Haploinsufficiency of ARFGEF1 is associated with developmental delay, intellectual disability, and epilepsy with variable expressivity

DOI:
10.1038/s41436-021-01218-6

Document Version
Accepted author manuscript

Published in:
Genetics in medicine : official journal of the American College of Medical Genetics

Haploinsufficiency of *ARFGEF1* is associated with developmental delay, intellectual disability and epilepsy with variable expressivity

Quentin Thomas [1,2,3], MS, Thierry Gautier [4], PhD, Dana Marafi [5,6], MD MSc, Thomas Besnard [7,8], PhD, Marjolaine Willems [9], MD, Sébastien Moutton [1,2], MD PhD, Bertand Isidor [7,8], MD PhD, Benjamin Cogné [7,8], PharmD PhD, Solène Conrad [7,8], MS, Romano Tenconi [10], MD PhD, Maria Iascone [10], MD, Arthur Sorlin [1,2,11], MD PhD, Alice Masurel [1,2], MD, Tabib Dabir [12], MD, Adam Jackson [13], MBChB MSc MRCP, Siddharth Banka [13,14], MBBS MRCPCH PhD, Julian Delanne [1,2], MD, James R. Lupski [5,15,16,17], MD PhD, Nebal Waill Saadi [18,19], MD, Fowzan S. Alkuraya [20,21], MD, Fatema Al Zahrani [20], MD, Pankaj B. Agrawal [22,23], MD, Eleina England [24], MS, Jill A. Madden [25], PhD MSc CGC, Jennifer E. Posey [26], MD, Lydie Burglen[27,28], MD PhD, Diana Rodriguez [29], MD PhD, Martin Chevarin [1,11], BS, Sylvie Nguyen [1,11], BS, Frédéric Tran Mau-Them [1,11], MD PhD, Yannis Duffourd [1,11], MSc, Philippine Garret [1,11], MSc PhD, Ange-Line Bruel [1,11], PhD, Patrick Callier [1,11], PharmD PhD, Nathalie Marle [1,11], MD PhD, Anne-Sophie Denomme-Pichon [1,11], MD, Laurence Duplomb [1,11], PhD, Christophe Philippe [1,11], MD PhD, Christel Thauvin-Robinet [1,2,11,30], MD PhD, Jérôme Govin [4], PhD, Laurence Faivre [1,2,11,31], MD PhD and Antonio Vitobello [1,11], PhD.

1 Inserm UMR1231 team GAD, University of Burgundy and Franche-Comté, F-21000 Dijon, France
2 Genetics Center, FHU-TRANSLAD and GIMI Institute, Dijon Bourgogne University Hospital, F-21000 Dijon, France
3 Department of Neurology, Dijon Bourgogne University Hospital, F-21000 Dijon, France
4 Institute for Advanced Biology, Centre de Recherche UGA, INSERM U1209, CNRS UMR 5309, Site Santé, Allée des Alpes, 38700 La Tronche, France
5 Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA
6 Department of Pediatrics, Faculty of Medicine, Kuwait University, P.O. Box 24923, 13110 Safat, Kuwait
7 Service de génétique médicale, CHU Nantes, Nantes, France
8 Université de Nantes, CNRS, INSERM, l'institut du thorax, F-44000 Nantes, France
9 Unité INSERM U 1051, Département de Génétique Médicale, CHRU de Montpellier, 34295 Montpellier Cedex 05, France
10 University of Padova, Laboratorio Genetica Medica Bergamo, Italy
11 Functional Unit of Innovative Diagnosis for Rare Diseases, Dijon Bourgogne University Hospital, F-21000 Dijon, France
12 Medical genetics Department, Belfast City Hospital, UK BT9 7AB
13 Division of Evolution & Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9WL, UK
14 Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester, UK
15 Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, 77030, USA
16 Texas Children's Hospital, Houston, Texas, 77030, USA
17 Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA
18 College of Medicine, University of Baghdad, Baghdad, Iraq
19 Children Welfare Teaching Hospital, Baghdad, Iraq
20 Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia

21 Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University 11533, Riyadh, Saudi Arabia
22 Divisions of Newborn Medicine and Genetics & Genomics, Manton Center for Orphan Disease Research
23 Department of Pediatrics, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
24 Center for Mendelian Genomics, Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
25 The Manton Center for Orphan Disease Research, Division of Genetics & Genomics; Boston Children's Hospital, Boston, MA, USA
26 Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA
27 Centre de Référence des Malformations et Maladies Congénitales du Cervelet, et Département de Génétique, AP-HP.Sorbonne Université, Hôpital Trousseau, 75012, Paris, France
28 Developmental Brain Disorders Laboratory, Imagine Institute, INSERM UMR 1163, 75015 Paris, France
29 APHP, Service de Neuropédiatrie, Hôpital Armand Trousseau, UPMC Université, Paris 06, Inserm U676, France
30 Centre de référence Déficiences Intellectuelles de Causes Rares, Dijon Bourgogne University Hospital, F-21000 Dijon, France
31 Centre de Référence Anomalies du Développement et Syndromes Malformatifs, Dijon Bourgogne University Hospital, F-21000 Dijon, France

Corresponding authors:

Quentin THOMAS, MS
Laboratoire Team GAD, 2 boulevard Maréchal de Lattre de Tassigny, 21 000 Dijon, FRANCE
Phone: 03.80.39.32.38 /Fax: 03.80.29.32.66
Email: quentin.thomas@chu-dijon.fr

Antonio Vitobello, PhD
Laboratoire Team GAD, 2 boulevard Maréchal de Lattre de Tassigny, 21 000 Dijon, FRANCE
Phone: 03.80.39.32.38 /Fax: 03.80.29.32.66
Email: antonio.vitobello@u-bourgogne.fr

Word Count:

Abstract: 193 Article: 3727
Figures: 4
Table: 1
Supplemental Figures and Tables: 6

ABSTRACT

Purpose: ADP Ribosylation Factor Guanine nucleotide Exchange Factors (ARFGEFs) are a family of proteins implicated in cellular trafficking between the Golgi apparatus and the plasma membrane through vesicle formation. Amongst them is ARFGEF1/BIG1, a protein involved in axon elongation, neurite development and polarization processes. *ARFGEF1* has been previously suggested as a candidate gene for different types of epilepsies, although its implication in human disease has not been well characterized.

Methods: International data sharing, *in-silico* predictions and *in-vitro* assays with mini-gene study, western blot analyses and RNA-sequencing.

Results: we identified 13 individuals with heterozygous likely-pathogenic variants in *ARFGEF1*. These individuals displayed congruent clinical features of developmental delay, behavioral problems, abnormal findings on brain MRI, and epilepsy for almost half of them. While nearly half of the cohort carried *de novo* variants, at least 40% of variants were inherited from mildly affected parents who were clinically reevaluated by reverse phenotyping. Our *in-silico* predictions and *in vitro* assays support the contention that *ARFGEF1*-related conditions are caused by haploinsufficiency, and are transmitted in an autosomal dominant fashion with variable expressivity.

Conclusion: We provide evidence that loss-of-function variants in *ARFGEF1* are implicated in sporadic and familial cases of developmental delay with or without epilepsy.

INTRODUCTION

Vesicular trafficking is a critical process in eukaryotic cells. It allows numerous membrane-enclosed compartments to exchange proteins and lipids and is directly involved in countless cellular mechanisms including organelle biogenesis, cellular signaling or membrane dynamics[1]. Its paramount importance is shown by the substantial number of human disorders that derive from the dysfunction of inter-organellar trafficking[2,3]. Amongst those involved in vesicle formation are ADP Ribosylation Factors (ARF) which promote the coating of secretory vesicles in Golgi traffic and their molecular switches: Guanine nucleotide Exchange Factors (GEFs) of the GBF1/BIG family[4]. These proteins are composed of a Sec7 domain responsible for the catalysis of nucleotide exchange on ARFs[5]. Among them is ADP-ribosylation factor guanine nucleotide exchange factor 1 (ARFGEF1), also known as Brefeldin A-inhibited guanine nucleotide-exchange protein 1 (BIG1). ARFGEF1 is a 200 kDa protein encoded by *ARFGEF1* (MIM *604141), a 39-exon gene mapping to the chromosome *8q13* locus. This ubiquitously expressed small GTPase is highly conserved among mammals and eukaryotes[4,6] and has been shown to be implicated in vesicle formation and to be essential for the maintenance of the Golgi apparatus' structure and function[7,8]. Its implication in cell trafficking between the Golgi apparatus and the plasma membrane has been shown to regulate axon elongation, neurite development and maintenance, and the polarization process[9].

After initial studies implicated the *8q13* locus, *ARFGEF1* was identified as the candidate gene within that interval that accounts for rolandic epilepsy[10], familial febrile convulsions[11,12] and epileptic encephalopathy[13,14]. Furthermore, following the discovery of an *ARFGEF1* nonsense variant in a patient diagnosed with Lennox-Gastaut Syndrome (LGS), Teoh and collaborators generated a haploinsufficient mouse model using CRISPR/Cas9 technology. The mice displayed developmental delay, altered cerebral and neuronal morphology and had a high susceptibility to seizures[15]. These findings were congruent with previous investigations in which BIG1-deficient mice showed altered axonal projection, delayed neural polarization, and had a smaller neocortex and hippocampus due to increased neuronal apoptosis[16].

Our data therefore suggest *ARFGEF1* as an interesting candidate gene for developmental delay and epilepsy.

MATERIAL AND METHODS

In silico analysis

ARFGEF1 sequences were all obtained from NCBI RefSeq database as FASTA files and aligned with EMBOSS Clustal Ω tool keeping its default settings[17]. Human and yeast sequences were also pairwise-aligned with the EMBOSS water tool keeping its default settings. Modelling of the structure was performed using the resolved fragment 3LTL available at the RCSB Protein Data Bank (https://www.rcsb.org/). All structures were modelled using PyMol version 2.3.4 (https://pymol.org/). Different scripts were written to generate the structures. After calculation, all were rendered with shaders on and saved as png files. The PyMol "Wizard" and "Protein contact potential" tools were used to generate the putative mutant structures and to render the false red/blue charge-smoothed surface representation respectively.

Splicing reporter minigenes analysis

In this study, we used the reference transcript NM_006421.4.

To evaluate the impact on splicing of the NM_006421.4:c.2392G>A variant, we performed splicing minigene reporter assays using the pCAS2 vector based on a previously described protocol[18]. This functional assay is based on the comparative analysis of the splicing pattern of wild-type and mutant after transfection of the pCAS2 plasmid in human cell lines. Briefly, the exon of interest and its intronic flanking regions were amplified from a DNA sample, using a high-fidelity polymerase (CloneAmp HiFi PCR, Takara). The resulting fragment was introduced into the pCAS2 plasmid according manufacturer's instructions using the In-Fusion HD Cloning kit (CloneAmp HiFi PCR, Takara) and was transformed into the kit's competent cells. Plasmids were purified (NucleoSpin Plasmid Mini kit, Macherey-Nagel) and were Sanger sequenced between the cloning sites. Next, minigenes

representing the wild-type and the mutant allele (1µg/well) were transfected into a human cell line using the FuGENE 6 transfection reagent (Promega). Total RNA was isolated 24 hours after transfection using the NucleoSpin RNA Plus kit (Macherey Nagel) according to the manufacturer's instructions. The concentration of total RNA was measured on a NanoDrop spectrophotometer (NanoDrop Technologies). Reverse transcription was performed with the ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs). PCRs were performed on cDNA with the Platinum DNA Polymerase (Thermo Fisher Scientific) according to the PCR amplicons (PCR conditions and protocol details are available upon request). PCR amplicons were verified by electrophoresis (agarose gel or Caliper LabChip) and Sanger sequenced to evaluate the splicing effect.

Cell cultures, proteins extraction and western blot analysis

Cell cultures, protein extraction and western blot analyses were performed as previously described[19]. Detailed information is available in supplemental methods.

RESULTS

Patient cohort and clinical features

At the time of the first evaluation, individual 1 was a 5-year-old boy with global developmental delay and without any relevant family history. Pregnancy and delivery had been uneventful. He presented with severe language delay and was speaking only a couple of words at the age of 5 years. Clinical examination showed mild gait impairment suggestive of subtle cerebellar ataxia. Brain MRI performed at the age of 6 years revealed a thin corpus callosum. Psychometric evaluations revealed moderate intellectual disability. He had mild facial dysmorphism with a wide mouth, a high forehead and low-set ears. Trio exome sequencing (proband and both parents) identified a *de novo* guanine to adenine transition, lying within a CG dinucleotide, in the *ARFGEF1* gene (NM_006421.4:c.2392G>A), and predicted to result in a missense variant (p.(Asp798Asn)) within the Sec7 domain at a genomic position showing intolerance to amino acid change (MetaDome

non-synonymous over synonymous ratio dN/dS score of 0.22)[20] (Supplemental Figure 1) and associated with deleterious *in silico* scores (Combined Annotation Dependent Depletion - CADD[21]= 34; PolyPhen-2[22]= 1.0; Genomic Evolutionary Rate Profiling - GERP[23]= 4.87).

Through an international collaboration facilitated by GeneMatcher[24], Decipher and RD-connect[25] platforms, we gathered 12 additional individuals from 10 unrelated families with rare likely-pathogenic *ARFGEF1* variants (Supplemental Table 1) identified by exome sequencing. These affected individuals displayed overlapping clinical features including developmental delay, intellectual disability and behavioral problems with or without epilepsy (Supplemental Table 2). Overall, 6/11 variants occurred *de novo* and 3/11 were inherited from an affected or mildly affected father. Interestingly, Sanger sequencing results indicated that paternal variants were germline rather than mosaic (Supplemental Figure 2). This observation was corroborated by the fact that in one family (individuals 8's and 9's family), also a paternal aunt carried the variant. In 2/11 cases, the mode of inheritance could not be established due to lack of parental DNA.

Of note, 2 out of 11 variants had been observed once in the gnomAD cohort: NM_006421.4:c.4033C>T/p.(Arg1345*) and NM_006421.4:c.3697C>T/p.(Gln1233*) but both were absent from healthy gnomAD controls (individuals who were not selected as a case in a case/control study of common disease) and one (p.(Gln1233*)) was absent in the gnomAD non-neurological cohort (individuals who were not ascertained for having a neurological condition in a neurological case/control study).

Interestingly, 12/13 (92%) individuals were males and only 1/13 (8%) was a female (X-squared = 9.3077, df = 1, p-value = 0.002282 - Figure 1A). Twelve individuals were children with a mean age on referral of 8.6 years (Table 1). All individuals were born following uneventful pregnancies and delivered without perinatal complications with normal Apgar scores. Delivery was at term for 12/13 individuals and at 34 weeks of gestation for one individual. Only one individual was born to consanguineous parents. Two children (15%) had a birth weight under the 5th percentile with

normal height and occipital-frontal circumference (OFC) while birth parameters were within normative values for the other 11 individuals.

All individuals had developmental delay of variable degree. First symptoms were noticed at a mean age of 22 months old, and before 3 years of age for all individuals. Almost all individuals (12/13) had predominant language delay, which was often severe, with 7/12 having a vocabulary of only few words at 6 years of age. Gross and fine motor skills were impaired for most children (12/13 and 10/13, respectively) but with slow progressive improvement following adequate rehabilitation. Of note, all children were ambulatory by 3 years of age. Behavioral problems were frequent (12/13 individuals) and often severe with variable occurrence of autism spectrum disorders, anxiety, aggressivity, anger bursts, stereotypies, attention disorders and psychomotor agitation. Ten individuals had intellectual disability (ID) of moderate (4/10) or mild (6/10) degree. None of the children displayed severe ID and only one child had intellectual abilities at the lower limit of normal (borderline Intelligence Quotient).

Associated clinical features included hyperopia (4/13), astigmatism (3/13), frequent otitis media (2/13), strabismus (1/13), occipital meningocele (1/13), interstitial lung disease with bronchiectasis and chronic atelectasis (1/13) and alcohol addiction in the only adult patient (1/13). Most individuals (8/13) displayed facial dysmorphisms that were mostly mild. Some dysmorphic features were shared amongst the individuals including large (5/8) and low-set (4/8) ears, wide mouth (3/8), high forehead (3/8), bulbous nasal tip (3/8) and thin long facies (2/8) (Figure 2). We performed a computer assisted facial visualization[26] from all available photographs in order to generate a "typical" *ARFGEF1* patient's face using the Facer program[27]. These analyses suggested that morphological features were mild in *ARFGEF1*-related disorders and highlighted some of the morphological features identified in the individuals: high forehead, wide mouth and bulbous nasal tip (Figure 2). Most recent available growth parameters were within normal ranges for 9 individuals (69%) with 3 displaying macrocephaly (between +2.1 and 2.5 SD deviation) and 1 displaying a small head circumference (-3.4 SD).

Neurological examinations were mostly unremarkable or limited to signs related to motor-development delay while cerebellar ataxia was noted in two individuals and hypotonia in another two individuals. Of note, one patient displayed an action tremor and another presented with clumsiness and a balance disorder, raising the possibility of a mild underlying cerebellar ataxia in these two additional individuals. One patient presented with left laryngeal paralysis and limited left eye abduction. Six individuals were followed for epilepsy, while a seventh individual was treated with valproic acid but could not specify whether it was for neurologic or psychiatric reasons. Many seizure types were observed including tonic-clonic seizures (4/6), atonic seizures (2/6), myoclonic seizures (2/6), focal motor seizures (1/6) and typical absence seizures (1/6). Three patients were suspected to have Lennox-Gastaut Syndrome (one with generalized drug-resistant epilepsy with slow spike and wave on EEG, and two others who had generalized tonic-clonic seizures along with myoclonic and atonic seizures but without drug-resistance and for whom access to EEG was not possible) and one had childhood absence epilepsy. Brain MRIs were available for 10 individuals and revealed abnormalities in eight of them including white matter T2-weighted signal hyperintensity compatible with hypomyelination (3/8), type 1 Arnold-Chiari malformation (1/8), low-set cerebellar tonsils (1/8), thin corpus callosum (1/8) and meningocele with periventricular heterotopia and disrupted superior cerebellar vermis (1/8) (Figure 3).

Family history was unremarkable in seven individuals (54%), while six (46%) had at least one relative with a history of developmental delay, learning disorder, behavioral problems, or milder phenotypes with only difficulties at school requiring special education or classroom accommodations (Figure 1). Furthermore, almost half (6/13) of the individuals inherited the variant from a parent who either presented with a milder version of their child's phenotype, or for whom retrospective phenotyping revealed a history of difficulties at school, learning disorder or a need for special education. For example, patient 7 is a child with severe drug-resistant epilepsy. Detailed clinical characterization following his molecular diagnosis revealed that his father, from whom he inherited the variant, had a self-limited childhood onset epilepsy. In individuals 8 and 9, revisiting the family

history revealed that their father had global developmental delay, with marked behavioral problems (aggressivity and temper tantrums) resulting in difficulties at school. For individuals 10 and 11, further questioning revealed that their father had mild developmental delay and borderline intellect. Additionally, two of individuals 8's and 9's paternal cousins displayed prominent difficulties at school, behavioral problems and delayed development requiring speech therapy, however none of them had genetic testing as no diagnostic procedure was initiated. Yet, the familial segregation study revealed that the mother of these two relatives (individuals 8's and 9's paternal aunt) carried the variant and displayed intellectual abilities at the lower limit of normal. Thus, we suspect that these two cousins also carry the familial *ARFGEF1* variant, although their biological samples were not available for confirmation.

Molecular findings and functional studies

Our cohort consisted of 9 loss-of-function (LoF) variants (6 stop-gain and 3 frameshift), 1 variant impacting a splice site (Supplemental Table 1, Figure 1) and 1 missense variant. Given that all of the loss-of-function variants identified in our cohort introduced premature stop codons before the last exon or upstream to the 50-55 nucleotides preceding the last exon-intron junction, all LoF variants were predicted to result in nonsense-mediated mRNA decay (NMD)[28] rendering the locus haploinsufficient.

ARFGEF1 is a gene under high mutational constraint (Supplemental Figure 1). The gnomAD missense Z score was 5.37, with two regional missense constraint regions corresponding to amino acid positions 482-1301 and 1302-1850 with a calculated score of 0.37 (p-Value = 3.39×10^{-21}) and 0.67 (p-Value = 2.38×10^{-5}) respectively (source Decipher https://decipher.sanger.ac.uk/, UniProt[29] accession Q9Y6D6 https://www.uniprot.org). The probability of loss-of-function intolerance (pLI) score is 1, with an observed/expected (oe) ratio equal to 0.08 and an upper bound confidence interval equal to 0.14 (source gnomAD[30] v2.1.1, https://gnomad.broadinstitute.org/).

These observations suggested that the mechanism responsible for *ARFGEF1*-related disorders is haploinsufficiency. However, a functional assessment of the impact of the only missense variant identified in our cohort on pre-mRNA splicing or protein stability was required to corroborate our hypothesis. Splicing prediction using the Alamut Visual v.2.15 software predicted that no obvious splicing defect results from the c.2392G>A transition. However, this variant was predicted to create a new SRp40 Exonic Splice Enhancer (ESE) (Supplemental Figure 3). We thus performed functional assays in order to investigate the experimentally observed functional effect of the c.2392G>A variant.

To that end, we performed a minigene reporter assay using the pCAS2 vector as previously described (Supplemental Figure 4)[18]. This assay allows a rapid screening of aberrant splicing events driven by DNA variants, cloned within a reporter vector containing two exons (named A and B) derived from the human *SERPING1/C1NH* gene, separated by an intron with BamHI and MluI cloning sites, and under the CMV-promoter-driven transcriptional control. This assay showed that the c.2392A allele induced exon skipping in approximately 22% of the reporter transcripts, corresponding to a 2.4-fold increase as compared to the clones carrying the wild-type c.2392G allele associated with 9% baseline skipped transcripts (Supplemental Figure 4) suggesting that a splicing effect of this variant was unlikely. Then, we obtained a fibroblast cell line from patient 1 and performed targeted splicing analysis, using next-generation sequencing, on a PCR-derived amplicon obtained from retrotranscribed RNA. These analyses allowed us to exclude any possible splicing defect associated with c.2392G>A *in vitro* (Supplemental Figure 5). Of note, these experiments also revealed that the variant allelic frequency (VAF) of c.2392G>A in retrotranscribed RNA was close to 50%, indicating a biallelic expression of *ARFGEF1* in fibroblasts (Supplemental Figure 6). Using the same approach, we also investigated the splicing defect associated with the *de novo* c.3592-2A>G variant in a blood sample obtained from patient 3. As expected, the variant triggered exon 26 skipping, inducing the loss of 152 nucleotides and resulting in NMD (Supplemental Figure 7).

Structurally, the polypeptide surrounding the Asp798 residue is conserved across species (Figure 4). *In silico* three-dimensional modelling of the Asp798 residue within the Sec7 domain, based

on the human resolved fragment (PDB 3LTL), allowed us to predict the corresponding steric changes induced by the Asp798Asn variant (Figure 4A) in support of a possible impact on protein misfolding or catalytic activity.

Using an ARFGEF1-specific antibody, we then performed Western Blot analyses on protein extracts obtained from fibroblast cell lines obtained from Patient 1 and the symptomatic father of Individuals 2 and 3 carrying the variant NM_006421.4:c.2392G>A; p.(Gln648*), and unrelated and healthy control individuals (Figure 4B). Interestingly, both cell lines derived from affected individuals revealed comparable reduced levels of protein expression when compared with healthy controls, indicating that c.2392G>A (p.(Asp798Asn)) resulted in protein instability and degradation.

Taken together, these data suggest that the pathogenic effect of these variants is compatible with haploinsufficiency.

DISCUSSION

Although high-density SNP array[10], linkage analysis[11] and exome sequencing meta-analysis[14] had previously linked the *ARFGEF1* locus to increased risk for rolandic epilepsy[10] and epileptic encephalopathy[14], there had been no substantial evidence to support an association between *ARFGEF1* and high penetrance of a neurodevelopmental disorder conforming to Mendelian inheritance expectations. More recently, one patient with Lennox-Gastaut syndrome harboring a truncating variant (p.Cys1455*) in *ARFGEF1* has been reported by Teoh and collaborators[15], requiring additional observations in order to confirm its implication as a disease-causing gene. Here, we describe a cohort of 13 affected individuals gathered through an international collaboration that harbor likely-pathogenic variants in *ARFGEF1* identified via exome sequencing.

Interestingly, while almost half (6/13) of the individuals were followed for epilepsy, all of them displayed developmental delay with impaired motor skills, speech delay, or both (Table 1). When analyzing the cohort as a whole, features of this disorder including mild motor delay, mild or moderate intellectual disability, severe speech delay and numerous behavioral problems such as

attention disorders, psychomotor agitation, anxiety or autism spectrum disorders. Regarding epilepsy, 5 different seizure types were observed in our cohort, ranging from focal motor seizures to generalized tonic-clonic seizures and although EEGs were unavailable for some individuals, two patients were suspected to have Lennox-Gastaut Syndrome (LGS) and a third one was formally diagnosed with LGS (EEG available) leading to up to half of the individuals with epilepsy to display a LGS phenotype. *ARFGEF1* thus appears as a possible gene contributing to Lennox-Gastaut syndrome.

This increased susceptibility to seizures may be related to ARFGEF1's involvement in cellular trafficking of GABAA receptors. Indeed, previous work has shown that ARFGEF1 is a binding partner of GABAA receptors, and that depletion of ARFGEF1, either through siRNA[31] or haploinsufficient mouse models[15], led to a decrease in GABAA receptors at the neuronal surface, which in turn might lead to impaired neuronal inhibition and higher seizure susceptibility. This point may prove critical in the future management of epileptic patients carrying pathogenic variants in *ARFGEF1* and will require special attention from physicians when using GABA-based drugs, whose efficiency might be reduced in such individuals.

As shown by our genetic and functional data, the pathophysiology of variants in *ARFGEF1* is consistent with haploinsufficiency. This is consistent with previously published functional studies, particularly in yeast, which showed that all protein domains are critical for the overall functioning of the protein[32]. Based on the molecular data from our study, the variable clinical features and clinical severity observed in different individuals in our cohort does not correlate with the localization of the variants along the gene. However, identification of further individuals carrying deleterious variants might bring additional information about the clinical spectrum associated with *ARFGEF1*-related dysfunction.

Family pedigrees and histories show *ARFGEF1*-related disorders display an autosomal dominant mode of inheritance (Figure 1A). As no rare copy number nor additional likely-pathogenic variants were identified in our cohort to account for inter-generational variability, variants in this gene seems to be subject to highly variable expressivity. Indeed, as explained above, one third of the

children had inherited the variant from an apparently healthy parent for whom the question of a genetic neurodevelopmental disorder had not been raised during the initial assessment but only once the *ARFGEF1* variant had been identified as part as the familial segregation analysis (reverse phenotyping, Figure 1). This point is not particularly surprising as mild difficulties at school or in learning in general can sometimes be tolerated. The situation of individuals 8's and 9's cousins was particularly illustrative of this and also shows how two individuals carrying expected-pathogenic variants in *ARFGEF1* can be found in gnomAD, including one in the non-neurological cohort.

Regarding the findings on brain MRI, the most common abnormality was a delayed myelination shown by subcortical white matter signal hyperintensity. This is consistent with what was observed in mice which displayed altered neurodevelopment with delayed neural polarization and increased neural apoptosis[15,16]. One individual - patient 3 with the focal motor seizures- displayed periventricular nodular heterotopies (PH), a feature that echoes the brain abnormalities observed in a Turkish family with *ARFGEF2* (Periventricular heterotopia with microcephaly – OMIM 608097 - Autosomal Recessive) pathogenic variants who displayed severe developmental delay, microcephaly, early-onset refractory epilepsy, bilateral nodular periventricular heterotopia and frequent infections[33], and a phenotype that partially overlaps the one observed in our cohort. Two hypotheses have been put forward to explain PH in *ARFGEF2*-related disorders. First, BIG2 inhibition has been shown to increase phosphorylation of Filamin A (a protein well known to be associated with PH[34]), thus disrupting cell intrinsic neuronal migration which favors PH[35] . Second, BIG2 has been shown to be involved in neuron migration through i) regulation of actin dynamics and ii) its critical importance for vesicle and membrane trafficking, mechanisms that are fundamental regulators of proliferation and migration during human cerebral cortical development[35,36]. We may thus hypothesize that *ARFGEF1* pathogenic variants could also generate PH as BIG1 has also been shown to be involved in intracellular vesicle trafficking[7], axon elongation[9] and actin dynamics/directed migration[36].

Surprisingly, our cohort mostly consisted of male individuals (12/13) and the ascertained inherited variants were inherited from mildly symptomatic fathers who carried Sanger-confirmed heterozygous variants. This appears surprising as no clear difference is observed between male and female individuals regarding tissue expression of ARFGEF1 according to databases such as GTEx[37] or the human protein atlas[38]. Furthermore, our RNA and protein expression data show that *ARFGEF1* expression is biallelic in clinically accessible tissues (i.e. skin-derived fibroblasts and whole blood) although we cannot exclude that it might be selectively imprinted in other tissues during development. Thus, no clear biological explanation appears to account for this unbalanced sex ratio, although members of the ARF family have been implicated in asymmetrical cell division in female meiosis in the mouse[39,40] and potentially provide an explanation for this observation. Further studies and reports should help untangle whether a real sex-dependent incidence of the disorders exists or whether this happened in our cohort simply by chance.

To conclude, we describe a cohort of 13 individuals with *de novo* or inherited likely-pathogenic *ARFGEF1* variants affected with developmental disorders of varying degree, offering a description of clinical phenotypes associated with this gene. We provide evidence that *ARFGEF1* should be considered as a new gene responsible for intellectual disability, developmental delay and syndromic epilepsy.

DATA AVAILABILITY

Data is available upon request.

ACKNOWLEDGMENTS

We thank the families and patients for taking part in the study. We thank the University of Burgundy Centre de Calcul (CCuB) for technical support and management of the informatics platform, and the Genematcher platform for data sharing. We thank the Centre de Ressources Biologiques Ferdinand Cabanne (CHU Dijon) for sample biobanking. This work was supported by grants from Dijon

University Hospital, the ISITE-BFC (PIA ANR) and the European Union through the FEDER programs.

Also supported in part by the US National Institutes of Health, National Human Genome Research Institute (NHGRI) to the Baylor Hopkins Center for Mendelian Genomics (UM1HG006542).

This work was supported by the National Institute for Health Research (NIHR) Manchester Biomedical Research Centre. The Deciphering Developmental Disorders (DDD) study presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009–003).

This study makes use of DECIPHER (http://decipher.sanger.ac.uk), which is funded by the Wellcome.

See *Nature* PMID 25533962 or www.ddduk.org/access.html for full acknowledgement.

D.M. is also supported by a Medical Genetics Research Fellowship Program through the United States National Institute of Health (T32 GM007526-42).

Several authors of this publication are members of the European Reference Network for Developmental Anomalies and Intellectual Disability (ERN-ITHACA) ". A.J., A-S.D-P. and A.V. are supported by Solve-RD.

The Solve-RD project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 779257.

J.E.P. was supported by NHGRI K08 HG008986.

AUTHOR INFORMATION

Conceptualization: A.V., C.T-R, L.F., L.D.; Data curation: Q.T., A.V., Y.D., P.G.; Investigation: Q.T., D.M., M.W., S.M., B.I., B.C., S.C., R.T., M.I., A.S., A.M., T.D., A.J., S.B., J.D., N.S., F.A., F.AZ., P.A., E.E., J.M. L.B., D.R., F.TM-T. A-L.B., P.C., N.M., A-S.D-P., C.P.; Methodology: A.V., L.D., M.C., S.N., T.B., T.G., J.G.; Supervision: Q.T., A.V., D.M.; Visualization: T.G., Q.T., A.V, Writing - original draft: Q.T., A.V., L.F., C.T-R; Writing – review & editing: all authors.

ETHICS DECLARATION

All affected individuals or their legal representative gave their informed consent for the sequencing procedures and the publication of their results alongside with clinical and molecular data. Special consent forms were signed authorizing publication of pictures when relevant.

The study was performed within the framework of the GAD ("Génétique des Anomalies du Développement") collection and approved by the appropriate institutional review board of Dijon University Hospital (DC2011-1332).

REFERENCES

1. Muro S. Alterations in cellular processes involving vesicular trafficking and implications in drug delivery. *Biomimetics*. 2018;3(3). doi:10.3390/biomimetics3030019
2. Aridor M, Hannan LA. Traffic Jams II: An update of diseases of intracellular transport. *Traffic*. 2002;3(11):781-790. doi:10.1034/j.1600-0854.2002.31103.x
3. Aridor M. Visiting the ER: The endoplasmic reticulum as a target for therapeutics in traffic related diseases. *Adv Drug Deliv Rev*. 2007;59(8):759-781. doi:10.1016/j.addr.2007.06.002
4. Wright J, Kahn RA, Sztul E. Regulating the large Sec7 ARF guanine nucleotide exchange factors: The when, where and how of activation. *Cell Mol Life Sci*. 2014;71(18):3419-3438. doi:10.1007/s00018-014-1602-7
5. Cherfils J, Ménétrey J, Mathieu M, et al. Structure of the Sec 7 domain of the Arf exchange factor ARNO. *Nature*. 1998;392(6671):101-105. doi:10.1038/32210
6. Mansour SJ, Skaug J, Zhao XH, Giordano J, Scherer SW, Melancon P. p200 ARF-GEP1: A Golgi-localized guanine nucleotide exchange protein whose Sec7 domain is targeted by the drug brefeldin A. *Proc Natl Acad Sci U S A*. 1999;96(14):7968-7973. doi:10.1073/pnas.96.14.7968
7. Boal F, Stephens DJ. Specific functions of BIG1 and BIG2 in endomembrane organization. *PLoS One*. 2010;5(3). doi:10.1371/journal.pone.0009898
8. Zhao X, Lasell TKR, Melançon P. Localization of large ADP-ribosylation factor-guanine nucleotide exchange factors to different Golgi compartments: evidence for distinct functions in protein traffic. *Mol Biol Cell*. 2002;13(1):119-133. doi:10.1091/mbc.01-08-0420
9. Zhou C, Li C, Li D, et al. BIG1, a brefeldin A-inhibited guanine nucleotide-exchange protein regulates neurite development via PI3K-AKT and ERK signaling pathways. *Neuroscience*. 2013;254:361-368. doi:10.1016/j.neuroscience.2013.09.045
10. Addis L, Sproviero W, Thomas S V., et al. Identification of new risk factors for rolandic epilepsy: CNV at Xp22.31 and alterations at cholinergic synapses. *J Med Genet*. 2018;55(9):607-616. doi:10.1136/jmedgenet-2018-105319
11. Wallace RH, Berkovic SF, Howell RA, Sutherland GR, Mulley JC. Suggestion of a major gene for familial febrile convulsions mapping to 8q 13-21. *J Med Genet*. 1996;33(4):308-312. doi:10.1136/jmg.33.4.308
12. Piro RM, Molineris I, Ala U, Di Cunto F. Evaluation of candidate genes from orphan FEB and GEFS+ loci by analysis of human brain gene expression atlases. *PLoS One*. 2011;6(8). doi:10.1371/journal.pone.0023149
13. Appenzeller S, Balling R, Barisic N, et al. De novo mutations in synaptic transmission genes including DNM1 cause epileptic encephalopathies. *Am J Hum Genet*. 2014;95(4):360-370. doi:10.1016/j.ajhg.2014.08.013
14. Takata A, Nakashima M, Saitsu H, et al. Comprehensive analysis of coding variants highlights genetic complexity in developmental and epileptic encephalopathy. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-10482-9
15. Teoh JJ, Subramanian N, Pero ME, et al. Arfgef1 haploinsufficiency in mice alters neuronal endosome composition and decreases membrane surface postsynaptic GABAA receptors. *Neurobiol Dis*. 2020;134(September 2019):104632. doi:10.1016/j.nbd.2019.104632
16. Teoh JJ, Iwano T, Kunii M, et al. BIG1 is required for the survival of deep layer neurons, neuronal polarity, and the formation of axonal tracts between the thalamus and neocortex in developing brain. *PLoS One*. 2017;12(4):1-24. doi:10.1371/journal.pone.0175888
17. Madeira F, Park YM, Lee J, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 2019;47(W1):W636-W641. doi:10.1093/nar/gkz268
18. Soukarieh O, Gaildrat P, Hamieh M, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet*. 2016;12(1):1-26. doi:10.1371/journal.pgen.1005756
19. Da Costa R, De Almeida S, Chevarin M, et al. Neutralization of HSF1 in cells from PIK3CA-related overgrowth spectrum patients blocks abnormal proliferation. *Biochem Biophys Res*

Commun. 2020;(xxxx):1-7. doi:10.1016/j.bbrc.2020.04.146
20. Wiel L, Baakman C, Gilissen D, Veltman JA, Vriend G, Gilissen C. MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum Mutat*. 2019;40(8):1030-1038. doi:10.1002/humu.23798
21. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894. doi:10.1093/nar/gky1016
22. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;(SUPPL.76). doi:10.1002/0471142905.hg0720s76
23. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15(7):901-913. doi:10.1101/gr.3577405
24. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. *Hum Mutat*. 2015;36(10):928-930. doi:10.1002/humu.22844
25. Lochmüller H, Badowska DM, Thompson R, et al. RD-Connect, NeurOmics and EURenOmics: Collaborative European initiative for rare diseases. *Eur J Hum Genet*. 2018;26(6):778-785. doi:10.1038/s41431-018-0115-5
26. Ferry Q, Steinberg J, Webber C, et al. Diagnostically relevant facial gestalt information from ordinary photos. *Elife*. 2014;2014(3):2020. doi:10.7554/eLife.02020.001
27. GitHub - johnwmillr/Facer: Simple face averaging in Python. https://github.com/johnwmillr/Facer. Accessed August 5, 2020.
28. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem Sci*. 1998;23(6):198-199. doi:10.1016/S0968-0004(98)01208-0
29. Bateman A, Martin MJ, O'Donovan C, et al. UniProt: A hub for protein information. *Nucleic Acids Res*. 2015;43(D1):D204-D212. doi:10.1093/nar/gku989
30. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7
31. Li C, Chen S, Yu Y, et al. BIG1, a brefeldin A-inhibited guanine nucleotide-exchange factor, is required for GABA-gated Cl- influx through regulation of GABA A receptor trafficking. *Mol Neurobiol*. 2014;49(2):808-819. doi:10.1007/s12035-013-8558-8
32. Ramaen O, Joubert A, Simister P, et al. Interactions between conserved domains within homodimers in the BIG1, BIG2, and GBF1 Arf guanine nucleotide exchange factors. *J Biol Chem*. 2007;282(39):28834-28842. doi:10.1074/jbc.M705525200
33. Sheen VL, Ganesh VS, Topcu M, et al. Mutations in ARFGEF2 implicate vesicle trafficking in neural progenitor proliferation and migration in the human cerebral cortex. *Nat Genet*. 2004;36(1):69-76. doi:10.1038/ng1276
34. Liu W, Yan B, An D, Xiao J, Hu F, Zhou D. Sporadic periventricular nodular heterotopia: Classification, phenotype and correlation with Filamin A mutations. *Epilepsy Res*. 2017;133:33-40. doi:10.1016/j.eplepsyres.2017.03.005
35. Zhang J, Neal J, Lian G, Shi B, Ferland RJ, Sheen V. Brefeldin A-inhibited guanine exchange factor 2 regulates Filamin a phosphorylation and neuronal migration. *J Neurosci*. 2012;32(36):12619-12629. doi:10.1523/JNEUROSCI.1063-12.2012
36. Le K, Li CC, Ye G, Moss J, Vaughan M. Arf guanine nucleotide-exchange factors BIG1 and BIG2 regulate nonmuscle myosin IIA activity by anchoring myosin phosphatase complex. *Proc Natl Acad Sci U S A*. 2013;110(34). doi:10.1073/pnas.1312531110
37. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653
38. Uhlen M, Fagerberg L, Hallstrom BM, et al. Tissue-based map of the human proteome. *Science*

(80-). 2015;347(6220):1260419-1260419. doi:10.1126/science.1260419

39. Duan X, Zhang HL, Pan MH, Zhang Y, Sun SC. Vesicular transport protein Arf6 modulates cytoskeleton dynamics for polar body extrusion in mouse oocyte meiosis. *Biochim Biophys Acta - Mol Cell Res*. 2018;1865(2):455-462. doi:10.1016/j.bbamcr.2017.11.016
40. Wang S, Hu J, Guo X, Liu JX, Gao S. ADP-ribosylation factor 1 regulates asymmetric cell division in female meiosis in the mouse1. *Biol Reprod*. 2009;80(3):555-562. doi:10.1095/biolreprod.108.073197

LEGENDS

Figure 1: Family trees and visual representation of *ARFGEF1* variants on a protein model.

Table 1: Clinical features and variants found in the 13 individuals of the cohort

Figure 2: Pictures of affected individuals carrying *ARFGEF1* expected-pathogenic variants

Figure 3: Brain MRI

Figure 4: In silico analyses regarding the missense p.Asp798Asn and Western Blot analyses

Haploinsufficiency of *ARFGEF1* is associated with developmental delay, intellectual disability and epilepsy with variable expressivity

Quentin Thomas, MS, Thierry Gautier, PhD, Dana Marafi, MD MSc, Thomas Besnard, PhD, Marjolaine Willems, MD, Sébastien Moutton, MD PhD, Bertand Isidor, MD PhD, Benjamin Cogné, PharmD PhD, Solène Conrad, MS, Romano Tenconi, MD PhD, Maria Iascone, MD, Arthur Sorlin, MD PhD, Alice Masurel, MD, Tabib Dabir, MD, Adam Jackson, MBChB MSc MRCP, Siddharth Banka, MBBS MRCPCH PhD, Julian Delanne, MD, James R. Lupski, MD PhD, Nebal Waill Saadi, MD, Fowzan S. Alkuraya, MD, Fatema Al Zahrani, MD, Pankaj Agrawal, MD, Eleina England, MS, Jill A. Madden, PhD MSc CGC, Jennifer E. Posey, MD, Lydie Burglen, MD PhD, Diana Rodriguez, MD PhD, Martin Chevarin, BS, Sylvie Nguyen, BS, Frédéric Tran Mau-Them, MD PhD, Yannis Duffourd, MSc, Philippine Garret, MSc PhD, Ange-Line Bruel, PhD, Patrick Callier, PharmD PhD, Nathalie Marle, MD PhD, Anne-Sophie Denomme-Pichon, MD, Laurence Duplomb, PhD, Christophe Philippe, MD PhD, Christel Thauvin-Robinet, MD PhD, Jérôme Govin, PhD, Laurence Faivre, MD PhD and Antonio Vitobello, PhD.

1. Disclosure: The authors declare no conflict of interest related to this study

2. Disclosure James R. Lupski has stock ownership in 23andMe, is a paid consultant for Regeneron Genetics Center, and is a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting.

3. Disclosure: The Department of Molecular and Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing conducted at Baylor Genetics (BG) Laboratories.

4. Disclosure: This work was supported by grants from Dijon University Hospital, the ISITE-BFC (PIA ANR) and the European Union through the FEDER programs. Also supported in part by the US National Institutes of Health, National Human Genome Research Institute (NHGRI) to the Baylor Hopkins Center for Mendelian Genomics (UM1HG006542).

5. Disclosure: This work was supported by the National Institute for Health Research (NIHR) Manchester Biomedical Research Centre. The Deciphering Developmental Disorders (DDD) study presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009–003).

6. Disclosure: Dana Marafi is also supported by a Medical Genetics Research Fellowship Program through the United States National Institute of Health (T32 GM007526-42).

7. Disclosure: Several authors of this publication are members of the European Reference Network for Developmental Anomalies and Intellectual Disability (ERN-ITHACA) “.

8. Disclosure: Adam Jackson is supported by Solve-RD.

9. Disclosure: The Solve-RD project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 779257.

10. Jennifer E. Posey was supported by NHGRI K08 HG008986.

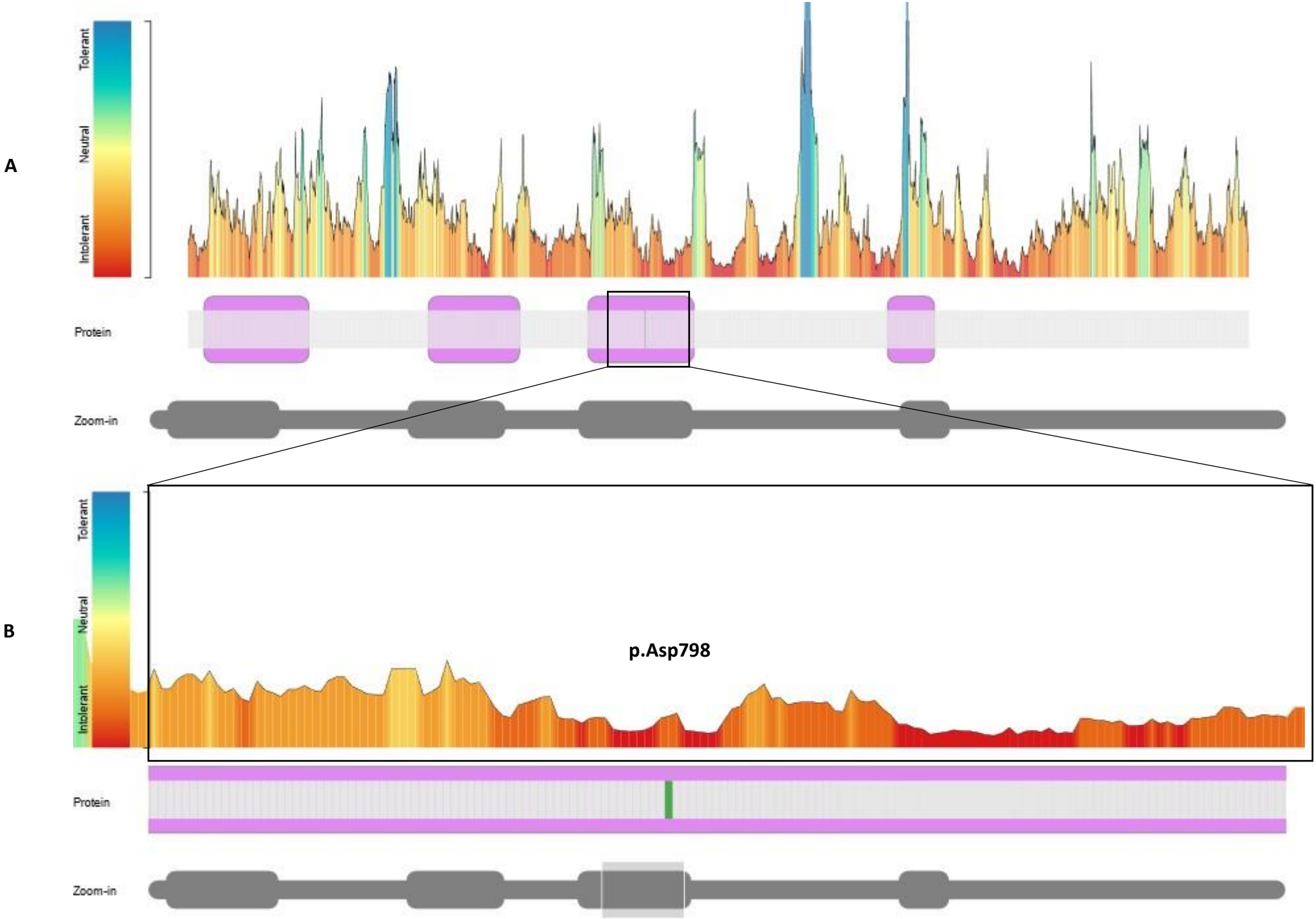
Table 1 - final

	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6	Individual 7	Individual 8	Individual 9	Individual 10	Individual 11	Individual 12	Individual 13	Total
Age on follow-up	6 years old	10 years old	14 years old	11 years old	5 years old	3 months old	9 years old	11 years old	10 years old	13 years old	11 years old	32 years old	10 years old	
Gender	M	M	M	M	M	M	M	M	F	M	M	M	M	9M/1F
Motor delay	+	+	+	+	+	+	+	+	-	+	+	+	+	12/13
Speech Delay	+	+	+	+	-	+	+	+	+	+	+	+	+	12/13
Degree of delay	severe	moderate	severe	severe	-	mild	severe	mild	severe	mild	severe	mild	severe	
Behavioral problems	+	+	-	+	+	+	+	+	+	+	+	+	+	12/13
Type of disorder	psychomotor agitation	autism spectrum disorders		aggressivity, temper tantrums	hyperactivity, aggressivity, obsessive compulsive behaviors	ADHD	temper tantrums, anger bursts, severe psychomotor agitation, distractibility, attention disorders, enuresis, encopresis	autism disorders, anxiety, aggressivity	anxiety, attention disorders,	autism spectrum disorder, hyperactivity	attention disorder	aggressive behavior, alcohol addiction	oppositional defiant disorder, aggressivity, ADHD	
Intellectual Disability (ID)	+	+	-	+	-	+	+	+	+	+	-	+	-	10/13
Degree of ID	moderate	mild	intellectual functions at the lower limit	moderate	-	mild	moderate	mild	mild	moderate	-	mild	mild	
Neurological features	cerebellar ataxia	-	Impaired fine motor skills, slight balance disorder, left laryngeal paralysis; limitation of abduction of the left eye with strabismus	-	hypotonia	-	impaired fine and gross motor skills, balance disorders, dysarthria	impaired fine motor skills mild cerebellar ataxia, dysarthria	action tremor	-	impaired fine motor skills	-	-	7/13
Neurosensory disorders	-	-	-	-	-	hyperopia and high astigmatism	-	hyperopia	hyperopia, astigmatism	-	astigmatism, hyperopia	strabismus	mild unilateral hearing impairment	6/13
MRI findings	thin corpus callosum	mild myelination delay	occipital meningocele, disrupted superior cerebellar vermis, bilateral asymmetrical nodular heterotopias	-	-	-	pineal cyst, low-set cerebellar tonsils	subcortical white matter T2 signal hyperintensity congruent with myelination delay	subcortical white matter T2 signal hyperintensity congruent with myelination delay	UN	UN	UN	Type 1 Arnold-Chiari malformation	6/13

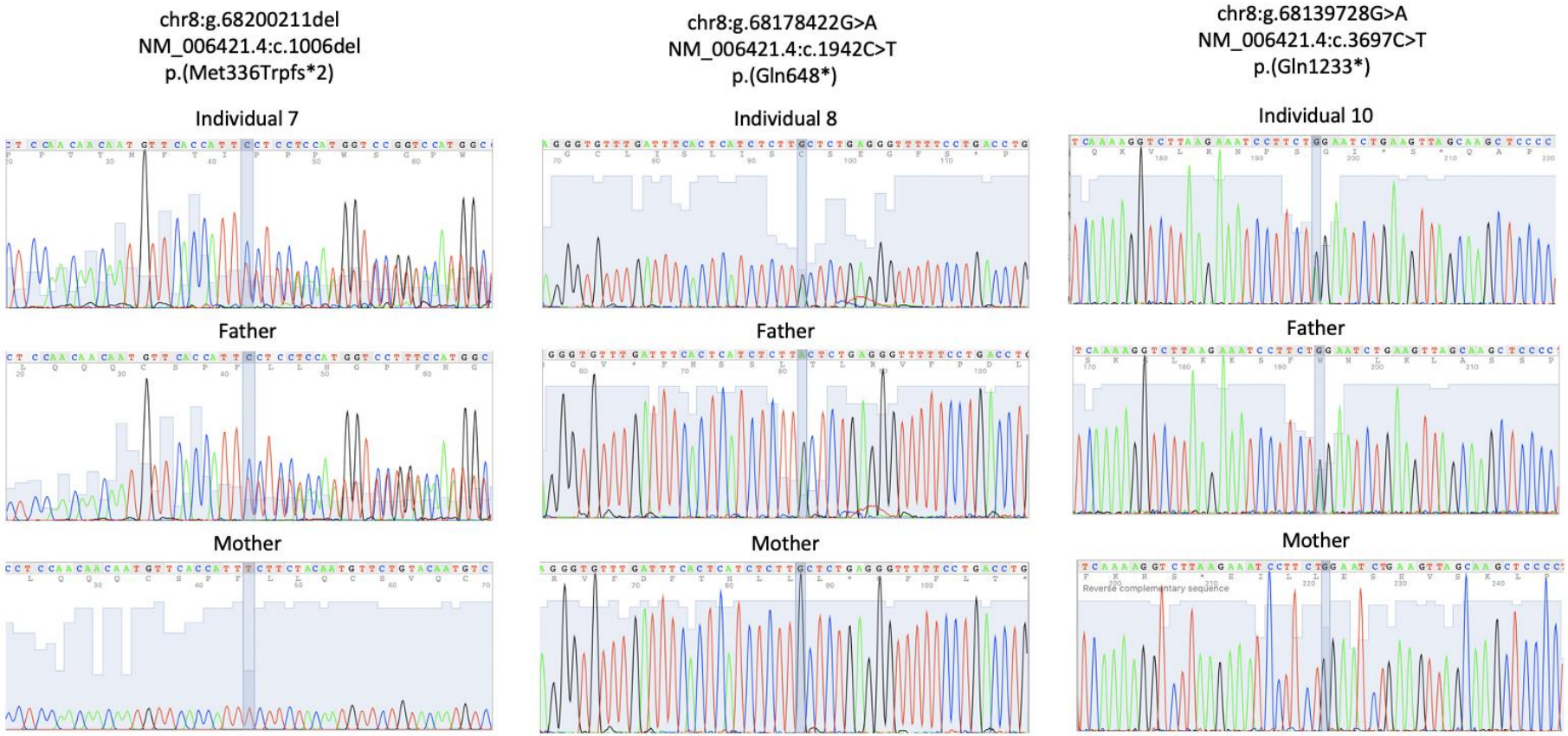
Table 1: Clinical features and variants found in the 13 individuals of the cohort. UN: unavailable

Feature	Individual 1	Individual 2	Individual 3	Individual 4	Individual 5	Individual 6	Individual 7	Individual 8	Individual 9	Individual 10	Individual 11	Individual 12	Individual 13	Total
Dysmorphic features	wide mouth, high forehead, low-set ears	frontal bossing, triangular face, facial hypertrichosis, pointed frontal hairline, thick eyebrows, short philtrum, protruding incisor teeth, generalized hirsutism, long face, wide mouth, low-set large ears	-	-	-	UN	hypotonic long face, wide mouth, large ears, bulbous nose tip	high forehead, large ears	high forehead, bulbous nose tip	brachycephaly, long and thin fingers, plagiocephaly, low-set large ears	long and thin fingers, large ears	small and low-set ears, long and thin fingers	-	8/13
Epilepsy	-	-	focal motor seizures	febrile seizures, generalized tonic-clonic seizures, atonic seizures, myoclonic and absence seizures. Drug-resistant epilepsy	Generalized tonic-clonic seizures, febrile seizures, myoclonic, atonic. Drug-resistant epilepsy	generalized tonic-clonic	generalized tonic-clonic seizures, febrile seizures, myoclonic, atonic	-	-	-	-	-	absence seizures	6/13
Exome Sequencing strategy	trio	trio	trio	trio	trio	solo	trio	solo	solo	trio	trio	solo	duo (mother)	
GrCh37/Hg19 genomic variants	chr8:g.68170369C>T	chr8:g.68112696G>A	chr8:g.68139835T>C	chr8:g.68172127del	chr8:68169969	chr8: 68152451	chr8:g.68200211delT	chr8:g.68178422G>A	chr8:g.68178422G>A	chr8:g.68139728G>A	chr8:g.68139728G>A	chr8:g.68138302G>A	chr8:68170366G>A	
cDNA variants.	NM_006421.4:c.2392G>A	NM_006421.4:c.5320C>T	NM_006421.4:c.3592-2A>G	NM_006421.4:c.2158del	NM_006421:c.C2524T	NM_006421.4:c.2923_2924dup	NM_006421.4:c.1006delA	NM_006421.4:c.1942C>T	NM_006421.4:c.1942C>T	NM_006421.4:c.3697C>T	NM_006421.4:c.3697C>T	NM_006421.4:c.4033C>T	NM_006421.5 c.2395C>T	
aminoacid variants	p.(Asp798Asn)	p.(Arg1774*)	p.?	p.(Leu720Serfs*24)	p.(Gln842*)	p.(Cys976Phefs*6)	p.(Met336Trpfs*2)	p.(Gln648*)	p.(Gln648*)	p.(Gln1233*)	p.(Gln1233*)	p.(Arg1345*)	p.(Arg799*)	
Familial segregation	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	paternally inherited	paternally inherited	paternally inherited	paternally inherited	paternally inherited	UN	not inherited from mother, paternal sample unavailable	

Table 1 continuation: Clinical features and variants found in the 13 individuals of the cohort.



Supplemental Figure 1: ***Visualization of protein intolerance to amino acid change according to MetaDome****. A: visualization of the overall protein (transcript NM_006421.4) intolerance to amino acid change. B: focus on the protein position p.Asp798. showing high intolerance to amino acid change.*



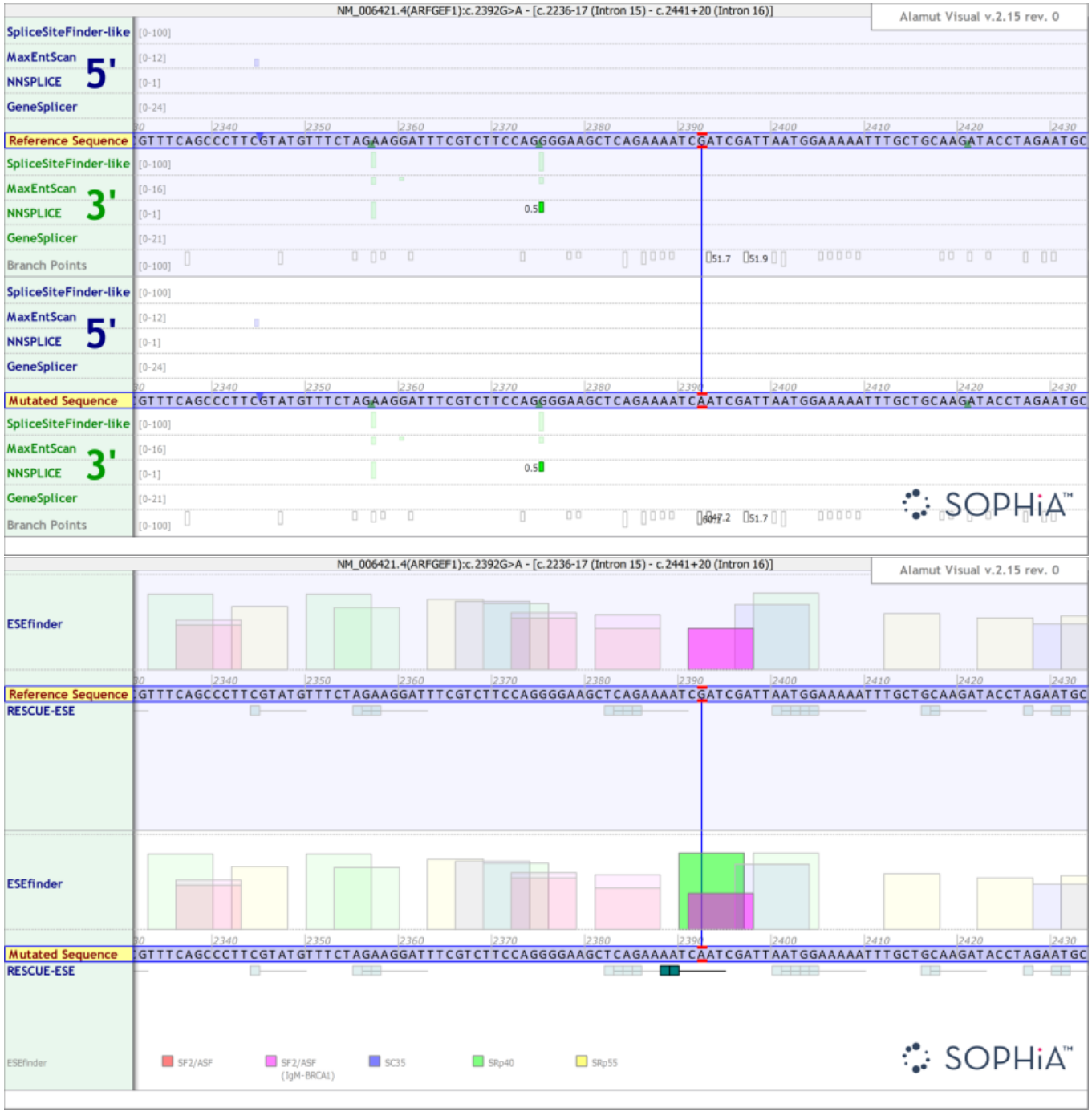
Supplemental Figure 2: Sanger Sequencing results of inherited variants. Causative variants were inherited from mildly affected fathers. Variant allele frequencies in fathers and affected children were close to 50%.

Genomic variant	Transcript variant	Protein variant	Familial segregation	ES Strategy	GnomAD	Number of individuals
chr8(GrCh37):g.68170369C>T	NM_006421.4:c.2392G>A	p.(Asp798Asn)	de novo	Trio	0	1
chr8(GrCh37):g.68112696G>A	NM_006421.4:c.5320C>T	p.(Arg1774*)	de novo	Trio	0	1
chr8(GrCh37):g.68139835T>C	NM_006421.4:c.3592-2A>G	p.?	de novo	Trio	0	1
chr8(GrCh37):g.68172127del	NM_006421.4 :c.2158del	p.(Leu720Serfs*24)	de novo	Trio	0	1
chr8(GrCh37):g.68169969G>A	NM_006421.4:c.2524C>T	p.(Gln842*)	de novo	Trio	0	1
chr8(GrCh37):g.68152452_68152453dup	NM_006421.4:c.2923_2924dup	p.(Cys976Phefs*6)	de novo	Proband-only	0	1
chr8(GrCh37):g.68200211delT	NM_006421.4:c.1006delA	p.(Met336Trpfs*2)	paternally inherited	Trio	0	1
chr8(GrCh37):g.68178422G>A	NM_006421.4:c.1942C>T	p.(Gln648*)	paternally inherited	Proband-only	0	2
chr8(GrCh37):g.68139728G>A	NM_006421.4:c.3697C>T	p.(Gln1233*)	paternally inherited	Trio	1	2
chr8(GrCh37):g.68138302G>A	NM_006421.4:c.4033C>T	p.(Arg1345*)	unknown	Proband-only	1	1
chr8(GrCh37):g.68170366G>A	NM_006421.4:c.2395C>T	p.(Arg799*)	not inherited from mother, paternal sample unavailable	Duo	0	1

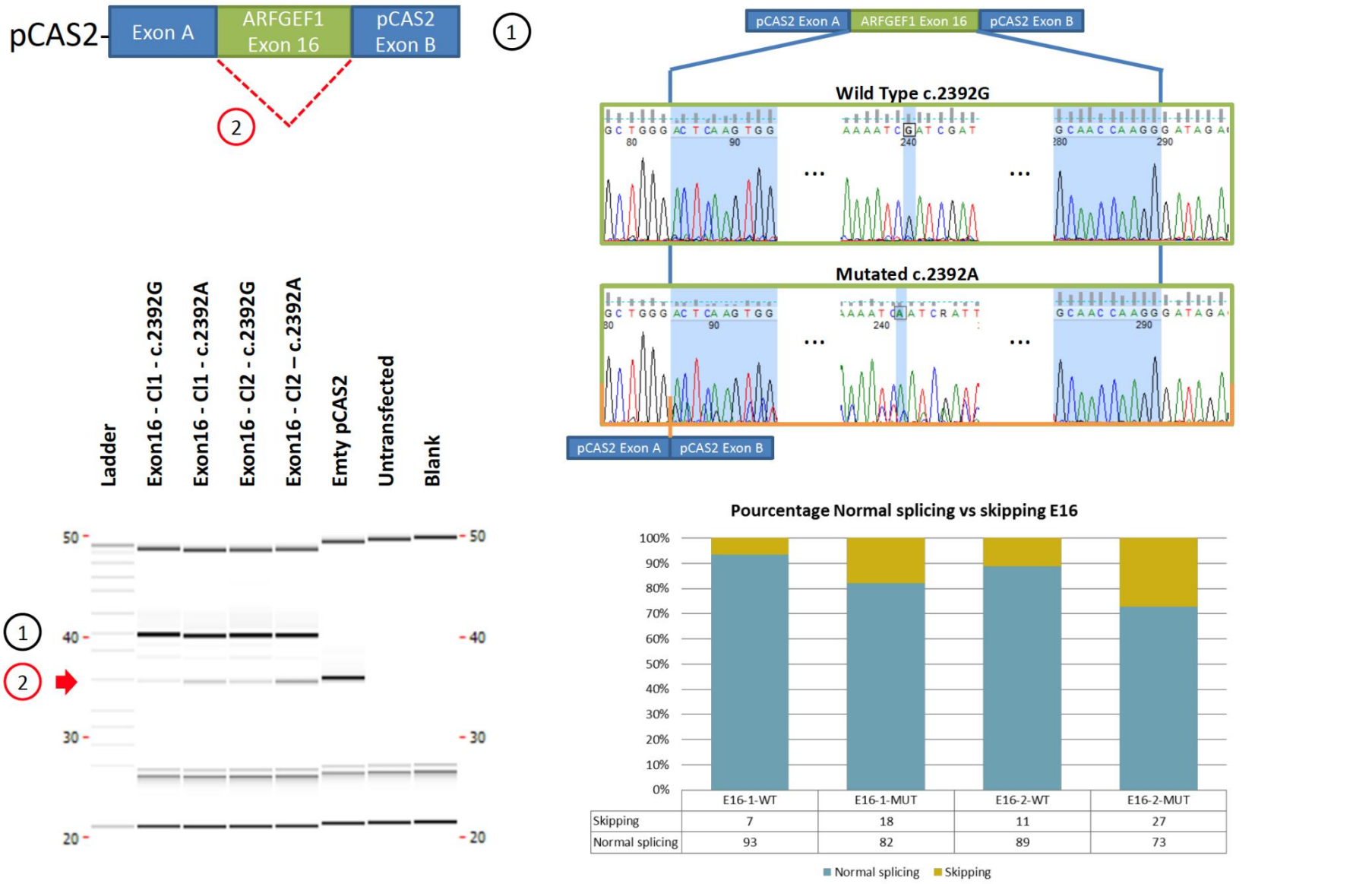
Supplemental Table 1 : Detailed molecular findings on the 11 ARFGEF1 variants identified in the cohort.

Development	Amount	Percentage
Psychomotor delay	12/13	92%
Speech delay	12/13	92%
Degree of speech delay		
Mild	4/12	33%
Moderate	1/12	8%
Severe	7/12	58%
Intellectual disability	10/13	77%
Degree of ID		
Borderline IQ	1/13	8%
Mild	5/10	50%
Moderate	4/10	40%
Severe	0/10	0%
Unavailable Data	1/10	
Behavioral Problems	12/13	92%
Type of disorder		
Aggressive behaviors	6/13	46%
Attention disorders	5/13	38%
Psychomotor agitation	4/13	31%
Autism spectrum disorder	3/13	23%
Anxiety	2/13	15%
Neurological features		
Epilepsy	6/13	46%
Hypotonia	2/13	15%
Cerebellar ataxia	2/13	15%
Morphological Features	8/13	62%
Large ears	5/8	63%
Low-set ears	4/8	50%
Bulbous nose-tip	3/8	38%
High forehead	3/8	38%
Wide mouth	3/8	38%
Long face	2/8	25%

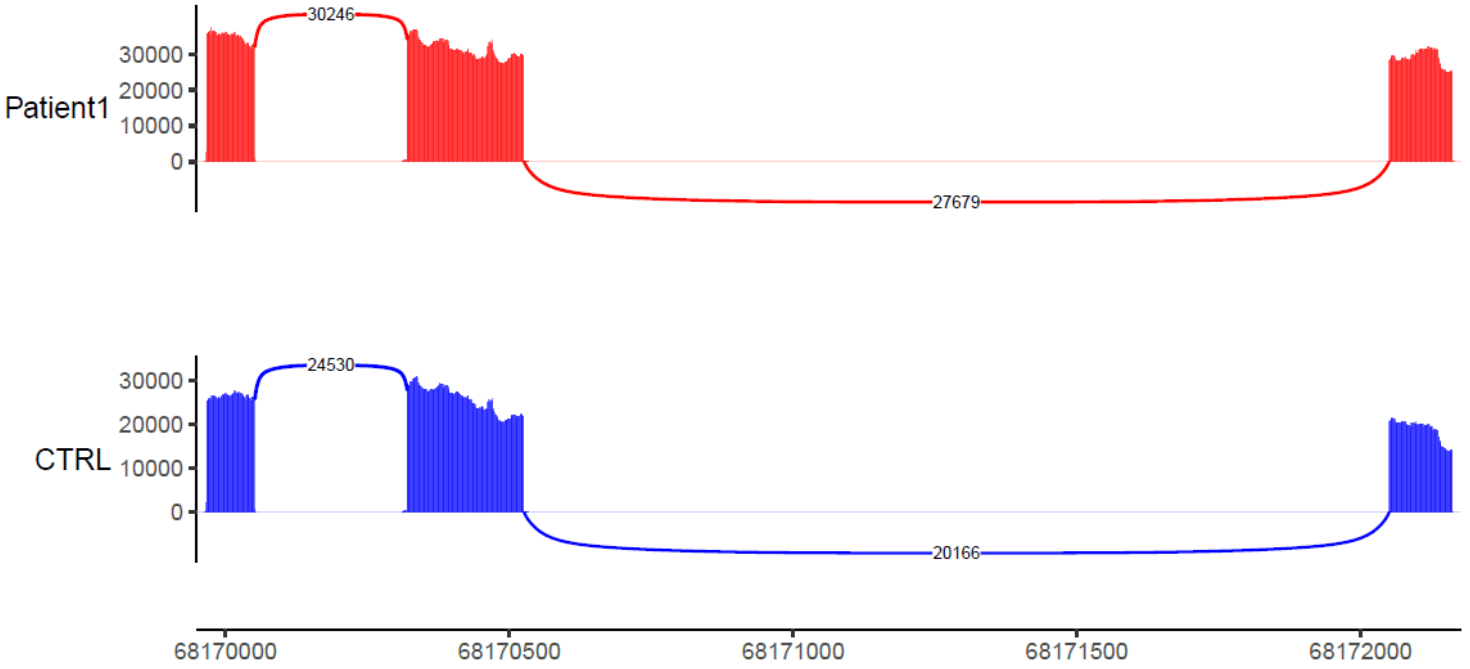
Supplemental table 2: Overall clinical phenotypes of patients with ARFGEF1 expected-pathogenic variants



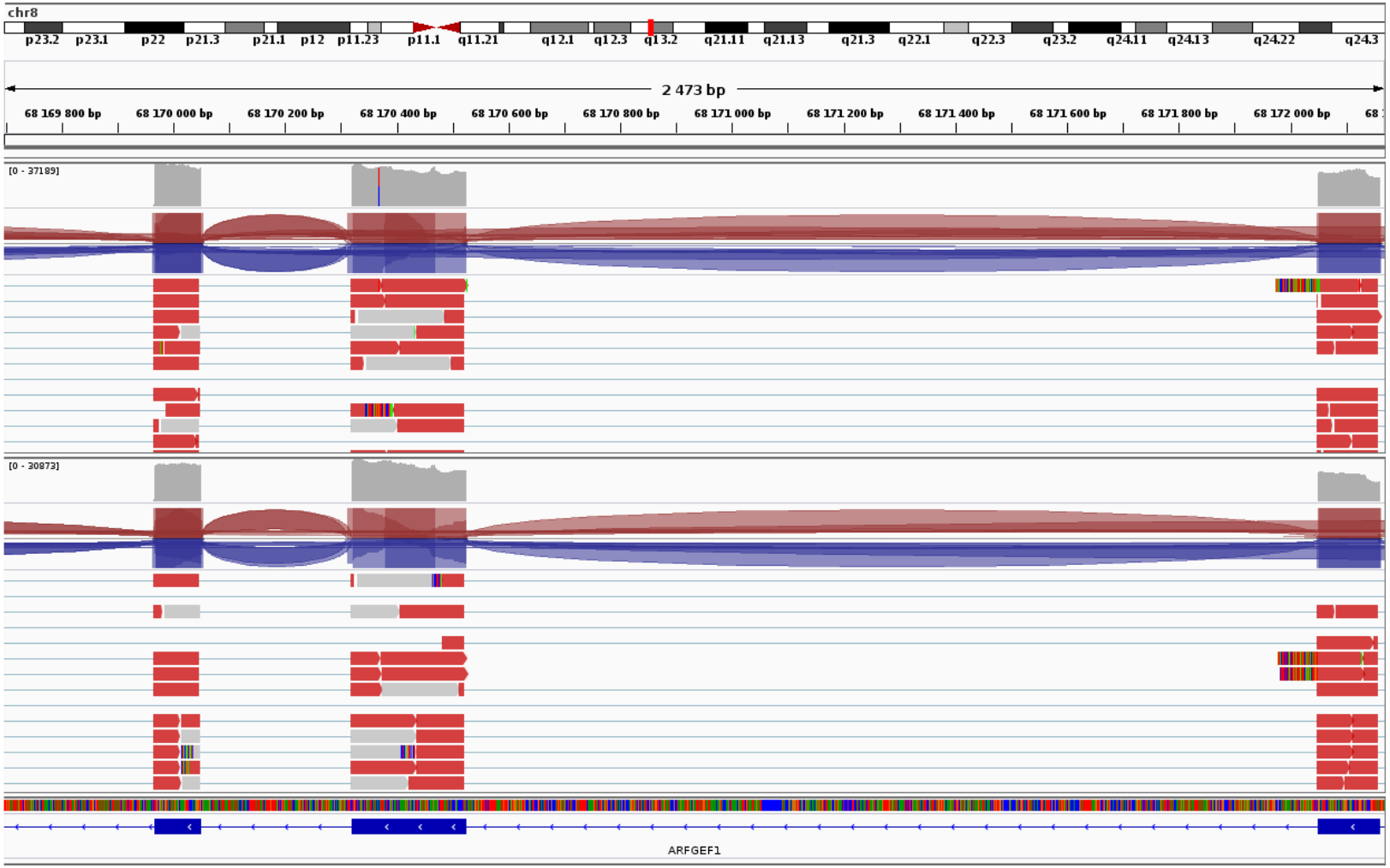
Supplemental Figure 3: Alamut prediction evaluating the splicing effect of the p.(Asp798Asn) variant. Note the prediction that c.2392G>A generates the formation of a new SRp40 exonic splice enhancer (green).



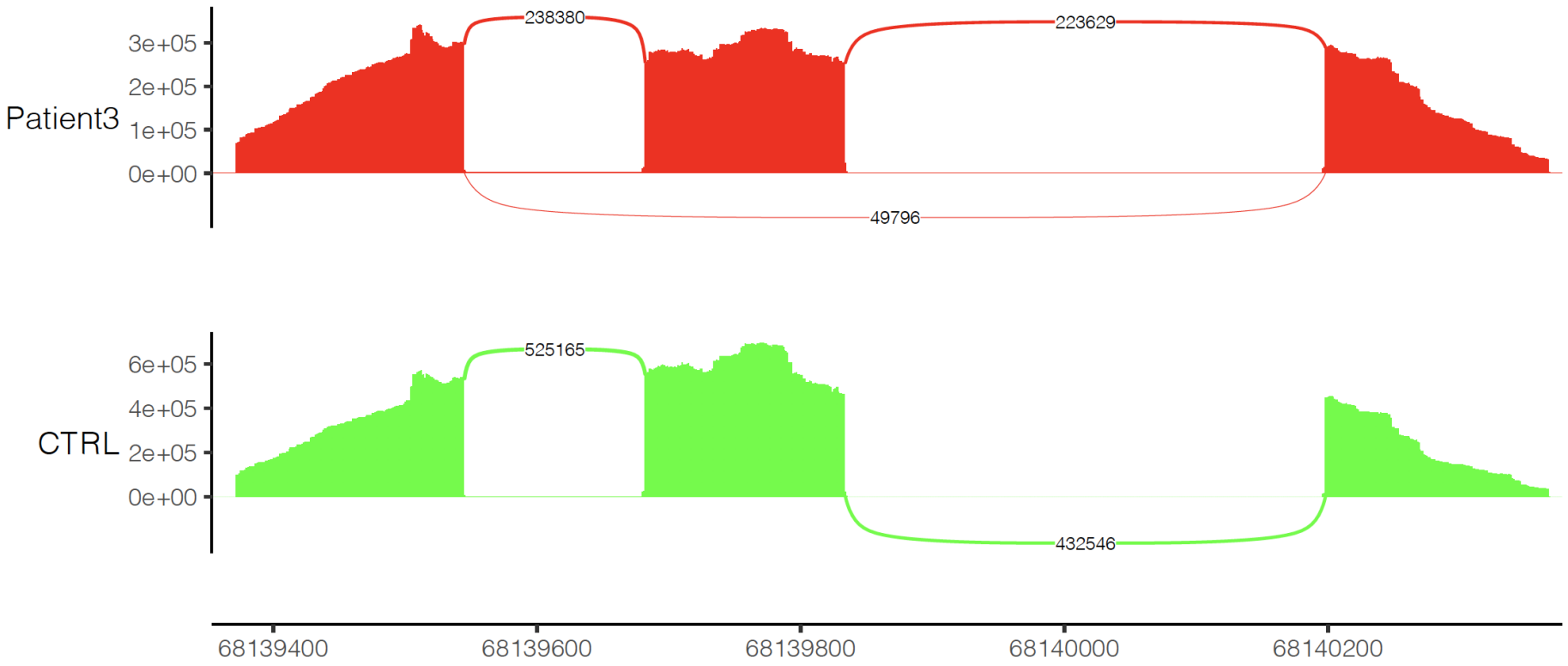
Supplementary Figure 4: Results of the minigene study on the missense variant p.(Asp798Asn). According to prediction programs compiled by Alamut Visual v2.15 (Interactive Biosoftware, Rouen, France), the variant NM_006421.4(ARFGEF1-Exon16):c.2392G>A does not affect canonical splice sites. However, the variant could have an effect by the prediction of the creation of a SRp40 binding site (top right panel). Two different transcripts were identified on Caliper LabChip electropherogram with minigene constructs of exon 16. Transcript 1 corresponds to the physiological splicing of the exon 16. Transcript 2 corresponds to the skipping of the exon 16 (top left panel). It is predominantly occurring in mutant constructs and was confirmed by sanger sequencing (only in mutant). These results suggest an effect of the c.2392G>A variant on splicing of the exon 16.



Supplemental Figure 5: Targeted amplicon RNA-seq studies evaluating splicing events for the c.2392G>A/p.(Asp798Asn) variant in fibroblast cell lines. Note the lack of visible difference between splicing events in the patient against the wild type control (CTRL).



Supplemental Figure 6: Targeted amplicon RNA-seq studies evaluating the allelic frequency in cDNA for the c.2392G>A/p.(Asp798Asn) variant in fibroblast cell lines. Patient 1 data are shown in the top panel. The allelic frequency of the c.2392G>A variant is close to 50%. A control sample obtained from a healthy individual is shown in the lower panel.



Supplemental Figure 7: Targeted amplicon RNA-seq studies evaluating splicing events for the c.3592-2A>G variant in blood. In Patient 3 (top panel), the intronic variant is responsible for exon 26 (center) skipping and consequently NMD. Residual read counts in support of the splicing anomaly between exon 24 (left) and 25 (right) are close to 20% of the wild-type allele. A control sample obtained from a healthy individual in shown in the lower panel.

SUPPLEMENTAL MATERIAL AND METHODS

Cell cultures

Fibroblasts from healthy donors and patients were obtained after consent for skin biopsy (patient 1 and symptomatic father of patient 2 and 3 from whom they inherited the variant). Fibroblasts were cultured in DMEM High Glucose medium (HyClone Thermo Scientific, Waltham, MA, USA) supplemented with 10% Fetal Bovine Serum (FBS, Thermo Scientific, Inc) and 1% ZellShield (Minerva, Biovalley, France). Cells were cultured at 37°C in a humidified 5% CO_2 atmosphere.

Transplanting and seeding were performed weekly as followed: cultures were washed with PBS (5mL) before being trypsinated (4 mL) and let to incubate for 5 minutes. Trypsin was then inhibited thanks to fetal veal serum by adding culture medium (8 mL). The resulting product was then centrifugated for 5 minutes at 500G and cell pellet was resuspended. Cell count was done thanks to a Malassez device with Trypan blue.

Total protein extraction and dosage

Once washed with cold PBS, each well was thoroughly scraped with 100µL RIPA and lysate was recovered and let to incubate on ice for 15 minutes before centrifugation for 15 minutes at 4°C and 16 000G. The final lysate was then transferred on a clean vial and stored at -20°C.

Protein dosage was performed using a BSA calibration scale (0, 50, 100, 200, 300, 400, 500 and 700 µg/mL) from the BCA protein dosage kit (Thermo Scientific©, Waltham, USA) and a 96-well-plate spectrophotometer reader Multiskan Go™ (Thermo scientific©) by measuring samples' absorbance at 562 nm. In brief, the BCA reagent (bicinchoninic acid) forms a colored complex with Cu_1 ions that has a 562 nm absorbance peak. This kit contains a BCA/Cu_2 mix that will be reduced to Cu_1 in presence of protein in an alkaline medium. Absorbance at 562 nm is then proportionate to the amount of protein in the medium.

Western Blot analyses

An equal quantity of protein (patients and control) was denaturized at 95°C for 5 minutes in Laemmli buffer containing β-mercaptoethanol (355mM final, Bio-Rad©, Hercules, California, USA). Proteins were run on a 10% SDS-PAGE in 25 mM Tris, 192 mM Glycine (and 1% SDS) at 100V for migration. This latter was made with 5% or 10% acrylamide.

Transfer was then performed from migration gel to PVDF Immobilon-P membrane (Millipore©) previously activated in methanol, in a buffer containing 6g/L Trisand 3g/L boric acid at 100V during one hour at 4°C. Membranes were then saturated in a washing buffer (PBS tween 0.05%) containing 5% non-fat dried milk during 45 minutes. Membranes were blotted overnight at 4°C with specific antibodies (1/1000) in PBS 0.05% Tween, 5% milk. Membranes were then washed 3 times in PBS 0.05% Tween and probed with an anti-rabbit for ARFGEF1 (Atlas Antibodies, Inc, reference HPA023822) or anti-mouse antibody for vinculin (Sigma, Inc, reference 073M4760V) coupled to horseradish peroxidase. The labeled proteins were detected using the Clarity™ Western ECL substrate (Bio-Rad Laboratories, Inc) according to the manufacturer's recommendations and visualized using ChemiDoc™ Imaging System (Bio-Rad Laboratories, Inc). Vinculin was used as cytoplasmic loading control.

Total RNA extraction and cDNA synthesis

RNA extraction was either performed from cell cultures or whole blood. Cells were washed in PBS and lysed in 1mL Trizol (Ambion). 200µL of chloroform were added, mixed, and incubated at room temperature (RT) for 3 minutes before centrifugation for 15 minutes at 4°C at 12 000G. The RNA-containing supernatant was then retrieved and 500µL of isopropanol were added and mixed before being incubated for 15 minutes at room temperature and then centrifugated at 4°C at 12000 G. The

RNA-containing pellet was then washed twice with 75% ethanol and centrifugated (5 minutes,7500 G, 4°C). The ethanol was then carefully eliminated and the resulting dried-up pellet was resuspended in 20µL nuclease-free water.

Total RNA was extracted from whole blood collected in a PAXgene tube (Preanalytics GmbH, Hombrechtikon, Switzerland) using the PAXgene Blood RNA kit (Preanalytics GmbH, Hombrechtikon, Switzerland) following the standard protocol.

RNA concentration was measured out on a Multiskan Go™ (Thermo Scientific, Waltham, MA, USA) via absorbance reading à 260nm. The purity of nucleic acid was verified thanks to the evaluation of the absorbance ratio A260/A280 which evaluates protein contamination, and the A260/A230 ratio which evaluates the organic solvent contamination. The sample is considered pure if both ratios are between 1.8 and 2. cDNA was obtained using the QuantiTect Reverse Transcription kit (Qiagen GmbH, Hilden, Germany).

Targeted amplicon RNA sequencing

Primer pairs for cDNA sequencing were designed using the Primer 3 software (https://primer3.ut.ee).

Exons:	Primer:	Sequence:
13-23	Forward	ATCAGTATGTGAATCCCAACTCC
13-23	Reverse	CTTGCTATCTGTTTCCAGTCCAC
24-28	Forward	TCTACAAGGCTAGATGGAAATGC
24-28	Reverse	CGCTGGAAAGTGTTTTTCAAAT

PCRs were performed with the Prime Star GXL DNA polymerase kit (Takara Bio Inc) using the following touchdown conditions: initial denaturation 98°C for 3 min; 5 thermal cycles of denaturation at 94°C for 30 sec, primer annealing 65°C for 30 sec, primer extension 68°C for 1 min per Kb; 35 thermal cycles using 60°C instead of 65°C for primer annealing; followed by a final incubation at 10 °C.

PCR products were verified by agarose gel electrophoresis and amplicons were purified with AMpure XP magnetic beads (Beckman Coulter Inc., Brea, CA, USA). Sequencing libraries were prepared with the Nextera XT kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina Miseq using a paired-end protocol (2x150 bases in length).

RNA sequencing data were aligned on the human genome reference (GrCh37/Hg19) using the STAR aligner (version 2.5.2)[53] for each sample.

Data analysis was performed using direct inspection of spliced reads in the Integrative Genomics Viewer software (Broad Institute) and Sashimi plots were generated in Python (2.7.18) using the package ggsashimi (0.6.0).