

<https://doi.org/10.1038/s41525-025-00502-7>

Cracking rare disorders: a new minimally invasive RNA-seq protocol



Laurenz De Cock^{1,2,3,4,5}, Erika D'haenens^{2,5}, Lies Vantomme^{1,2}, Lynn Backers^{1,2,3}, Aude Beyens^{1,2}, Kathleen BM Claes^{1,2,3}, Griet De Clercq^{1,2}, Robin de Putter², Candy Kumps², Nika Schuermans², Jo Sourbron², Hannes Syryn^{1,2}, Simon Tavernier², Eva Vanbelleghem^{1,2}, Olivier Vanakker^{1,2}, Bart Vandekerckhove^{3,4}, Tim Van Damme², Bert Callewaert^{1,2}, Annelies Dheedene^{1,2}, Sarah Vergult^{1,2,6} ✉ & Björn Menten^{1,2,3,6} ✉

RNA sequencing (RNA-seq) has become key to complementing exome and genome sequencing for variant interpretation. We present a minimally invasive RNA-seq protocol using short-term cultured peripheral blood mononuclear cells (PBMCs) with and without cycloheximide treatment, enabling detection of transcripts subject to nonsense-mediated decay. While broadly applicable, this protocol is particularly suited for neurodevelopmental disorders, as up to 80% of the genes in our intellectual disability and epilepsy gene panel are expressed in PBMCs. Applied to 46 affected individuals and 15 parents, RNA-seq revealed splicing defects in six of nine individuals with splice variants, allowing reclassification of seven variants. Targeted cDNA analysis confirmed aberrant splicing in four individuals but missed intron retention in two. Global analyses (FRASER, OUTRIDER, and monoallelic expression) supported findings but did not yield new diagnoses. We propose a flowchart integrating RNA-seq into diagnostic workflows. Overall, our protocol is easily implementable, captures complex splicing events, and enhances variant classification.

The introduction of exome sequencing (ES) and genome sequencing (GS) in genetic diagnostics has tremendously changed the field in the last decade. Several new disease-causing variants in novel genes have emerged and are still being discovered, increasing the diagnostic yield for Mendelian disorders to approximately 50%. Part of the missing heritability results from the inconclusive classification of variants detected by ES or especially GS, which causes great challenges for clinical interpretation. These variants often require extra (functional) experiments to provide evidence for pathogenicity^{1,2}.

Investigation of the RNA can validate and confirm aberrant expression of specific genes resulting from suspicious variants. Targeted RNA analysis using reverse transcription PCR (RT-PCR) or quantitative reverse transcription PCR (RT-qPCR) has proven its usefulness but requires targeted, designed assays and often fails to reveal the entire spectrum of the molecular consequences caused by a variant. Over the last years, several studies indicated that transcriptome analyses (i.e., RNA-seq) may complement ES and GS by providing an extra layer of information on the effect of a specific variant at the RNA level, providing up to a 7–36% increase in diagnostic yield^{3–9}.

Nevertheless, as pointed out by Rowlands et al., the implementation of RNA-seq in clinical testing comes with significant hurdles¹⁰. Firstly, tissues from affected organs are rarely available, hence, in most cases, transcriptomics can only be performed on clinically accessible tissues (CATs) such as fibroblasts, whole blood, peripheral blood mononuclear cells (PBMCs), lymphoblastoid cell lines (LCLs), or muscle biopsies. Though obtainable in routine clinical care, each CAT has its own advantages and disadvantages regarding the invasiveness of sample collection, processing, and/or culture time. Secondly, compared to data on 'static' DNA, the interpretation of RNA data is much more complicated due to the existence of multiple isoforms per gene. In addition, nonsense-mediated decay (NMD) of aberrant transcripts can mask the underlying event. NMD is a cellular surveillance mechanism that eliminates transcripts containing premature stop codons 50–55 nucleotides upstream of the last exon-exon junction to prevent translation of truncated, potentially harmful proteins¹¹.

The predominant focus in the initial RNA-seq studies pertained to autosomal recessive diseases. Given that both alleles are affected in recessive disorders, the impact on gene expression is anticipated to be more

¹Department of Biomolecular Medicine, Ghent University, Ghent, Belgium. ²Center for Medical Genetics Ghent, Ghent University Hospital, Ghent, Belgium.

³Cancer Research Institute Ghent (CRIG), Ghent, Belgium. ⁴Department of Diagnostic Sciences, Ghent University, Ghent, Belgium. ⁵These authors contributed equally: Laurenz De Cock, Erika D'haenens. ⁶These authors jointly supervised this work: Sarah Vergult, Björn Menten. ✉ e-mail: sarah.vergult@ugent.be; bjorn.menten@ugent.be

pronounced compared to scenarios where only one allele is affected. Haploinsufficiency is often much more difficult to assess at the functional level. Therefore, only a limited number of studies have focused on dominant disorders^{12,13}.

At the Center for Medical Genetics Ghent, we established a minimally invasive RNA-seq protocol for transcriptome analysis on short-term cultured PBMCs in individuals with a Mendelian neurodevelopmental disease. FRASER¹⁴ and OUTRIDER¹⁵ were used as bioinformatics tools for the detection of aberrant splicing and outlier detection, respectively.

We show the added value of RNA-seq analysis to (re-) classify genomic variants and compare the results with targeted RT-PCR-based Sanger sequencing, highlighting the added value of RNA-seq. In addition, we illustrate the difficulty of interpreting RNA-seq results in a clinical setting and propose a decision tree to estimate whether RNA-seq can be of added value.

Results

Up to 80% of our ID/epi panel genes are expressed in fibroblasts and PBMCs

The Center for Medical Genetics Ghent conducts ES-based gene panel analysis for various Mendelian disorders. We assessed the expression of all genes represented in the 35 disease gene panels (see CMGG for the lists¹⁶), in four different CATs, i.e., cultured fibroblasts ($n = 10$, FIB), whole blood ($n = 9$, WB), peripheral blood mononuclear cells ($n = 10$, PBMCs), and lymphoblastoid cell lines ($n = 10$, LCLs) (Supplementary Tables 3, 4, and 5). For each gene panel, the percentage of genes with expression levels of >1 TPM, was quantified per CAT. Fibroblasts expressed most genes, with an average of 72.2% of genes/panel, followed by whole blood (69.4%), PBMCs (69.4%), and LCLs (64.3%), thereby aligning with previous studies^{7,9,10,17}. When focusing on the larger gene panels (i.e., >100 genes), the difference in expressed genes per panel became smaller among fibroblasts 70.6%, PBMCs 68.9%, and whole blood 68.8%), while LCLs (63.7%) performed less. Interestingly, for our Mendeliome gene panel (4732 disease-associated

genes in version 6), PBMCs had the most expressed genes (63.7%), although only slightly better than whole blood (63.5%) and fibroblasts (62.5%). Moreover, for the ID&Epi gene panel (version 6, $n = 1689$ genes), we observed that fibroblasts (80.4%) and PBMCs (79.7%) express more genes compared to whole blood (77.3%) and LCLs (75.4%).

Subsequently, we investigated whether the expression patterns of the genes in our Mendeliome and ID&Epi panel were similar between different CATs. The Upset Plots in Fig. 1A, B demonstrate that, respectively, 64.0% (Mendeliome) and 79.6% (ID&Epi) of genes expressed at >1 TPM in one CAT are also expressed in at least one other CAT. Moreover, Fig. 1A, B also show that, respectively, 50.5% and 70.4% of the genes in our Mendeliome and ID&Epi gene panel are expressed in all CATs.

Considering the results above, the less invasive nature, and the shorter culture time required for PBMCs compared to fibroblasts, we decided to proceed with PBMCs as the preferred cell type for our subsequent experiments.

Cycloheximide treatment can successfully inhibit NMD in CATs

Since NMD is anticipated to occur for protein-truncating variants (PTVs) residing >55 bp from the last exon boundary, we evaluated the use of NMD inhibitors in our workflow. The effectiveness of two NMD inhibitors^{18–20}, i.e., PUR and CHX, was determined using four commercially available LCLs harboring known and confirmed PTVs in *XPC*, *PPT1*, or *TPP1*, in a homozygous or (compound) heterozygous state (Table 1). To test the efficacy, custom RT-qPCR primers to target the wildtype and/or NMD transcripts (“Materials and methods”, Supplementary Table 2, Fig. 1C, and Supplementary Fig. 1) were designed. The untreated LCLs harboring a truncating *XPC* variant in a heterozygous or homozygous state both show a clear drop in expression of *XPC* (Fig. 1C). As expected, the decrease in expression level was most pronounced in the LCL with the homozygous *XPC* variant. Treatment with PUR and CHX resulted in increased *XPC* expression in both cell lines, indicating NMD was (partially) inhibited by PUR, as well as CHX. However, CHX treatment resulted in an overall higher

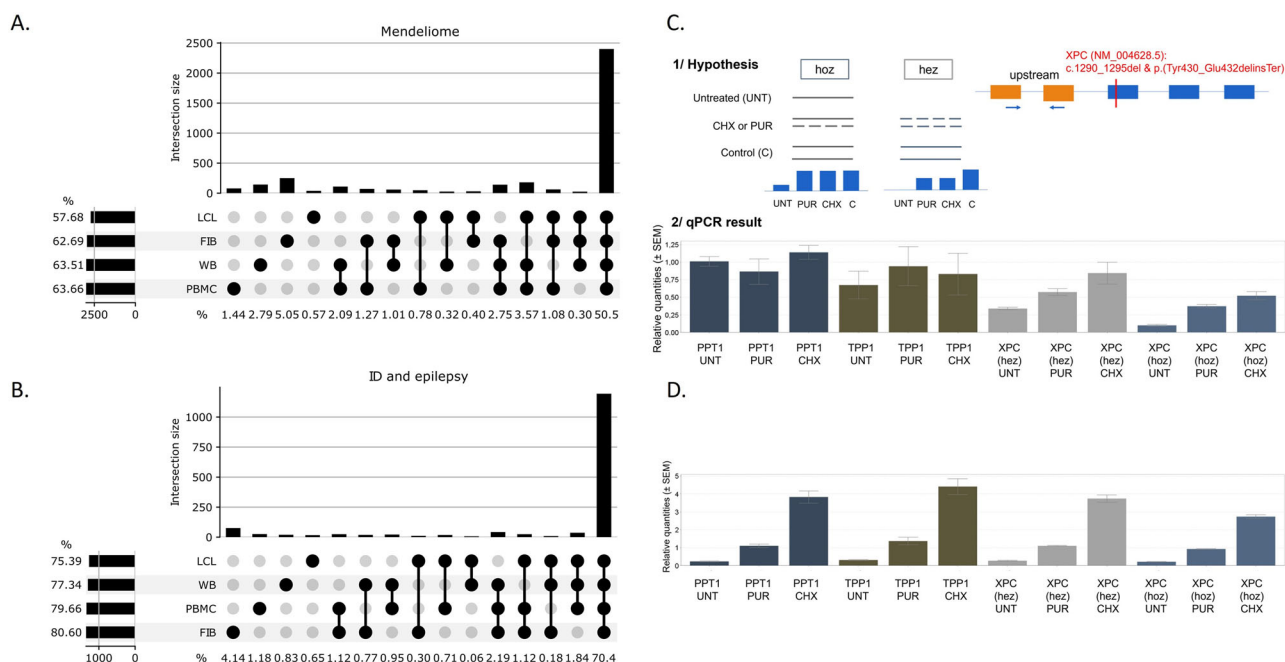


Fig. 1 | Gene expression in four different CATs and upon PUR or CHX treatment in LCLs. A, B Upset plot of the genes within **A** the Mendeliome gene panel ($n = 4732$) and **B** the ID and epilepsy gene panel ($n = 1689$) and their expression (>1 TPM) in four different CATs. **C, D** Relative expression levels of CXPC (NM_004628.5) and the NMD sensitive *SRSF2* transcript (ENST00000452355.7) in LCL samples with, respectively, *PPT1* (darkblue bars), *TPP1* (brown bars), and *XPC* (heterozygous variant: gray bars; homozygous variant: blue bars) protein

truncating variants. For each sample, the relative expression is given in the untreated situation, after 6 h treatment with puromycin, and after 6 h treatment with cycloheximide. The theoretical effect of the heterozygous (left) and homozygous (right) PTV in *XPC* (XPC (NM_004628.5): c.1290_1295del & p.(Tyr430_Glu432delinsTer) is given in (C) based on the position of the qPCR primers (as indicated by the blue convergent arrows). UNT untreated, CHX treated with cycloheximide, PUR treated with puromycin, C control (i.e., samples without *XPC* PTV).

Table 1 | Validation samples to verify the effect of NMD inhibitors

	Gene	Variants	Inheritance	Coriell ID	PubMed ID/family or proband nr.	Average expression LCLs
Proband	<i>XPC</i>	<i>XPC</i> (NM_004628.5):c.1290_1295del (p.Tyr430_Glu432delinsTer) exon 9/16	Homozygous	GM14870	16081512/XP69TMA	54.7
Mother	<i>XPC</i>	<i>XPC</i> (NM_004628.5):c.1290_1295del (p.Tyr430_Glu432delinsTer) exon 9/16	Heterozygous	GM15704	16081512/XP69TMA	
Proband	<i>TPP1</i>	<i>TPP1</i> (NM_000391.4):c.622C > T (p.Arg208Ter) exon 6/13	Heterozygous	GM20377	11339651/not mentioned	66
	<i>TPP1</i>	<i>TPP1</i> (NM_000391.4): c.376_380 + 6del exon 4/13	Heterozygous			
Proband	<i>PPT1</i>	<i>PPT1</i> (NM_000310.3):c.451C > T (p.Arg151Ter) exon 5/9	Homozygous	GM16083	9664077/ref no. 9 (UT8-01)	72

expression level of *XPC* in both LCLs (Fig. 1C). Similar results were obtained for *TPP1* and *PPT1* expression levels, respectively, in the LCL harboring compound heterozygous *TPP1* variants (Supplementary Fig. 1A, B) and in the LCL harboring a heterozygous *PPT1* nonsense variant (Supplementary Fig. 1C, D).

In these four LCLs, specific genes were targeted to determine the effectiveness of NMD inhibitors. However, no *prior* knowledge of possible NMD events will be available in most cases. Therefore, an endogenous control of NMD inhibition was included. Recently, Häuser et al.²¹ included *SRSF2* in their protocol to enable targeted confirmation of mis-splicing in a candidate variant suspected to undergo NMD. *SRSF2* encodes the serine/arginine-rich splicing factor 2 that autoregulates its expression by promoting splicing events from an NMD-insensitive to an NMD-sensitive transcript²¹. For the four commercial LCLs, we assessed the expression level of the NMD-sensitive *SRSF2* transcript in untreated and PUR or CHX-treated LCLs. In line with the results for *XPC*, *TPP1*, and *PPT1*, CHX treatment resulted in a higher expression level of the *SRSF2* NMD-sensitive transcript compared to PUR (Fig. 1D and Supplementary Fig. 2B, C). Therefore, we chose CHX as an NMD inhibitor for further RNA-seq experiments and confirmed the validity of the *SRSF2* transcripts as an internal control in LCLs, PBMCs, and fibroblasts (Supplementary Fig. 3A). To further corroborate the effectiveness of NMD inhibition in 46 CHX-treated samples, the percentage of *SRSF2* reads encompassing exon 3 of the *SRSF2* NMD-sensitive transcript was calculated. We observed an apparent increase in exon 3 spanning reads for all CHX-treated samples (8.58%) compared to their untreated counterpart (4.55%), except for individual 11, indicating CHX treatment was effective (Supplementary Fig. 3B).

RNA-seq outperforms in silico prediction tools and targeted cDNA analysis in capturing complex splicing events

To evaluate the effectiveness of RNA-seq in detecting (complex) splicing events, we selected nine individuals harboring either a VUS or a (likely) pathogenic variant that may potentially affect splicing, including an intronic splice donor variant (individuals 1–3), an intronic splice acceptor variant (individuals 4–6) or a synonymous variant near the splice donor region (individuals 7–9) (Table 2).

TPM values of each gene of interest (i.e., *WDR26*, *NIPBL*, *SMC1A*, *STAG1*, *SMARCC2*, *DDX3X*, *SETD1A*, *CNOT1*, and *PHF8*) ranged from 14 to 157, indicating sufficient gene expression in PBMCs. (Supplementary Table 4).

RNA-seq revealed aberrant splicing in six out of nine individuals (1, 2, and 4–7). An effect on splicing was observed for two of the three intronic splice donor region variants. In both individual 1 (*WDR26* variant, Fig. 2A) and individual 2 (*NIPBL* variant, Fig. 3A), exon skipping was observed due to the variant (Figs. 2B–D and 3B–D). In individual 1, 27% of the reads contained an alternative junction, skipping *WDR26* exon 11. In individual 2, a skip of *NIPBL* exon 2 was observed in 14% of the reads (Fig. 3D). The *SMC1A* intronic splice region variant in individual 3 did not result in aberrant splicing (Supplementary Fig. 4C, D).

The splice acceptor region variants in individuals 4, 5, and 6 all resulted in aberrant splicing. In individual 4, a skip of *STAG1* exon 6 was detected (Fig. 4C, D), while in individual 5, a complex splicing pattern consisting of a

skip of exon 22 and retention of intron 21 of *SMARCC2* was observed (Fig. 5C, D). The intronic *DDX3X* splice region variant in individual 6 resulted in a cryptic exonic acceptor (Fig. 6C, D).

Of the three synonymous variants, only the *SETD1A* synonymous variant in individual 7 resulted in a complex splicing pattern with both a skip of exon 5 and retention of intron 5 (Fig. 7C, D). In individual 8, in 3% of the junction reads from the splice donor at *CNOT1* exon 16, a skip of exon 17 was observed. This skip was also observed in <1% of reads of three randomly inspected individuals (not harboring this variant) (Supplementary Fig. 5C, D). This suggests that the aberrant splicing in individual eight is not due to the variant. For individual 9, the synonymous *PHF8* variant did not result in aberrant splicing (Supplementary Fig. 6C, D).

Importantly, aberrant splicing was detectable in both untreated and CHX-treated PBMCs. However, for individuals 1 and 4, a higher percentage of aberrant splicing events was observed after CHX treatment (respectively 49% vs 27% in individual 1 and 44% vs 31% in individual 4), indicating activation of the NMD pathway (Table 2 and Figs. 2D and 4D). In the absence of other SNVs in these individuals, VAFs could not be compared between untreated and CHX-treated PBMCs. Nevertheless, using the expression levels of the *SRSF2* NMD-sensitive and insensitive transcripts, we could ensure that CHX treatment was successful (Supplementary Fig. 3B).

In contrast, cDNA analysis detected aberrant splicing in only four individuals (1 (Fig. 2B), 4 (Fig. 4B), 5 (Fig. 5B), and 7 (Fig. 7B), and Table 2). Moreover, in individuals 5 and 7, cDNA analysis failed to detect the intron retention due to the difficult-to-interpret Sanger results. In the other five individuals (individuals 2, 3, 6, 8, and 9), no clear aberrant splicing events were observed with targeted analysis (Supplementary Figs. 4B–6B and Table 2). While RNA-seq showed skipping of the second exon of *NIPBL* in individual 2, this effect was challenging to decipher with targeted cDNA analysis due to high background noise (Fig. 3B).

For all nine splice region variants, the scores of three in silico splicing prediction algorithms (i.e., ADA^{22,23}, RF²³, and SpliceAI²⁴) were compared with the results obtained from RNA-seq (Supplementary Table 6). According to the guidelines, respectively, an ADA or RF-score of ≥ 0.5 and a SpliceAI-score of $\Delta \geq 0.20$ can be assumed as highly suggestive for an effect on splicing^{24,25}. Based on these values, respectively seven, seven, and six out of the nine variants were predicted by ADA, RF, and SpliceAI to affect splicing (Supplementary Table 6). All three tools correctly predicted that the *SMC1A* and *PHF8* variants in individuals 3 and 9 did not cause a splice event.

RNA-seq improves variant classification for splice region variants

Following RNA-seq, we reclassified the variants for all nine targeted cases according to the guidelines by Walker et al.²⁶ For individuals 1, 2, and 7, we could now apply PVS1, reclassifying the variants to class 5. For individual 5, in-frame aberrant splicing allowed to application of PS4_PM, switching the variant from class 3 to 4. For individual 4, the *STAG1* variant remained class 5 after RNA-seq. Although we observed an effect on splicing for the *DDX3X* variant in individual 6, the variant remained class 3 as RNA-seq indicated a cryptic splice acceptor site 8 bp upstream of the variant, resulting in an in-

Table 2 | Observations on the cDNA and RNA level for individuals 1–9 with splice (region) variants

Individual	Gene	c. position	Inheritance	Sanger sequencing	Targeted RNA-seq analysis				FRASER		Variant classification		Class of variant before cDNA analysis/ RNA-seq analysis/ (Walker et al., 2023)	Criteria after cDNA analysis/ RNA-seq analysis/ (Walker et al., 2023)	Class of variant after cDNA analysis/ RNA-seq
					Untreated	CHX treatment	% Alternative junction reads untreated	% Alternative junction reads CHX	Reading frame	p-adj	deltaPSI	Criteria before cDNA analysis/ RNA-seq			
1	WDR26	c.1644 + 5G > A	De novo	Exon 11 skip	Exon 11 skip	Exon 11 skip	27%	49%	Frameshift	1.2322E-09	0.33	PM2, PS3, PS2_PP	3	PM2, PVS1, PS2_PP	5
2	NIPBL	c.64 + 6T > C	De novo	No effect	Exon 2 skip	Exon 2 skip	14%	15%	Start-loss (no alternative start codon)	0.46345	0.13	PM2, PP4_PM, PP3_PS2	4	PM2, PP4_PM, PVS1_PS2	5
3	SMC1A	c.1254 + 4 G > A	Inherited from the mother (X-linked)	No effect	No effect	No effect						BP4	3	BP4, BP7	2
4	STAG1	c.395-2A > T	Inherited from the father	Exon 6 skip	Exon 6 skip	Exon 6 skip	31%	44%	Frameshift	0.074355	0.29	PM2, PVS1	5	PM2, PVS1	5
5	SMARCC2	c.2186-3C > G	De novo	Not interpretable	Exon 22 skip intron retention	Exon 22 skip intron retention	Exon 22 skip: 38%	Exon 22 skip: 45%	In-frame	4.7782E-09	0.25	PM2, PP3, PS2_PM	3	PM2, PM4_PM, PS2_PM	4
6	DDX3X	c.1770-2A > G	Inherited from the mother (X-linked)	Not interpretable	Cryptic exonic acceptor	Cryptic exonic acceptor	37%	40%	In-frame	2.6358E-08	0.2	PM2, PVS1_PM	3	PM2, PM4_PM	3
7	SETD1A	c.639G > A	Inherited from the mother	Exon 5 skip	Exon 5 skip intron retention	Exon 5 skip intron retention	Exon 5 skip: 27%	Exon 5 skip: 33%	Frameshift	6.2414E-10	0.28	PM2, PP3, PP1_PM	3	PM2, PVS1, PP1_PM	5
8	CNOT1	c.2130G > A	Inherited from the father	No effect	No effect	No effect						PM2, PP3	3	PM2, BP7_strong	2
9	PHF8	c.1140C > T	Inherited from the mother (X-linked)	No effect	No effect	No effect						BP4	3	BP4, BP7	2

RNA-seq analysis led to the reclassification of seven variants.

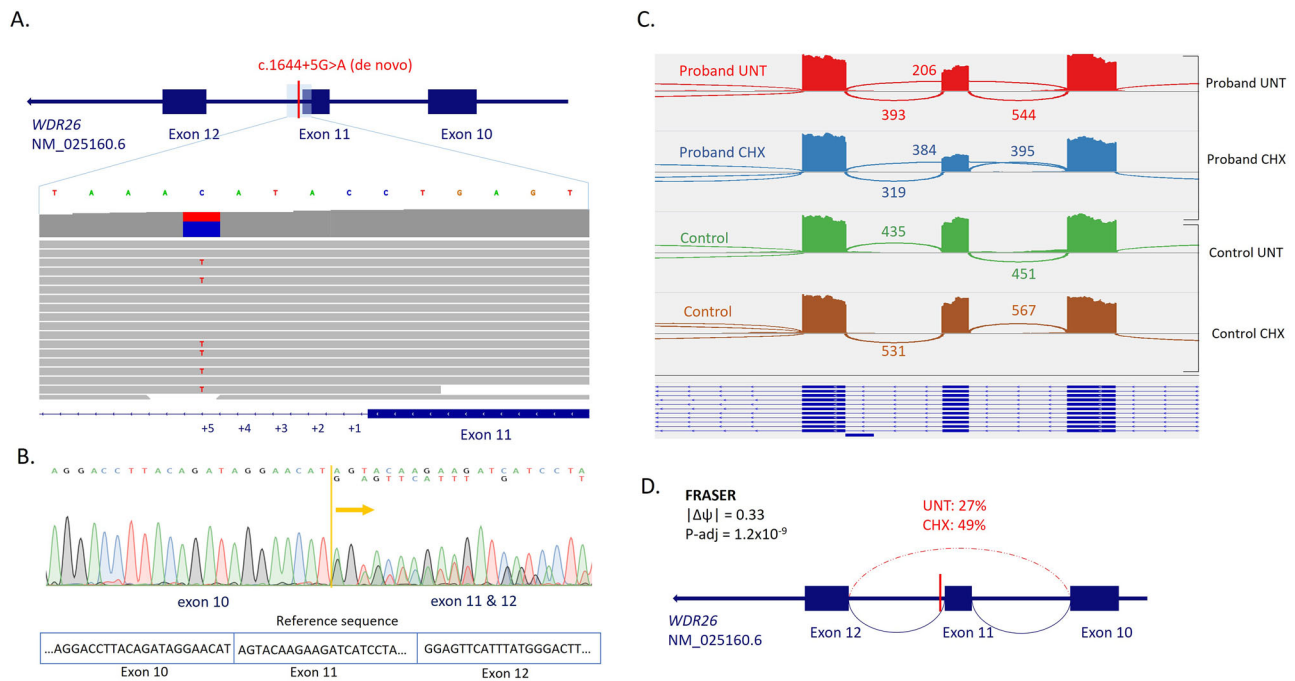


Fig. 2 | Effect on splicing of intronic *WDR26* variant: *WDR26* (NM_025160.6): c.1644 + 5G > A. **A** Zoomed in IGV plot of exome data showing the heterozygous +5 intronic variant downstream of exon 11. **B** Sanger profile showing the heterozygous exon 11 skip. **C** Sashimi plot of exons 10–12 of *WDR26* in untreated and treated PBMCs of the affected individual (red = untreated (UNT) and blue = CHX treated (CHX)) and a control PBMC sample, both untreated (green) and CHX

treated (brown). The skipping of exon 11 can clearly be observed in both the untreated and CHX-treated PBMCs. **D** Schematic representation of the splicing effect(s) caused by the variant. While in the untreated PBMCs, the exon skip was observed in 27% of the splicing events from the exon 10 donor site, the aberrant splicing was observed in approximately 50% of the splicing events at that site in the CHX-treated PBMCs.

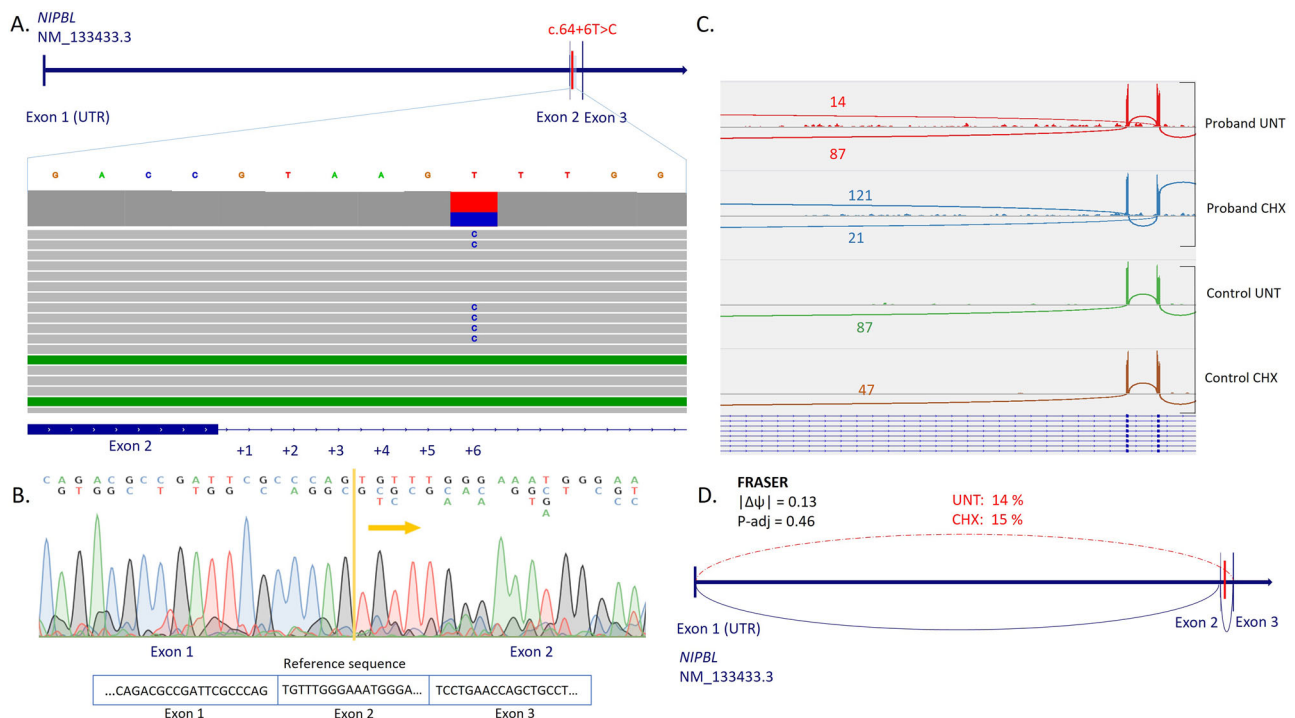


Fig. 3 | Effect on splicing of intronic *NIPBL* variant: *NIPBL* (NM_133433.3): c.64 + 6T > C. **A** Zoomed in IGV plot of exome data showing the heterozygous +6 intronic variant downstream of exon 2. **B** Sanger profile with background noise, skip of exon 2 is suggestive but difficult to interpret due to the background noise. **C** Sashimi plot of exons 2–3 of *NIPBL* in untreated and treated PBMCs of the affected individual (red = untreated (UNT) and blue = CHX treated (CHX)) and a control

PBMC sample, both untreated (green) and CHX treated (brown). The skipping of exon 2 can be observed in both the untreated and CHX-treated PBMCs. **D** Schematic representation of the splicing effect(s) caused by the variant. The exon skip was observed in 14% and 15% of the splicing events starting from the exon 1 donor site in the untreated and CHX-treated PBMCs.

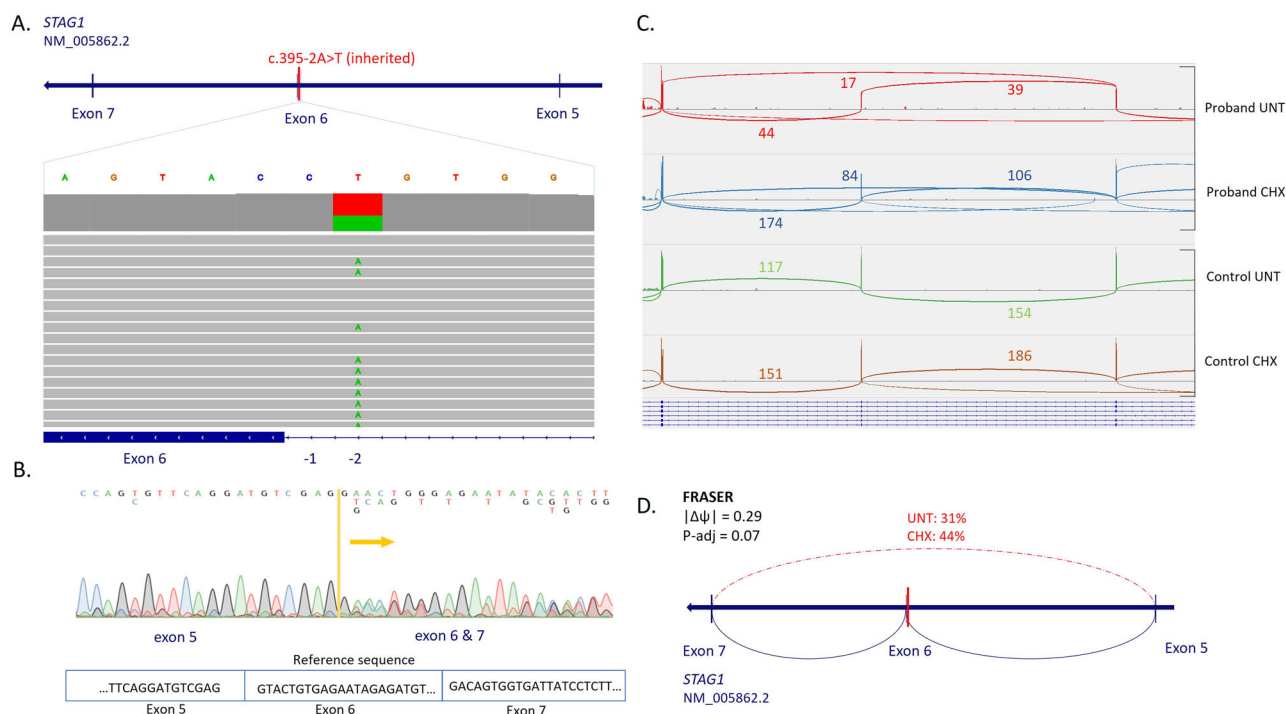


Fig. 4 | Effect on splicing of intronic *STAG1* variant: *STAG1* (NM_005862.2): c.395-2A > T. **A** Zoomed in IGV plot of exome data showing the heterozygous-2 intronic variant upstream of exon 6. **B** Sanger profile showing the heterozygous exon 6 skip. **C** Sashimi plot of exons 5–7 of *STAG1* in untreated and treated PBMCs of the affected individual (red = untreated (UNT) and blue = CHX treated (CHX)) and a control PBMC sample, both untreated (green) and CHX treated (brown). The

skipping of exon 6 can clearly be observed in both the untreated and CHX-treated PBMCs. **D** Schematic representation of the splicing effect caused by the variant. While the effect was observed in 31% of the splicing events from the exon 5 donor site in the untreated PBMCs, the aberrant splicing was observed in approximately 44% of the splicing events at that site, in the CHX-treated PBMCs.

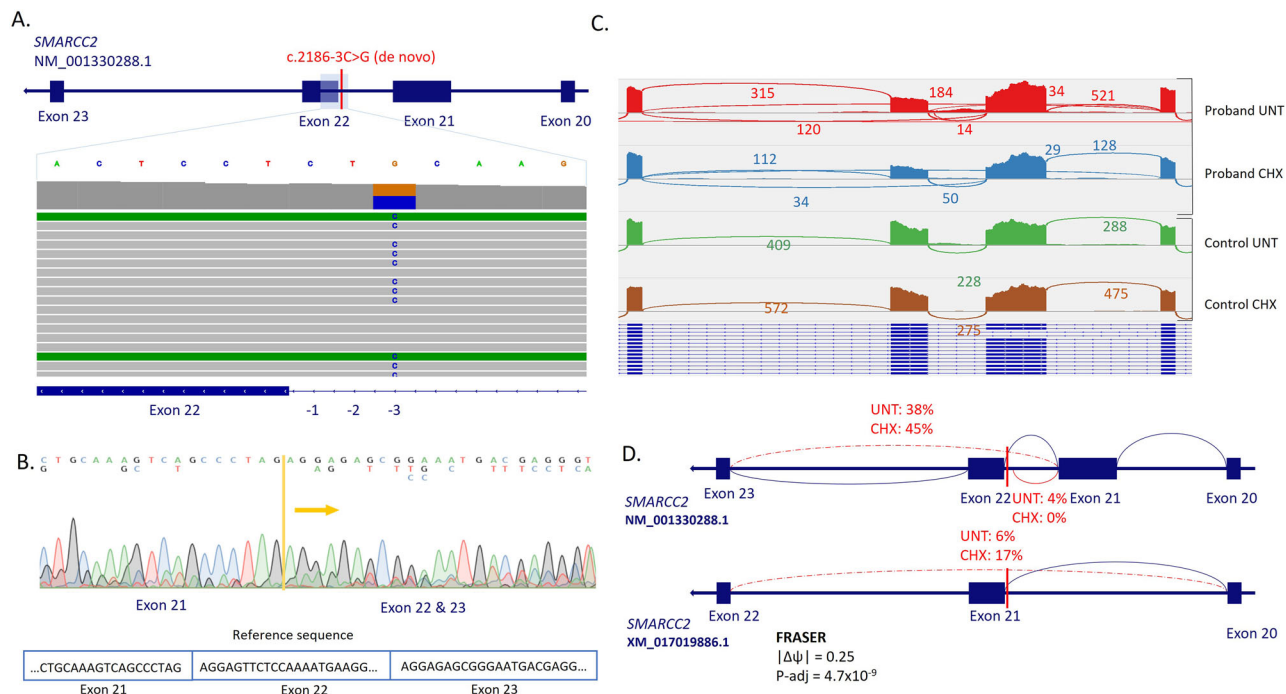


Fig. 5 | Effect on splicing of intronic *SMARCC2* variant: *SMARCC2* (NM_001330288.1): c.2186-3C > G. **A** Zoomed in IGV plot of exome data showing the heterozygous-3 intronic variant upstream of exon 22. **B** The Sanger profile is difficult to interpret due to the background noise, exon 22 skip is suggestive. **C** Sashimi plot of exons 20–23 of *SMARCC2* in untreated and treated PBMCs of the affected individual (red = untreated (UNT) and blue = CHX treated (CHX)) and a control PBMC sample, both untreated (green) and CHX treated (brown). Different split reads are observed, for the canonical transcript (NM_001330288.1): skip of

exon 22, as well as intron retention in intron 21 in untreated PBMCs. Intron retention is not present in CHX-treated cells. **D** Schematic representation of the splicing effects caused by the variant. Transcript NM_001330288.1: from the exon 21 splice donor, an exon 22 skip was observed in 45% of the reads of CHX-treated PBMCs vs 38% for the untreated cells. Intron retention was only observed in the untreated cells in 4%. Transcript XM_017019886.1: from the exon 20 donor site, 6% of splice events appeared to skip exon 21 in untreated PBMCs, and 17% in cells treated with CHX.

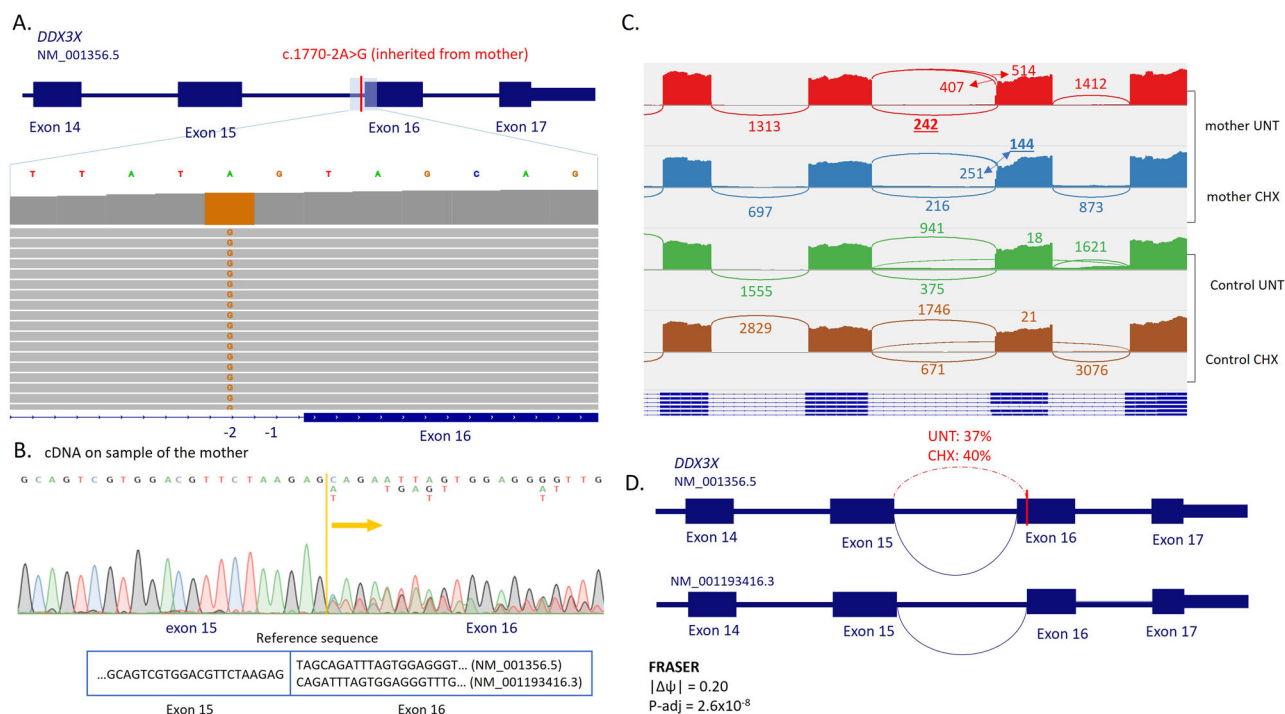


Fig. 6 | Effect on splicing of intronic *DDX3X* variant: *DDX3X* (NM_001356.5): c.1770-2 A > G. A Zoomed-in IGV plot of exome data of the proband showing the hemizygous-2 intronic variant downstream of exon 16. **B** Sanger sequencing on cDNA from the mother, a heterozygous carrier of the *DDX3X* variant: Normal sequence until the end of exon 15, from then on, multiple peaks per nucleotide, the exact effect is difficult to decipher. **C** Sashimi plot of exons 14–17 of *DDX3X* in untreated and treated PMBCs of the mother (red = untreated (UNT) and blue = CHX-treated (CHX)) and a control PMBC sample, both untreated (green) and CHX-treated (brown). From the exon 15 donor, three arches can be distinguished. A cryptic splice acceptor site is recognized in both untreated and treated PMBCs from the patient's mother (junctions indicated in bold). In the controls, two arches starting from the exon 15 donor are present. **D** Schematic representation of the splicing effects caused by the variant. A cryptic acceptor was recognized in both untreated and CHX-treated PMBCs, in 37% and 40% of the reads, respectively. In an alternative transcript (NM_001193416.3), exon 16 is shorter and starts 3 bp later than the canonical transcript.

frame deletion of two serine amino acids not located in a critical region of the protein.

In the absence of clear aberrant splicing, the variants detected in individuals 3, 8, and 9 were reclassified from VUS to likely benign (class 2). In total, RNA-seq allowed for to reclassify of seven of nine splice variants (Table 2).

The findings of targeted RNA-seq analysis were recapitulated in global splicing analysis with FRASER

We used FRASER to determine aberrantly spliced genes transcriptome-wide in individuals 1–9 (Supplementary Table 1). When applying standard filtering (i.e., delta percent-splice-in (PSI) > 0.15 and adjusted p -value ($p\text{-adj}$) $\leq 0.05^{27}$), FRASER indicated a significant effect on splicing in only four (individuals 1 and 5–7) out of the six individuals in which manual analysis of RNA-seq data indicated aberrant splicing (1, 2, and 4–7). In individual 4, FRASER yielded a deltaPSI of 0.29 for the *STAG1* aberrant splicing ($p\text{-adj}$ was 0.074). In individual 2, the effect on splicing of the *NIPBL* variant was not identified as significant by FRASER (deltaPSI of 0.13 and adjusted p -value of 0.46).

We subsequently compared the deltaPSI values from the untreated samples of these nine individuals vs the CHX-treated samples. In line with manually analyzed data, we observed no difference for individuals 5–7 (Supplementary Table 6) and observed a more significant effect on splicing for the *WDR26* and *STAG1* variants in the CHX-treated samples (individuals 1 and 4, respectively). Remarkably, although the deltaPSI value for the aberrant splicing in individual 2 remained 0.13 in the CHX-treated sample, this was now called significant with a $p\text{-adj}$ of 0.026.

Although FRASER detected all abnormal splicing events identified through the manual RNA-seq analysis, only in four out of the six, the aberrant splicing reach significance.

FRASER analyses did not result in novel causal variants in our cohort

We performed a blind analysis to detect aberrant splicing in 37 individuals without a suspicious splicing variant based on exome analysis. Using FRASER with standard filtering (deltaPSI > 0.15 and $p\text{-adj} \leq 0.05^{27}$), on average, 18.67 and 20.42 splicing events were observed in the untreated and CHX-treated samples, with an average overlap of 13.7%. Most splicing events were detected in complex genomic loci, such as the HLA region. We then filtered the results from FRASER for genes included in the ID&Epi gene panel, taking only deltaPSI > 0.15 into account. This identified 2.38 aberrant splicing events in the untreated samples and 3.02 in the CHX-treated ones.

In individual 10, we identified a splicing event that involved a pseudo-exon within intron 15 of *BLM* (Supplementary Fig. 7A, B). This is consistent with earlier findings by Backers et al., who identified a deep intronic *BLM* variant c.3020-258 A > G, creating a novel donor splice site²⁸. For the remaining 36 affected individuals, we did not detect any aberrant splicing events related to the phenotype.

Expression outlier analysis with OUTRIDER reveals insights into structural variants, but does not result in novel genetic diagnoses

When taking a p -adjusted value of <0.05 into account, OUTRIDER identified on average 1.6 outliers in untreated samples (range 0–30) and 1.5 outliers (range 0–25) for the CHX-treated samples.

For the nine cases with known splice region variants, we did not detect outlier expression for the splice-affected gene nor in other genes that could be related to the phenotype of these individuals. Despite the compound heterozygous class 5 variants in *BLM* (NM_000057.4), c.3020-258A > G and a nonsense variant: c.3379C > T, p.(Gln1127Ter) in individual 10, no aberrant expression of *BLM* was observed in PMBCs (Supplementary Fig. 7A–C). Similarly, the class 4 frameshift variant affecting *CNOT3* (*CNOT3*

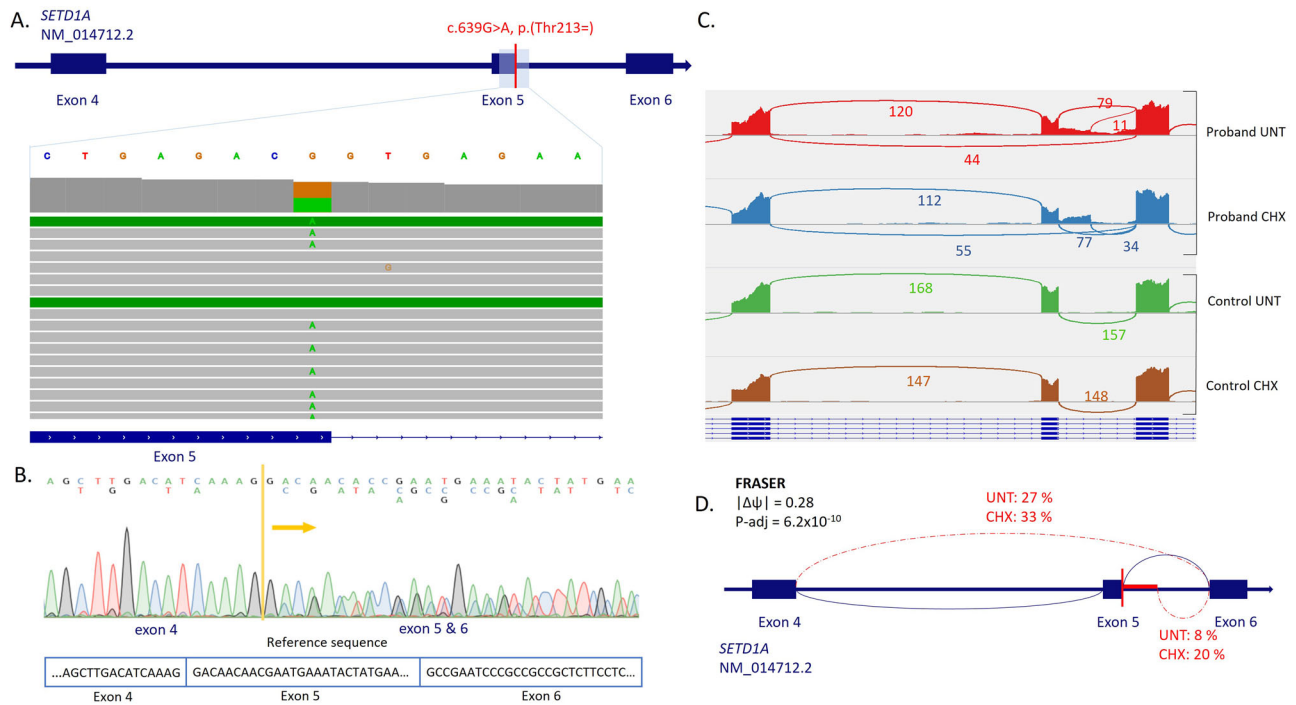


Fig. 7 | Effect on splicing of exonic *SETD1A* variant: *SETD1A* (NM_014712.2): c.639G>A, p.(Thr213=). **A** Zoomed in IGV plot of exome data showing the heterozygous c.639G>A in exon 5. **B** Sanger's profile is indicative of exon 5 skip. **C** Sashimi plot of exons 4–6 of *SETD1A* in untreated and treated PBMCs of the affected individual (red = untreated (UNT) and blue = CHX treated (CHX)) and a control PBMC sample, both untreated (green) and CHX treated (brown). The skipping of exon 5 can clearly be observed in both the untreated and CHX-treated

PBMCs. In addition, intron 5 retention can be observed. **D** Schematic representation of the splicing effects caused by the variant. The exon 5 skip effect was observed in 27% and 33% of the splicing events starting from the exon 4 donor site in the untreated and CHX-treated PBMCs, respectively. A cryptic splice donor was also observed, resulting in intron retention in 8% and 20% of the splicing events starting from exon 6 in untreated cells and cells treated with CHX.

(NM_014516.4): p.(Ser245GlnfsTer8)dn, Supplementary Table 1) in individual 12, did not result in aberrant expression of *CNOT3* (Supplementary Fig. 7D) in untreated PBMCs, nor were other relevant gene expression outliers detected.

In four samples with known copy number variants (CNVs) (individual 13–16, Supplementary Table 1), OUTRIDER identified aberrant expression ($p\text{-adj} < 0.05$) for several genes located within the CNV (Supplementary Figs. 8–11). Strikingly, OUTRIDER gave a $p\text{-adj} = 1$ for almost all genes in all samples. When we used a very lenient cut-off (i.e., $p\text{-adj} < 1$), we identified, on average, 5.0 gene expression outliers when excluding the samples with larger CNVs and 9.9 when taking them into account. For the CHX-treated samples, we counted, on average, 12.5 outliers ($p\text{-adj} < 1$), ranging from 0 to 283.

When only considering genes from the ID&Epi gene panel, an average of 0.08 significant gene expression outliers were identified (ranging from 0 to 3) ($p\text{-adj} < 0.05$), while using the $p\text{-adj} < 1$ cut-off, on average, 0.93 gene expression outliers could be identified.

In individual 17, diagnosed with Mowat-Wilson syndrome (#OMIM 235730) due to a de novo translocation disrupting *ZEB2* on chromosome 2 (46, XY, t(2;21)(q22;q21)dn (Supplementary Table 1), OUTRIDER did not detect aberrant expression of *ZEB2*, although the PBMCs of individual 17 exhibited the lowest expression of *ZEB2* in our cohort (Supplementary Fig. 7E).

Individual 11 was diagnosed with Beckwith-Wiedemann syndrome (#OMIM 130650) due to a loss of maternal methylation of the imprinting center 2 (IC2) on chromosome 11p15.2 (Supplementary Table 1), probably affecting *KCNQ1OT1*, *KCNQ1*, and *CDKN1C* expression^{29,30}. Though *KCNQ1OT1* has a very low expression in all CATs (<0.5 TPM), *KCNQ1* is clearly expressed in whole blood and PBMCs, and *CDKN1C* has the highest expression in PBMCs (Supplementary Tables 4 and 5). Nevertheless, OUTRIDER analysis did not reveal aberrant expression of these genes in individual 11.

In individuals 16–46, no clear aberrant expression could be observed in relation to the phenotype.

Monoallelic expression analysis did not result in novel genetic diagnoses

As stated by Yépez et al.⁹, mono-allelic expression (MAE) refers to the phenomenon in which, at heterozygous loci, only one of the two alleles is expressed. MAE can result from transcriptional silencing or post-transcriptional degradation of one allele and may arise due to genetic or epigenetic mechanisms^{9,11,31}. For individuals 1–9 with known splice region variants, no MAE of the affected gene was observed in six individuals (2, 4, and 6–9). In comparison, the remaining three individuals (1, 3, and 5) did not have any additional SNVs in the affected gene for MAE analysis. In individual 10, biallelic *BLM* expression was clearly observed (Supplementary Fig. 12C). For individual 11 with BWS, we manually curated the *KCNQ1OT1*, *CDKN1C*, and *KCNQ1* genes. All three genes had low expression in the patient's PBMCs, respectively 0.38, 0.44, and 1.54 TPM. Consequently, only *KCNQ1* could be evaluated for MAE and showed biallelic expression (Supplementary Fig. 12A), which is in line with the analysis of 13 other informative individuals without BWS (Supplementary Fig. 12B). For individual 12, biallelic expression of *CNOT3* in a near 50–50 ratio was observed in the untreated, as well as in the CHX-treated sample (Supplementary Fig. 12D). For Individual 17, where *ZEB2* is disrupted by a de novo translocation, we observed biallelic expression in a 90–10 ratio in the untreated sample, but not in the CHX-treated sample (Supplementary Fig. 12E).

In the next step, we determined the MAE transcriptome-wide in our cohort. Unfortunately, we did not have ES variant data for individuals 10, 11, 17, and 46 to perform this analysis (Supplementary Table 1). The total number of MAE events per individual in the untreated samples varied between 269 and 647, with an average of 405.9 events (Supplementary Table 7).

When relating these MAE events to the gene level, the number of genes with MAE ranged from 183 to 520, and on average, 252.5 genes presented with MAE. This is in concordance with previously reported numbers⁹. In the treated samples, the total number of MAE events fluctuated between 284 and 628, with an average of 398.14, while for the gene level, the range was 183 to 512, with an average of 242.7. The overlap between the treated and untreated samples was on average 43.13%.

Subsequently, MAE events were filtered for those occurring in genes from the ID and Epi gene panel. On average, 19.9 ID and Epi genes are present with MAE in our patient cohort. As we observed that genes such as *FANCD2*, *ATXN2*, and *TUBB2A* presented with MAE events in multiple patients, we limited our MAE analysis further to ID and Epi genes that were unique for each patient. In this manner, we had on average 3.6 ID and Epi genes with MAE per patient.

For the nine individuals with known splice region variants, we detected MAE events in zero to nine unique ID and Epi genes (Supplementary Table 7). However, none of these genes were related to the reported phenotypes (Supplementary Table 1). Similarly, for the four individuals with known unbalanced structural variants (i.e., 13 – 16), none of the ID and Epi genes with MAE events were related to the phenotype. For the remaining 29 individuals, we also did not detect any ID and Epi genes with MAE events that were related to the phenotype.

We propose a flow diagram and decision tree for combined ES/WGS and RNA-seq analysis

Our diagnostic RNA-seq workflow (Fig. 8A) and decision tree to assess aberrant splicing (Fig. 8B), aberrant expression (Fig. 8C), and monoallelic expression (Fig. 8D) are shown in Fig. 8.

1. Aberrant splicing

When a class 3 or 4 variant that may affect splicing is detected, a gene-specific analysis (section 2/) can be proposed if the gene is adequately expressed in a CAT. If a relevant CAT is obtained, but RNA seq analysis by FRASER does not show a splice event for the specific gene, inspect sashimi plots and coverage in IGV manually. Conversely, if FRASER identifies a splice event, follow the decision tree outlined in section 1 of Fig. 8B, beginning from the step 'Evaluate splice pattern in IGV'.

If no relevant exonic variants are detected, discuss with the physician the possibility of assessing a CAT for RNA-seq analysis. Using FRASER in a transcriptome-wide setting with ($|\Delta\psi| \geq 0.15$ and a width >2 and $p\text{-adj} < 0.05$) parameter settings, correlate flagged genes with the phenotype and assess the splice pattern with three or more (unrelated) individuals, matching age and sex, using IGV. If splicing is aberrant, re-evaluate the ES or WGS data for coverage and variants near the splice event. If no associated variant can be found at the DNA level, we recommend discussing the results with the clinical geneticist before reporting this splice event. If a splice site or region variant is detected at the DNA level, classify this variant according to the guidelines provided by Walker et al.²⁶.

2. Expression outlier detection

If OUTRIDER (stringent filtering $p\text{-adj} < 0.05$, less stringent filtering $p\text{-adj} < 1$) identifies an expression outlier for a gene that correlates with the phenotype, reanalyze variants and coverage for this gene in detail. Discuss the results with the clinical geneticist if no causal variants at the DNA level could be found.

3. Assess monoallelic expression

To evaluate MAE transcriptome-wide, MAE events can be assessed through heterozygous, biallelic SNVs using ASEReadCounter. MAE events are defined as $\text{abs}(z\text{-score}) > 2$ of the normal distribution fitted to the VAF of all variants at the RNA level of all patients combined. Subsequently, only those genes with MAE events that are part of a gene panel of interest and unique for the patient are retained for further analysis. The remaining genes are screened for phenotypic overlap with the patient, after which the gene data is manually reviewed.

Discussion

Although several studies have highlighted the importance of RNA-seq in clinical testing^{3–9,12,13,17}, they often fail to mention the hurdles that need to be overcome to do so. Here, we provide a workflow based on our in-house experience with RNA-seq and highlight the difficulties we encountered during wet lab experiments and data analysis. Furthermore, we propose a decision tree on how RNA-seq data analysis can be implemented to aid in variant interpretation and classification.

We opted for PBMC as our CAT of choice as these are: (i) minimally invasive, (ii) have a short culture period, thus allowing a shorter turnaround time, (iii) can be treated with an NMD inhibitor, and iv) can be frozen. Furthermore, PBMCs express ~80% of the genes within our ID and Epi panel at >1 TPM. For a targeted screen, however, we recommend assessing the expression level of the gene of interest in different CATs (Supplementary Tables 2 and 3, or the web-accessible Minimum Required Sequencing Depth (MRSD) tool <https://mcgm-mrds.github.io/>)¹⁰.

To validate our workflow, we compared targeted cDNA analysis, manual targeted, and global RNA-seq analysis in nine individuals for whom a splice variant was found via ES-based panel analysis. Our RNA-Seq workflow detected an effect on splicing in six out of nine cases and enabled the reclassification of the variant in seven cases. It should, however, be noted that for accurate detection of aberrant splicing patterns via Sanger sequencing at the cDNA level, several optimizations may be required, like primer design (especially in case naturally occurring isoforms are present), choice of cDNA synthesis kit and PCR program (e.g., in case of intron retentions longer elongation times may be required). Comprehensive optimization, as suggested by Bournazos et al.³², was not carried out in this study. The main advantage of the proposed RNA-seq protocol is that a large spectrum of variants can be characterized at the cDNA level with a single protocol.

In addition, similarly to previous observations^{33,34}, we found that in silico predictions of putative splice variants perform well but may fail to predict the exact consequence at the RNA level. For example, for the *NIPBL* variant in individual 2, a donor loss and exon skip were observed in the RNA-seq data. Although SpliceAI predicted a donor loss for the *NIPBL* variant in individual 2, it had a score of 0.18, which was below the threshold of 0.2.

Treatment with CHX may enable the detection of transcripts that undergo NMD¹³. Although we observed that aberrant splicing events were detectable in both treated and untreated samples, CHX treatment increased the number of transcripts with aberrant splicing (individuals 1 and 4). It may thus facilitate the identification of a deleterious defect. Interestingly, in some cases, the wild-type allele appeared to be upregulated to partially compensate for the loss of the mutant allele. An event exemplifying this is the mis-splicing caused by the *NIPBL* splice variant in individual 3. Mis-splicing was observed in 14% and 15% of junction reads in untreated and CHX-treated PBMCs of this individual. Although *NIPBL* is a haploinsufficient gene, Kaur et al.³⁵ reported that in all individuals with genomic aberrations affecting *NIPBL*, only a reduction of 25–35% in *NIPBL* expression could be observed, even in patients with large deletions. Moreover, although not explicitly reported by the authors, in recent data published by Rentas et al.¹² and Coursimault et al.³⁶, an upregulation of the wild-type *NIPBL* allele could also be observed.

Since we only observed moderately reduced levels of *NIPBL* expression and aberrant splicing in only 14% of the *NIPBL* transcripts, this was not identified by OUTRIDER or FRASER, respectively. This clearly indicates that clinical phenotyping remains important, and that aberrant expression and splicing should be investigated with caution. As such, for targeted analysis of aberrant splicing of a small set of genes, we recommend manually inspecting the RNA-seq data and/or investigating the FRASER results for these genes without filter parameters. Clinical interpretation is also relevant for variants causing low levels of splice aberrations, as abnormal splicing may occur in healthy individuals³⁴. In individual 6, we observed an exon 17 skip of *CNOT1* in 3% of the reads, as well as in some control individuals,

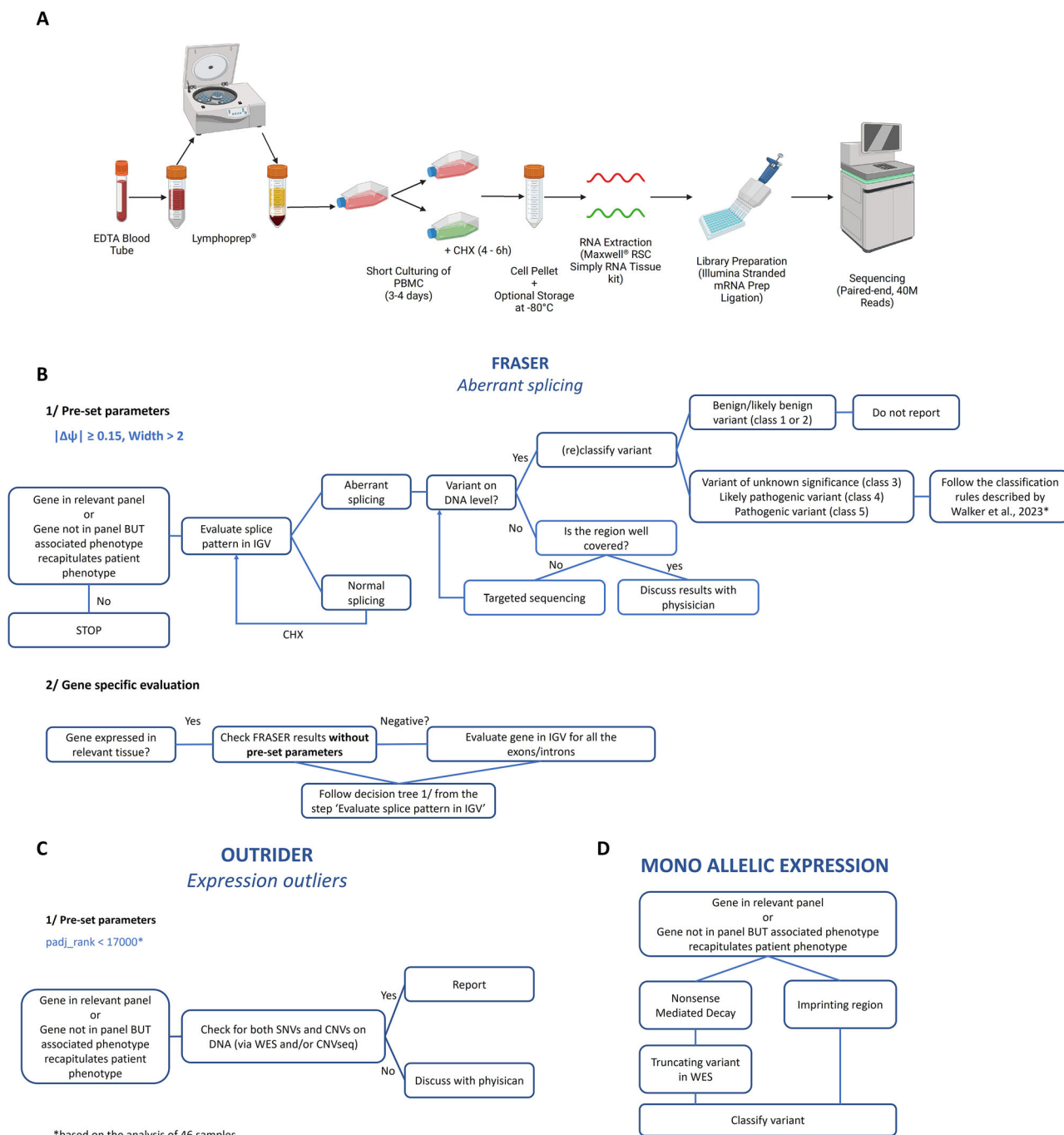


Fig. 8 | Overview of our proposed diagnostic workup for RNA-sequencing. A Wet lab workflow. **B–D** Decision trees for variant classification/evaluation for respectively FRASER (**B** 1/ pre-set parameters and 2/gene-specific evaluation), OUTRIDER (**C**), and mono allelic expression (**D**).

in the absence of a convincing genotype-phenotype correlation, exemplifying the challenges of interpreting RNA-seq data.

OUTRIDER¹⁵ and FRASER¹⁴ are most widely used for a non-targeted approach to detect aberrant expression and splicing, respectively. Although FRASER and OUTRIDER analyses did not result in new diagnoses in our cohort, FRASER was able to detect all mis-splicing events identified in the targeted approach, though only four were statistically significant (i.e., $p\text{-adj} < 0.05$). When considering the guidelines by Smirnov et al.²⁷ for single allele splice variants (i.e., $\Delta\text{psi} > 0.15$ and $p\text{-value} < 0.05$ or $p\text{-adj} < 0.1$), five out of six were retained.

OUTRIDER confirmed the structural variants in individuals 13–16 at the RNA level. Not all genes located within a deletion or duplication exhibited an outlier expression pattern, and even though the PBMCs of

individual 17 have the lowest *ZEB2* expression in our cohort, this was not called an outlier expression by OUTRIDER. Although previous power analyses showed that at least a cohort of 50–60 samples ($n = 61$ in this study) is needed to detect aberrant expression, Brechtmann et al. suggest that expanding the cohort is advisable for detecting more subtle outliers¹⁵. Consequently, due to our relatively modest cohort size, the detection of additional aberrant splice variants and/or expression outliers through a blind analysis approach was suboptimal. Increasing the sample size in future studies will likely enhance our power to uncover more subtle transcriptional alterations.

Individual 11 was diagnosed with Beckwith-Wiedemann Syndrome and presented with loss of maternal methylation of the IC2. We did not detect aberrant expression of *KCNQ1* or *CDKN1C*. Still, we observed

biallelic expression of both genes, similar to previous observations by Algar et al.³⁷ We also assessed the expression of the genes regulated by the IC1 for Beckwith–Wiedeman Syndrome: *IGF2* and *H19*. Both genes' expression levels fall below the 0.10 TPM in PBMCs to reliably assess allelic expression. These findings underscore that it remains vital to integrate epigenomics as an additional layer within the diagnostic framework until long-read DNA sequencing³⁸ becomes clinically accessible. Finally, none of the genes in our ID and Epi gene panel that presented with MAE could be correlated to the phenotype of the included individuals.

Variant calling algorithms on RNA-seq data have significantly improved in the last couple of years, and hence, not only can splicing defects or aberrant expression be investigated in the RNA-seq data. This is exemplified for individuals 20 and 23, where a causal missense variant could be identified after reanalysis of the exome data. In individual 20, a heterozygous class 5 missense *SMAD4* variant (*SMAD4*: (NM_005359.5) c.1486C > T, p.(Arg496Cys), absent in the mother (father not available)) was identified. In individual 23, a homozygous class 5 missense variant: *HMBS* (NM_000190.4): c.499C > T, p.(Arg167Trp) was identified, explaining the observed phenotype. Both variants could also be readily identified in the RNA-seq data. Unfortunately, not all relevant genes are expressed in CATs, and hence, DNA-based ES or even GS analysis remains essential.

This study and all other studies focusing on RNA-seq in a diagnostic context are based on short-read sequencing. Although we and others clearly show the added value of short-read RNA-seq in a diagnostic setting, it may only partially capture complex splicing events, especially those involving long introns or multiple exon skipping, leading to potential misinterpretation of gene isoforms. Therefore, in the (near) future, long-read RNA sequencing might offer a promising solution for studying aberrant splicing events in greater detail. By providing full-length transcripts, long-read sequencing enables more accurate detection of alternative splicing events, including exon skipping, intron retention, novel splice junctions, and phasing of variants. This can be particularly relevant in a clinical setting where precise characterization of RNA isoforms is crucial for disease diagnosis and prognosis. Furthermore, transactivation and transdifferentiation techniques³⁹ may also be increasingly used in the near future to study variants affecting tissue-specific transcripts that are not expressed in CATs.

It is worth noting that since the submission of this manuscript, Oquendo et al. have reported similar findings using blood-based RNA-seq in a diagnostic framework⁴⁰. Their approach closely mirrors ours, although they recommend using rMATS-turbo⁴¹, a recently developed tool, for detecting aberrant splicing events and conducting RNA analysis on whole blood instead of PBMCs. This distinction is particularly relevant, as the use of PBMCs permits CHX treatment to assess the impact of genetic variants on NMD.

In conclusion, PBMCs are an excellent alternative to other CATs for detecting deleterious defects at the RNA level in Mendelian disorders. We show that RNA sequencing demonstrates superiority over targeted cDNA analysis in investigating the effect of splice variants. Notably, FRASER effectively identified a significant portion of these validated splice events. At the same time, OUTRIDER precisely characterized the impact of structural variants on RNA-level dynamics, enhancing the interpretation of different genetic variants. Finally, through the application of RNA-seq, we reclassified seven out of nine variants, highlighting the potential of our PBMC-based RNA-seq workflow.

Methods

Recruitment, exome/GS, and CNV-seq

Forty-six individuals with a rare congenital disorder who either lack a molecular diagnosis or have a suspicious variant of uncertain significance that can be further investigated by RNA-seq, were included in this study (Supplementary Table 1). All individuals were evaluated at the Center for Medical Genetics, Ghent, Belgium, and examined by a clinical geneticist who provided pre-test counseling. ES or GS and molecular karyotyping (i.e., CNV-seq) were performed in a diagnostic setting. Seventeen individuals

(individuals 1–17) were selected for targeted evaluation of variants identified upon ES analysis and/or CNV sequencing. The remaining individuals (individuals 18–46) underwent analysis of the full transcriptome following normal routine workup with ES.

Legal guardians of affected probands gave informed consent for genomic investigations and publication of pseudonymized data following the rules of the Helsinki declaration and according to EC2019/1430 of Ghent University Hospital.

Exome sequencing. Genomic DNA (gDNA) was extracted from whole blood using a MagCore Automated Nucleic Acid Extractor (MagCore HF16 SUPER; Atrida) with the MagCore Genomic DNA Large Volume Whole Blood Kit (104; Atrida) according to the manufacturer's instructions. Afterwards, ES libraries were prepared with either the SureSelect Human All Exon v7 low input kit (5191-4030; Agilent Technologies) or the KAPA HyperExome kit (09062602001; Roche) following the manufacturer's instructions. Subsequently, the libraries were paired-end sequenced 2 × 150bp on a HiSeq3000 or NovaSeq6000 platform (Illumina) to a median depth of 50X or until at least 90% of all exons had a coverage of at least 20X. The raw sequencing data was then processed with the BCBio pipeline (v1.1.5; <https://zenodo.org/records/5781867>) according to their small germline variants workflow variant2. Data analysis was performed in trio with both parents for evaluation of de novo, recessive, and X-linked variants, limited to a gene panel of 1689 selected genes associated with intellectual disability (ID) and epilepsy (Epi) (ID&Epi panel, version 6, ISO 15189 accredited diagnostic workflow).

Genome sequencing. The gDNA was extracted using the same method described for ES. GS libraries were prepared with the DNA PCR-Free Library Prep Tagmentation kit (20041795; Illumina) in accordance with the manufacturer's instructions. Subsequently, the libraries were sequenced paired-end 2 × 150bp on a NovaSeq6000 platform (Illumina) to a median depth of 30X. The raw sequencing data was then processed with the BCBio pipeline according to their small germline variants workflow 3. Data analysis was limited to trio analysis of the individual and his/her parents for de novo, recessive, and X-linked variants in a panel of 1689 selected ID&Epi genes (ID&Epi panel, version 6)¹².

CNV-seq. Molecular karyotyping was performed through shallow whole-genome sequencing (sWGS) on gDNA extracted from blood. Whole genome libraries were prepared with the Illumina DNA PCR-Free Prep kit (20041795, Illumina) according to the manufacturer's instructions. The libraries were subsequently sequenced paired-end 2 × 50 bp on a NovaSeq6000 system (Illumina) to a depth of 75 M reads (high resolution). The raw sequencing data were then aligned to the human reference genome GRCh38/hg38 with Bowtie. Further processing/analysis was done through WisecondorX⁴² and Vivar⁴³. In individuals 12, 25, and 28, CNV detection was performed with arrayCGH (comparative genomic hybridization) using a 180 k Agilent array with an average genome-wide resolution of 100 kb (as described earlier⁴⁴).

RNA extraction

RNA was extracted using the Maxwell® RSC instrument (AS4500, Promega). Depending from which tissue RNA was isolated a different extraction kit was used: (1) PBMCs, fibroblasts, LCLs: Maxwell® RSC simply RNA Tissue kit (AS1340, Promega); or (2) Whole blood (from PAXgene® blood RNA tubes): Maxwell® RSC Blood simplyRNA kit (AS1380, Promega).

PBMC culture and treatment

From EDTA blood tubes, PBMCs were isolated with Lymphoprep™ (1114545, STEMCELL Technologies), a density gradient medium. These mononuclear cells were transferred into PB-max™ culture medium (12557, Life Technologies Europe) and were kept in culture for 72 h at 37 °C. Subsequently, the PBMCs were divided into two cultures, of which one was

treated with cycloheximide (CHX) (11 mg/mL, C7698, Sigma Aldrich) to allow the detection of NMD-sensitive transcripts. RNA was extracted after a 4–6 h incubation period at 37 °C.

PAXgene® blood RNA tubes—removal of Globulin RNA

Blood was drawn and collected in 9 mL PAXgene® Blood RNA tubes (762165, BD Biosciences), which were optionally stored at –80 °C. After RNA extraction, we used the QIAseq FastSelect-Globulin Kit (334376, Qiagen) to remove globulin mRNA, which pairs with the TruSeq Stranded mRNA library preparation from Illumina according to the manufacturer's instructions.

Fibroblast culture and treatment

Fibroblasts derived from skin biopsies were cultured in DMEM medium supplemented with Fetal Calf Serum (FCS), Penicillin/Streptomycin, Kanamycin, Non-Essential Amino Acids (NEAA), and Fungizone until we had at least 600,000 cells. Afterwards, the cells were split into two P60 plates, 300,000 cells/plate on Day 1. On Day 2, the culture medium was replenished with or without CHX (5 µg/mL, C7698, Sigma Aldrich), and on Day 3, RNA was extracted with the Maxwell device. In short, the cells were loosened with a cell scraper and transferred to a 15 mL tube, then centrifuged for 3 min at 300rcf. Afterwards, the supernatant was removed, and 200 µL 1-thioglycerol/homogenization solution (Promega) was added. The cells were then vortexed, and 200 µL lysis buffer (Promega) was added. Finally, the sample (400 µL) was transferred to a well from the Maxwell, and RNA was extracted with the RSC Simply RNA Tissue kit (Promega) following the RSC Simply RNA Cells method.

LCL culture and treatment

LCLs were cultured in RPMI (Roswell Park Memorial Institute). To assess the effect of treatment with puromycin (PUR) and cycloheximide (CHX), the following cell lines were purchased from Coriell: GM14870, GM15704, and GM20377 (Table 1). Following standard culturing, these cell lines were treated with PUR (10 mg/mL, Sigma Aldrich) or CHX (11 mg/mL, C7698, Sigma Aldrich) for 4–6 h prior to RNA isolation.

RT-PCR followed by targeted cDNA analysis (i.e., PCR and Sanger sequencing)

Single-stranded complementary DNA (cDNA) was synthesized with the iScript cDNA Synthesis kit (Bio-Rad, Hercules, USA) according to the manufacturer's instructions. The primer sequences can be found in Supplementary Table 2.

RT-qPCR

RT-qPCR was performed with a PCR mix containing 5 ng cDNA, 2.5 µL of SsoAdvanced SYBR qPCR mix (BioRad), and 0.25 µL of forward and reverse primers (250 µM concentration, Integrated DNA Technologies). The RT-qPCR was performed on a LC-480 device (Roche), and gene expression levels were analyzed using the qBase+ 187 software 3.2 (Cell-Carda) and normalized against housekeeping genes *SDHA* and *GAPDH*. In Fig. 1C and Supplementary Fig. 1A–D, the strategies and results of the impact of different truncating variants on expression level via RT-qPCR are shown. All RT-qPCR primers used in this study can be found in Supplementary Table 2.

RNA-sequencing—library preparation

Library preparation of RNA samples (RIN > 7) was performed using the Illumina Stranded mRNA Prep Ligation kit (20040534, Illumina) according to manufacturer's instructions with some minor modifications: (i) six minutes fragmentation time, (ii) 50 min first strand cDNA synthesis and (iii) an additional purification step at the end. The first two adjustments were implemented to achieve longer read fragments, and the third one to eliminate adapter dimers. The libraries were subsequently paired-end sequenced (2 x 150bp) on a NovaSeq6000 (Illumina) with an average of 40 M read pairs per sample.

RNA-sequencing—data analysis

RNA-seq data were aligned to the human reference genome, GRCh38/hg38, using the genome annotation data from Gencode⁴⁵ (version 41) and STAR⁴⁶ (version 2.7.6a).

Percentage gene panel expression calculation. Transcript expression was calculated with Kallisto⁴⁷ (version 0.46.1) for 10 fibroblast, 9 whole blood, 10 LCL, and 10 PBMC samples. Afterwards, the data were loaded into Python (version 3.8.5) to calculate gene expression and the percentage of gene panel expression for each CAT and generate the UpSet plots. Only genes where gene expression was >1 TPM (transcript per million)⁴⁸ were included in the calculation. The number of genes that fulfilled this criterion was then divided by the total number of genes in the panel to determine the percentage of gene panel expression.

Determination of the effectiveness of CHX treatment. To determine the effectiveness of NMD inhibition with CHX, splicing of *SRSF2*, particularly the splicing pattern of isoform ENST00000452355.7 at exon 3, was assessed. This exon is exclusively present in this NMD-sensitive transcript. After CHX treatment, this transcript should be present at a much higher level than in the untreated cells³¹. Therefore, to assess the effectiveness of CHX treatment, the percentage of *SRSF2* reads mapping to exon 3, was calculated for the untreated and treated PBMCs. These percentages are given in Supplementary Fig. 3A.

Aberrant expression—OUTRIDER. To identify outlier expression in one sample within a larger population, OUTRIDER¹⁵ was used. The workflow was adapted from Dekker et al.¹³. In summary, gene counts were determined through bedtools multicov⁴⁹ (version v2.29.2) using the genome annotation data from Gencode⁴⁵ (version 41). Raw gene counts were subsequently normalized and processed with OUTRIDER (version 1.12.0), after which, z-scores and p-values were calculated. To filter for aberrant expression, we applied the following filtering strategy: significantly aberrant expressed genes (SAEGs) had a *p*-adj ≤ 0.05, and putatively aberrantly expressed genes (PAEGs) had a *p*-adj value between 0.05 and 1. Non-aberrant genes had a *p*-adj value of 1.

Aberrant splicing—FRASER. For the aberrant splicing analysis, the R package, Find RARE Splicing Events in RNA-seq (FRASER)¹⁴ (version 1.9.1) was used. First, the number of splits and non-spliced reads per sample was determined using the countRNA function. Subsequently, the ψ -values (percent spliced in) were calculated with the calculatePSIvalues function, and afterwards, the data were filtered with filter-ExpressionAndVariability. minDeltaPsi was set to 0.0 and filter to FALSE. This was done to include as many splice junctions as possible in the FRASER analysis. Then, the optimal value for each hyperparameter (psi5, psi3, and theta) was established via the optimHyperParams function. Finally, the splicing model was fit to the data using FRASER and setting q to a list of optimal hyperparameter values. The outliers were then called through the results function. To filter for aberrant splicing events, we applied the following strategy: Significant Aberrant Splicing Event (SASE): $|\Delta\psi| \geq 0.15$ and a *p*-adj ≤ 0.05; Putative Aberrant Splicing Event (PASE): $|\Delta\psi| \geq 0.15$ and a *p*-adj between 0.05 and 1. Additionally, all FRASER calls with a width ≤ 2 were also filtered out.

Aberrant splicing—targeted manual analysis—calculation percentage of aberrant splicing. Splicing patterns for regions of interest (targeted manual analysis) were visualized by Sashimi plots available in the integrative genomics viewer (IGV, version 2.7.2)^{50,51}.

Aberrant splicing at a splice acceptor/donor, was calculated as the ratio of split reads aligning with aberrantly spliced transcripts to the total number of split reads at that splice acceptor/donor. This resultant ratio was subsequently represented as a percentage. If a variant caused multiple splice events, the percentage of aberrant splicing was calculated separately per event.

Mono allelic expression—ASEReadCounter. For this step, the variant call format (VCF) files of the BCBio germline variant workflow were used. These VCF files were first filtered to include only heterozygous single-nucleotide variants (SNVs) using bcftools and the SelectVariants function from GATK⁵² (version 4.4.0). Afterwards, ASEReadCounter was used to determine the number of reference and alternative reads at each variant locus in the RNA data. The resulting count files were then processed through Python. Each mutant locus's allele frequency was calculated by dividing the altCounts by the totalCounts. Loci where the lowMAPQDepth was more than half of the totalCounts were excluded from the analysis, as the mono allele expression (MAE) at these loci might be due to an alignment issue. Furthermore, only genes with an expression level of at least 1 TPM were included in the analysis. Subsequently, a normal distribution was fitted to the VAFs (variant allele frequencies) of all variants from all patients through Scipy⁵³ (version 1.13.1), and the mean and standard deviation (SD) of this fit were used to define MAE events as $\text{mean} \pm 2 \times \text{SD}$.

Alternatively, manual targeted analysis was performed using the RNA-seq alignment files and the IGV^{51,54}.

Variant classification after RNA-seq

Variants are classified based on the ACMG⁵⁵ and ACGS²⁵ guidelines. For the classification of non-canonical splice variants (outside the $-/+$ 1,2 position), the guidelines described by Walker et al. were used²⁶.

Data availability

All RNA-seq data are stored in a repository at the Center for Medical Genetics Ghent. The consent agreement only allows us to share anonymized data of affected individuals, and therefore we are unable to share the sequencing data.

Code availability

We made use of the publicly available tools FRASER, OUTRIDER, and ASE ReadCounter. We recommend using the latest update of these tools.

Received: 18 November 2024; Accepted: 7 May 2025;

Published online: 28 May 2025

References

- Starita, L. M. et al. Variant interpretation: functional assays to the rescue. *Am. J. Hum. Genet.* **101**, 315–325 (2017).
- Brnich, S. E., Rivera-Muñoz, E. A. & Berg, J. S. Quantifying the potential of functional evidence to reclassify variants of uncertain significance in the categorical and Bayesian interpretation frameworks. *Hum. Mutat.* **39**, 1531–1541 (2018).
- Montgomery, S. B., Bernstein, J. A. & Wheeler, M. T. Toward transcriptomics as a primary tool for rare disease investigation. *Cold Spring Harb. Mol. Case Stud.* **8**, a006198 (2022).
- Lee, H. et al. Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet. Med.* **22**, 490–499 (2020).
- Frésard, L. et al. Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat. Med.* **25**, 911–919 (2019).
- Cummings, B. B. et al. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* **9**, eaal5209 (2017).
- Murdock, D. R. et al. Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J. Clin. Invest.* **131**, e141500 (2021).
- Gonorazky, H. D. et al. Expanding the boundaries of RNA sequencing as a diagnostic tool for rare Mendelian disease. *Am. J. Hum. Genet.* **104**, 466–483 (2019).
- Yépez, V. A. et al. Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med.* **14**, 1–26 (2022). 2022 14:1.
- Rowlands, C. F. 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.* **109**, 210–222 (2022).
- Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* **16**, 665–677 (2015).
- Rentas, S. et al. Diagnosing Cornelia de Lange syndrome and related neurodevelopmental disorders using RNA sequencing. *Genet. Med.* **22**, 927–936 (2020).
- Dekker, J. et al. Web-accessible application for identifying pathogenic transcripts with RNA-seq: Increased sensitivity in diagnosis of neurodevelopmental disorders. *Am. J. Hum. Genet.* **110**, 251–272 (2023).
- Mertes, C. et al. Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat. Commun.* **12**, 529 (2021).
- Brechtmann, F. et al. OUTRIDER: a statistical method for detecting aberrantly expressed genes in RNA sequencing data. *Am. J. Hum. Genet.* **103**, 907–917 (2018).
- Center for Medical Genetics Gent. *Constitutional Genetic Conditions*. <https://www.cmgg.be/en/professionals/labguide/constitutional-genetic-conditions> (2024).
- Yépez, V. A. et al. Detection of aberrant gene expression events in RNA sequencing data. *Nat. Protoc.* **16**, 1276–1296 (2021).
- Carter, M. S. et al. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. *J. Biol. Chem.* **270**, 28995–29003 (1995).
- Noensie, E. N. & Dietz, H. C. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. *Nat. Biotechnol.* **19**, 434–439 (2001).
- Dang, Y. et al. Inhibition of nonsense-mediated mRNA decay by the natural product pateamine A through eukaryotic initiation factor 4AIII. *J. Biol. Chem.* **284**, 23613–23621 (2009).
- Häuser, F. et al. A non-invasive diagnostic assay for rapid detection and characterization of aberrant mRNA-splicing by nonsense mediated decay inhibition. *Mol. Genet. Metab.* **130**, 27–35 (2020).
- Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res.* **42**, 13534–13544 (2014).
- Xiong, H. Y. et al. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **347**, 1254806 (2015).
- Jaganathan, K. et al. Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548.e24 (2019).
- Ellard, S. et al. ACGS best practice guidelines for variant classification in rare disease 2020. *Association for Clinical Genomic Science* <https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf> (ACGS, 2020).
- Walker, L. C. 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.* **110**, 1046–1067 (2023).
- Smirnov, D., Schlieben, L. D., Peymani, F., Berutti, R. & Prokisch, H. Guidelines for clinical interpretation of variant pathogenicity using RNA phenotypes. *Hum. Mutat.* **43**, 1056–1070 (2022).
- Backers, L. et al. Missing heritability in Bloom syndrome: first report of a deep intronic variant leading to pseudo-exon activation in the BLM gene. *Clin. Genet.* **99**, 292–297 (2021).
- Diaz-Meyer, N. et al. Silencing of CDKN1C (p57KIP2) is associated with hypomethylation at KvDMR1 in Beckwith–Wiedemann syndrome. *J. Med. Genet.* **40**, 797 (2003).
- Valente, F. M. et al. Transcription alterations of KCNQ1 associated with imprinted methylation defects in the Beckwith–Wiedemann locus. *Genet. Med.* **21**, 1808–1820 (2019). 2019 21:8.
- Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* **12**, 565–575 (2011).
- Bournazos, A. M. et al. Standardized practices for RNA diagnostics using clinically accessible specimens reclassifies 75% of putative splicing variants. *Genet. Med.* **24**, 130–145 (2022).

33. 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).
34. Maddirevula, S. et al. Analysis of transcript-deleterious variants in Mendelian disorders: implications for RNA-based diagnostics. *Genome Biol.* **21**, 145 (2020).
35. Kaur, M. et al. NIPBL expression levels in CdLS probands as a predictor of mutation type and phenotypic severity. *Am. J. Med. Genet. C. Semin. Med. Genet.* **172**, 163–170 (2016).
36. Coursimault, J. et al. Deep intronic NIPBL de novo mutations and differential diagnoses revealed by whole genome and RNA sequencing in Cornelia de Lange syndrome patients. *Hum. Mutat.* **43**, 1882–1897 (2022).
37. Algar, E., Brickell, S., Deeble, G., Amor, D. & Smith, P. Analysis of CDKN1C in Beckwith Wiedemann syndrome. *Hum. Mutat.* **15**, 497–508 (2000).
38. Geysens, M. et al. Nanopore sequencing-based episignature detection. *medRxiv* <https://doi.org/10.1101/2024.04.19.24305959> (2024).
39. Nicolas-Martinez, E. C. et al. RNA variant assessment using transactivation and transdifferentiation. *Am. J. Hum. Genet.* **111**, 1673–1699 (2024).
40. Jaramillo Oquendo, C. et al. Identification of diagnostic candidates in Mendelian disorders using an RNA sequencing-centric approach. *Genome Med.* **16**, 110 (2024).
41. Wang, Y. et al. rMATS-turbo: an efficient and flexible computational tool for alternative splicing analysis of large-scale RNA-seq data. *Nat. Protoc.* **19**, 1083–1104 (2024).
42. Raman, L., Dheedene, A., De Smet, M., Van Dorpe, J. & Menten, B. WisecondorX: improved copy number detection for routine shallow whole-genome sequencing. *Nucleic Acids Res.* **47**, 1605–1614 (2019).
43. Sante, T. et al. ViVar: a comprehensive platform for the analysis and visualization of structural genomic variation. *PLoS One* **9**, e113800 (2014).
44. Buysse, K. et al. Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. *Eur. J. Med. Genet.* **52**, 398–403 (2009).
45. Frankish, A. et al. GENCODE 2021. *Nucleic Acids Res.* **49**, D916–D923 (2021).
46. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
47. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
48. Aicher, J. K., Jewell, P., Vaquero-Garcia, J., Barash, Y. & Bhoj, E. J. Mapping RNA splicing variations in clinically accessible and nonaccessible tissues to facilitate Mendelian disease diagnosis using RNA-seq. *Genet. Med.* **22**, 1181–1190 (2020).
49. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
50. Katz, Y. et al. Quantitative visualization of alternative exon expression from RNA-seq data. *Bioinformatics* **31**, 2400–2402 (2015).
51. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
52. Van der Auwera, G. A. & O'Connor, B. D. *Genomics in the Cloud* (O'Reilly Media, Inc., 2020).
53. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
54. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
55. 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).

Acknowledgements

We extend our gratitude to the participating families for their contribution to this study. Financial support has been provided by grants G055422N, G063423N, and T003819N from the Research Foundation—Flanders (FWO) and BOF/STA/201909/009 and BOF20/GOA/023 from the Special Research Fund (BOF) from Ghent University.

Author contributions

L.D.C., E.D., S.V. and B.M. conceived the overall project. L.D.C., E.D. and S.V. designed the experiments. E.D. and L.V. performed the wet lab experiments. L.D.C. designed and performed bioinformatic analyses. E.D., G.D.C. and S.V. performed bioinformatic analyses as well. L.B., A.B., K.B.M.C., R.d.P., C.K., N.S., J.S., H.S., S.T., E.V., O.V., T.V.D. and B.C. provided biological samples of affected individuals and their parents and gathered clinical information. B.V., A.D., S.V. and B.M. supervised the study. L.D.C., E.D., A.D., S.V. and B.M. interpreted the results. L.D.C., E.D. and S.V. wrote the initial manuscript. All authors reviewed and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41525-025-00502-7>.

Correspondence and requests for materials should be addressed to Sarah Vergult or Björn Menten.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025