

Translating transcriptomics analysis into diagnostic workflows: clinical variant identification and interpretation in hypothesis-driven and hypothesis-free approaches



Chingyiu Pang,^a Martin Man-Chun Chui,^a Wenshu Tang,^b Anna Ka-Yee Kwong,^a Hiu Yu Cherie Leung,^a Sze-Shing Fan,^a Alice Wing-Sze Kwok,^a Godfrey Chi-Fung Chan,^{a,c} Ho Ming Luk,^d Rosanna Ming-Sum Wong,^c Wanling Yang,^a Ivan Fai-Man Lo,^d Cheuk-Wing Fung,^c Joanna Yuet-Ling Tung,^c Anthony Pak-Yin Liu,^{a,c} Kit-San Yeung,^c Sheila Suet-Na Wong,^c Christopher Chun-Yu Mak,^{a,e} and Brian Hon-Yin Chung^{a,b,c,e,*}



^aDepartment of Paediatrics and Adolescent Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

^bHong Kong Genome Institute, Hong Kong SAR, China

^cDepartment of Paediatrics and Adolescent Medicine, Hong Kong Children's Hospital, Hong Kong SAR, China

^dDepartment of Clinical Genetics, Hong Kong Children's Hospital, Hong Kong SAR, China

Summary

Background Despite significant advancements in genetic diagnosis, there are still bottlenecks in DNA-level testing. Challenges to diagnosis include the inability to identify the causal variant, and the lack of functional evidence leading to an accumulation of variants of uncertain significance (VUS). Recently, there has been growing evidence demonstrating the diagnostic value of RNA sequencing (RNA-seq).

Methods This diagnostic study implemented RNA-seq analysis of blood (and fibroblasts, if available) in 102 patients with genetically undiagnosed diseases. An outlier analysis for gene expression and splicing was adopted through a multi-modal machine learning algorithm (i.e., Detection of RNA Outliers Pipeline—DROP).

Findings Our analysis aided the interpretation of 22/102 (21.6%) patients through both hypothesis-driven (i.e., with a prior genetic candidate; $n = 10$) and hypothesis-free (i.e., without a prior genetic candidate; $n = 12$) approaches. Not only did this workflow aid genetic diagnosis ($n = 12$), but it also provided additional information to known findings ($n = 4$) and guided the discovery of unestablished disease mechanisms ($n = 6$).

Interpretation This study demonstrated the clinical and scientific value of blood transcriptomics through both hypothesis-driven and hypothesis-free approaches. We have proposed an initial framework for RNA-seq implementation into the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines using PVS1 (null variants) and PP4 (phenotypic specificity). However, other considerations are required, including further clarifications for thresholds, and detailed guidance on the incorporation of different aspects of RNA-seq results (e.g., degree of nonsense-mediated decay and completeness of splicing).

Funding This study was supported by grants from the Society for the Relief of Disabled Children, the Health and Medical Research Fund (HMRF) and Commissioned Paediatric Research at HKCH under HMRF both by the Health Bureau, The Government of the Hong Kong Special Administrative Region.

Copyright © 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Transcriptomics; Hypothesis-driven; Hypothesis-free; Outlier approach; ACMG variant curation; Genetic diagnosis

Introduction

Rare diseases, individually affecting fewer than one in 2000 people by definition, have a significant burden

worldwide.^{1,2} Despite individual rarity, rare diseases consist of more than 7000 types and collectively affect more than 400 million individuals globally.² In Hong

*Corresponding author. Room 115, 1/F, New Clinical Building, Queen Mary Hospital, Hong Kong SAR, China.

E-mail address: bhychung@hku.hk (B.H.-Y. Chung).

^cCo-last authors.

Research in context

Evidence before this study

Current sequencing technologies, including whole exome and genome sequencing, have shown a diagnostic yield capped at roughly 40%. To tackle this bottleneck, cumulative evidence has demonstrated the diagnostic value of incorporating transcriptomics into the genetic diagnostic workflow. However, a consensus has yet to be reached on when transcriptomics should be incorporated and how different aspects of the results can be considered during the variant curation process. A PubMed search was performed, using the keywords '(RNA-seq OR transcriptomics) AND (ES OR GS) AND (diagnosis OR adult OR child OR infant OR neonatal) NOT (carrier OR animal) in Titles and Abstracts' for publications dated before September 2025.

Added value of this study

Our study aided the interpretation of genetically undiagnosed cases through implementing blood (and fibroblast) RNA-seq analysis using an outlier approach. Our findings support and reinforce the diagnostic value of RNA-seq in clinical genetics aligning with previous studies. Furthermore, we have demonstrated that blood RNA-seq is beneficial in both hypothesis-driven and hypothesis-free approaches, addressing a heated debate in the field. Through

this work, we have aided not only diagnosis with currently known gene-disease associations, but also the discovery of unestablished disease mechanisms and the downstream effects of certain variants (e.g., RNA signatures). We have also proposed an initial framework for the application of RNA-seq results into the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines, where we identified areas that require further guidance.

Implications of all the available evidence

Given the growing clinical focus on genetic diseases and rapid advances in technologies, there is an increasing demand to tackle challenging interpretations of variants that remain unresolved at the DNA-level. Our study highlights the diagnostic value of transcriptomics and sheds light on aspects of its clinical implementation. Its applicability in a heterogeneous cohort, regardless of the presence of a prior candidate variant, supported the utility of RNA-seq in genetically unresolved cases. An initial framework has also been proposed to incorporate RNA-seq into the variant curation workflow, in which our study further highlighted considerations that are required for a more comprehensive evaluation of the results.

Kong, one in every 67 people is diagnosed with a rare disease.³ These patients often suffer from a long diagnostic odyssey owing to individually low disease prevalence.¹ With an annual socio-economic cost of HK\$484,253 (US\$62,084) per patient on average in the Hong Kong population, the total cost and catastrophic health expenditure incidence rates increase with the duration of the diagnostic odyssey.⁴ Simultaneously, patients and caretakers shoulder a significantly lower health-related quality of life compared to counterparts with chronic diseases.⁵ These factors demonstrate the importance of tackling current hurdles in rare disease diagnosis in order to provide the most timely clinical management for patients.

A majority of rare diseases (80%) are known to have an underlying genetic cause with 70% presenting clinically in childhood,² and ongoing efforts have been made to diagnose rare disease and advance genomic medicine. Clinical integration of genetic tests, exome reanalysis, newborn screening etc., have been eagerly proposed and performed in different countries. Large-scale genome projects have also been established worldwide. Despite advancements in sequencing technologies, however, whole exome sequencing (WES) only yields an overall diagnostic rate of 38%.⁶ According to our experiences, the diagnostic yield of WES ranges from 10 to 65% among different paediatric cohorts with a predominant Chinese ethnicity.^{7–15} Surprisingly, whole genome sequencing (WGS) was found to have a

similar diagnostic yield as WES, despite higher clinical utility.⁶ Regardless of the sequencing technology, a large proportion of patients remain genetically undiagnosed owing to challenges in (1) causal variant identification due to the intrinsic property of WES selecting for coding regions, and limitations of current analytic protocols in prioritising challenging variants like non-coding variants, and (2) variant interpretation of likely causal variants that are identified but remain classified as variants of uncertain significance (VUS) due to the lack of functional evidence. This has led to an accumulation of VUS, which has been reported in 21–29% of cases remaining genetically undiagnosed after WES/WGS.^{6,16} This phenomenon is exacerbated in non-European populations owing to the limited data and studies available.¹⁶

While the first difficulty has been improved by multiple variant prioritisation algorithms like Exomiser¹⁷ and Genomiser,¹⁸ the issue with accumulating VUS remains a global challenge. RNA sequencing (RNA-seq) has emerged as a complementary diagnostic tool for DNA-based analysis by providing additional functional evidence. According to the latest Simple ClinVar data available (as of July 14, 2021),¹⁹ one-fourth of ClinVar VUS has potential RNA phenotypes. Thus, RNA-seq can guide the identification of missed causal variants from previous tests, and/or support the reclassification of VUS according to the American College of Medical Genetics and Genomics and the

Association for Molecular Pathology (ACMG/AMP) guidelines.^{20,21} Several studies have shown the diagnostic utility of RNA-seq with an incremental diagnostic yield of 2.6–64.2% for unresolved ES/GS cases.^{22–38} The significant variability among these studies can be attributed to the differences in disease cohort, tissue samples used, analysis approach and recruitment criteria (Supplementary Table S1). RNA-seq analysis for rare disease diagnosis are typically performed using an ‘outlier’ approach, differing from the conventional differential expression approaches for case–control studies. In 2021, Yepez et al.³⁹ proposed a multi-modular computational workflow–Detection of RNA Outliers Pipeline (DROP)–which investigates multiple RNA phenotypes including aberrant expression (AE) using OUTRIDER,⁴⁰ and aberrant splicing (AS) using FRASER/FRASER2.^{41,42}

Owing to the intrinsic differences between genomics and transcriptomics, the clinical implementation of RNA-seq is met with challenges in accessing clinically relevant tissues for analysis. As one of the factors for the effectiveness of RNA-seq, efforts have been made in the field to identify favourable clinically accessible tissues (CATs) as functional proxies for clinically relevant but inaccessible tissues like the brain. Fibroblasts and blood are common CATs being considered owing to their accessibility. Despite fibroblasts having a higher coverage of disease-causing genes,^{26,35,43} blood has demonstrated a coverage of 56.8–86.7% in different disease-specific gene panels⁴³ while being the least invasive CAT. Blood RNA-seq alone has achieved an 8.5% incremental diagnostic yield,^{24,28,29,36,38,44} (Supplementary Table S1) showing that blood can serve as a readily available CAT with a promising diagnostic utility.

Growing interest lies in the clinical application of transcriptomics in the genetic diagnostic workflow. Recently, Zhao S. et al.⁴³ suggested a clinical validation framework throughout the entire RNA-seq workflow for benchmarking between different laboratories. However, the direct application of RNA-seq results into the variant curation pipeline has yet to be investigated, apart from the PVS1 decision tree proposed by the ClinGen splice variant working group.⁴⁵ In our study, we recruited 102 patients that remained genetically undiagnosed after routine genetic testing and/or WES. RNA was extracted from blood (and fibroblast tissues; $n = 11$) and subjected to RNA-seq analysis using the DROP pipeline to detect aberrant RNA phenotypes in expression and splicing. By using hypothesis-driven (i.e., with a prior genetic candidate) and hypothesis-free (i.e., without a prior genetic candidate) approaches, this study aims to shed light on the value of RNA-seq in genetic diagnosis, aspects of its clinical implementation and more importantly, considerations required when incorporating transcriptomics interpretation into the current ACMG/AMP guidelines.

Methods

Patient recruitment

Patients with congenital disorders who remained genetically undiagnosed after routine genetic testing/WES ($n = 102$) were recruited from two hospitals affiliated with The University of Hong Kong (HKU): Queen Mary Hospital and Hong Kong Children’s Hospital during the period of 2021–2024. Clinical information including sex, age and clinical phenotypes was retrieved from patients’ clinical records. Our study recruitment ensured inclusivity and sufficient sample size, resulting in a heterogeneous cohort. The sample sizes recommended for the applied analytical algorithms are a minimum of 50–60 and 30 samples for OUTRIDER⁴⁰ and FRASER2.0,⁴¹ respectively, to ensure sufficient power.³⁹

Ethics

The study was approved by the institutional review boards of the University of Hong Kong/Hospital Authority Hong Kong West Cluster and Hong Kong Children’s Hospital (UW 12–211 and HKCH-REC-2020-035) and conformed to the Declaration of Helsinki. Written informed consent was obtained from all participants and/or their guardians.

Sample preparation and RNA-seq

Peripheral blood collected from patients was treated with Red Blood Cell Lysis Solution (Qiagen). Parental samples were also collected if possible after clinical evaluation. Among the 102 patients, fibroblast samples were also available from 11 participants. Fibroblasts were cultured into a monolayer at 80% confluency in high-glucose Dulbecco’s Modified Eagle Medium (DMEM; Gibco) supplemented with 10% foetal bovine serum and 1% penicillin/streptomycin (Thermo Fisher Scientific).

The subsequent RNA extraction and RNA-seq processes were similar to those in our previous studies.^{46,47} In brief, blood and fibroblast samples were subjected to further lysis and RNA extraction using TRIzol™ Reagent (Thermo Fisher Scientific). Only extracted RNA with RNA integrity (RIN) ≥ 7 was used for RNA-seq at the Centre for PanorOmic Sciences (CPOS) of HKU. Library preparation through poly-A-tail capture was performed using the KAPA mRNA HyperPrep Kit (Roche Diagnostics) according to the manufacturer’s protocol. RNA-seq was performed in paired-end runs (151 bp) on an Illumina NovaSeq 6000 platform (RRID: SCR_016387) with a sequencing depth of at least 100 million reads to achieve higher sensitivity for outlier detection, while balancing the associated decrease in specificity.^{39,40,42,43} Only samples with 90% of sequenced bases reaching a Phred Score of 30 (99% base accuracy) were included.

RNA-seq data processing and analysis

RNA-seq data were subjected to data cleaning by AfterQC (v0.9.6; RRID: SCR_016390) before being

mapped to the GRCh37 human reference genome, and output as BAM files using STAR (v2.5.2a; RRID: [SCR_004463](#)) in two-pass mode. The sequencing data from different tissues were processed independently using the computational module workflow DROP. It is a protocol that automates RNA-seq analysis using an outlier approach from raw input files, allowing a standardised workflow with multiple modules.³⁹ In our study, DROP v1.3.3 was utilised to detect AE outliers using OUTRIDER,⁴⁰ and AS outliers using FRASER2.0 (v1.99.3).⁴¹ Denoising autoencoders to control for confounders not known *a priori* were integrated in the modules, including but not limited to technical confounders like sequencing batches and biological confounders like sex and age. This outlier approach is substantially different from a conventional differential expression analysis, where we compare expression levels between cases and controls. Here, we compare individual target samples with the rest of the cohort, on the basis of the rarity of two unrelated individuals having the same disease-causing variant leading to identical RNA phenotypes.³⁹ As suggested by the DROP developers,³⁹ publicly available RNA-seq datasets from The Genotype-Tissue Expression (GTEx) project (dbGaP accession number phs000424.v8.p2; RRID: [SCR_013042](#))⁴⁸ were included as controls to increase statistical power (sensitivity) by increasing sample size along with our internal controls; the details of GTEx controls are provided in [Supplementary Table S2](#).

To increase the sensitivity of detecting expression outliers, we defined significant AE outliers at a false discovery rate (FDR) ≤ 0.1 , which is slightly looser than the default threshold (FDR ≤ 0.05) used in other studies.^{26,35,39} Significant AS outliers were filtered based on the default thresholds (FDR ≤ 0.1 and effect size $|\Delta J| \geq 0.1$).⁴¹ $|\Delta J|$ (J = Intron Jaccard Index) represents the absolute value of differences between the expected (modelled by an autoencoder under the hypothesis that all individuals in the cohort have the same splicing specificity and efficiency), and the observed splicing specificity and efficiency of a particular splice site of an intron.⁴¹ The FDR for p-values derived from both modules (AE: negative-binomial distribution; AS: beta-binomial distribution) were generated using the Benjamini-Yekutieli method to correct for multiple testing.^{40–42} In this study, p-value correction was performed for 16,946 genes per sample (OUTRIDER) and 206,605 junctions per sample (FRASER2.0). Manual inspection of sequencing reads using the Integrative Genomics Viewer (IGV; RRID: [SCR_011793](#)) was performed in cases with candidate VUS and significant aberrant RNA outliers. In perfect scenarios, fold changes of 0.5 and 0 from AE, and $|\Delta J|$ of 0.5 and 1 from AS would indicate heterozygous and homozygous events respectively. However, these values are highly subjected to variability owing to transcriptomics processes being dynamic and affected by different factors

such as nonsense-mediated decay (NMD) efficiency, and therefore their interpretation was made along with manual inspection of IGV in case-specific scenarios.

Aberrant RNA outliers from both modules with candidate VUS and/or reported gene–disease associations in the Online Mendelian Inheritance in Man (OMIM; RRID: [SCR_006437](#)) database,⁴⁹ specifically those that phenotypically matched with patient phenotypes, were prioritised for manual inspection. Potential candidate non-morbid genes or aberrant results (e.g., large numbers of outliers) also prompted deeper investigation if highly suspicious, especially those that might indicate the presence of an RNA signature (i.e., a specific or non-specific group of aberrant events caused by the defective gene). Causative variants were searched for using existing DNA/RNA-seq data. When causative variants remained unidentifiable in specific cases, short-read WGS (srWGS) and long-read WGS (lrWGS) were performed according to an optimised protocol as previously described.⁵⁰ In brief, for srWGS, following genomic DNA extraction, polymerase chain reaction (PCR)-free srWGS libraries were prepared using the KAPA HyperPlus Kit (Kapa Biosystems Inc.) and sequenced on an Illumina NovaSeq 6000 (Illumina Inc.; RRID: [SCR_016387](#)) or NovaSeq X Plus sequencer (Illumina Inc.; RRID: [SCR_024568](#)). For lrWGS, the processes were performed following the instructions of the manufacturer, Oxford Nanopore Technologies (ONT; RRID: [SCR_003756](#)). Using 1 μ g of extracted genomic DNA, libraries were prepared with the ONT Ligation Sequencing Kit (SQK-LSK114) with sequencing performed for up to 72 h on a PromethION P-24 device using an R10.4.1 flow cell (ONT, Oxford UK; RRID: [SCR_017987](#)). The generated srWGS and lrWGS data were then subjected to quality control, secondary analysis (including alignment to the GRCh38/hg38 human reference genome), and tertiary analysis. A wide range of variant types were investigated including single nucleotide variants, small insertion/deletion variants, copy number variants, large structural variants, short tandem repeat expansions, as well as methylation changes (only for lrWGS).

Manual curation of identified variants was then performed according to the ACMG/AMP guidelines.^{20,21,45,51} Sanger sequencing or real-time PCR was done in cases where variant confirmation and segregation analysis were required. The DNA and RNA datasets generated will not be made publicly available due to ethical restrictions regarding minor participants and patient anonymity. Results were returned to clinicians for further clinical follow-up.

Statistics

In cases with an aberrant number of outliers, statistical significance was evaluated as follows: With the large variety of overexpressed genes attributed to the heterogeneity of our cohort, mean and standard deviation

of the number of outliers from the control dataset (unrelated samples) were compared with the patient data for the derivation of Z-scores. Student's T distribution with a degree of freedom of 122 was then used to derive two-tailed p-values from Z-scores. The p-value significance threshold was taken as 0.05.⁵²

Other statistical tests performed in DROP are in-built executions of the pipeline.^{39–42}

Role of the funding source

The funding sources are detailed in the Acknowledgement section. The funders had no role in the design and conduct of this study; collection, management, analysis and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. All authors had full access to the data in the study and accept responsibility for the decision to submit for publication.

Results

Cohort characteristics

A total of 102 probands were recruited for blood transcriptomics analysis with most cases analysed as singletons (48.0%) or trios (41.2%). While the current study did not identify explicit advantages of trio analysis, further studies are required to investigate the implementation of parental data to enhance the analysis. The mean age of the cohort was 9.6 years (standard deviation of 6.6), with a majority of probands aged 0–18 years old (89.2%). The remaining 11 cases were >18 years of age, having been recruited after being referred to the adult clinic with minimum follow-up on their genetic condition. Our cohort comprises a heterogeneous range of genetic conditions, with 57.8% of patients presenting with conditions associated with the central nervous system, followed by cardiac diseases and suspected mitochondrial diseases, among others. Further details of the cohort are provided in [Table 1](#), [Supplementary Information S1–S3](#) and [Supplementary Tables S3 and S4](#).

Hypothesis-driven blood RNA-seq analysis

In our cohort, hypothesis-driven analysis aided the interpretation of ten cases (9.8%), in which prior VUS candidates were identified from previous genetic tests. Among these cases, six achieved genetic diagnoses with RNA-seq providing further supporting evidence to upgrade the previously identified VUS to a likely pathogenic/pathogenic (LP/P) variant according to the ACMG/AMP guidelines ([Table 2](#); [Supplementary Table S3](#); [Supplementary Information S1](#)).^{20,45,51} All upgraded variants were intronic variants outside the $\pm 1,2$ dinucleotide canonical splice positions, where PVS1 criteria (i.e., “Null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss-of-function

Total	n = 102	
Sex		
Female	46	45.1%
Male	56	54.9%
Analysis		
Singleton	49	48.0%
Duos	10	9.8%
Trio	42	41.2%
Quadro	1	1.0%
Age		
Mean (SD)	9.6 years (6.6)	
≤18 years old	91	89.2%
>18 years old	11	10.8%
Disease categories		
Central nervous system ^a	59	57.8%
Suspected mitochondrial diseases	8	7.8%
Cardiac diseases	9	8.8%
Congenital anomalies	2	2.0%
Metabolic disorders	6	5.9%
Endocrine disorders	3	2.9%
Immunodysregulatory disorders	2	2.0%
Others ^b	13	12.7%

Sex-disaggregated demographics are provided in [Supplementary Table S4](#).

^aIncludes neurodevelopmental disorder, neurodegenerative disorder, epilepsy, neurotransmitter disease, movement disorder and neurological immune disorder. ^bConditions associated with other systems or those that cannot be classified under other disease categories.

Table 1: Cohort demographics.

(LoF) is a known mechanism of disease”)²⁰ or PS3 criteria under gene specific guidelines (i.e., case B29) can only be applied if we have functional data supporting pathogenicity through RNA-related findings.⁴⁵

In the remaining four cases, RNA-seq clarified uncertain findings from the DNA-level investigation ([Table 2](#); [Supplementary Table S3](#); [Supplementary Information S1](#)). In one case (B37), RNA-seq pinpointed candidate underlying genes within a large duplication region. Overexpression of 14 genes on chromosomal Xq28 was detected, aligning with the chromosomal microarray finding of *de novo* duplication. Four candidates were flagged to possibly explain the patient's phenotype: *RPL10*, *FAM50A*, *LAGE3* and specifically *GDI1* supported by previously reported patients with recurrent but variable 0.3 Mb copy-number gain at Xq28.⁵⁴ In three other cases (B12, B48,⁵³ B80), RNA-seq refined or validated *in silico* predictions, refined prior experimental results (B48⁵³) and suggested limited damaging effects of variants in two cases (B12, B80). This demonstrated that the diagnostic value of RNA-seq lies not only in upgrading variants but also in re-evaluating variants for potential downgrading or exclusion in subsequent analyses. This was exemplified with case B12, where the proband was experiencing gross motor delay, verbal expression delay, brain and genitalia abnormalities and multiple ocular phenotypes.

Cases with new diagnosis (n = 6)

Case No.	Indication and phenotypes	Previously identified VUS	Gene-disease association	RNA-seq finding	Original ACMG classification (before RNA-seq)	New ACMG classification (after RNA-seq)
B10 ⁵³ (Trio)	<u>Cardiac disease:</u> dilated cardiomyopathy with neutropenia and DD	Maternally inherited hemizygous TAFAZZIN variant NM_000116.5: c.284+5G>A	Barth Syndrome (OMIM# 302060)	AS: TAFAZZIN ($ \Delta J $ = 0.5; FDR = 0.0000079)	VUS (PM2_supporting, PP3)	Likely pathogenic (PVS1_strong (RNA), PM2_supporting, PP4_strong)
B11 ⁵³ (Singleton)	<u>Cardiac disease:</u> hypertrophic cardiomyopathy, cardiac structural anomalies, DD and facial features of Noonan Syndrome	Compound heterozygous LZTR1 variants: 1) Maternally inherited NM_006767.4: c.1261-3C>G and 2) Paternally inherited c.1943-256C>T	Noonan syndrome 2 (OMIM# 605275)	AS: LZTR1 ($ \Delta J $ = 0.63; FDR = 0.00095) LZTR1 ($ \Delta J $ = 0.38; FDR = 0.032)	c.1261-3C>G VUS (PM2_supporting, PM3, PP3) c.1943-256C>T Likely Pathogenic (PM2_supporting, PM3, PVS1 (RNA))	c.1261-3C>G Likely Pathogenic (PM2_supporting, PM3, PVS1_strong (RNA)) c.1943-256C>T Likely Pathogenic (PM2_supporting, PM3, PVS1 (RNA))
B29 (Trio)	<u>Neurodevelopmental disorder:</u> macrocephaly, history of global DD and dysmorphic features	<i>De novo</i> heterozygous PTEN variant NM000314.8: c.209+3A>T	Cowden syndrome 1 (OMIM# 158350)	AS: PTEN ($ \Delta J $ = 0.55; FDR = 2.7×10^{-11})	VUS (PM2_supporting, PS2_moderate, PP3)	Likely Pathogenic (PM2_supporting, PS2_moderate, PS3_strong)
B56 ^a (Singleton)	<u>Suspected mitochondrial disease:</u> multiple motor phenotypes, global DD, and epilepsy	Compound heterozygous RARS2 variants: 1) Maternally inherited NM_020320.5: c.1238-28T>G and 2) Paternally inherited c.685C>T p.(Arg229*)	Pontocerebellar hypoplasia type 6 (OMIM# 611523)	AS: RARS2 (FDR = 0.31; unadjusted p-value = 0.000016; $ \Delta J $ = 0.46)	c.1238-28T>G VUS (PM2_supporting, PM3, PP3) c.685C>T Pathogenic (PVS1, PM2_supporting, PM3)	c.1238-28T>G Likely Pathogenic (PVS1_strong(RNA), PM2_supporting, PM3) c.685C>T Likely Pathogenic (PVS1_strong(RNA), PM2_supporting, PM3)
B68 (Trio)	<u>Neurodevelopmental disorder:</u> bilateral hearing loss, hydronephrosis, spastic diplegia and history of global DD, ID	<i>De novo</i> hemizygous MSL3 variant NM_078629.4: c.749+5G>A	Basilicata-Akhtar syndrome (OMIM# 301032)	AS: MSL3 ($ \Delta J $ = 0.95; FDR = 7.9×10^{-24})	VUS (PP3, PS2_supporting, PM2_supporting)	Likely pathogenic (PVS1_strong(RNA), PS2_supporting, PM2_supporting)
B119 (Singleton)	<u>Neurodevelopmental disorder:</u> microcephaly, failure to thrive, global DD, ID, autism and bilateral hearing loss	<i>De novo</i> heterozygous NIPBL variant NM_133433.4:c.64+5G>C	Cornelia de Lange syndrome 1 (OMIM# 122470)	AS: NIPBL ($ \Delta J $ = 0.22; FDR = 0.0055)	VUS (PP3, PS2_supporting, PM2_supporting)	Likely Pathogenic (PVS1_moderate (RNA), PS2_strong, PM2_supporting, PP4_Moderate)

Cases where RNA-seq provided further information for interpretation (n = 4)

Case no.	Indication	RNA-seq finding	Information provided by RNA-seq
B12 (Trio)	<u>Congenital Anomalies:</u> type 1 jejunal atresia and inguinal hernia, also involving DD, genital, and ocular phenotypes.	AS: SON ($ \Delta J $ = 0.45; FDR = 0.00065)	Refined in silico prediction and indicate limited damaging effects of variant - Variant: NM_138927.4 (SON): c.78-2A>C (paternally inherited heterozygous VUS) - RNA-seq: in-frame 6 bp deletion - Interpretation: Limited damaging effect
B37 (Trio)	<u>Neurodevelopmental disorder:</u> global DD with mild ID and indications of Dandy-Walker variant or vermian-cerebellar hypoplasia	AE: 14 genes on chromosomal Xq28 were detected as significant overexpression (FDR < 0.1; FC > 2).	Confirming duplication and narrow down possible genes - Variant: Duplication on chromosomal Xq28 (<i>de novo</i> heterozygous) - RNA-seq: confirmed duplication and suggested <i>GDI1</i>, <i>RPL10</i>, <i>FAM50A</i>, <i>LAGE3</i> associated with X-linked intellectual developmental disorder as candidate genes underlying patient phenotype - Interpretation: <i>GDI1</i> on patient phenotypes are highly suspected⁵⁴
B48 ⁵³ (Singleton)	<u>Cardiac disease:</u> Dilated cardiomyopathy	MYBPC3 is not detected as RNA splicing outlier	Validate in silico prediction and refined prior experimental results - Variant: NM_000256.3 (MYBPC3):c.1224-80G>A (paternally inherited heterozygous VUS) - RNA-seq: exon 14 elongation (78 bp) - Interpretation: RNA-seq supported SpliceAI prediction⁷⁰ and provided a more refined picture to the minigene splice report assay⁸¹ but did not change the classification of VUS

(Table 2 continues on next page)

Cases where RNA-seq provided further information for interpretation (n = 4)

Case no.	Indication	RNA-seq finding	Information provided by RNA-seq
(Continued from previous page)			
B80 (Duos)	Neurodevelopmental disorder; severe global DD with ID, failure to thrive, hypotonia, Dandy-walker malformation, and facial dysmorphism	AS: KMT2A ($ \Delta J $ = 0.47; FDR = 0.00097)	Validated in silico prediction and indicate limited damaging effects of variant - Variant: NM_001197104.2 (KMT2A): c.4480-18_4480del (heterozygous, not maternally inherited) - RNA-seq: deletion of 3 bp in exon 12 without frameshift - Interpretation: limited damaging effect

Further description of individual cases is available in [Supplementary Information S1](#). Thresholds for AE: FDR $\leq$ 0.1. Thresholds for AS: FDR $\leq$ 0.1 and $|\Delta J| \geq$ 0.1. ACMG: American College of Medical Genetics; AE = Aberrant Expression; AS = Aberrant Splicing; DD = developmental delay; FC = fold change; FDR = false discovery rate; ID = intellectual disability; OMIM = Online Mendelian Inheritance in Man; RNA-seq: RNA sequencing; VUS = Variants of uncertain significance; $|\Delta J|$ = difference in Intron Jaccard Index. *This case was diagnosed through manual inspection of Integrative Genomics Viewer with insignificant RNA sequencing (RNA-seq) results based on thresholds set in this study.

Table 2: Cases aided by hypothesis-driven blood RNA-seq analysis (n = 10).

A paternally inherited heterozygous variant NM_138927.4(SON): c.78-2A>C was identified in a prior genetic test. This variant was originally classified as a VUS owing to the presence of three potential cryptic splice sites in the ± 20 bp window based on *in silico* predictions thereby leading to ambiguous interpretation (Fig. 1A). One cryptic splice site (A) leads to in-frame splicing while the other two (B and C) lead to frameshift, which are much more deleterious than the former. However, based solely on DNA-level data, no explicit conclusions could be drawn on the variant's pathogenicity without knowing the preferred cryptic splice site. In our RNA-seq results, SON was detected as

an AS outlier ($|\Delta J|$ = 0.45; FDR = 0.00065), showing that cryptic splice site A was used in almost half of the detected transcripts leading to an in-frame deletion of 6 bp without utilising other cryptic splice sites (B and C) (Fig. 1B). The insignificant expression results (fold change = 1.1; unadjusted p-value = 0.28; FDR = 1), further suggested that NMD was unlikely to be occurring at the other cryptic splice sites (B and C). This supported that the pathogenicity of this variant is limited as no frameshift events were observed, and the in-frame deletion is not situated within a known functional domain. Often overlooked during analysis, downgrading a variant's pathogenicity is just as

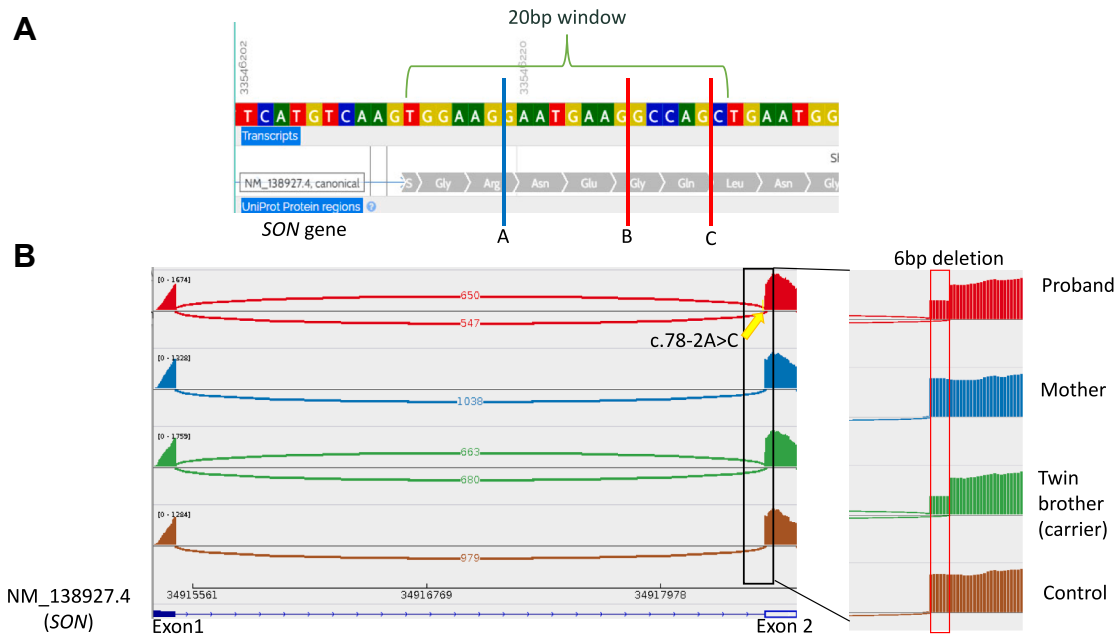


Fig. 1: In silico prediction and RNA-seq data of case B12. (A) Three cryptic splice sites were identified within ± 20 bp window from the affected splice site in exon 2 of SON (NM_138927). Blue line indicates in-frame deletion while red line indicates a frameshift. (B) The Integrative Genomics Viewer showed a clear usage of cryptic splice site in almost half of the transcripts in both proband and carrier twin brother who is not affected, which is absent in non-carrier mother and control. A zoom-in of the splice event showed a deletion of 6 bp in both the proband and the carrier twin brother.

important as upgrading a variant to establish a genetic diagnosis.

Hypothesis-free blood RNA-seq analysis

The interpretation of another 11 cases (10.8%) was aided by hypothesis-free blood RNA-seq analysis, where diagnostic discoveries were made in the absence of guidance from prior genetic tests. On average, 8.6 (AE) and 12.5 (AS) outliers per sample were filtered down to 2.0 (AE) and 2.9 (AS) outliers per sample after selecting for morbid genes, resulting in a handful of candidate genes for further interpretation (Supplementary Table S5). Five of these cases achieved diagnoses through the identification of previously undetected causative P/LP variants of currently known disease genes, including intronic variants and deletion mutations (Table 3; Supplementary Table S3; Supplementary Information S2). RNA-seq guided variant identification in DNA-level tests, where variants were identified directly from RNA data in two cases (B47 and B107), and provided functional evidence for variant pathogenicity classification.

In the remaining six cases, RNA-seq guided the discovery of unestablished disease mechanism by prioritising relevant genes for further functional studies through (1) discovering new disease genes; (2) widening the variant type by suggesting tandem repeat variants in a disease that only had single nucleotide variants reported; and (3) broadening the phenotypic spectrum of existing diseases (Table 3; Supplementary Table S3; Supplementary Information S2).

Regarding the discovery of new disease genes, RNA-seq guided and supported the establishment of two new disease genes. Our RNA-seq analysis directly detected a new disease gene as an outlier in case B22, who was suspected to have a mitochondrial or neurodegenerative disease. We detected *PSMF1* with both aberrant expression (Fold change (FC) = 0.23; FDR = 0.020) and aberrant splicing ($|\Delta J|$ = 0.44; FDR = 8.3×10^{-24}) in this proband which guided the identification of biparentally inherited NM_006814.4 (*PSMF1*):c.282+5G>A in both RNA-seq and exome data. The role of *PSMF1* in proteasome transport in axons⁵⁵ sparked interest in the possibility of this gene being associated with neurological disorders. This subsequently led to further collaboration with an international research team for the establishment of disease associations between *PSMF1* and a neurodegenerative syndrome with movement disorders.⁵⁶ The downregulation of this gene from our RNA-seq results is suggestive of a LoF mechanism for this gene–disease association. In a separate proband, B31, who also presented with neurodevelopmental phenotypes, RNA-seq indirectly detected the downstream effect (e.g., RNA signature) of the aberrant causal gene. A significantly larger number ($n = 106$) of aberrant splicing outliers were detected in this case compared to controls (Z-score = 4.8, Student T

two-tailed $p = 4.7 \times 10^{-6}$). The *de novo* variant NM_004640.7 (*DDX39B*): c.368G>A was subsequently identified in exome data guided by the RNA-seq finding. *DDX39B* is a key member of the TREX-complex which is important for mRNA transcription and export processes.⁵⁷ Further collaboration with an international research team identified five additional individuals with variants in this gene sharing similar phenotypes establishing its association with TREX-complex-related neurodevelopmental syndrome.⁵² Additionally, a consistent RNA signature (i.e., large number of non-specific aberrant splice events) was detected in affected individuals, B66 and B67 who share neurodevelopmental disorder phenotypes, which might serve as future diagnostic candidates.

In some complex cases, the detection of an aberrant expression outlier by RNA-seq can serve as a catalyst, guiding subsequent advanced genomic testing to uncover previously unrecognised variant types. Demonstrating this, we identified *HECTD4* as an aberrant expression outlier (FC = 0.51; FDR = 0.013) in proband B93, who presented with global developmental delay and autism spectrum disorder (ASD). This aberrant expression prompted subsequent lrWGS analysis revealing a heterozygous GCC expansion of >1000 bp in the 5' UTR of *HECTD4*, accompanied by simultaneous hypermethylation on the same allele. No other likely variants were identified in the gene. *HECTD4* is previously associated with “Neurodevelopmental disorder with seizures, spasticity, and complete or partial agenesis of the corpus callosum” (MIM #620250) in an autosomal recessive form. However, our patient's phenotypes were found to be less severe than the previous (and limited) reported cases with LoF and missense variants.⁵⁸ The SFARI Gene database, which evaluates the strength of association between specific genes and ASD,⁵⁹ along with previous literature^{60,61} has indicated possible association between heterozygous *HECTD4* LoF variants with ASD. Together, these multi-omic findings suggest a possible expansion of the pathogenic variant spectrum for *HECTD4*, wherein tandem repeat expansions may lead to a milder form of the neurodevelopmental disorder following an autosomal dominant inheritance pattern.

Apart from variant types, RNA-seq also suggested possible widening of phenotypic spectrum, showing RNA-seq's extended capability of providing more in depth understanding of reported diseases. For example, *WDR11* was detected as an aberrant expression outlier (FC = 0.55; FDR = 0.0014) in proband B114 with both genital and intellectual developmental phenotypes. *WDR11* has been associated with two diseases with conflicting and limited understanding,^{62–65} including “Hypogonadotropic hypogonadism 14 with or without anosmia” (OMIM #614858) in an autosomal dominant form with mainly missense variants and “Intellectual developmental disorder” (OMIM #620237) in an autosomal recessive form with

Cases with new diagnosis (n = 5)

Case no.	Indication	RNA-seq Finding	Gene-disease association	Variant identified	ACMG classification (after RNA-seq)	Role of RNA-seq in diagnosis
BF4 (Trio)	<u>Suspected mitochondrial disease</u> : Complex I and IV deficiency mitochondrial disease	AE : <i>GFM1</i> (FC = 0.57; FDR = 0.016) <i>MFSD1</i> (FC = 0.03; FDR = 1.1×10^{-10}) Results later confirmed with fibroblast samples ⁴⁷	Combined oxidative phosphorylation deficiency 1 (OMIM# 609060)	Biparentally inherited homozygous 104.5 kb deletion (chr3:158435847-158540316) (GRCh37)	Likely Pathogenic (PP4_strong, PM3_supporting, PM2_supporting)	- Guided variant identification in ES, further confirmed in GS. - Provided evidence for PP4_strong.
B47 (Trio)	<u>Neurodevelopmental disorder</u> : global DD with ID, autism, and central hypotonia	AS : <i>KMT5B</i> ($ \Delta J $ = 0.44; FDR = 0.057)	Intellectual developmental disorder 51 (MRD51; OMIM# 617788)	Maternally inherited heterozygous <i>KMT5B</i> variant NM_017635.5:c.977+2T>A	Likely Pathogenic (PVS1_strong (RNA), PM2_supporting, PP4_supporting)	- Variant identified in RNA data, confirmed with sanger sequencing - Provided evidence for PVS1_strong(RNA) - Prompted further investigation (e.g., DNA methylation signature)
B84 (Singleton)	<u>Neurodevelopmental disorder</u> : suspected distal myopathy, microcephaly, speech delay, learning disability and ADHD	AE : <i>CTCF</i> (FC = 0.66; FDR = 0.000021)	Intellectual developmental disorder, autosomal dominant 21 (OMIM# 615502)	<i>De novo</i> heterozygous ~280 kb deletion [NC_000016.10:g.(67429006_67710074)del]	Pathogenic (1 point for CNV loss)	Guided variant identification in GS.
B105 (Trio)	<u>Multisystem disorder</u> : Neurofibromatosis with distinctive features.	AS : <i>NF1</i> ($ \Delta J $ = 0.32; FDR = 0.021)	Neurofibromatosis type 1 (OMIM# 162200)	<i>De novo</i> heterozygous <i>NF1</i> deletion NC_000017.11:g.31327741_31329914delinsT (NM_001042492.3:c.5511_5610-382delinsT)	Pathogenic (PVS1_moderate (RNA), PS2, PM2_supporting, PP4_strong)	- Guided variant identification in GS - Provided evidence for PVS1_moderate (RNA)
B107 (Trio)	<u>Multisystem disorder</u> : Neurofibromatosis with distinctive features	AS : <i>NF1</i> ($ \Delta J $ = 0.56; FDR = 0.0046)	Neurofibromatosis type 1 (OMIM# 162200)	<i>De novo</i> heterozygous <i>NF1</i> variant NM_001042492.3:c.2410-16A>G	Pathogenic (PVS1_strong (RNA), PM2_supporting, PP4_strong, PS2)	- Variant identified in RNA data, confirmed in GS - Provided evidence for PVS1_strong(RNA)

Cases with disease mechanism discovery (n = 6)

Case no.	Indication	RNA-seq finding	Information provided by RNA-seq
B22 ⁵⁶ (Trio)	<u>Neurodegenerative disorder</u> : DD and progressive motor phenotypes.	AE : <i>PSMF1</i> (FC = 0.23; FDR = 0.020) AS : <i>PSMF1</i> ($ \Delta J $ = 0.44; FDR = 8.3×10^{-24})	Discovered potential new disease gene - RNA-seq: 17 bp exon 2 elongation (majority) leading to NMD - Variant: NM_006814.4 (<i>PSMF1</i>):c.282+5G>A (biparentally inherited homozygous) - Investigation: RNA-seq flagged <i>PSMF1</i> as possible disease gene, subsequently aiding new gene-disease association through international collaboration
B31 ⁵² (Trio)	<u>Neurodevelopmental disorder</u> : global DD with severe ID and hypotonia, short stature and dysmorphic features	AS : Significant aberrant splice events in 106 genes (FDR<0.1), which is significantly larger (Z-score = 4.8, Student T two-tailed p-value = 4.7×10^{-6}) than other internal controls and external GTEx controls—RNA signature	Discovered potential new disease gene - RNA-seq: RNA signature - Variant: NM_004640.7 (<i>DDX39B</i>): c.368G>A p.(Arg123Gln) (<i>de novo</i> heterozygous) - Investigation: RNA-seq flagged <i>DDX39B</i> as possible disease gene, leading to new gene-disease association through international collaboration and establishment of RNA signature verifying gene function.
B66 ⁵² (Singleton)	<u>Neurodevelopmental disorder</u> : global DD with hypotonia, seizure, autistic behaviour, short stature and dysmorphic features.	AS : Significant aberrant splice events in 303 genes with FDR<0.1 which is significantly larger (Z-score = 14.9, Student T two-tailed p-value = 2.4×10^{-29}) than other control and external GTEx controls—RNA signature	Validating functionality of newly discovered disease-genes - RNA-seq: RNA signature - Variant: NM_004640.7 (<i>DDX39B</i>): c.109G>T p.(Gly37Cys) (<i>de novo</i> heterozygous) - Investigation: RNA signature helped verify the function of <i>DDX39B</i>

(Table 3 continues on next page)

Cases with disease mechanism discovery (n = 6)

Case no.	Indication	RNA-seq finding	Information provided by RNA-seq
(Continued from previous page)			
B67 ⁵² (Duos)	Neurodevelopmental disorder; delayed motor development, history of seizures, autism and short stature.	AS: Signature aberrant splice events in 297 genes with FDR<0.1 which is significantly larger (Z-score = 14.6, Student T two-tailed p-value = 1.2×10^{-28}) than other control and external GTEx controls—RNA signature AS: DDX39B (ΔJ = 0.48; FDR = 1.8×10^{-10})	Validating functionality of newly discovered disease-genes - RNA-seq: RNA signature and 33 bp in-frame deletion in exon 5 of DDX39B - Variant: NM_004640.7 (DDX39B): c.433-1G>T (maternally inherited heterozygous) - Investigation: RNA signature helped verify the function of DDX39B. Aberrant splice event supported the pathogenicity of the variant.
B93 (Trio)	Neurodevelopmental disorder; global DD and autism	AE: HECTD4 (FC = 0.51; FDR = 0.013)	Suggested possible new disease mechanism - RNA-seq: HECTD4 downregulation by half - Variant: GCC repeat expansion of >1000 bp in HECTD4 5' UTR (heterozygous) and is hypermethylated - Investigation: possible new disease mechanism where milder form of the associated AR neurodevelopmental disorder is manifested in monoallelic form through methylation-related events
B114 (Singleton)	Multisystem disorder; moderate ID, bilateral sensorineural hearing loss, and genitalia abnormalities.	AE: WDR11 (FC = 0.55; FDR = 0.0014) 26 genes (including WDR11) within the chr10 deletion region are under expressed (FDR<0.1; FC = 0.32–0.61)	Suggested possible expansion of disease spectrum - RNA-seq: downregulation of WDR11 by half - Variant: NC_000010.11:g.(119969782_133655950) del (heterozygous) - Interpretation: Under expression of 26 genes (including WDR11) with this region supported the deletion. Patient phenotype lies between the established WDR11-associated AD and AR disease, suggesting a spectrum of disease associated with this gene.

Further description of individual cases is available in [Supplementary Information S2](#). Thresholds for AE: FDR ≤ 0.1. Thresholds for AS: FDR ≤ 0.1 and |ΔJ| ≥ 0.1. ACMG = American College of Medical Genetics; AD = Autosomal dominant; ADHD = Attention deficit hyperactivity disorder; AE = Aberrant Expression; AR = autosomal recessive; AS = Aberrant Splicing; CNV = copy number variant; DD = developmental delay; ES = exome sequencing; FC = Fold change; FDR = false discovery rate; GS = genome sequencing; GTEx = Genotype-Tissue Expression Project; ID = intellectual disability; NMD = nonsense mediated decay; OMIM = Online Mendelian Inheritance in Man; RNA-seq = RNA sequencing; UTR = untranslated region; |ΔJ| = difference in Intron Jaccard Index.

Table 3: Cases aided by hypothesis-free blood RNA-seq analysis (n = 11).

LoF variants. Our proband's phenotype overlapped with both diseases, suggesting a possible expansion of the phenotypic spectrum of *WDR11*-related diseases. This prompted the identification of a heterozygous deletion NC_000010.11:g.(119969782_133655950)del from WGS, supported by 26 under expressed genes from our AE result (FDR < 0.1; FC = 0.32–0.61). This deletion contained *WDR11* and none of the remaining genes within the deletion were found to provide likely explanations of the patient's phenotypes. For both cases discussed (B93 with *HECTD4*, and B114 with *WDR11*), WGS had no positive findings from gene panels associated with patient phenotypes, therefore we highly suspect the clinical relevance of *HECTD4* and *WDR11* in these two cases. Nevertheless, further studies and deeper understanding are required to confirm our hypothesis and establish a diagnosis.

Fibroblast RNA-seq

Eleven of the 102 patients in our cohort had fibroblast samples available for RNA-seq analysis; for one of them (BF4), fibroblast RNA-seq confirmed the positive blood RNA-seq finding. Although the other ten cases had negative results in blood RNA-seq analysis, one of these cases was solved by hypothesis-free fibroblast RNA-seq, as reported in our previous publication.⁴⁷ Interestingly,

fibroblast RNA-seq detected an RNA signature absent in blood RNA-seq. Proband BF1 presented with global developmental delay, mild intellectual disability, leukoencephalopathy, progressive spasticity, liver derangement, progressive renal failure with protein-losing enteropathy, and suspected bilateral optic atrophy. A significant number (n = 26) of histone genes were detected as aberrantly expressed in fibroblasts, but not in blood, when compared to controls (FC = 2.3–7.9; FDR = 2.9×10^{-7} –0.054). This guided the identification of compound heterozygous variants (NR_023317: n.[28C>T]; [35G>A]) in a small nuclear RNA gene *RNU7-1*, which is responsible for mRNA processing and associated with Aicardi-Goutières syndrome (OMIM #619487).⁶⁶ The RNA-seq finding combined with an abnormal interferon signature (interferon score of 18.6; decimalised age of 19.3 years) consistent with the disease enabled the application of PP4_strong evidence. This led to a classification of likely pathogenic for both variants (PM3, PM2_supporting, PP4_strong) and the establishment of a diagnosis ([Supplementary Information S3](#) and [Supplementary Table S3](#)).

Discussion

Through the application of blood (and fibroblast) RNA-seq analysis on 102 patients with genetically

undiagnosed disorders, our pipeline aided the interpretation of 21.6% (22/102) of cases. With an incremental diagnostic yield of 11.8% (12/102) based on currently known gene–disease associations, nine of these cases demonstrated aberrant splice events. Blood RNA-seq alone led to a diagnosis in 10.8% ($n = 11$) of cases, which is slightly higher than reported in previous studies ($\sim 8.5\%$) (Supplementary Table S1).^{24,28,29,36,38,44} This demonstrates the diagnostic value of blood RNA-seq in rare genetic diseases, even for clinically well-defined disorders like Noonan syndrome, with particular benefits in identifying mutations with splicing phenotypes. In addition to facilitating diagnoses, our findings show that blood RNA-seq can serve both clinical and scientific value through new disease mechanism discoveries, re-evaluating VUS candidates (e.g., downgrading pathogenicity), and facilitating hypothesis-free research to aid more patients suffering from similar diseases.

Regarding the clinical integration of RNA-seq and translatability, our analysis addressed an ongoing debate on the “most appropriate” practice for RNA-seq integration into the diagnostic workflow. Two approaches are possible: (1) hypothesis-driven, where RNA-seq is used to validate candidate findings from DNA-level tests, and (2) hypothesis-free, where the analysis is primarily driven by RNA-seq results with no prior genetic candidates. More specifically, our findings address whether a hypothesis-free approach (i.e., RNA-centric or RNA-first) is beneficial for case interpretation. Upon reviewing 18 studies that used RNA-seq to aid genetic diagnosis (Supplementary Table S1), we observed that the hypothesis-driven approach and hypothesis-free approach aided interpretation of 72.1% (147/204) and 27.9% (57/204) cases, respectively. For cases achieving diagnosis ($n = 150$), a hypothesis-free approach was responsible for 25.3% of the diagnoses. Despite the seeming advantage of a hypothesis-driven approach in increasing the diagnostic rate owing to the detection of “low-hanging fruit” given prior candidate genes, the hypothesis-free approach in our study has demonstrated its importance in reprioritising variants missed in previous genetic tests. The hypothesis-free approach also served to identify genes with limited understanding, enabling the establishment of new gene–disease associations. Our results support the utility of blood RNA-seq in both hypothesis-driven and hypothesis-free approaches. The proportion of cases aided by each type of analysis was approximately equal (50:50), showing that they are equally important and useful in solving undiagnosed cases. Aligning with the literature, hypothesis-free analysis reprioritised/guided variant identification in DNA data ($n=3$), identified underlying variants in RNA ($n = 2$), and aided discovery of disease mechanisms ($n = 6$). This supports the application of RNA-seq regardless of whether prior candidates have been indicated in undiagnosed cases.

The applicability of hypothesis-free approach has also been demonstrated by our workflow, with a resultant handful of candidate genes identified for further manual inspection. The capability of our pipeline to analyse samples from heterogeneous diseases, facilitated by the autoencoder, further enhanced its generalisability and translatability without having to accumulate samples from the same disease cohort for resource-demanding patient-specific studies. Our pipeline maximises benefits in healthcare management of patients with rare diseases by providing more timely diagnoses, which are essential for clinical management decisions regarding precision treatments and supportive care planning.

Despite the promising clinical diagnostic utility of transcriptomics, at present, there is a lack of a consensual framework regarding its clinical application. Zhao et al.⁴³ recently proposed a validation framework for benchmarking RNA-seq workflows across different laboratories, while our study demonstrated different considerations required when applying RNA-seq results into the variant curation pipeline. According to the ACMG/AMP guidelines,^{20,45,51} results from RNA-seq can be applied to variant curation mainly through two criteria: (1) PVS1 (i.e., a null variant identified in a gene with LoF as an established disease mechanism), and (2) PP4 (i.e., phenotypic specificity).

For PVS1, RNA-seq can provide further information in three aspects to determine the final strength of this criterion.^{45,51} First, it provides functional data on whether the transcript has undergone nonsense-mediated decay (NMD) through aberrant expression (AE) findings. The most widely used rule to predict NMD is that a transcript will escape NMD if the premature termination codon (PTC) is situated in the last exon or the last 50–55 bp of the second-to-last exon.^{51,67} However, our data showed that out of nine newly diagnosed cases with aberrant splicing events, only one case had concordant results between NMD prediction and RNA-seq (Supplementary Table S6 and Supplementary Informations S1 and S2). This discordance was also reported in previous studies,^{38,43} suggesting that this might be partially attributed to variable NMD efficiencies in different tissues. To remain conservative, we suggest that the “predicted to undergo NMD” pathway in the PVS1 decision tree should only be selected if RNA-seq clearly demonstrates NMD (i.e., a statistically significant decrease in fold change that matches the zygosity).⁵¹ Evidently, RNA-seq is able to provide a more accurate understanding of NMD efficiency and hence avoid the overestimation of “pathogenicity” for the variant under investigation. Second, RNA-seq results act as functional evidence (PVS1 criterion) through aberrant splicing (AS) findings that replaces *in silico* predictions (PP3 criterion), the latter of which has a lower strength of pathogenicity (Tables 2 and 3 and Supplementary Informations S1 and S2).

Additionally, aligning with previous studies,^{38,68} RNA-seq can provide further refined information on splicing events compared to *in silico* predictions, demonstrated by 4/8 newly diagnosed cases with splice variants in our study (Supplementary Table S6 and Supplementary Informations S1 and S2). An accurate understanding of aberrant events is of paramount importance for future development of specific nucleic acid therapies.⁶⁹ Third, RNA-seq provides more quantifiable results for assessing the “completeness”⁴⁵ of the splice event as opposed to more traditional methods like gel electrophoresis. While explicit thresholds have yet to be established, our study considered an aberrant event complete when RNA-seq results match the zygosity of the variant.

As an illustration of the RNA-seq application, two examples are described in cases B68 (Fig. 2) and B105 (Fig. 3):

In B68, RNA-seq identified *MSL3* as a significant splicing outlier (FDR = 7.9×10^{-24} , $|\Delta J| = 0.95$). For the curation of the hemizygous NM_078629.4(*MSL3*): c.749+5G>A variant, PP3 is replaced with PVS1_strong (RNA) by choosing “not predicted to undergo NMD” and “complete” for the aberrant events. This is due to the uncertainty of NMD in AE module (FC = 0.81; unadjusted p-value = 0.027; FDR = 1) and events conforming with hemizygosity, supported by a $|\Delta J|$ value near 1.

In B105, RNA-seq flagged *NF1* as a significant splicing outlier (FDR = 0.021; $|\Delta J| = 0.32$), clearly showing 22/90 transcripts skipping exon 38. This guided the identification of the heterozygous NC_000017.11:g.31327741_31329914delinsT mutation through srWGS. The variant is classified as pathogenic with PVS1_mod(RNA) by choosing “not predicted to undergo NMD” and “near complete” for the splice event. This is attributed to the insignificant fold change (FC = 0.81; unadjusted p-value = 0.023; FDR = 1) and the aberrant event not conforming to heterozygosity.

For PP4 (phenotypic specificity), this criterion can be applied in situations where downstream events of the mutation are detected instead of the variant-carrying gene. Demonstrated by BF1, RNA-seq analysis identified enrichment of overexpressed histone genes, which can be regarded as “RNA phenotypes”. This enables the application of PP4 in pathogenicity classification and hence, achieving diagnosis (Supplementary Information S3). Similarly, for *DDX39B* cases (Table 3 and Supplementary Information S2), the significantly increased number of aberrant RNA events (i.e., RNA signature) can serve as an “RNA phenotype” for the application of PP4 in variant curation for future diagnosis. However, PP4_supporting is suggested, as multiple other

disease genes have been reported to present similar aberrant RNA phenotypes.^{71,72} In another case (BF4), RNA-seq flagged *GFM1* as significantly downregulated, which later guided the identification of a 104.5 kb homozygous deletion that included 11 expression enhancers of *GFM1*. This RNA finding acts as a “phenotype” specific to loss of *GFM1* enhancers and supported the LP classification of the deletion mutation through PP4_strong (Table 3 and Supplementary Information S2).

Despite the strategies applied by our study, the current curation framework still requires further work to better incorporate multiple aspects of RNA-seq data. Several questions remain. For example, does the “completeness” of an event refer to expression and/or splicing data? How should we interpret isoform switching-related data? Which algorithms and thresholds should be implemented for analysis? The current thresholds, which are based heavily on adjusted p-values, are likely leading to the overlooking of aberrant events, which might also explain the substantial inconsistency between NMD and expression data. As exemplified by case B56, manual curation identified a statistically insignificant aberrant splice event in *RARS2*, resulting in the upgrading of a previously detected VUS to LP. In addition, continuous amendments are required to determine the specificity of an RNA signature and the threshold of aberration. Further benchmarking and determination of the “best” workflow for transcriptomics implementation are required to maximise its benefit in clinical genetics.

We acknowledged several limitations in our study. First, while the inclusion of 2 additional patients with known *DDX39B* variants might have slightly inflated the number of cases our pipeline aided, they are not included in our incremental diagnostic yield. More importantly, their inclusion led to the discovery of a potential RNA signature that might act as a future diagnostic entity. RNA signatures have already reported in multiple different genes including *RNU4ATAC* variants in *RNU4atac-opathy*,⁷¹ *RNU4-2* variants in ReNU syndrome⁷³ or *RNU12* variants in CDAGS syndrome.⁷² Another case (BF1) in our cohort has also indicated the potential presence of an RNA signature in *RNU7-1* for Aicardi-Goutières syndrome. Not only does it potentially guide the identification of variants, but it also supports the variant’s pathogenicity through the application of PP4. Second, other omics technologies will be required to complement transcriptomics for variants that do not have an RNA phenotype. ClinVar reports that 64% of VUS are missense variants, which likely affect protein stability/folding.¹⁹ The clinical value of incorporating proteomics into the RNA-seq diagnostic pipeline has been proven in a mitochondrial cohort and our own neurodevelopmental cohort, yielding an incremental diagnostic rate of 22.3% and 14.7% upon DNA-level analysis, respectively.^{47,74} preprint Approximately half of these diagnosed cases were only

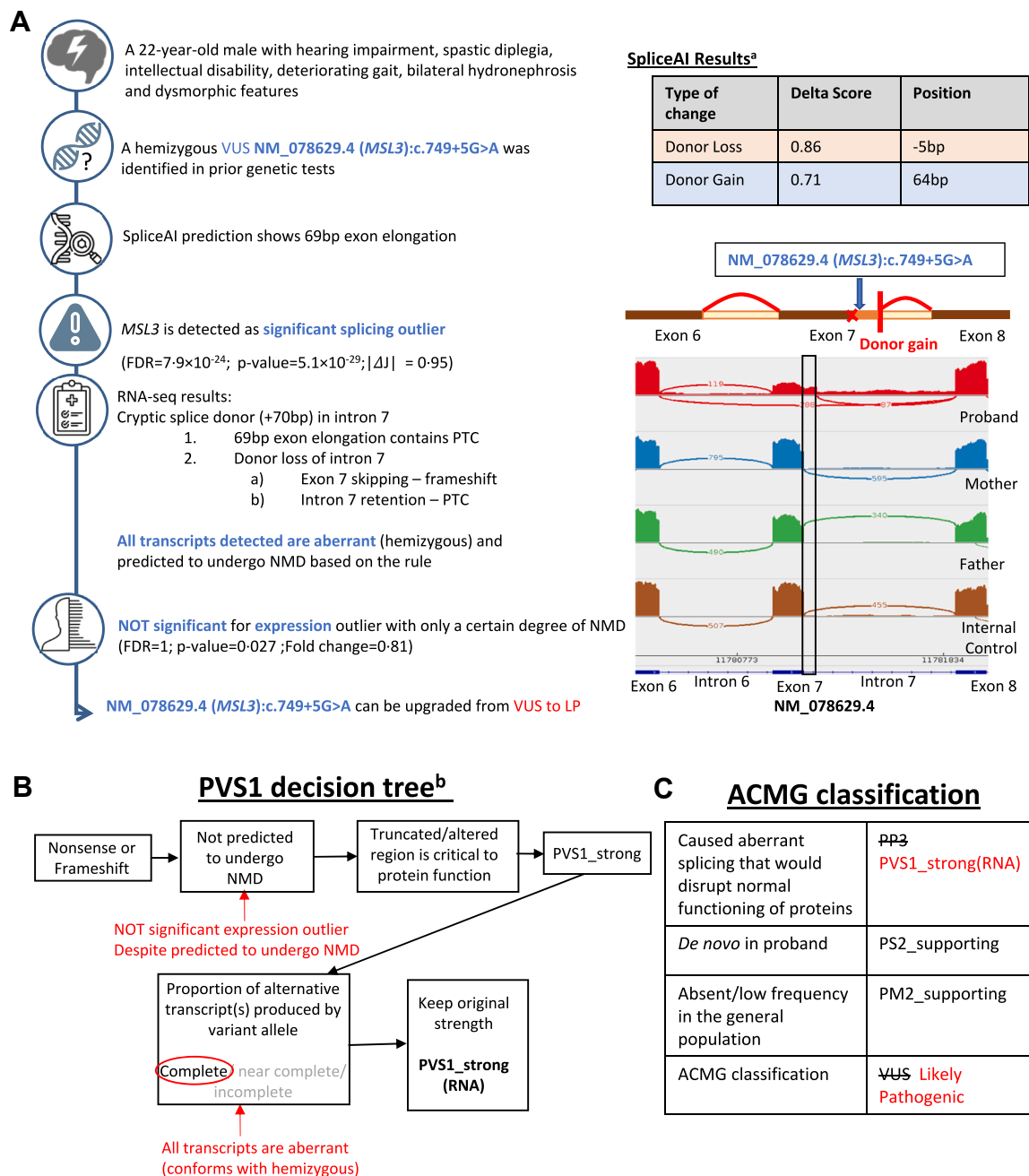


Fig. 2: Upgrading a VUS variant to likely Pathogenic in case B68. (A) The timeline shows the interpretation and sequence of analysis of this case, from identifying a MSL3 VUS candidate variant to upgrading the variant to likely pathogenic. A resultant intron retention was initially predicted based on SpliceAI results. Statistical tests performed are in-built executions of the DROP pipeline.³⁹⁻⁴² ^aThe results are directly derived from SpliceAI⁷⁰ webtool (<https://spliceailookup.broadinstitute.org/>). (B) This shows the pathway taken for PVS1 decision tree, indicating how RNA-seq results were incorporated. ^bThe decision trees are referenced from Abou Tayoun et al.⁵¹ and Walker et al.⁴⁵ (C) Criteria in black are used before RNA-seq results were available. Criteria in red are changes and final classification made based on RNA-seq results. J: Intron Jaccard Index; FDR: false discovery rate; LP: likely pathogenic; NMD: nonsense-mediated decay; PTC: premature termination codon; VUS: variant of uncertain significance.

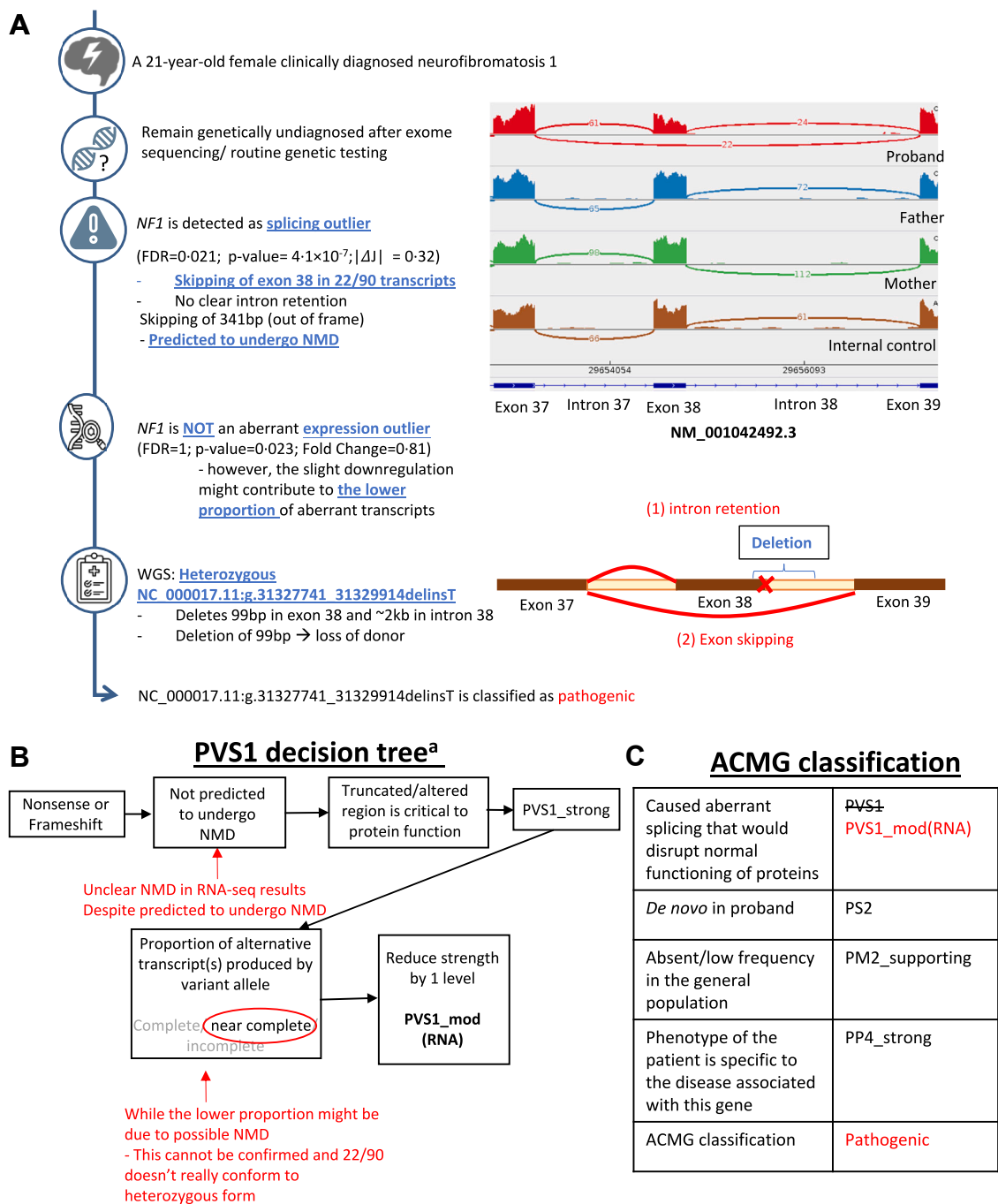


Fig. 3: RNA-seq results guided the identification of a pathogenic variant for case B105. (A) The timeline shows the interpretation and sequence of analysis of a case with clinically diagnosed neurofibromatosis 1. RNA-seq results guided the identification of pathogenic NF1 variant in whole genome sequencing. 2 outcomes (i.e., intron retention and exon skipping) are possible in the event of donor loss based on the prediction. Statistical tests performed are in-built executions of the DROP pipeline.^{39–42} (B) This shows the pathway taken for PVS1 decision tree, indicating how RNA-seq results were incorporated. ^a The decision trees are referenced from Abou Tayoun et al.⁵¹ and Walker et al.⁴⁵ (C) Criteria in black are used before RNA-seq results were available. Criteria in red are changes and final classification made based on RNA-seq results. J: Intron Jaccard Index; FDR: false discovery rate; NMD: nonsense-mediated decay.

detected in proteomics in both papers. Further studies are required to establish similar systematic workflows incorporating multi-omics technologies (e.g., proteomics, epigenomics) to advance clinical genetics in both diagnosis and precision treatments. Third, other tissues might need to be considered in cases where blood transcriptomics might not be informative, as in our case BF1. Multi-omics data are more subject to tissue-specific variabilities than “static” DNA. Multiple studies have shown that fibroblasts, while more invasive to obtain, have a higher coverage of Mendelian disease genes.^{35,43} Our previous study has also demonstrated the diagnostic utility of amniocyte transcriptomics, with a comparable coverage of expressed genes to fibroblasts.⁴⁶ There are currently emerging resources to aid the decisions on clinically accessible tissues (CAT) for analysis. For example, a score to estimate the minimum required sequencing depth (MRSD)^{75,76} for a gene in the CATs, and MAJIQ-CAT⁷⁷ which estimates how representative a CAT is to the tissue of interest in splicing. The choice of CATs should consider all these aspects, including the invasiveness of tissue sampling. If genes are not expressed in all CATs, transactivation of silent genes or transdifferentiation of CATs into the tissue of interest have also been shown to aid case interpretation.^{78–80} Nevertheless, further validation and benchmarking are still required for the application of these data into the clinical diagnostic workflow.

In conclusion, our study highlights the clinical diagnostic and scientific value of blood RNA-seq in both hypothesis-driven and hypothesis-free approaches within a heterogeneous rare disease cohort. It enabled diagnosis, discoveries, and the identification of potential diagnostic signatures. While suggesting an initial framework for its clinical implementation, our study also directed aspects where further guidance is required for the comprehensive incorporation of RNA-seq results into variant curation workflows.

Contributors

C.P., M.M.C.C. and W.T. wrote the manuscript and both C.P. and M.M.C.C. have accessed and verified the data reported. C.P., M.M.C.C., A.K.Y.K., and K.S.Y. performed data analysis. C.P. was responsible for the visualisation of the data. C.P. and M.M.C.C. performed data curation. C.P., A.K.Y.K., and H.Y.C.L., performed experiments. S.S.F., A.W.S.K., G.C.F.C., H.M.L., R.M.S.W., I.F.M.L., C.W.F., J.Y.L.T., A.P.Y.L., S.S.N.W. and B.H.Y.C. were responsible for patient recruitment. G.C.F.C., H.M.L., R.M.S.W., I.F.M.L., C.W.F., J.Y.L.T., A.P.Y.L., S.S.N.W. and B.H.Y.C. characterised clinical data from patients. W.Y. helped with the setting up of server and bioinformatics pipeline. C.C.Y.M. and B.H.Y.C. conceptualised and supervised the entire project. All authors read and approved the final version of the manuscript.

Data sharing statement

Clinical, genetic and transcriptomics data are provided in [Supplementary tables and informations](#). DNA and RNA datasets generated from this study are not made publicly available owing to ethical restrictions regarding minor participants and patient anonymity. Requests for the datasets can be directed to the corresponding author, and data will be de-identified before being transmitted to the qualified

researchers after review. The codes for DROP pipeline can be accessed through <https://github.com/gagneurlab/drop>. GTEx datasets were accessed through NCBI dbGaP (the Database of Genotypes and Phenotypes).

Declaration of interests

There are no existing conflicts of interest to declare.

Acknowledgements

We would like to thank the medical staff for the referrals, the healthcare professionals who managed the patients, and the patients and their families for their participation in the study. This study was supported by grants from the Society for the Relief of Disabled Children, the Health and Medical Research Fund (HMRF) (Project no.: 10212106) and Commissioned Paediatric Research at Hong Kong Children's Hospital under HMRF (PR-HKU-4) both by the Health Bureau, The Government of the Hong Kong Special Administrative Region. None of the sponsors played any role in the design and conduct of the study; data collection, management, analysis and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. We also thank the Centre for Pan-Omic Sciences at the University of Hong Kong for providing RNA-seq services and Mr. Christopher Brandon Preusch for English editing.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2026.106313>.

References

- 1 Boycott KM, Giugliani R. The RDI-Lancet Commission on Rare Diseases: improving visibility to address health-care disparities for 400 million people. *Lancet*. 2025;405:605–607. [https://doi.org/10.1016/s0140-6736\(25\)00211-9](https://doi.org/10.1016/s0140-6736(25)00211-9).
- 2 Health TLG. The landscape for rare diseases in 2024. *Lancet Glob Heal*. 2024;12:e341. [https://doi.org/10.1016/s2214-109x\(24\)00056-1](https://doi.org/10.1016/s2214-109x(24)00056-1).
- 3 Chiu ATG, Chung CCY, Wong WHS, Lee SL, Chung BHY. Healthcare burden of rare diseases in Hong Kong – adopting ORPHA codes in ICD-10 based healthcare administrative datasets. *Orphanet J Rare Dis*. 2018;13:147. <https://doi.org/10.1186/s13023-018-0892-5>.
- 4 Chung CCY, Ng NYT, Ng YNC, et al. Socio-economic costs of rare diseases and the risk of financial hardship: a cross-sectional study. *Lancet Reg Health West Pac*. 2023;34:100711. <https://doi.org/10.1016/j.lanwpc.2023.100711>.
- 5 Ng YNC, Ng NYT, Fung JLF, et al. Evaluating the health-related quality of life of the rare disease population in Hong Kong using EQ-5D 3-level. *Value Health*. 2022;25:1624–1633. <https://doi.org/10.1016/j.jval.2022.04.1725>.
- 6 Chung CCY, Hue SPY, Ng NYT, et al. Meta-analysis of the diagnostic and clinical utility of exome and genome sequencing in pediatric and adult patients with rare diseases across diverse populations. *Genet Med*. 2023;25:100896. <https://doi.org/10.1016/j.gim.2023.100896>.
- 7 Tsang MH, Leung GK, Ho AC, et al. Exome sequencing identifies molecular diagnosis in children with drug-resistant epilepsy. *Epilepsia Open*. 2019;4:63–72. <https://doi.org/10.1002/epi4.12282>.
- 8 Tsang MHY, Chiu ATG, Kwong BMH, et al. Diagnostic value of whole-exome sequencing in Chinese pediatric-onset neuromuscular patients. *Mol Genet Genom Med*. 2020;8:e1205. <https://doi.org/10.1002/mgg3.1205>.
- 9 Chau JFT, Lee M, Chui MMC, et al. Functional evaluation and genetic landscape of children and young adults referred for assessment of Bronchiectasis. *Front Genet*. 2022;13–2022:933381. <https://doi.org/10.3389/fgene.2022.933381>.
- 10 Chung CCY, Leung GKC, Mak CCY, et al. Rapid whole-exome sequencing facilitates precision medicine in paediatric rare disease patients and reduces healthcare costs. *Lancet Reg Health West Pac*. 2020;1:100001. <https://doi.org/10.1016/j.lanwpc.2020.100001>.
- 11 Kwong AK-Y, Tsang MH-Y, Fung JL-F, et al. Exome sequencing in paediatric patients with movement disorders. *Orphanet J Rare Dis*. 2021;16:32. <https://doi.org/10.1186/s13023-021-01688-6>.

- 12 Mak CC, Leung GK, Mok GT, et al. Exome sequencing for paediatric-onset diseases: impact of the extensive involvement of medical geneticists in the diagnostic odyssey. *NPJ Genom Med*. 2018;3:19. <https://doi.org/10.1038/s41525-018-0056-5>.
- 13 Yeung KS, Tso WWY, Ip JJK, et al. Identification of mutations in the PI3K-AKT-mTOR signalling pathway in patients with macrocephaly and developmental delay and/or autism. *Mol Autism*. 2017;8:66. <https://doi.org/10.1186/s13229-017-0182-4>.
- 14 Kwok SY, Liu AP, Chan CY, et al. Clinical and genetic profile of congenital long QT syndrome in Hong Kong: a 20-year experience in paediatrics. *Hong Kong Med J*. 2018;24:561–570. <https://doi.org/10.12809/hkmj187487>.
- 15 Tsang MHY, Kwong AKY, Chan KLS, et al. Delineation of molecular findings by whole-exome sequencing for suspected cases of paediatric-onset mitochondrial diseases in the Southern Chinese population. *Hum Genom*. 2020;14:28. <https://doi.org/10.1186/s40246-020-00278-0>.
- 16 Shickh S, Mighton C, Uleryk E, Pechlivanoglou P, Bombard Y. The clinical utility of exome and genome sequencing across clinical indications: a systematic review. *Hum Genet*. 2021;140:1403–1416. <https://doi.org/10.1007/s00439-021-02331-x>.
- 17 Robinson PN, Köhler S, Oellrich A, et al. Improved exome prioritization of disease genes through cross-species phenotype comparison. *Genome Res*. 2014;24:340–348. <https://doi.org/10.1101/gr.160325.113>.
- 18 Smedley D, Schubach M, Jacobsen JOB, et al. A whole-genome analysis framework for effective identification of pathogenic regulatory variants in mendelian disease. *Am J Hum Genet*. 2016;99:595–606. <https://doi.org/10.1016/j.ajhg.2016.07.005>.
- 19 Pérez-Palma E, Gramm M, Nürnberg P, May P, Lal D. Simple ClinVar: an interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database. *Nucleic Acids Res*. 2019;47:W99–W105. <https://doi.org/10.1093/nar/gkz411>.
- 20 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;17:405–423. <https://doi.org/10.1038/gim.2015.30>.
- 21 Riggs ER, Andersen EF, Cherry AM, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med*. 2020;22:245–257. <https://doi.org/10.1038/s41436-019-0686-8>.
- 22 Kremer LS, Bader DM, Mertes C, et al. Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat Commun*. 2017;8:15824. <https://doi.org/10.1038/ncomms15824>.
- 23 Riquin K, Isidor B, Mercier S, et al. Integrating RNA-Seq into genome sequencing workflow enhances the analysis of structural variants causing neurodevelopmental disorders. *J Méd Genet*. 2024;61:47. <https://doi.org/10.1136/jmg-2023-109263>.
- 24 Deshwar AR, Yuki KE, Hou H, et al. Trio RNA sequencing in a cohort of medically complex children. *Am J Hum Genet*. 2023;110:895–900. <https://doi.org/10.1016/j.ajhg.2023.03.006>.
- 25 Cummings BB, Marshall JL, Tukiainen T, et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med*. 2017;9. <https://doi.org/10.1126/scitranslmed.aal5209>.
- 26 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. <https://doi.org/10.1172/jci141500>.
- 27 Estévez-Arias B, Matalonga L, Yubero D, et al. Phenotype-driven genomics enhance diagnosis in children with unresolved neuromuscular diseases. *Eur J Hum Genet*. 2025;33:239–247. <https://doi.org/10.1038/s41431-024-01699-4>.
- 28 Lunke S, Bouffier SE, Patel CV, et al. Integrated multi-omics for rapid rare disease diagnosis on a national scale. *Nat Med*. 2023;29:1681–1691. <https://doi.org/10.1038/s41591-023-02401-9>.
- 29 Frésard L, Smail C, Ferraro NM, et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat Med*. 2019;25:911–919. <https://doi.org/10.1038/s41591-019-0457-8>.
- 30 Cloney T, Gallacher L, Pais LS, et al. Lessons learnt from multifaceted diagnostic approaches to the first 150 families in Victoria's Undiagnosed Diseases Program. *J Méd Genet*. 2022;59:748. <https://doi.org/10.1136/jmedgenet-2021-107902>.
- 31 Colin E, Duffourd Y, Chevarin M, et al. Stepwise use of genomics and transcriptomics technologies increases diagnostic yield in Mendelian disorders. *Front Cell Dev Biol*. 2023;11:1021920. <https://doi.org/10.3389/fcell.2023.1021920>.
- 32 Lee H, Huang AY, Wang L, et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet Med*. 2020;22:490–499. <https://doi.org/10.1038/s41436-019-0672-1>.
- 33 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;104:466–483. <https://doi.org/10.1016/j.ajhg.2019.01.012>.
- 34 Colin E, Duffourd Y, Tisserant E, et al. OMIXCARE: OMICS technologies solved about 33% of the patients with heterogeneous rare neuro-developmental disorders and negative exome sequencing results and identified 13% additional candidate variants. *Front Cell Dev Biol*. 2022;10:1021785. <https://doi.org/10.3389/fcell.2022.1021785>.
- 35 Yépez VA, Gusic M, Kopajtich R, et al. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med*. 2022;14:38. <https://doi.org/10.1186/s13073-022-01019-9>.
- 36 Oquendo CJ, Wai HA, Rich WI, et al. Identification of diagnostic candidates in Mendelian disorders using an RNA sequencing-centric approach. *Genome Med*. 2024;16:110. <https://doi.org/10.1186/s13073-024-01381-w>.
- 37 Marchant RG, Bryen SJ, Bahlo M, et al. Genome and RNA sequencing boost neuromuscular diagnoses to 62% from 34% with exome sequencing alone. *Ann Clin Transl Neurol*. 2024;11:1250–1266. <https://doi.org/10.1002/acn3.52041>.
- 38 Luo X, Xiao B, Liang L, et al. Blood RNA-seq in rare disease diagnostics: a comparative study of cases with and without candidate variants. *J Transl Med*. 2025;23:586. <https://doi.org/10.1186/s12967-025-06609-w>.
- 39 Yépez VA, Mertes C, Müller MF, et al. Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc*. 2021;16:1276–1296. <https://doi.org/10.1038/s41596-020-00462-5>.
- 40 Brechtman F, Mertes C, Matuszevičiūtė A, et al. OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA sequencing data. *Am J Hum Genet*. 2018;103:907–917. <https://doi.org/10.1016/j.ajhg.2018.10.025>.
- 41 Scheller IF, Lutz K, Mertes C, Yépez VA, Gagneur J. Improved detection of aberrant splicing with FRASER 2.0 and the intron Jaccard index. *Am J Hum Genet*. 2023;110:2056–2067. <https://doi.org/10.1016/j.ajhg.2023.10.014>.
- 42 Mertes C, Scheller IF, Yépez VA, et al. Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat Commun*. 2021;12:529. <https://doi.org/10.1038/s41467-020-20573-7>.
- 43 Zhao S, Macakova K, Sinson JC, et al. Clinical validation of RNA sequencing for Mendelian disorder diagnostics. *Am J Hum Genet*. 2025;112:779–792. <https://doi.org/10.1016/j.ajhg.2025.02.006>.
- 44 Blake B, Brady LI, Rouse NA, Nagy P, Tarnopolsky MA. The efficacy of whole genome sequencing and RNA-seq in the diagnosis of whole exome sequencing negative patients with complex neurological phenotypes. *J Pediatr Genet*. 2023;12:206–212. <https://doi.org/10.1055/s-0041-1736610>.
- 45 Walker LC, de La Hoya M, Wiggins GAR, et al. Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: recommendations from the ClinGen SVI Splicing Subgroup. *Am J Hum Genet*. 2023;110:1046–1067. <https://doi.org/10.1016/j.ajhg.2023.06.002>.
- 46 Lee M, Kwong AKY, Chui MMC, et al. Diagnostic potential of the amniotic fluid cells transcriptome in deciphering mendelian disease: a proof-of-concept. *npj Genom Med*. 2022;7:74. <https://doi.org/10.1038/s41525-022-00347-4>.
- 47 Chui MM-C, Kwong AK-Y, Leung HYC, et al. An outlier approach: advancing diagnosis of neurological diseases through integrating proteomics into multi-omics guided exome reanalysis. *npj Genom Med*. 2025;10:36. <https://doi.org/10.1038/s41525-025-00493-5>.
- 48 The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–585. <https://doi.org/10.1038/ng.2653>.
- 49 Online mendelian inheritance in man. <https://omim.org/>.
- 50 Lam WKJ, Lau CS, Luk HM, et al. The implementation of genome sequencing in rare genetic diseases diagnosis: a pilot study from the Hong Kong genome project. *Lancet Reg Health West Pac*. 2025;55:101473. <https://doi.org/10.1016/j.lanwpc.2025.101473>.
- 51 Tayoun ANA, Pesaran T, DiStefano MT, et al. Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant

- criterion. *Hum Mutat.* 2018;39:1517–1524. <https://doi.org/10.1002/humu.23626>.
- 52 Booth KTA, Jangam SV, Chui MMC, et al. De novo and inherited variants in DDX39B cause a novel neurodevelopmental syndrome. *Brain.* 2025;148:2658–2670. <https://doi.org/10.1093/brain/awaf035>.
 - 53 Kwok SY, Kwong AKY, Shi JZ, et al. Whole genome sequencing in paediatric channelopathy and cardiomyopathy. *Front Cardiovasc Med.* 2024;11:1335527. <https://doi.org/10.3389/fcvm.2024.1335527>.
 - 54 Vandewalle J, Esch HV, Govaerts K, et al. Dosage-dependent severity of the phenotype in patients with mental retardation due to a recurrent copy-number gain at Xq28 mediated by an unusual recombination. *Am J Hum Genet.* 2009;85:809–822. <https://doi.org/10.1016/j.ajhg.2009.10.019>.
 - 55 Liu K, Jones S, Minis A, Rodriguez J, Molina H, Steller H. PI31 is an adaptor protein for proteasome transport in axons and required for synaptic development. *Dev Cell.* 2019;50:509–524.e10. <https://doi.org/10.1016/j.devcel.2019.06.009>.
 - 56 Magrinelli F, Tesson C, Angelova PR, et al. Variants in the proteasome regulator PSMF1 cause a phenotypic spectrum from parkinsonism to perinatal lethality. *Nat Commun.* 2026. <https://doi.org/10.1038/s41467-026-71351-w>.
 - 57 Chi B, Wang Q, Wu G, et al. Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. *Nucleic Acids Res.* 2013;41:1294–1306. <https://doi.org/10.1093/nar/gks1188>.
 - 58 Faqeih EA, Alghamdi MA, Almahroos MA, et al. Biallelic variants in HECT E3 paralogs, HECTD4 and UBE3C, encoding ubiquitin ligases cause neurodevelopmental disorders that overlap with Angelman syndrome. *Genet Med.* 2023;25:100323. <https://doi.org/10.1016/j.gim.2022.10.006>.
 - 59 Abrahams BS, Arking DE, Campbell DB, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism.* 2013;4:36. <https://doi.org/10.1186/2040-2392-4-36>.
 - 60 Rubes SD, He X, Goldberg AP, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature.* 2014;515:209–215. <https://doi.org/10.1038/nature13772>.
 - 61 Yuen RKC, Merico D, Bookman M, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci.* 2017;20:602–611. <https://doi.org/10.1038/nn.4524>.
 - 62 Haag N, Tan EC, Begemann M, et al. Biallelic loss-of-function variants in WDR11 are associated with microcephaly and intellectual disability. *Eur J Hum Genet.* 2021;29:1663–1668. <https://doi.org/10.1038/s41431-021-00943-5>.
 - 63 Kim HG, Ahn JW, Kurth I, et al. WDR11, a WD protein that interacts with transcription factor EMX1, is mutated in idiopathic hypogonadotropic hypogonadism and Kallmann syndrome. *Am J Hum Genet.* 2010;87:465–479. <https://doi.org/10.1016/j.ajhg.2010.08.018>.
 - 64 Castro S, Brunello FG, Sansó G, et al. Delayed puberty due to a WDR11 truncation at its N-terminal domain leading to a mild form of ciliopathy presenting with dissociated central hypogonadism: case report. *Front Pediatr.* 2022;10:887658. <https://doi.org/10.3389/fped.2022.887658>.
 - 65 Globa E, Zelinska N, Shcherbak Y, Bignon-Topalovic J, Bashamboo A, McElreavey K. Disorders of sex development in a large Ukrainian cohort: clinical diversity and genetic findings. *Front Endocrinol.* 2022;13:810782. <https://doi.org/10.3389/fendo.2022.810782>.
 - 66 Uggetti C, Lepellet A, Depp M, et al. cGAS-mediated induction of type I interferon due to inborn errors of histone pre-mRNA processing. *Nat Genet.* 2020;52:1364–1372. <https://doi.org/10.1038/s41588-020-00737-3>.
 - 67 Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci.* 1998;23:198–199. [https://doi.org/10.1016/s0968-0004\(98\)01208-0](https://doi.org/10.1016/s0968-0004(98)01208-0).
 - 68 Oh RY, AlMail A, Cheerie D, et al. A systematic assessment of the impact of rare canonical splice site variants on splicing using functional and in silico methods. *Hum Genet Genom Adv.* 2024;5:100299. <https://doi.org/10.1016/j.xhgg.2024.100299>.
 - 69 Cheerie D, Meserve MM, Beijer D, et al. Consensus guidelines for assessing eligibility of pathogenic DNA variants for antisense oligonucleotide treatments. *Am J Hum Genet.* 2025;112:975–983. <https://doi.org/10.1016/j.ajhg.2025.02.017>.
 - 70 Jaganathan K, Panagiotopoulou SK, McRae JF, et al. Predicting splicing from primary sequence with deep learning. *Cell.* 2019;176:535–548.e24. <https://doi.org/10.1016/j.cell.2018.12.015>.
 - 71 Arriaga TM, Mendez R, Ungar RA, et al. Transcriptome-wide outlier approach identifies individuals with minor spliceopathies. *Am J Hum Genet.* 2025;112:2458–2475. <https://doi.org/10.1016/j.ajhg.2025.08.018>.
 - 72 Xing C, Kanchwala M, Rios JJ, et al. Biallelic variants in RNU12 cause CDAGS syndrome. *Hum Mutat.* 2021;42:1042–1052. <https://doi.org/10.1002/humu.24239>.
 - 73 Nava C, Cogne B, Santini A, et al. Dominant variants in major spliceosome U4 and U5 small nuclear RNA genes cause neurodevelopmental disorders through splicing disruption. *Nat Genet.* 2025;57:1374–1388. <https://doi.org/10.1038/s41588-025-02184-4>.
 - 74 [preprint] Kopajtich R, Smirnov D, Stenton SL, et al. Integration of proteomics with genomics and transcriptomics increases the diagnostic rate of Mendelian disorders. *medRxiv*; 2021, 2021.03.09.21253187 <https://doi.org/10.1101/2021.03.09.21253187>.
 - 75 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.* 2022;109:210–222. <https://doi.org/10.1016/j.ajhg.2021.12.014>.
 - 76 Zhao S, Sinson JC, Li S, et al. The utility of ultra-deep RNA sequencing in Mendelian disorder diagnostics. *Am J Hum Genet.* 2025;112:2578–2590. <https://doi.org/10.1016/j.ajhg.2025.09.013>.
 - 77 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;22:1181–1190. <https://doi.org/10.1038/s41436-020-0780-y>.
 - 78 Nicolas-Martinez EC, Robinson O, Pflueger C, et al. RNA variant assessment using transactivation and transdifferentiation. *Am J Hum Genet.* 2024;111:1673–1699. <https://doi.org/10.1016/j.ajhg.2024.06.018>.
 - 79 Herdy J, Schafer S, Kim Y, et al. Chemical modulation of transcriptionally enriched signaling pathways to optimize the conversion of fibroblasts into neurons. *eLife.* 2019;8:e41356. <https://doi.org/10.7554/elife.41356>.
 - 80 Li S, Zhao S, Sinson JC, et al. The clinical utility and diagnostic implementation of human subject cell transdifferentiation followed by RNA sequencing. *Am J Hum Genet.* 2024;111:841–862. <https://doi.org/10.1016/j.ajhg.2024.03.007>.
 - 81 Janin A, Chanavat V, Rollat-Farnier P-A, et al. Whole MYBPC3 NGS sequencing as a molecular strategy to improve the efficiency of molecular diagnosis of patients with hypertrophic cardiomyopathy. *Hum Mutat.* 2020;41:465–475. <https://doi.org/10.1002/humu.23944>.