Integrating RNA-Seq into genome sequencing workflow enhances the analysis of structural variants causing neurodevelopmental disorders

Kevin Riquin, Bertrand Isidor, Sandra Mercier, Mathilde Nizon, Estelle Colin, Dominique Bonneau, Laurent Pasquier, Sylvie Odent, Xavier Maximin Le Guillou Horn, Gwenaël Le Guyader, et al.

Integrating RNA-Seq into genome sequencing workflow enhance the analysis of structural variants causing neurodevelopmental disorders.

Kevin Riquin[1], Bertrand Isidor[1,2], Sandra Mercier[1,2], Mathilde Nizon[1,2], Estelle Colin[3,4], Dominique Bonneau[3,4], Laurent Pasquier[5], Sylvie Odent[5,6], Xavier Le Guillou Horn[7,8], Gwenael le Guyader[7], Annick Toutain[9,10], Vincent Meyer[11], Jean-François Deleuze[11], Olivier Pichon[2], Martine Doco-Fenzy[1,2], Stéphane Bézieau[1,2], Benjamin Cogné[1,2]

1 Nantes Université, CHU de Nantes, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France
2 Nantes Université, CHU de Nantes, Service de Génétique médicale, F-44000 Nantes, France
3 CHU Angers, Service de Génétique médicale, 49933 Angers Cedex 9, France
4 UMR CNRS 6214-INSERM 1083, Université d'Angers, 49933 Angers Cedex 9, France
5 Service de Génétique Clinique, ERN ITHACA, CHU Rennes, Rennes, France
6 Institut de Génétique et Développement de Rennes, IGDR UMR 6290 CNRS, INSERM, IGDR Univ Rennes, Rennes, France
7 Service de génétique médicale, CHU de Poitiers, Poitiers France
8 LabCom I3M-Dactim mis / LMA CNRS 7348, Université de Poitiers, Poitiers, France
9 UF de Génétique Médicale, Centre Hospitalier Universitaire, 37044 Tours, France
10 UMR 1253, iBrain, Université de Tours, INSERM, 37032 Tours, France
11 Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine (CNRGH), 91057, Evry, France

Abstract

Background: Molecular diagnosis of neurodevelopmental disorders (NDD) is mainly based on exome sequencing (ES), with a diagnostic yield ranging from 31% for isolated and 53% for syndromic NDD. As sequencing costs decrease, genome sequencing (GS) is gradually replacing ES for genome-wide molecular testing. As many variants detected by GS only are in deep intronic or non-coding regions, the interpretation of their impact may be difficult. Here, we showed that integrating RNA-Seq into the genome sequencing workflow can enhance the analysis of the molecular causes of NDD, especially structural variants, by providing valuable complementary information such as aberrant splicing, aberrant expression, and monoallelic expression.

Methods: We performed trio-GS on a cohort of 33 individuals with NDD for whom ES was inconclusive. RNA-Seq on skin fibroblasts was then performed in nine individuals for whom GS was inconclusive and optical genome mapping (OGM) was performed in two individuals with a structural variant (SV) of unknown significance.

Results: We identified pathogenic or likely pathogenic variants in 16 individuals (48%) and six variants of uncertain significance. RNA-seq contributed to the interpretation in three individuals, and OGM helped to characterize two SVs.

Conclusion: Our study confirmed that GS significantly improves the diagnostic performance of NDDs. However, most variants detectable by GS alone are structural or located in non-coding regions, which can pose challenges for interpretation. Integration of RNA-Seq data overcame this limitation by confirming the impact of variants at the transcriptional or regulatory level. This result paves the way for new routinely applicable diagnostic protocols.

Key messages:

WHAT IS ALREADY KNOWN ON THIS TOPIC:

Genome sequencing has shown that it can outperform the standard diagnostic procedures for neurodevelopmental disorders. Routine implementation in diagnostic laboratories poses new challenges, particularly in the detection and interpretation of structural or non-coding variants.

WHAT THIS STUDY ADDS:

The application of complementary techniques, such as RNA-Seq and optical genome mapping, on a case-by-case basis, may help enhance the diagnostic yield of genome sequencing for complex or difficult-to-interpret cases.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY:

Clinical RNA-Seq is a growing subject of research. However, its application in routine diagnostics has not yet been standardized. Our study provides novel insights into its gradual introduction, in association with genome sequencing.

Introduction

Neurodevelopmental disorders (NDD) are a group of behavioral and cognitive disorders that occur during the developmental period and involve significant difficulties in the acquisition and performance of specific intellectual, motor, language, or social functions [1]. Among this disorders, intellectual disability have frequently been associated with monogenic causes. Owing to cumulative efforts to improve molecular diagnosis, more than 1400 causative genes have been identified [2,3]. Although next-generation sequencing has significantly improved the diagnosis, most individuals still face many years of diagnostic odyssey.

Currently, molecular diagnosis is mainly based on exome sequencing (ES), which is a cost-effective technique with a diagnostic yield ranging from 31% for isolated NDD, and 53%

for syndromic NDD [4]. Thus, the molecular cause is still not identified in most of the cases. Possible reasons for this high proportion of unexplained cases are the inability of ES to detect variants in deep intronic and non-coding regions or difficulties in detecting structural variants (SVs). Over the past decade, studies have shown that genome sequencing (GS) outperforms the standard ES + array-CGH procedure with GS-unique diagnostic yields ranging from 19% to 25% [5–8].

Nevertheless, many variants detected in deep intronic or non-coding regions raise interpretation issues. Transcriptomic analysis can provide valuable complementary information to interpret noncoding variants, including aberrant splicing, aberrant expression, and monoallelic expression. To date, RNA-Seq studies have been particularly successful for neuromuscular disorders [9] for which appropriate tissue is accessible, with a diagnostic rate of up to 35%. Recent studies have shown an increased diagnostic rate of up to 17% after applying RNA-Seq to GS-negative individuals with diverse disorders [10]. To our knowledge, this approach has rarely been specifically applied to a cohort of individuals with neurodevelopmental disorders[11,12].

Here, we performed trio GS on a cohort of 33 individuals with NDD for whom previous genetic analyses were negative, including array-CGH and ES. RNA-Seq was then performed in nine probands with inconclusive GS. We also characterized two SVs by optical genome mapping (OGM). Finally, we report interesting cases.

Methods

Recruitment, Inclusion Criteria

33 families from the University Hospitals of Western France (HUGO network) were recruited based on suspected neurodevelopmental disorders. Phenotypic data were collected as Human Phenotype Ontology (HPO) terms. All probands presented with syndromic or nonsyndromic moderate-to-severe intellectual disability or global developmental delay. Parents were asymptomatic. Individuals were excluded if a non-genetic etiology was suspected or if a parent was symptomatic. The detailed clinical phenotypes are listed in Sup.Table 2. Informed consent for genetic analysis was obtained from the legal guardians according to the French law on bioethics and following the Helsinki Declaration. The study was approved by the ethics committee of Nantes University Hospital (number CCTIRS:14.556). Solo- or trio-ES and array-CGH have already been performed and were negative for each proband. RNA-seq was proposed to any individual for whom a variant of uncertain significance or no variant was identified after the GS.

Genome sequencing variant calling and annotation

Genome sequencing was performed on 33 trio at the French National Research Center for Human Genomics (CNRGH, Evry, France). Library generation was performed using DNA extracted from blood samples with TruSeq®DNA PCR-free kit (Illumina, San Diego, CA, USA). The genomes were sequenced on a HiSeq X5 (Illumina, San Diego, CA, USA) at a minimum mean depth of 30X. DNA sequences were mapped to the reference human

genome sequence (GRCh37) using bwa 0.7.12, and single nucleotide variants (SNVs) and small insertions/deletions (INDELs) were identified following GATK's best practices (v3.4). Copy number variations (CNVs), structural variants (SV), and short tandem repeats (STRs) were detected using a custom pipeline (https://gitlab.univ-nantes.fr/kriquin/sv-genome). Briefly, this pipeline combines algorithms with complementary calling methods and annotation tools. It is based on four algorithms: CNVs are detected using the read depth method GATK 4.1.4.1 (Broad Institute) with a sliding window of 1000 bp. SV are called by MANTA 1.6.0 (Illumina Inc.) [13] and DELLY 0.8.7 (European Molecular Biology Laboratory) [14]. Short tandem repeats were genotyped by Expansion Hunter de novo 0.9.0 (Illumina Inc.) [15] and *de novo* mobile element insertions were called in individuals negative after GS using MELT 2.2.2 [16]. All structural variants were annotated using AnnotSV 2.2 [17] and a prioritization tag was assigned based on inheritance status, allelic frequency of the variant in databases and localization. Identified variants were classified using the ACMG standard for the interpretation of sequence variants [18]. The degree of consanguinity was estimated using Automap 1.2 [19]. All SVs were confirmed by karyotyping, quantitative PCR, or optical genome mapping (OGM).

Diagnostic yield calculations

The overall diagnostic yield was determined as the proportion of probands in the cohort with a definitive molecular diagnosis. The GS-unique diagnostic yield was calculated as the percentage of probands in the cohort who received a conclusive molecular diagnosis that would have been unattainable solely through the reanalysis of ES data.

Optical genome mapping (OGM)

High-molecular-weight DNA was extracted from frozen whole peripheral blood EDTA according to the manufacturer's instructions (Bionano Genomics, San Diego, CA, USA). Labeled DNA was loaded onto a Saphyr chip for linearization and imaging on the Saphyr instrument. After checking QC metrics, the *de novo* assembly and variant annotation pipeline were run using Bionano Solve v.7.1 software at 80X coverage. Each optical genome map was compared with a reference genome map. SVs were visualized using Bionano Access software V7.1 and compared to an optical genome map dataset of 300 human population control samples from apparently healthy individuals (provided by Bionano Genomics) to filter out common SVs and potential artifacts.

Prediction tools

The impact of missense mutations was predicted using Mobidetails [20]. The effect of the mutation on protein stability was predicted using I-Mutant 3.0 [21] and DUET [22]. The parameters used were: T=25°C and pH = 7. The PDB accession number of the structure of the APP protein used was 4PWQ [23].

RNA-Sequencing

RNA-Sequencing was performed on skin fibroblasts of nine individuals and one parent in two separate batches. The library was generated using NEBNext Ultra™ II Directional RNA Library Prep (New England Biolabs, Ipswich, MA, USA). RNA was sequenced using NextSeq 500 (Illumina, San Diego, CA, USA) with approximately 60-70 million reads per

sample. Reads were aligned using STAR 2.6.1d [24] in two pass-mode (GRCh37 release75, Gencode34) and duplicate marked with Picard (Broad Institute). Analysis of RNA-Seq data was performed using Drop 1.1.4 [25]. Aberrant expression was detected using OUTRIDER [26] with external matrix of 127 samples (https://doi.org/10.5281/zenodo.7510845). Aberrant splicing was detected using FRASER [27] without external matrix. Splicing events were filtered as follows: pvalue ≤ 0.01, padjGene < 1, DeltaPsi ≤ -0.10 or DeltaPsi ≥ 0.10. Monoallelic expression was detected using the MAE module of Drop and filtered as follows: padj ≤ 0.05, cohort_freq ≤ 0.1, AFmax ≤ 0.01 [28,29]. We quantified the gene expression in whole blood, cultured fibroblasts, and cultured urine-derived stem cells (sup. Method) from donors without NDD. Using Kallisto 0.46.2 [30] (Kallisto index: Ensembl 99; GRCh38), we generated TPM values for the expressed genes in each sample. Genes with TPM > 10 were considered well expressed [31].

Results

Cohort

Our cohort was composed of 33 individuals, 19 of whom were female (57%). The median age was 12 years (range, 2–31 years). All individuals have syndromic or nonsyndromic

intellectual disability (ID) (HP:0001249) or global developmental delay (GDD) (HP:0001263). Clinical data are available in Sup.Table 2. All individuals had a previous negative array-CGH and ES (Trio-ES 26/33, Solo-ES for 7/33). GS with a minimum mean depth of 30X was performed for each parent–child trio. RNA-seq was then proposed to all individual for whom a variant of uncertain significance or no variant was identified by GS. Nine families accepted.

Diagnostic yield

We report a diagnosis in 16 individuals (16/33; 48%) (Table 1; Sup.Table 1; Figure 1). Seven *de novo* variants and one inherited variant were identified in genes associated with an autosomal dominant neurodevelopmental disorder (8/33; 24%). Five cases (5/33; 15%) were resolved through the identification of biallelic variants in genes associated with autosomal recessive inheritance. In addition, we identified a hemizygous pathogenic variant inherited from the mother in an X-linked gene in one male participant, and a *de novo* variant in an X-linked gene in two female participants. No candidate variants involving short tandem repeats or mobile element insertion were identified.

SNV/INDELs:

SNVs or INDELs supported a diagnosis in ten individuals (10/16; 62%). Of these, seven were identified thanks to the reanalysis effect. They were detected by ES but not classified as pathogenic at the time of analysis. Better support for pathogenicity, a recent publication of the genes involved, or additional investigations allowed us to revise this

classification. Among these variants, we reconsidered a maternally inherited nonsense variant in *MN1*. Indeed, the variant was identified as de novo in another affected individual [32] with very similar dysmorphic facial features. After analysis in maternal grand-parents, the variant was found *de novo* in the mother. Although the variant appeared heterozygous in the blood, we suspected mosaicism in the healthy mother. Additionally, a low-level mosaic frameshift variant in *NIPBL* was identified through targeted NGS analysis of saliva following inconclusive GS and ES. Notably, two cases were resolved by SNVs in exons of *H3-3A* that were not covered by the kit at the time of ES. These cases may have been resolved by renewing ES.

SV:

Pathogenic or likely pathogenic SVs were identified in six individuals (6/16; 37%). One balanced event was identified, a *de novo* inversion with a breakpoint in FOXP1, as previously reported [33]. The five others were CNVs that were too small to be detected by ES or Array-CGH. We also found 4 *de novo* SV of uncertain significance. Two were *de novo* balanced events with breakpoints in the intergenic regions and no genes directly involved. One was a complex event involving the duplication of *FGF18* associated with an inverted insertion.

Contribution of complementary techniques

OGM:

OGM helped in the characterization of two SVs. In the first case (P25), we suspected an inversion of 2.2Mb following a 500kb duplication at the *TENM2/FGF18* locus. The OGM

showed an inverted insertion, rather than an inversion/duplication. In the second case (P24), GS detected a duplication in *PUM1* with uncertainty in the breakpoint coordinates, and the OGM characterized this as an intragenic tandem duplication.

RNA-Seq:

We additionally collected skin biopsies from 10 individuals (nine probands and one parent) with inconclusive GS and performed RNA-Seq on cultured fibroblasts. We identified two cases with an abnormal splicing event caused by an intragenic CNV: skipping of two exons in *CBX3* (P23) and an abnormal splicing event in *PUM1* (P24). We also detected one sample with significant underexpression of the long non-coding RNA (lncRNA) *CHASERR*, which could be linked to *de novo* CNV. RNA-seq contributed to the diagnosis or interpretation of the variant in three individuals (3/9; 33%) (Sup Table 3).

To evaluate the relevance of fibroblasts as a source of biological material for clinical RNA-seq analysis, we quantified and compared gene expression levels among blood samples, fibroblasts, and an alternative cell type known as Urine Derived Stem Cells (USCs) (sup. Method). We observed that approximately 60% of NDD-associated genes were expressed in fibroblasts with TPM>10 and 37% were expressed in blood. In addition, *CBX3*, *PUM1* and *CHASERR* were expressed at a level that theoretically allowed their detection in blood samples (https://kriquin-univ-nantes.shinyapps.io/expression-rna-seq). Interestingly, USCs obtained by non-invasive means expressed approximately 63% of the NDD-associated genes. This could make them an interesting alternative to the fibroblasts.

Case/Sex	Type	Gene	Variant	Transmission	Interpretation	ACMG	Reason not reported by ES	Technique applied
P2/F	SNV indel	*NIPBL*	NC_000005.9:g.37059021_37059022del (Mosaicism 11%)	*De novo* [MOS]	Solved [MIM :122470]	5 (PVS1, PS2, PM2)	Mosacism Variant not found in blood	GS Panel
P4/M	SNV indel	*H3-3A*	NC_000001.10 :g.226259121G>C NM_002107.7 :c.352G>C :p.(Val118Leu)	*De novo* [HET]	Solved [MIM:619720]	5 (PP2, PS2, PM2, PP3, PM1)	Exon not covered	GS
P5/M	SNV indel	*MN1*	NC_000022.10 :g.28192754C>A NM_002430.3 :c.3778G>T :p.(Glu1260*)	Inherited from mother [MOS]	Solved [MIM :618774]	5 (PVS1,PS1,PM2)	Nonconcordant segregation	GS
P6/F	SNV indel	*H3-3A*	NC_000001.10:g.226259146A>G NM_002107.7:c.377A>G:p.(Gln126Arg)	*De novo* [HET]	Solved [MIM :619720]	5 (PS2,PM1,PM2, PP2,PP3)	Exon not covered	GS
P7/F	SNV indel	*TFE3*	NC_000023.10:g.48895930A>G NM_006521.6:c.572T>C:p.(Leu191Pro)	*De novo* [HET]	Solved [MIM :301066]	4 (PS2,PM1,PM2, PP3)	Gene now linked to ID	GS
P9/F	SNV indel	*BAP1*	NC_000003.11:g.52442077C>G NM_004656.4:c.272G>C:p.(Cys91Ser)	*De novo* [HET]	Solved [MIM :619762]	5 (PS2,PM1,PM2, PM5,PP3)	More support for pathogenicity	GS
P13/M	SNV indel	*HACE1*	NC_000006.11:g.105232328_105232331del NM_020771.4:c.1439_1442del:p.(Val480Alafs*7) - NC_000006.11:g.105297081_105297084del NM_020771.4:c.259_262del:p.(Lys87Glufs*27)	Autosomal recessive [COMP HET]	Solved [MIM:616756]	5 (PVS1,PM2,PM3) - 5 (PVS1,PM2,PM3)	Gene now linked to ID	GS
P15/F	SNV indel	*INTS11*	NC_000001.10:g.1248089C>T NM_017871.6:c.1295-9G>A:p.? - NC_000001.10:g.1248275T>C NM_017871.6:c.1186A>G:p.(Lys396Glu)	Autosomal recessive [COMP HET]	Solved [34]	4 (PM2,PM3,PP3, PP4) - 4 (PS3,PM2,PP3)	More support for pathogenicity	GS
P17/F	SNV indel	*TMEM147*	NC_000019.9:g.36036812_36036830del NM_032635.4:c.100_118del:p.(Lys34Serfs*33) - NC_000019.9:g.36038077C>G NM_032635.4:c.486C>G:p.(Tyr162*)	Autosomal recessive [COMP HET]	Solved [MIM:613585]	5 (PVS1,PM2,PM3) - 5 (PVS1,PM2,PM3)	More support for pathogenicity	GS

P28/M	SNV indel	*FITM2*	NC_000020.10:g.42935539A>C NM_001080472.4:c.515T>G:p.(Val172Gly) - NC_000020.10:g.42935585A>C NM_001080472.4:c.469T>G:p.(Phe157Val)	Autosomal recessive [COMP HET]	Solved [MIM :618635]	4 (PM1,PM2,PP3 PP4) - 4 (PM1,PM2,PP3, PP4)	Gene now linked to ID	GS
P1/F	SV	*FOXP1*	NC_000003.11:g.68954396_71064931inv NM_001244813.1:c.570-127_*2054002inv	*De novo* [HET]	Solved	5 (PVS1,PS2,PM2)	Copy neutral inversion not detected by ES	GS
P10/F	SV	*RSPRY1*	NC_000016.9:g.57230856_57252612del NM_133368.1:c.-155-7560_901+1665del	Autosomal recessive [HOM]	Solved [MIM :616723]	5 (PVS1,PM2,PM3)	Not detected	GS
P14/F	SV	*STEEP1*	NC_000023.10:g.118673832_118675628del NM_022101.3:c.514-245_607-80del	*De novo* [HET]	Solved [MIM : 301013]	4 (PVS1,PS2)	Not detected	GS
P24/F	SV	*PUM1*	NC_000001.10:g.3140901_31422000dup NM_001020658.1:c.2856+974_3435+489dup	*De novo* [HET]	Solved	4 (PS2,PM1,PM2)	Not detected	GS RNA-seq OGM
P27/M	SV	*SLC6A8*	NC_000023.10:g.152957273_152958417del NM_001142805.1:c.645-157_778-79del	Inherited from mother [HEM]	Solved [MIM :300352]	5 (PVS1,PS3,PM2)	Not detected	GS
P32/M	SV	*CHASERR*	NC_000015.9:g.93422237_93430600del	*De novo* [HET]	Solved	N/A	Non coding gene	GS RNA-seq

Table 1: Variants identified in this study. ACMG classe and evidence categories are determined following ACMG/AMP Standards and Guidelines. HOM: Homozygous; HEM: Hemizygous; HET: Heterozygous; COMP HET: Compound heterozygous; MOS: Mosaicism. NC_000001.10:g.31409001_31422000dup

Case reports

Individual P19 - Homozygous VUS in *APP*

This case involved a male presenting with severe ID, lack of language, abnormal behavior, autism, and epilepsy. Parents are related (their fathers are half-brothers). A homozygous variant inherited from healthy heterozygous parents was identified in the *APP* gene NM_000484.4:c.(440A>G) p.(His147Arg). This variant is located on chromosome 21 in a 7.08 Mb region with 99.96% homozygosity (Sup.figure 1). No other homozygous candidate variants have been identified in this region.

APP is a ubiquitous cell-surface receptor capable of dimerization. The most studied isoform is the 695 amino acid protein which is expressed at the surface of neurons. It has been observed that APP dimerization at the neuron membrane participates in neurite growth, neuronal adhesion, axonogenesis, and synaptogenesis. Mutations have been identified in the Aβ domain and are associated with early-onset Alzheimer's disease, but few damaging variants outside this domain have been documented. However, it is interesting to note that a homozygous nonsense variant NM_000484.3:c.1075C>T p.(Arg359*) associated with decreased somatic growth, microcephaly, hypotonia, developmental delay, thinning of the corpus callosum, and seizures has already been reported [35].

Nine in silico prediction tools classified this variant as damaging. Furthermore, the algorithms for predicting the impact on the stability of protein, I-Mutant and DUET, both

predicted a strong decrease in the stability of the H147R mutated protein (respectively ΔΔG = -0.75 Kcal/mol; ΔΔG = -1.07 Kcal/mol).

Histidine 147 is located in the CuBD subdomain of the E1 domain (Figure 2). It is one of the histidines involved in the chelation of copper ions. Copper binding to histidine 147 induces histidine-bridging between APP molecules. Previous experimental studies have shown the impact of His147 mutations or CuBD deletion on APP dimerization and processed fragment secretion levels. Collectively, these elements make this variant an interesting candidate for further functional studies.

Individual P14 - Deletion in *STEEP1*

This case involved a female presenting with severe GDD, no language, and axial hypotonia. MRI showed a slight widening of the pericerebral space. The SNP array showed a 6q22.33 deletion inherited from the father. Using GS we identified a *de novo* 1.7kb deletion of *STEEP1* (*CXorf56*) on chromosome X. This deletion in-frame spans exon 6 and parts of its flanking intronic sequences NM_022101.3:c.514-245_607-80del (Sup.figure 2). According to the ENCODE ChIP-seq assay [36], H3K4me1 and H3K2Ac histone mark enrichment in this region suggests the presence of regulatory elements in the vicinity of this deletion. The absence of this SV in the gnomAD database (gnomAD SVs v2.1) and the pLI score of 0.95, support its pathogenicity. Moreover, *STEEP1* loss of function has been associated with ID in several families. In these studies, male phenotypes were described as mild or severe ID with additional features such as epilepsy, abnormal gait, or abnormal

reflexes. In females, penetrance is incomplete, and expressivity is variable. Reported female cases showed non-syndromic mild intellectual disabilities and asymptomatic carriers, depending on the skewed X-inactivation pattern [37,38]. According to the ACMG guidelines, this case was classified as solved.

Individual P32 - Deletion in *CHASERR* lncRNA

This individual is a male presenting with GDD and facial dysmorphism. Clinical examination revealed widely spaced eyes, anteverted nares, a long philtrum, axial hypotonia, peripheral hypertonia, bilateral optic atrophy, recurrent unexplained fevers, and dystonic movements. At 33 months, he could not sit and did not speak, had recurrent unexplained fevers, and dystonic movements. Brain MRI at one month was normal, but MRI at four years showed global dysmyelination, thinning of the corpus callosum, reduced brainstem volume, cortical atrophy, and ventriculomegaly.

By GS we identified a *de novo* heterozygous deletion NC_000015.9:g.93422237_93430600del. This deletion was also detected by RNA-Seq (Figure 3). *CHASERR* is located approximately 1.5kb upstream of *CHD2*, whose haploinsufficiency has already been implicated in developmental and epileptic encephalopathy (MIM:615369). However, the deletion does not cover the *CHD2* promoter and the individual's phenotype is not consistent with the loss of expression of this gene. The regulatory role of *Chaserr* in *Chd2* expression has been previously reported in a mouse model. The absence of Chaserr results in a significant increase in *Chd2* expression at both

the mRNA and protein levels, subsequently causing transcriptional interference [39]. Therefore, a deleterious mechanism by the overexpression of *CHD2* was suspected. RNA-Seq allowed us to confirm skewed allelic expression. We noted an allelic imbalance skewed towards the *CHD2* allele in cis with the *CHASERR* deletion, consistent with what has been found in animal models (Sup.figure 3). No aberrant expression of *CHD2* was detected by Drop. However, the low level of expression of this gene in the blood and fibroblasts could hamper its ability to detect significant variations. We confirmed here that *CHASERR* regulates *CHD2*, an NDD-associated gene. As another case with concording phenotype was described we classified this case as resolved [10]. Nevertheless, further investigation is needed to clarify the underlying pathological mechanisms.

We also observed significant overexpression of *LIMD1* (FC=1.82; padj=4.25E-05). We did not detect any GS variants in this gene, and manual verification with IGV provided no explanation for this event. Moreover, this gene is not associated with neurodevelopment. These elements raise the question of possible regulation by *CHASERR*. However, this hypothesis remains unconfirmed.

Individual P23 - Deletion in *CBX3*

This case involved a female, the first child of non-consanguineous parents, presenting with moderate GDD, infantile muscular hypotonia, talipes valgus, strabismus, hypertrichosis, and epicanthus. She sat at 11 months, stood at 21 months, and had bi-syllabic language at 18 months.

Trio genome sequencing revealed a balanced event with a breakpoint in the *NUS1* promoter, a gene associated with an autosomal dominant ID. However, karyotyping did not confirm this SV (Sup.figure 4). Subsequent RNA-seq revealed aberrant splicing of *CBX3* (Figure 4. A). A junction was detected between exons 2 and 5, which was absent in the mother and controls (deltaPsi: 0.47). This event was consistent with an in-frame deletion of exons 3-4. This deletion was detected in GS, but it appeared dubious in the IGV overview at first glance. Moreover, this gene was not associated with NDD; therefore, this variant seemed less relevant than the putative translocation involving *NUS1*. Close examination of the IGV overview confirmed that there was indeed a *de novo* 3.8kb deletion of exons 3-4 of *CBX3* in this individual NM_007276.4:c.25-1502_330+141del (Figure 4. B). This deletion results in a highly truncated protein (83 aa vs. 183 aa) with loss of the chromodomain, which is the main functional domain of CBX3 (Figure 4. C). In an animal model, it was shown that CBX3 is required for H4K20me3 epigenetic mark deposition and the regulation of protocadherin genes, which are involved in the mechanisms of neuronal self-avoidance [40]. We classified this event as a strong candidate VUS.

Individual P24 - Intragenic duplication in *PUM1*

This individual is a female born to non-consanguineous parents. Pregnancy was notable for oligohydramnios. Examination revealed severe ID and dysmorphism, including a wide mouth, protruding tongue, and macrocephaly. Unilateral renal hypoplasia and gastrointestinal anomalies, such as esodeviation, gastroesophageal reflux, and chronic constipation, were also observed. Brain MRI showed a periventricular leukomalacia.

Using GS, we identified a *de novo* heterozygous duplication of 13 kb in *PUM1* NM_001020658.1:c.2856+974_3435+489dup involving two L2c retrotransposons with 531bp of homology (Figure 5). OGM allowed us to characterize this event as intragenic tandem duplication. RNA-Seq analysis confirmed that the mRNA contains a 3'exon21 - 5'exon18 junction and a single last exon, as shown by a heterozygous SNP. PUM1 belongs to the PUF protein family. These RNA-binding proteins act as post-transcriptional repressors by binding to specific sequences located in the 3'-UTR of mRNA [41,42]. This in-frame duplication spans five protein domains involved in RNA binding, suggesting a deleterious impact. In addition, haploinsufficiency of *PUM1* has been reported to be responsible for GDD with concordant phenotype [43]. Therefore, we classified this case as resolved.

Discussion:

In this study, 33 individuals with NDD who were negative after ES and array-CGH were enrolled for trio-GS complemented by RNA-seq for 9 of them and OGM for 2 individuals. GS identified pathogenic or likely pathogenic variants in 16 individuals and achieved an

overall diagnostic yield of 48%. Nevertheless, it is interesting to note that a large majority of these pathogenic variants (10/16, 62,5%) were SNVs/indels; therefore, it was theoretically possible to detect them in ES. Seven of the 16 solved cases were supported by variants detected in ES but not reported at the time of the first analysis. Therefore, it is necessary to qualify the net contribution of the trio-GS technique by considering the reanalysis effect. This effect is essentially due to the publication of new gene-disease associations, and its intensity is directly correlated to the time elapsed since the last analysis [44]. Excluding the reanalysis effect, we reported a diagnostic for eight individuals reaching a GS-unique diagnostic rate of 24% which is in line with the 19%-25% reported by previous studies [6–8]. Among them, six cases were solved by SVs (6/8, 75%).

Our study highlights that short-read GS may present limitations in the interpretation of SVs. For example, duplication involving *PUM1* required OGM to be fully characterized. The use of long read GS could also overcome this limitation. Moreover, unlike OGM, it allows the precise identification of breakpoints and provides additional information on DNA methylation status. However, the current cost of these techniques does not facilitate their implementation in diagnostic laboratories.

RNA-Seq contributed to the identification of molecular causes in two probands and a strong candidate in one. This was the case for the *CBX3* deletion in individual P23. Even though it was detected in GS, the IGV preview was dubious, and this variant was eliminated. Doubts regarding the IGV preview can be attributed to three reasons: the overall number of false-positive SVs detected in this individual, the absence of this gene in

the NDD databases, and the detection of a more compelling variant. In this individual, RNA-seq confirmed the variant in *CBX3* and eliminated other candidates. Furthermore, its effect on the transcript was immediately delineated, confirming the existence of a truncated transcript rather than a total loss of expression. This case highlights one of the advantages of using RNA-seq in addition to inconclusive. Recent algorithms for the analysis of GS data have shown high sensitivity at the cost of many false positives [45]. Verification of these numerous variants is a time-consuming task prone to error and can therefore lead to discarding the causal variant. RNA-seq appears here as a technique that can counterbalance a low specificity of GS in SVs detection by proposing a short list of variants with proven transcriptional effects.

The deletion in the *CHASERR* lncRNA was detected by GS; however, to our knowledge, no guidelines for the prediction of the impact of variants in non-coding RNAs have been validated for diagnosis. In this case, the identification of an allelic bias in the *CHD2* gene in -cis of the deletion is consistent with a pathogenic mechanism, previously observed in mouse and cell models. As GS implementation becomes progressively more widespread, an increasing number of variants, especially SVs, will be identified in non-coding RNAs generating a significant number of VUS. Interpretation of these events based on GS data alone is challenging.

However, clinical RNA-seq has several limitations. Notably, RNA-seq provides information that may be unusual for biologists and clinicians. In particular, we identified highly significant aberrant expression or aberrant splicing confirmed in IGV, but it has not been

possible to link these to a genomic variant. Understanding these phenomena and their medical relevance is time-consuming and can only be envisaged for a limited number of patients. In addition, the small number of families that accepted to benefit from RNA-seq did not allow us to draw any conclusions on the systematic application of RNA-seq in patients with inconclusive GS. However, it appears from our study that integration on a case-by-case basis is useful for filtering and interpreting SVs.

The choice of the biological tissue must also be considered. The first diagnostic RNA-seq tests were performed on blood samples. However, this tissue offered a low diagnostic yield for most diseases, except for hematological and immune diseases. A comparative study showed better performance of RNA-seq in fibroblasts [10]. Explanations for the low yield of whole-blood assays include the cellular heterogeneity of this tissue and the large transcriptional variations that can affect leukocytes. However, we cannot rule out that the expression and splicing defects identified in *CBX3*, *PUM1*, and *CHASERR*, which are expressed in blood, could have been identified in blood samples.

We observed that approximately 60% of NDD-associated genes [2] were expressed in fibroblasts with TPM>10, while only 37% were expressed in blood, making fibroblasts more relevant for NDD diagnosis. However, the skin biopsy required to collect fibroblasts may compromise their widespread use. The identification of alternative clinically accessible tissues (CAT), such as buccal cells, hair follicles, or urine cells, could help overcome this limitation. We performed gene expression analysis on one alternative cell type, urine-derived stem cells [46], and observed that they expressed approximately 60% of

the NDD-associated genes (TPM >10). This might be a promising alternative CAT for RNA-Seq studies.

It should be noted that the choice of tissue remains a compromise between biological relevance and practical application. Even if a gene associated with neurodevelopment is expressed in a cell type, tissue-specific differences exist in gene expression or splicing. Furthermore, it is illusory to attempt to recapitulate the expression and the splicing pattern complexity of neurodevelopment with a single cell type.

The application of RNA-seq on a case-by-case basis in patients with inconclusive GS presenting with a severe phenotype seems to be an interesting approach. Diagnostic RNA-seq is still in its infancy, and there are several areas for improvement; however, it is on the way to becoming a standard tool to enhance molecular diagnosis in NDD.

Acknowledgments:

This study was performed thanks to a collaboration between CEA and Nantes hospital. This study was supported by the program "High throughput sequencing and rare diseases" of the Foundation for Rare Diseases. This work has been carried out within the framework of the FHU GenOMedS thanks to the support of the Health cooperation group of University Hospitals of the Great West (GCS HUGO) and the National Alliance for Life Sciences and Health (Aviesan). The CEA-CNRGH sequencing platform 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)". We are most grateful to the Bioinformatics Core Facility of Nantes BiRD, member of Biogenouest, Institut Français de Bioinformatique (IFB) (ANR-11-INBS-0013) for the use of its resources and for its technical support.

Funding:

This work was supported by the Foundation for Rare Diseases and Groupama Foundation.

Competing of interests:

None declared.

Contributorship statement:

Conceptualisation, supervision of the overall project and edition of the final manuscript draft : KR, BC, SB.

Validation, writing—original draft: KR, BC, SB.

Data collection : BI, SM, MN, EC, DB, LP, SO, XGH, GG, AT, KR, OP, MDF, BC, SB.

Data analysis and interpretation : KR, OP, MDF, BC.

All authors critically revised the manuscript for important intellectual content. They take full responsibility for the integrity of the data and the accuracy of the data analysis.

Guarantors: KR, SB, BC.

Ethics approval statement :

Informed consent was obtained form participants or the legal guardians according to the French law on Bioethics and following the Helsinki Declaration.This study was approved by the CHU de Nantes ethics committee (Research Programme "Génétique Médicale" DC-2011-1399).

Figures :

Figure 1: Performance of GS for NDD diagnosis after inconclusive ES. A) Overall diagnostic yield of the combined approach. The GS-unique diagnostic yield was calculated as the percentage of probands in the cohort who received a conclusive molecular diagnosis that would have been unattainable solely through the reanalysis of ES data. Solved by the reanalysis effect are cases solved owing to variants that were detected by ES but were not classified as pathogenic at the time of analysis. Better support for

pathogenicity, a recent publication of the genes involved, or additional investigations allowed us to revise their classification. B) Types of variants identified at the case level. No compound heterozygous involving both SNV/indel and SVs were reported.

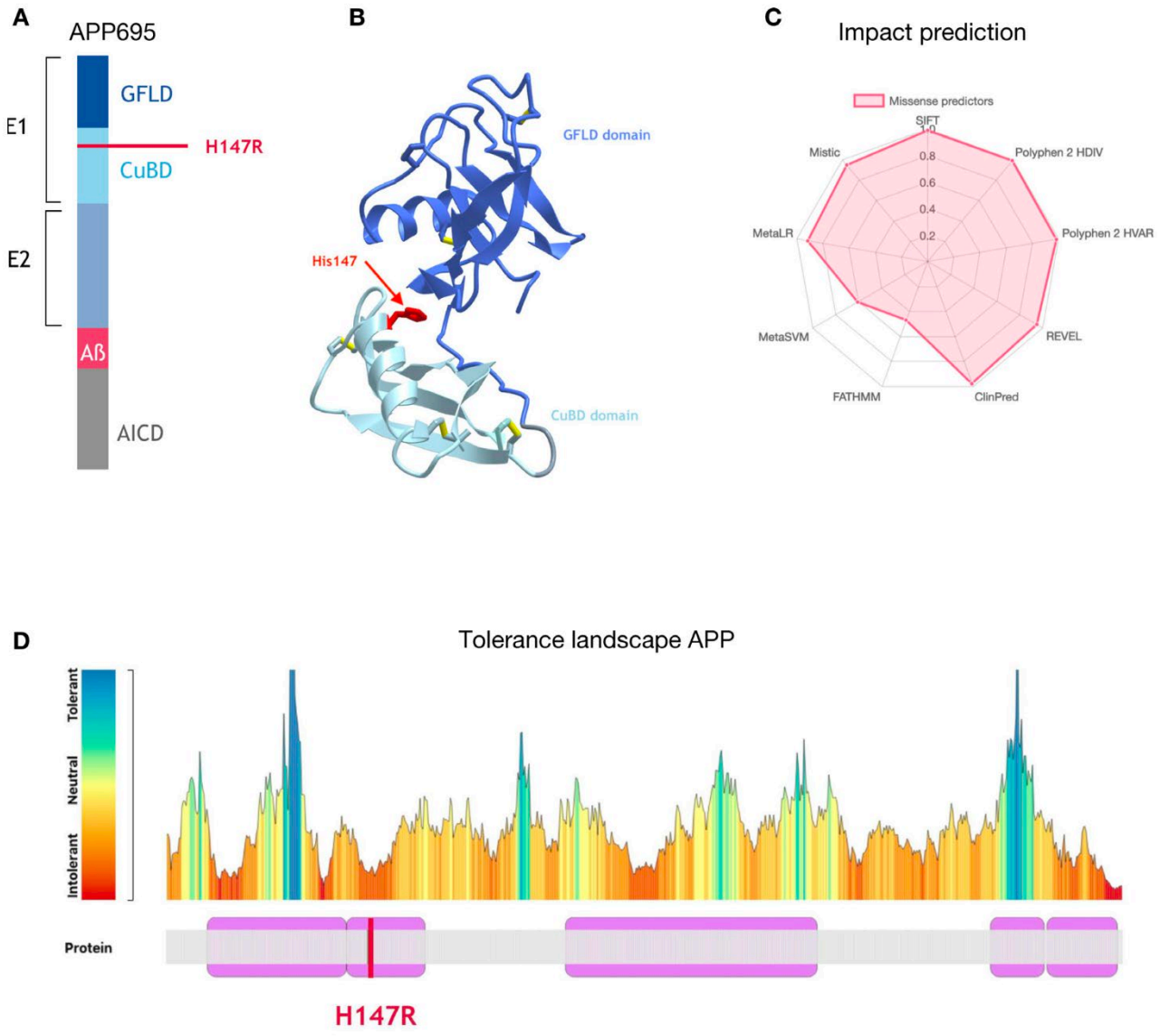


Figure 2: Homozygous VUS in *APP;* NM_000484.3:c.[440A>G] [440A>G p.[H147R] [H147R]. A) The H147 variant is located in the CuBD (copper-binding domain) subdomain of the E1 extracellular domain. Histidine 147 is directly involved in copper binding. B) Protein structure (PDB accession number 4PWQ). C) H147 is predicted pathogenic by nine prediction tools aggregated by Mobidetails. D) Tolerance landscape of APP (695aa) generated by Metadome. H147 is located in a region predicted to be intolerant to mutation.

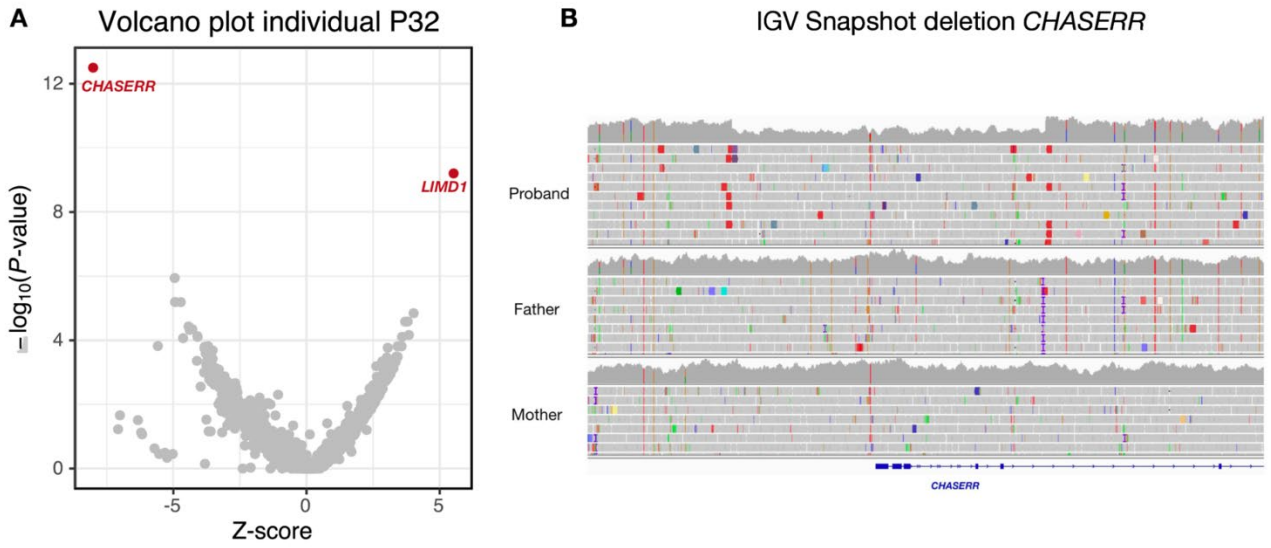


Figure 3: *De novo* heterozygous deletion in lncRNA *CHASERR*; NR_037600.1:g.92,879,007_92,887,374del. A) Volcano plot from OUTRIDER module in individual P32 showing aberrant expression of *CHASERR* in fibroblasts. B) IGV snapshot showing a 8.4 kilobases deletion covering the *CHASERR* promoter and exons 1-3.

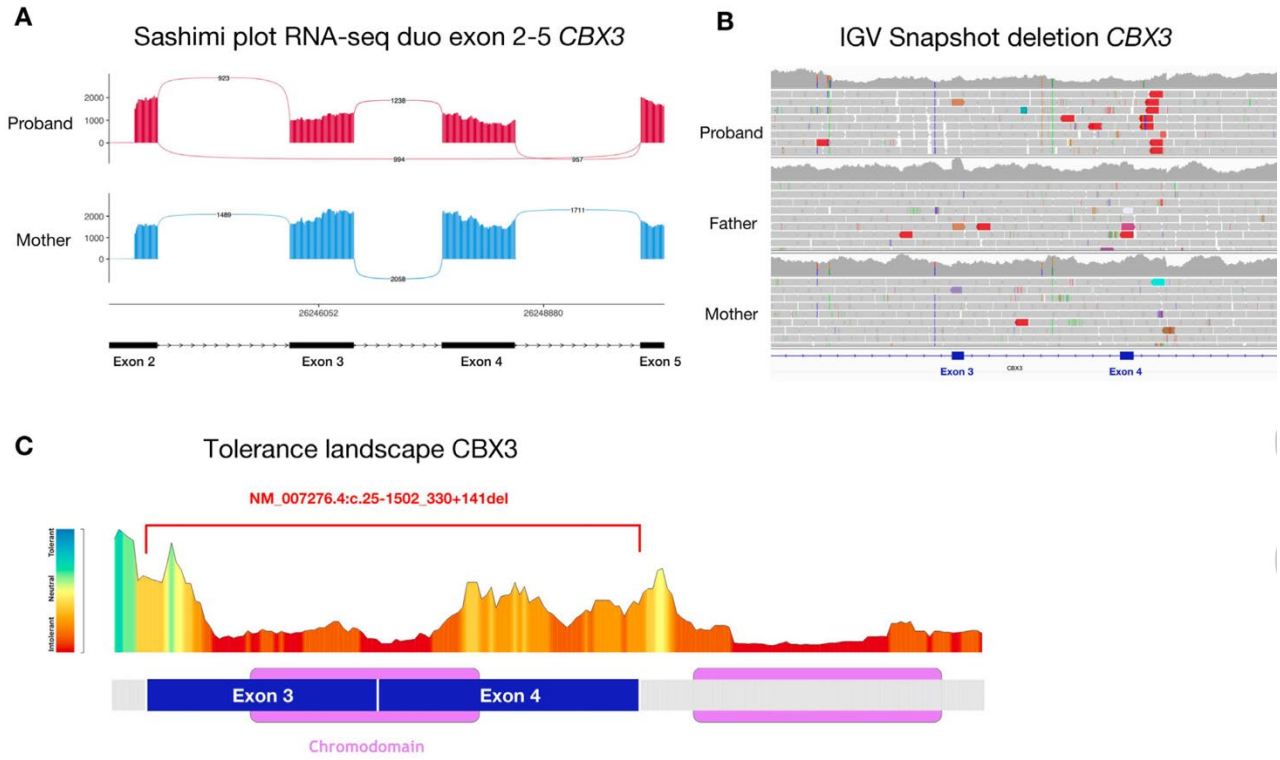


Figure 4: *De novo* heterozygous deletion in *CBX3*; NM_007276.4:c.25-1502_330+141del. A) Sashimi plot RNA-Seq on fibroblasts from P23 individual and healthy mother showing a heterozygous aberrant junction between exons 2 and 5 of CBX3. B) IGV snapshot showing a 3.8 kilobases deletion covering exons 3-4. C) Tolerance landscape of CBX3 (183 aa) generated by Metadome. The exons 3-4 codes for a region predicted to be highly intolerant to mutation. This region contains the functional chromodomain.

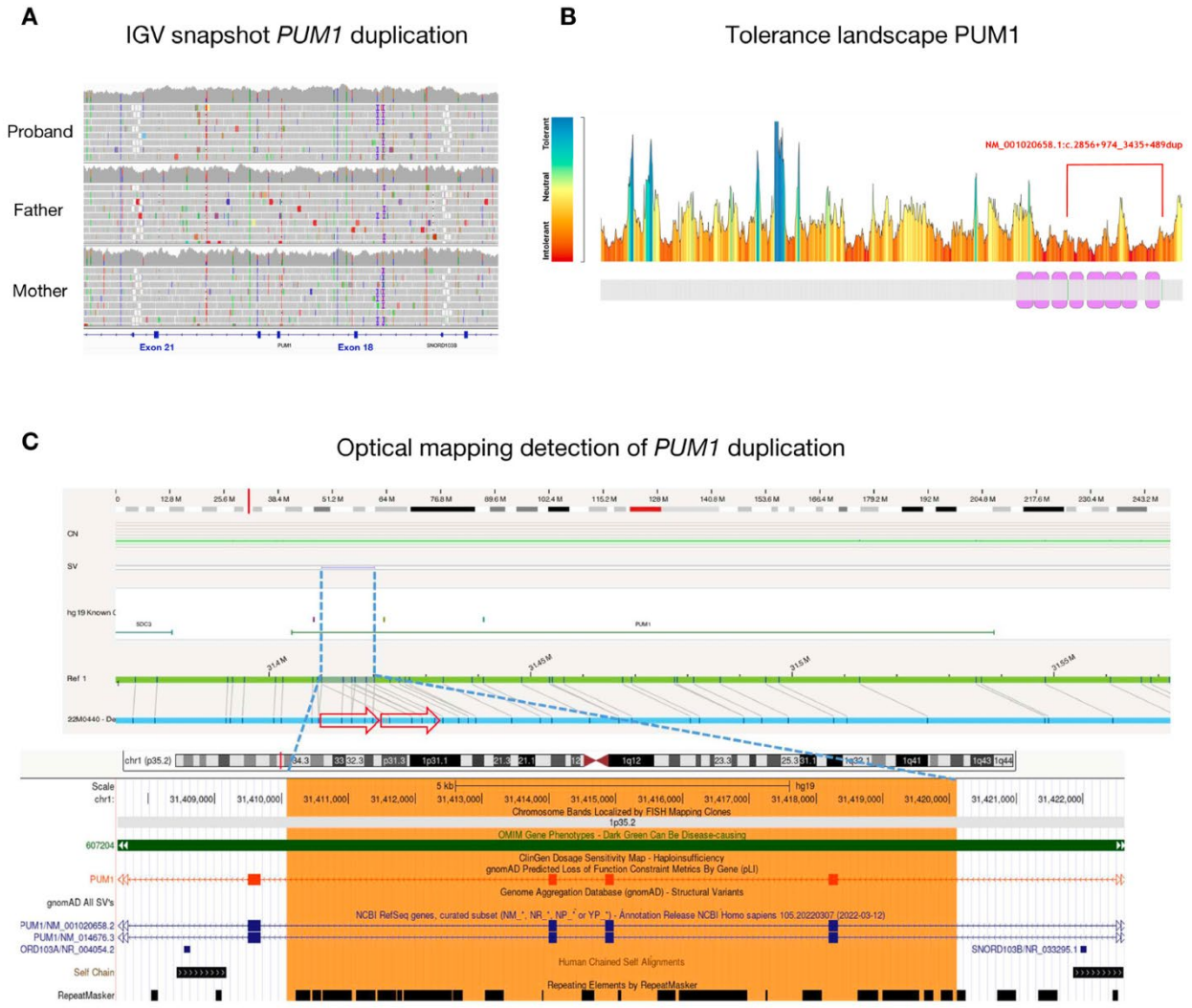
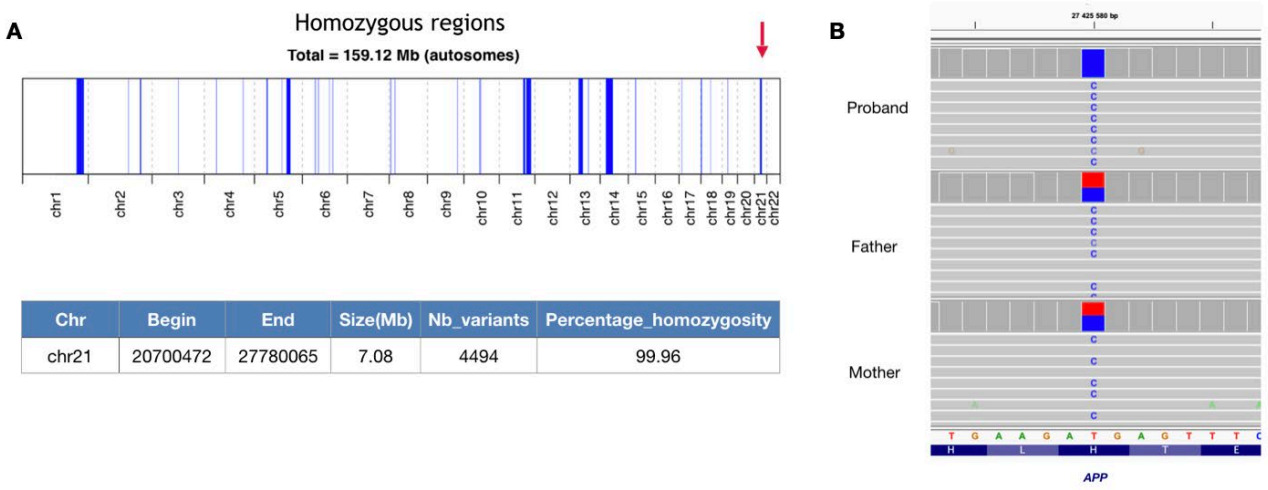


Figure 5: *De novo* heterozygous duplication in *PUM1*; NM_001020658.1:c.2856+974_3435+489dup. A) IGV snapshot showing a 13 kb duplication involving exons 18-21. B) Tolerance landscape of PUM1 (1224 aa) generated by Metadome. The exons 18-21 codes for a region predicted to be highly intolerant to mutation. This region contains functional domains involved in RNA binding. C) Optical map of *PUM1* locus showing an intragenic tandem duplication between two L2c retrotransposons.

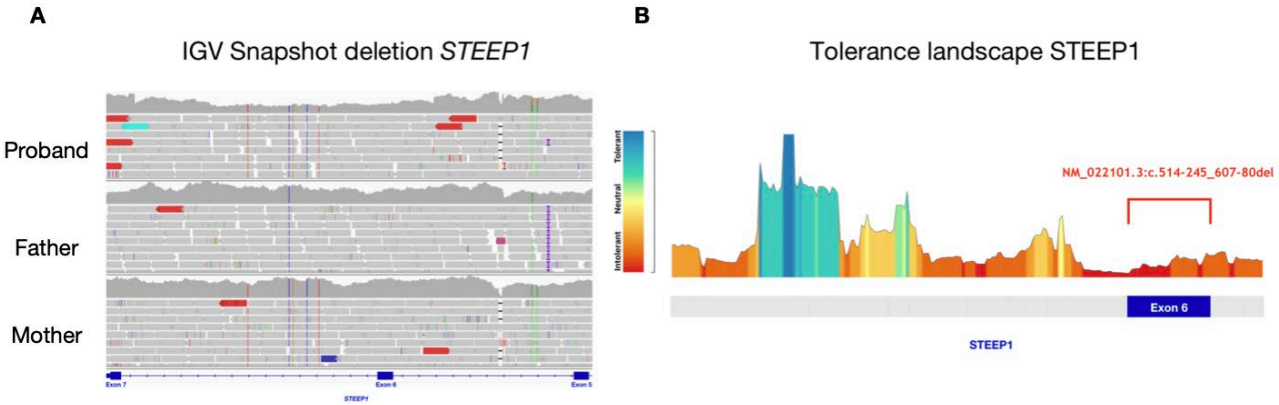
Supplemental material

Supplemental figures :

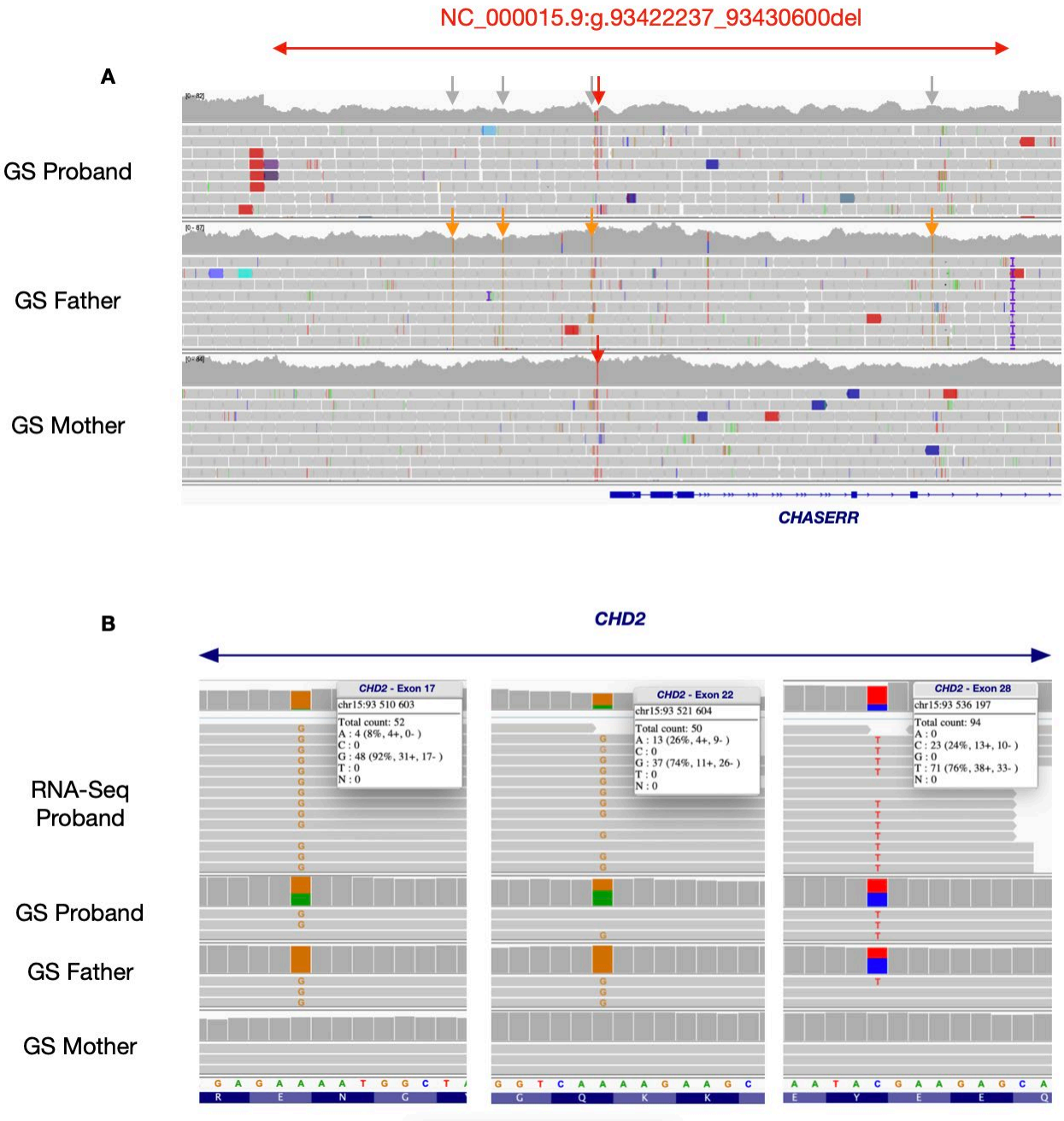


Chr	Begin	End	Size(Mb)	Nb_variants	Percentage_homozygosity
chr21	20700472	27780065	7.08	4494	99.96

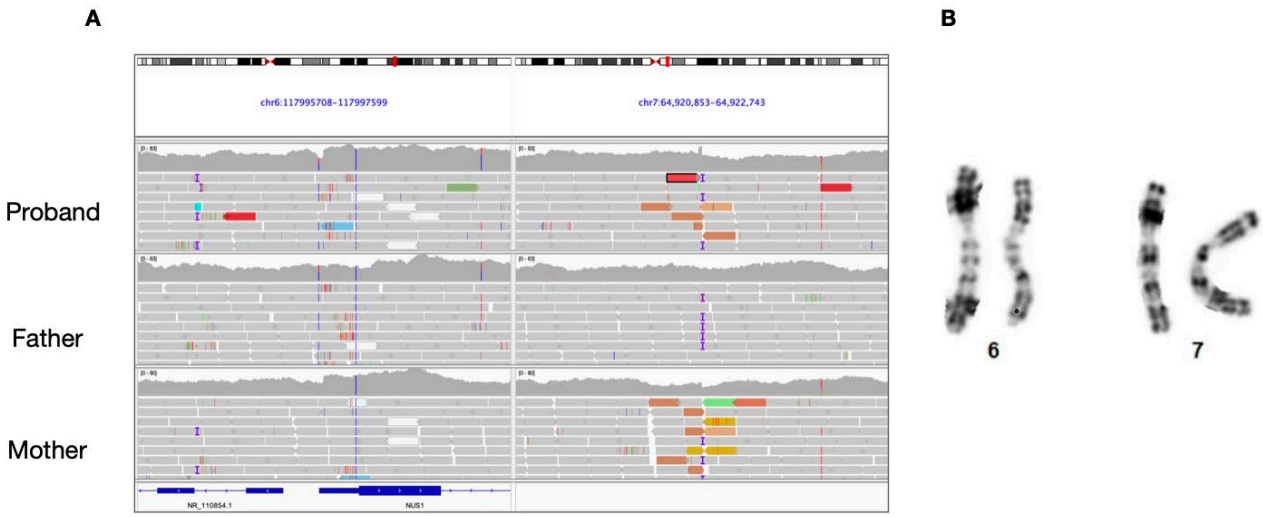
Sup Figure 1: P19. A) Automap estimation shows 159.12 Mb of homozygous regions, confirming the consanguinity of the parents. High homozygosity of the *APP* locus (7.08 Mb ; 99.96%). No other homozygous candidate variants were identified in the LOH region. B) IGV snapshot showing homozygous variant NM_000484.3:c.[440A>G] in *APP*.



Sup Figure 2: P14. *De novo* heterozygous deletion in *STEEP1*; NM_022101.3:c.514-245_607-80del A) IGV snapshot showing a 1.7kb deletion leading to the lost of the exon 6 (in-frame). B) Tolerance landscape of STEEP1 generated by Metadome. The exon 6 codes for a region predicted to be highly intolerant to mutation.



Sup Figure 3: P32. A) IGV snapshot of trio-GS at the *CHASERR* locus. Four homozygous SNP (orange arrow) present in the father were absent from the deleted locus of the proband. One homozygous SNP (red arrow) in the mother was found at the deleted locus of the proband. This indicates that the deleted allele was paternal. B) IGV snapshot showing allelic imbalance skewed towards the CHD2 allele in cis- with *CHASERR* deletion.



Sup Figure 4: P23. A) IGV snapshot showing a balanced event in the NUS1 promoter. This event was also observed in the mother. As NUS1 one is already associated with NDD we suspected a mosaicism in the mother. B) Karyotyping of the proband appears normal, invalidating the t(6;7) hypothesis.

Sup Table 1 :

Case/Sex	Type	Gene	Variant	Transmission	Interpretation	ACMG	Reason not reported by ES	Technique applied
P19/M	SNV indel	*APP*	NC_000021.8:g.27425580T>C NM_000484.4:c.440A>G:p.(His147Arg)	Autosomal recessive [HOM]	Strong candidate VUS	NA	More support for pathogenicity	GS
P23/F	SV	*CBX3*	NC_000007.13:g.26244486_26248316del NM_007276.4:c.25-1502_330+141del	*De novo* [HET]	Strong candidate VUS	N/A	Not detected	GS RNA-seq
P31/M	SNV indel	*SYT7*	NC_000011.9:g.61290738C>G NM_001252065.1:c.1141G>C:p.(Val381Leu)	*De novo* [HET]	VUS	N/A	More support for pathogenicity	GS
P3/F	SV	-	NC_000001.10:g.pter_69716317delins[NC_000006.11:g.108848601_qterinv] NC_000006.11:g.108848601_qterdelins[NC_000001.10:g.69716317_pterinv]	*De novo* [HET]	VUS	N/A	Copy neutral translocation not detected by ES	GS
P8/F	SV	-	NC_000005.9:g.105706517_170958911inv	*De novo* [HET]	VUS	N/A	Copy neutral inversion not detected by ES	GS
P25/M	SV	*FGF18*	Complex structural variant duplication and inverted insertion	*De novo* [HET]	VUS	N/A	Complex event not detected by ES	GS RNA-seq OGM

Sup Table 1: Variants of uncertain significance identified in this study. ACMG classe and evidence categories are determined following ACMG/AMP Standards and Guidelines. HOM: Homozygous; HEM: Hemizygous; HET: Heterozygous; COMP HET: Compound heterozygous; MOS: Mosaicism.

Sup Table 2 :

Case/Sex/Age range	Phenotype	ES date	ES kit
P1/F/20-40 y	Intellectual disability severe (IQ 20), global developmental delay, hypotonia, absent speech, autism, epilepsy (since the age of 7 year), cerebral MRI : small periventricular heterotopia and dilated, cortical veins, extensive naevus, height –1.7 SD, weight +0.5 SD	TRIO/2016	1
P2/F/10-20 y	Mild pulmonary valve stenosis, left renal cyst, pyelonephritis, stridor, severe laryngomalacia, small external auditory canals, gastrostomy until the age of 5 years, sitting without support at 1 year, walking acquired at 2.5 years, MRI at 2 months: delayed myelination of brainstem and internal capsules, bruxism, stereotypies, height -1DS , weight -3DS , head circumference -4DS	TRIO/2016	1
P3/F/2-10 y	Axial hypotonia, severe global developmental delay, strabismus, head circumference -2,2 SD	TRIO/2016	1
P4/M/2-10 y	Moderate global developmental delay, severe language delay with first words at 3 years old, behavior troubles, stereotypy	TRIO/2016	1
P5/M/10-20 y	Postnatal growth delay (- 2.5 SD), microcephaly (- 3 SD), global developmental delay, facial dysmorphism, cryptorchidism, inguinal hernia, laryngomalacia, Turricephaly, unilateral pachygyria, patent sagittal suture	TRIO/2016	1
P6/F/20-40 y	Epileptic encephalopathy, facial dysmorphism, ventricular dilatation, myelination delay	TRIO/2016	1
P7/F/10-20 y	Hypotonia, severe global developmental delay, behavior troubles, trunk hyperpigmentation, short stature -3 SD, bilateral hearing impairement, short corpus callosum	TRIO/2016	1
P8/F/10-20 y	Axial hypotonia, peripheral hypertonia, severe global developmental delay, scoliosis, dysmorphic ventricles, thin corpus callosum	TRIO/2016	1
P9/F/10-20 y	Axial hypotonia, facial dysmorphism, severe global developmental delay predominantly on language, nystagmus	TRIO/2016	1
P10/F/10-20 y	Craniosynostosis, mild global developmental delay, bilateral microtia, strabismus, choanal stenosis, short stature (-2 SD)	TRIO/2016	1
P11/M/10-20 y	Epileptic encephalopathy, intellectual disabibility, behavior troubles	TRIO/2016	1
P12/F/10-20 y	Hypotonia, global developmenatal delay, seizures, no speech, facial dysmorphism	TRIO/2016	1
P13/M/10-20 y	Myoclonic epilepsy, encephalopathy, axial hypotonia, languange acquisition then regression, cortical atrophy on cerebral MRI, myelinisation delay, thin corpus callosum	TRIO/2016	1
P14/F/10-20 y	Severe global developmental delay, sat at 15 months, walked at 3.5 years old, no language, axial hypotonia	TRIO/2016	1
P15/F/20-40 y	Severe global developmental delay, walked at 4 years old, short stature, cerebellar atiophy	TRIO/2016	1
P16/M/10-20 y	Severe global developmental delay, macrocephaly +2 SD	TRIO/2016	1
P17/F/10-20 y	Global developmental delay, behavior troubles, clinodactyly, enlarged cerebral ventricles	TRIO/2016	1

Sup Table 2. Exome sequencing kit : 1:SureSelectXT Clinical Research Exome system (Agilent Technologies); 2:SureSelectXT Clinical Research Exome V2 system (Agilent Technologies); 3:Human Comprehensive Exome (Twist Bioscience).

Case/Sex/Age	Phenotype	ES date	ES kit
P18/F/10-20 y	Epilepsy, short stature, microcephaly, facial dysmorphism, cerebral MRI : cortical atrophy	TRIO/2016	1
P19/M/10-20 y	Severe intellectual disability, no language, abnormal behavior, autism, epilepsy	TRIO/2016	1
P20/M/10-20 y	Severe global developmental delay, hypotonia, severe intellectual disability, no speech	TRIO/2016	1
P21/M/10-20 y	Severe global developmental delay, Plagiocephaly, Hypertrichosis, Hypertelorism, Global developmental delay, Epicanthus, Corneal astigmatism, Abnormality of earlobe, Abnormal facial shape	SOLO/2018	2
P22/F/2-10 y	Short stature, Seizures, Moderate global developmental delay, Global developmental delay, Epicanthus, Delayed speech and language development, Abnormal facial shape	SOLO/2020	3
P23/F/2-10 y	Relative microcephaly -1.5SD, microcephaly relative, Talipes valgus, Strabismus, Moderate global developmental delay, Infantile muscular hypotonia, Hypertrichosis, Epicanthus, Delayed speech and language development, Decreased fetal movement	TRIO/2020	3
P24/F/2-10 y	Wide mouth, Unilateral renal hypoplasia, Skin plaque, Protruding tongue, Periventricular leukomalacia, Oligohydramnios, Macrocephaly, Intellectual disability, severe, Gastroesophageal reflux, Failure to thrive, Esodeviation, Chronic constipation	TRIO/2018	1
P25/M/20-40 y	Nystagmus, Nevus, Neonatal hypotonia, Macrocephaly, Intellectual disability, severe, Global developmental delay, Cerebellar hypoplasia, Autism, Absent speech	SOLO/2020	3
P26/F/2-10 y	Severe global developmental delay, Scoliosis, Ptosis, Pachygyria, Neonatal hypotonia, Laryngomalacia, Global developmental delay, Focal clonic seizures	SOLO/2020	3
P27/M/2-10 y	Synophrys, Status epilepticus, Seizures, Recurrent singultus, Moderate global developmental delay, Infantile spasms, Hypoplasia of the corpus callosum, Fused teeth, Delayed speech and language development, Abnormal behavior	SOLO/2018	2
P28/M/10-20 y	Specific learning disability, Spasticity, Postural instability, Loss of speech, Loss of ability to walk, Global developmental delay, Dysphagia, Delayed gross motor development, Delayed ability to walk, Congenital sensorineural hearing impairment, Cerebellar ataxia associated with quadrupedal gait, Autistic behavior, Ataxia, Abnormal pyramidal sign	SOLO/2017	2
P29/F/2-10 y	duplication pituitary, Wide anterior fontanel, Vertebral segmentation defect, Umbilical hernia, Tufted angioma, Tuberous angioma, Thin vermilion border, Sleep apnea, Posterior pharyngeal cleft, Polyhydramnios, moderate intellectual disability, Hypertelorism, Enlarged cisterna magna, Cerebellar dysplasia, Agenesis of cerebellar vermis, Abnormal facial shape	TRIO/2018	1
P30/M/10-20 y	Severe global developmental delay, Seizures, Periventricular white matter hyperdensities, Long fingers, Severe intellectual disability, Global developmental delay, Epileptic encephalopathy, Dystonia, Decreased thalamic volume, Decreased serum ceruloplasmin, Cerebral cortical atrophy	TRIO/2018	1
P31/M/10-20 y	Strabismus, Sleep disturbance, Short stature, Short neck, Retinal detachment, Premature birth, Preaxial polydactyly, Micropenis, Microcephaly, Low-set ears, Involuntary movements, Hypotelorism, Hypermetropia, Duplication of thumb phalanx, Duodenal atresia, Decreased body weight, Clinodactyly of the 5th finger, Arrhythmia	TRIO/2017	1
P32/M/2-10 y	Short 1st metacarpal, Scoliosis, Ptosis, Prominent nose, Optic atrophy, Moderate global developmental delay, Long philtrum, Infantile muscular hypotonia, Global developmental delay, Feeding difficulties, Dystonia, Camptodactyly of toe, Anteverted nares, Adducted thumb, Abnormal facial shape	SOLO/2019	3
P33/F/2-10 y	Short stature, Severe global developmental delay, Retinopathy, Infantile muscular hypotonia, Global developmental delay, Feeding difficulties, Failure to thrive, Abnormal facial shape	TRIO/2020	3

Sup Table 2. Exome sequencing kit : 1:SureSelectXT Clinical Research Exome system (Agilent Technologies); 2:SureSelectXT Clinical Research Exome V2 system (Agilent Technologies); 3:Human Comprehensive Exome (Twist Bioscience).

Sup Table 3: DROP pipeline calls:

OUTRIDER calls

hgncSymbol	geneID	sampleID	pValue	padjust	zScore	l2fc	rawcounts	normcounts	meanCorrected	theta	aberrant	AberrantBySample	AberrantByGene	padj_rank	foldChange
CEP164	ENSG00000110274.16_7	P16	7.91706977897483e-09	0.00112830715723481	-6.12	-0.67	1039	1003.45	1598.74	206.2	TRUE	1	1	1	0.63
ADAMTSL1	ENSG00000178031.17_5	P22	2.693399238081e-08	0.00383851819228388	-6.46	-2.38	651	2074.02	10784.15	19.18	TRUE	1	2	1	0.19
PCGF6	ENSG00000156374.16_7	P23	7.19813936306138e-14	1.02584824800767e-08	-7.83	-0.85	418	271.8	489.56	290.76	TRUE	4	1	1	0.55
AL355802.1	ENSG00000219470.1_5	P23	1.51973280652435e-10	1.08292932268939e-05	-7.93	-3.88	13	2.44	38.56	14.64	TRUE	4	2	2	0.07
SLC39A6	ENSG00000141424.13_7	P23	2.72051890679604e-08	0.0012923893307439	-5.94	-0.77	3576	3669.39	6237.4	135.08	TRUE	4	1	3	0.59
SURF6	ENSG00000148296.7_5	P23	7.59799534826747e-08	0.00270708506158862	-5.68	-0.9	1552	395.06	735.48	96.22	TRUE	4	1	4	0.54
PAIP1	ENSG00000172239.14_9	P24	2.3756323223411e-16	3.38565028108532e-11	-8.71	-0.73	2337	2165.65	3580.37	347.75	TRUE	13	1	1	0.6
MRPL13	ENSG00000172172.8_6	P24	2.04664622206452e-15	1.45839663230934e-10	-8.34	-0.69	1211	881.39	1418.08	406.85	TRUE	13	1	2	0.62
PTMA	ENSG00000187514.16_5	P24	2.34183707520426e-13	1.11249557677174e-08	-7.85	-0.9	24920	8638.92	16134.07	169.44	TRUE	13	1	3	0.54
SAP18	ENSG00000150459.12_6	P24	7.47108499340372e-12	2.66186824977642e-07	-6.91	-0.37	4734	3168.89	4101.33	891.14	TRUE	13	1	4	0.77
BTBD1	ENSG00000064726.10_7	P24	4.29234874458904e-10	1.2234546227518e-05	-6.45	-0.44	3214	2955.35	4015.07	521.56	TRUE	13	1	5	0.74
SNHG7	ENSG00000233016.7_7	P24	6.23995635611959e-10	1.48215345954975e-05	-6.86	-1.67	290	234.66	748.71	44.25	TRUE	13	3	6	0.31
RAD51AP1	ENSG00000111247.14_3	P24	7.93506082024162e-07	0.0161553023431344	3.15	0.45	922	476.65	348.89	323.03	TRUE	13	1	7	1.37
NOL7	ENSG00000225921.7_4	P24	1.23842316079013e-06	0.0220618510859235	-4.68	-0.3	2807	1495.67	1846.92	691.48	TRUE	13	1	8	0.81
HNRNPA3	ENSG00000170144.20_6	P24	2.01649871083776e-06	0.0319314256903383	-4.86	-0.34	14157	9521.67	12033.88	457.98	TRUE	13	1	9	0.79
MPPE1	ENSG00000154889.17_8	P24	2.38569897354456e-06	0.0339999684480061	-5.08	-0.73	660	576.68	955.75	117.52	TRUE	13	1	10	0.6
CLN5	ENSG00000102805.16_7	P24	3.15565370648955e-06	0.0408845784223026	-4.9	-0.62	1014	1026.79	1579.98	151.32	TRUE	13	1	11	0.65
AC026271.1	ENSG00000174977.8_5	P24	3.69323516717925e-06	0.0438620157543023	-3.4	-0.75	363	70.47	117.92	118.15	TRUE	13	1	12	0.59
GNAQ	ENSG00000156052.11_7	P24	4.37610158212878e-06	0.0479741084648332	-4.81	-0.47	2884	2794.9	3864.16	240.77	TRUE	13	1	13	0.72
ZWINT	ENSG00000122952.17_7	P25	3.52101712240169e-12	5.01800404804134e-07	-7.57	-1.02	209	530.46	1079.33	191.25	TRUE	2	1	1	0.49
GLA	ENSG00000102393.12_5	P25	1.16395556966866e-10	8.29410019505052e-06	5.71	0.89	2621	1443.13	780.68	92.67	TRUE	2	1	2	1.85
PAWR	ENSG00000177425.11_7	P26	1.98513234596548e-08	0.00282912630119728	-6.42	-1.46	286	770.33	2118.03	46.7	TRUE	1	1	1	0.36
LINC01578	ENSG00000272888.7_7	P32	2.95917169321535e-13	4.21728580668628e-08	-8.02	-0.97	631	763.15	1500.14	171.14	TRUE	2	1	1	0.51
LIMD1	ENSG00000144791.10_5	P32	5.96167815423907e-10	4.24816524190683e-05	5.53	0.86	7904	2574.43	1422.53	89.17	TRUE	2	1	2	1.82
PSAP	ENSG00000197746.14_6	P33	1.90417407966811e-10	2.71374801876351e-05	-6.88	-0.89	26055	25526.16	47142.42	132.47	TRUE	14	1	1	0.54
SIAE	ENSG00000110013.13_5	P33	3.87722427435622e-09	0.000276282767557415	-6.68	-1.24	384	757.1	1788.13	68.87	TRUE	14	1	2	0.42
SRSF3	ENSG00000112081.17_4	P33	1.22411463731532e-08	0.000581518728990168	-6	-0.53	4770	6494.96	9381.54	284.52	TRUE	14	1	3	0.69
OS9	ENSG00000135506.16_7	P33	9.24079886421774e-08	0.00303969293380115	-5.63	-0.53	9416	9552.66	13753.14	248.28	TRUE	14	1	4.5	0.69
DDOST	ENSG00000244038.10_7	P33	1.06644103554827e-07	0.00303969293380115	-5.54	-0.44	11196	9977.23	13487.14	352.99	TRUE	14	1	4.5	0.74
KPNA6	ENSG00000025800.14_5	P33	4.03359166635939e-07	0.0092147195929882	-5.44	-0.35	1982	3652.53	4653.46	599.04	TRUE	14	1	6.5	0.78
HEXA	ENSG00000213614.10_8	P33	4.52602859784344e-07	0.0092147195929882	-5.39	-0.7	1821	4213.41	6834.76	135.41	TRUE	14	2	6.5	0.62
NPIPA1	ENSG00000183426.17_6	P33	6.95495162257014e-07	0.0123898770521256	-5.44	-0.79	958	2919.55	5053.73	106.47	TRUE	14	1	8	0.58
EVI5	ENSG00000067208.14_5	P33	1.45747241951895e-06	0.0230791976257957	-5.54	-0.76	298	636.39	1078.43	132.05	TRUE	14	1	9	0.59
ZC3H11A	ENSG00000058673.16_7	P33	1.83547322700989e-06	0.0230869143383949	-5.11	-0.45	2329	6438.39	8763.01	294.34	TRUE	14	1	11	0.73
CDCA2	ENSG00000184661.14_4	P33	1.94394631596136e-06	0.0230869143383949	4.64	0.75	245	999.94	592.41	115.35	TRUE	14	1	11	1.68
ZBED6	ENSG00000257315.2_5	P33	1.87895281858377e-06	0.0230869143383949	-5.11	-0.45	2328	6427.63	8753.07	292.65	TRUE	14	1	11	0.73
PRCP	ENSG00000137509.11_8	P33	2.37838874908613e-06	0.0260736817184878	-5.03	-0.63	3184	3352.01	5184.04	140.88	TRUE	14	1	13	0.65
QTRT2	ENSG00000151576.10_5	P33	2.63144340318087e-06	0.0267872954853666	-5.56	-0.55	266	576.13	846.95	341.09	TRUE	14	1	14	0.68

FRASER calls 1/4

sampleID	seqnames	start	end	width	strand	hgncSymbol	addHgncSymbols	type	pValue	padjust	zScore	psiValue	deltaPsi	meanCounts	meanTotalCounts	counts	totalCounts	pValueGene	padjustGene	STRAND_SPECIFIC	PAIRED_END
P16	chr17	16285552	16285553	2	+	UBB		theta	2.1935e-07	0.1036	-2.79	0.77	-0.23	7657.4	7831.8	5750	7494	1.3161e-06	0.0814028970125354	reverse	TRUE
P16	chr2	189864035	189864036	2	+	COL3A1		theta	9.0462e-07	0.1036	-2.73	0.88	-0.12	7119.7	7175.6	3986	4545	4.5231e-06	0.0814028970125354	reverse	TRUE
P16	chr6	29913059	29913227	169	+	HLA-A		psi5	1.7356e-06	0.35532	-2.44	0.8	-0.19	4260.6	4304.4	1608	2002	2.77696e-05	0.121896308925597	reverse	TRUE
P16	chr6	29913059	29977311	64253	+	HLA-J	HCG9;HLA-A;HLA-W;DDX39BP2	psi5	1.7356e-06	0.35532	2.71	0.2	0.2	39.3	4304.4	393	2002	2.6034e-05	0.121896308925597	reverse	TRUE
P16	chr22	33256107	33256108	2	+	TIMP3		theta	2.6081e-06	0.16183	-2.73	0.86	-0.14	2057.3	2094.1	2192	2560	1.04324e-05	0.104825209300007	reverse	TRUE
P16	chr2	69476241	69476242	2	+	ANTXR1		theta	2.7833e-06	0.16183	-2.77	0.19	-0.8	304.1	339	81	430	1.39165e-05	0.104825209300007	reverse	TRUE
P16	chr12	125397202	125398113	912	-	UBC		psi5	3.3066e-06	0.35532	2.63	0.19	0.19	81.5	5046.9	815	4213	0.0002347686	0.456865118361179	reverse	TRUE
P16	chr15	40328574	40328575	2	-	SRP14		theta	4.4688e-06	0.18777	-2.76	0.9	-0.1	2136	2157.1	1939	2149	2.2344e-05	0.121896308925597	reverse	TRUE
P16	chr7	135357554	135357555	2	+	STMP1		theta	5.7539e-06	0.18777	2.76	0.27	0.27	11.1	626.5	110	411	2.87695e-05	0.121896308925597	reverse	TRUE
P16	chr17	55917117	55917118	2	-	MRPS23		theta	7.0723e-06	0.19333	-2.78	0.62	-0.38	352.6	360.3	123	200	2.82892e-05	0.121896308925597	reverse	TRUE
P16	chr1	154574511	154574512	2	-	ADAR		theta	8.1803e-06	0.2117	-2.77	0.47	-0.52	255.1	268.2	118	249	4.09015e-05	0.158618133685881	reverse	TRUE
P16	chr19	1036534	1036535	2	+	CNN2		theta	1.0581e-05	0.23961	-2.77	0.88	-0.12	1033.2	1047.7	1026	1171	9.5229e-05	0.273019937287605	reverse	TRUE
P16	chr17	78183516	78183517	2	-	SGSH		theta	1.1448e-05	0.25395	-2.78	0.4	-0.59	181.4	195.3	91	230	5.724e-05	0.194020604401221	reverse	TRUE
P16	chr6	29857163	29857164	2	+	HLA-H		theta	2.5976e-05	0.43437	-2.16	0.8	-0.2	2342.3	2372.2	687	863	0.000467568	0.7043856352098	reverse	TRUE
P16	chr1	225707024	225707025	2	-	ENAH		theta	2.792e-05	0.45294	-2.68	0.85	-0.15	635.6	649.4	780	917	0.00030712	0.533853034669166	reverse	TRUE
P16	chr14	105235989	105235990	2	-	AKT1		theta	2.9492e-05	0.46872	-2.77	0.89	-0.11	678.8	684.9	482	543	0.000176952	0.383965216894817	reverse	TRUE
P16	chr1	170633237	170633238	2	+	PRRX1		theta	4.3561e-05	0.57046	-2.67	0.82	-0.18	357	361.8	212	260	0.000566293	0.789301850818804	reverse	TRUE
P16	chr2	86737600	86756340	18741	-	CHMP3	RNU6-640P;RNF103-CHMP3	psi5	4.4716e-05	1	2.72	0.15	0.14	7.4	626.2	74	510	0.000178864	0.383965216894817	reverse	TRUE
P16	chr19	51227674	51227675	2	+	CLEC11A		theta	4.5311e-05	0.58632	-2.67	0.88	-0.11	779.4	789.7	750	848	0.000181244	0.383965216894817	reverse	TRUE
P16	chr12	6639125	6639856	732	+	NCAPD2		psi5	4.9871e-05	1	-2.71	0.06	-0.89	127.7	241.2	71	1205	0.000149613	0.34974342377262	reverse	TRUE
P16	chr19	54379026	54379027	2	+	MYADM		theta	6.0653e-05	0.74916	-2.76	0.88	-0.12	414.9	419	289	330	0.000242612	0.456865118361179	reverse	TRUE
P16	chr5	141381205	141381206	2	-	GNPDA1		theta	7.5173e-05	0.88814	-2.77	0.78	-0.21	222.2	228.6	232	296	0.000375865	0.60668159523198	reverse	TRUE
P16	chr10	32307085	32307243	159	-	KIF5B		psi3	9.1017e-05	1	-2.6	0.87	-0.13	664.6	672.2	459	529	0.000364068	0.601972809421552	reverse	TRUE
P16	chr7	75601780	75608768	6989	+	POR		psi5	0.0001428	1	-2.72	0.77	-0.23	230.8	233.7	96	125	0.0005712	0.789301850818804	reverse	TRUE
P16	chr9	5522602	5534744	12143	+	PDCD1LG2		psi5	0.00016276	1	-2.64	0.32	-0.61	59.5	61.9	7	22	0.00048828	0.719596973209888	reverse	TRUE
P16	chr15	42824531	42824532	2	+	SNAP23		theta	0.00019405	1	-2.76	0.85	-0.14	213.5	217	202	237	0.00058215	0.789301850818804	reverse	TRUE
P23	chr15	72500999	72501000	2	-	PKM		theta	7.6951e-07	0.30936	-2.46	0.87	-0.13	18295.1	18724.1	27706	31996	8.46461e-06	0.306615540331036	reverse	TRUE
P23	chr7	100781247	100781248	2	+	SERPINE1		theta	4.3135e-06	0.47991	-2.37	0.89	-0.11	4926.1	4999.5	6008	6742	3.4508e-05	0.313864586210536	reverse	TRUE
P23	chr11	69467908	69467909	2	+	CCND1		theta	8.238e-06	0.47991	-2.43	0.84	-0.16	1909.3	1975.6	3534	4197	9.0618e-05	0.392409792910298	reverse	TRUE
P23	chr2	242194986	242194987	2	-	HDLBP		theta	8.3797e-06	0.47991	-2.45	0.82	-0.17	2163	2207.2	2029	2471	4.18985e-05	0.326643644505013	reverse	TRUE
P23	chr7	26242643	26251281	8639	+	CBX3		psi3	1.0544e-05	1	2.32	0.48	0.47	84.7	1348.8	847	1748	5.272e-05	0.331014901029088	reverse	TRUE
P23	chr7	135613199	135613200	2	-	MTPN		theta	1.4302e-05	0.6327	-2.45	0.62	-0.37	584.4	614.7	500	803	5.7208e-05	0.331014901029088	reverse	TRUE
P23	chr12	63359268	63359269	2	+	RPL14P1		theta	1.5164e-05	0.6327	-2.45	0.89	-0.11	2030.2	2058	2175	2453	6.0656e-05	0.331014901029088	reverse	TRUE
P23	chr1	1735926	1735927	2	-	GNB1		theta	1.5649e-05	0.6327	-2.45	0.89	-0.11	1861.5	1890	2325	2610	7.8245e-05	0.388184050307861	reverse	TRUE
P23	chr9	90342645	90342941	297	+	CTSL		psi5	2.9257e-05	1	-2.25	0.86	-0.13	2007.8	2036.1	1647	1916	0.000175542	0.532209122450543	reverse	TRUE
P23	chr2	110969869	110969870	2	-	MTLN		theta	3.107e-05	0.91274	-2.46	0.49	-0.49	245.9	265.9	193	393	0.00012428	0.477645726928432	reverse	TRUE
P23	chr17	78120803	78120804	2	-	EIF4A3		theta	3.6483e-05	0.94415	-2.43	0.4	-0.56	221.6	244.7	157	388	0.000328347	0.7790745079034 3	reverse	TRUE
P23	chr12	6688051	6688052	2	-	AC006064.6	CHD4	theta	4.6739e-05	0.99612	-2.45	0.78	-0.21	512.9	529.5	589	755	9.3478e-05	0.392409792910298	reverse	TRUE
P23	chr11	66082212	66082213	2	-	CD248		theta	6.2574e-05	1	-2.3	0.87	-0.12	1751.2	1790.4	1197	1371	0.00031287	0.7790745079034 3	reverse	TRUE
P23	chr8	71495514	71495515	2	-	TRAM1		theta	6.3474e-05	1	-2.45	0.89	-0.11	856.9	865.3	653	737	0.00031737	0.7790745079034 3	reverse	TRUE
P24	chr17	48266372	48272592	6221	-	COL1A1		psi5	5.9411e-07	0.23598	2.82	0.13	0.13	258.7	14127.9	2587	19913	3.267605e-05	0.192812324782637	reverse	TRUE
P24	chr11	62292738	62292739	2	-	AHNAK		theta	9.6299e-07	0.18043	-2.81	0.83	-0.17	3037.3	3075.4	1885	2266	0.00022630265	0.662585646488052	reverse	TRUE
P24	chr11	75283172	75283173	2	+	SERPINH1		theta	2.1846e-06	0.18043	-2.8	0.83	-0.17	2266.3	2322.8	2725	3290	1.52922e-05	0.14991259227975	reverse	TRUE
P24	chr22	33256816	33256817	2	+	TIMP3		theta	2.5247e-06	0.18043	-2.79	0.88	-0.12	2641.6	2689	3538	4012	1.00988e-05	0.143016535145336	reverse	TRUE

FRASER calls 2/4

sampleID	seqnames	start	end	width	strand	hgncSymbol	addHgncSymbols	type	pValue	padjust	zScore	psiValue	deltaPsi	meanCounts	meanTotalCounts	counts	totalCounts	pValueGene	padjustGene	STRAND_SPECIFIC	PAIRED_END
P24	chr17	74684246	74684247	2	-	MXRA7		theta	4.117e-06	0.2092	-2.82	0.81	-0.19	1085.2	1105.8	868	1074	1.6468e-05	0.14991259227975	reverse	TRUE
P24	chr15	48784765	48784766	2	-	FBN1		theta	4.2343e-06	0.2092	-2.79	0.81	-0.19	1241.6	1257.1	669	823	1.69372e-05	0.14991259227975	reverse	TRUE
P24	chr14	70352711	70352712	2	-	RPL7AP6		theta	4.7985e-06	0.21043	-2.82	0.86	-0.14	1275	1294.8	1190	1388	1.9194e-05	0.151011306930833	reverse	TRUE
P24	chr19	11275414	11275415	2	-	KANK2		theta	8.3149e-06	0.29352	-2.58	0.12	-0.86	245.1	295.3	34	272	7.48341e-05	0.374507097072762	reverse	TRUE
P24	chr2	47131356	47131357	2	-	MCFD2		theta	9.1837e-06	0.30249	-2.66	0.62	-0.38	465.4	494.5	185	300	5.51022e-05	0.300131840602742	reverse	TRUE
P24	chr12	113828235	113828499	265	+	PLBD2		psi3	2.0113e-05	1	2.81	0.93	0.88	2.8	76.7	28	30	0.00020113	0.61920649448194	reverse	TRUE
P24	chr10	120802132	120802133	2	-	EIF3A		theta	2.3015e-05	0.51052	-2.81	0.9	-0.1	780.6	787.7	637	708	0.000161105	0.543220573462508	reverse	TRUE
P24	chr2	85628759	85628760	2	-	CAPG		theta	2.362e-05	0.51348	-2.73	0.9	-0.1	898.7	906.9	742	824	0.00014172	0.519098417213162	reverse	TRUE
P24	chr6	24977085	24977086	2	+	PPIAP29		theta	2.6445e-05	0.55729	-2.81	0.64	-0.35	193.7	203.6	177	276	7.9335e-05	0.374507097072762	reverse	TRUE
P24	chrX	102841846	102841847	2	+	TCEAL4		theta	2.7393e-05	0.55729	-2.81	0.77	-0.22	318.7	326.2	253	328	0.000136965	0.519098417213162	reverse	TRUE
P24	chr1	170633239	170633240	2	+	PRRX1		theta	3.0415e-05	0.56998	-2.47	0.69	-0.3	340.2	354.5	308	444	0.00030415	0.717881978790448	reverse	TRUE
P24	chr15	90350041	90357195	7155	-	ANPEP		psi3	3.5013e-05	1	2.77	0.14	0.14	11.9	1055.7	119	829	0.000455169	0.856443667672075	reverse	TRUE
P24	chr6	75812371	75812372	2	-	COL12A1		theta	3.6655e-05	0.66404	-2.68	0.9	-0.1	714.5	719.7	459	511	0.00014662	0.519098417213162	reverse	TRUE
P24	chr3	196197186	196197187	2	-	RNF168		theta	4.7389e-05	0.80342	-2.8	0.42	-0.56	79.5	84.9	39	93	0.000236945	0.662585646488052	reverse	TRUE
P24	chr12	120998605	120998606	2	+	RNF10		theta	5.053e-05	0.83218	-2.82	0.84	-0.15	345.3	350.7	293	347	0.00025265	0.662585646488052	reverse	TRUE
P24	chr11	61165448	61165732	285	+	TMEM216		psi5	5.5312e-05	1	-2.82	0	-0.92	33	38.8	0	57	0.000442496	0.856443667672075	reverse	TRUE
P24	chr16	2964291	2979613	15323	+	FLYWCH1		psi5	5.8964e-05	1	-2.72	0.33	-0.55	108.7	128.5	48	144	0.000471712	0.856443667672075	reverse	TRUE
P24	chr16	129505	129506	2	+	MPG		theta	7.2714e-05	1	-2.81	0.87	-0.12	368.6	373.7	351	402	0.000290856	0.717881978790448	reverse	TRUE
P24	chr11	33720046	33720047	2	-	AL049629.2	C11orf91	theta	0.00012507	1	-2.81	0.67	-0.32	107.7	110.2	51	76	0.00025014	0.662585646488052	reverse	TRUE
P24	chr7	6523718	6523719	2	-	DAGLB	KDELR2	theta	0.00014171	1	-2.77	0.89	-0.1	365.7	369.4	315	352	0.00042513	0.856443667672075	reverse	TRUE
P24	chr8	100904280	100904281	2	-	COX6C		theta	0.00016769	1	2.56	0.18	0.17	62	1114.1	160	874	0.00033538	0.766058528145834	reverse	TRUE
P25	chr2	216248908	216249582	675	-	FN1		psi3	8.9082e-07	0.45558	-2.64	0.86	-0.14	12046.4	12404.3	22441	26017	2.048886e-05	0.113650461635756	reverse	TRUE
P25	chr17	40555726	40555727	2	-	CAVIN1		theta	1.0148e-06	0.2001	-2.64	0.72	-0.27	3678.6	3852.6	4573	6312	1.0148e-05	0.108631733641734	reverse	TRUE
P25	chr12	49523506	49525080	1575	-	TUBA1B		psi3	2.0771e-06	0.53114	-2.59	0.66	-0.34	6210.3	6489	5302	8034	1.24626e-05	0.108631733641734	reverse	TRUE
P25	chr12	49580617	49582759	2143	-	TUBA1A		psi5	2.5995e-06	0.33621	-2.61	0.73	-0.27	4824.8	5098.7	7453	10181	7.7985e-06	0.108631733641734	reverse	TRUE
P25	chr9	135896130	135896131	2	+	EEF1A1P5		theta	3.0237e-06	0.20541	-2.65	0.87	-0.13	2849.4	2892.5	2867	3298	1.81422e-05	0.113650461635756	reverse	TRUE
P25	chr20	1350248	1350249	2	-	AL136531.2	FKBP1A	theta	3.2851e-06	0.21004	-2.66	0.83	-0.17	2146.1	2181.6	1775	2130	6.5702e-06	0.108631733641734	reverse	TRUE
P25	chr2	65496065	65496066	2	+	ACTR2		theta	1.2073e-05	0.30165	-2.66	0.9	-0.1	1612.8	1633.1	1792	1995	4.8292e-05	0.196440046042754	reverse	TRUE
P25	chr2	38709444	38709445	2	+	RPLP0P6		theta	1.6535e-05	0.35946	-2.66	0.87	-0.13	939.8	953.4	880	1016	6.614e-05	0.224125189614505	reverse	TRUE
P25	chr8	144993371	144993372	2	-	PLEC		theta	1.7333e-05	0.36714	-2.64	0.87	-0.12	1132.1	1153	1453	1662	0.000242662	0.477624034964986	reverse	TRUE
P25	chrX	51639743	51639744	2	+	MAGED1		theta	2.5098e-05	0.45467	-2.65	0.89	-0.1	934.6	950.4	1323	1481	0.000200784	0.445567403957804	reverse	TRUE
P25	chrX	48436404	48436405	2	+	RBM3		theta	3.1717e-05	0.51197	-2.65	0.89	-0.11	840.3	851.1	867	975	0.000126868	0.368619998600066	reverse	TRUE
P25	chr10	120801787	120801788	2	-	EIF3A		theta	4.0933e-05	0.60947	-2.41	0.82	-0.17	1252.6	1312.6	1419	1729	0.000286531	0.526152544900932	reverse	TRUE
P25	chr11	6340520	6340521	2	-	CAVIN3		theta	4.3281e-05	0.63505	-2.63	0.87	-0.12	672.8	686.2	907	1041	0.000346248	0.555967968941773	reverse	TRUE
P25	chr6	52362713	52362714	2	-	TRAM2		theta	5.1117e-05	0.7018	-2.65	0.86	-0.14	517.6	526.3	525	612	0.000204468	0.445567403957804	reverse	TRUE
P25	chr19	41808779	41808780	2	+	HNRNPUL1		theta	5.6264e-05	0.7458	-2.64	0.89	-0.11	606.4	614.6	636	718	0.000393848	0.600460232094169	reverse	TRUE
P25	chr2	189918675	189949885	31211	-	COL5A2		psi5	6.3497e-05	1	2.68	0.12	0.12	19.5	796.4	195	1618	0.000253988	0.484294251774754	reverse	TRUE
P25	chr1	1431081	1431082	2	+	ATAD3B		theta	7.653e-05	0.85075	-2.56	0.33	-0.62	82.3	87.3	23	69	0.00084183	0.959886742607655	reverse	TRUE
P25	chr8	119123170	119123171	2	-	EXT1		theta	8.3097e-05	0.87918	-2.6	0.89	-0.1	572.5	580.9	712	796	0.000332388	0.54916017805483	reverse	TRUE
P25	chr6	116598279	116598280	2	-	TSPYL1		theta	8.3252e-05	0.87918	-2.64	0.84	-0.15	351	358.6	400	476	0.000333008	0.54916017805483	reverse	TRUE
P25	chr1	32157650	32157651	2	-	COL16A1		theta	0.00010049	0.98651	-2.6	0.88	-0.11	479.9	487.9	598	678	0.00060294	0.782748660372482	reverse	TRUE
P25	chr9	133499029	133499030	2	+	FUBP3		theta	0.00011461	1	-2.65	0.89	-0.1	474.2	481.1	585	654	0.00045844	0.639789519550122	reverse	TRUE
P25	chr16	55513234	55513235	2	+	MMP2		theta	0.00012497	1	-2.63	0.88	-0.12	444.6	459.6	1100	1249	0.00074982	0.933699302819695	reverse	TRUE
P25	chr17	37075287	37075288	2	+	LASP1		theta	0.00018183	1	-2.29	0.08	-0.83	285.4	349.2	41	536	0.00090915	0.973210490676073	reverse	TRUE
P25	chr6	33268335	33268336	2	-	TAPBP		theta	0.00019863	1	-2.63	0.77	-0.22	159	165.4	215	279	0.00079452	0.950562664026005	reverse	TRUE

FRASER calls 3/4

sampleID	seqnames	start	end	width	strand	hgncSymbol	addHgncSymbols	type	pValue	padjust	zScore	psiValue	deltaPsi	meanCounts	meanTotalCounts	counts	totalCounts	pValueGene	padjustGene	STRAND_SPECIFIC	PAIRED_END
P25	chr2	86732201	86732202	2	-	CHMP3		theta	0.00020833	1	-2.65	0.88	-0.12	296	300.9	353	402	0.00083332	0.959886742607655	reverse	TRUE
P26	chr9	75775203	75775204	2	+	ANXA1		theta	1.1644e-06	0.17294	-2.74	0.9	-0.1	5804.9	5880.1	6884	7636	4.6576e-06	0.162570787228326	reverse	TRUE
P26	chr21	47610354	47610355	2	-	LSS		theta	2.8576e-06	0.17294	-2.71	0.6	-0.4	810.8	863.6	784	1311	6.28672e-05	0.214952997977593	reverse	TRUE
P26	chr19	3055707	3055708	2	-	TLE5		theta	3.4923e-06	0.1898	-2.73	0.86	-0.14	1938.1	1973.9	2118	2476	3.14307e-05	0.193439850166093	reverse	TRUE
P26	chr6	29912394	29912835	442	+	HLA-A		psi3	4.1512e-06	0.70766	-1.82	0.72	-0.26	3097.3	3190.7	1308	1823	0.0001079312	0.272616716363847	reverse	TRUE
P26	chr6	24977114	24977115	2	+	PPIAP29		theta	8.8095e-06	0.2603	-2.74	0.86	-0.14	1111.7	1130	1106	1289	2.64285e-05	0.180726525343925	reverse	TRUE
P26	chr1	89475178	89476586	1409	-	GBP3		psi3	9.5036e-06	1	-2.56	0	-0.86	99.8	101.5	0	7	0.0001520576	0.311945532665151	reverse	TRUE
P26	chr19	4838841	4838842	2	-	PLIN3		theta	1.2786e-05	0.31566	-2.74	0.84	-0.16	837.9	861.1	1207	1439	5.1144e-05	0.204129869122249	reverse	TRUE
P26	chr7	98015303	98015304	2	-	BAIAP2L1	RPS3AP26	theta	1.3247e-05	0.31566	-2.74	0.82	-0.18	749.2	768.5	883	1076	5.2988e-05	0.204129869122249	reverse	TRUE
P26	chr15	73994708	73994709	2	+	CD276		theta	1.4192e-05	0.31566	-2.72	0.74	-0.25	455.7	477.3	622	838	9.9344e-05	0.272616716363847	reverse	TRUE
P26	chr20	18548187	18548188	2	+	SMIM26	AL121900.1	theta	2.1659e-05	0.36672	-2.73	0.71	-0.28	335.5	342.4	170	238	0.000129954	0.285642952914603	reverse	TRUE
P26	chr11	62494370	62494371	2	-	HNRNPUL2-B	HNRNPUL2	theta	4.2083e-05	0.53813	-2.71	0.53	-0.46	132.7	143.8	123	234	0.000126249	0.285642952914603	reverse	TRUE
P26	chr7	44146335	44146336	2	+	AEBP1		theta	5.3051e-05	0.6479	-2.46	0.79	-0.21	857.5	950.6	1382	1754	0.000477459	0.624994485390682	reverse	TRUE
P26	chr2	178082450	178082451	2	+	HNRNPA3		theta	5.6282e-05	0.67973	-2.67	0.8	-0.19	382.7	393.6	445	553	0.000393974	0.551070012970355	reverse	TRUE
P26	chrX	20146155	20146156	2	-	EIF1AX		theta	7.0712e-05	0.79976	-2.72	0.89	-0.11	565	574	721	811	0.000212136	0.354528420566436	reverse	TRUE
P26	chr12	52467687	52470569	2883	+	ATG101		psi3	9.2813e-05	1	-1.11	0.8	-0.14	338	376.6	445	559	0.000371252	0.544015664801658	reverse	TRUE
P26	chr4	57326959	57326960	2	+	PAICS		theta	0.0001017	1	-2.71	0.76	-0.23	204.4	207.9	109	144	0.0003051	0.481470291096863	reverse	TRUE
P26	chr7	128410012	128410013	2	+	CALU		theta	0.0001244	1	-2.08	0.89	-0.11	3740.8	3844.7	4950	5551	0.0004976	0.624994485390682	reverse	TRUE
P26	chr6	33287857	33287858	2	-	DAXX		theta	0.00012763	1	-2.72	0.82	-0.17	237.1	244.3	324	396	0.00051052	0.628397791374635	reverse	TRUE
P26	chr17	4575419	4575420	2	-	PELP1		theta	0.0001436	1	-2.71	0.81	-0.18	210.7	216	226	279	0.0005744	0.68594560646302	reverse	TRUE
P26	chr6	10724867	10749854	24988	+	TMEM14C	TMEM14B;AL024498.2	psi3	0.00020148	1	2.78	0.12	0.12	5.4	376.6	54	453	0.00040296	0.551113840381317	reverse	TRUE
P26	chr6	10749502	10749854	353	+	TMEM14B	AL024498.2	psi3	0.00020148	1	-2.72	0.88	-0.12	371.1	376.6	399	453	0.00020148	0.354528420566436	reverse	TRUE
P26	chr10	89720693	89720694	2	+	PTEN		theta	0.00021309	1	-2.72	0.89	-0.11	299.2	302.7	273	308	0.00085236	0.794824085270282	reverse	TRUE
P26	chr5	864221	864222	2	-	BRD9		theta	0.00023616	1	-2.72	0.64	-0.34	77.8	81.2	61	95	0.0011808	0.995509429621542	reverse	TRUE
P26	chr19	50335415	50335564	150	+	MED25		psi5	0.000613	1	-2.75	0.8	-0.19	125.6	129.5	145	182	0.000613	0.68594560646302	reverse	TRUE
P26	chr19	50335415	50361805	26391	+	PTOV1	MED25;MIR4749	psi5	0.000613	1	2.78	0.2	0.19	3.7	129.5	37	182	0.000613	0.68594560646302	reverse	TRUE
P26	chr2	37431503	37431504	2	+	CEBPZOS		theta	0.0006428	1	-2.71	0.84	-0.15	127.6	129.8	116	138	0.0006428	0.686340100576837	reverse	TRUE
P29	chr21	47424104	47424105	2	+	COL6A1		theta	2.6922e-07	0.16015	-2.73	0.85	-0.15	11667.3	11793	7137	8394	1.373022e-05	0.177392788163101	reverse	TRUE
P29	chr21	47518088	47531371	13284	+	COL6A2		psi3	2.9889e-06	1	-2.41	0	-0.88	736	745.2	0	9	8.66781e-05	0.399953788891534	reverse	TRUE
P29	chr6	29856416	29856417	2	+	HLA-H		theta	3.3632e-06	0.20845	-2.68	0.1	-0.89	1050.7	1143.6	102	1027	8.07168e-05	0.399953788891534	reverse	TRUE
P29	chr16	31201682	31201683	2	+	FUS		theta	6.139e-06	0.27803	-2.71	0.84	-0.16	1336.7	1357.3	1082	1288	5.5251e-05	0.377263065917384	reverse	TRUE
P29	chr17	78183459	78183460	2	-	SGSH		theta	8.5956e-06	0.30917	-2.72	0.27	-0.72	192.5	211.3	68	256	4.2978e-05	0.377263065917384	reverse	TRUE
P29	chr19	13885488	13885489	2	+	C19orf53		theta	9.3504e-06	0.30917	-2.73	0.87	-0.13	1204.1	1221.5	1170	1344	4.6752e-05	0.377263065917384	reverse	TRUE
P29	chr1	1717167	1717168	2	-	GNB1		theta	1.6131e-05	0.41745	-2.73	0.89	-0.11	970	981.4	885	999	0.000145179	0.446594092529209	reverse	TRUE
P29	chr17	6917814	6917815	2	+	RNASEK	RNASEK-C17orf49;C17orf49	theta	2.0803e-05	0.43347	-2.73	0.78	-0.22	399.5	407.1	266	342	0.000104015	0.446594092529209	reverse	TRUE
P29	chr5	151042818	151042819	2	-	SPARC		theta	2.1852e-05	0.43347	-2.26	0.9	-0.1	11317.9	11475.7	10138	11294	0.000262224	0.62738934942049	reverse	TRUE
P29	chr13	31036696	31036697	2	-	HMGB1		theta	2.9075e-05	0.50722	-2.71	0.88	-0.11	782.9	790.1	550	622	0.0001163	0.446594092529209	reverse	TRUE
P29	chr16	418620	418621	2	-	MRPL28		theta	3.4746e-05	0.54734	-2.73	0.81	-0.19	394.4	404.4	429	529	0.000138984	0.446594092529209	reverse	TRUE
P29	chr4	119686019	119686020	2	-	SEC24D		theta	5.4658e-05	0.79433	-2.72	0.83	-0.16	338.1	344.7	327	393	0.00027329	0.630513191718366	reverse	TRUE
P29	chr1	24082876	24083442	567	+	ELOA		psi5	8.074e-05	1	-2.74	0.61	-0.38	158.2	165.4	112	184	0.0004037	0.899268165799337	reverse	TRUE
P29	chr2	183606083	183606084	2	+	DNAJC10		theta	9.9969e-05	1	2.49	0.23	0.22	5.5	324.8	45	198	0.000199938	0.516634974235724	reverse	TRUE
P29	chr15	83041045	83041046	2	-	RPL9P8		theta	0.00012867	1	-2.12	0.56	-0.44	1168.1	1501.2	2344	4187	0.00012867	0.446594092529209	reverse	TRUE
P29	chr11	68777413	68777414	2	-	MRGPRF		theta	0.00012911	1	-2.71	0.85	-0.15	276	279	169	199	0.00051644	0.981226849787763	reverse	TRUE
P29	chr15	82664505	82664506	2	-	RPL9P9		theta	0.00012922	1	-2.12	0.56	-0.44	1167.9	1501	2342	4185	0.00012922	0.446594092529209	reverse	TRUE
P29	chr5	52394292	52394293	2	-	MOCS2		theta	0.00013346	1	-2.72	0.85	-0.15	268.3	272.6	239	282	0.00053384	0.985306926025636	reverse	TRUE

FRASER calls 4/4

sampleID	seqnames	start	end	width	strand	hgncSymbol	addHgncSymbols	type	pValue	padjust	zScore	psiValue	deltaPsi	meanCounts	meanTotalCounts	counts	totalCounts	pValueGene	padjustGene	STRAND_SPECIFIC	PAIRED_END
P32	chr15	60639860	60639861	2	-	ANXA2		theta	3.3373e-07	0.33185	-2.65	0.89	-0.11	22180.2	22438.3	20944	23525	3.3373e-06	0.170177005784621	reverse	TRUE
P32	chr12	49578888	49578889	2	-	TUBA1A		theta	2.4866e-06	0.33185	-2.61	0.89	-0.11	5525	5577.7	4313	4840	9.9464e-06	0.170177005784621	reverse	TRUE
P32	chr6	29925272	29925273	2	+	HLA-W		theta	5.1993e-06	0.33185	2.06	0.79	0.73	4.6	974	11	14	1.03986e-05	0.170177005784621	reverse	TRUE
P32	chr2	47130231	47130232	2	-	MCFD2		theta	1.2109e-05	0.43734	-2.64	0.76	-0.23	634.6	653.8	617	809	4.8436e-05	0.274686128757168	reverse	TRUE
P32	chr11	33729536	33729537	2	-	AL049629.2	CD59	theta	1.3574e-05	0.43734	-2.6	0.6	-0.39	387.7	416.7	437	727	5.4296e-05	0.284232796591204	reverse	TRUE
P32	chr3	154900965	154900966	2	+	MME		theta	1.368e-05	0.43734	-2.58	0.78	-0.21	1173	1208.4	1288	1642	0.0001368	0.489984340668497	reverse	TRUE
P32	chr15	72492984	72494794	1811	-	PKM		psi3	1.8613e-05	1	-1.28	0	-0.57	68.5	582.7	0	74	0.000260582	0.611499518486965	reverse	TRUE
P32	chr3	127410913	127410914	2	-	MGLL		theta	1.9989e-05	0.53046	-2.63	0.74	-0.26	417.4	435.5	504	685	7.9956e-05	0.362751740208244	reverse	TRUE
P32	chr7	35840881	35871104	30224	+	SEPTIN7		psi5	2.0677e-05	1	2.52	0.35	0.34	56.3	903.2	339	974	0.000227447	0.552804902440613	reverse	TRUE
P32	chr6	122765253	122765254	2	-	SERINC1		theta	2.6291e-05	0.5832	-2.65	0.89	-0.11	904.1	914.3	789	891	0.000105164	0.420986218971093	reverse	TRUE
P32	chr1	55352610	55352611	2	-	DHCR24		theta	3.4951e-05	0.71678	-2.54	0.77	-0.23	527.5	531.5	131	171	0.000104853	0.420986218971093	reverse	TRUE
P32	chr20	53691404	53691405	2	+	RPL12P4		theta	3.7685e-05	0.74476	-2.64	0.78	-0.21	388.7	402.9	508	650	0.00037685	0.827287628946788	reverse	TRUE
P32	chr2	61719334	61719459	126	-	XPO1		psi5	3.9413e-05	1	-2.64	0.75	-0.24	394.7	410.8	482	639	0.000157652	0.496365025638091	reverse	TRUE
P32	chr6	18258631	18258632	2	-	DEK		theta	3.9924e-05	0.76132	-2.63	0.84	-0.15	477.3	487.2	531	630	0.00019962	0.522492917116695	reverse	TRUE
P32	chr8	146016696	146017257	562	-	RPL8		psi3	4.3404e-05	1	2.7	0.65	0.63	14.7	145.3	147	225	0.000173616	0.496365025638091	reverse	TRUE
P32	chr16	56660967	56660968	2	+	MT1E		theta	5.3894e-05	0.89195	-2.53	0.88	-0.12	545	551.3	443	506	0.000431152	0.862980916862352	reverse	TRUE
P32	chr17	7485153	7485154	2	+	AC016876.3	CD68	theta	7.376e-05	1	-2.61	0.89	-0.11	602	610.5	695	780	0.00029504	0.669282314216821	reverse	TRUE
P32	chr13	114287601	114288204	604	+	TFDP1		psi3	8.2223e-05	1	-2.7	0.85	-0.15	447.8	457.7	544	643	0.000411115	0.862980916862352	reverse	TRUE
P32	chr5	52388758	52388759	2	+	ITGA2		theta	0.00012473	1	-2.31	0	-0.9	131.2	164.5	0	247	0.00049892	0.94314424091947	reverse	TRUE
P32	chr20	35826803	35826804	2	+	RPN2		theta	0.00023955	1	-2.6	0.45	-0.51	56.5	63.3	56	124	0.0004791	0.931553561961416	reverse	TRUE
P33	chr2	38709324	38709325	2	+	RPLP0P6		theta	8.2868e-06	0.70064	-2.34	0.72	-0.27	1501.6	1544.2	1121	1547	3.31472e-05	0.369706619764706	reverse	TRUE
P33	chr13	113976750	113976751	2	+	LAMP1		theta	1.0286e-05	0.70064	-2.32	0.87	-0.13	2309.7	2338.6	1921	2210	5.143e-05	0.422178548822565	reverse	TRUE
P33	chr11	69468818	69468819	2	+	CCND1		theta	1.0432e-05	0.70064	-2.31	0.81	-0.19	2071.8	2117.9	1914	2374	0.000135616	0.555738712105515	reverse	TRUE
P33	chr19	10504000	10504001	2	-	CDC37		theta	3.114e-05	1	-2.33	0.84	-0.15	935.6	953.1	937	1112	0.00012456	0.555738712105515	reverse	TRUE
P33	chr14	75598980	75598981	2	-	TMED10		theta	3.1896e-05	1	-2.33	0.56	-0.41	299.7	307.7	101	181	0.000127584	0.555738712105515	reverse	TRUE
P33	chr8	144995405	144995406	2	-	PLEC		theta	4.7404e-05	1	-2.31	0.89	-0.11	971.8	987.8	1281	1441	0.000379232	0.713544177357312	reverse	TRUE
P33	chr3	47958377	47958378	2	-	MAP4		theta	5.3937e-05	1	-2.33	0.9	-0.1	956.7	967.1	930	1034	0.000323622	0.683449403528176	reverse	TRUE
P33	chr18	8376043	8376044	2	+	PTPRM		theta	5.596e-05	1	2.32	0.56	0.52	5.9	154.7	58	103	0.0002798	0.627699736224168	reverse	TRUE
P33	chr9	139687451	139687452	2	+	TMEM141	AL355987.3	theta	5.9099e-05	1	-2.33	0.57	-0.4	203.8	213	122	214	0.000236396	0.593403019787128	reverse	TRUE
P33	chr18	346892	346893	2	-	COLEC12		theta	6.4203e-05	1	-2.3	0.88	-0.11	873.5	877.7	301	343	0.000256812	0.599173901520976	reverse	TRUE
P33	chr19	13885389	13885443	55	+	C19orf53		psi5	6.4232e-05	1	2.3	0.19	0.18	18	1100.8	180	970	0.000513856	0.833041452650438	reverse	TRUE
P33	chr6	170627851	170627852	2	+	FAM120B		theta	0.00011279	1	-2.32	0.41	-0.53	108.3	116.4	56	137	0.00056395	0.833041452650438	reverse	TRUE
P33	chr10	70742383	70742384	2	+	DDX21		theta	0.00012355	1	-2.31	0.72	-0.26	247.3	253.7	162	226	0.0004942	0.833041452650438	reverse	TRUE
P33	chr19	11777223	11777224	2	-	HNRNPA1P10		theta	0.00013595	1	-2.32	0.77	-0.21	286.9	291.8	162	211	0.0005438	0.833041452650438	reverse	TRUE
P33	chr9	130648354	130648355	2	-	AL157935.3	ST6GALNAC6	theta	0.00023198	1	-2.32	0.82	-0.17	265.7	272.4	297	364	0.00046396	0.820057271436598	reverse	TRUE

MAE calls

gene_name	ID	contig	position	variantID	refAllele	altAllele	refCount	altCount	totalCount	pvalue	padj	log2FC	altRatio	AF	AF_afr	AF_amr	AF_eas	AF_nfe	MAX_AF	rare	gene_type	other_names	N_var	cohort_freq	MAE	MAE_ALT
SNHG14	P16	chr15	25304758	.	C	T	0	29	29	0.000446527622348894	0.0114957111285566	7.40017797023297	1	1E-04	3E-04	0	0	0	3E-04	FALSE	lncRNA		1	0.1	TRUE	TRUE
FLNB	P16	chr3	58104626	.	G	T	9	104	113	8.52570006153225e-08	1.57364192661163e-05	3.63004788189686	0.92	7E-04	0	0	0	0.001	0.001	FALSE	protein_coding		1	0.1	TRUE	TRUE
HGS	P22	chr17	79660663	.	A	G	4	132	136	2.11598843118188e-09	4.42818669870973e-07	5.16637994719267	0.971	7E-05	0	0	0	1E-04	1E-04	FALSE	protein_coding		1	0.1	TRUE	TRUE
AIFM1	P24	chrX	129299753	.	C	A	11	58	69	0.000125679530701733	0.00335863029808936	2.52407833497455	0.841	5E-05	0	0	0	9E-05	9E-05	FALSE	protein_coding		1	0.1	TRUE	TRUE
COLEC12	P24	chr18	318029	.	T	A	2	21	23	0.00244550996681754	0.0346201110523429	3.51784265785704	0.913	0.002	7E-04	0.001	0	0.004	0.004	FALSE	protein_coding	AP000915.1	1	0.1	TRUE	TRUE
MRI1	P24	chr19	13875821	.	G	T	7	30	37	0.00333868088230662	0.0425943532312332	2.22506440597205	0.811	0.003	9E-04	0.001	0	0.005	0.005	FALSE	protein_coding		1	0.1	TRUE	TRUE
PCSK9	P24	chr1	55505651	.	C	T	0	12	12	0.00389602023969395	0.0478888895578904	6.1531594501359	1	0.008	0.002	0.01	0	0.01	0.01	FALSE	protein_coding		1	0.1	TRUE	TRUE
MEST	P25	chr7	130145727	.	A	C	0	13	13	0.00327964644383008	0.0455053706289042	6.258582626624	1	0.001	7E-04	0.002	0	0.002	0.002	FALSE	protein_coding		1	0.1	TRUE	TRUE
MMP1	P25	chr11	102660962	.	T	G	1557	7366	8923	2.63744488010503e-07	3.05636173866895e-05	2.35758782913522	0.826	0.003	0.002	0.002	0	0.003	0.003	FALSE	protein_coding	WTAPP1	1	0.1	TRUE	TRUE
NGFR	P25	chr17	47590459	.	G	T	2	128	130	5.41376804562646e-08	8.11592576300303e-06	6.11546863161224	0.985	0.002	7E-04	0.006	0	0.003	0.006	FALSE	protein_coding	AC006487.1	1	0.1	TRUE	TRUE
AL669831.3	P26	chr1	565283	.	T	C	1	1799	1800	5.17069099079231e-13	1.03252583215058e-10	10.9284305115231	0.999	3E-04	3E-04	0	0	5E-04	5E-04	FALSE	transcribed_processed_pseudogene	MTND2P28	1	0.1	TRUE	TRUE
SSNA1	P26	chr9	140083122	.	G	A	14	73	87	5.71830943664438e-05	0.00174089712373644	2.497945062888	0.839	0.006	0.002	0	0	0.008	0.008	FALSE	protein_coding		1	0.1	TRUE	TRUE
DNAJC10	P29	chr2	183606088	.	G	C	0	45	45	0.000130148652156937	0.00351256290670052	8.04142698755107	1	7E-04	0	0	0	0.001	0.001	FALSE	protein_coding		1	0.1	TRUE	TRUE
PRMT1	P29	chr19	50190833	.	G	C	1	23	24	0.00268652132820438	0.0374456003221514	4.63046524888601	0.958	0.004	8E-04	0.001	0	0.005	0.005	FALSE	protein_coding		1	0.1	TRUE	TRUE
BLZF1	P29	chr1	169337581	.	G	A	7	70	77	2.77371593712368e-06	0.000161796696463842	3.4288406279806	0.909	0.006	0.002	0.007	0	0.01	0.01	FALSE	protein_coding		1	0.1	TRUE	TRUE
MEG3	P32	chr14	101298291	.	T	G	1	160	161	9.10468391365233e-07	8.27885061218866e-05	7.45049428130272	0.994	0.007	0.002	0.007	0	0.01	0.01	FALSE	lncRNA		1	0.1	TRUE	TRUE
MUC20-OT1	P32	chr3	195393086	.	G	A	1	161	162	8.82971447182377e-07	8.06669386274919e-05	7.45948304398127	0.994	0.007	0.004	0.01	0	0.01	0.01	FALSE	lncRNA	SDHAP2	1	0.1	TRUE	TRUE
MSMO1	P33	chr4	166249087	.	G	A	23	99	122	8.23941554571518e-05	0.00239190233292112	2.2260873492767	0.811	9E-04	0.002	0.002	0	6E-04	0.002	FALSE	protein_coding		1	0.1	TRUE	TRUE

Supplemental Method[LSEP]:

Urine derived stem cell isolation and culture.

Approximately 200 ml of fresh urine samples were collected from donors without NDD. Each urine sample was centrifuged at 400g for 10 min at room temperature and the supernatant was discarded. Next, 5 ml of washing buffer (Dulbecco's phosphate-buffered saline buffer supplemented with 100 U/ml penicillin, 100 µg/ml streptomycin, and 500 ng/ ml amphotericin B) was added and centrifuged at 400 × g for 10 min at room temperature. After carefully discarding the supernatant, 3 ml of proliferation medium (1:1 mixture of DMEM high glucose (Gibco, Thermo Fisher Scientific, Waltham, USA) / REGM with SingleQuot Kit (LONZA, Basel, Switzerland), supplemented with 5% (v/v) FBS, 0.5% (v/v) NEAA, 0.5% (v/v) GlutaMax, 50 U/ml penicillin, 50µg/ml streptomycin, and 1.5 µg/ml amphotericin B) was added to suspend the cell pellet, and the volume was transferred into a single well of a 6-well plate (coated beforehand with 0.1% (w/v) gelatin). The cells were then incubated at 37 °C. Approximately 96 h after plating, most of the medium was aspirated, leaving approximately 1 ml, and then 2 ml of proliferation medium was added. Half of the proliferation medium was replaced daily.

References:

1. International Classification of Diseases, Eleventh Revision (ICD-11), World Health Organization (WHO) 2019/2021.https://icd.who.int/browse11.

2. Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., Keerthikumar, S., Oortveld, M.A.W., Kleefstra, T., Kramer, J.M., et al. (2016). Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am. J. Hum. Genet. *98*, 149–164. 10.1016/j.ajhg.2015.11.024.

3. Leblond, C.S., Le, T.-L., Malesys, S., Cliquet, F., Tabet, A.-C., Delorme, R., Rolland, T., and Bourgeron, T. (2021). Operative list of genes associated with autism and neurodevelopmental disorders based on database review. Mol. Cell. Neurosci. *113*, 103623. 10.1016/j.mcn.2021.103623.

4. Srivastava, S., Love-Nichols, J.A., Dies, K.A., Ledbetter, D.H., Martin, C.L., Chung, W.K., Firth, H.V., Frazier, T., Hansen, R.L., Prock, L., et al. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. Genet. Med. *21*, 2413–2421. 10.1038/s41436-019-0554-6.

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

6. Palmer, E.E., Sachdev, R., Macintosh, R., Melo, U.S., Mundlos, S., Righetti, S., Kandula, T., Minoche, A.E., Puttick, C., Gayevskiy, V., et al. (2021). Diagnostic Yield of Whole Genome Sequencing After Nondiagnostic Exome Sequencing or Gene Panel in Developmental and Epileptic Encephalopathies. Neurology *96*, e1770–e1782. 10.1212/WNL.0000000000011655.

7. Sun, Y., Peng, J., Liang, D., Ye, X., Xu, N., Chen, L., Yan, D., Zhang, H., Xiao, B., Qiu, W., et al. (2022). Genome sequencing demonstrates high diagnostic yield in children with undiagnosed global developmental delay/intellectual disability: A prospective study. Hum. Mutat. *43*, 568–581. 10.1002/humu.24347.

8. Ewans, L.J., Minoche, A.E., Schofield, D., Shrestha, R., Puttick, C., Zhu, Y., Drew, A., Gayevskiy, V., Elakis, G., Walsh, C., et al. (2022). Whole exome and genome sequencing in mendelian disorders: a diagnostic and health economic analysis. Eur. J. Hum. Genet. *30*, 1121–1131. 10.1038/s41431-022-01162-2.

9. Gonorazky, H.D., Naumenko, S., Ramani, A.K., Nelakuditi, V., Mashouri, P., Wang, P., Kao, D., Ohri, K., Viththiyapaskaran, S., Tarnopolsky, M.A., et al. (2019). Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. Am. J. Hum. Genet. *104*, 466–483. 10.1016/j.ajhg.2019.01.012.

10. Murdock, D.R., Dai, H., Burrage, L.C., Rosenfeld, J.A., Ketkar, S., Müller, M.F., Yépez, V.A., Gagneur, J., Liu, P., Chen, S., et al. (2021). Transcriptome-directed analysis for Mendelian disease diagnosis overcomes limitations of conventional genomic testing. J. Clin. Invest. *131*, e141500. 10.1172/JCI141500.

11. 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., et al. (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. 10.3389/fcell.2022.1021785.

12. Dekker, J., Schot, R., Bongaerts, M., de Valk, W.G., van Veghel-Plandsoen, M.M., Monfils, K., Douben, H., Elfferich, P., Kasteleijn, E., van Unen, L.M.A., et al. (2023). Web-accessible application for identifying pathogenic transcripts with RNA-seq: Increased sensitivity in diagnosis of neurodevelopmental disorders. Am. J. Hum. Genet. *110*, 251–272. 10.1016/j.ajhg.2022.12.015.

13. Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics *32*, 1220–1222. 10.1093/bioinformatics/btv710.

14. Rausch, T., Zichner, T., Schlattl, A., Stutz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics *28*, i333–i339. 10.1093/bioinformatics/bts378.

15. Dolzhenko, E., Deshpande, V., Schlesinger, F., Krusche, P., Petrovski, R., Chen, S., Emig-Agius, D., Gross, A., Narzisi, G., Bowman, B., et al. (2019). ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. Bioinformatics *35*, 4754–4756. 10.1093/bioinformatics/btz431.

16. Gardner, E.J., Lam, V.K., Harris, D.N., Chuang, N.T., Scott, E.C., Pittard, W.S., Mills, R.E., The 1000 Genomes Project Consortium, and Devine, S.E. (2017). The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res. *27*, 1916–1929. 10.1101/gr.218032.116.

17. Geoffroy, V., Herenger, Y., Kress, A., Stoetzel, C., Piton, A., Dollfus, H., and Muller, J. (2018). AnnotSV: an integrated tool for structural variations annotation. Bioinformatics *34*, 3572–3574. 10.1093/bioinformatics/bty304.

18.; on behalf of the ACMG Laboratory Quality Assurance Committee, Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–423. 10.1038/gim.2015.30.

19. Quinodoz, M., Peter, V.G., Bedoni, N., Royer Bertrand, B., Cisarova, K., Salmaninejad, A., Sepahi, N., Rodrigues, R., Piran, M., Mojarrad, M., et al. (2021). AutoMap is a high performance homozygosity mapping tool using next-generation sequencing data. Nat. Commun. *12*, 518. 10.1038/s41467-020-20584-4.

20. Baux, D., Van Goethem, C., Ardouin, O., Guignard, T., Bergougnoux, A., Koenig, M., and Roux, A.-F. (2021). MobiDetails: online DNA variants interpretation. Eur. J. Hum. Genet. *29*, 356–360. 10.1038/s41431-020-00755-z.

21. Capriotti, E., Fariselli, P., and Casadio, R. (2005). I-Mutant2.0: predicting stability changes

upon mutation from the protein sequence or structure. Nucleic Acids Res. *33*, W306–W310. 10.1093/nar/gki375.

22. Pires, D.E.V., Ascher, D.B., and Blundell, T.L. (2014). DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res. *42*, W314–W319. 10.1093/nar/gku411.

23. Hoefgen, S., Dahms, S.O., Oertwig, K., and Than, M.E. (2015). The Amyloid Precursor Protein Shows a pH-Dependent Conformational Switch in Its E1 Domain. J. Mol. Biol. *427*, 433–442. 10.1016/j.jmb.2014.12.005.

24. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. 10.1093/bioinformatics/bts635.

25. Yépez, V.A., Mertes, C., Müller, M.F., Klaproth-Andrade, D., Wachutka, L., Frésard, L., Gusic, M., Scheller, I.F., Goldberg, P.F., Prokisch, H., et al. (2021). Detection of aberrant gene expression events in RNA sequencing data. Nat. Protoc. *16*, 1276–1296. 10.1038/s41596-020-00462-5.

26. Brechtmann, F., Mertes, C., Matusevičiūtė, A., Yépez, V.A., Avsec, Ž., Herzog, M., Bader, D.M., Prokisch, H., and Gagneur, J. (2018). OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data. Am. J. Hum. Genet. *103*, 907–917. 10.1016/j.ajhg.2018.10.025.

27. Mertes, C., Scheller, I.F., Yépez, V.A., Çelik, M.H., Liang, Y., Kremer, L.S., Gusic, M., Prokisch, H., and Gagneur, J. (2021). Detection of aberrant splicing events in RNA-seq data using FRASER. Nat. Commun. *12*, 529. 10.1038/s41467-020-20573-7.

28. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. 10.1186/s13059-014-0550-8.

29. Kremer, L.S., Bader, D.M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., Haack, T.B., Graf, E., Schwarzmayr, T., Terrile, C., et al. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. Nat. Commun. *8*, 15824. 10.1038/ncomms15824.

30. Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. 10.1038/nbt.3519.

31. 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. 10.1038/s41436-020-0780-y.

32. Mak, C.C.Y., Doherty, D., Lin, A.E., Vegas, N., Cho, M.T., Viot, G., Dimartino, C., Weisfeld-Adams, J.D., Lessel, D., Joss, S., et al. (2020). MN1 C-terminal truncation syndrome is a novel neurodevelopmental and craniofacial disorder with partial rhombencephalosynapsis. Brain *143*, 55–68. 10.1093/brain/awz379.

33. Vuillaume, M.-L., Cogné, B., Jeanne, M., Boland, A., Ung, D.-C., Quinquis, D., Besnard, T., Deleuze, J.-F., Redon, R., Bézieau, S., et al. (2018). Whole genome sequencing identifies a de

novo 2.1 Mb balanced paracentric inversion disrupting FOXP1 and leading to severe intellectual disability. Clin. Chim. Acta *485*, 218–223. 10.1016/j.cca.2018.06.048.

34. Tepe, B., Macke, E.L., Niceta, M., Weisz Hubshman, M., Kanca, O., Schultz-Rogers, L., Zarate, Y.A., Schaefer, G.B., Granadillo De Luque, J.L., Wegner, D.J., et al. (2023). Bi-allelic variants in INTS11 are associated with a complex neurological disorder. Am. J. Hum. Genet. *110*, 774–789. 10.1016/j.ajhg.2023.03.012.

35. Klein, S., Goldman, A., Lee, H., Ghahremani, S., Bhakta, V., UCLA Clinical Genomics Center, Nelson, S.F., and Martinez-Agosto, J.A. (2016). Truncating mutations in APP cause a distinct neurological phenotype. Ann. Neurol. *80*, 456–460. 10.1002/ana.24727.

36. The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. 10.1038/nature11247.

37. Verkerk, A.J.M.H., Zeidler, S., Breedveld, G., Overbeek, L., Huigh, D., Koster, L., van der Linde, H., de Esch, C., Severijnen, L.-A., de Vries, B.B.A., et al. (2018). CXorf56, a dendritic neuronal protein, identified as a new candidate gene for X-linked intellectual disability. Eur. J. Hum. Genet. *26*, 552–560. 10.1038/s41431-017-0051-9.

38. Rocha, M.E., Silveira, T.R.D., Sasaki, E., Sás, D.M., Lourenço, C.M., Kandaswamy, K.K., Beetz, C., Rolfs, A., Bauer, P., Reardon, W., et al. (2020). Novel clinical and genetic insight into CXorf56-associated intellectual disability. Eur. J. Hum. Genet. *28*, 367–372. 10.1038/s41431-019-0558-3.

39. Rom, A., Melamed, L., Gil, N., Goldrich, M.J., Kadir, R., Golan, M., Biton, I., Perry, R.B.-T., and Ulitsky, I. (2019). Regulation of CHD2 expression by the Chaserr long noncoding RNA gene is essential for viability. Nat. Commun. *10*, 5092. 10.1038/s41467-019-13075-8.

40. Newman, A.G., Sharif, J., Bessa, P., Zaqout, S., Brown, J., Mueller, S., Böhm-Sturm, P., Ohara, O., Koseki, H., Singh, P.B., et al. HP1 deficiency results in De-Repression of Endogenous Retroviruses and Induction of Neurodegeneration via Complement. 55.

41. Wickens, M., Bernstein, D.S., Kimble, J., and Parker, R. (2002). A PUF family portrait: 3'UTR regulation as a way of life. Trends Genet. *18*, 150–157. 10.1016/S0168-9525(01)02616-6.

42. Lu, G., and Hall, T.M.T. (2011). Alternate Modes of Cognate RNA Recognition by Human PUMILIO Proteins. Structure *19*, 361–367. 10.1016/j.str.2010.12.019.

43. Gennarino, V.A., Palmer, E.E., McDonell, L.M., Wang, L., Adamski, C.J., Koire, A., See, L., Chen, C.-A., Schaaf, C.P., Rosenfeld, J.A., et al. (2018). A Mild PUM1 Mutation Is Associated with Adult-Onset Ataxia, whereas Haploinsufficiency Causes Developmental Delay and Seizures. Cell *172*, 924-936.e11. 10.1016/j.cell.2018.02.006.

44. Li, J., Gao, K., Yan, H., Xiangwei, W., Liu, N., Wang, T., Xu, H., Lin, Z., Xie, H., Wang, J., et al. (2019). Reanalysis of whole exome sequencing data in patients with epilepsy and intellectual disability/mental retardation. Gene *700*, 168–175. 10.1016/j.gene.2019.03.037.

45. 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. 10.1186/s13059-019-1828-7.

46. Lang, R., Liu, G., Shi, Y., Bharadwaj, S., Leng, X., Zhou, X., Liu, H., Atala, A., and Zhang, Y. (2013). Self-Renewal and Differentiation Capacity of Urine-Derived Stem Cells after Urine Preservation for 24 Hours. PLoS ONE *8*, e53980. 10.1371/journal.pone.0053980.