031885.1) resulting in the deletion of a conserved amino acid residue Valine, that was found in compound heterozygous state in a patient with Joubert syndrome (Romani et al., 2014).
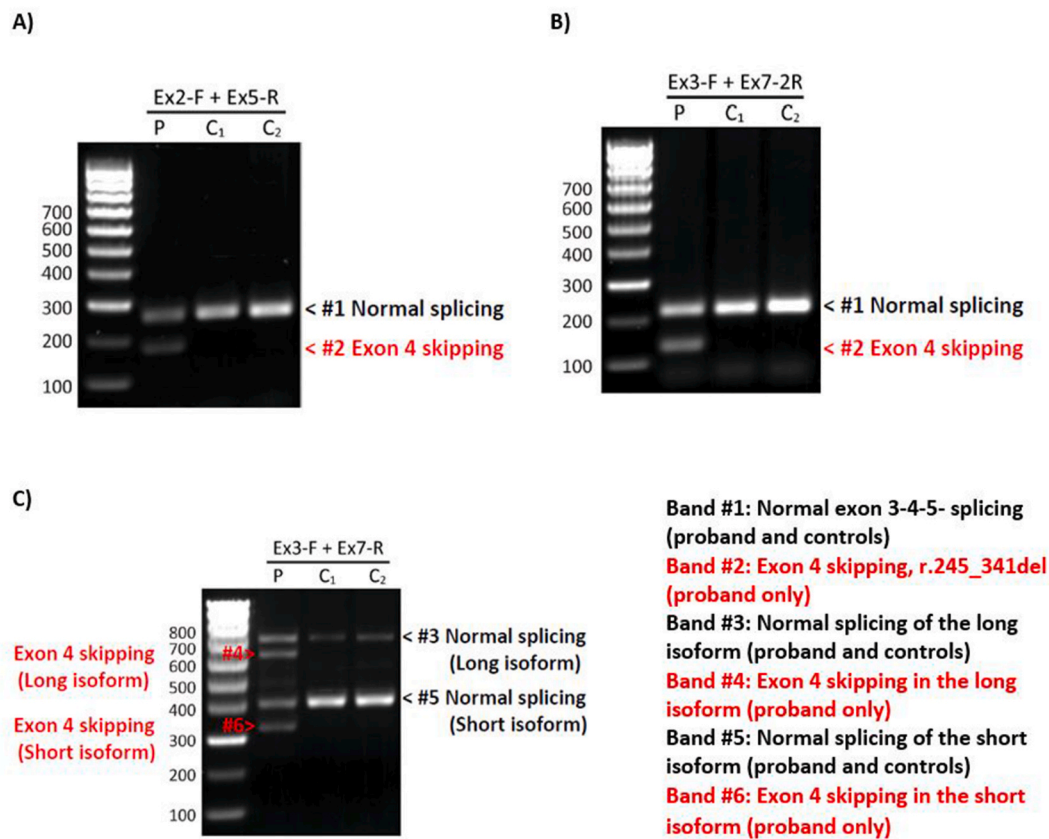
Fig. 2. RT-PCR of *B9D1* mRNA: A) and B) detected two bands when using primers flanking the c.324G > A variant (#1 normal splicing; #2 exon four skipping). C) detected four bands using primers to amplify over both variants (#3 normal splicing long isoform, #4 exon four skipping long isoform, #5 normal splicing short isoform, #6 exon four skipping short isoform). Bands #2, #4 and #6 were only present in the proband's sample.

Table 1
Summary of published *B9D1* variants and clinical phenotype.

Gene	Allele 1	Allele 2	Clinical phenotype	Variant	Reference
B9D1	c.505+2T > C	1.713 Mb deletion	MKS	*CEP290* heterozygous variant	Hopp et al. (2011)
	c.467G > A, p.Arg156Gln	c.467G > A, p.Arg156Gln	JS	–	Romani et al. (2014)
	c.95A > G, p.Tyr32Cys	c.520_22delGTG, p.Val174del		–	
	c.151T > C, p.Ser51Pro	NM_001321214.2:c.510G > C		*CC2D2A* heterozygous variant	Kroes et al. (2016)

5. Conclusion

This report outlined the case of a 24-year-old female with a relatively mild presentation of JS, as diagnosed by a positive MRI showing the molar tooth sign, global delay, ataxia, nystagmus, and abnormal eye movements. Molecular diagnosis confirmed two novel variants in *B9D1*, which is implicated in both JS and MKS. The more severely mutated allele caused a loss-of-function mutation at the splice site for exon four, and activation of the NMD pathway, which would have resulted in a clinical presentation similar to those seen in MKS; however, as is the case in recessive conditions, the phenotypic presentation is dependent on the less severe variant, which in this case cause a missense mutation and amino acid substitution in exon seven. This substitution occurred in a highly conserved region of the gene which is alternately spliced in 10–30% of brain transcripts, and was therefore significant enough to result in the presentation of this autosomal recessive ciliopathy. mRNA studies and splice site analysis were used to confirm and explain the role of these novel mutations in this case.

CRediT authorship contribution statement

Disha Katiyar: Conceptualization, Visualization, Writing - original draft. **Neil Anderson:** Writing - original draft. **Shobhana Bommireddipalli:** Methodology, Formal analysis, Investigation. **Adam Bournazos:** Methodology, Resources, Formal analysis, Investigation. **Sandra Cooper:** Methodology, Resources, Formal analysis, Investigation. **Himanshu Goel:** Conceptualization, Supervision, Visualization, Writing - original draft.

Acknowledgements

References

Bachmann-Gagescu, R., Dempsey, J.C., Phelps, I.G., O'Roak, B.J., Knutzen, D.M., Rue, T. C., Doherty, D., 2015. Joubert syndrome: a model for untangling recessive disorders with extreme genetic heterogeneity. J. Med. Genet. 52 (8), 514–522. https://doi. org/10.1136/jmedgenet-2015-103087.

Bialas, N.J., Inglis, P.N., Li, C., Robinson, J.F., Parker, J.D.K., Healey, M.P., Leroux, M.R., 2009. Functional interactions between the ciliopathy-associated Meckel syndrome 1 (MKS1) protein and two novel MKS1-related (MKSR) proteins. J. Cell Sci. 122 (5), 611. https://doi.org/10.1242/jcs.028621.

Chih, B., Liu, P., Chinn, Y., Chalouni, C., Komuves, L.G., Hass, P.E., Peterson, A.S., 2012. A ciliopathy complex at the transition zone protects the cilia as a privileged membrane domain. Nat. Cell Biol. 14 (1), 61–72. https://doi.org/10.1038/ncb2410.

Dowdle, William E., Robinson, Jon F., Kneist, A., Sirerol-Piquer, M.S., Frints, Suzanna G. M., Corbit, Kevin C., Reiter, Jeremy F., 2011. Disruption of a ciliary B9 protein complex causes Meckel syndrome. Am. J. Hum. Genet. 89 (1), 94–110. https://doi.org/10.1016/j.ajhg.2011.06.003.

Furu, L., Onuchic, L.F., Gharavi, A., Hou, X., Esquivel, E.L., Nagasawa, Y., Somlo, S., 2003. Milder presentation of recessive polycystic kidney disease requires presence of amino acid substitution mutations. J. Am. Soc. Nephrol. 14 (8) https://doi.org/10.1097/01.ASN.0000078805.87038.05, 2004.

Hopp, K., Heyer, C.M., Hommerding, C.J., Henke, S.A., Sundsbak, J.L., Patel, S., Harris, P.C., 2011. B9D1 is revealed as a novel Meckel syndrome (MKS) gene by targeted exon-enriched next-generation sequencing and deletion analysis. Hum. Mol. Genet. 20 (13), 2524–2534. https://doi.org/10.1093/hmg/ddr151.

Kroes, H.Y., Monroe, G.R., van der Zwaag, B., Duran, K.J., de Kovel, C.G., van Roosmalen, M.J., van Haaften, G., 2016. Joubert syndrome: genotyping a Northern European patient cohort. Eur. J. Hum. Genet. : EJHG (Eur. J. Hum. Genet.) 24 (2), 214–220. https://doi.org/10.1038/ejhg.2015.84.

Parisi, M., Glass, I., 1993. Joubert syndrome. In: Adam, M.P., Ardinger, H.H., Pagon, R. A., Wallace, S.E., Bean, L.J.H., Stephens, K., Amemiya, A. (Eds.), GeneReviews((R)). Seattle (WA).

Parisi, M.A., 2009. Clinical and molecular features of Joubert syndrome and related disorders. Am. J. Med. Genet. Part C, Seminars in medical genetics 151C (4), 326–340. https://doi.org/10.1002/ajmg.c.30229.

Romani, M., Micalizzi, A., Kraoua, I., Dotti, M.T., Cavallin, M., Sztriha, L., Valente, E.M., 2014. Mutations in B9D1 and MKS1 cause mild Joubert syndrome: expanding the genetic overlap with the lethal ciliopathy Meckel syndrome. Orphanet J. Rare Dis. 9 (1), 72. https://doi.org/10.1186/1750-1172-9-72.

Romani, M., Micalizzi, A., Valente, E.M., 2013. Joubert syndrome: congenital cerebellar ataxia with the molar tooth. Lancet Neurol. 12 (9), 894–905. https://doi.org/10.1016/S1474-4422(13)70136-4.