

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

Estelle Colin, Yannis Duffourd, Emilie Tisserant, Raissa Relator, Ange-Line Bruel, Frédéric Tran Mau-Them, Anne-Sophie Denommé-Pichon, Hana Safraou, Julian Delanne, Nolwenn Jean-Marçais, et al.

▶ **To cite this version:**

Estelle Colin, Yannis Duffourd, Emilie Tisserant, Raissa Relator, Ange-Line Bruel, 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. Frontiers in Cell and Developmental Biology, 2022, 10, pp.1021785. 10.3389/fcell.2022.1021785 . hal-03886419

HAL Id: hal-03886419

https://univ-rennes.hal.science/hal-03886419

Submitted on 6 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

frontiers | Frontiers in Cell and Developmental Biology

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fcell.2022.1021785

Check for updates

OPEN ACCESS

EDITED BY
Ilaria Parenti,
University Hospital Essen, Germany

REVIEWED BY
Palma Finelli,
University of Milan, Italy
Beatriz Puisac,
University of Zaragoza, Spain

*CORRESPONDENCE
Estelle Colin,
escolin@chu-angers.fr
Antonio Vitobello,
antonio.vitobello@u-bourgogne.fr

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Molecular and Cellular Pathology, a section of the journal Frontiers in Cell and Developmental Biology

RECEIVED 17 August 2022
ACCEPTED 11 October 2022
PUBLISHED 28 October 2022

CITATION
Colin E, Duffourd Y, Tisserant E, Relator R, Bruel A-L, Tran Mau-Them F, Denommé-Pichon A-S, Safraou H, Delanne J, Jean-Marçais N, Keren B, Isidor B, Vincent M, Mignot C, Heron D, Afenjar A, Heide S, Faudet A, Charles P, Odent S, Herenger Y, Sorlin A, Moutton S, Kerkhof J, McConkey H, Chevarin M, Poë C, Couturier V, Bourgeois V, Callier P, Boland A, Olaso R, Philippe C, Sadikovic B, Thauvin-Robinet C, Faivre L, Deleuze J-F and Vitobello A (2022), 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. 10:1021785.
doi: 10.3389/fcell.2022.1021785

COPYRIGHT
© 2022 Colin, Duffourd, Tisserant, Relator, Bruel, Tran Mau-Them, Denommé-Pichon, Safraou, Delanne, Jean-Marçais, Keren, Isidor, Vincent, Mignot, Heron, Afenjar, Heide, Faudet, Charles, Odent, Herenger, Sorlin, Moutton, Kerkhof, McConkey, Chevarin, Poë, Couturier, Bourgeois, Callier, Boland, Olaso, Philippe, Sadikovic, Thauvin-Robinet, Faivre, Deleuze and Vitobello. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

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

Estelle Colin[1,2*†], Yannis Duffourd[2,3†], Emilie Tisserant[2], Raissa Relator[4], Ange-Line Bruel[2,3], Frédéric Tran Mau-Them[2,3], Anne-Sophie Denommé-Pichon[2,3], Hana Safraou[2,3], Julian Delanne[2,5], Nolwenn Jean-Marçais[5], Boris Keren[6], Bertrand Isidor[7], Marie Vincent[7], Cyril Mignot[8,9], Delphine Heron[10], Alexandra Afenjar[11], Solveig Heide[10], Anne Faudet[10], Perrine Charles[10], Sylvie Odent[12,13], Yvan Herenger[14], Arthur Sorlin[5], Sébastien Moutton[5], Jennifer Kerkhof[4], Haley McConkey[4], Martin Chevarin[2,3], Charlotte Poë[2,3], Victor Couturier[2,3], Valentin Bourgeois[2,3], Patrick Callier[2], Anne Boland[15], Robert Olaso[15,16], Christophe Philippe[2,3], Bekim Sadikovic[4,17], Christel Thauvin-Robinet[2,3,18†], Laurence Faivre[2,5†], Jean-François Deleuze[15,16†] and Antonio Vitobello[2,3*†]

[1]Service de Génétique Médicale, CHU d'Angers, Angers, France, [2]UFR des Sciences de Santé, GAD "Génétique des Anomalies du Développement", INSERM-Université de Bourgogne UMR1231, Fédération Hospitalo-Universitaire (FHU)-TRANSLAD, Dijon, France, [3]Unité Fonctionnelle Innovation en Diagnostic Génomique des Maladies Rares, Fédération Hospitalo-Universitaire-TRANSLAD, CHU Dijon Bourgogne, Dijon, France, [4]Molecular Diagnostics Program and Verspeeten Clinical Genome Centre, London Health Sciences and Saint Joseph's Healthcare, London, ON, Canada, [5]Centre de Génétique et Centre de Référence "Anomalies du Développement et Syndromes Malformatifs", Hôpital d'Enfants, Centre Hospitalier Universitaire de Dijon, Dijon, France, [6]Assistance publique - Hôpitaux de Paris (APHP), Département de Génétique, Groupe Hospitalier Pitié Salpêtrière, Paris, France, [7]Service de Génétique Médicale, CHU Nantes, Nantes, France, [8]Sorbonne Université/INSERM U1127/CNRS UMR 7225/Institut du Cerveau, Paris, France, [9]Service de Neurologie, Hôpital la Pitié Salpêtrière, Sorbonne Université, Paris, France, [10]Département de Génétique, Assistance publique - Hôpitaux de Paris Sorbonne Université, Hôpital Pitié-Salpêtrière et Trousseau, Paris, France, [11]Assistance publique - Hôpitaux de Paris, Département de Génétique, Sorbonne Université, GRC No. 19, ConCer-LD, Centre de Référence Déficiences Intellectuelles de Causes Rares, Hôpital Armand Trousseau, Paris, France, [12]Service de Génétique Clinique, European Reference Network (ERN) ITHACA, CHU Rennes, Rennes, France, [13]IGDR (Institut de Génétique et Développement de Rennes)—UMR 6290, ERL U1305, CNRS, INSERM, Univ Rennes, Rennes, France, [14]Service de Génétique Médicale, CHU de Tours, Tours, France, [15]Commissariat à l'énergie atomique et aux énergies alternatives (CEA), Centre National de Recherche en Génomique Humaine (CNRGH), Université Paris-Saclay, Evry, France, [16]LabEx GENMED (Medical Genomics) ParisFrance, [17]Department of Pathology and Laboratory Medicine, Western University, London, ON, Canada, [18]Centre de Référence Maladies Rares "Déficiences Intellectuelles de Causes Rares", Centre de Génétique, Fédération Hospitalo-Universitaire-TRANSLAD, CHU Dijon Bourgogne, Dijon, France

Purpose: Patients with rare or ultra-rare genetic diseases, which affect 350 million people worldwide, may experience a diagnostic odyssey. High-throughput sequencing leads to an etiological diagnosis in up to 50% of individuals with heterogeneous neurodevelopmental or malformation disorders. There is a growing interest in additional omics technologies in translational research settings to examine the remaining unsolved cases.

Methods: We gathered 30 individuals with malformation syndromes and/or severe neurodevelopmental disorders with negative trio exome sequencing and array comparative genomic hybridization results through a multicenter project. We applied short-read genome sequencing, total RNA sequencing, and DNA methylation analysis, in that order, as complementary translational research tools for a molecular diagnosis.

Results: The cohort was mainly composed of pediatric individuals with a median age of 13.7 years (4 years and 6 months to 35 years and 1 month). Genome sequencing alone identified at least one variant with a high level of evidence of pathogenicity in 8/30 individuals (26.7%) and at least a candidate disease-causing variant in 7/30 other individuals (23.3%). RNA-seq data in 23 individuals allowed two additional individuals (8.7%) to be diagnosed, confirming the implication of two pathogenic variants (8.7%), and excluding one candidate variant (4.3%). Finally, DNA methylation analysis confirmed one diagnosis identified by genome sequencing (Kabuki syndrome) and identified an episignature compatible with a BAFopathy in a patient with a clinical diagnosis of Coffin-Siris with negative genome and RNA-seq results in blood.

Conclusion: Overall, our integrated genome, transcriptome, and DNA methylation analysis solved 10/30 (33.3%) cases and identified a strong candidate gene in 4/30 (13.3%) of the patients with rare neurodevelopmental disorders and negative exome sequencing results.

KEYWORDS

undiagnosed neurodevelopmental diseases, genome sequencing, transcriptome sequencing, DNA methylation analysis, translational research

1 Introduction

Rare and ultra-rare genetic diseases, defined as having an average global prevalence of 1 in 2,500 and 1 in 50,000, respectively, collectively affect about 350 million of the general population (Ferreira, 2019). Affected individuals and their families experience a diagnostic odyssey lasting on average 5 years ("Global Commission | Ending the Diagnostic Odyssey for Children with a Rare Disease" n.d.). However, early molecular diagnosis is fundamental for a better understanding of the disease, informed care in general medicine, and genetic counseling. Over the past decade, high-throughput sequencing, and in particular whole exome sequencing (ES), which enriches coding regions, representing ~1.5% of the human genome, has rapidly become the first-line genomics assay in clinical settings. Its diagnostic yield ranges from 30% to 50% in patients presenting with heterogeneous rare syndromic genetic disorders with suspected Mendelian inheritance (McInerney-Leo et al., 2013; Veeramah et al., 2013; Clark et al., 2018). However, molecular diagnosis remains elusive in 50%–75% due to 1) the challenge of interpreting the data, 2) technological limitations [i.e., mosaic variants, repeat expansions, or structural variants (SVs) not correctly detected through ES], 3) non-coding regulatory variants affecting promoters, enhancers, deep intronic regions, or distant-acting regulatory sequences located in intergenic regions, and 4) complex inheritance (Frésard and Montgomery, 2018; Boycott et al., 2019; Hartley et al., 2020).

There is growing interest in whole genome sequencing (GS) coupled with total RNA sequencing (RNA-seq) in translational research settings. Indeed, GS explores variants in the coding and non-coding regions with fewer technological limitations although the challenge of interpreting the variants remains. GS analysis detects more than three million single nucleotide variants (SNV) and more than 1,500 SVs per individual on average. Of these three million SNVs, 30,000 are rare, and

some are expected to have a significant impact on gene expression or alternative splicing. RNA-seq is able to measure variations in RNA abundance, allele-specific expression, and aberrant splicing, which assists with interpretation of variants. Thus, some recent studies reported an increased diagnostic yield of 7.5%–35% using RNA-Seq as a complementary approach to ES or GS in well-defined diseases, with homogeneous cohorts of patients and appropriate sample tissues (Cummings et al., 2017; Kremer et al., 2018; Frésard et al., 2019; Gonorazky et al., 2019; Hamanaka et al., 2019; Lee et al., 2020; Murdock et al., 2020; Stenton and Prokisch, 2020; Yépez et al., 2022). Furthermore, the study of genome-wide DNA methylation profiles in peripheral blood as biomarkers associated with rare developmental disorders has been demonstrating its utility for the assessment and the reclassification of variants of unknown significance in diagnostic settings (Aref-Eshghi et al., 2019; Aref-Eshghi et al., 2020; Sadikovic et al., 2021; Levy et al., 2022).

In this context, our project aimed to integrate short-read genome sequencing, messenger RNA-seq analysis, and methylation studies as complementary translational research tools to examine several individual-derived samples and look for rare diseases associated with neuro-developmental disorders, when the first line and high-quality trio ES had produced negative results.

2 Materials and methods

2.1 Recruitment of individuals and data sharing

Thirty individuals were recruited from four genetics centers belonging to the French network for rare diseases (CHU Dijon, CHU Nantes, CHU Rennes, APHP Paris) and carefully evaluated by our interdisciplinary clinical-biological team. Affected individuals with malformation syndromes and/or severe neurodevelopmental disorders, with negative trio exome sequencing and array comparative genomic hybridization results were enrolled. Informed consent was obtained from all subjects participating in the study.

2.2 DNA extraction—quantity and quality controls

DNA was extracted from blood collected in EDTA tubes. 3–5 ml of whole blood was incubated for 10 min in RBC lysis buffer (Qiagen GmbH, Hilden, Germany) and then centrifuged for 2 min at 2000 rpm to obtain white blood cell pellet, which was resuspended in 180 µl of residual supernatant and 20 µl of RNAse A (Qiagen GmbH, Hilden, Germany). Purification was then performed using the QiAamp DNA Blood mini kit on a QiaCube extraction device following the standard protocol.

Quantification was obtained using the Qubit dsDNA HS Assay (Life Technologies, CA, United States) and gel electrophoresis. The purity of DNA was verified through an evaluation of the 260/280 and 260/230 absorbance ratios on a Multiskan Go device (Thermo Scientific, Waltham, MA, United States).

At least 4 µg of DNA was needed per sample to use for quality control before sequencing at the CNRGH platform and to potentially prepare a second library in the event of technical problems. If the quantity or quality of DNA from a sample was insufficient, a new sample was requested from the center.

2.3 RNA extraction—quantity and quality control

Total RNA was extracted from whole blood collected in PAXgene tubes (Preanalytics GmbH, Hombrechtikon, Switzerland) using the PAXgene Blood RNA kit (Preanalytics GmbH, Hombrechtikon, Switzerland) automated on a QiaCube extraction device (Qiagen GmbH, Hilden, Germany) following the standard protocol. Alternatively, RNA was extracted from fibroblast cell cultures using TRIzol® RNA isolation reagent (ThermoFisher).

RNA was then quantified by measuring absorbance using a NanoDrop device. The quality was assessed by determining the RNA Integrity Number (RIN) on the bioanalyzer device (Agilent Technologies, Santa Clara, CA, United States). RNA was suitable for RNA-Seq if the RIN was at least 7.

2.4 Short-read genome sequencing

The genomic DNA libraries were prepared following the TruSeq DNA PCR-free protocol (Illumina, CA, United States). A minimum of 1 µg of genomic DNA was sheared by sonication and then purified. Oligonucleotide adaptors to sequence both ends were ligated on end-repaired fragments and then purified. DNA libraries were barcoded (indexed) and then multiplexed. GS was performed at the *Centre National de Recherche en Génomique Humaine* (CNRGH, CEA) using the Illumina NovaSeq6000 platform (Illumina, CA, United States), generating 150 base pairs paired-end reads. Data sequencing was required to meet minimum quality standards, with an average of over ×35 depth of coverage and more than 97% of the genome covered by at least 10 reads.

2.5 RNA sequencing

RNA-seq sequencing was performed by the CNRGH (CEA). After complete RNA quality control (quantified in duplicate on a

NanoDrop™ 8,000 spectrophotometer and RNA 6000 Nano LabChip analysis on a Bioanalyzer from Agilent), libraries were prepared using the TruSeq Stranded mRNA Library Prep Kit (Illumina). All libraries were prepared on an automated platform using an input of 1 μg of total RNA, in line with the manufacturer's instructions. Library quality was checked on a LabChip GX (Perkin Elmer) for profile analysis and quantification, and sample libraries were pooled before sequencing, to reach the expected sequencing depth. Sequencing was performed on an Illumina HiSeq 4,000 as paired-end 100 bp reads, using dedicated Illumina sequencing reagents. Libraries were generally pooled using four samples per lane. FASTQ files produced after RNA-seq sequencing were then processed by in-house CNRGH tools to assess the quality of raw and aligned nucleotides.

2.6 DNA methylation data analysis

Methylation analysis was performed with version 3 of the clinically validated EpiSign™ assay as previously described (Aref-Eshghi et al., 2020, 2019; Sadikovic et al., 2021; Levy et al., 2022).

2.7 Bioinformatics analysis

2.7.1 Short-read genome sequencing

Variants were identified using the FHU Translad computational platform, hosted by the University of Burgundy Computing Cluster (CCuB). Raw data quality was evaluated by FastQC software (v0.11.4). Reads were aligned to the GRCh37/hg19 human genome reference sequence using the Burrows-Wheeler Aligner (v0.7.15) and subsequently to GRCh38 for reanalysis. Aligned read data underwent the following steps: 1) duplicate paired-end reads were removed by Picard software (v2.4.1), and 2) base quality score recalibration was done by the Genome Analysis Toolkit (GATK v3.8) Base recalibrator. Using GATK Haplotype Caller, Single Nucleotide Variants with a quality score >30 and an alignment quality score >20 were annotated with SNPEff (v4.3). Rare variants were identified by focusing on nonsynonymous changes at a frequency of less than 1% in the gnomAD database.

Copy Number Variants were detected using two approaches: the first based on read depth analysis using Control-FREEC (v11.4) and the second on anomalous read pairs combined with split-read detection using Lumpy (v0.2.12). The resulting CNVs and SVs were annotated using in-house python scripts and were filtered in terms of their frequency in public databases (DGV, ISCA, DDD).

2.7.2 RNA-sequencing

Aberrant splice events and expression outliers were identified using the FHU Translad computational platform, hosted by the University of Burgundy Computing Cluster (CCuB). Raw data quality was evaluated by FastQC software (v0.11.4). Reads were aligned to the GRCh37/hg19 human genome reference sequence using the STAR2 Aligner (v2.5.2b) with the 2-pass mapping method using the human RefSeq genome annotation (Build GCF_000001405.25). Read counts were also collected using STAR2. Uniquely mapped reads are counted when overlapping only one gene.

Outlier expressed genes were detected using two parallel methods: DESeq2 (v1.26.0) and Outrider (v1.4.2). After a normalization step, the expression analysis was performed using the following analysis design: one versus the whole analysis batch, allowing computation of the expression variance for the whole cohort. A Z-score was computed, and filters were applied to only keep genes with a z-score superior to 3 or inferior to −3.

Aberrant splice events were detected using three parallel methods: rMATS (v4.0.2), LeafCutter (v0.2.9), and a custom method derived from Cummings et al. (2017)

rMATS allowed us to compute a Percent Spliced In (PSI) value, indicating the proportion of the junction involved in a splice event. LeafCutter performs an intron analysis using a clustering method. For both methods, a Z-score was computed and the same filters were applied as for expression. The custom method considered each splice junction as a rare variant and applied a filter based on frequency in the cohort to select only rare events.

2.7.3 DNA methylation data analysis

Briefly, methylated and unmethylated signal intensity generated from the EPIC array was imported into R 3.5.1 for normalization, background correction, and filtering. Beta values ranging from 0 (no methylation) to 1 (complete methylation) were calculated as a measure of methylation level and processed through the established support vector machine (SVM) classification algorithm for EpiSign disorders. The EpiSign Knowledge Database, composed of over 10,000 methylation profiles from reference disorder-specific and unaffected control cohorts, was used by the classifier to generate disorder-specific methylation variant pathogenicity (MVP) scores. MVP scores represent confidence of prediction for each disorder, ranging from 0 (discordant) to 1 (highly concordant). A positive classification typically generates MVP scores greater than 0.5. These scores, in combination with the assessment of hierarchical clustering and multidimensional scaling, are used in generating the final matched EpiSign result.

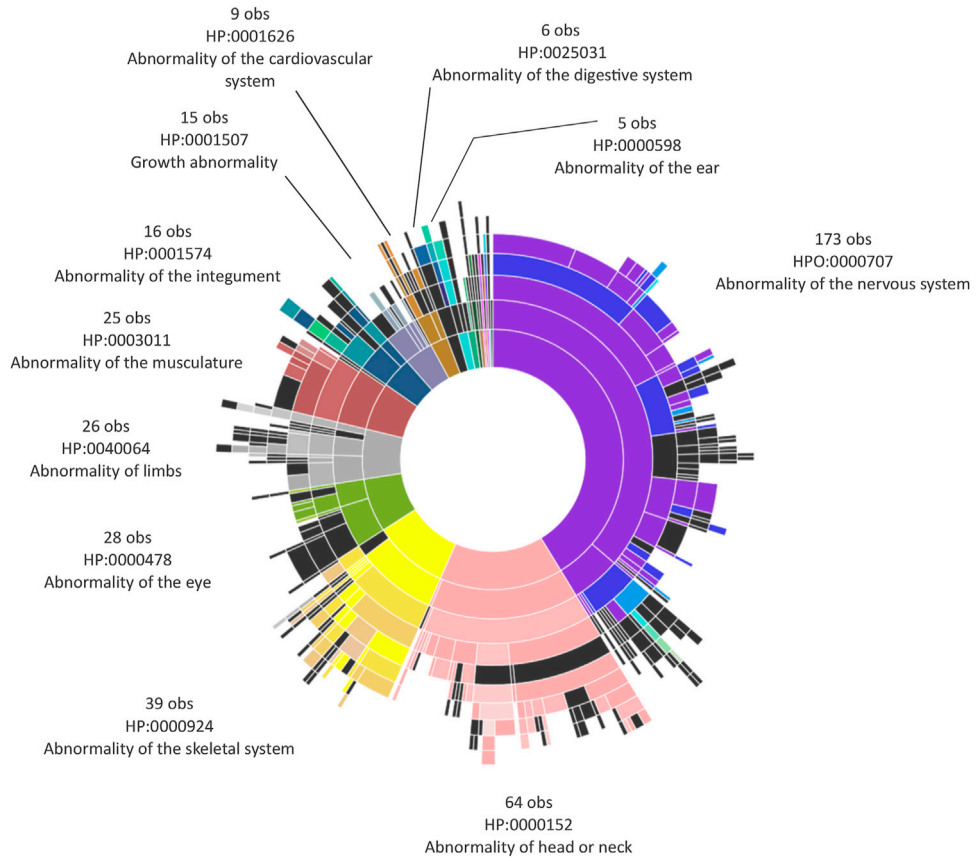


FIGURE 1
Human Phenotype Ontology (HPO) terms observed in the cohort. Sunburst plot depicting the hierarchical organization of ontologies described in our cohort, based on the Human Phenotype Ontology (http://purl.obolibrary.org/obo/hp.obo; format-version: 1.2; data-version: hp/releases/2019-02-12). The phenotypic abnormalities, representing the roots or the topmost terms in the hierarchy, are depicted as semi-circular sections at the center of the sunburst. For each phenotypic abnormality and its corresponding HP code, the number of observations stemming from each root is reported. The Sunburst plot was obtained using the JavaSript library D3.js—https://d3js.org

3 Results

3.1 Characteristics of the cohort

The cohort was mainly composed of pediatric individuals (22/30; 73%), and the sex distribution was mostly female (19/30; 63%). Only two individuals (6%) came from consanguineous unions. The median age of our cohort was 13.7 years (4 years and 6 months to 35 years and 1 month), including eight adult patients aged 18–35 years and 1 month. Phenotypic data were collected as Human Phenotype Ontology (HPO) terms. For each individual, at least two HPO terms and at most 11 HPO terms were collected, giving rise to a global dataset of 417 observations (Figure 1). The most represented terms, accounting for 66% of the available HPO terms, included abnormalities of the nervous system (41.3%), head and neck (15.3%), and skeletal system (9.3%). Clinical data of the individuals are available in Supplementary Table S2 and Supplementary Data.

3.2 Diagnostic rate of genome sequencing

In eight out of 30 individuals (26.7%), we identified at least one causative variant [class 4 or 5 of ACMG Guidelines (Richards et al., 2015)]. These included three single nucleotide variants (SNP) and three indels: a missense variant in *CYFIP2* in individual 9, a nonsense variant in *KMT2D* for individual 6 and in *TMEM147* for individual 12, and frameshift variants in *FOXG1*, *PURA* and *TMEM147* in individuals 7, 8, and 12, respectively. Three SVs were identified: one intragenic heterozygous deletion-inversion of 9.4 kb in *CASK* in individual 1, one partial intragenic heterozygous deletion of 37 kb in *GATAB2D* in individual 2, and one heterozygous

TABLE 1 Causative, candidate and excluded candidate genes of the cohort SNV, single nucleotide variant; indel, insertion-deletion; SV, structural variant. GRCh37-hg19 Genome Reference Consortium Human Build 37,NM_ c./r. Human Genome Variation Society nomenclature at the transcript or the RNA level p. nomenclature at the transcrip level ACMG American College of Medical Genetics and Genomics OMIM Online Mendelian Inheritance in Man.

	Gene Individual	(GRCh37—hg19) g.	(NM_) c./r.	p.	ACMG class	Inheritance	OMIM
Positive results							
SNV/ indel	*KMT2D Individual 6*	NC_000012.11:g.49426598G>A	NM_003482.3: c.11890C>T	p.(Gln3964*)	5	*De novo*	Kabuki syndrome # 147920
	FOXG1 Individual 7	NC_000014.8:g.29236741dup	NM_005249.4: c.256dup	p.(Gln86Profs*35)	5	*De novo*	Rett syndrome # 613454
	PURA Individual 8	NC_000005.9:g.139493864dup	NM_005859.4: c.98dup	p.(Gly34Argfs*167)	5	*De novo*	Mental retardation, autosomal dominant 31 # 616158
	CYFIP2 Individual 9	NC_000005.9:g.156754997A>G	NM_014376.2: c.2096A>G	p.(Asp699Gly)	5	*De novo*	Developmental and epileptic encephalopathy 65 # 618008
	TCF4 Individual 30	NC_000018.9:g.52926128C>T	NM_001083962.2:c.1069+1052G>A NM_001083962.2:r.1069_1070ins [1069+833_1069+1,049]	p.(Ala357Glyfs*7)	5	*De novo*	Pitt-Hopkins syndrome # 610954
	TMEM147 Individual 12	NC_000019.9:g.36036812_36036830del NC_000019.9: g.36038077C>G	NM_032635.3: c.100_118delc.486C>G	p.(Lys34Serfs*33) p.(Tyr162*)	5	Recessive mode of inheritance	*613585
SV	*CASK Individual 1*	NC_000023.10: g.41387135_41396533delins41391688_41391989inv	NM_003688.3:r.2156_2505del	p.(Asp719Glyfs*28)	5	*De novo*	Mental retardation, with or without nystagmus # 300422
	GATAD2B Individual 2	NC_000001.10: g.[?_ 153753742)_(153791156 _?]del			5	*De novo*	GAND syndrome # 615074
	SPTAN1 Individual 5	NC_000009.11: g.[?_ 131382516)_(131393966 _?]del	NM_001130438.2:r.5734_6762del	p.(Gly1912_Lys2254del)	5	*De novo*	Developmental and Epileptic Encephalopathy 5 # 613477
	MEF2C Individual 3	NC_000005.9:g.[88625547_88635553delinsTA; 88635554_90795688inv; 90795689_90795690del]			3	*De novo*	Mental retardation, stereotypic movements, epilepsy, and/or cerebral malformations # 613443
Candidate							
SNV/ indel	*POLA1 Individual 10*	NC_000023.10:g.25013973A>T	NM_016937.3 c.4295A>T	p.(Lys1432Ile)	3	Maternal	Van Esch-O'Driscoll # 301030
	ARI5B Individual 13	NC_000010.10:g.63845563del	NM_032199.2 c.1302del	p.(Asn434fs)	3	*De novo*	* 608538
	GRIN2B Individual 14	NC_000012.11:g.13893083A>G	NM_000834.3 c.1010+13168T>C	p.?	3	*De novo*	Intellectual developmental disorder, autosomal dominant 6, with or without seizures # 613970
SV	Chromoanagenesis *Individual 4*	Complex rearrangement involving chromosomes 6 and 11			3	*De novo*	

(Continued on following page)

10.3389/fcell.2022.1021785

TABLE 1 (*Continued*) Causative, candidate and excluded candidate genes of the cohort SNV, single nucleotide variant; indel, insertion-deletion; SV, structural variant. GRCh37-hg19 Genome Reference Consortium Human Build 37,NM_ c./r. Human Genome Variation Society nomenclature at the transcript or the RNA level p. nomenclature at the transcrip level ACMG American College of Medical Genetics and Genomics OMIM Online Mendelian Inheritance in Man.

	Gene Individual	(GRCh37—hg19) g.	(NM_) c./r.	p.	ACMG class	Inheritance	OMIM
Excluded candidate							
SNV/indel	*SENP6 Individual 15*	NC_000006.11:g.76350410C>T	NM_015571.2 c.469C>T	p.(Arg157*) p.(Arg157*)	2	Recessive mode of inheritance	
	FGD1 Individual 11	NC_000023.10:g.54492281G>A	NM_004463.3 c.1345C>T	p.(Arg449Cys)	2	Maternal	Intellectual developmental disorder, X-linked, syndromeic 16 # 305400

balanced inversion of about 2.2 Mb of a regulatory region of *MEF2C* in individual 3 (Table 1; Figure 2). The variants in *PURA*, *KMT2D*, and *FOXG1* had not been identified by ES because the capture kits utilized did not cover these regions. All these variants occurred *de novo* but the *TMEM147* variants followed a recessive mode of inheritance. Furthermore, *TMEM147* was initially identified as a new candidate gene, and data sharing and functional studies allowed us to confirm its causal role (Thomas et al., 2022).

We also identified at least a candidate disease-causing variant in seven additional individuals (23.3%). Six SNVs, including a hemizygous missense variant in *POLA1* in individual 10 and in *FGD1* in individual 11, both inherited from healthy mothers, a homozygous nonsense variant in *SENP6* in individual 15, two *de novo* deep intronic non-coding variants in *GRIN2B* and *TCF4* in individuals 14 and 30 respectively, and one *de novo* indel in *ARI5B* in individual 13 were identified. Furthermore, *de novo* complex structural variants involving two chromosomes (i.e., chromoanagenesis) were identified in individual 4 (Table 1; Figure 2). Data sharing allowed us to corroborate the suspected involvement of the *de novo ARI5B* variant in individual 13. Clinical and molecular data of the individuals are available in the Supplementary Data.

3.3 Diagnostic rate from RNA sequencing data

RNA-seq from whole blood was performed in 23 individuals (76.3% of the cohort): 11 undiagnosed individuals, eight with candidate genes, and five with positive GS. For the remaining seven individuals (23.3%), RNA-seq was not performed either because GS alone had already identified the causative variant (*KMT2D*, *PURA*, *CYFIP2*) or because the RNA was not available or did not pass the quality control standards (RIN ≥ 7). RNA-seq analysis confirmed the causal role of two variants in *CASK* and *GATAD2B* (2/23; 8.7%). In particular, an aberrant splicing event was found in individual 1, who harbored a *de novo* deletion-inversion of 9.4 kb in Xp11.4 involving *CASK*, while the partial deletion of *GATAD2B* was identified in RNA-seq data as an expression outlier due to nonsense-mediated mRNA decay, accompanied by gene expression down-regulation. RNA-seq also led to the identification of one additional diagnosis consisting of a *de novo* deletion in *SPTAN1* of about 11 kb, not detected by our CNV pipeline, associated with a splicing anomaly (Figure 3). The blood RNA-seq data from individual 30 did not allow us to confirm the pathogenic effect of the *de novo* intronic variant in *TCF4*, which was predicted to create a donor splice site. Indeed, *TCF4* expression was barely detectable in blood-derived RNA-seq data. However, we also obtained a fibroblast cell culture from the same patient, and the RNA-seq data from this sample revealed the retention of a cryptic exon of 218 nt, causing a frameshift variant. The nonsense-mediated

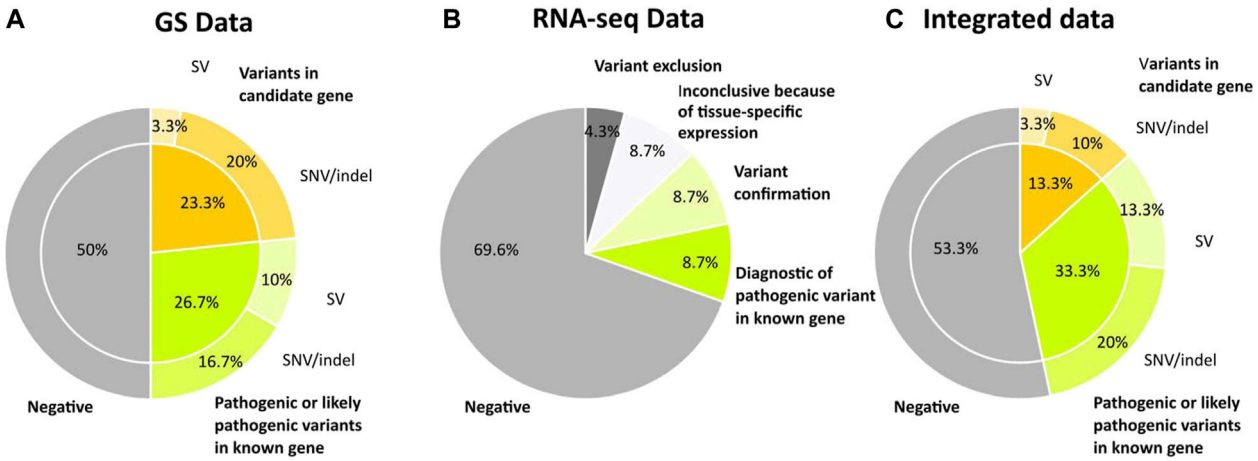


FIGURE 2
Diagnostic yield obtained with the different approaches deployed. Schematic representation of the evolution of the diagnostic yield in our cohort of 30 individuals with heterogeneous rare neurodevelopmental disorders. **(A)** The initial diagnostic yield with the genome sequencing (GS) data alone. **(B)** Contribution of RNA sequencing (RNA-seq) to the diagnostic rate in 23/30 patients. **(C)** Diagnostic yield obtained by integrating GS, RNA-seq and DNA methylation results. GS, genome sequencing; RNA seq, RNA sequencing; SV, structural variant; SNP, single nucleotide variants; indel, insertion deletion.

decay of the transcript carrying the cryptic exon was supported by the observation of a skewed allelic expression of an informative polymorphism in the 3′ end of the transcript (Supplementary Figure S1). The *TCF4* gene is responsible for Pitt-Hopkins syndrome, which is characterized by intellectual disability, wide mouth and distinctive facial features, and intermittent hyperventilation followed by apnea (MIM 602272) (Amiel et al., 2007). Reverse phenotyping was consistent with this syndrome.

Overall, RNA-seq identified two additional diagnoses (2/23; 8.7%) and independently confirmed two pathogenic variants already identified by GS (i.e., *CASK* and *GATAD2B*) (2/23; 8.7%). RNA-seq also allowed us to exclude the candidate variant in *SENP6* (1/23; 4.3%). Indeed, *SENP6* was not identified as a transcriptome outlier as the RNA-seq did not show any significant down-regulation of this gene, indicating that the nonsense variant p.(Arg157*) affected a minor isoform. These results were corroborated by a more accurate analysis of GTEx data, revealing that the RefSeq transcript NM_015571.4, corresponding to the MANE select transcript ENST00000447266.7 is ranked third in terms of abundance in all tissues and in particular in the central nervous system. Furthermore, this exon was also alternatively spliced in the computed GTEx gene model. Finally, RNA-seq did not show any monoallelic expression secondary to the *MEF2C* regulatory inversion or an aberrant splice event in *GRIN2B* in blood and fibroblast cell lines because the respective genes showed a neural-specific expression (2/23; 8.7%) (Figure 2), hence the analysis remained inconclusive for these variants. Splicing and expression abnormalities were validated by visual inspection of the RNA-seq alignment in the Integrative Genomics Viewer (IGV) (Robinson et al., 2017).

3.4 Analysis of DNA methylation profiles

DNA methylation profiles from whole blood were performed for all individuals from the same sample as was used for GS. EpiSign™ analysis revealed a genome-wide DNA methylation profile consistent with one of the 59 established episignatures in 10% (3/30) of cases assessed. All positive cases obtained a high confidence methylation variant pathogenicity (MVP) score of 1.0 (Supplementary Figure S2) with supportive multidimensional scaling (MDS) and hierarchical clustering. The patients positive for Episign episignatures were: individual 6 with a molecular diagnosis of Kabuki syndrome made by GS analysis (Kabuki syndrome due to variants in *KMT2D* or *KDM6A*), individual 11 with a clinical diagnosis of Coffin-Siris and negative GS and blood RNA-seq results (BAFopathy due to variants in *ARID1A*, *ARID1B*, *SMARCB1*, *SMARCA2* or *SMARCA4*), and individual 13 with a *de novo* variant in *ARID5B* (Wolf-Hirschhorn syndrome caused by deletions at 4p16.3). Interestingly, for individual 11, DNA methylation analysis also allowed us to exclude the implication of the variant of unknown significance (VUS) in *FGD1* identified by GS analysis. Furthermore, the analysis of genes involved in BAFopathies did not reveal any aberrant hypermethylation at promoters or gene body regions (Supplementary Figure S3). The visual inspection of genes involved in BAFopathies did not reveal any obvious structural variants. In addition, in individual 13, whose *ARID5B* candidate variant was found by GS, reverse phenotyping was not consistent with Wolf-Hirschhorn syndrome, and nor was a deletion in 4p16.3 found by array CGH and GS, suggesting that *ARID5B*

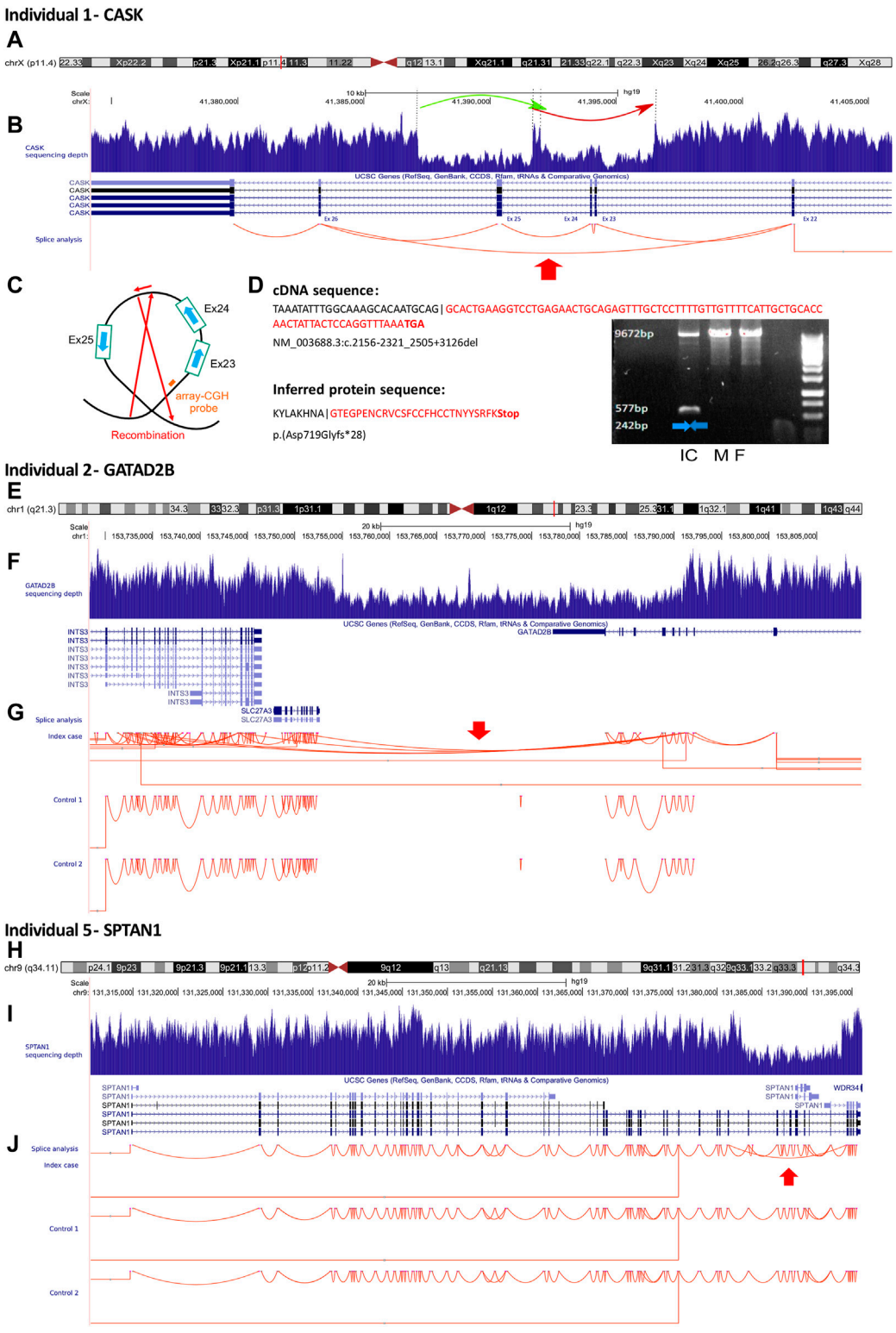


FIGURE 3
Illustration of individuals 1—CASK, 2—GATAD2B, and 5—SPTAN1. **(A)** Ideogram showing chromosome X and *CASK* localization. Under the ideogram, the green and red arrows represent the deletion-inversion of 9.4 kb in Xp11.4. **(B)** UCSC genome browser snapshot with visualization of *CASK* sequencing depth and *CASK* splice analysis. *CASK* sequencing depth demonstrates an intragenic deletion encompassing exons 23 through 25. The red arrow in the splicing analysis demonstrates the presence of a heterozygous transcript lacking exons 23 through 25. **(C)** Graphic

(*Continued*)

FIGURE 3 (Continued)
illustrating the mechanism of the loss of three exons from the *CASK* gene. The deletion was not detected by array CGH as the deleted region contained only one probe. **(D)** PCR confirmed deletion in proband but this was absent in two controls. Splice reads defined the readout of the event, identifying a frameshift variant. **(E)** Ideogram showing chromosome 1 and *GATAD2B* localization **(F)** UCSC genome browser snapshot showing sequencing depth, and **(G)** splice analysis at the *GATAD2B* locus. The sequencing depth shows a partial deletion of ~37 Kb of *GATAD2B* encompassing exons 5 through 11. The red arrow indicates the presence of a heterozygous transcript showing a fusion transcript between *GATAD2B* and *SLC27A3*. **(H)** Ideogram showing chromosome 9 and *SPTAN1* localization. **(I)** UCSC genome browser snapshot with visualization of sequencing depth and **(J)** splicing analysis at the *SPTAN1* locus. The sequencing depth shows the partial deletion of ~11 kb of *SPTAN1* including exons 44 through 51. The red arrow indicates the presence of a heterozygous transcript showing exon skipping in *SPTAN1*.

mutation may share some molecular biomarkers in common with Wolf-Hirschhorn syndrome.

Finally, two cases were inconclusive for the episignatures for Velocardiofacial syndrome (individual 23) and Rubinstein-Taybi syndrome (individual 25), as MVP elevation (<0.5) was insufficient and MDS and hierarchical clustering were inconsistent. Reverse phenotyping in individual 23 was not consistent with Velocardiofacial syndrome. Indeed, this individual showed a severe intellectual disability with microcephaly, myoclonic absence seizure, and hypoplasia of the corpus callosum. However, the clinical diagnostic hypothesis for individual 25 was Coffin-Siris syndrome.

All in all, we were able to diagnose ten out of 30 individuals (33.3%) and to have a candidate gene in four out of 30 individuals (13.3%) (Figure 2; Table 1; Supplementary Table S1).

3.5 Illustrative cases

3.5.1 Individual 1—*CASK*

Individual 1 was a 7-year-old girl, the only child of unaffected, non-consanguineous French parents. The pregnancy had been uncomplicated. She was born at 39 WG with normal birth length (48.5 cm, p37), weight (3290 g, p58), and OFC (32.5 cm, p11). The neonatal period was marked by poor feeding. All motor development milestones were delayed: she was able to sit independently at 9.5 months and walk at 2 years of age. She presented with delayed speech and language development. A brain MRI was performed and it was normal. Physical examination revealed no obvious dysmorphic features or microcephaly (−3.5 SD). She presented with bruxism. Previous genetic investigations, consisting of array CGH and trio ES, had been normal. GS identified a rearrangement of the *CASK* gene. This was a *de novo* deletion-inversion of 9.4 kb in Xp11.4 (Figure 3). RNA-seq identified an aberrant splicing event involving exons 23 through 25 skipping. This deletion was verified in qPCR. The *CASK* gene is involved in X-linked dominant intellectual disability with or without nystagmus (MIM 300422).

3.5.2 Individual 2—*GATAD2B*

Individual 2 was a 24-year-old male, the child of unaffected, non-consanguineous French parents. The pregnancy had been marked by ventriculomegaly at 22 WG. He was born at 41 WG with normal birth length (50.5 cm, p37) and weight (3040 g, p8), and macrocephaly with an OFC of 37.2 cm, (p92). During the neonatal period, he presented with hypotonia and poor feeding, followed by global developmental delay with language impairment and severe intellectual disability. Brain MRI was normal. His facial dysmorphisms included macrocephaly, prominent forehead, hypertelorism, and small, low-set ears. Physical examination revealed long toes, finger swelling and excessive wrinkling of palmar skin. He experienced hyperactivity in infancy, and subsequently short attention span, restricted behaviors, and sleep disturbance. Previous genetic investigations, including array CGH, screening for Sotos syndrome (*NSD1*) and Cowden syndrome (*PTEN*), intellectual disability panel, and trio ES, had returned normal results. GS identified a *de novo* partial deletion of ~37 kb of the *GATAD2B* gene with breakpoints within two AluY elements flanking the deleted region (Figure 3). This deletion was confirmed by a high-resolution array CGH but had not been identified by the first array CGH because of the lack of probes in this region. Transcriptome outlier detection confirmed the partial deletion of *GATAD2B*. The *GATAD2B* gene is responsible for the neurodevelopmental syndrome GAND, which combines hypotonia, psychomotor retardation, language disorders, intellectual disability, macrocephaly, and shared facial features (MIM. 615074). Reverse phenotyping was consistent with GAND syndrome.

3.5.3 Individual 3—*MEF2C*

Individual 3 was an 11-year-old girl, the second child of unaffected, non-consanguineous French parents. The pregnancy was uncomplicated. She was born at 38 WG with intrauterine growth retardation, birth length 45.5 cm (p7), birth weight 2630 g (p16), and OFC of 34.5 cm (p71). All motor development milestones were delayed: she was able to sit independently at 19 months, and was still unable to walk at 11 years of age. She presented with language impairment and behavioral problems such as abnormally aggressive, impulsive or violent behavior. A first EEG at 11 months of age showed some spike-wave discharges. At 8 years of age, EEG showed typical absence seizures. A brain MRI showed enlargement of the pericerebral spaces and slight hyperintensity of posterior cerebral white matter. She had facial dysmorphisms, including a prominent forehead, deep philtrum, and wide mouth with full lips. Previous genetic investigations, consisting of array CGH,

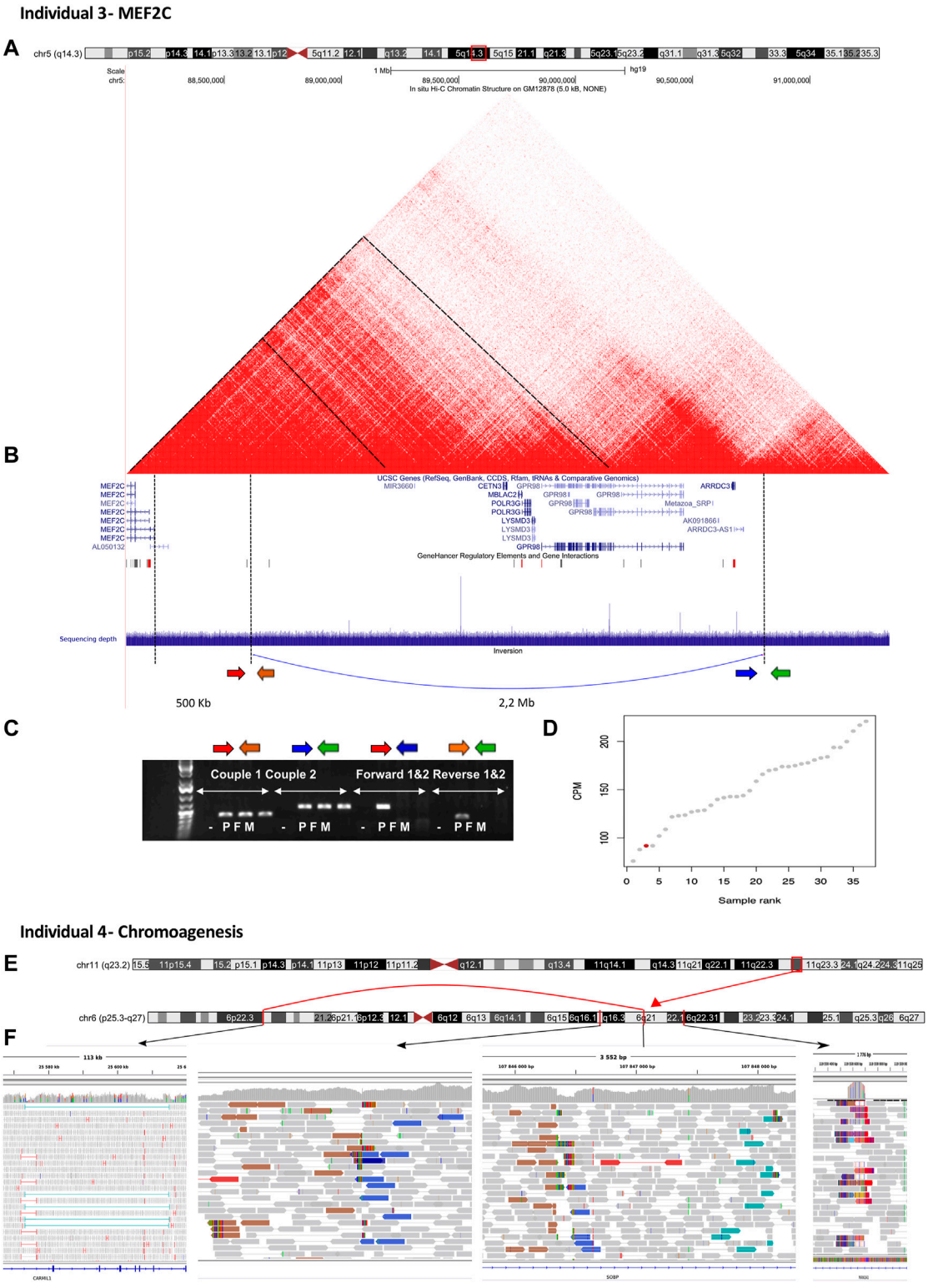


FIGURE 4
Illustration of individual 3—MEF2C and individual 4—Chromoanagenesis. **(A)** Ideogram showing chromosome 5 and *MEF2C* localization. **(B)** UCSC genome browser snapshot with visualization of three-dimensional (3D)-genome map at the 5q14.3 locus derived from Hi-C data of gM1287825 (10 kb resolution) and sequencing depth of this region. The structural variant characterized by a *de novo* inversion of 2.2 Mb in 5q14.3 is shown in blue under the sequencing depth track. It encompasses part of the regulatory region responsible for the neuronal expression of the *MEF2C* gene. The inversion did not include the gene body as it was located 500 Kb away from its proximal promoter region. This inversion is expected to deregulate *MEF2C via* its topologically associating domain dysfunction. **(C)** PCR analysis confirming the inversion. **(D)** *MEF2C* expression with RNA sequencing data. There was low expression of the MEF2C transcript in the individual; however, its biallelic expression in blood suggested that the regulatory regions affected by this inversion were specific to the neuronal lineage. **(E)** Ideogram showing chromosomes 11 and 6 affected by the complex rearrangement. **(F)** IGV visualization of the breakpoint located in chromosome 6.

screening for Angelman syndrome (methylation and sequencing of *UBE3A*) and Fragile X Syndrome (*FMR1*), sequencing of *FOXG1*, *CDKL5*, *STK9*, *RAI1*, *MECP2*, *MEF2C*, and trio ES, had returned normal results. GS identified a pathogenic structural variant characterized by a *de novo* inversion of 2.2 Mb in 5q14.3 encompassing part of the regulatory region responsible for the neuronal expression of the *MEF2C* gene. This rearrangement was confirmed on qPCR. MEF2C expression with RNA sequencing data showed low expression of the MEF2C transcript in this individual, although there was biallelic expression in blood, confirming that the regulatory regions affected by the inversion were specific for the neuronal lineage (Figure 4). The *MEF2C* gene is responsible for neurodevelopmental disorders with hypotonia, stereotypic hand movements, and impaired language (MIM 613443). Reverse phenotyping was consistent with this diagnosis.

3.5.4 Individual 4—chromoanagenesis

Individual 4 was a 15-year-old boy, the first child of unaffected, non-consanguineous French parents. The pregnancy had been uncomplicated. He was born at 41 WG with normal birth length (52 cm, p70) and weight (3700 g, p60), and macrocephaly, with an OFC of 37 cm (p90). All motor development milestones were delayed: he was able to sit independently at 18.5 months and to walk at 3 years and 3 months of age. He presented with language impairment. He had a severe intellectual disability. A brain MRI was performed and showed a retrocerebellar cyst. His facial dysmorphisms included brachycephaly, synophrys, epicanthus, small mouth, and pointed chin. Physical examination revealed global hypotonia, pectus excavatum, joint laxity, short fingers, and pes planovalgus. Previous metabolic and genetic investigations, including extensive metabolic screening, chromosome analysis, array CGH, Fragile X Syndrome testing (FMR1), intellectual disability panel, and trio ES, had returned normal results. GS led to the identification of a *de novo* complex rearrangement involving chromosomes 6 and 11 (Figure 4).

3.5.5 Individual 5—*SPTAN1*

Individual 5 was a 23-year-old man, the third child of unaffected, consanguineous Algerian parents. The pregnancy had been uneventful. He was born at 41 WG with normal birth length (53 cm, p85), weight (3520 g, p43), and OFC (35.5 cm, p55). He had severe gastroesophageal reflux requiring Nissen fundoplication. He had limited acquisition of motor skills for his age: he walked at 18 months. He presented with delayed speech and language development followed by severe intellectual disability. Brain MRI was normal. Physical examination revealed no obvious dysmorphic features, microcephaly (−2.5 SD), slender build, high palate, hypermobile finger joints, and myopia. He experienced attention deficit hyperactivity disorder. Previous genetic investigations, consisting of array CGH and trio ES, had returned normal results. GS detected no obvious anomalies. RNA-seq evidenced a splicing event in *SPTAN1*. This RNA splicing alteration consisting of exon skipping was validated by visual inspection of the RNA-seq alignment and then of the genome sequencing alignment in IGV. This SV of ~11 kb was associated with breakpoints at AluSx elements flanking the deleted region and was *de novo* (Figure 3). The *SPTAN1* gene is responsible for a broad spectrum of neurodevelopmental phenotypes characterized by moderate intellectual disability, with or without epilepsy and behavioral disorders (Syrbe et al., 2017). Reverse phenotyping was consistent with developmental and epileptic encephalopathy-5 (MIM 613477).

4 Discussion

Thirty individuals with malformation syndromes and/or severe neuro-developmental disorders and negative first-line trio ES were recruited from four centers in France. Short-read GS is becoming more affordable compared to other next-generation sequencing-based genomics technologies in diagnostics settings. In our study, the main explanation for the diagnostic yield of GS was the identification, with higher sensitivity, of genomic variations in coding and non-coding regions, such as indels (small insertion-deletions) not enriched by ES, copy-number variations (CNVs), and complex structural chromosomal rearrangements (Gilissen et al., 2014; Belkadi et al., 2015; Boycott et al., 2019; Burdick et al., 2020). Unbalanced structural variants below the detection limit of comparative chromosomal hybridization techniques are probably underdiagnosed in Mendelian disorders. GS represents a good candidate to overtake array CGH in the future, although identifying structural variants from NGS data still represents a challenge for bioinformatics (Mahmoud et al., 2019; Kobren et al., 2021). The use of GRCh38, which can be a better reference than GRCh37, can improve SV detection although it is not routine (Guo et al., 2017; Pan et al., 2019; Wagner et al., 2022). However, in our study, reanalyzing sequencing data using the GRCh38 reference genome did not lead to further diagnoses. We expect the use of the latest reference genome, obtained from the Telomere-2-Telomere consortium (Nurk et al., 2022) the optimization of bioinformatics pipelines, and the implementation of long-read sequencing technology and optical mapping approaches to improve CNV/SV detection (Chaisson et al., 2015; Chan et al., 2018; Logsdon et al., 2020). Finally, GS is far from being considered a comprehensive method to detect all types of genetic variants (mosaic variants, for example, require very deep sequencing of target regions) or to interpret the clinical implication of deep intronic variants (Boycott et al., 2019). In this respect, the integration of RNA-seq data is essential because they can identify variations in RNA abundance and sequence (i.e., gene expression outliers, allele-specific expression, splicing aberrations, and gene fusions). Thus far, several computational approaches have been developed

either for transcript abundance or differential splicing (Cummings et al., 2020; Mehmood et al., 2020; Shahjaman et al., 2020). Moreover, integrating ES or GS and transcriptome analyses has shown an increased diagnostic yield of 7.5%–35% depending on the tissue analyzed and the homogeneity of the disease studied (Kremer et al., 2018, 2017; Cummings et al., 2017; Frésard et al., 2019; Gonorazky et al., 2019; Hamanaka et al., 2019; Lee et al., 2020; Murdock et al., 2020; Stenton and Prokisch, 2020; Yépez et al., 2022). In our study, RNA-seq was performed on 23 out of 30 individuals with a combined diagnostic yield of 17.4% including the identification of one structural variant not detected by GS alone, the confirmation of an intronic variant of unknown significance observed by GS, and the confirmation of two causal variants identified by GS. Of note, it was possible to confirm the pathogenic role of the intronic *TCF4* variant due to the availability of a fibroblast cell line, utilized as a second-tier approach after RNA-seq in blood. This result, together with the failure to validate the effect of the *de novo* inversion in *MEF2C* regulatory region and the deep intronic *GRIN2B* variant, emphasizes the need to perform RNA-seq in clinically accessible samples that adequately represent splicing events in relevant but non-accessible tissues (Aicher et al., 2020). Often, clinically accessible tissues deployed in these studies are blood, skin, or muscle biopsies (e.g., whole blood, Epstein-Barr virus-transformed lymphocytes, fibroblasts, and myocytes). The expression of *MEF2C* in the brain is controlled by tissue-specific regulatory elements, and perturbation of their activity cannot be modelled in peripheral tissues. To overcome these limitations, iPS-derived cell lines are sometimes used to obtain a more suitable tissue for further analysis. In most cases, RNA-seq derived from fibroblasts exhibits higher and less variable gene expression in clinically relevant genes, as Murdock et al. showed in their cohort of 115 undiagnosed patients with diverse phenotypes (Murdock et al., 2020). Furthermore, RNA-seq allowed us to exclude one candidate variant, preventing a misdiagnosis.

Finally, using DNA methylation episignatures, which are highly sensitive and specific DNA methylation biomarkers, can result in the diagnosis of rare neurodevelopmental disorders (Aref-Eshghi et al., 2020, 2019; Sadikovic et al., 2021; Levy et al., 2022), allowing VUS in genes with an established episignature to be assessed or reclassified. In our analysis, DNA methylation corroborated one patient's molecular diagnosis of Kabuki syndrome. In another patient with a clinical diagnosis of Coffin-Siris syndrome, it found a positive episignature for a BAFopathy. However, he had negative GS and RNA-seq results, apart from a variant of unknown significance in *FGD1*, which was excluded from involvement following examination of its specific episignature. Further analyses will be required to identify the associated causal variants, including RNA sequencing using patient-derived fibroblasts and long-read sequencing or optical genome mapping. DNA methylation found also found the episignature for Wolf-Hirschhorn syndrome (WHS) in an individual with a *de novo* heterozygous *ARID5* variant. Further studies will be required to investigate the extent to which ARID5B shares differentially methylated regions with WHS. Moreover, our findings in two individuals were inconclusive. These results might be due to fewer penetrant variants, interference from a yet to be defined episignature or technical artifact.

Overall, the combined diagnostic yield of GS, RNA-seq, and DNA methylation analysis in our approach was 33.3%. We identified strong candidate variants for 13.3% additional patients that will require further functional validation. We expect the deployment of new bioinformatics pipelines for detecting SV/CNV, mobile element insertions or mitochondrial DNA genome variants (Garret et al., 2019; Niu et al., 2022) in combination with the development of new disease-associated episignatures and the advent of third generation genome sequencing or optical mapping to improve the identification of pathogenic genetic variants.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: ClinVar accession numbers: VCV000827810.2, SUB12094393, SUB12094652, VCV001708019.1, VCV001708028.1.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional review board of Dijon University Hospital (DC2011-1332). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

Author contributions

AV, YD, CT-R, and LF conceived and designed the study. AV, LF, and CT-R provided funding. EC, BI, MV, CM, DH, AA, SH, NJ-A, PC (19th author), SO, YH, AS, SM, and LF clinically evaluated the patients. AF provided technical assistance. EC, A-LB, and AV collected the data. MC, CP (27th author), VC, and VB performed the wet-lab work. AB, RO, and J-FD performed genome sequencing and RNA sequencing, YD and ET performed bioinformatic analyses, RR, JK, HM, and BS performed DNA Methylation analysis. FT, A-SD-P, HS, BK, PC (30th author), CP

(33rd author), CT-R, EC, and AV performed variant interpretation. EC, YD, and AV organized tables and figures. EC and AV wrote the paper with contributions from all authors who read and approved the submitted version.

Funding

The study was performed within the framework of the GAD (*Génétique des Anomalies du Développement*) collection and approved by the appropriate institutional review board of Dijon University Hospital (DC2011-1332). This work was supported by grants from Dijon University Hospital, the ISITE-BFC (PIA ANR), the European Union through the FEDER programs (PERSONALISE), and the Burgundy-Franche-Compté regional council (INTEGRA). The sequencing platform at the CNRGH was supported by the France Génomique National infrastructure, funded as part of the "Investissements d'Avenir" program, managed by the Agence Nationale pour la Recherche (contract ANR-10-INBS-09). The whole genome sequencing performed at the CNRGH was funded by the Laboratory of Excellence GENMED (Medical Genomics) Grant No. ANR-10-LABX-0013, managed by the National Research Agency (ANR) as part of the Investment for the Future program.

Acknowledgments

We are grateful to the families who have participated in this study. We thank the University of Burgundy Computing Cluster (CCuB). We also thank Dr. Agnès Guichet and Dr. Marine Tessarech for their cytogenetic advice, Dr. Celine Bris for proofreading the text and Dr Elke De Boer for critical reading.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcell.2022.1021785/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Illustration of TCF4 findings in individual 30. **(A)** Ideogram showing chromosome 18 and TCF4 localization. **(B)** IGV (integrative genomics viewer) visualization of GS results showing a *de novo* heterozygous deep intronic variant in *TCF4*. **(C)** Sashimi plot visualization showing the inclusion of a cryptic exon (chr18:g.52926121-52926348), predicted to result in a frameshift variant p.Ala357Glyfs*7. **(D)** IGV visualization of RNA-seq data in individual 30 and an unaffected control showing the monoallelic expression of an informative single nucleotide polymorphism located in the 3′ end of TCF4.

SUPPLEMENTARY FIGURE S2
EpiSign (DNA methylation) MVP scores from this cohort. A multi-class supervised classification system capable of discerning between multiple episignatures by generating a probability score (MVP) for each episignature. A positive score is typically greater than 0.5, and three patients produced an MVP of 1.0, indicating a methylation profile match for BAFopathy (red), Kabuki syndrome (green), and Wolf-Hirschhorn syndrome (purple). Two cases (dark grey) were inconclusive for Rubinstein-Taybi syndrome and Velocardiofacial syndrome as MVP elevation was insufficient. All remaining cases (light grey) were negative for all 59 episignatures analyzed.

SUPPLEMENTARY FIGURE S3
DNA methylation analysis of BAF complex gene promoters. DNA methylation analysis of promoter regions in individual 11 for **(A,B)** ARID1A **(D,D)** ARID1B **(E–G)** SMARCA2 **(H,I)** SMARCA4 **(J)** SMARCB1. All of them were within normal range.

SUPPLEMENTARY TABLE S1
Phenotype and genotype of all 30 individuals in the OMIXCARE cohort. F, female; M, male; ID, intellectual disability; N, no; Y, yes; NA, not available; /, absent; m, months; y, years; W, weight; H, height; HC, head circumference; EEG, electroencephalogram. Dark green signifies the identified genes, light green is for the candidate genes, and orange is for the rejected genes.

References

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

Amiel, J., Rio, M., de Pontual, L., Redon, R., Malan, V., Boddaert, N., et al. (2007). Mutations in TCF4, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* 80, 988–993. doi:10.1086/515582

Andrews, S.FastQC: a quality control tool for high-throughput sequence data. (2010). Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aref-Eshghi, E., Bend, E. G., Colaiacovo, S., Caudle, M., Chakrabarti, R., Napier, M., et al. (2019). Diagnostic utility of genome-wide DNA methylation testing in genetically unsolved individuals with suspected hereditary conditions. *Am. J. Hum. Genet.* 104, 685–700. doi:10.1016/j.ajhg.2019.03.008

Aref-Eshghi, E., Kerkhof, J., Pedro, V. P., Di France, G., Barat-Houari, M., Ruiz-Pallares, N., et al. (2020). Evaluation of DNA methylation episignatures for

diagnosis and phenotype correlations in 42 mendelian neurodevelopmental disorders. *Am. J. Hum. Genet.* 106, 356–370. doi:10.1016/j.ajhg.2020.01.019

Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q. B., Antipenko, A., et al. (2015). Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5473–5478. doi:10.1073/pnas.1418631112

Boycott, K. M., Hartley, T., Biesecker, L. G., Gibbs, R. A., Innes, A. M., Riess, O., et al. (2019). A diagnosis for all rare genetic diseases: The horizon and the next Frontiers. *Cell* 177, 32–37. doi:10.1016/j.cell.2019.02.040

Burdick, K. J., Cogan, J. D., Rives, L. C., Robertson, A. K., Koziura, M. E., Brokamp, E., et al.Undiagnosed Diseases Network (2020). Limitations of exome sequencing in detecting rare and undiagnosed diseases. *Am. J. Med. Genet. A* 182, 1400–1406. doi:10.1002/ajmg.a.61558

Chaisson, M. J. P., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F., et al. (2015). Resolving the complexity of the human genome using single-molecule sequencing. *Nature* 517, 608–611. doi:10.1038/nature13907

Chan, S., Lam, E., Saghbini, M., Bocklandt, S., Hastie, A., Cao, H., et al. (2018). Structural variation detection and analysis using bionano optical mapping. *Methods Mol. Biol.* 1833, 193–203. doi:10.1007/978-1-4939-8666-8_16

Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., et al. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genom. Med.* 3, 16. doi:10.1038/s41525-018-0053-8

Cummings, B. B., Karczewski, K. J., Kosmicki, J. A., Seaby, E. G., Watts, N. A., Singer-Berk, M., et al. (2020). *Genome Aggregation Database Production Team, Genome Aggregation Database Consortium*, 581, 452–458. doi:10.1038/s41586-020-2329-2. Transcript expression-aware annotation improves rare variant interpretation*Nature*

Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., Bolduc, V., Waddell, L. B., Sandaradura, S. A., O'Grady, G. L., Estrella, E., Reddy, H. M., Zhao, F., Weisburd, B., Karczewski, K. J., O'Donnell-Luria, A. H., Birnbaum, D., Sarkozy, A., Hu, Y., Gonorazky, H., Claeys, K., Joshi, H., Bournazos, A., Oates, E. C., Ghaoui, R., Davis, M. R., Laing, N. G., Topf, A., Kang, P. B., Beggs, A. H., North, K. N., Straub, V., Dowling, J. J., Muntoni, F., Clarke, N. F., Cooper, S. T., Bönnemann, C. G., and MacArthur, D. G.Genotype-Tissue Expression Consortium (2017). Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* 9, eaal5209. doi:10.1126/scitranslmed.aal5209

Ferreira, C. R. (2019). The burden of rare diseases. *Am. J. Med. Genet. A* 179, 885–892. doi:10.1002/ajmg.a.61124

Frésard, L., and Montgomery, S. B. (2018). Diagnosing rare diseases after the exome. *Cold Spring Harb. Mol. Case Stud.* 4, a003392. doi:10.1101/mcs.a003392

Frésard, L., Smail, C., Ferraro, N. M., Teran, N. A., Li, X., Smith, K. S., Bonner, D., Kernohan, K. D., Marwaha, S., Zappala, Z., Balliu, B., Davis, J. R., Liu, B., Prybol, C. J., Kohler, J. N., Zastrow, D. B., Reuter, C. M., Fisk, D. G., Grove, M. E., Davidson, J. M., Hartley, T., Joshi, R., Strober, B. J., Utiramerur, S., Lind, L., Ingelsson, E., Battle, A., Bejerano, G., Bernstein, J. A., Ashley, E. A., Boycott, K. M., Merker, J. D., Wheeler, M. T., and Montgomery, S. B.Undiagnosed Diseases Network; Care4Rare Canada Consortium (2019). Identification of rare-disease genes using blood transcriptome sequencing and large control cohorts. *Nat. Med.* 25, 911–919. doi:10.1038/s41591-019-0457-8

Garret, P., Bris, C., Procaccio, V., Amati-Bonneau, P., Vabres, P., Houcinat, N., et al. (2019). Deciphering exome sequencing data: Bringing mitochondrial DNA variants to light. *Hum. Mutat.* 40, 2430–2443. doi:10.1002/humu.23885

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature* 511, 344–347. doi:10.1038/nature13394

Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., et al. (2019). Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* 104, 466–483. doi:10.1016/j.ajhg.2019.01.012

Guo, Y., Dai, Y., Yu, H., Zhao, S., Samuels, D. C., and Shyr, Y. (2017). Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics* 109, 83–90. doi:10.1016/j.ygeno.2017.01.005

Hamanaka, K., Miyatake, S., Koshimizu, E., Tsurusaki, Y., Mitsuhashi, S., Iwama, K., et al. (2019). RNA sequencing solved the most common but unrecognized NEB pathogenic variant in Japanese nemaline myopathy. *Genet. Med.* 21, 1629–1638. doi:10.1038/s41436-018-0360-6

Hartley, T., Lemire, G., Kernohan, K. D., Howley, H. E., Adams, D. R., and Boycott, K. M. (2020). New diagnostic approaches for undiagnosed rare genetic diseases. *Annu. Rev. Genomics Hum. Genet.* 21, 351–372. doi:10.1146/annurev-genom-083118-015345

Kobren, S. N., Baldridge, D., Velinder, M., Krier, J. B., LeBlanc, K., Esteves, C., Pusey, B. N., Züchner, S., Blue, E., Lee, H., Huang, A., Bastarache, L., Bican, A., Cogan, J., Marwaha, S., Alkelai, A., Murdock, D. R., Liu, P., Wegner, D. J., Paul, A. J., Sunyaev, S. R., and Kohane, I. S.Undiagnosed Diseases Network (2021). Commonalities across computational workflows for uncovering explanatory variants in undiagnosed cases. *Genet. Med.* 23, 1075–1085. doi:10.1038/s41436-020-01084-8

Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., et al. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8, 15824. doi:10.1038/ncomms15824

Kremer, L. S., Wortmann, S. B., and Prokisch, H. (2018). Transcriptomics": Molecular diagnosis of inborn errors of metabolism via RNA-sequencing. *J. Inherit. Metab. Dis.* 41, 525–532. doi:10.1007/s10545-017-0133-4

Lee, H., Huang, A. Y., Wang, L.-K., Yoon, A. J., Renteria, G., Eskin, A., Signer, R. H., Dorrani, N., Nieves-Rodriguez, S., Wan, J., Douine, E. D., Woods, J. D., Dell'Angelica, E. C., Fogel, B. L., Martin, M. G., Butte, M. J., Parker, N. H., Wang, R. T., Shieh, P. B., Wong, D. A., Gallant, N., Singh, K. E., Tavyev Asher, Y. J., Sinsheimer, J. S., Krakow, D., Loo, S. K., Allard, P., Papp, J. C., Palmer, C. G. S., Martinez-Agosto, J. A., and Nelson, S. F.Undiagnosed Diseases Network (2020). Diagnostic utility of transcriptome sequencing for rare Mendelian diseases. *Genet. Med.* 22, 490–499. doi:10.1038/s41436-019-0672-1

Levy, M. A., McConkey, H., Kerkhof, J., Barat-Houari, M., Bargiacchi, S., Biamino, E., et al. (2022). Novel diagnostic DNA methylation episignatures expand and refine the epigenetic landscapes of Mendelian disorders. *HGG Adv.* 3, 100075. doi:10.1016/j.xhgg.2021.100075

Logsdon, G. A., Vollger, M. R., and Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nat. Rev. Genet.* 21, 597–614. doi:10.1038/s41576-020-0236-x

Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., and Sedlazeck, F. J. (2019). Structural variant calling: The long and the short of it. *Genome Biol.* 20, 246. doi:10.1186/s13059-019-1828-7

McInerney-Leo, A. M., Marshall, M. S., Gardiner, B., Coucke, P. J., Van Laer, L., Loeys, B. L., et al. (2013). Whole exome sequencing is an efficient, sensitive and specific method of mutation detection in osteogenesis imperfecta and Marfan syndrome. *Bonekey Rep.* 2, 456. doi:10.1038/bonekey.2013.190

Mehmood, A., Laiho, A., Venäläinen, M. S., McGlinchey, A. J., Wang, N., and Elo, L. L. (2020). Systematic evaluation of differential splicing tools for RNA-seq studies. *Brief. Bioinform.* 21, 2052–2065. doi:10.1093/bib/bbz126

Murdock, D. R., Dai, H., Burrage, L. C., Rosenfeld, J. A., Ketkar, S., Müller, M. F., et al. (2020). Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. *J. Clin. Invest.* 131, 141500. doi:10.1172/JCI141500

Niu, Y., Teng, X., Zhou, H., Shi, Y., Li, Y., Tang, Y., et al. (2022). Characterizing mobile element insertions in 5675 genomes. *Nucleic Acids Res.* 50, 2493–2508. doi:10.1093/nar/gkac128

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., et al. (2022). The complete sequence of a human genome. *Science* 376, 44–53. doi:10.1126/science.abj6987

Pan, B., Kusko, R., Xiao, W., Zheng, Y., Liu, Z., Xiao, C., et al. (2019). Similarities and differences between variants called with human reference genome HG19 or HG38. *BMC Bioinforma.* 20, 101. doi:10.1186/s12859-019-2620-0

Picard, T.Broad Institute (2018). Available at: http://broadinstitute.github.io/picard/

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine : official journal of the American College of Medical Genetics* 17 (5), 405–424. doi:10.1038/gim.2015.30

Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., and Mesirov, J. P. (2017). Variant review with the integrative genomics viewer. *Cancer Res.* 77, e31–e34. –e34. doi:10.1158/0008-5472.CAN-17-0337

Sadikovic, B., Levy, M. A., Kerkhof, J., Aref-Eshghi, E., Schenkel, L., Stuart, A., et al. (2021). Clinical epigenomics: Genome-wide DNA methylation analysis for the diagnosis of mendelian disorders. *Genet. Med.* 23, 1065–1074. doi:10.1038/s41436-020-01096-4

Shahjaman, M., Manir Hossain Mollah, M., Rezanur Rahman, M., Islam, S. M. S., and Nurul Haque Mollah, M. (2020). Robust identification of differentially expressed genes from RNA-seq data. *Genomics* 112, 2000–2010. doi:10.1016/j.ygeno.2019.11.012

Stenton, S. L., and Prokisch, H. (2020). The clinical application of RNA sequencing in genetic diagnosis of mendelian disorders. *Clin. Lab. Med.* 40, 121–133. doi:10.1016/j.cll.2020.02.004

Syrbe, S., Harms, F. L., Parrini, E., Montomoli, M., Mütze, U., Helbig, K. L., et al. (2017). Delineating SPTAN1 associated phenotypes: From isolated epilepsy to encephalopathy with progressive brain atrophy. *Brain* 140, 2322–2336. doi:10.1093/brain/awx195

Thomas, Q., Motta, M., Gautier, T., Zaki, M. S., Ciolfi, A., Paccaud, J., et al. (2022). Bi-allelic loss-of-function variants in TMEM147 cause moderate to profound intellectual disability with facial dysmorphism and pseudo-Pelger-Huët anomaly. *Am. J. Hum. Genet.* S0002-9297 (22), 1909–1922. –3. doi:10.1016/j.ajhg.2022.08.008

Veeramah, K. R., Johnstone, L., Karafet, T. M., Wolf, D., Sprissler, R., Salogiannis, J., et al. (2013). Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia* 54, 1270–1281. doi:10.1111/epi.12201

Wagner, J., Olson, N. D., Harris, L., McDaniel, J., Cheng, H., Fungtammasan, A., et al. (2022). Curated variation benchmarks for challenging medically relevant autosomal genes. *Nat. Biotechnol.* 40, 672–680. doi:10.1038/s41587-021-01158-1

Yépez, V. A., Gusic, M., Kopajtich, R., Mertes, C., Smith, N. H., Alston, C. L., et al. (2022). Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med.* 14, 38. doi:10.1186/s13073-022-01019-9