UNIVERSITÀ DEGLI STUDI
DI GENOVA

Tesi di dottorato in Pediatria

XXXV Ciclo (01/11/2019 - 31/10/2022)

Translational and clinical research applications of exome sequencing to neurodevelopmental disorders of childhood

Relatore

Prof. Pasquale Striano

Candidato

Dott. Marcello Scala

TABLE OF CONTENTS

Background

1. Neurodevelopmental disorders

Neurodevelopmental disabilities are a group of chronic clinically distinct disorders sharing a documented disturbance, quantitative, qualitative, or both, in developmental progress in one or more developmental domains compared with established norms (*Kinsbourne et al., 2001*). These domains are not mutually independent or exclusive and include: (1) motor (gross or fine), (2) speech and language, (3) cognition, (4) personal-social, and (5) activities of daily living (*Shevell et al., 2008*). Whole Exome Sequencing (WES) has provided a huge contribution to the discovery of disease-causing variants for rare diseases, especially neurodevelopmental disorders. Furthemore, the use of WES in clinical practice has improved the diagnostic rate in several rare genetic conditions, including neurodevelopmental disorders, which previously remained unexplained (*Xue et al., 2014*). This has led to a relevant improvement in patient management in selected disorders. Indeed, the better understanding of the pathophysiology underlying a specific condition has helped clinicians in developing a disease-specific approach in patient care. Furthermore, the identification of several new possible therapeutic targets has promoted the development of new therapeutic strategies or specific drugs.

According to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), the diagnosis of intellectual disability (ID) requires deficits in intellectual function (IF), deficits in adaptive behavior (AB), and onset before the age of 18 years (*American Psychiatric Association, 2013*). These deficiencies can be associated with several related and co-occurring problems, including: psychiatric (e.g., depression, anxiety), neurodevelopmental (e.g., autism spectrum disorders - ASD, attention deficit hyperactivity disorder - ADHD), neurological (e.g., infantile cerebral palsy), and medical conditions (e.g., meningitis) (*Lee et al., 2019*). The IF

indicates the global ability allowing the individual to understand the reality and interact with it and includes the common mental activities, such as: logical reasoning and practical intelligence (problem-solving), learning ability, and verbal skills. The IF is measured through the calculation of the intelligence quotient (IQ), a total score obtained from expressly developed standardized tests (IQ tests) derived from Stanford-Binet Intelligence Scales and adapted by Lewis Terman to measure intelligence. Scores were reported as "mental age" divided by chronological age, multiplied by 100. IQ test score has a median of 100 and a standard deviation of 15. A score of 70 or below, corresponding to two standard deviations below the median, is consistent with the diagnosis of ID (*American Psychiatric Association, 2013*). The severity of the ID depends on the IQ score obtained.

Level of ID	IQ	Equivalent Mental Age as Adult	Functioning	
Mild	50-69	9-12 yrs	• Skills acquired late • Sufficient speech • Independent self care at home • Struggling with academic work – reading, math • Unskilled or semi-skilled manual labour • Can live independently with minimal support	**85**
Moderate	35-49	6-9 yrs	Slow acquisition of language, but usually acquire adequate functional speech. Adults require some degree of support in community – e.g. group homes. May be able to read but with difficulties.	**10**
Severe	20-34	3-6 yrs	Daily assistance needed with self-care and safety supervision.	**4%**
Profound	<20	<3 yrs	Continuous care needed, severe limitation in self-care, continence, communication and mobility.	**1%**

Schematic classification of the degree of ID according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), and Relative percentage.

The disabilities in the AB manifest instead as lack of competence in conceptual (the ability to understand time, finance, and language), social (interpersonal skills, social responsibility, self-

esteem, gullibility, naivety, resolution of social problems, and the ability to follow the rules), and practical skills (the ability to use tools, carry out activities of daily living, and interact with other people). The skills defining the AB are learned throughout psychomotor development and become progressively more complex with age. Specific tools are available for assessing AB limitations, such as the Adaptive Behaviour Assessment System (*Milne et al., 2013; Lee et al., 2019*).

Severity level	Conceptual domain	Social domain	Practical domain
Mild	Preschool = no obvious differences. School-aged children and adults = academic skills involving reading writing math time or money. In adults abstract thinking planning cognitive flexibility are somewhat impaired impaired. Tendency toward concrete thinking	Immaturity and social interactions; some difficulty picking up social cues communication conversation in language more concrete than peers. Possible difficulties in emotional regulation and age-appropriate behavior. Perhaps impairment in risk assessment	Personal care may be age-appropriate, but more complex tasks might require support. For example grocery shopping, transportation home and childcare organization food prep banking and money management
Moderate	Conceptual skills lag markedly language development and pre-academic skills slow to develop. School-age children = progress in reading writing mass understanding of time and money but slower than peers. Adults = academic skill development is at an elementary level. Ongoing assistance needed in conceptual decision-making	Marked differences in social and communication from peers. Spoken language is much less complex than peers. Capacity for relationships evident in familial friendship ties. Problems with perceiving social cues in social situations accurately. Social judgment and decision-making limited. Help is needed with life decisions	Personal care is okay in adulthood. Adults typically can participate in all household tasks with teaching. Can work with considerable support in the workplace
Severe	Limited attainment of conceptual skills. Little or no understanding of written language math, time and money. Extensive support for problem solving is needed	Spoken language is limited in terms of vocabulary and grammar. Communication is focused on the here and now an everyday event. Relationships and relational ability is considerable.	Support needed for all activities of daily living. Supervision required at all times. We will not make responsible decisions regarding well-being .skill acquisition is very limited
Profound	No concept of symbolic processes, perhaps some functional use of objects, although this might be limited by disturbance and motor skills.	Might understand simple instructions and cues. Social expression is often nonverbal. Can respond and enjoy relationships with people who were well known to them. Can initiate limited social interaction with such people through gestures. Sensory and physical impairments may prevent social activities	Dependent on others for all aspects of daily physical care. Participation in these activities is limited.. Some simple concrete tasks such as carrying dishes to the table might be accomplished. Co-occurring physical and sensory impairments are often barriers to participation

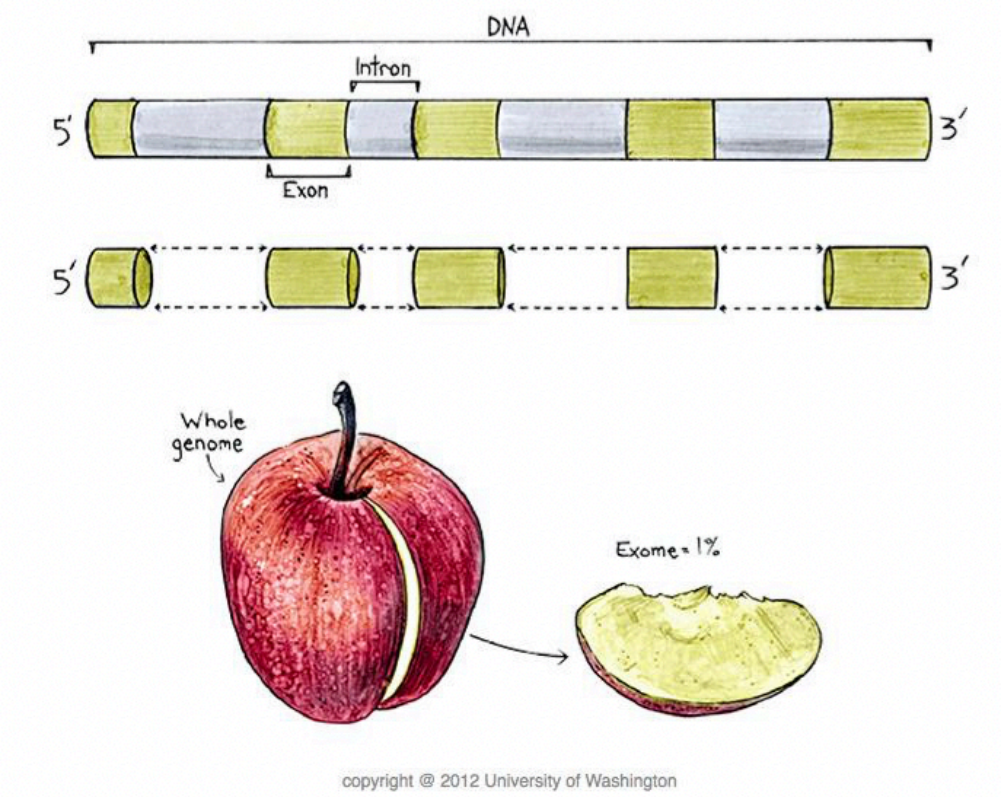
Adapted from Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)

In contrast with previous classifications, the DSM-V has changed the standards for the diagnosis of ID, which must be made considering both the IF and AB in each case and cannot exclusively rely on the IQ score alone (*Obi et al., 2011*). A child with IQ<70 but efficient AB does not have an ID, whereas this diagnosis can be made in a patient with a normal (or higher than normal) IQ score but severe deficits in AB. A distinct diagnostic category in the DSM-V is the “Unspecified Intellectual Disability”, also known as Intellectual Developmental Disorder – IDD. This term is

used to describe individuals over the age of 5 suspected of having an ID but having difficulties in completing required tests, usually because of limitations resulting from blindness, deafness, or concurrent mental illness. Many causes of ID are not known yet. A part from the ID conditions related to environmental exposure during pregnancy (e.g., maternal exposure to a toxin, infectious agent, uncontrolled maternal condition, birth complications), many genetic abnormalities can cause syndromic or nonsyndromic ID. These genetic abnormalities include single gene mutations, copy number variations (CNVs), or chromosomal abnormalities which all result in an inborn error of metabolism, neurodevelopmental defect, and neurodegeneration (*Lee et al., 2019*).

2. Whole Exome Sequencing

Whole exome sequencing (WES) consists in the massive parallel sequencing of the exons of human genes. Exons are the genetic regions that will be translated into proteins and thus be expected to have a functional or structural effect. Since each exon is sequenced multiple times, the sequencing is checked within each subject, increasing confidence that the final sequence is accurate. This technique can be applied to all exons or can be used to focus on a particular region of the genome which has previously been identified as being of interest (e.g., a section of a chromosome). Furthermore, WES can be used to sequence a set of genes which are not necessarily genomic neighbours (*Tetreault et al., 2015; Yang et al., 2014; Harding et al., 2014*). Despite WES only covers 1–1.5% of the human genome, this portion of the genome houses approximately 85% of the known disease-causing variants (*Ng et al., 2009; Rosenfeld et al., 2012*).



Each gene is composed of several exons and intronic sequences. Exome represent around 1% of the whole human genome.

2.1 Technique

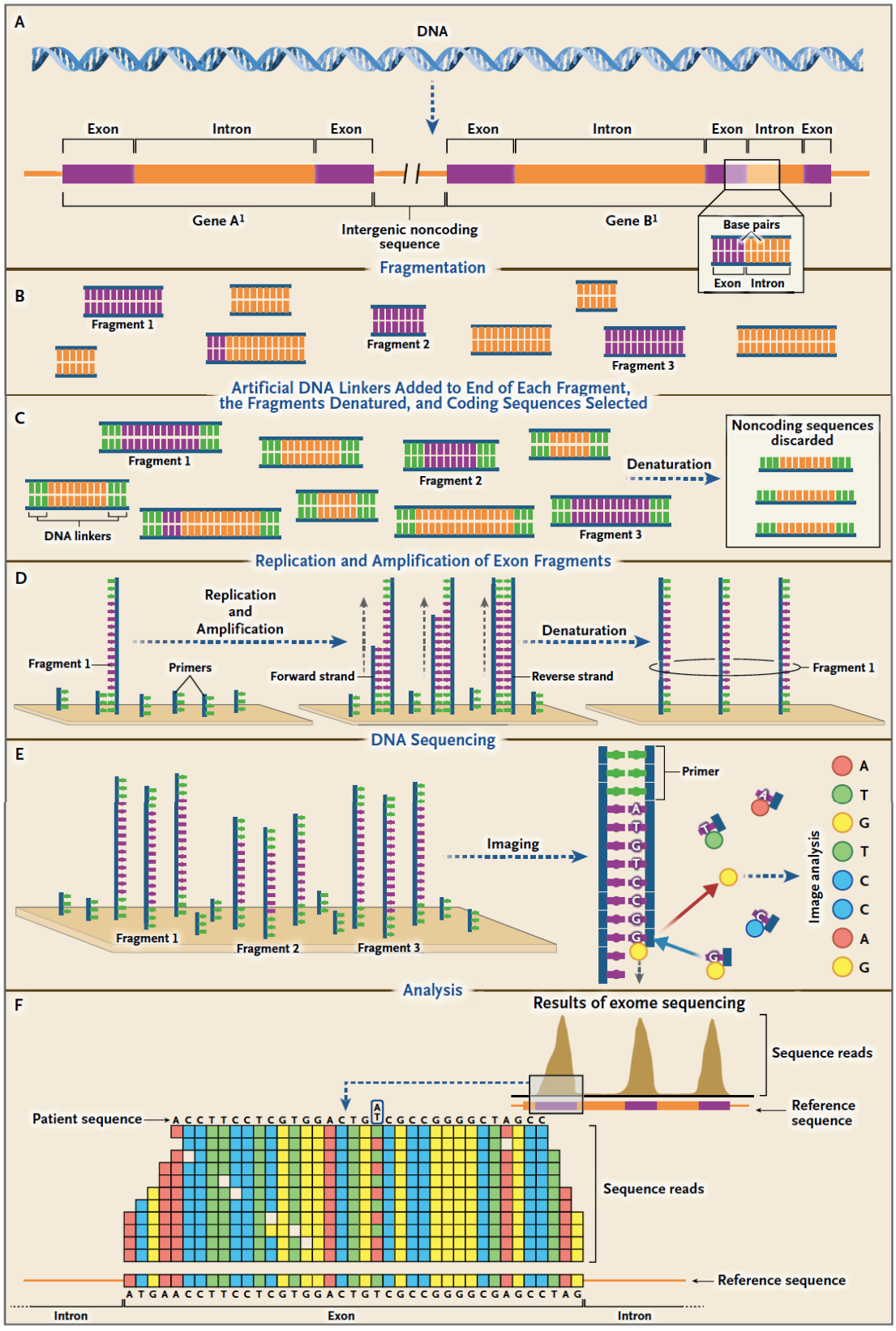
Next Generation Sequencing (NGS) consists in the sequencing of millions of sequence reads in parallel, allowing a massively through-put. Initially used to study the whole genome, NGS techniques has been further employed to address specific regions of the genome, namely the exonic sequences in WES. Two technical preparatory approaches can be adopted for this purpose (*Koboldt et al., 2013*):

1. PCR-based: this approach typically involves multiple primer pairs in a mixture that are combined with genomic DNA of interest in a multiplex approach to preserve precious DNA. The use of multiplex primer pairs couples the high throughput of NGS platforms and the fact that each sequence read represents a single DNA product in the mixture due to the nature of the sequencing

platforms. Following the PCR, the resulting fragments have platform-specific adapters ligated to their ends to form a library that is suitable for sequencing (*Mardis et al., 2013*);

2. Hybrid capture-based: hybrid capture takes advantage of the hybridization of DNA fragments from a whole-genome library to complementary sequences that were synthesized and combined into a mixture of probes designed with high specificity for the matching regions in the genome. These probes typically have covalently linked biotin moieties, enabling a secondary "capture" by mixing the probe:library complexes with streptavidin-coated magnetic beads.

Hence, the targeted regions of the genome can be selectively captured from solution by applying a magnetic field, whereas most of the remainder of the genome is washed away in the supernatant. Subsequent denaturation releases the captured library fragments from the beads into solution, ready for postcapture amplification, quantitation, and sequencing. When the probes are designed to capture essentially all of the known coding exons in a genome, the capture approach is referred to as "exome sequencing." Additional probes may be designed, synthesized, and added to an exome reagent, typically referred to as "exome plus." When only a subset of the exome or of the genome outside of the exome is targeted, this is called a "targeted panel" (*Albert et al., 2007; Gnirke et al., 2009; Hodges et al., 2007*).



Schematic representation of NGS-based exon sequencing, adapted from Biesecker et al., 2014.

2.2 Data analysis

The complex process of data analysis plays the pivotal role in WES. This analysis encompasses several steps in sequence: base calling, quality control, mapping and variant identification, annotation, filtering and prioritization (*Tetreault et al., 2015*):

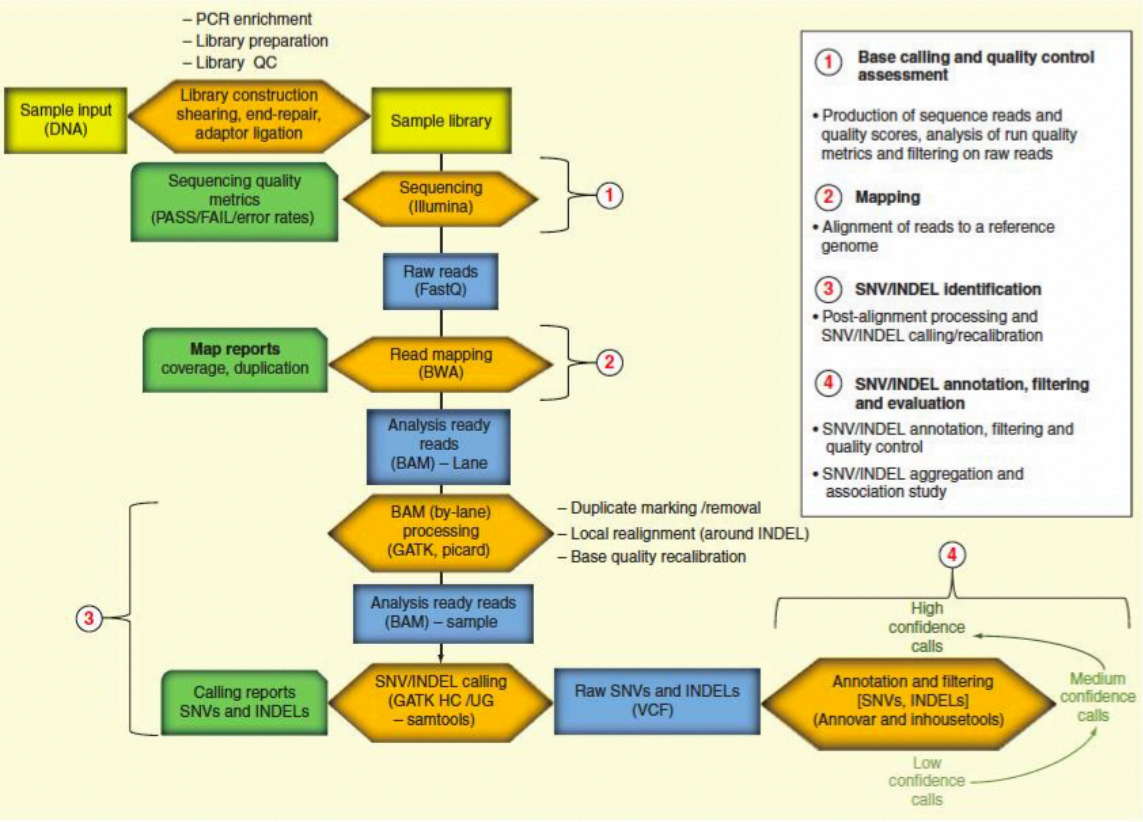


Figure 1. A clinical exome sequencing analysis workflow for mapping, QC control, single nucleotide variant and small INDEL detection, annotation, filtering and evaluation. This workflow is mainly recommended for use in clinical setting rather than research, as it does not cope for complex variations.

Figure adapted from Tetreault et al., 2015.

2.2.1 Base calling and quality control assessment

Base calling indicates the identification of each nucleotide in the analyzed sequence. The accuracy of variant detection is critically dependent upon the accuracy of this process, which is prone to several errors (such as those due to nucleotide misincorporation during the sequencing

process and optical detection errors occurring during image analysis). Thanks to the relevant improvements in sequencing chemistry and the use of high-performance base calling algorithms, current NGS technologies have reached a base-calling accuracy of more than 99.5% (*Massingham et al., 2012*).

Quality control assessment (QCA) is an essential step to identify and correct the base-calling errors before proceeding with the further analysis, avoiding mistakes in the variant interpretation process. QCA is a two-step process:

1. Pre-alignment step: assesses the intrinsic quality of the raw reads using metrics generated by the sequencing platform (% error and per-base quality scores, % duplication). Reads with insufficient quality may be discarded or trimmed. Duplicate reads are generally discarded to guard against PCR-generated errors.
2. After mapping step: involves the calculation of metrics from the alignment profiles. Tools such as FASTQC [Babraham bioinformatics. Available from: www.bioinformatics.babraham.ac.uk/projects/fastqc/] and Picard toolkit [Picard. Available from: http://picard.sourceforge.net/command-line-overview.shtml] are used for QC assessment. They generate a large number of metrics, of which some of the most relevant are: proportion of reads aligned to reference genome, average depth of coverage, overrepresented sequences, error rate and duplication rate, which can be used to gauge the success of sequencing. Sequencing runs with unacceptable final quality may be repeated or supplemented with additional data.

2.2.2 Mapping

This process consists in the alignment of the generated sequence reads to a ‘reference’ genome to reconstruct a patient’s exome or genome. Even though the comparison of patient data with a reference genome may sound easy to perform, some relevant limitations arise. For example, the current human reference genomes are still incomplete and there is no consensus on the optimal reference(s) to be used (*Rosenfeld et al., 2012; Tetreault et al., 2015*). Many alignment programs have been developed to efficiently process millions of short reads. Few generations of algorithmic improvements with implementations such as MAQ and BWA had remarkably improved the mapping process and contributed to exemplify the current lineup (*Li, Ruan, et al., 2008; Li, Durbin, et al., 2009*). The performance of the aligners is generally efficient in the presence of single base mismatches, but difficulties may arise in case of insertions or deletions. When the alignment stage is complete, a catalogue of the ‘exomic’ regions been sequenced at a high redundancy is produced. In the best scenario, each exonic base may be represented by around 100 independent sequencing reactions (100-fold or 100_ coverage). This level of redundancy guarantees good coverage of virtually all targeted regions and allows to guard against random sequencing errors that may appear in individual reads.

Visualization of sequence read alignment data (BAM or SAM) on the *Integrative Genomics Viewer* (IGV), a non-GATK tool developed at the Broad Institute allowing the interactive exploration of large genomic datasets.

Adapted from https://gatkforums.broadinstit ute.org

2.2.3 Variant identification

Once appropriate QC and sequence alignment are completed, Variant identification is performed. This crucial stage in WES data analysis indicates the process of determining the variations between a sample and the reference genome. These variations include: single nucleotide variants (SNVs), multinucleotide variants, small insertions or deletions (INDELs, generally less than 50 bp), larger structural variants (SVs) and copy number variants (CNVs) (*Kadalayil et al., 2014*).

Table 1. Efficacy of technologies to detect different types of variants.

	WES	WGS	RNAseq	aCGH	SNP arrays
SNVs	✓✓	✓✓✓	✓✓	–	✓
Coding INDELs	✓✓	✓✓	✓✓	–	✓
Splicing variants	✓✓	✓✓✓	✓✓✓	–	✓
Deep splicing variants	✓	✓✓	✓✓✓	–	✓
Regulatory variants	✓	✓✓	✓	–	✓
Intronic SNPs	✓	✓✓✓	✓	–	✓
CNVs	✓✓	✓✓✓	–	✓✓	✓✓
Gene fusion	–	✓	✓✓	✓	–

–: None are detected.
✓: A few variants of this type are detected.
✓✓: Most variants of this type are detected.
✓✓✓: All variants of this type are detected.
CNVs: Copy number variants; SNVs: Single nucleotide variants.

Adapted from Tetreault et al., 2015.

Probabilistic methods based on Bayesian statistical models, such as SAMtools and Genome Analyzer Tool Kit have proved to be robust in variant identification thanks to their accurate statistical modeling of various sequencing errors (*Abecasis et al., 2012; Li, Handsaker, et al., 2009; McKenna et al., 2010*). However, the accuracy of these tools is highly variable and depends on the average depth of sequencing and variant type of interest (e.g., SNVs are identified much more consistently than INDELs). Furthermore, the sensitivity of the whole process (i.e., the ability to detect all true variants within the exome) depends more on the sequencing coverage than the software that is used. For example, a 100–150_ mean exome coverage allows to cover 95% of all known coding exons at a depth of 20_ or more, with a rough exome-wide sensitivity of around 95%. The selectivity (i.e., the probability that a given identified variant is real) is instead more dependent on the software used, as shown by the slightly better results obtained using Genome Analyzer Tool Kit in comparison to other software kits. Depending on the evaluation method used, the selectivity of WES analysis in detecting SNVs has been reported to be 99-99.7% and around 80% for INDELs (*O'Rawe et al., 2013*).

Eventually, additional improvements in detecting false positives can still be obtained by visual inspection of the data by experienced analysts thanks to tools such as the Integrated Genomics Viewer (*Thorvaldsdottir et al., 2013*).

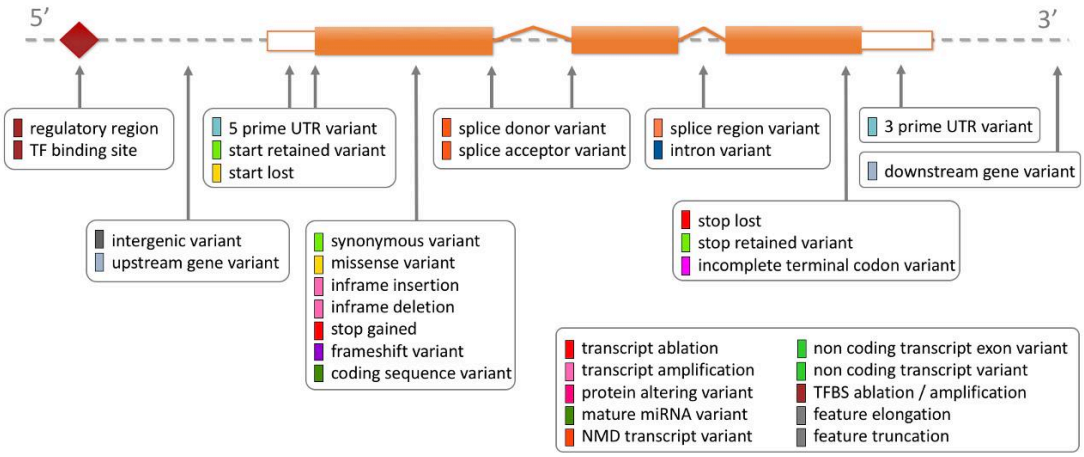
2.2.4 Variant annotation, filtering, and evaluation

The final stage of WES data analysis encompasses the three following essential steps:

A. Annotation

This process assigns functional information to DNA variants, aiming to provide a link between the identified variants and the patient's phenotype. Currently, the most commonly used annotation tools that can provide all the relevant information requested by the user to the raw variant lists are ANNOVAR, SnpEff and VEP (*Cingolani et al., 2012; Wang et al., 2010*) (*The Bioinformatics Knowledge blog. Available from: http://bioinformatics. knowledgeblog.org38*). The annotation process can be in turn divided in substeps:

a) **Variant classification:** classify the identified variants according to their functional category (e.g., synonymous, missense, nonsense, splicing, etc.) using known gene annotation. The changes directly affecting the amino acid sequence will be the most relevant to disease studies.



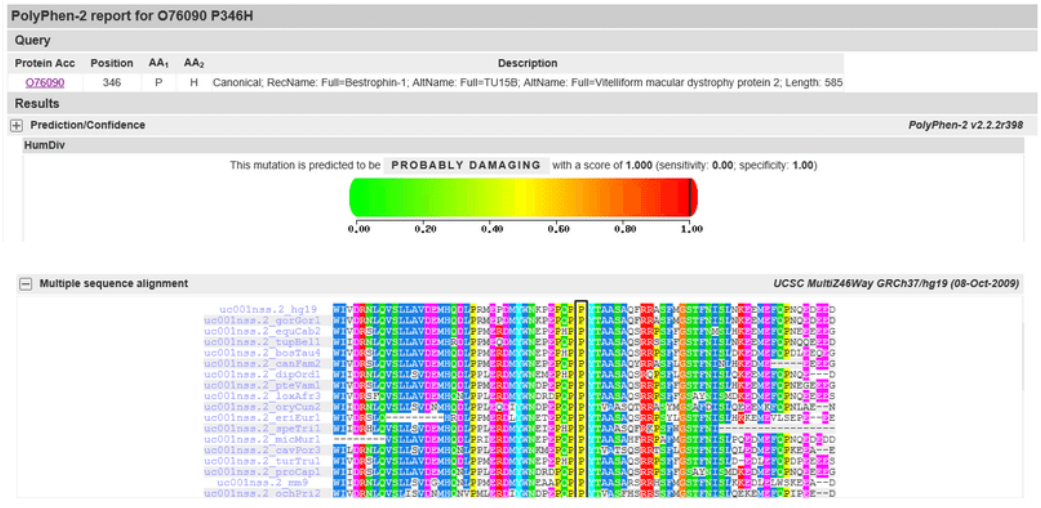
Adapted from Ensembl.

b) Frequency analysis: the presence and frequency of a selected variant is assessed in public databases. A polymorphism present in more than 5% in the normal population is unlikely to be of relevance to a rare genetic disease and this threshold falls to 1% in case of presumably dominant genetic conditions.

Many NGS studies annotate their variants using Minor Allele Frequency (MAF) data from major population repositories such as NHLBI Exome Sequencing Project, dbSNP, The 1000 Genomes Project Consortium (1000G), Genome Aggregation Database (GnomAD) and ExAC database (*Sherry et al., 2001; Fu et al., 2013*) (*ExAC Browser (Beta). Available from:* http://exac.broad.institute.org).

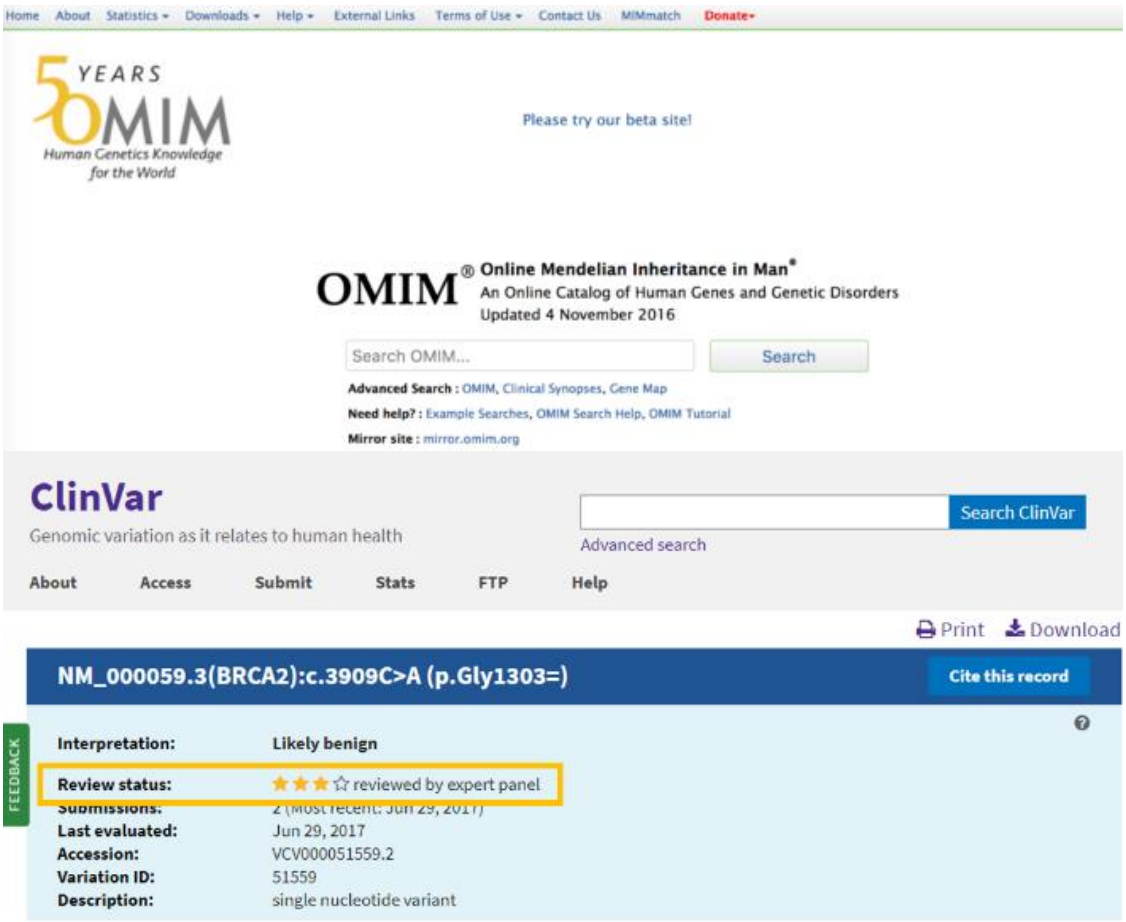
Adapted from GnomAD.

c) Conservation and effect predition: in this step, important additional information on the possible pathogenicity of a selected variant are retrieved through the measures of evolutionary sequence conservation (thanks to tools such PhastCons, GERP++, and Phylp) and the predictions of the possible effects produced by that variant on protein function (*Davydov et al., 2010; Margulies et al., 2003; Pollard et al., 2010*). For the last purpose, several in silico tools are available to estimate the computed pathogenicity, including the most well-known SIFT and Poly-Phen (*Adzhubei et al., 2010; Adzhubei et al., 2013; Ng et al., 2003*).



Sample use of Polyphen-2. The amino acid substitution p.P346H in protein bestrophin 1 is predicted to be damaging with a score of 1. Multiple sequenced alignment (Basic Local Alignment Search Tool) shows that the residue at position 346 of bestrophin-1 is highly conserved across species. *Adapted from Li et al., 2017.*

d) Reference to disease-specific databases: these large genomic databases contains extremely relevant information concerning both genes and individual variants and their use is of priceless value in establishing the final link between genotype and phenotype. Some of the most common databases are OMIM, ClinVar, Human Gene Mutation Database (HGMD), and Gen2Phen (*Amberger et al., Landrum et al., 2014; Stenson et al., 2014; Webb et al., 2011*).



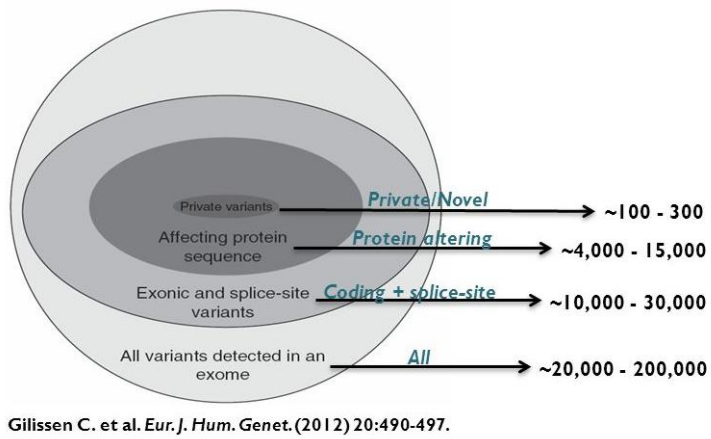
Two of the most common disease- specific databases used to screen potentially pathogenic variants. *Adapted from OMIM and ClinVar.*

B. Filtering

Typically, after the comparison of a patient's exome with the human reference genome, nearly 100,000 variant positions are identified. Most of such variants are functionally neutral and do not have any pathogenic impact. Since the list is too long to be evaluated manually, the filtering process allow to discard those variants which are clearly benign in order to narrow the list down to a few candidate genes. These genes can then be followed-up by a researcher or a clinical geneticist to determine their involvement in the studied disorder.

The filtering process is usually a semi-automated process starting excluding non-coding variants, at least not in the first pass of the analysis, and variants with high MAF in public databases. For rare recessive disorders, the appropriate MAF cut-off may be around 1% and only genes with rare homozygous or multiple heterozygous variants are retained (usually 5-10 per exome, but typically higher in consanguineous families). For a dominant or X-linked disorder, a cut-off much closer to zero should be used. In cases with suspect dominant mode of inheritance, the final list may consist of 50–100 variants per exome, making sequencing of additional family members relevant in order to further filter these variants comparing the probands with the parents and healthy siblings, when available.

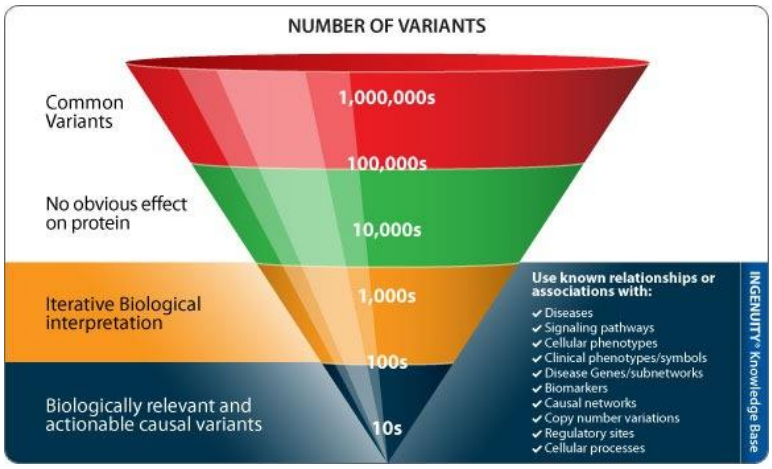
Despite the extra cost associated with this procedure, it is especially important when looking for *de novo* mutations (i.e., variants carried by the proband but absent in the parents, usually 1-2 per exome). *In silico* tools predicting the functional impact on the protein (such as SIFT or PolyPhen) can be use in the prioritization of variants or may be left for the final step of 'manual analysis' of the most interesting variants (*Tetreault et al., 2015*).



Adapted from INGENUITY Knowledge base.

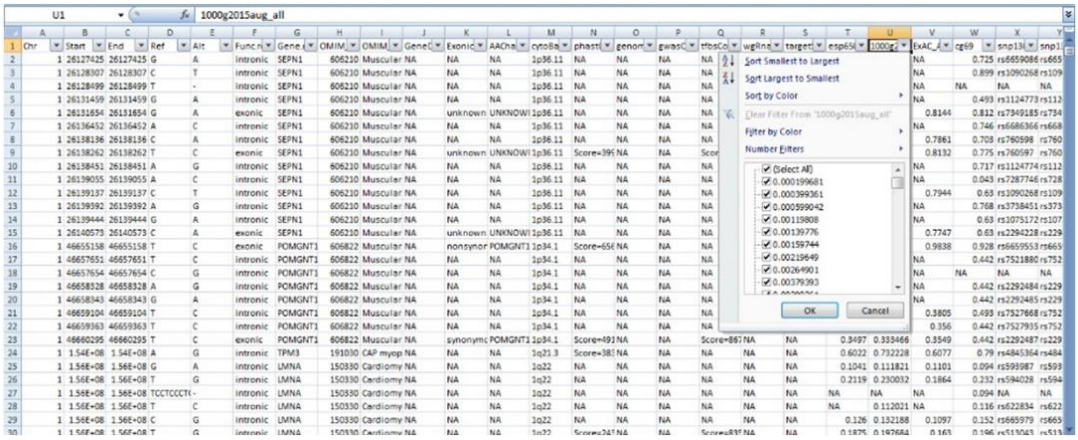
C. Evaluation

The manual examination of the filtered candidates (representing the most interesting variants which might account for the phenotype showed by the patient) by a qualified geneticist represents the final step of the interpretation process. Indeed, despite the efforts made to improve automated diagnosis and keyword matching, the role of the expert intervention still remains essential and irreplaceable for both the diagnostic process and generation of a final report. This analysis is performed through the visual inspection of detailed annotations, especially OMIM descriptors and review of the relevant literature.



Variants processing and analysis. *Adapted from INGENUITY Knowledge base*.

In silico tools for the prediction of the functional effect of the studied variant are of relevant value. A part from the most common bioinformatics tools (Polyphen, Polyphen-2, SIFT, Mutation Assessor, Mutation Taster, etc.), the use of the Combined Annotation Dependent Depletion (CADD) may provide important a relevant support in the final choice of the right variant. Once again, the analysis of the conservation of the affected nucleotide residue remains fundamental in this stage and several scores may be useful, especially GERP+ and Phylp). Eventually, a pivotal role in the evaluation process is played by the 'deep phenotyping', a term indicating the thorough clinical characterization of the patient. Indeed, correct and detailed clinical information are essential for establishing the possible pathogenicity of a specific variant in the studied subjects and the possible relevant genotype-phenotype correlations.



Snapshot of a sample list of variant file in excel format. *Adapted from Exome Data Analysis for Clinicians: How and Why by Aneek Das Bhowmik and Ashwin Dalal*

In the end of the interpretation process, the variants of interest should be classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines, which provide a definite characterization: pathogenic/likely pathogenic (V-IV classes, variants which are most likely predicted to explain the clinical phenotype of the studied subjects on the basis of several frequency and biocomputational criteria), variant of unknown significance (VUS) (III class,

variants whose relevance as the cause of the studied phenotype is uncertain due to mostly benign biological predictions or conflicting interpretations of pathogenicity), and likely benign/benign variants (II-I classes, variants predicted to have no significant impact on the protein function/expression). This classification provides a valuable way of standardise the variant interpretation process among different clinical and research centers worldwide (*Richards et al., 2015*).

Interpretation of sequence variants | RICHARDS *et al*

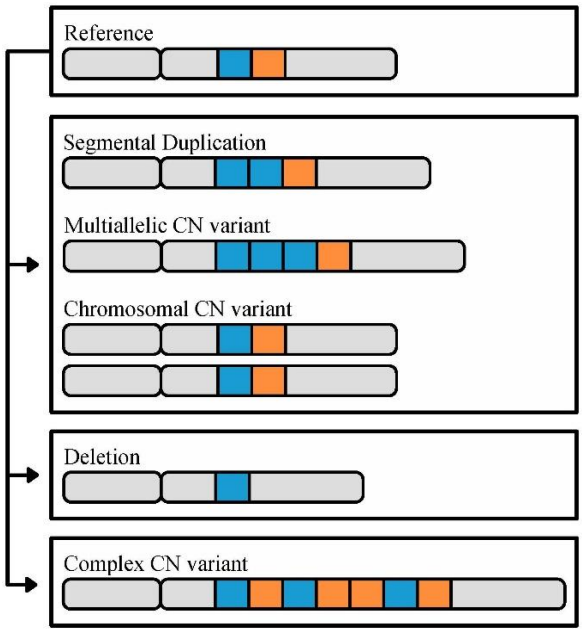
ACMG STANDARDS AND GUIDELINES

	Benign		Pathogenic			
	Strong	**Supporting**	**Supporting**	**Moderate**	**Strong**	**Very strong**
Population data	MAF is too high for disorder BA1/BS1 **OR** observation in controls inconsistent with disease penetrance BS2			Absent in population databases PM2	Prevalence in affecteds statistically increased over controls PS4	
Computational and predictive data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein length changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in gene with low rate of benign missense variants and path. missenses common PP2	Mutational hot spot or well-studied functional domain without benign variation PM1	Well-established functional studies show a deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1	Increased segregation data →		
De novo data				De novo (without paternity & maternity confirmed) PM6	De novo (paternity and maternity confirmed) PS2	
Allelic data		Observed in *trans* with a dominant variant BP2 Observed in *cis* with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other data		Found in case with an alternate cause BP5	Patient's phenotype or FH highly specific for gene PP4			

ACMG guidelines for variant classification. *Adapted from Richards et al., 2015.*

Copy Number Variants (CNVs)

This term indicates a genomic structural alteration consisting in the variation of the number of specific sections of the genome resulting from duplication and deletion processes affecting a large number of base pairs, with a varying number of repeats in human population (*McCarroll et al., 2007; Sharp et al., 2005*).

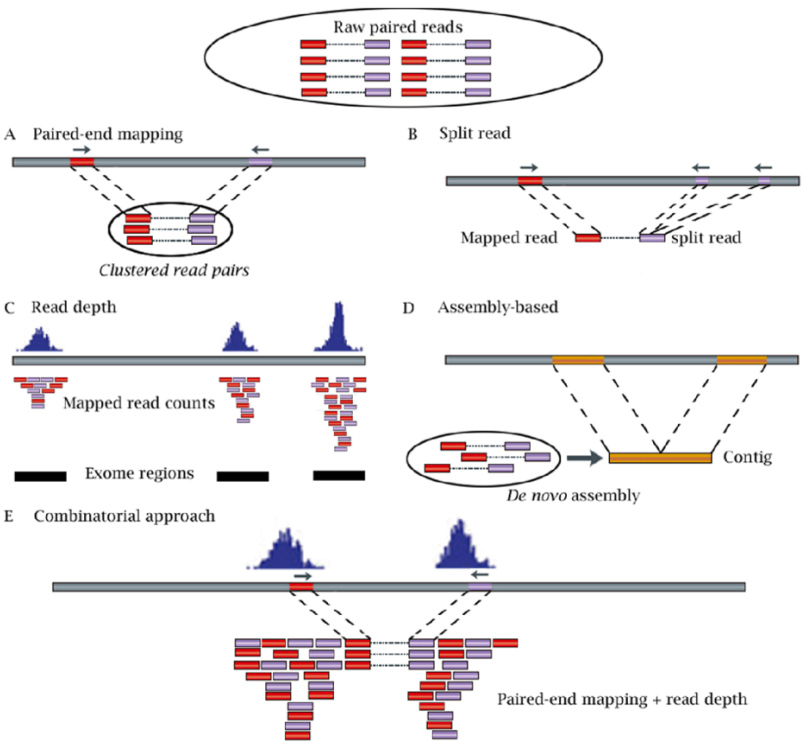


The different types of CN variation. CN variants range in size (50 base pairs or greater) to whole chromosomes, and are identified through comparison to a reference genome. *Adapted from Steenwyk et al., 2018.*

Genomic changes can be extensive (involving several megabases) and complex, leading to duplication, deletion, insertion, inversion or translocation of large DNA segments. Some of the most extreme changes occur as somatically acquired rearrangements in tumor genomes, including whole arm and chromosome events, as well as whole genome duplications (*Weinstein et al., 2013*). The potential consequences of the detected CNVs can vary significantly, ranging from neutral (common CNVs in human populations, neutral polymorphysms) to potentially damaging (associated with a variety of clinical phenotypes). The clinical impact may be

determined determined by mode of inheritance, total rare CNV burden, size and number of dosage-sensitive genes in the CNV region (*Redon et al., 2006*).

Thanks to the recent remarkable technical advancements, it is possible to detect these variations using WES. The detection process differs a lot from pipelines used to identify SNVs and small INDELs, usually involving five main approaches: read depth, allele frequency, paired-end mapping, split read mapping, and *de novo* assembly (*Alkodsi et al., 2014; Tan et al., 2014*). Due to the discrete nature of WES data, the first two are the most popular approaches for CNV detection using WES data. Through the combination of these two methods, WES can identify both copy neutral genomic rearrangements that would be lost with read depth-based approach alone and balanced copy number changes (e.g., AB to AABB aberrations, where normal and aberrant genotypes both show allele ratio of 0.5) not detectable with the allele frequency-based approach alone (*Miyatake et al., 2015; Nadaf et al., 2015; Wong et al., 2004*).



Five approaches to detect CNVs from NGS short reads:
A. Paired-end mapping (PEM) strategy detects CNVs through discordantly mapped reads (distance between two ends of a read pair significantly different from the average insert size); B. Split read (SR)-based methods use incompletely mapped read from each read pair to identify small CNVs;
C. Read depth (RD)-based approach detects CNV by counting the number of reads mapped to each genomic region (three exome regions in the figure);
D. Assembly (AS)-based approach detects CNVs by mapping contigs to the reference genome.
E. Combinatorial approach combines RD and PEM information to detect CNVs.
Adapted from Zhao et al., 2013.

The annotation and the following interpretation of the identified CNVs should be performed according to the available phenotypic information. Likewise to the case of the abovementioned smaller variants, 'deep phenotyping' is of paramount importance for the correct interpretation of the genetic variations in the single patient. Specific databases are available for the evaluation of the potential clinical consequences of CNVs, including Database of Genomic Variants and the Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (*Church et al., 2010; Firth et al., 2009*). Due to the relatively small size of exome, WES can quickly and efficiently identify recurrent aberrations across large number of samples, being particularly interesting in the field of cancer genome studies (*Nadaf et al., 2015*). Even though the perspectives on the efficiency of WES in detecting CNVs are very promising, some technical limitations will need to be overcome in the future. For example, it is still challenging to determine the clinical consequences of a given CNV despite the help provided by databases. Indeed, clinical CNVs may not be identical, even for those causing similar phenotypes and different technologies, segmentation algorithms, or computational approaches would report different breakpoints for a given sample. Despite these challenges, however, detection of CNVs is becoming more and more important since their involvement in human diseases is increasingly being recognized (*Ionita-Laza et al., 2009*).

2.3 Applications of WES in research and clinical practice

Through the comparison of an individual sequence with a normal reference WES allow to identify variations in an individual's DNA which may be causative of the studied specific medical conditions. The first diagnosis of a patient using WES was reported in 2009 and in 2013

more than 180 distinct novel disease-causing genes had already been identified thanks to WES (*Boycott et al., 2013*). Over the past 10 years, WES has remarkably accelerated the process of identification of disease-causing genes in research setting (*Frebourg, 2014*). The huge success of WES in gene identification has further led to the incorporation of this technique in the clinical practice with a relevant impact on diagnosis and patient management. Three relevant fields have gained advantage from the use of WES: rare diseases, cancer, complex disorders.

2.3.1 *Rare diseases*

According to the European Medicines Agency (EMA), rare diseases are defined as life-threatening or chronically debilitating conditions that affect no more than 5 in10,000 people in the EU (*Richter et al., 2015*). Although individually rare, overall they affect a large number of individuals (*Frebourg, 2014*). Molecular characterization is available for more than a half of the around 7000 rare inherited disorders identified so far (*Frebourg, 2014*). In recent years, WES has provided a huge contribution to the discovery of disease-causing variants for rare diseases, overcoming the limitations of linkage studies in extremely rare or heterogeneous disorders, small families, and sporadic conditions (*Ku et al., 2011*). The number of distinct novel disease-causing genes identified thanks to WES is continuously increasing and, at the same time, this technique has allowed to expand the phenotypic spectrum of a large number of known genes (*Boycott et al., 2013; Zhang et al., 2014*). In some cases, WES has allowed to detect a disease variant which had been missed by a previous genetic test due to limitations of the test or the technique (especially sensitivity) used. Indeed, in a case reported by Landoure et al., WES identified a pathogenic variant in TRPV4 which had been previously missed by Sanger sequencing due to a SNP in the primer (*Landoure et al., 2012*). The application of WES in clinical setting has been revealed to be a valuable tool in the hand of clinicians to improve the diagnostic yield in patients

without a definite diagnosis after common genetic tests. This was especially true for rare diseases, which are clinically unspecific or involve a large number of genes, such as pediatric polyneuropathies (over 70 genes) (*Harding et al., 2014; Williams et al., 2013*). The value of WES is especially high in cases when an overlap of symptoms between different syndromes occurs. In these circumstances, the list of the potential candidate genes is longer and the diagnosis is more difficult, but WES allows to obtain the results faster and at a lower cost than targeted or single-gene sequencing (performed in combination with clear clinical information and relevant testing results) (*Tetreault et al., 2015*). A peculiar group of disorders is the subgroup of rare diseases caused by mutations in the mitochondrial genome (*McCormick et al., 2013*). Based on the clinical and genetic heterogeneity, patients with mitochondrial disorders could potentially be good candidates for the application of clinical WES. Despite the good results obtained by some research groups, the use of WES in mitochondrial disorders still has limits and it is not enough robust to be used in clinical practice (*Calabrese et al., 2014; Dinwiddie et al., 2013; Wong, 2013*). In particular, sequence homologs between nuclear and mitochondrial genome create a significant background which makes it difficult to discern the true mitochondrial sequence and tissue specificity of mitochondrial variants can also be a limitation in the analysis of WES data (DNA extracted from blood is usually sequenced whereas mitochondrial diseases mainly affect muscle and brain, so that the variant might not be detected unless it is homoplasmic or have a high heteroplasmy level in blood) (*Wong, 2013*). Therefore, the proposed methodology in this disorder is whole mitochondrial DNA sequencing where the mitochondrial genome is specifically captured and sequenced (*McCormick et al., 2013*).

2.3.2 *Cancer*

Genomic alterations such as point mutations, CNVs, and structural rearrangements play a pivotal

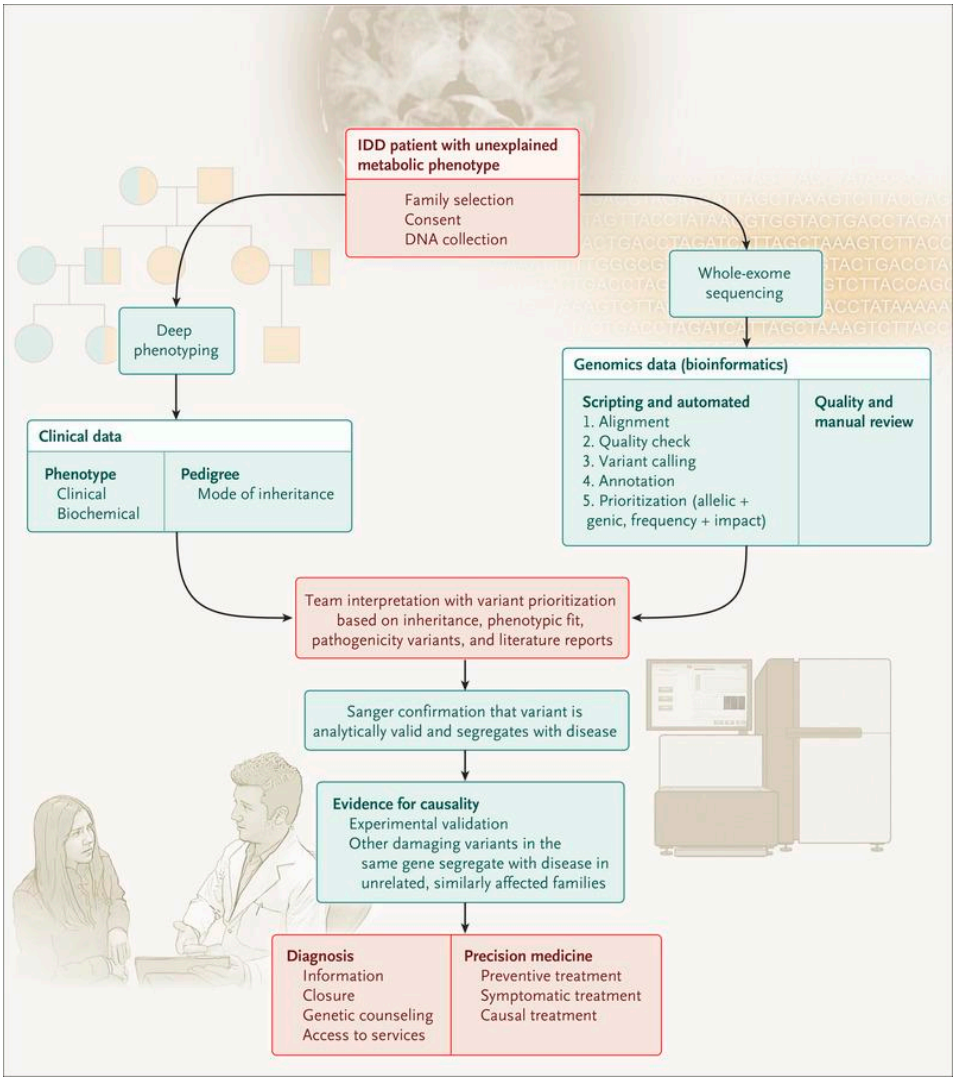
role in the pathogenesis of cancer (*Mardis et al., 2009*). In this field, WES has relevantly contributed to the identification of numerous cancer susceptibility genes and has made possible the detection of somatic driver mutations in sporadic cancers, allowing the identification of novel driver genes and pathways (e.g., chromatin remodeling in pediatric glioblastoma, pre-mRNA splicing in myelodysplasia, DNA methylation in acute monocytic leukemia) (*Schwartzentruber et al., 2012; Wang et al., 2014; Yan et al., 2011; Yoshida et al., 2011*). In clinical setting, WES allows to test all candidate genes for a specific genetically heterogeneous condition in a single experiment (e.g., more than 10 genes in case of patients with multiple colorectal polyps), obtaining a diagnosis more rapidly than single-gene testing (*Vasen et al., 2015*). WES has also been studied as screening tool for the identification of at-risk individuals for selected type of cancers to increase the number of patients diagnosed during the early stages of cancer, with a potentially relevant role in cancer prevention (*Genovese et al., 2014*). The incorporation of WES in clinical practice has also allowed to improve patients' management and guide treatment choice, as brilliantly shown by the advancements achieved in personalized medicine (*Parsons et al., 2014*). Indeed, thanks to WES, it is possible to perform a multiplex analysis of the cancer biomarkers and determine the tumor's mutational profile for each patient, which strongly influences the response to a specific treatment for a specific cancer subtype (*Catenacci, 2014*). The genetic information can be further used to develop pre-defined treatment priority algorithms allowing clinicians to determine the best therapy for the single patient (*Catenacci, 2014*).

2.3.3 *Complex disorders*

Complex disorders are conditions caused by a combination of many genetic and environmental factors, making them difficult to study. These conditions include autism, intellectual disabilities (ID), and other relevant neurological disorders (such as Alzheimer's disease) (*Jiang et al., 2014*).

The mode of inheritance is usually unclear, although some of them can cluster in families. Genome- wide association studies (GWAS) have been used to study these disorders in research setting, based the common variant/common disease hypothesis (*Jiang et al., 2014*). Thanks to GWAS, a large number of susceptibility alleles have been identified, but each variant only confers a small increase in risk with limited impact in the single patient and most of these variants only explain a small proportion of familial cases (*Frebourg, 2014; Manolio et al., 2009*). WES has been more recently used to find rare genetic variations with a strong effect, which have been hypothesized to explain the missing heritability resulting from GWAS. This approach is particularly interesting in the determination of the contribution of rare variants, especially in monogenic disorders with a clear mode of inheritance resembling Mendelian diseases (*Jiang et al., 2014; Marian et al., 2012*). Indeed, the identification of possibly causative genes in this disease subgroup has allowed the detection of mutations in the same gene or genes from the same pathway in patients with complex disorders. Accordingly, *de novo* mutations are a frequent cause of neuropsychiatric diseases and trio-based WES has led to the identification of rare alleles associated with sporadic diseases such as autism and ID (*Ku et al., 2011; Kuechler et al., 2015; Rabbani et al., 2014*). Most of the strong variants causing complex disorders are found in coding regions and therefore can be easily identified through WES (without requiring WGS). Their identification may play a relevant role in the improvement of the knowledge available on disorders affecting a large proportion of general population, with possible relevant impact on diagnosis and patient management (*Tetreault et al., 2015*). With regard to the use in clinical setting, WES results very useful in the diagnosis of common monogenic diseases, such as some forms of diabetes mellitus (maturity-onset diabetes of the young, MODY) and hypercholesterolemia (*Bonnefond et al., 2010; Repas et al., 2014*). Despite the application of

WES for clinical diagnosis of complex diseases has proved to be suitable in some type of disorders, the interpretation of the variants can be challenging. Indeed, in many cases, the identified variant is only conferring a small risk and multiple susceptibility alleles are involved (*Tetreault et al., 2015*).



Flow Chart Showing Gene-Discovery Approach with the Use of Collaborative Phenomics and Semiautomated Genomics. intellectual developmental disorder (IDD). *Adapted from Tarailo-Graovac et al., 2016.*

2.3.4 Focus on research implications

In the following sections, we will focus on three of the main research fields in which the use of WES reveals especially crucial. The genetic information obtained from the sequencing of patients with neurodevelopmental conditions may indeed be the starting point for different translational and medical research strategies:

A. Pathophysiology

This approach implies the investigation of the pathophysiological mechanisms underlying a specific genetic condition, with the purpose of expanding the current knowledge in this field towards the implementation of patient management and the potential development of novel targeted therapies. Selected variants in a specific gene of interest can be investigated through *in vitro* and/or *in vivo* experimental procedure, leading at the delineation of the mechanisms of the molecular machinery impaired in the studied subjects and the definition of genotype-phenotype correlation. The functional information obtained can be also helpful to predict the potential impact of different variants affecting the same or similar regions within the protein.

B. Gene discovery

Aiming at the identification of novel genes to be associated with medical conditions with unknown aetiology, the gene discovery field relies on WES as the most powerful tool to detect potentially interesting variants in non-morbid genes. Subsequently, these variants are functionally characterized through *in vitro* and/or *in vivo* experiments to establish their pathogenic potential. The functional results are eventually reviewed in relation to the clinical

phenotype observed in the patients, in order to predict the possible pathophysiological link between the novel disease gene and the medical condition described in the study cohort.

C. Case series study: natural history

The identification of causative variants in patients with rare conditions of known genetic aetiology may reveal helpful to conduct natural history studies. This may be especially relevant for disorders that are poorly characterized in terms of clinical phenotype, response to specific therapies, overall outcome, and availability of outcome predictors or diagnostic markers. With the use of statistical analyses, the case series study approach allows to depict the pattern followed by a selected disease of interest in its natural course. From a clinical perspective, this information may be extremely helpful in the case of very rare genetic disorders, paving the way to an improvement of several aspects of patient management: 1) improved clinical diagnostic sensibility; 2) enhanced awareness of clinical/molecular/biochemical markers and/or predictors of outcome; 3) revision of therapeutic strategies in light of the drug responses observed in similar patients; 4) optimization of the follow-up schedule and assessments; 5) adoption of symptomatic treatment strategies based on the benefits observed in other patients.

D. Case series study: genotype-phenotype correlations

These studies involve the collection of patients harboring heterogeneous variants in a specific gene of interest to define the genotype and phenotype spectrum of a disorder, with the purpose of establishing potential links between genetic changes, functional consequences, observed phenotypes, and (when possible) response to specific treatment strategies. Therefore, the main aim of this research approach is to refine the genotype-phenotype correlations of a certain

medical condition. This information may be very helpful in the clinical management of patients, especially in relation to: 1) improvement of follow-up strategies based on the genetic background of the subject; 2) implementation of targeted therapies (precision medicine); 3) significant enhancement of genetic counseling, with beneficial impact on family planning and support.

References

Abecasis G.R., et al. An integrated map of genetic variation from 1,092 human genomes. Nature, 2012. 491(7422):56-65.

Adzhubei I.A., et al. A method and server for predicting damaging missense mutations. Nat Methods, 2010. 7(4):248-249.

Adzhubei I., et al. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet, 2013. Chapter 7;Unit 7.20.

Albert T.J., et al. Direct selection of human genomic loci by microarray hybridization. Nat Methods, 2007. 4:903-905.

Alkodsi A., et al. Comparative analysis of methods for identifying somatic copy number alterations from deep sequencing data. Brief Bioinformatics, 2014. 16(2):242-54.

Amberger J.S., et al. OMIM.org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic. Nucleic Acids Res, 2015. 43(Database issue):D789-798.

American Psychiatric Association. DSM 5 Am J Psychiatry, 2013. [Internet].
Biesecker L.G., et al. Diagnostic clinical genome and exome sequencing. N Engl J Med, 2014. 371(12):1170.

Bonnefond A., et al. Molecular diagnosis of neonatal diabetes mellitus using next-generation sequencing of the whole exome. PLoS One, 2010. 5(10): e13630.

Boycott K.M., et al. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet, 2013. 14(10): 681-691

Calabrese C., et al. MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. Bioinformatics, 2014. 30(21):3115-3117.

Catenacci D.V.T. Next-generation clinical trials: Novel strategies to address the challenge of tumor molecular heterogeneity. Mol Oncol, 2014. Epub ahead of print.

Church D.M., et al. Public data archives for genomic structural variation. Nat Genet, 2010. 42(10): 813-814.

Cingolani P., et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 2012. 6(2): 80-92.

Davydov E.V., et al. Identifying a high fraction of the human genome to be under selective constraint using

Dinwiddie D.L., et al. Diagnosis of mitochondrial disorders by concomitant next-generation sequencing of the exome and mitochondrial genome. Genomics, 2013. 102(3):148-1456.

Firth H.V., et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet, 2009. 84(4):524-533.

Frebourg T. The challenge for the next generation of medical geneticists. Hum Mutat, 2014. 35(8):909-911.

Fu W., et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature, 2013. 493(7431):216-20.

Genovese G., et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med, 2014. 371(26):2477-2487.

GERP++. PLoS Comput Biol, 2010. 6(12):e1001025.

Gnirke A., et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol, 2009. 27:182-189.

Harding K.E., et al. Applications of next-generation whole exome sequencing. J Neurol, 2014. 261(6):1244-1246.

Hodges E., et al. Genome-wide in situ exon capture for selective resequencing. Nat Genet, 2007. 39:1522-1527.

Ionita-Laza I., et al. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics, 2009. 93(1):22-26.

Jiang T., et al. Application of next-generation sequencing technologies in Neurology. Ann Transl Med, 2014. 2(12):125.

Kadalayil L., et al. Exome sequence read depth methods for identifying copy number changes. Brief Bioinformatics, 2015. 16(3):380-392.

Kinsbourne M. Disorders of mental development. In: Menkes J, Sarnat B, Maria B, editors. Child neurology. 6th edition. Baltimore (MD): Lippincott Williams & Wilkins, 2001. p. 1155-1211.

Koboldt D.C., et al. The next-generation sequencing revolution and its impact on genomics. Cell, 2013. 26;155(1):27-38.

Ku C-S., et al. Revisiting Mendelian disorders through exome sequencing. Hum Genet, 2011. 129(4): 351-370.

Kuechler A., et al. De novo mutations in beta-catenin (CTNNB1) appear to be a frequent cause of intellectual disability: expanding the mutational and clinical spectrum. Hum Genet, 2015. 134(1):97-109.

Landoure G., et al. Exome sequencing identifies a novel TRPV4 mutation in a CMT2C family. Neurology, 2012. 79(2):192-194.

Landrum M.J., et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res, 2014. 42: Database issue D980-985.

Lee K., et al. Intellectual Disability. Source StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing, 2019.

Li H., et al. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 2009. 25(14):1754-1760.

Li H., et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 2009. 25(16): 2078-2079.

Li H., et al. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res, 2008. 18(11):1851-1858.

Manolio T.A., et al. Finding the missing heritability of complex diseases. Nature, 2009. 461(7265):747-753.

Mardis E.R., et al. Cancer genome sequencing: a review. Hum Mol Genet, 2009. 18(R2):R163-168.

Mardis E.R. Next-generation sequencing platforms. Annu Rev Anal Chem, 2013. 6:287-303.

Margulies E.H., et al. Identification and characterization of multi-species conserved sequences. Genome Res, 2003. 13(12):2507-2518.

Marian A.J. Challenges in medical applications of whole exome/genome sequencing discoveries. Trends Cardiovasc Med, 2012. 22(8):219-223.

Massingham T., et al. All Your Base: a fast and accurate probabilistic approach to base calling. Genome Biol, 2012. 13(2):R13.

McCarroll S.A., et al. Copy-number variation and association studies of human disease. Nat Genet, 2007. 39(7 Suppl):S37-42.

McCormick E., et al. Molecular genetic testing for mitochondrial disease: from one generation to the next. Neurotherapeutics, 2013. 10(2):251-261.

McKenna A., et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res, 2010. 20(9):1297-1303.

Milne S.L., et al. Adaptive function in preschoolers in relation to developmental delay and diagnosis of autism spectrum disorders: insights from a clinical sample. Autism, 2013. 17(6):743-753.

Miyatake S., et al. Detecting copy-number variations in whole-exome sequencing data using the eXome Hidden Markov Model: an 'exome-first' approach. J Hum Genet, 2015. 60(4):175-182.

Nadaf J., et al. Detection of recurrent allelic imbalance in tumors using whole-exome sequencing data. Bioinformatics, 2015. 31(3): 429-431.

Ng P.C., et al. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res, 2003. 31(13): 3812-3814.

Obi O., et al. Effect of incorporating adaptive functioning scores on the prevalence of intellectual disability. Am J Intellect Dev Disabil, 2011. 116(5):360-370.

O'Rawe J., et al. Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. Genome Med, 2013. 5(3):28.

Parsons D.W., et al. Evaluating the implementation and utility of clinical tumor exome sequencing in the pediatric oncology clinic: Early results of the BASIC3 study. Cancer Res, 2014. 74(20 Suppl):IA16-16.

Pollard K.S., et al. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res, 2010. 20(1): 110-121.

Rabbani B., et al. The promise of whole-exome sequencing in medical genetics. J Hum Genet, 2014. 59(1):5-15.

Redon R., et al. Global variation in copy number in the human genome. Nature, 2006. 444(7118): 444-454.

Repas T.B., et al. Preventing early cardiovascular death in patients with familial hypercholesterolemia. J Am Osteopath Assoc, 2014. 114(2):99-108.

Richards S., et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med, 2015. 17(5): p. 405-424.

Richter T., et al. Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value Health, 2015. 18(6):906-914.

Rosenfeld J.A., et al. Limitations of the human reference genome for personalized genomics. PLoS One, 2012. 7(7):e40294.

Schwartzentruber J., et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature, 2012. 482(7384): 226-231.

Sharp A.J., et al. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet, 2005. 77(1):78-88.

Sherry S.T., et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res, 2001. 29(1): 308-311.

Shevell M. Global developmental delay and mental retardation or intellectual disability: conceptualization, evaluation, and etiology. Pediatr Clin North Am, 2008. 55(5):1071-1084, xi.

Stenson P.D., et al. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum Genet, 2014. 133(1):1-9.

Tan R., et al. An evaluation of copy number variation detection tools from whole-exome sequencing data. Hum Mutat, 2014. 35(7):899-907.

Tarailo-Graovac M., et al. Exome Sequencing and the Management of Neurometabolic Disorders. N Engl J Med, 2016. 374(23):2246-2255.

Tetreault M., et al. Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities. Expert Rev Mol Diagn, 2015. 15(6):749-760.

Thorvaldsdottir H., et al. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinformatics, 2013. 14(2):178-192.

Vasen H.F.A., et al. Clinical management of hereditary colorectal cancer syndromes. Nat Rev Gastroenterol Hepatol, 2015. 12:88-97.

Wang K., et al. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res, 2010. 38(16):e164.

Wang L., et al. Genomic sequencing for cancer diagnosis and therapy. Annu Rev Med, 2014. 65:33-48.

Webb A.J., et al. An informatics project and online 'Knowledge Centre' supporting modern genotype-to-phenotype research. Hum Mutat, 2011. 32(5):543-550.

Weinstein J.N., et al. Cancer Genome Atlas Research Network. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet, 2013. 45(10):1113-1120.

Williams E.S., et al. Implementing genomic medicine in pathology. Adv Anat Pathol, 2013. 20(4):238-244.

Wong L.-J.C. Next generation molecular diagnosis of mitochondrial disorders. Mitochondrion, 2013. 13(4):379-387.

Wong K.-K., et al. Allelic imbalance analysis by high-density single-nucleotide polymorphic allele (SNP) array with whole genome amplified DNA. Nucleic Acids Res, 2004. 32(9):e69.

Xue Y., et al. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next- generation sequencing: single-gene, gene panel, or exome/genome sequencing. Genet Med, 2015. 17(6):444-451.

Yan X.-J., et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. Nat Genet, 2011. 43(4):309-315.

Yang Y., et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA, 2014. 312(18):1870-1879.

Yoshida K., et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature, 2011. 478(7367):64-69.

Zhang X. Exome sequencing greatly expedites the progressive research of Mendelian diseases. Front Med, 2014. 8(1): 42-57.

A. Pathophysiology:

Variant-specific changes in RAC3 function disrupt corticogenesis in neurodevelopmental phenotypes

Abstract

Variants in *RAC3*, encoding a small GTPase RAC3 which is critical for the regulation of actin cytoskeleton and intracellular signal transduction, are associated with a rare neurodevelopmental disorder with structural brain anomalies and facial dysmorphism. We investigated a cohort of 10 unrelated participants presenting with global psychomotor delay, hypotonia, behavioral disturbances, stereotyped movements, dysmorphic features, seizures, and musculoskeletal abnormalities. Magnetic resonance imaging of brain revealed a complex pattern of variable brain malformations, including callosal abnormalities, white matter thinning, gray matter heterotopia, polymicrogyria/dysgyria, brainstem anomalies, and cerebellar dysplasia. These patients harbored eight distinct *de novo RAC3* variants, including six novel variants (NM_005052.3): c.34G>C p.G12R, c.179G>A p.G60D, c.186_188delGGA p.E62del, c.187G>A p.D63N, c.191A>G p.Y64C, and c.348G>C p.K116N. We then examined the pathophysiological significance of these novel and previously reported pathogenic variants p.P29L, p.P34R, p.A59G, p.Q61L, and p.E62K. *In vitro* analyses revealed that all tested RAC3 variants were biochemically and biologically active to variable extent, and exhibited a spectrum of different affinities to downstream effectors including p21-activated kinase 1. We then focused on the four variants p.Q61L, p.E62del, p.D63N, and p.Y64C, in the Switch II region, which is essential for the biochemical activity of small GTPases and also a variation hot spot common to other Rho family genes, *RAC1* and *CDC42*.

Acute expression of the four variants in embryonic mouse brain using *in utero* electroporation caused defects in cortical neuron morphology and migration ending up with cluster formation during corticogenesis. Notably, defective migration by p.E62del, p.D63N, and p.Y64C were rescued by a dominant negative version of p21-activated kinase 1. Our results indicate that *RAC3* variants result in morphological and functional defects in cortical neurons during brain development through variant-specific mechanisms, eventually leading to heterogeneous neurodevelopmental phenotypes.

1. Introduction

1.1. Rho GTPases

Rho family guanosine triphosphatases (GTPases) are key regulators of the cellular signaling and cytoskeletal dynamics, which influence cell adhesion, morphology, migration, and cell cycle progression, and play crucial roles in various cell types (*Stankiewicz et al., 2014*). The Rac subfamily consists of three proteins, Rac1-3, which share about 90% identity in their amino acid (aa) sequences, with the largest degree of divergence at their carboxy-terminal ends. While Rac1 is ubiquitously expressed in most organs and Rac2 is specifically expressed in hematopoietic cells, Rac3 is abundantly expressed both in the developing and adult nervous system (*Moll et al., 1991; Shirsat et al., 1990; Malosio et al., 1997; Haataja et al., 1997*). Accordingly, Rac3 has been reported to participate in different aspects of neuronal development, such as neuritogenesis, axon and dendrites formation, synaptogenesis, and regulation of migration (*Albertinazzi et al., 1997; Bolis et al., 2003; Corbetta et al., 2009; Pennucci et al., 2011; Vaghi et al., 2014*).

Like other small GTPases, Rac3 is thought to act as a molecular switch cycling between GTP-bound (active) and GDP-bound (inactive) states via conformational changes in the Switch I

and Switch II regions, which are highly conserved in small GTPases. The Switch II region determines the selective interaction with functionally diverse guanine nucleotide-exchange factors (GEFs) and GTPase-activating proteins (GAPs), which are the major classes of regulators controlling the guanine nucleotide-binding state (*Dvorsky et al., 2004*). GEFs promote GTP-loading to activate small GTPases, while GAPs stimulate intrinsic GTP-hydrolysis to inactivate them. In addition to GEFs and GAPs, the Switch I region interacts with downstream effector molecules whose activities are strictly regulated in a spatio-temporal manner by activated small GTPases. Although Rac1 and Rac3 share common downstream effectors, their distinct expression profiles strongly suggest isoform-specific functions.

1.1.a. Rho GTPases in human disease

Accumulating evidence supports critical roles for Rho family GTPases in neural development (*Stankiewicz et al, 2014; Azzarelli et al., 2015; de Curtis, 2019*), and indicates that the disruption of Rho GTPase signaling significantly contributes to the pathogenesis of neurodevelopmental disorders (NDDs) in humans (*Guo et al., 2020; Zamboni et al., 2018*). In the Simons Foundation Autism Research Initiative (SFARI) database, 7/82 RhoGEFs (9 %), 7/57 RhoGAPs (12 %), and 6/73 downstream effectors (8 %) are categorized as autism spectrum disorders (ASD) candidate genes (*Guo et al., 2020*). At least, 1.6 % of all 1,231 listed ASD candidate genes are therefore directly involved in the signaling pathways of Rho family GTPases. More specifically, the link between the abnormal function of Rho GTPases *per se* and NDDs has been recently highlighted as follows. As for the Rac subfamily, *de novo* missense *RAC1* variants affecting protein function have been associated with heterogeneous conditions characterized by variable psychomotor delays

and brain malformations, while variations in *RAC2* have been identified in patients with different forms of primary immunodeficiencies (*Reijnders et al., 2017; Lougaris et al., 2020*).

1.1.b. Rac3 in human disease

Recently, a novel NDD with structural brain anomalies and dysmorphic facies (NEDBAF, OMIM #618577) has been reportedly in association with *RAC3* variations (*Costain et al., 2019; Hiraide et al., 2019*). This condition is characterized by global developmental delay (GDD), intellectual disability (ID), abnormal muscle tone, heterogeneous neurological phenotypes, and structural brain abnormalities (*Costain et al., 2019; Hiraide et al., 201*). Although the reported variants appear to hamper RAC3's basic functions, such as actin cytoskeletal organization and signal transduction, the exact molecular mechanisms involved in the pathogenesis of NEDBAF remain to be elucidated.

1.2 Study outline

In the present study, we investigated a cohort of 10 unrelated participants presenting with a syndromic NDD and peculiar brain malformations, including one partially reported case (White et al., 2018). We identified six novel *de novo* variants in *RAC3*, including five missense and one in-frame deletion. *In vitro* analyses revealed that the novel and previously reported 11 variants were biochemically and biologically active to variable extent and seemed to interact with various downstream effectors in variant-specific manners. *In vivo* analyses revealed that the four variants in the Switch II region caused morphological and migration defects of excitatory neurons during corticogenesis, with the involvement of a functional dysregulation of the downstream effector PAK1 in three out of four cases. In addition, the four variants prevented axonal development *in*

vivo. Collectively, RAC3 was found to play a crucial role in brain development and pathogenic variants affecting its function are most likely to cause impaired corticogenesis, leading to the neurodevelopmental phenotypes observed in patients with NEDBAF associated with *RAC3* abnormalities.

2. Materials and methods

2.1. Clinical study

2.1.a. Ethics statement

The study involving human participants was conducted according to the guidelines of the Declaration of Helsinki. Ethical review and approval were obtained locally and informed consent was obtained from the parents or legal guardians of all the enrolled participants. For the animal experiments, we followed the fundamental guidelines for proper conduct of animal experiments and related activity in academic research institutions under the jurisdiction of the Ministry of Education, Culture, Sports, Science, and Technology (Japan). All protocols for animal handling and treatment were reviewed and approved by the Animal Care and Use Committee of Institute for Developmental Research, Aichi Developmental Disability Center (approval number: 2019-013).

2.1.b. Patient enrolment and ascertainment

Patients were recruited through international collaboration, also using GeneMatcher (*Sobreira et al., 2015*), from several clinical and research centers in Canada, France, Germany, Italy, Spain, and the USA: Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'interrégion Est, Hôpital d'Enfants, CHU de Dijon, France;

Department of Human Genetics, Emory Healthcare, USA; Departments of Human Genetics, Pediatrics and Psychiatry, David Geffen School of Medicine at UCLA, USA; Department of Pediatrics, University of California San Diego, USA; Genetics/Dysmorphology, Rady Children's Hospital San Diego, USA; Dipartimento Universitario Scienze della Vita e Sanità Pubblica, Università Cattolica Sacro Cuore, Rome, Italy; Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, Canada; Genetica Medica, Fondazione Policlinico Universitario A. Gemelli IRCCS, Italy; Neuropediatric Department Helios-Klinikum Hildesheim Hildesheim, Germany; Parc Taulí Hospital Universitari, Institut d'Investigació i Innovació Parc Taulí I3PT, Universitat Autònoma de Barcelona, Spain; Unit of Child Neuropsychiatry, IRCCS Istituto Giannina Gaslini, Italy.

Patients were evaluated by pediatricians with expertise in neurological disorders, pediatric neurologists, and geneticists with expertise in the field of neurogenetics. Vineland Adaptive Behavior Scale (VABS, third edition) tests were also administered for adaptive behavior assessment to Patients #4, #5, and #8. The overall level of adaptive functioning is described by the score on the Adaptive Behavior Composite (ABC), based on scores for three specific adaptive behavior domains: Communication, Daily Living Skills, and Socialization. The domain scores are also expressed as standard scores with a mean of 100 and standard deviation of 15. The Communication domain measures how well the patient listens and understands, expresses him/herself through speech, and reads and writes. The Daily Living Skills domain assesses the patient’s performance of the practical, everyday tasks of living that are appropriate for his/her age. The individual’s score for the Socialization domain reflects his/her functioning in social situations.

Clinical and molecular data of Patient #6, who was partially described in a previous report (*White et al., 2018*), were critically reviewed. Brain MRI scans, locally performed for patient care,

were reviewed by an expert pediatric neuroradiologist (MS). Previously published neuroimaging studies were also thoroughly reviewed for comparison and delineation of the neuroradiological spectrum.

2.1.c. Previously reported cases assessment

All the articles indexed in PubMed (https://pubmed.ncbi.nlm.nih.gov/?term=RAC3&sort=date) between April 2019, when *RAC3* variants were first associated with NEDBAF by Costain *et al*. (2019) (*Costain et al., 2019*), and May 2021 (*Pellegatta et al., 2021*) were retrieved using the terms "RAC3", "Rac3 GTPase", and "neurodevelopmental syndrome". All the articles concerning the molecular, clinical, and neuroradiological spectrum associated with NEDBAF were thoroughly reviewed. Inclusion criteria for previously published patients included clinical data availability, unambiguous identification of pathogenic/likely pathogenic *RAC3* variants, and lack of duplication from other previous reports. Ambiguous clinical presentation not consistent with NEDBAF and inconclusive genetic testing were exclusion criteria.

2.2. Genetic investigation

2.2.a. Outline of genetic testing

We investigated 10 individuals presenting with syndromic NDD and neuroimaging abnormalities. Different Next Generation Sequencing (NGS) approaches were employed, including trio-exome sequencing (Trio-ES) (Patients #1-3, #8, and #10), singleton ES (#4, #6, and #9), trio-genome sequencing and RNA sequencing (#5), and a NGS panel including 480 ID genes (#7). Sanger sequencing was performed for the validation of the candidate variants and the parental segregation analysis. The presence of copy number variants (CNVs) was investigated in all subjects, either

through array comparative genomic hybridization (aCGH) (#1, #2, #4, #5, #6, #7, #8, #9, and #10) or CNV detection on ES data (#3).

2.2.b. Next Generation Sequencing: variant detection and annotation

Genomic DNA was extracted from peripheral blood leukocytes using standard protocols. After standard DNA extraction, trio-ES (#1-3, #8, and #10), singleton exome sequencing (ES) (#4, #6, and #9), trio-genome sequencing and RNA sequencing (#5), or intellectual disability next-generation sequencing gene panel (480 genes) (#7) were performed. NGS panels and ES were carried out as previously described. QC statistics with FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) was used to assess the quality of the sequence reads. Reads alignment to the reference human genome (hg19, UCSC assembly, February 2009) was performed through BWA with default parameters. Recalibration of the quality score and for indel realignment and variant calling was performed through the HaplotypeCaller algorithm within the GATK package. Variants were annotated with ANNOVAR.

2.2.c. Next Generation Sequencing: variant analysis

The identified variants were filtered according to minor allele frequency ≤ 0.001 in Genome Aggregation Database (gnomAD, https://gnomad.broadinstitute.org) (*Lek et al., 2016*), conservation (Genomic Evolutionary Rate Profiling – GERP, http://mendel.stanford.edu/SidowLab/downloads/gerp/), and predicted impact on protein function. *In silico* prediction tools were used for the interpretation of candidate variants, including Combined Annotation Dependent Depletion (CADD, https://cadd.gs.washington.edu), Mutation Taster (http://www.mutationtaster.org), Sorting Intolerant From Tolerant (SIFT, https://sift.bii.a-

star.edu.sg), and Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/). Ultimately, the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines were used to classify the candidate variants based on their predicted pathogenicity (*Richards et al., 2015*). All *RAC3* variants are reported according to the NM_005052.3 transcript (NP_005043.1) (https://www.ncbi.nlm.nih.gov/nuccore/NM_005052.3).

2.2.d. Array Comparative Genomic Hybridization

Array Comparative Genomic Hybridization (Array-CGH) was performed in #1, #2, #4, #5, #6, #7, #8, and #10 as previously described (*Redon et al., 2009*) and the detected rearrangements were interpreted according to the Decipher database (https://decipher.sanger.ac.uk). In patient #3, CNVs analysis was performed using XHMM software, which uses principal component analysis (PCA) normalization and a hidden Markov model to genotype CNVs from normalized read-depth information (*Fromer et al., 2012*). Depth of coverage was calculated with GATK (Genome Analysis Toolkit, v3.5), and a matrix of the mean per-exon (RefSeq n = 193342) depth of coverage across all samples (n = 846) was obtained. The target read depths were first mean-centered and PCA was used on the read depth matrix to identify the components with high variance. These components, which explain most of the variation due to systematic noise, were subtracted from the matrix to obtain the normalized read depth matrix. In order to scale target depth values, XHMM used a z-score calculation to transform normalized read depths for each sample. The z-scores were then used to call CNVs with a hidden Markov model (HMM) algorithm to determine deletion or duplication regions based on below- average and above-average read depth. These calculations take into account the distance between exome targets, CNV rates and length distribution. CNVs were then filtered using the following criteria: estimated CNV length ≥1kb, XHMM quality score

(SQ) ≥65, exons span >1, minor allele frequency (MAF) ≤1%. CNVs were annotated using RefSeq, (https://www.ncbi.nlm.nih.gov/refseq/), DGV17 and DECIPHER18 databases.

2.3. Molecular studies

2.3.a. Plasmids

RAC3 was amplified by PCR from a cDNA pool of human glioblastoma U251MG cells and cloned into pCAG-Myc vector. The 11 variants, *RAC3*-G12R, -P29L, -P34R, -A59G, -G60D, -Q61L, -E62del, -E62K, -D63N, -Y64C, and -K116N, were generated by site-directed mutagenesis using KOD-Plus Mutagenesis kit (Toyobo Inc., Osaka, Japan) with pCAG-Myc-RAC3 as a template. *RAC3* and the 11 variants were also constructed into pTriEx-4 vector (Merck, Darmstadt, Germany). pRK5-Myc-PAK1 (p21-activated kinase 1)-KA, a kinase-negative variant with a single amino acid substitution p.K299A, and pRK5-Myc-MLK2 (mixed lineage kinase 2/MAPKKK10)-KN, a kinase-negative variant lacking aa 139 - 183, were kindly provided by Prof. Alan Hall (Univ. College London, UK), and subcloned into pCAG-Flag vector. RAC-binding regions (RBRs) in human *PAK1* (aa 67-150), human *MLK2* (aa 401-550), human *IRSp53* (aa 2-228), rat *N-WASP* (aa 191-270), mouse *RTKN* (Rhotekin) (aa 10-100), and human *ROCK* (Rho-associated coiled-coil containing protein kinase 1) (aa 67-150) were amplified by PCR from a cDNA pool of U251MG cells, rat brain or mouse brain, and constructed into pGS21a vector (GenScript, Piscataway, NJ). For gene transcription analysis, pGL4.74[hRluc/TK] (control reporter plasmid), pGL4.44[luc2P/AP1-RE/Hygro] (AP1-luciferase reporter plasmid), and pGL4.32[luc2P/NF-κB-RE/Hygro] (NFkB-luciferase reporter plasmid) were purchased by Promega (Madison, WI). The SRF reporter element, 5'-CAGACAGACGTGTTCTTAAGTCCATATTAGGACATCTACCATGTCCATATTAGGACA

TCTACTATGTCCATATTAGGACATCTTGTATGTCCATATTAGGACATCTAAAATGTCC ATATTAGGAC-3', was inserted into pGVB (Toyo, Inc., Tokyo, Japan) to yield SRF-luciferase reporter plasmid. All constructs were verified by DNA sequencing.

2.3.b. Antibodies and histochemical reagents

The following antibodies were used: anti-GFP (Medical & Biological Laboratories, Nagoya, Japan, Cat# 598, RRID: AB_591819 or Nacalai Tesque, Kyoto, Japan, Cat# 04404-84, RRID: AB_10013361), anti-NeuN (Millipore, Temecula, CA, Cat# MAB377, RRID:AB_2298772), and anti-Myc (Medical & Biological Laboratories, Nagoya, Japan, Cat# M047-3, RRID: AB_591112). Additional employed antibodies included: anti-Rac3 (Novus Biologicals, Littleton, CO, Cat# NBP2-32058), anti-PAK1 (Cell Signaling Technology Cat# 2602, RRID:AB_330222), anti-phospho-PAK1 (Ser199/204) (Cell Signaling Technology Cat# 2605, RRID:AB_2160222), and anti-phospho-PAK1 (Thr423) (Cell Signaling Technology Cat# 2601, RRID:AB_330220).

Alexa Fluor 488 and 568 (Invitrogen, Carlsbad, CA, USA) were used as secondary antibodies. 4', 6-diamidino-2-phenylindole (DAPI; Nichirei Bioscience, Tokyo, Japan) was used for staining DNA. Rhodamine-phalloidin (Invitrogen, Carlsbad, CA) was used for staining filamentous actin.

2.3.c. GTP-hydrolysis and GTP/GDP-exchange assays

Preparation and purification of His-tag-fused RAC3 and the 11 variants were carried out according to the manufacturer's instructions. To assess the basal GTP/GDP-exchange reactions, release of methylanthraniloyl (mant)-GDP (Sigma-Aldrich, St Louis, MO) was measured (*Kanie et al., 2018*). Briefly, loading of RAC3 proteins with ^{mant}GDP was performed by incubation for 90 min

at 20°C in ^{mant}GDP loading buffer [20 mM HEPES-NaOH (pH 7.5), 50 mM NaCl, 0.5 mM $MgCl_2$, 5 mM EDTA, 1 mM dithiothreitol, 20-fold excess of ^{mant}GDP] and subsequently adding 10 mM $MgCl_2$ (final concentration) to stop the reaction. Remaining non-bound nucleotide was removed from the incubation mixture with a NAP-5 column (GE Healthcare Life Sciences, Buckinghamshire, England). ^{mant}GDP loaded GTPase was incubated with GppNHp (Abcam, Tokyo, Japan), which has 100 times higher concentration than that of RAC3 proteins, and fluorescence (365/450 nm) change was measured using an MTP-800 microplate reader (Corona, Ibaraki, Japan). Intrinsic GTP-hydrolysis activity was then undertaken by directly measuring changes in the GTP concentration with a luciferase-coupled GTPase assay kit (GTPase-Glo Assay Kit, Promega, Madison, WI) according to the manufacturer's instructions (*Mondal et al., 2015*). In short, 5 μM GTP and serially diluted RAC3 proteins were incubated in GTPase/GAP buffer for 60 min at room temperature. Once the GTP-hydrolysis reaction was completed, the remaining GTP was converted to ATP by adding GTPase-Glo Reagent. Subsequently, generated ATP was detected by the luciferase/luciferin-based reagent using a Turner Designs TD-20/20 luminometer (BMG Labtech, Saitama, Japan).

2.3.d. Cell culture and transfection

COS7 (monkey kidney fibroblast-like cell) and primary hippocampal neurons derived from embryonic day (E) 16.5 mice were cultured as described (*Nishikawa et al., 2021*). Transient transfection was carried out using polyethyleneimine "MAX" reagent (for COS7 cells) (Polysciences Inc., Warrington, PA) or Neon transfection system (for primary neurons) (Invitrogen, Carlsbad, CA).

2.3.e. Pull-down assay of GTP-bound RAC3 and the variants

Glutathione S-transferase (GST)-fused RBRs of PAK1, MLK2, IRSp53, N-WASP, RTKN, and ROCK were expressed in *Escherichia coli* BL21 (DE3) strain. Recombinant GST-proteins were purified according to the manufacturer's instructions. COS7 cells were transfected with pCAG-Myc-RAC3, RAC3-G12R, -P29L, -P34R, -A59G, -G60D, -Q61L, -E62del, -E62K, -D63N, -Y64C, or -K116N (1.0 μg/60 mm-dish). After 24 h, cells were lysed with the pull-down buffer (50 mM Tris-HCl, pH7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.1% SDS, 1% Nonidet P-40, and 0.5% deoxycholate), and insoluble materials were removed by centrifugation. The supernatant was then incubated for 30 min at 4℃ with Glutathione Sepharose 4B beads (GE Healthcare Life Sciences) with which GST-fused RBR of PAK1, MLK2, IRSp53, N-WASP, RTKN, or ROCK was bound. The beads were washed twice with the pull-down buffer, and bound proteins were analyzed by western blotting. Images were captured with LAS-4000 luminescent image analyzer (GE Healthcare Life Sciences).

2.3.f. Assay of SRF-, AP1- and NFkB-mediated gene transcription

Assays were performed as described (*Nishikawa et al., 2019*). The control reporter vector was co-transfected into COS7 cells seeded on 24-well plates with the indicated RAC3 expression plasmid (0.1 μg/well) together with the SRF-, AP1-, or NFkB-luciferase reporter plasmid. After transfection, cells were washed once with phosphate-buffered saline and lysed with the passive lysis buffer according to the manufacturer's instructions. Luciferase activities were determined with the dual-luciferase reporter assay system (Promega).

2.3.g. *In utero* electroporation

The surgery on pregnant mice and embryo manipulation in the uterus were performed as previously described (*Tabata et al., 2001; Hamada et al., 2018*). At E14, pregnant ICR mice provided by Japan SLC (Shizuoka, Japan) were deeply anesthetized with a mixture of three drugs: medetomidine (0.75 mg/kg), midazolam (4 mg/kg), and butorphanol (5 mg/kg) (*Kawai et al., 2011*). *In utero* electroporation was then performed in the specific-pathogen-free animal facilities as described (*Tabata et al., 2016*).[32]

Briefly, 1 μl of solution containing 0.1 μg of pCAG-Myc vector (control), pCAG-Myc-RAC3, -D63N, -E62del, -Y64C, or -Q61L was injected with pCAG-EGFP (0.5 μg) into the lateral ventricle of mouse embryos with a glass micropipette made from a microcapillary tube (GD-1; Narishige, Tokyo, Japan). After the embryo in the uterus was placed between the tweezers-type disc electrode (5 mm in diameter) (CUY650-5; NEPA Gene, Chiba, Japan), electronic pulses (50 ms of 35 V) were charged five times at 450 ms intervals with an electroporator (NEPA21; NEPA Gene). In this method, plasmids are introduced into the somatosensory area which is included in the parietal lobe. Brains were fixed at indicated postnatal days, sectioned, and then analyzed.

As for the quantification of distribution of GFP-positive cells in brain slices, coronal sections of cerebral cortices containing the labeled cells were classified into three bins (Fig. S1). The number of labeled cells in each region of at least three slices per brain was counted. All experimental procedures were carried out in the day time. Animals were neither excluded nor died during experimentation.

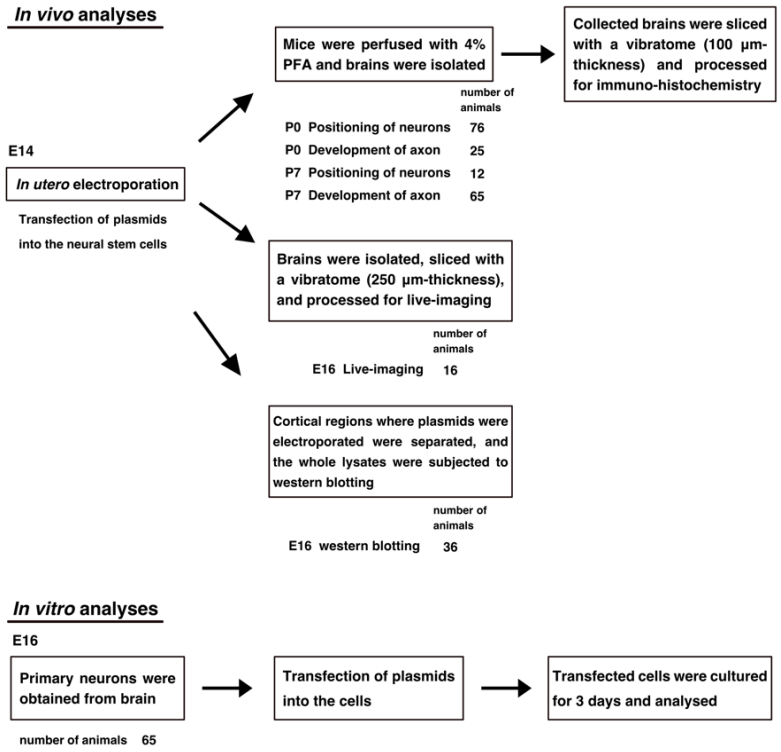


Figure S1. Graphical time-line of *in vivo* and *in vitro* analyses design.

2.3.h. Time-lapse imaging

After *in utero* electroporation at E14, organotypic coronal slices (250 µm-thickness) were prepared at E16 from the interventricular foramen with a microtome, placed on an insert membrane (pore size, 0.4 µm; Millipore, Bedford, MA), mounted in agarose gel, and cultured. The dishes were then mounted in an incubator chamber (5% CO_2 and 40%O_2, at 37 °C) fitted onto an FV1000 confocal laser microscope (Olympus, Tokyo, Japan), and the primary somatosensory cortex was examined as described (*Hamada et al., 2017*). Approximately 10 - 15 optical Z sections were acquired automatically every 15 min for 15 h, and about 10 focal planes (~50 µm-thickness) were merged to visualize the entire shape of the cells.

2.3.i. Quantitative analysis of axon growth

For estimation of axon growth, GFP signal intensity of the callosal axons was measured at P0 or P7 using imageJ software in distinct regions (bin 1-3 or bin 1-4). The relative intensities of bins were normalized with bin 1 as 1.0, and compared using R software.

2.3.j. Immunofluorescence

Immunofluorescence analysis was conducted essentially as described (*Nishikawa et al., 2021*). Images of cultured cells were captured with a BZ-9000 microscope (Keyence, Osaka, Japan) or an LSM-880 confocal laser microscope (Carl Zeiss, Oberkochen, Germany). As for the cortical slice staining, brains were embedded in 3% agarose, cut into sections (100 μm-thickness) with a vibratome, and photographed with an LSM-880 confocal laser microscope. Acquired images were analyzed with ImageJ to determine cell morphological descriptor and fluorescence intensity.

2.4. Statistical analysis

For all cell imaging experiments, cell counting and traces were assessed in a blinded manner by a technical staff of the lab who was not aware of experimental conditions. Statistical significance for multiple comparisons was determined by Dunnett's or Tukey's test (https://cran.r-project.org/web/packages/multcomp/multcomp.pdf). Comparisons between two groups were performed with Welch's t-test (*Welch et al., 1938*). $p < 0.05$ was considered statistically significant. Statistical analyses were performed using R (https://intro2r.com/citing-r.html; https://cran.r-project.org/doc/FAQ/R-FAQ.html#Citing-R).

3. Results

3.1. Phenotype dissection

3.1.a. Core clinical phenotype

The study cohort consisted of 10 patients presenting with a syndromic NDD characterized by a moderate to severe psychomotor delay affecting the achievement of all the developmental milestones and resulting in a global cognitive impairment (Table 1).

Table 1. Clinical and neuroimaging features of individuals with RAC3-related disorder

	Costain *et al.*	Hiraide *et al.*	This study	Total
Clinical features				
Gender	4 M, 1 F	1 M	8 F, 2 M	**M:F = 0.77**
Global DD/ID	+ (5/5)	+	+ (10/10)	**16/16 (100%)**
Hypotonia	+ (5/5)	-	+ (10/10)	**15/16 (93.7%)**
Dysmorphic features	+ (5/5)	-	+ (8/10)	**13/16 (81.3%)**
Dysphagia	+ (4/5)	+	+ (6/10)	**10/16 (62.5%)**
Failure to thrive	+ (4/5)	+	+ (6/10)	**10/16 (62.5%)**
Seizures	+ (2/5)	+	+ (5/10)	**7/16 (43.7%)**
Musculoskeletal features	+ (2/5)	-	+ (5/10)	**7/16 (43.7%)**
Dyspraxia	NA	NA	+ (4/10)	**4/10 (40%)**
Genitourinary abnormalities	+ (1/5)	+	+ (4/10)	**6/16 (37.5%)**
Abnormal cranial shape	+ (1/5)	-	+ (4/10)	**5/16 (31.2%)**
Respiratory problems	-	+	+ (4/10)	**5/16 (31.2%)**
Behavioral disorders	NA	-	+ (3/10)	**3/11 (27.3%)**
Stereotyped movements	NA	-	+ (3/10)	**3/11 (27.3%)**
Spasticity	-	+	+ (3/10)	**4/16 (25%)**
Eye abnormalities	-	-	+ (4/10)	**4/16 (25%)**
Endocrinological features	-	-	+ (3/10)	**3/16 (18.7%)**
Microcephaly	+ (1/5)	-	+ (1/10)	**2/16 (12.5%)**
Hearing loss	-	-	+ (1/10)	**1/16 (6.2%)**
Neuroimaging features				
CCH/CCA	+ (5/5)	+	+ (10/10)	16/16 (100%)
White matter thinning	+ (5/5)	+	+ (8/10)	14/16 (87.5%)
MCD	+ (5/5)	+	+ (6/10)	12/16 (75%)
Polymicrogyria	+ (3/5)	-	+ (5/10)	8/16 (50%)
Dysgyria	+ (2/5)	+	+ (6/10)	8/16 (50%)
Brainstem abnormalities	+ (3/5)	+	+ (4/10)	8/16 (50%)
Cerebellar dysplasia	-	+	+ (6/10)	7/16 (43.7%)
Gray matter heterotopia	+ (2/5)	-	+ (1/10)	3/16 (18.7%)

CCA= corpus callosum agenesis; CCH = corpus callosum hypoplasia; DD = developmental delay; F = female; ID = intellectual disability; M = male; MCD= Malformation of cortical development; NA = not available.

Facial dysmorphism was present in 8/10 individuals and recurrent features included wide forehead with frontal bossing and high anterior hairline, prominent eyes with upslanted palpebral fissures, arched eyebrows, long eyelashes, midface hypoplasia, broad nasal bridge, and anteverted nares (Fig. 1A). Gross motor development was impaired in all individuals, who also showed a significant deficiency in expressive language, with three patients (#1, #7, and #9) being nonverbal. A mild improvement of motor and verbal skills over time was common, although a regression limited to the phonological skills was observed in Patient #1. Behavioral abnormalities mainly consisted of ASD, aggressive/self-injurious behavior, and hypersensitivity. Feeding difficulties due to weak sucking or dysphagia were particularly relevant in the neonatal period, leading to failure to thrive in most individuals.

In Patients #4, #5, and #8, VABS administration revealed an overall low level of communication skills (scores range 34-68), daily living skills (scores range 27-56), and social skills and relationships (scores range 40-64). Patient #4 had an ABC of 59, which is well below the normative mean of 100, with a percentile rank for this overall score <1. Her Communication standard score was 55, corresponding to a percentile rank of <1. Her standard score for Daily Living Skills was 47, which corresponds to a percentile rank of <1 and represents a relative weakness for this individual. Her Socialization standard score was 64, corresponding to a percentile rank of 1 and representing a domain of relative strength for her. Patient #5 had an ABC score of 39, which is well below the normative mean of 100. The percentile rank for this overall score is <1. Her Communication standard score was 34, which corresponds to a percentile rank of <1. Her standard score for Daily Living Skills was 27, which corresponds to a percentile rank of <1. This domain is a relative weakness for her, whereas socialization is a relative strength for this individual, with a standard score of 40 (percentile rank of <1). Patient #8 had an ABC score of 63,

which is well below the normative mean of 100. The percentile rank for this overall score is 1. His Communication standard score was 68, corresponding to a percentile rank of 2 and representing a relative strength for him. His standard score for Daily Living Skills was 56, which corresponds to a percentile rank of <1. This domain is a relative weakness for him, such as his Socialization, due to a standard score of 58 (percentile rank of <1).

Neurological examination revealed hypotonia in all patients (Fig. 1B), frequently of neonatal onset, whereas appendicular spasticity was detected in only two patients (#2 and #7). Additional neurological features included dyspraxia with poor visuospatial coordination (#1-4), sleep disorders (#5 and #7), irritability (#6), and eyelid clonus (#9). Stereotyped movements of the mouth and upper limbs were present in three patients (#1, #2, and #9). Five out of 10 individuals (#1, #5, #6, #9, and #10) experienced recurrent seizures. Age at onset ranged from the first day of life to 10 months and semiology was variable, including infantile spasms, clonic, myoclonic, and tonic seizures. EEG showed focal/multifocal epileptic discharges, sometimes with secondary generalization. Hypsarrhythmia or slowed background were also observed in the most severely affected individuals (#1 and #5). In all individuals, seizures were responsive to antiseizure medications, e.g. valproate, vigabatrin, phenobarbital, or levetiracetam.

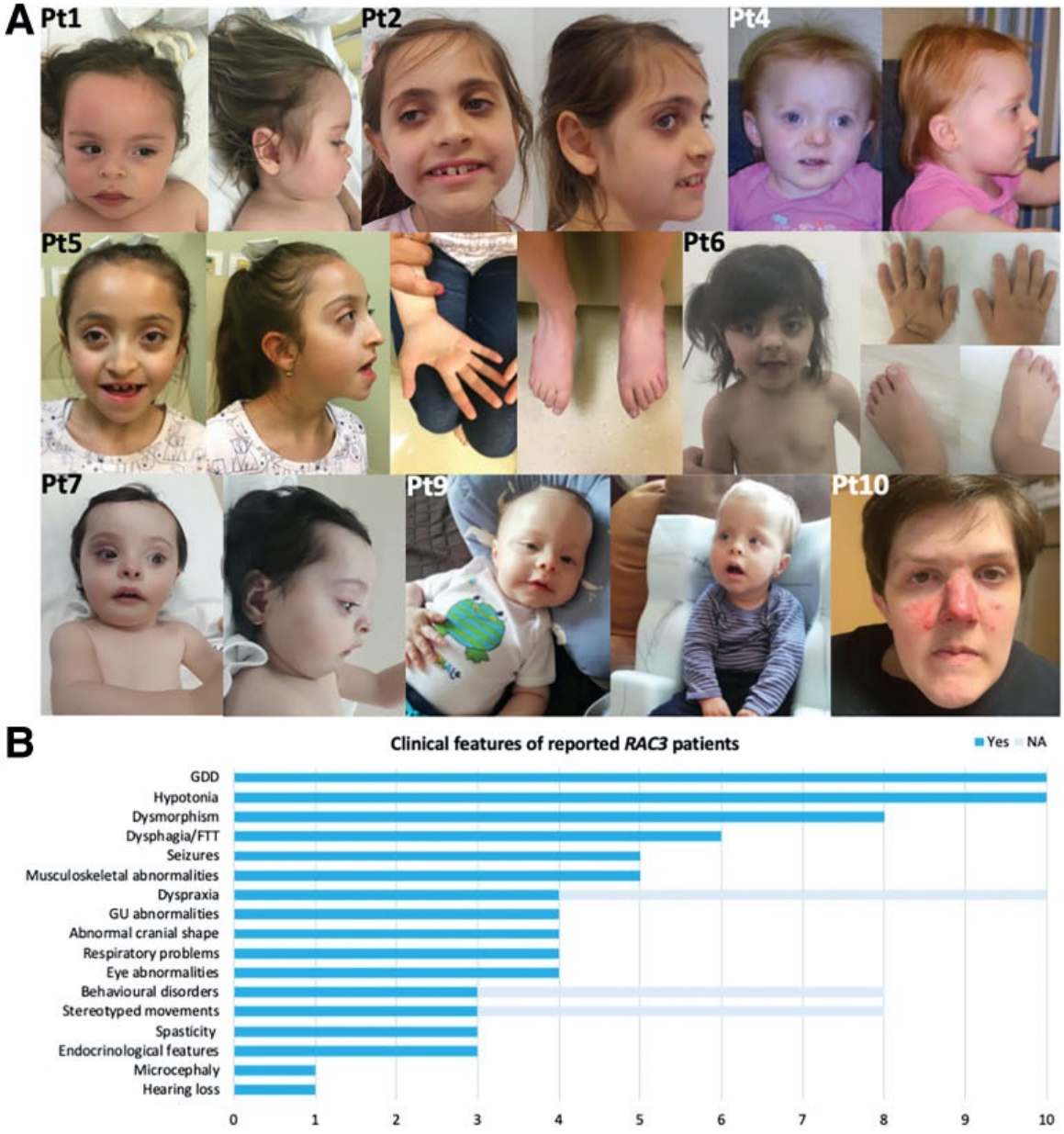


Figure 1. Clinical and neuroimaging features of individuals with *RAC3*-related disorder. (A) Dysmorphic features in subjects harboring pathogenic *RAC3* variants. Patient #1 shows wide forehead with frontal bossing, upslanted palpebral fissures, depressed nasal bridge, anteverted nares, prominent philtrum and upper lip, and thick earlobes. Patient #2 is microcephalic and presents with palpebral ptosis, synophrys, deep-set eyes, prominent nasal tip with hypoplastic nasal wings, large and prominent ears, and pointed chin. A round face with facial asymmetry, frontal bossing, high anterior hairline, upslanted palpebral fissures, depressed nasal bridge, micrognathia, and posteriorly rotated ears can be observed in Patient #4. Patient #5 has facial hypotonia, frontal bossing, bilateral ptosis, proptosis, hypertelorism, shallow orbits, midface hypoplasia, and pointed chin. Skeletal features consist of arachnodactyly and pes planus. Patient #6 has prominent eyes, long eyelashes, hypertelorism, midface hypoplasia, short nose, wide nasal bridge, anteverted nares, long philtrum, micrognathia, and dental anomalies. She also has clinodactyly and fetal finger and toe pads. Patient #7 displays scaphocephaly with a wide forehead and a high anterior hairline, broad nasal base, anteverted nares, retrognathia, and full lips. A prominent forehead with frontal bossing, epicanthal folds, deep philtrum, micrognathia with horizontal chin crease, and posteriorly rotated ears are

evident in Patient #9. Patient #10 shows down-slanted palpebral fissures, simplified ear helices, sharp tip of the nose, flat and long philtrum, and pointed chin. **(B)** Graphic illustration of the most common clinical findings in the reported study participants: the bar graph shows the distribution of the cardinal clinical features, with the grey lines representing not available (NA) data.

Progressive microcephaly was detected in one individual (#2), with others having abnormal cranial shape in the form of plagiocephaly (#1 and #4), brachycephaly (#5), and scaphocephaly (#7). Additional syndromic features were also evident in most individuals. Among the musculoskeletal abnormalities, scoliosis (#1, #3, and #10), vertebral defects (#1), pes planus (#3 and #5), and joint laxity (#3 and #4) were recurrent. Ocular involvement consisted of horizontal nystagmus (#1 and #5), strabismus (#1, #3, and #4), and astigmatism with hypermetropia (#1 and #3). Some individuals presented with genitourinary abnormalities, such as genital hypoplasia (#6), follicular cysts (#5), and enuresis (#3 and #4). Endocrinological features were less common and encompassed truncal obesity (#4), precocious puberty (#5), and hyperthyroidism with advanced bone age (#10). Respiratory involvement was observed in four patients, who experienced recurrent respiratory tract infections (#3 and #7) and apneic episodes (#9 and #10).

Although clinical features of *RAC3*-related disorder have been well documented in infancy, limited follow-up information is available and little is known about the natural history of the condition. Herein we report the oldest known affected individual (#10, aged 29 years), who seems to confirm three aspects of NEDBAF. First, the psychomotor delay may mildly improve over time, whereas regression is unusual (only observed in Patient #1 from our cohort, and limited to phonological skills). Second, seizures are responsive to antiepileptic treatment and have a generally benign course. Third, premature death does not appear to be part of the core clinical spectrum, although the apneic episodes reported in a minority of subjects are potentially life-threatening. Additional studies reporting large case series will determine if these observations are valid.

Affected individuals present with hypotonia, weak cry, or poor sucking in the neonatal period. Feeding difficulties are common and may fail to thrive (62.5%). During the first year of life, GDD with a severe involvement of language skills becomes apparent, but psychomotor arrest or regression are not typical. Neurological examination shows hypotonia (93.7%), dyspraxia (40%), or spasticity (25%) (Fig. S2).

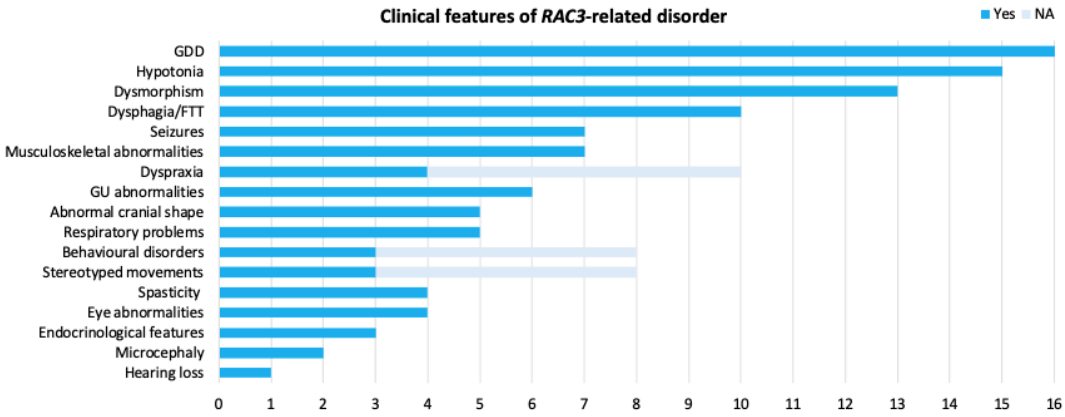


Figure S2. Phenotypic spectrum of RAC3-related disorder. Bar graph illustrating the distribution of clinical features in the whole population of *RAC3* patients.

Seizures occur in 43.7% of individuals, as early as the newborn period in two patients. Behavioral disorders and stereotyped movements are seen (both in 27.3% of subjects). Dysmorphic features are prominent (81.3%) and suggestive of a peculiar facial gestalt. Although microcephaly is rare (12.5%), an abnormal cranial shape is observed in 31.2% of individuals, but rarely reflects true craniosynostosis. Associated syndromic features are particularly relevant in some patients and include musculoskeletal (43.7%), genitourinary (37.5%), ocular (25%), and endocrinological (18.7%) abnormalities. Of note, respiratory problems are present in 31.2% of individuals.

3.1.b. Neuroimaging spectrum

Callosal anomalies were present in 10/10 participants, including hypoplasia with prevalent posterior involvement (eight cases), thick and short morphology (one case, #3), and partial agenesis (one case, #10) (Fig. 1C, extract).

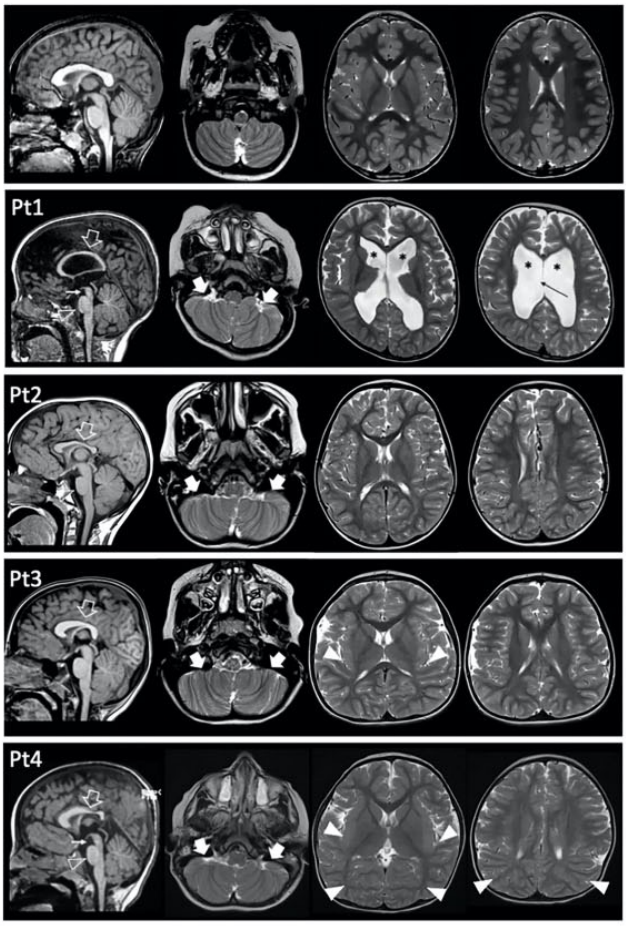


Figure 1C. Clinical and neuroimaging features of individuals with *RAC3*-related disorder. (C) Brain MRI, sagittal T1-weighted and axial T2-weighted images, of Patients #1-4 and of a normal individual (first row) for comparison. The corpus callosum is shorter and thinner in all patients, especially at the level of the splenium (empty arrows), with thick genu in Patients #3 and #4. Note the absence of the anterior commissure in all patients. The volume of the cerebral white matter is reduced in all patients, with marked ventricular enlargement (asterisks) and septal defect (black thin arrow) in Patient #1. Diffuse dysgyria is present in all subjects, associated with polymicrogyria in the posterior regions in Patients #3 and #4 (arrowheads). Bilateral cerebellar folia abnormalities are noted in the inferior portions of the cerebellum in all patients, especially on the left side (thick arrows). There is hypoplasia of the midbrain in Patients #1 and #4 (white thin arrows) and a small pons in Patients #1, #2, and #4 (empty arrowheads). Mild enlargement of the subarachnoid spaces is noted in all patients, while brachycephaly is present only in Patients #3 and #4.

White matter thinning and enlarged subarachnoid spaces with global reduction of the brain volume were noted in 8/10 and 9/10 cases, respectively. Malformations of cortical development were noted in 7/10 participants, including polymicrogyria involving the posterior regions (five cases), diffuse dysgyria (four cases), and small gray matter heterotopias dispersed in the cerebral white matter (one case, #1). Cerebellar dysplasia with foliation abnormalities involving the inferior cerebellar hemispheres was detected in 6/10 participants, while brainstem anomalies characterized

by a small midbrain and/or pons were seen in 5/10 cases. Chiari I anomaly was not detected in this series, but mild (<5mm) caudal descent of the cerebellar tonsils was noted in three participants.

3.2. Genotype dissection

3.2.a. Identification and analyses of *RAC3* variants

Based on allele frequency (<0.001 in gnomAD), involvement of conserved amino acid residues, and predicted impact on protein function, variants in *RAC3* were the most plausible candidate variants in the study cohort. NGS led to the identification of seven distinct variants, including the previously reported c.184G>A p.E62K in Patient #7 (ClinVar ID: 425149)[18]. Most of the novel variants were missense: c.187G>A p.D63N (#1 and #9); c.191A>G p.Y64C (#3); c.179G>A p.G60D (#4); c.34G>C p.G12R (#5); c.348G>C p.K116N (#8). All these variants are rare, affect conserved residues (GERP scores 2.5-3.91), and are predicted to have a high pathogenic impact (CADD score 24.3-27.6) (Table 2).

The missense variant c.176C>G p.A59G previously detected in Patient #6 [20] is also predicted to be pathogenic by several *in silico* tools. In Patient #2, the novel in-frame deletion c.186_188delGGA p.E62del was detected. This variant results in a protein coding length change through the deletion of the conserved Glu62 (GERP score 3.79). No CNVs with a pathogenic potential could be detected in the reported subjects.

Table 2. *In silico* analysis of *RAC3* variants.

RAC3 variant (NM_005052.3)	ID	Consequence	gnomAD_AF	GERP	CADD_PHRED	CADD_RAW	SIFT	PolyPhen
c.34G>C; p.(Gly12Arg)	17_79989671_G/C	missense&splice_region_variant	0	2.5	24.3	3.293	deleterious_low_confidence(0.02)	benign(0.434)
c.176C>G; p.(Ala59Gly)	17_79990655_C/G	missense_variant	0	3.79	25.8	3.700	deleterious_low_confidence(0.01)	possibly_damaging(0.805)
c.179G>A; p.(Gly60Asp)	17_79990658_G/A	missense_variant	0	3.79	25.8	3.703	deleterious_low_confidence(0)	probably_damaging(0.995)
c.186_188delGGA; p.(Glu62del)	17_79990661_AGG/-	inframe_deletion	0	3.79	NA	NA	NA	NA
c.184G>A; p.(Glu62Lys)	17_79990663_G/A	missense_variant	0	3.79	27.4	3.960	deleterious_low_confidence(0)	probably_damaging(0.979)
c.187G>A; p.(Asp63Asn)	17_79990666_G/A	missense_variant	0	3.79	27.6	3.988	deleterious_low_confidence(0)	probably_damaging(0.977)
c.191A>G; p.(Tyr64Cys)	17_79990670_A/G	missense_variant	0	3.79	25.5	3.632	deleterious_low_confidence(0.01)	probably_damaging(1)
c.348G>C; p.(Lys116Asn)	17_79991375_G/T	missense_variant	0	3.91	25.2	3.558	deleterious_low_confidence(0)	probably_damaging(0.998)

3.2.b. Structural and biochemical features of *RAC3* variants

A common structural and biochemical feature across the small GTPase superfamily is the G domain, which is defined by five G boxes with certain structurally conserved amino acid residues (Fig. S3). The G1 box/P-loop (GxxxxGKS/T, where x is any amino acid) recognizes the β-phosphate and Mg^{2+} ion of guanine nucleotides. The G2 box (T) positioned in the Switch I region interacts with the γ-phosphate and Mg^{2+}, while the G3 box (DxxGQ/H/T) localizing in the Switch II region is responsible for GTP-hydrolysis. Eventually, the G4 (T/NKxD) and G5 (C/SAK/L/T) boxes make specific contacts with the guanine base.

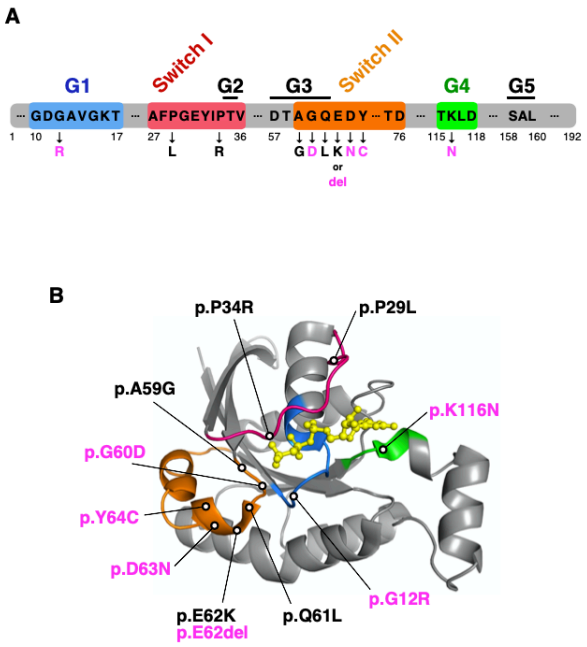


Figure S3. Schematic representation of RAC3 variants. (A) Schematic representation of RAC3 structure was shown with the positions of the variations identified in this study (p.G12R, p.G60D, p.E62del, p.D63N, p.Y64C, and p.K116N) in magenta, and previously (p.P29L, p.P34R, p.A59G, p.Q61L, and p.E62K) in black. **(B)** The 3D structure of GTP-bound RAC3 was predicted with the SWISS-MODEL program, based on the structure of RAC1 (PDB ID: 1E96). Structural domains are colored as follows: G1, blue; Switch I, red; Switch II, orange; G4, green. G2, G3 and G5 boxes are shown with black bars in (A). GTP is indicated in yellow in (B).

The previously reported p.P34R and p.P29L variants (*Costain et al., 2019; Hiraide et al., 2019*) lead to substitutions of evolutionarily conserved amino acids in the Switch I region, with p.P34R being also included in the effector binding region. It is of note that all variations identified in this study occurred at residues highly conserved across species (Fig. S3 and S4).

Among the *RAC3* variants observed in our cohort, seven variants (p.A59G, p.G60D, p.Q61L, p.E62del, p.E62K, p.D63N, and p.Y64C) are mapped to the Switch II region, which is recognized as a variation hot spot among *RAC1*, *RAC3*, and *CDC42* (*Reijnders et al., 2017; Martinelli et al., 2018; Hamada et al., 2020*). In addition, the p.G12R and p.K116N are the first variants affecting the conserved residues in the G1 and G4 boxes, respectively (Fig. S3 and S4).

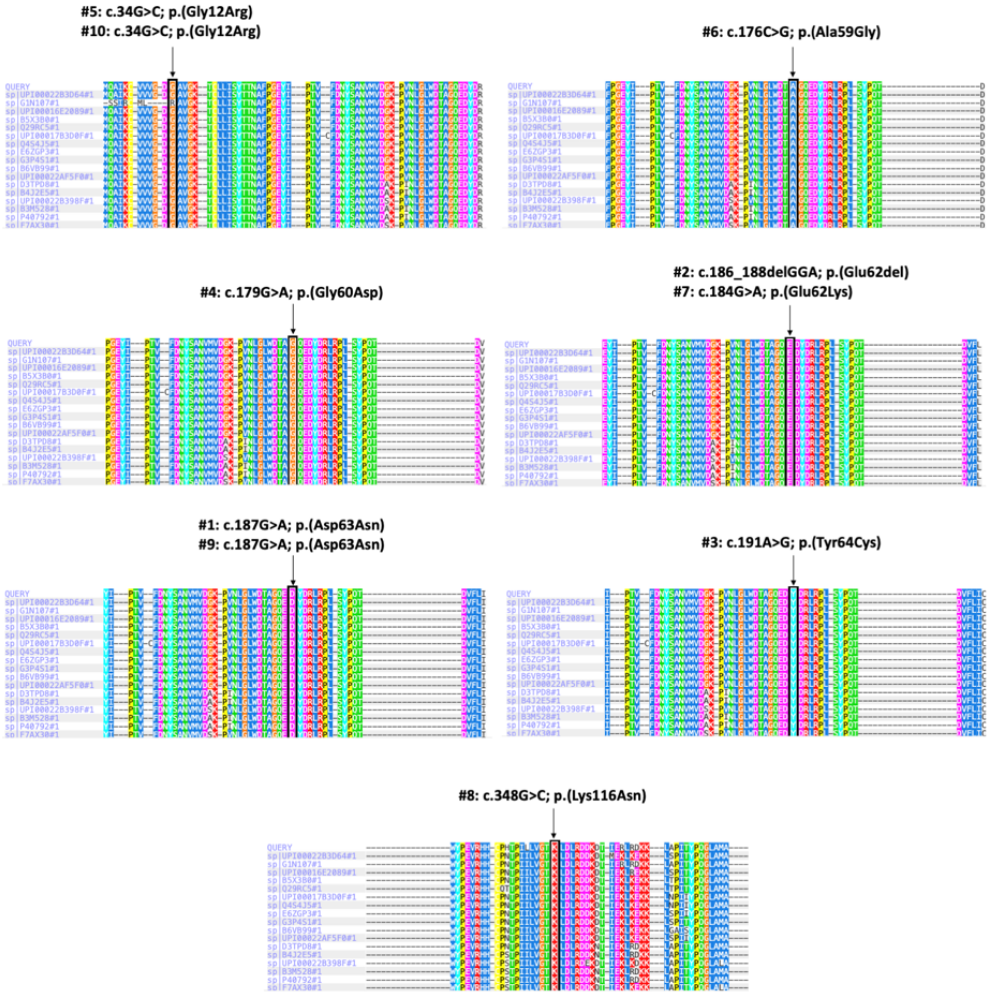


Figure S4. Conservation among different species of the RAC3 residues affected in the reported patients. *RAC3* variants are reported according to the NM_005052.3 transcript. UPI00022B3D64 (A0A087XGA4): Poecilia formosa (Amazon molly) (Limia formosa); G1N107_MELGA: Meleagris gallopavo (Wild turkey); UPI00016E2089 (A0A3B5K3K0): Takifugu rubripes (Japanese pufferfish) (Fugu rubripes); B5X3B0_SALSA: Salmo salar (Atlantic salmon); Q29RC5_DANRE: Danio rerio (Zebrafish) (Brachydanio rerio); UPI00017B3D0F: Tetraodon nigroviridis (Spotted green pufferfish) (Chelonodon nigroviridis); Q4S4J5_TETNG: Tetraodon nigroviridis (Spotted green pufferfish) (Chelonodon nigroviridis); E6ZGP3_DICLA: Dicentrarchus labrax (European seabass) (Morone labrax); G3P4S1_GASAC: Gasterosteus aculeatus (Three-spined stickleback); B6VB99_SCOMX: Scophthalmus maximus (Turbot) (Psetta maxima); UPI00022AF5F0: Oreochromis niloticus (Nile tilapia) (Tilapia nilotica); D3TPD8_GLOMM: Glossina morsitans morsitans (Savannah tsetse fly); B4J2E5_DROGR: Drosophila grimshawi (Hawaiian fruit fly) (Idiomyia grimshawi); UPI00022B398F (A0A3P8P4B1): Astatotilapia calliptera (Eastern happy) (Chromis callipterus); B3M528_DROAN: Drosophila ananassae (Fruit fly); RAC1_DROME: Drosophila melanogaster (Fruit fly); F7AX30_HORSE: horse. Source: http://genetics.bwh.harvard.edu/pph2/index.shtml.

3.2.c. Array-CGH

Patient #6 was found to harbor three maternally inherited duplication: arr[hg19] 3q12.2 (100348441_100438903) x3, 8p12 (29923503_29940466) x3, and 12p13.31 (7945545_7972180) x3. The first rearrangement (arr[hg19] 3q12.2 (100348441_100438903) x3) encompasses *GPR128* (OMIM * 612307) and *TFG* (* 602498). While *GPR128* is not a disease-causing gene, variants in *TFG* are associated with autosomal recessive spastic paraplegia 57 (OMIM # 615658) and Okinawa type hereditary motor and sensory neuropathy (OMIM # 604484). However, duplications involving *TFG* have not been linked to a clinical phenotype so far. The second duplication (arr[hg19] 8p12 (29923503_29940466) x3) includes *MIR548O2* (not OMIM) and *TMEM66* (OMIM * 614768), both not associated with a human disorder. The third rearrangement (arr[hg19] 12p13.31 (7945545_7972180) x3) encompasses *NANOG* (OMIM * 607937) and *SLC2A14* (OMIM * 611039), not associated with a human phenotype. According to the inheritance pattern and the involved genes, all these duplications are not predicted to have a pathogenic potential. Eventually, the array-CGH (patients #1, #2, #4, #5, and #10) and exome sequencing copy number variants (CNVs) analysis (patient #3) yielded negative results in the remaining cases.

Patient #7 was found to harbor two rearrangements: arr[hg19] 17p13.3 (799610_912880) x 1 and Xp22.33 (1378391_1431726) x 3. The first deletion encompasses 4 OMIM genes (*ADAP1*, *608114; *COX19*, * 610429; *GET4*, *612056; *SUN1*, *607723), none of which is currently associated with a human phenotype. Furthermore, no patient with an overlapping rearrangement could be identified in the Decipher database (https://www.deciphergenomics.org/search/patients/results?q=7%3A799610-912880). The Xp22.33 duplication involves five genes, of which three OMIM (*ASMTL*, *300162; *IL3RA*, *430000; *SLC25A6*, *403000), but none associated with human disease so far. In the Decipher

database, a single patient carried a partially overlapping rearrangement (ID 290028). Although reported to have seizures, this individual also carried two additional rearrangement and the Xp22.33 duplication (1378227_1387657) was interpreted as benign.

Patient #8 harbored the maternally-inherited arr[hg19] 15q11.2 (22750082_23672560) x 1. This deletion encompasses 17 genes, including six OMIM genes (*CYFIP1*, *606322; *MAGEL2*, *605283; *MKRN3*, *603856; *NIPA1*, *608145; *NIPA2*, *608146; *TUBGCP5*, *608147). Three of these genes are morbid genes (*MAGEL2* - Schaaf-Yang syndrome, #615547; *MKRN3* - Precocious puberty, central, 2, # 615346; *NIPA1* - Spastic paraplegia 6, autosomal dominant, #600363), but no clinical feature consistent with the conditions associated with their haploinsufficiency could be identified either in the proband or in his healthy mother. The deleted interval is part of the 15q11.2 BP1–BP2 microdeletion syndrome, a rare condition characterized by developmental delay and neuropsychiatric features (*Cox et al., 2015; Rafi et al., 2020*). Since the deletion was inherited from the healthy mother, this rearrangement was not considered pathogenic in this individual, although a partial contribution to the neurobehavioral phenotype cannot be completely excluded due to the characteristic incomplete penetrance reported in patients with the 15q11.2 BP1–BP2 microdeletion syndrome.

3.3. Molecular studies

3.3.A. *In vitro*

3.3.A.a. Biological and biochemical activities of the 11 disorder-causative *RAC3* variants

To get further insights into the pathophysiological significance of *RAC3* variants associated with NDDs, activation states were first examined *in vitro* for the eight variants identified in the enrolled patients, together with the recently reported p.P34R, p.P29L, and p.Q61L variants (*Costain et al.,*

2019; Hiraide et al., 2019). Of note, the p.Q61L variant is best characterized as a GTPase-defective constitutively active version of RAC1 involved in tumor cell metastasis and invasion (*Krengel et al., 1990; Alan et al., 2013; Kazanietz et al., 2017*).

When control pCAG-Myc vector was expressed in primary cultured hippocampal neurons, cells were normally differentiated in a time-dependent manner (Fig. 2A and N). Likewise, neurons expressing wild type RAC3 were normally differentiated with single axon and numerous dendrites, although its expression induced lamellipodia at the growth cone of axon and cell body (Fig. 2B).

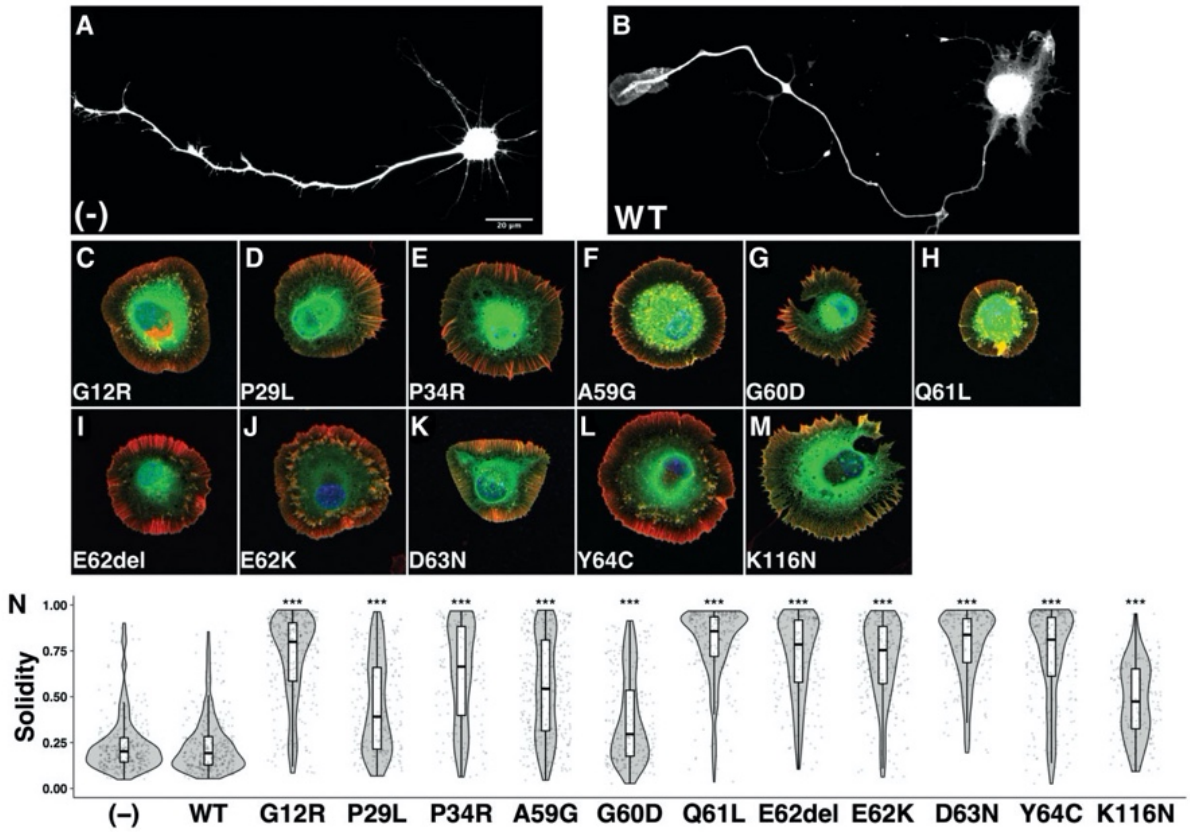


Figure 2. Effects of the disease-causative 11 RAC3 variants on neuronal morphology *in vitro*. (A - M) Dissociated hippocampal neurons from E16 mice were co-electroporated with pCAG-EGFP (0.1 μg) together with pCAG-Myc (-), pCAG-Myc-RAC3 (WT), -G12R, -P29L, -P34R, -A59G, -G60D, -Q61L, -E62del, -E62K, -D63N, -Y64C and -K116N (0.3 μg each). Cells were fixed at 3 div, and co-stained with anti-GFP (green), rhodamine-phalloidin (red) and DAPI (blue). Scale bar, 20 μm. (N) Quantification of (A-M). The morphological descriptor (solidity) of GFP-positive neurons (≧ 200 cells each) was shown in violin plot with boxplot. "Solidity" is the ratio of the area of a cell to the area of a convex hull of the cell. Values of the ratio were calculated by "Analyze Particles..." method in imageJ (https://imagej.nih.gov/ij/docs/menus/

analyze.html). The significance of difference between (-) and each variant was determined using Dunnett's test. WT vs. (-), p = 1; G12R vs. (-), p < 1e-10; P29L vs. (-), p < 1e-10; P34R vs. (-), p < 1e-10; A59G vs. (-), p < 1e-10; G60D vs. (-), p < 1e-10; Q61L vs. (-), p < 1e-10; E62del vs. (-), p < 1e-10; E62K vs. (-), p < 1e-10; D63N vs. (-), p < 1e-10; Y64C vs. (-), p < 1e-10; K116N vs. (-), p < 1e-10. ***p < 0.001.

These results indicate that the basal activity of exogeneous wild type RAC3 had little effect on cell differentiation *in vitro*. In contrast, neurons transfected with the 11 studied variants displayed cell rounding with typical lamellipodia formation and less neurite extension, resulting in a prominent increase in the solidity (Fig. 2C-N). While the median of solidity was shifted upward when each variant was expressed, the degree was variant-dependent because cells expressing some variants such as RAC3-P29L, -A59G, -G60D, and -K116N occasionally showed neurite extension (Fig. S5).

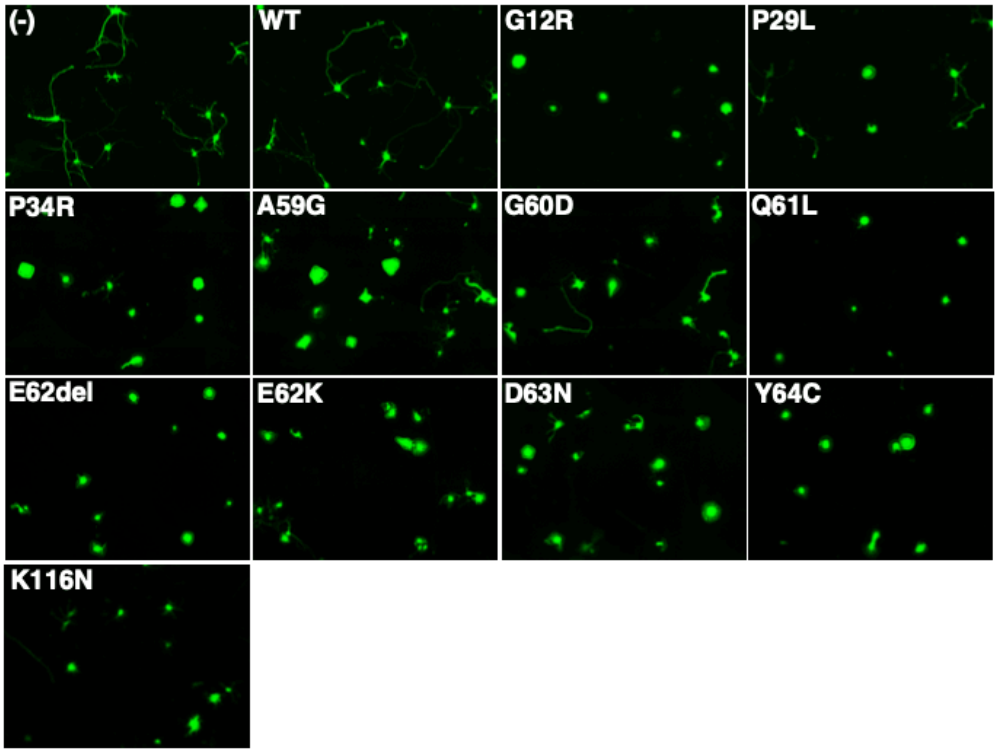


Figure S5. Effects of the disease-causative 11 *RAC3* variants on neuronal morphology *in vitro*. Hippocampal neurons dissociated at E16 were co-electroporated with pCAG-EGFP (0.1 μg) together with pCAG-Myc (-), pCAG-Myc-RAC3 (WT), -G12R, -P29L, -P34R, -A59G, -G60D, -Q61L, -E62del, -E62K, -D63N, -Y64C, and -K116N (0.3 μg each) and were cultured *in vitro* for three days. Control (-) and WT- expressing neurons began to extend an axon (the longest neurite) and dendrites. Although growth of axon and dendrites was severely suppressed in cells

expressing the 11 variants, the ratio of round cells was dependent on the variation types, which affects the quantification analyses in Fig. 2.

Taken together, all the 11 variants appeared to facilitate cytoskeletal reorganization to form lamellipodia to greater or lesser degrees. The respective variants should dysregulate intracellular signaling in variant-dependent manners and adversely affect neuronal morphology and function.

We then performed biochemical analyses and measured GTP/GDP-exchange and GTP-hydrolysis activity. When the effects of these 11 variations on GTP/GDP-exchange activity were examined with recombinant RAC3 proteins, the assay documented a variably increased or decreased activity. When compared with wild type RAC3, $\log_{10}(k_{obs}$: observed rate constant) was especially high for RAC3-K116N and moderately high for RAC3-Q61L, -P29L, -Y64C, -E62del, -P34R, and -D63N in this order, indicating an accelerated exchange reaction (Fig. 3A and S6A).

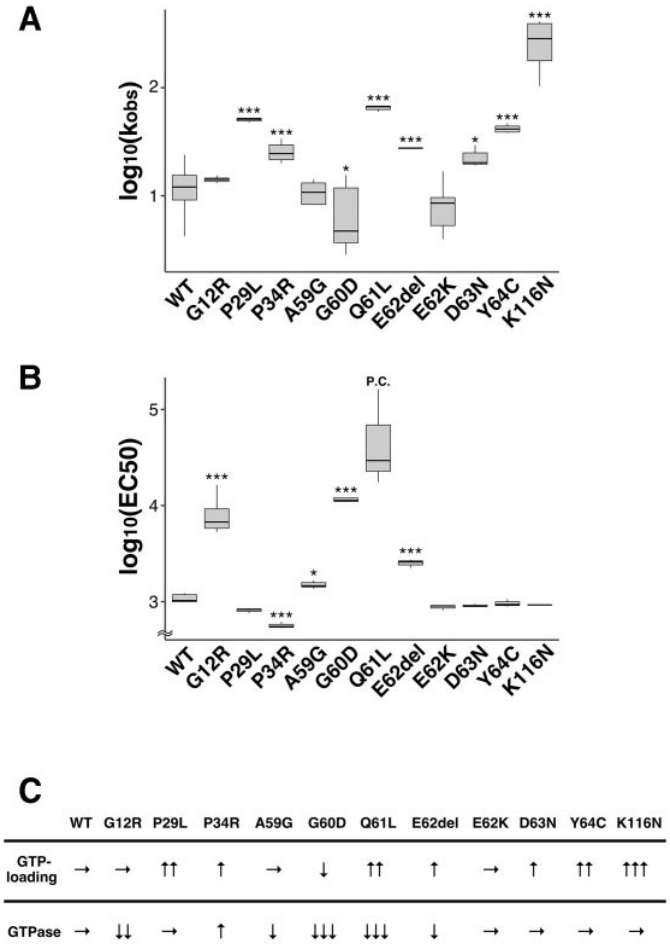


Figure 3. Characterization of activation states of the disease-causative 11 *RAC3* variants *in vitro*. **(A)** Measurement of GDP/GDP-exchange activity. Recombinant His-tag-fused RAC3 (WT) and the 11 variants (RAC3-G12R, -P29L, -P34R, -A59G, -G60D, -Q61L, -E62del, -E62K, -D63N, -Y64C, and -K116N) were preloaded with fluorescent mantGDP and incubated with unlabeled GTP. Intrinsic nucleotide exchange was measured and mantGDP-dissociation rates of WT and respective variants were calculated as observed rate constants (kobs [×10-5 s-1]) from the results in Supplementary Fig. 5A. Number of replicates, N ≧ 4. The significance of difference between WT and each variant was determined using Dunnett's test. G12R vs. WT, p = 0.8305; P29L vs. WT, p < 0.001; P34R vs. WT, p < 0.001; A59G vs. WT, p = 0.7655; G60D vs. WT, p = 0.0135; Q61L vs. WT, p < 0.001; E62del vs. WT, p < 0.001; E62K vs. WT, p = 0.1018; D63N vs. WT, p = 0.0132; Y64C vs. WT, p < 0.001; K116N vs. WT, p < 0.001. *p < 0.05, ***p < 0.001. **(B)** Measurement of GTP-hydrolysis activity. Intrinsic GTP-hydrolysis activity was analyzed by direct measuring of changes in the GTP concentration with GTPase-Glo assay kit. EC50 (half maximal effective concentration) was estimated from the sigmoidal fitting curve in Fig. S5B. Number of replicates, N ≧ 3. P value was calculated as in (A). G12R vs. WT, p < 1e-04; P29L vs. WT, p = 0.174; P34R vs. WT, p < 1e-04; A59G vs. WT, p = 0.012; G60D vs. WT, p < 1e-04; E62del vs. WT, p < 1e-04; E62K vs. WT, p = 0.514; D63N vs. WT, p = 0.743; Y64C vs. WT, p = 0.952; K116N vs. WT, p = 0.833. *p < 0.05, ***p < 0.001. P.C., positive control. **(C)** Summary of the results in (A) and (B). →, no change; ↑, activity increased; ↓, activity decreased.

In contrast, a statistically significant decrease of $\log_{10}(k_{obs})$ was observed for RAC3-G60D, indicating suppression of the exchange reaction (Fig. 3A and S6A). The exchange activity of RAC3-G12R, -A59G, and -E62K did not show statistical difference in comparison to the wild type (Fig. 3A and S6A). Then, GTP-hydrolysis activity was assayed for each variant. This was highly suppressed by p.Q61L, p.G60D, and p.G12R, while moderate to weak suppression was observed for p.E62del and p.A59G (Fig. 3B and S6B). We consider that p.G12R, p.G60D, and p.Q61L primarily activated the protein by eliminating the GTP-hydrolysis activity like oncogenic RAS proteins. In contrast, the activity of RAC3-P34R was higher than the wild type (Fig. 3B and S6B). Of note, RAC3-P29L, E62K, -D63N, -Y64C and -K116N did not affect GTP-hydrolysis activity when compared to the wild type (Fig. 3B and S6B).

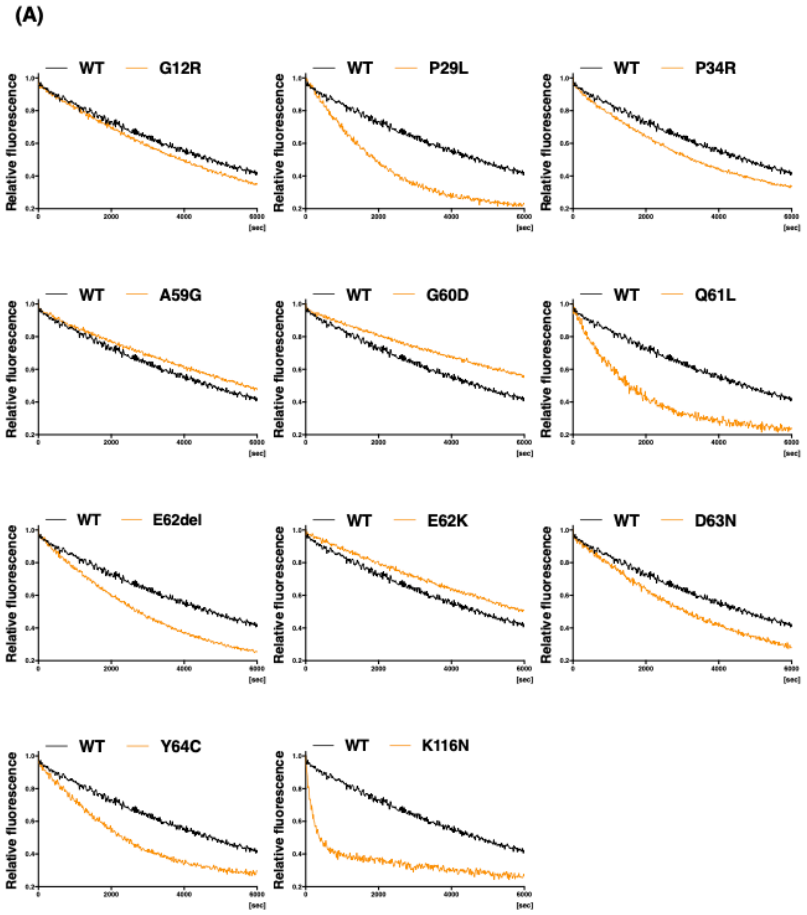


Figure S6. Characterization of activation state of the disease-causative 11 RAC3 variants *in vitro*. **(A)** Raw data of GDP/GDP-exchange activity. The release of a fluorescently labelled GDP (^{mant}GDP) from each variant was measured as described in the "Materials and methods" section and compared to that of WT. The data were quantified and shown in Fig. 3A.

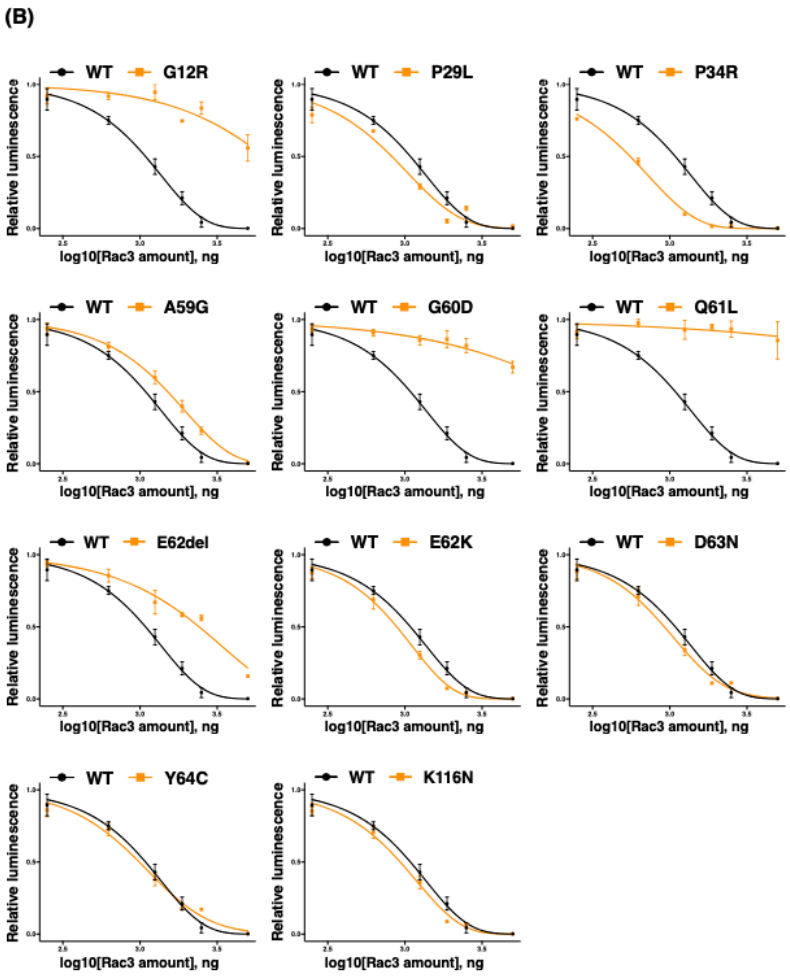


Figure S6. Characterization of activation state of the disease-causative 11 RAC3 variants *in vitro*. (B) Raw data of intrinsic GTP-hydrolysis activity. The activity was measured as described in the "Materials and Methods" section. The data were quantified and shown in Fig. 3B.

Based on the obtained results summarized in Fig. 3C, it is possible to categorize the biochemical properties of the 11 variants into three groups based on GTP/GDP-exchange activity. Group I encompasses RAC3-P29L, -P34R, -Q61L, -E62del, -D63N, -Y64C, and -K116N. These variants demonstrated a higher GTP/GDP-exchange activity. While GTP-hydrolysis activity of RAC3-P29L, -D63N, -Y64C, and -K116N was comparable to that of wild type, RAC3-Q61L and -E62del showed lower hydrolysis activity. We assume that these seven variants are prone to be in GTP-bound active conformation, since they exhibited high GTP/GDP-exchange activity with GTP-hydrolysis activity equivalent to or lower than that of wild type RAC3. On the other hand, although RAC3-P34R has a statistically high GTP-hydrolysis activity, we assume that the variant

is activated due to the accelerated GTP/GDP exchange activity in the context of an intracellular environment in which the GTP concentration is substantially higher than that of GDP. It is notable that the substitution of the corresponding Pro34 in RAC2 with His was suggested to activate the protein based on crystal structure modeling (*Lougaris et al., 2019*). Group II includes RAC3-G12R, -A59G, and -E62K, all showing a GTP/GDP-exchange activity similar to the wild type. Since RAC3-G12R and -A59G displayed suppressed GTP-hydrolysis activity, they were likely to prefer GTP-binding. As for RAC3-E62K with normal GTP-hydrolysis and GTP/GDP-exchange activities *in vitro*, it appears that abnormal interaction with unidentified GEFs and/or GAPs holds RAC3-E62K in an active state in primary neurons. In this context, RAC2-E62K has been reported to be hyper-activated through altered GEF specificity and impaired GAP function (*Arrington et al., 2020*). Eventually, Group III includes RAC3-G60D, which may prefer GTP-binding because of its very low GTP-hydrolysis activity even if the GTP/GDP-exchange rate is also low. Based on the results in Fig. 2 and Fig. 3, we concluded that all the 11 variations confer a gain-of-function phenotype with respect to the RAC3 function. Biological and biochemical diversity among variations may variously dysregulate RAC3 function *in vivo*, leading to variation-specific clinical phenotypes.

3.3.A.b. Possible interaction of the *RAC3* variants with various downstream effectors

While the patients described in this study share core NDD phenotypes associated with callosal and cortical malformations, clinical features peculiar to each individual may be attributable to the position and type of the involved amino acids changes. Since the 11 *RAC3* variants analyzed in this study are considered to be differently activated based on cell biological and biochemical analyses (Fig. 2 and 3), we looked into the downstream signaling pathways interacting with these

variants. To this end, a pull-down assay was performed to examine the interaction of respective RAC3 variants with recombinant RBRs of various effectors including PAK1, MLK2, IRSp53, N-WASP, ROCK, and RTKN. Consequently, the *RAC3* variants were observed to interact with these RBRs in variant-dependent manners (Fig. 4A).

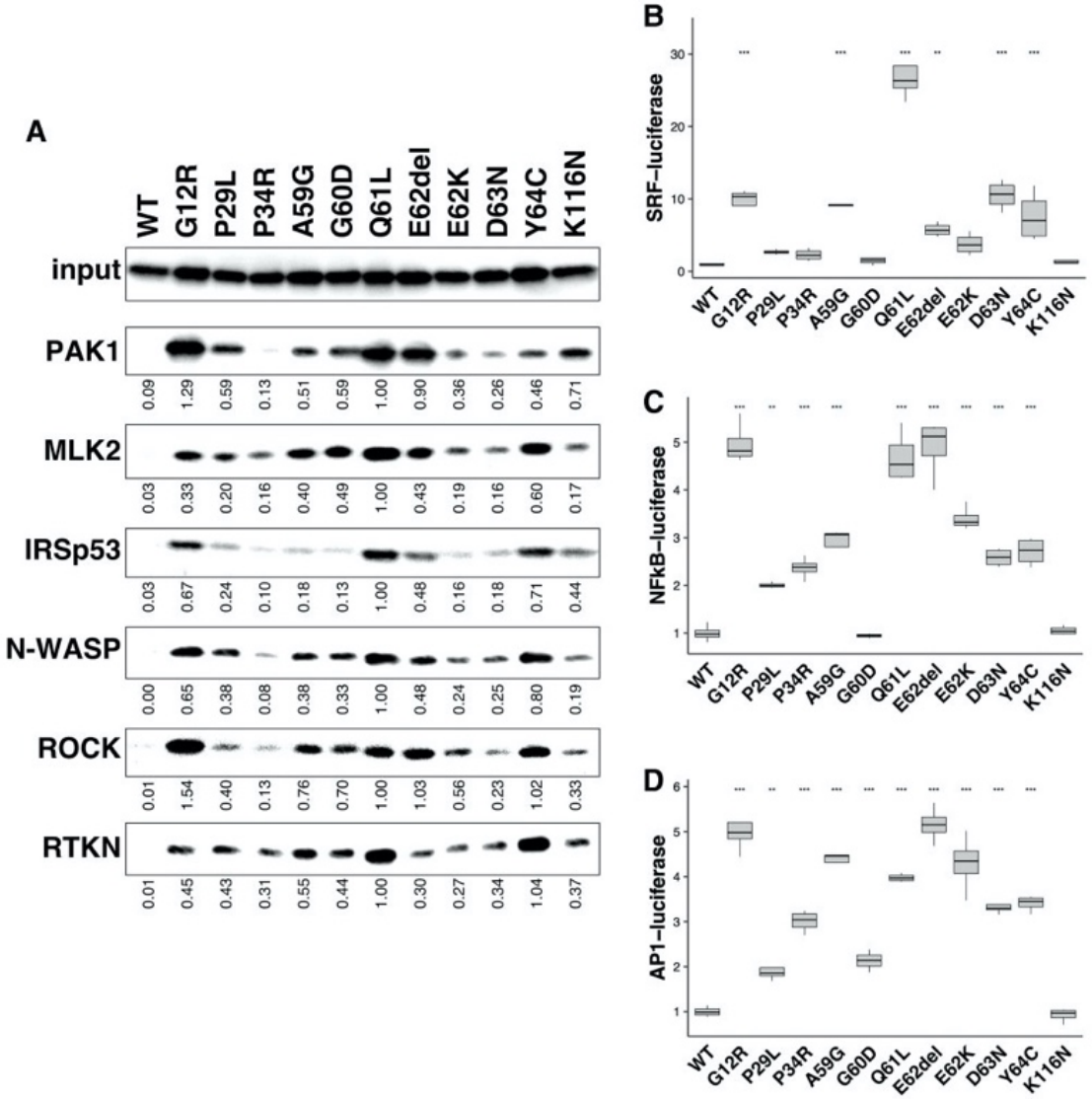


Figure 4. Possible interaction of the disease-causative 11 *RAC3* variants with various downstream effectors. (A) Interaction with the RBRs of PAK1, MLK2, IRSp53, N-WASP, ROCK and RTKN. COS7 cells were transfected with pCAG-Myc-RAC3 (WT), -G12R, -P29L, -P34R, -A59G, -G60D, -Q61L, -E62del, -E62K, -D63N, -Y64C, and -K116N (0.3 μg each). Cell lysates were prepared, and the pull-down assay was conducted as described in the "Materials and Methods" section with respective GST-fused RBRs (5 μg each). Bound RAC3 proteins were detected by western blotting with anti-Myc. RAC3 activity was indicated by the amounts of RAC3 bound to respective GST-RBRs, and total cell lysates were also immunoblotted with anti-Myc for normalization (*input*). Relative band intensity was shown when the value of RAC3-Q61L was taken as 1.0. **(B - D)** Involvement of the *RAC3* variations in SRF- (B),

NFkB- (C) or AP1-dependent (D) gene transcription. COS7 cells were co-transfected with each luciferase expression vector together with pCAG-Myc-RAC3 (WT) or the expression vector for the 11 variants. Luciferase activity obtained with WT was taken as 1.0, and relative activities are shown as box plot. Number of replicates, N = 4. The significance of difference between WT and each variant was determined using Dunnett's test.
(B) G12R vs. WT, $p < 0.001$; P29L vs. WT, $p = 0.81326$; P34R vs. WT, $p = 0.94861$; A59G vs. WT, $p < 0.001$; G60D vs. WT, $p = 0.99999$; Q61L vs. WT, $p < 0.001$; E62del vs. WT, $p = 0.00997$; E62K vs. WT, $p = 0.27663$; D63N vs. WT, $p < 0.001$; Y64C vs. WT, $p < 0.001$; K116N vs. WT, $p = 0.99999$. (C) G12R vs. WT, $p <$ 1e-04; P29L vs. WT, $p = 0.00159$; P34R vs. WT, $p <$ 1e-04; A59G vs. WT, $p <$ 1e-04; G60D vs. WT, $p =$ 1.00000; Q61L vs. WT, $p <$ 1e-04; E62del vs. WT, $p <$ 1e-04; E62K vs. WT, $p <$ 1e-04; D63N vs. WT, $p <$ 1e-04; Y64C vs. WT, $p <$ 1e-04; K116N vs. WT, $p =$ 1.00000. (D) G12R vs. WT, $p <$ 1e-04; P29L vs. WT, $p = 0.00275$; P34R vs. WT, $p <$ 1e-04; A59G vs. WT, $p <$ 1e-04; G60D vs. WT, $p = 0.00017$; Q61L vs. WT, $p <$ 1e-04; E62del vs. WT, $p <$ 1e-04; E62K vs. WT, $p <$ 1e-04; D63N vs. WT, $p <$ 1e-04; Y64C vs. WT, $p <$ 1e-04; K116N vs. WT, $p = 0.99999$. $^{**}p < 0.01$, $^{***}p < 0.001$.

The RBR of PAK1, which is crucial for neuronal migration and variated in patients with NDDs (*Causeret et al., 2009; Ohori et al., 2020; Harms et al., 2018; Zhang et al., 2020; Horn et al., 2019*), was associated with all the variants except RAC3-P34R. High affinity was displayed for RAC3-G12R, - Q61L, -E62del, and -K116N in this order, and low affinity was observed for RAC3-D63N (Fig. 4A). The RBR of MLK2, which activates the JNK-MAP kinase pathway downstream of RAC (*Nagata et al., 1998*), significantly interacted with RAC3-Q61L and then -Y64C, and moderate to weak affinity was observed for all other variants (Fig. 4A). The RBR of IRSp53, an adaptor protein acting at the membrane-actin interface and related to the formation of filopodia and lamellipodia (*Miki et al., 2000; Nakagawa et al., 2003*), showed relatively high affinity towards RAC3-Q61L, -Y64C, and -G12R (Fig. 4A). Although RAC3-E62del and -K116N had moderate affinity, a very weak interaction was observed with other variants (Fig. 4A). The RBR of N-WASP, a regulator of actin polymer reorganization in the cytoplasm and nucleus (*Burbelo et al., 1995; Tomasevic et al., 2007*), displayed relatively strong affinity to RAC3-Q61L, -Y64C, and -G12R in this order, while weak to moderate affinity was detected for all other variants but -P34R (Fig. 4A). The RBR of ROCK, a key regulator of actin cytoskeleton and cell polarity (*Soriano-Castell et al., 2017; Amano et al., 2010*), exerted strong binding to RAC3-G12R, -Q61L, -E62del, and -Y64C (Fig. 4A). While moderate interaction was shown for RAC3-A59G, -G60D, and -E62K, other variants except RAC3-P34R showed a weak affinity (Fig. 4A). RTKN has been

shown to regulate the septin cytoskeleton. (*Ito et al., 2005*). Although RTKN was first reported as a Rho-specific effector (*Reid et al., 1996*), we have found that RTKN interacts with activated RAC3 but not RAC1 through the Rho-binding domain. We thus asked whether this domain, which is referred to as RBR here, interacts with the 11 variants. We determined that RTKN-RBR interacted with the 11 variants, with particularly strong affinity to RAC3-Q61L and -Y64C (Fig. 4A). Taken together, given that the respective *RAC3* variants are activated versions, they are strongly suggested to hyper-activate different sets of effectors in variant-dependent manners and, thereby, dysregulate specific downstream signaling pathways in context-dependent manners. Full cropped data of western blotting were depicted in Fig. S7.

We next examined signal transduction pathways underlying the aberrant activation states caused by these variations. We focused on SRF-, NFkB- and AP1-mediated gene expression, since their related signaling pathways include Rho-family proteins and MAP kinases (*Hill et al., 1995; Perona et al., 1997; Coso et al., 1995*). When effects of respective variants on SRF-dependent gene transcription were assessed, prominent transcriptional activation was observed in cells expressing RAC3-Q61L, while moderate activation was induced by RAC3-G12R, -A59G, -E62del, -D63N, and -Y64C (Fig. 4B). In contrast, RAC3-P29L, -P34L, -G60D, and -K116N had marginal effects. Meanwhile, NFkB-dependent gene transcription was highly increased by RAC3-E62del, -G12R, and -Q61L (Fig. 4C). Noteworthy, these three variants were observed to show high affinity to PAK1-RBR (Fig. 4A). Other variants except for RAC3-G60D and -K116N, which showed no effects, demonstrated moderate NFkB-activation (Fig. 4C). Then, when we looked into the effects on AP1-mediated gene expression, each *RAC3* variant induced the expression similar to that of NFkB; the *RAC3* variants other than RAC3-A59G, -G60D, -Q61L and -E62K might share common signaling pathways in terms of NFkB and AP-1 (Fig. 4C and D). These results

further support the hypothesis that the disease-causing *RAC3* variants may drive different and/or common downstream signaling pathways, leading to variant-dependent dysregulation of cellular processes and, eventually, common as well as patient-specific clinical phenotypes.

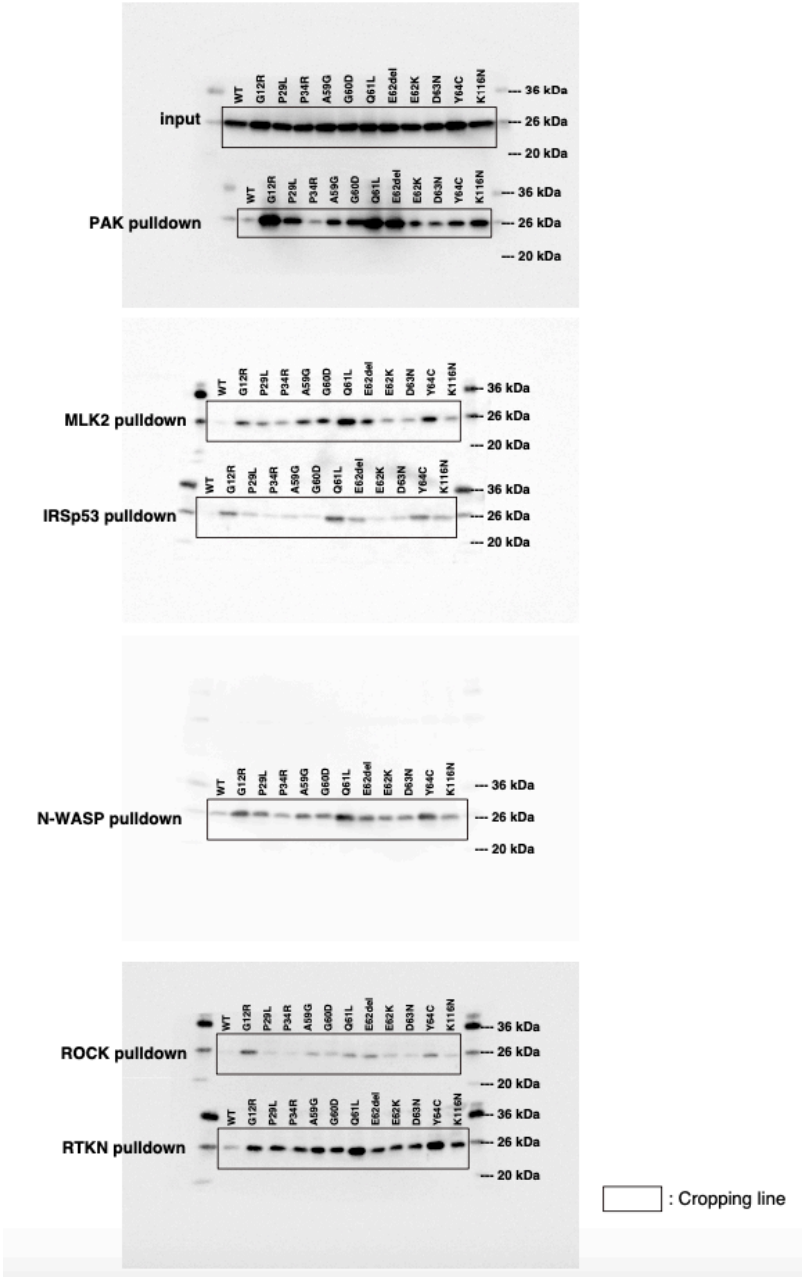


Figure S7. Uncropped blotting data of the pull-down assay. The data in Fig. 4A were prepared by cropping of the original blots. The band intensity was measured by ImageJ software.

3.3.B. *In vivo*

3.3.B.a. Effects of the four *RAC3* variants in the Switch II region on neuronal migration and morphology during corticogenesis

Based on the data obtained from *in vitro* analyses, the disease-causative *RAC3* variants are most likely to be activated and induce abnormal neuronal morphology. Since cell morphology is closely associated with migration, we examined the effects of the *RAC3* variants on the migration of newly generated excitatory cortical neurons *in vivo*. We especially focused on the Switch II region and selected the p.Q61L, p.E62del, p.D63N, and p.Y64C, since the region is a variation hotspot among *RAC1*, *RAC3*, and *CDC42*. Using an *in utero* electroporation-mediated acute gene transfer method, pCAG-Myc-RAC3 or pCAG-Myc vector harboring each variant was co-electroporated with pCAG-EGFP into the progenitor cells in the ventricular zone (VZ) of E14.5 embryonic brains, and localization of transfected cells and their progeny was observed at P0. Neurons expressing the control vector or pCAG-Myc-RAC3 migrated normally to the superficial layer (bin 3; layers II/III) of the cortical plate (CP) (Fig. 5A). In contrast, most cells transfected with pCAG-Myc-RAC3-Q61L, -E62del, -D63N, or -Y64C remained in the ventricular and subventricular zones (VZ/SVZ) and the intermediate zone (IZ) (bin 1) (Fig. 5A). Quantitative analyses confirmed that each variant exhibited statistically similar effects (Fig. 5B).

The result that the expression of the wild type did not statistically affect neuronal cell positioning indicates that the basal activity of RAC3 had no effects on neuronal cell migration and that physiological balance between GTP- and GDP-bound states of RAC3 should be essential for the establishment of cortical architecture during corticogenesis. Based on western blotting analyses with cortical neurons where Rac3 proteins were electroporated, expression level of each protein was found to be comparable and lower than endogenous Rac3 (Fig. S8).

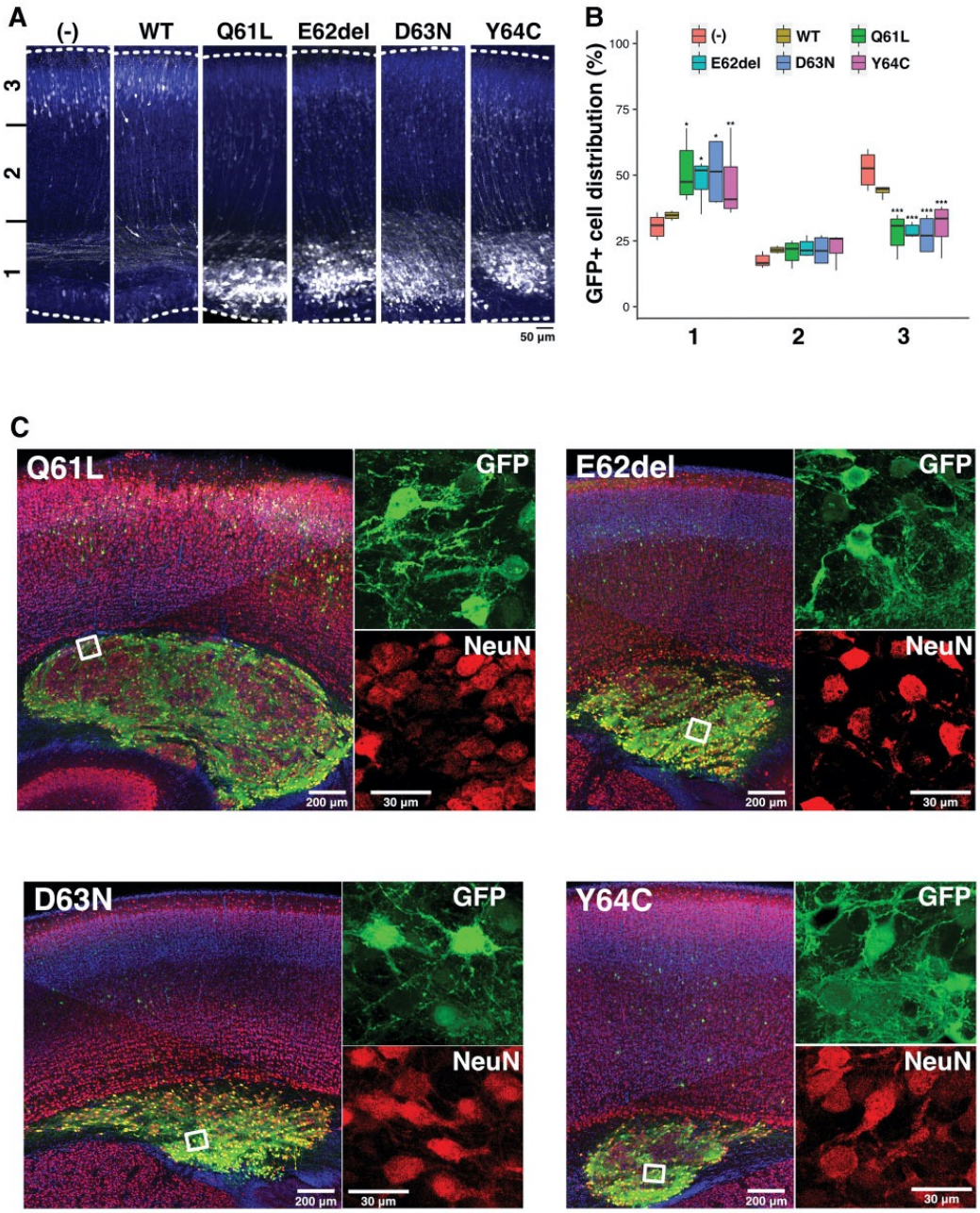


Figure 5. Effects of the four RAC3 variants in the Switch II region on excitatory neuron migration during corticogenesis. (A and **C)** Migration defects of neurons expressing each variant. pCAG-EGFP (0.5 μg) was co-electroporated *in utero* with pCAG-Myc (-), pCAG-Myc-RAC3 (WT), -RAC3-Q61L, -E62del, -D63N, or -Y64C (0.1 μg each) into the VZ progenitor cells at E14.5. Coronal sections were prepared at P0 (A) or P7 (C). Coronal slices were double-stained with anti-GFP (white) and DAPI (blue) for (A) or triple-stained with anti-NeuN (red), anti-GFP (green), and DAPI (blue) for (C). Boxed areas in the *left* panels were magnified in *right* panels (C). Scale bars; 50 μm (A), 200 μm (C, *left*), and 30 μm (C, *right*). **(B)** Quantification of the distribution of GFP-positive neurons in distinct regions of the cerebral cortex (bin 1-3) for each condition in (A). Number of replicates, N ≧ 5. The significance of difference between control (-) and each variant was determined using Dunnett's test and shown in boxplot. (bin 1) WT vs. (-), $p = 0.9498$; Q61L vs. (-), $p = 0.0045$; E62del vs. (-), $p = 0.0172$; D63N vs. (-), $p = 0.0119$; Y64C vs. (-), $p = 0.0356$. (bin 2) WT vs. (-), $p = 0.413$; Q61L vs. (-), $p = 0.535$; E62del vs. (-), $p = 0.185$; D63N vs. (-), $p = 0.439$; Y64C

vs. (-), $p = 0.133$. (bin 3) WT vs. (-), $p = 0.212$; Q61L vs. (-), $p < 0.001$; E62del vs. (-), $p < 0.001$; D63N vs. (-), $p < 0.001$; Y64C vs. (-), $p < 0.001$. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

Although cells expressing respective variants were dominantly distributed at bin 1, a small portion of such neurons still reached the superficial layer of CP (Fig. 5A and B), perhaps due to a relatively low amount of the expression vector incorporated. Given that transfection efficiency into each cell depends on the size of the cell surface area which is physically exposed to the ventricular lumen (cerebrospinal fluid) where plasmids are injected, neurons incorporating low amount of the expression vector are supposed to undergo partial effects of the variant.

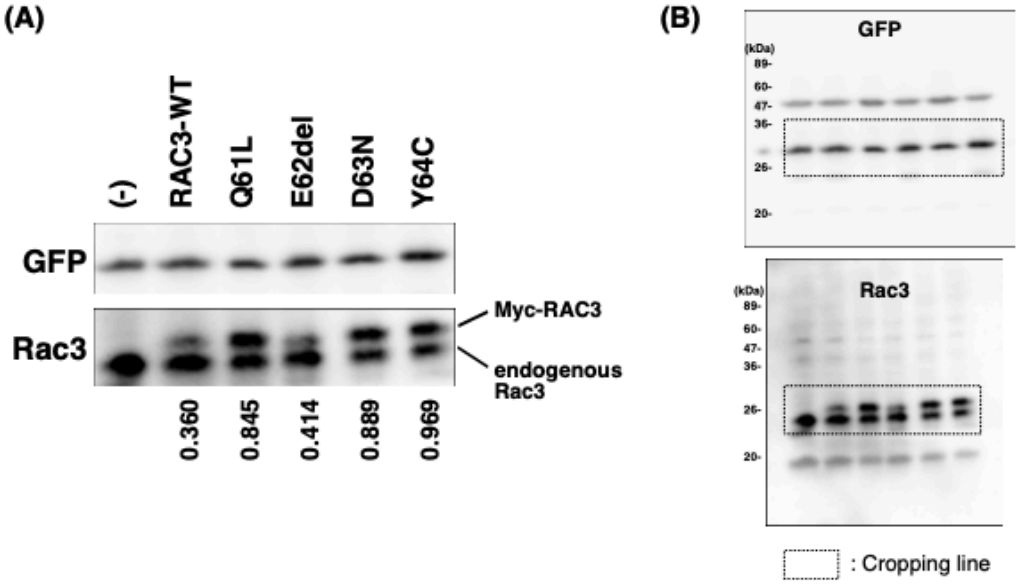


Figure S8. Expression of the disease-causative variants in the Switch II region in cortical neurons with migration defects. (A) pCAG-EGFP (0.5 μg) was co-electroporated *in utero* with pCAG-Myc (-), pCAG-Myc-RAC3 (WT), -RAC3-Q61L, -E62del, -D63N, or -Y64C (0.1 μg each) into the VZ progenitor cells at E14.5. Cortical regions where plasmids were electroporated were separated at E16, and the whole lysates were subjected to western blotting. Endogenous Rac3 and overexpressed RAC3 proteins were visualized with anti-Rac3. Relative band intensities of Myc-RAC3 proteins were calculated with ImageJ software based on densitometry, and normalized against endogenous Rac3. The data shown are representative of three independent experiments. Note that glial cells are hardly generated at E16. **(B)** Uncropped western blot images presented in the Fig. S8A.

When long-term effects of expression of the four RAC3 variants were examined, we noticed the formation of neuronal cell clusters in the VZ/SVZ in developing cerebral cortex at P7 (Fig.

5C). The cells incorporated in the cluster were positive for NeuN, indicating they were differentiated at abnormal positions. These cells also extended neurites in the cluster (Fig. 5C). The results obtained indicate that the four variants prevent, rather than delay, cortical neuron migration, and that RAC3 plays a pivotal role in neuronal migration.

3.3.B.b. Time-lapse imaging of migration of cortical neurons expressing the four *RAC3* variants in the Switch II region

Newborn cortical neurons generated at the VZ primarily exhibit multipolar shapes in the lower IZ, where cells show a slow and irregular movement termed multipolar movement for ~24 h.[59] Neurons then transform into a bipolar shape with a leading process and an axon in the upper IZ, move into the CP, and exhibit a saltatory movement termed radial migration toward pial surface. Given the tight correlation between cell movement and morphology, the abnormal accumulation and cluster formation of the *RAC3* variant-expressing neurons in the SVZ and IZ should be associated with impaired shape change into the bipolar status (multipolar-bipolar transition).

We thus carried out time-lapse imaging to further investigate the morphology of cells stuck in the SVZ and IZ during corticogenesis. To this end, VZ progenitor cells were co-electroporated with pCAG-EGFP together with wild type RAC3- or respective variant-expression vectors at E14.5, and cell migration was monitored from E16 for 15 h in the SVZ, IZ, and lower CP. Consequently, time-lapse imaging revealed clear differences in migration profiles between control cells and those expressing the respective variants. In the control experiments with wild type (WT) RAC3, while a considerable number of GFP-positive cells were positioned in the IZ, many cells advanced slowly toward the pial surface (Fig. 6A and B).

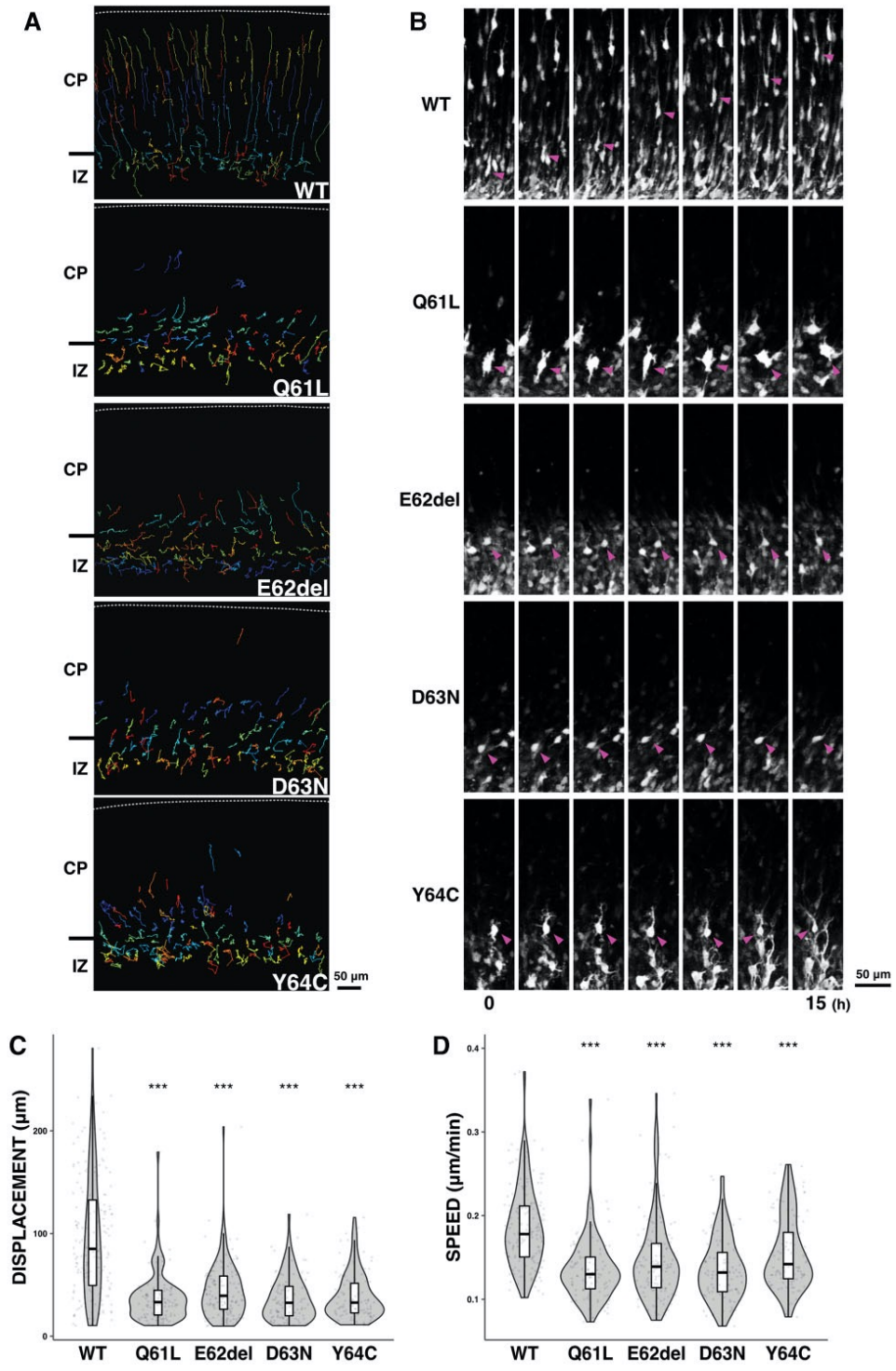


Figure 6. Time-lapse imaging analyses of migration of cortical neurons expressing the four disease-causative *RAC3* variants in the Switch II region. Electroporation was performed as in Fig. 5. Analyses were repeated three times for each case and representative results were shown in (A) and (B). **(A)** Tracing of neurons expressing Myc-RAC3 (WT), -RAC3-Q61L, -E62del, -D63N, or -Y64C in upper IZ - lower CP. Migratory tracks of over 100 cells

were demonstrated as color lines. **(B)** Time-lapse imaging of neurons, which express Myc-RAC3 (WT) or the variants, migrating in the IZ-CP boundary. **(C)** Migration distance of neurons expressing Myc-RAC3 or the 4 variants. Over 100 cells selected in (A) were analyzed. The distance was shown in violin plot with boxplot. The significance of difference between WT and each variant was determined using Dunnett's test. Q61L vs. WT, $p < 2e\text{-}16$; E62del vs. WT, $p < 2e\text{-}16$; D63N vs. WT, $p < 2e\text{-}16$; Y64C vs. WT, $p < 2e\text{-}16$. ***$p < 0.001$. **(D)** Migration velocity of cells expressing Myc-RAC3 or the 4 variants. Over 100 cells selected in (A) were analyzed. The velocity was shown in violin plot with boxplot. The significance of difference between WT and each variant was determined using Dunnett's test. Q61L vs. WT, $p < 1e\text{-}07$; E62del vs. WT, $p < 1e\text{-}07$; D63N vs. WT, $p < 1e\text{-}07$; Y64C vs. WT, $p = 1.16e\text{-}07$. ***$p < 0.001$.

In stark contrast, cells expressing RAC3-Q61L, -E62del, -D63N, or -Y64C were stuck in the IZ. When looking closer, cells expressing RAC3-E62del or -D63N remained round and appeared not to obtain the multipolar status (Fig. 6A and B). On the other hand, cells expressing RAC3-Q61L or -Y64C were observed to transform into the multipolar shape but then fail to undergo the multipolar-bipolar transition (Fig. 6A and B). Quantification analyses revealed that migration distance of the variant-expressing neurons at the IZ-CP boundary was much shorter than that of the control cells (Fig. 6C). The average migration speed of neurons expressing each *RAC3* variant was also reduced when compared to that of control cells (Fig. 6D).

3.3.B.c. Role of PAK1 as a downstream effector of RAC3 in neuronal migration *in vivo*

Since PAK1 is a downstream effector for RAC3 and activating *PAK1* variants cause NDD,[60] dysregulation of this kinase is most likely to be involved in the pathogenesis of *RAC3* variants. We thus investigated the possible involvement of PAK1 in the migration defects caused by the four *RAC3* variations localized in the Switch II region. When pCAG-EGFP was co-electroporated with pCAG-Myc-RAC3-D63N, -E62del, or -Y64C, together with pCAG-Flag-PAK1KA encoding a kinase-negative version of PAK1, the positional defects of GFP-positive cells were partially rescued at P0 (Fig. 7A-C).

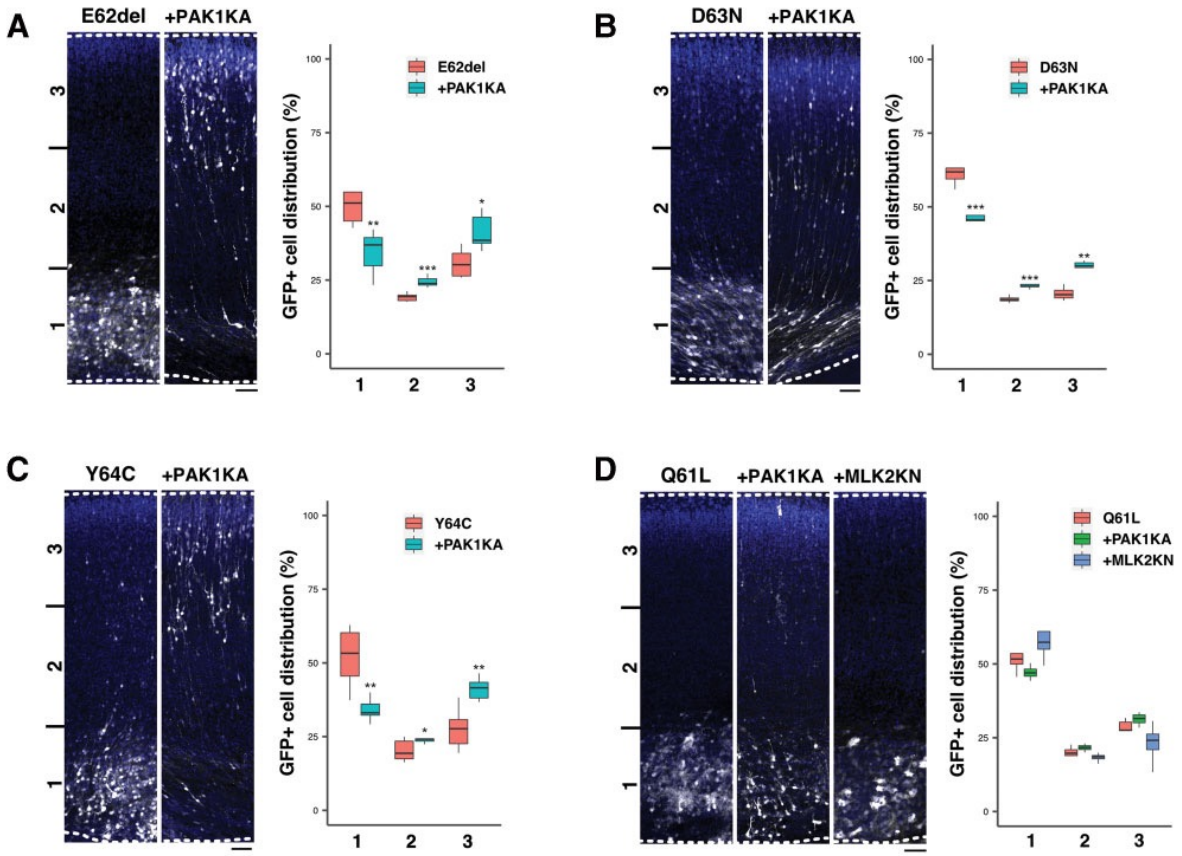


Figure 7. Rescue effects by a dominantly negative PAK1 on migration defects caused by the four variants in the Switch II region. pCAG-Myc-RAC3-E62del **(A)**, -D63N **(B)**, or -Y64C **(C)** (0.1 μg each) was co-electroporated with pCAG-EGFP (0.5 μg) together with pCAG-Flag vector (1.0 μg, control) (*left*) or pCAG-Flag-PAK1KA (1.0 μg) (*right*). **(D)** pCAG-Myc-RAC3-Q61L (0.1 μg) was co-electroporated with pCAG-EGFP (0.5 μg) together with 1.0 μg each of pCAG-Flag vector (control) (*left*), pCAG-Flag-MLK2KN (*middle*), or pCAG-Flag-PAK1KA (*right*). Analysis was done as in Fig. 5. Quantification results of the distribution of GFP-positive neurons in distinct regions (bin 1-3) of the cerebral cortex were shown in boxplot for each condition. Number of replicates, N ≧ 5. The significance of difference between control and each rescue condition was determined using Welch's t test (A-C) or Dunnett's test (D), and shown in boxplot.

(A, bin 1) E62del vs. +PAK1KA, $p = 0.005591$. (A, bin 2) E62del vs. +PAK1KA, $p = 0.0006076$. (A, bin 3) E62del vs. +PAK1KA, $p = 0.01177$. (B, bin 1) D63N vs. +PAK1KA, $p = 0.0007075$. (B, bin 2) D63N vs. +PAK1KA, $p = 0.0009054$. (B, bin 3) D63N vs. +PAK1KA, $p = 0.001021$. (C, bin 1) Y64C vs. +PAK1KA, $p = 0.005963$. (C, bin 2) Y64C vs. +PAK1KA, $p = 0.04256$. (C, bin 3) Y64C vs. +PAK1KA, $p = 0.003158$. (D, bin 1) Q61L vs. +PAK1KA, $p = 0.338$; Q61L vs. +MLK2KN, $p = 0.807$. (D, bin 2) Q61L vs. +PAK1KA, $p = 0.311$; Q61L vs. +MLK2KN, $p = 0.993$. (D, bin 3) Q61L vs. +PAK1KA, $p = 0.393$; Q61L vs. +MLK2KN, $p = 0.673$. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$; Scale bar, 50 μm.

The observed rescue effects by PAK1KA confirmed that these three variants cause a hyper-activation of PAK1, which is supposed to be a crucial pathogenic mechanism of *RAC3*-related disorder caused by these three variations. To confirm the variant-mediated activation of PAK1, these three variants were expressed with PAK1 in COS7 cells, or without PAK1 in primary

cultured cortical neurons. Consequently, activation of exogenous and endogenous PAK1 was observed in COS7 cells and cortical neurons, respectively (Fig. S9A, B, D and E).

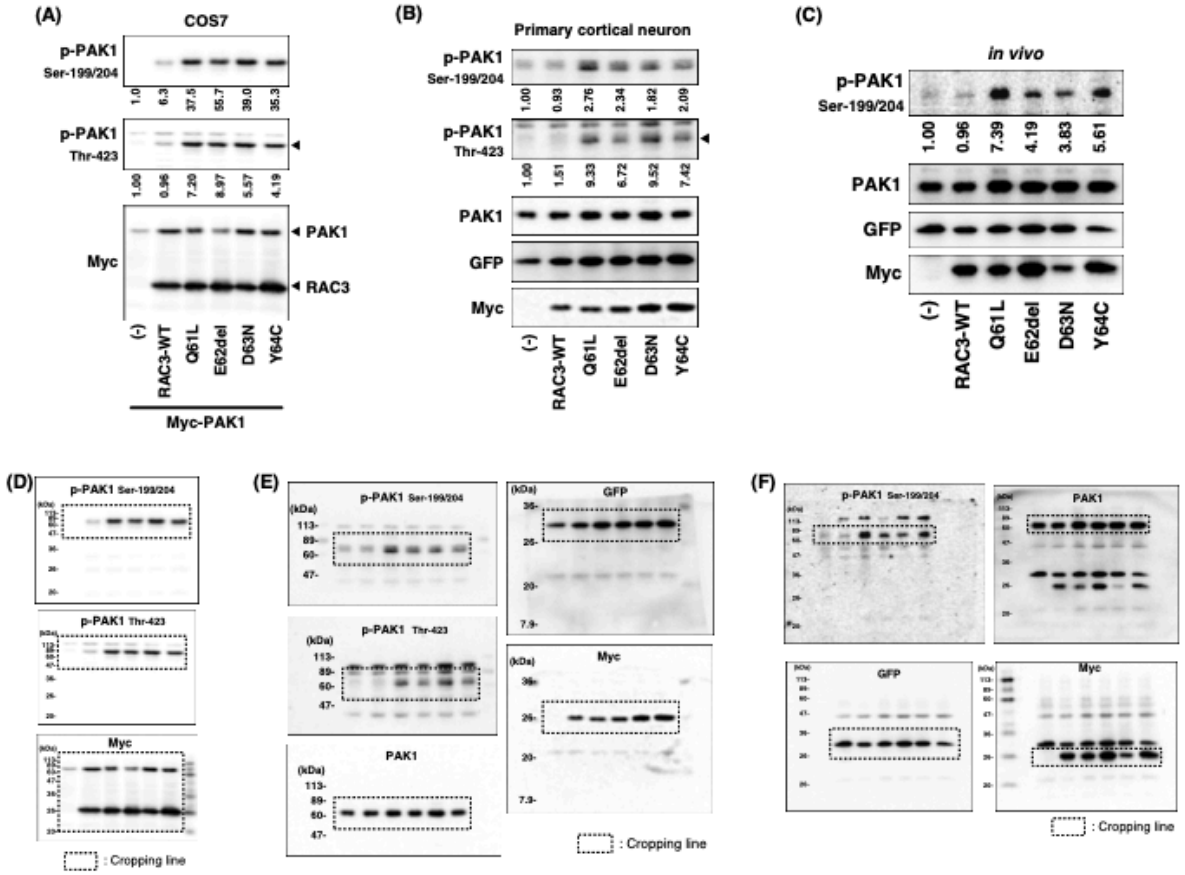


Figure S9. Activation of PAK1 by the disease-causative variants in the Switch II region. (A) Activation of exogenous PAK1 in COS7 cells. Cells were transfected with pCAG-Myc-PAK1 (0.1 μg) together with pCAG-Myc (-), pCAG-Myc-RAC3 (WT), -Q61L, -E62del, -D63N, or -Y64C (0.1 μg each). After 24h, cell lysates were prepared and subjected to western blot analyses. Activated Myc-PAK1 was detected by anti-phospho-Ser199/204-PAK1 or anti-phospho-Thr423-PAK1. Myc-PAK1 and Myc-RAC3 were detected with anti-Myc. Relative band intensities of activated PAK1 were calculated with ImageJ software based on densitometry, and normalized against Myc-PAK1. **(B)** Activation of endogenous PAK1 in primary cortical neurons. Dissociated neurons from E16 mice were transfected with pCAG-GFP (0.5 μg) together with pCAG-Myc (-), pCAG-Myc-RAC3 (WT), -Q61L, -E62del, -D63N, or -Y64C (2.0 μg each). After cell lysates were prepared at 3 div, activated endogenous PAK1 was detected as in (A). Endogenous PAK1, GFP, and Myc-RAC3 were detected with anti-PAK1, anti-GFP, and anti-Myc, respectively. The data shown are representatives of three independent experiments. Relative band intensities of endogenous activated PAK1 were calculated with ImageJ software based on densitometry, and normalized against endogenous PAK1. **(C)** Activation of endogenous PAK1 in cortical tissue. pCAG-EGFP (0.5 μg) was co-electroporated *in utero* with pCAG-Myc (-), pCAG-Myc-RAC3 (WT), -RAC3-Q61L, -E62del, -D63N, or -Y64C (0.2 μg each) into the VZ progenitor cells at E14.5. Cortical regions where plasmids were electroporated were separated at E16, and the whole lysates were subjected to western blotting. These lysates were prepared from different mouse (10 samples each). Endogens activated PAK1 was visualized with anti-phospho-Ser199/204-PAK1.

The blot was then reprobed for endogenous PAK1, GFP, and Myc. Relative band intensities of endogenous activated PAK1 were calculated as in (B). **(D - F)** Uncropped western blot images presented in Supplementary Fig. 9A - C.

In addition, when these variants were electroporated into embryonic mice brains, endogenous PAK1 activation was again detected in cortical neurons (Fig. S9C and F). On the other hand, PAK1KA had no effects on the migration defects produced by RAC3-Q61L under the same experimental conditions (Fig. 7D). This result implicates the possibility that other downstream effector(s) is also involved in the pathogenicity of *RAC3*-related disorder. To answer this question, we selected another kinase MLK2, an activator for broad MAP kinase cascades including JNK (c-Jun N-terminal kinase), ERK (extracellular signal-regulated kinase) and p38, as the next candidate effector molecule involved in RAC3-Q61L-mediated pathophysiological mechanism (*Nagata et al., 1998*). However, MLK2 failed to rescue the aberrant migration phenotype by RAC3-Q61L (Fig. 7D), indicating that yet unidentified downstream effector(s) plays a crucial role in the variation-mediated signaling dysregulation.

3.3.B.d. Effects of the four *RAC3* variants in the Switch II region on axon development during corticogenesis

Anomalies of corpus callosum in subjects with *RAC3*-related disorder, such as hypoplasia and agenesis, reflect defects in axon growth (Fig. 1C). Specifically, an anomaly at the level of the splenium strongly suggests impaired axon elongation of pyramidal neurons in layer II/III and V of the temporal, parietal, and occipital lobes. In this context, the *in utero* electroporation method is available to examine axon elongation from pyramidal neurons in layer II/III in the parietal lobe during mouse brain corticogenesis. When the wild type of RAC3 (as a control), RAC3-Q61L, -E62del, -D63N, or -Y64C were introduced into the VZ progenitor cells at E14.5 and axon bundle of corpus callosum was visualized at P0, neurons expressing RAC3-Q61L, -E62del and -Y64C did

not project axons under the conditions where control axon bundle reached the midline (Fig. 8A and B).

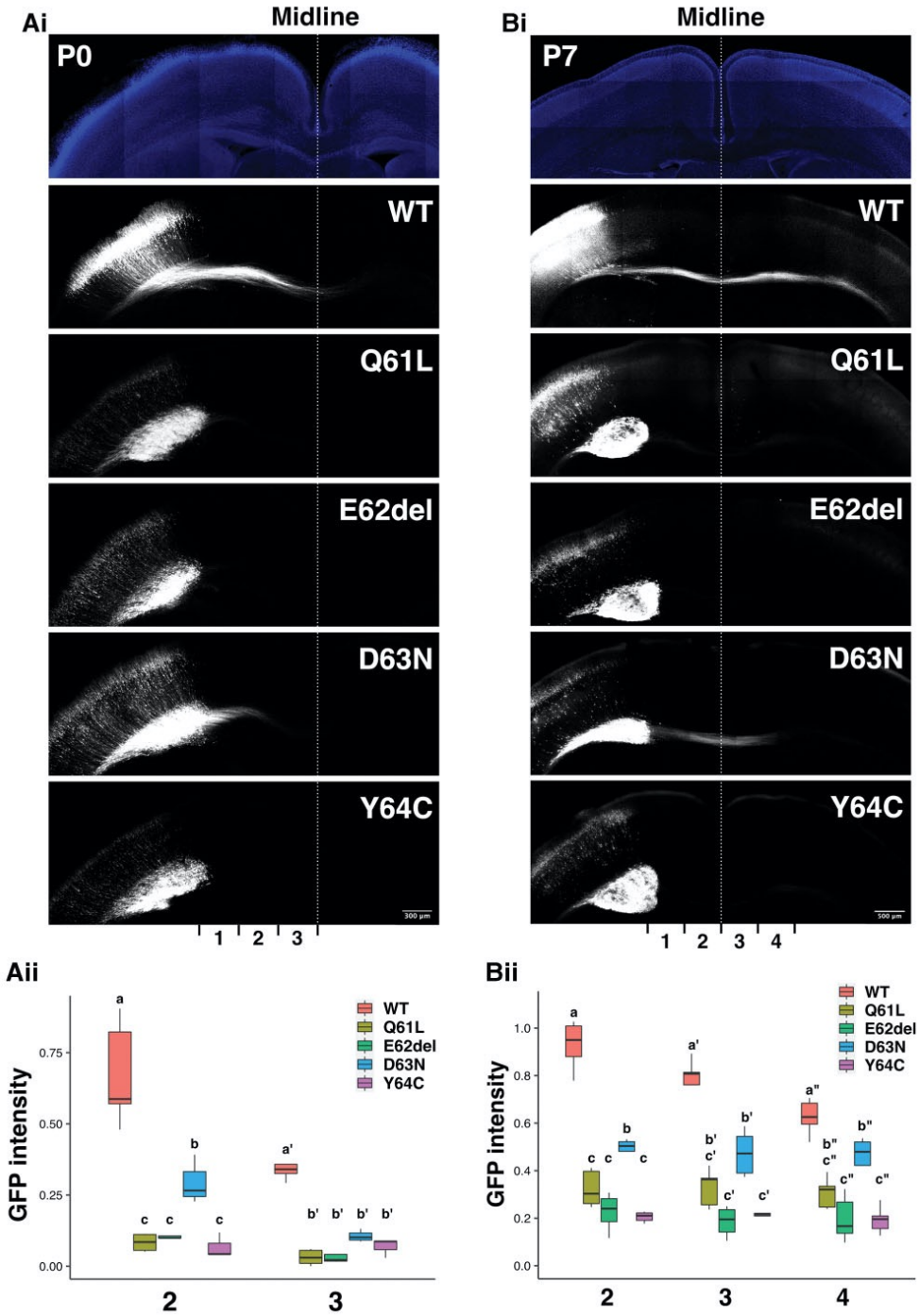


Figure 8. Role of the four disease-causative variants in the Switch II region in the axon growth *in vivo.* (A and **C)** pCAG-GFP was co-electroporated with pCAG-Myc-RAC3 (WT), -RAC3-Q61L, -E62del, -D63N, or -Y64C (0.1 μg each) into the VZ progenitor cells at E14.5. Coronal sections were prepared at P0 (A) or P7 (C), and stained with anti-GFP (white). DAPI staining (blue) of a slice was also shown (*top* panels). Scale bars; 300 μm (A) and 500 μm (C). **(B** and **D)** The GFP intensity of the callosal axon was measured at P0 (B) or P7 (D) in different regions [bins 1 -

3 for (A) and bins 1 - 4 for (C)], and then the relative intensities of bins were normalized with bin 1 as 1.0. Number of replicates, N ≧ 4. Different letters above bars represent significant differences, $p < 0.05$, according to a Tukey's test.

(B, bin 2) Q61L vs. WT, $p < 0.001$; E62del vs. WT, $p < 0.001$; D63N vs. WT, $p < 0.001$; Y64C vs. WT, $p < 0.001$; E62del vs. Q61L, $p = 1.0000$; D63N vs. Q61L, $p = 0.0205$; Y64C vs. Q61L, $p = 0.9756$; D63N vs. E62del, $p = 0.0444$; Y64C vs. E62del, $p = 0.9835$; Y64C vs. D63N, $p = 0.0235$. (B, bin 3) Q61L vs. WT, $p < 0.001$; E62del vs. WT, $p < 0.001$; D63N vs. WT, $p < 0.001$; Y64C vs. WT, $p < 0.001$; E62del vs. Q61L, $p = 0.997$; D63N vs. Q61L, $p = 0.646$; Y64C vs. Q61L, $p = 0.997$; D63N vs. E62del, $p = 0.537$; Y64C vs. E62del, $p = 0.976$; Y64C vs. D63N, $p = 0.936$. (C, bin 2) Q61L vs. WT, $p < 0.001$; E62del vs. WT, $p < 0.001$; D63N vs. WT, $p < 0.001$; Y64C vs. WT, $p < 0.001$; E62del vs. Q61L, $p = 0.1794$; D63N vs. Q61L, $p = 0.0132$; Y64C vs. Q61L, $p = 0.0755$; D63N vs. E62del, $p < 0.001$; Y64C vs. E62del, $p = 0.9784$; Y64C vs. D63N, $p < 0.001$. (C, bin 3) Q61L vs. WT, $p < 0.001$; E62del vs. WT, $p < 0.001$; D63N vs. WT, $p < 0.001$; Y64C vs. WT, $p < 0.001$; E62del vs. Q61L, $p = 0.31921$; D63N vs. Q61L, $p = 0.38265$; Y64C vs. Q61L, $p = 0.51491$; D63N vs. E62del, $p = 0.00505$; Y64C vs. E62del, $p = 0.99767$; Y64C vs. D63N, $p = 0.01394$. (C, bin 4) Q61L vs. WT, $p < 0.001$; E62del vs. WT, $p < 0.001$; D63N vs. WT, $p = 0.01014$; Y64C vs. WT, $p < 0.001$; E62del vs. Q61L, $p = 0.48422$; D63N vs. Q61L, $p = 0.34330$; Y64C vs. Q61L, $p = 0.45377$; D63N vs. E62del, $p = 0.00869$; Y64C vs. E62del, $p = 0.99996$; Y64C vs. D63N, $p = 0.00904$.

On the other hand, axon elongation was observed but resulted in significantly delayed in neurons expressing RAC3-D63N (Fig. 8A and B). When we further analyzed the long-term effects at P7, control neurons extended the axon bundle into the contralateral white matter, whereas axon elongation could not be detected yet for cells expressing RAC3-Q61L, -E62del and -Y64C (Fig. 8C and D). Notably, although axons from the hemisphere containing RAC3-D63N-expressing cells eventually reached the contralateral white matter at P7, the bundle was observed to be thinner compared to the control neurons (Fig. 8C). Then, rescue effects by dominant negative PAK1 on the defects in axon growth were examined *in vivo*. When PAK1KA was co-expressed with RAC3-Q61L, -E62del, -D63N, or -Y64C, the defective phenotypes were at least partially rescued for RAC3-E62del, -D63N, or -Y64C (Fig. S10). In contrast, the phenotype by RAC3-Q61L was not rescued, as in the case of cortical neuron migration (Fig. S10).

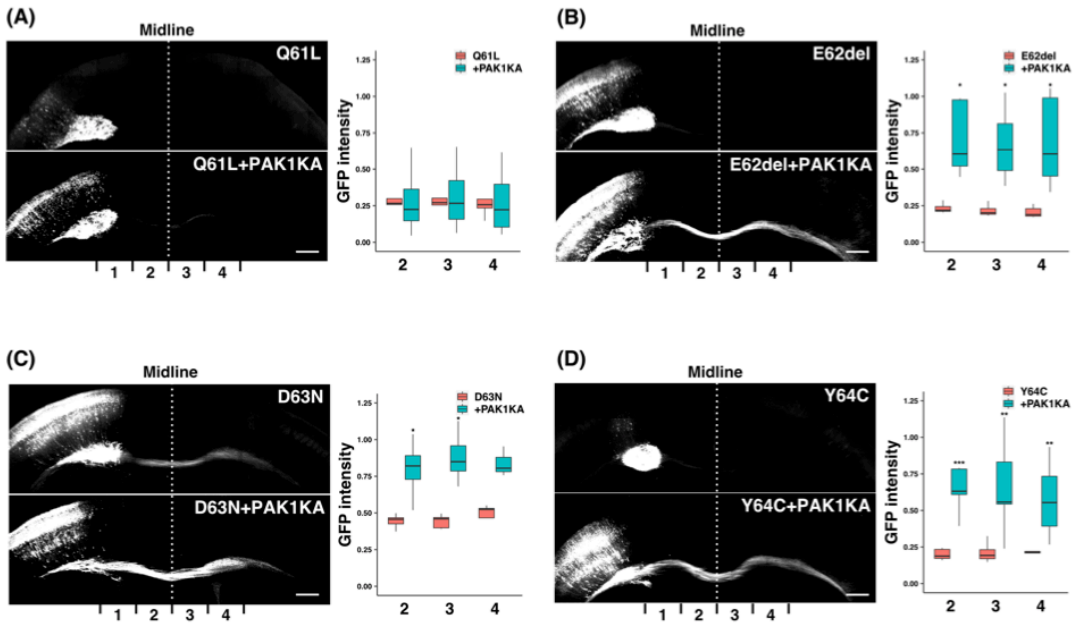


Figure S10. Rescue effects by a dominantly negative PAK1 on axon growth defects caused by the four variants in the Switch II region. pCAG-Myc-RAC3-Q61L **(A)**, E62del **(B)**, -D63N **(C)**, or - Y64C **(D)** (0.1 μg each) was co-electroporated with pCAG-EGFP (0.5 μg) together with pCAG-Flag vector (1.0 μg, control) (*upper*) or pCAG-Flag-PAK1KA (1.0 μg) (*lower*). Analyses were done at P7 as in Fig. 8. Scale bars, 500 μm. The GFP intensity of the callosal axon was measured in different regions (bins 1 - 4), and then the relative intensities of bins were normalized with bin 1 as 1.0. Number of replicates, N ≧ 4. The significance of difference between control and each rescue condition was determined using Welch's t test. (A, bin 2) Q61L vs. +PAK1KA, $p = 0.9833$. (A, bin 3) Q61L vs. +PAK1KA, $p = 0.8981$. (A, bin 4) Q61L vs. +PAK1KA, $p = 0.9638$. (B, bin 2) E62del vs. +PAK1KA, $p = 0.01392$. (B, bin 3) E62del vs. +PAK1KA, $p = 0.03876$. (B, bin 4) E62del vs. +PAK1KA, $p = 0.02716$. (C, bin 2) D63N vs. +PAK1KA, $p = 0.0421$. (C, bin 3) D63N vs. +PAK1KA, $p = 0.02379$. (C, bin 4) D63N vs. +PAK1KA, $p = 0.08895$. (D, bin 2) Y64C vs. +PAK1KA, $p = 0.0006848$. (D, bin 3) Y64C vs. +PAK1KA, $p = 0.007151$. (D, bin 4) Y64C vs. +PAK1KA, $p = 0.007304$. ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

These results strongly suggest that abnormal PAK1 activation contributes to the impaired axon elongation by p.E62del, p.D63N, and p.Y64C variations, which leads to the corpus callosum dysgenesis, while yet unidentified effector(s) is involved in the axon phenotype by p.Q61L variation. Further analyses are required to determine if the observed impaired axon growth is a primary phenotype or secondary to migration defects.

4. Discussion

4.1. *RAC3* variants are associated with heterogeneous neurodevelopmental phenotypes

We identified seven distinct *de novo* missense variants (five novel, one recurrent) and a novel *de novo* in-frame deletion in *RAC3* in 10 patients presenting with NEDBAF, which is a complex syndromic NDD characterized by a moderate to severe psychomotor delay leading to ID, associated with peculiar neurological and extra-neurological features. The potential contribution of pathogenic CNVs to the reported phenotypes was excluded in all subjects. Regarding the neuroimaging features of *RAC3* deficiency in this study, the most frequent findings were abnormalities in corpus callosum and cortical development, collectively noted in 100% and 90% of the individuals reported so far, respectively. In particular, the corpus callosum was found to be frequently dysplastic, with thinner splenium and/or a thick genu, indicating the presence of axonal growth anomalies. On the other hand, cortical malformations included diffuse dysgyria, polymicrogyria, and small gray matter heterotopias, showing remarkable evidence of abnormal neuronal migration. Although disruption of corticogenesis has been presumed to be crucial for the pathogenesis and pathophysiology of *RAC3*-related disorder, the underlying molecular mechanisms remained unknown. We thus investigated the pathophysiological significance of *RAC3* variants causing NEDBAF during cortical development *in vitro* and *in vivo*.

4.2. RAC3 variants impair the interactions with regulatory proteins in variant- and context-dependent manners

We assume that the pathogenic effects of the seven variants in the Switch II region (p.A59G, p.G60D, p.Q61L, p.E62K, p.E62del, p.D63N, and p.Y64C) are related to the aberrant interactions with the regulatory proteins in variant- and context-dependent manners, since the region is

essential for the physical association with GEFs and GAPs. Being positioned in the effector-binding loop (aa32 - 41), which is overlapping with the Switch I region, the p.P34R variant might impact the interaction with downstream effectors as well as GEFs and GAPs. Indeed, when compared to other variants, RAC3-P34R displayed weaker affinities to RBRs of all effector molecules tested (Fig. 4A). Notably, the patient harboring p.P34R presented syntelencephaly, a rare brain malformation characterized by an abnormal midline connection of the cerebral hemispheres, suggesting a role of RAC3 in the cleavage process of prosencephalon (*Hiraide et al., 2019*). Meanwhile, p.G12R and p.K116N affect residues conserved in most small GTPases within the G1 and G4 boxes, respectively. In line with the different roles of these boxes in the interaction with guanine nucleotides, these two variants may affect biochemical properties distinctly. Indeed, while the p.G12R variation activated the protein by suppressing GTP-hydrolysis activity, p.K116N facilitated GTP/GDP-exchange reaction. Collectively, biochemical analyses strongly suggest that the 11 *RAC3* variants analyzed show "gain-of-function" phenotypes; they preferentially bind GTP in variant type-specific modes and are hyper-activated in various degrees. This hypothesis is consistent with the results that all the *RAC3* variants exhibited lamellipodia formation with cell rounding when overexpressed in primary cultured hippocampal neurons.

Although the 11 *RAC3* variants are supposed to be activated, it is surprising that these variants displayed different affinities to RBRs of various downstream effectors, including PAK1, MLK2, IRSp53, N-WASP, ROCK, and RTKN. Pull-down assay with PAK1-RBR has so far been conducted to assess the "activation state" of RAC by measuring the amount of the GTP-bound form co-precipitated, under the tacit recognition that the interaction with an RBR of any effector molecule precisely reflects the GTP-bound activated state. However, the results obtained here indicate that the RAC3-GTP amount precipitated by the RBR of PAK1 or other effectors was not

necessarily correlated with the activation state of RAC3. We thus suggest that disease-causative *RAC3* variants even within the same structural/functional domains differently modulate the activation state of RAC3, and show specific spectra in the interaction with effectors, leading to abnormal upregulation of relevant effectors in variant-type- and context-dependent manners. The hyper-activation of certain signaling pathways may underlie the pathogenesis and pathophysiology of the divergent clinical features observed in each patient. Meanwhile, each variant is possible to not only activate particular downstream effectors differently but also exert dominant negative effects on some other signaling pathways simultaneously. In any case, intracellular signaling networks should be affected by each variant both in a qualitatively and quantitatively different manner, which may contribute to the variable phenotypes observed in patients with *RAC3*-related disorder.

4.3. *RAC3* variants affect neuronal migration and morphogenesis during corticogenesis

To clarify the pathophysiological significance of *RAC3* variations *in vivo*, we focused on p.D63N, p.E62del, p.Y64C, and p.Q61L in the Switch II region, a variation hotspot not only for *RAC3* but also for *RAC1* and *CDC42*. Although these four variants should affect different signaling pathways *in vitro*, their expression in cortical neurons resulted in similar phenotypes *in vivo*: severe defects in neuronal migration, mispositioning, and eventual cluster formation in the IZ and SVZ, as well as defects in axon extension to the contralateral hemisphere during corticogenesis. As for the migration phenotypes, time-lapse imaging revealed differences among the analyzed variants. New-born neurons expressing RAC3-Q61L or -Y64C appeared to become multipolar but did not transform into the bipolar status, whereas cells expressing RAC3-E62del or -D63N failed to become even multipolar. Further analyses are required to elucidate the pathophysiological

meaning of this phenotypic difference. Cluster formation by mislocalized NeuN-positive neurons should account for the pathogenic mechanisms underlying the heterotopia and, partially, polymicrogyria and dysgyria, which represent the main neuronal migration/positioning abnormality observed in *RAC3*-related disorder. Also, considering the corpus callosum anomalies observed in affected individuals, it is a reasonable conclusion that the four *RAC3* variants suppress axon elongation *in vivo*. The partial suppression of axon growth by RAC3-D63N may explain the absence of "white matter thinning" in the individual harboring this variant.

Since activating *PAK1* variants have been shown to cause NDDs (*Zhang et al., 2020*), dysregulation of PAK1 may be involved in the migration defects associated with the *RAC3* variants. Indeed, the abnormal phenotypes caused by p.D63N, p.E62del, and p.Y64C were rescued by a kinase-negative version of PAK1, PAK1KA. We thus concluded that these three variants hyper-activated PAK1 and provoked defects in neuronal migration during corticogenesis, irrespective of their different affinities to PAK1 *in vitro*. On the other hand, despite the strong affinity, PAK1 did not appear to be related to the RAC3-Q61L-mediated migration defects, since PAK1KA did not exert rescue effects. These results strongly suggest that PAK1 is not the sole molecule regulating cortical neuron migration downstream of RAC3.

4.4. Analogies and differences between RAC3 and RAC1

De novo RAC1 missense variants have also been identified in NDDs with GDD/ID and brain size abnormalities as core phenotypes (Mental Retardation autosomal dominant 48, MRD48, OMIM #617751) (*Reijnders et al., 2017*). While Rac3 is mainly, if not exclusively, expressed in developing and adult neurons, Rac1 is ubiquitously expressed from the early embryonic stage. Despite the different expression profiles, the precise functional difference between RAC1 and

RAC3 has not been investigated. Rather, RAC1-3 have seemingly been equated in various analyses, since they show very similar phenotypes *in vitro*, especially when overexpressed (*Aspenström et al., 2004*). RAC1 and RAC3 are primarily divergent in the last nine carboxy-terminal residues, which include a polybasic region and an adjacent CAAX box, where the C is a Cys, the two A residues are aliphatic amino acids, and the X can be any residue (*Haataja et al., 1997*). The CAAX box is post-translationally modified and is crucial, together with the polybasic region, for the subcellular localization of the protein (*Roberts et al., 2008*). Given the highly homologous structure and common effector molecules, the differences in the pathogenic mechanisms and clinical phenotypes associated with *RAC1* and *RAC3* variants may be attributable to diverse subcellular distribution, at least partially. We assume that differentially distributed *RAC1* and *RAC3* variants may spatiotemporally hyper-activate interactable effectors in context-dependent manners.

The *RAC1* missense variants associated with MRD48 exert their pathophysiological functions differently from *RAC3* (*Reijnders et al., 2017*). The p.Y64D variant was determined to be constitutively activated, since its expression in fibroblasts resulted in a rounding shape and formation of lamellipodia (*Reijnders et al., 2017*). The p.C18Y and p.N39S variants were instead considered as dominant negative alleles, based on fibroblast morphology analyses and *in vivo* zebrafish analyses (*Reijnders et al., 2017*). Meanwhile, three additional variants (p.V51M, p.P73L, and p.C157Y) were categorized as neither active nor inactive versions and considered to have context-dependent effects (*Reijnders et al., 2017*). Although GTP/GDP-binding states were not determined biochemically, these results show that 1) the variant position may determine the activation state of RAC1, ranging from dominant negative to neutral to constitutively active, and 2) the precise control of RAC1 activity is crucial for correct brain development. In this context,

pathogenic variants in *TRIO*, encoding a GEF which activates RAC1 through the first GEF domain (GEFD1), were shown to cause neurodevelopmental impairment, behavioral disturbances, microcephaly or macrocephaly, and skeletal features [autosomal dominant intellectual developmental disorder with microcephaly (MRD44, OMIM #617061) or macrocephaly (MRD63, OMIM #618825)] (*Ba et al., 2016; Pengelly et al., 2016*). Affected individuals with variations in *TRIO* present with variable neurodevelopmental phenotypes. For example, missense changes affecting the GEFD1 domain were found to inhibit TRIO function and thereby inactivate RAC1. In contrast, missense variations in the seventh spectrin domain of TRIO, a variation hotspot, resulted in RAC1 hyperactivation, further supporting the relevance of the tight control of RAC1 signaling during brain development (*Barbosa et al., 2020*).

4.5. *RAC3* variants act through gain of function mechanisms: analogies with *CDC42*

Unlike the case of RAC1, we concluded that all the 11 tested *RAC3* variants act as GTP-bound active versions in varying degrees. From the results obtained in this and previous studies, the variant-dependent spatio-temporal dysregulation (hyper-activation or inhibition) of subsets of common downstream effectors appears to cause diverse phenotypic features in *RAC1*- and *RAC3*-related disorders. Given that variants affecting RAC3 function cause polymicrogyria, heterotopia and dysgyria, which are only occasional in *RAC1*-related disorder (*Reijnders et al., 2017*), we assume that RAC3 plays a central role in neuronal migration *in vivo*, although Rac3 knockout mice displayed little phenotype in terms of cortical architecture, perhaps due to compensation by Rac1 (*Corbetta et al., 2005*). In contrast, *RAC1* variants are associated with abnormalities in brain size (extraordinary spread from -5 to +4.5 SD of occipital-frontal circumference), which are rare among *RAC3* patients, suggesting a major role of RAC1 in neurogenesis and/or apoptosis (*Zamboni et al.,*

2018; Reijnders et al., 2017). Since the disruption of synaptic functions is involved in the pathogenesis of GDD/ID, RAC1 and RAC3 are supposed to cooperate to build up the neuronal signaling network and functions (Fig. S11).

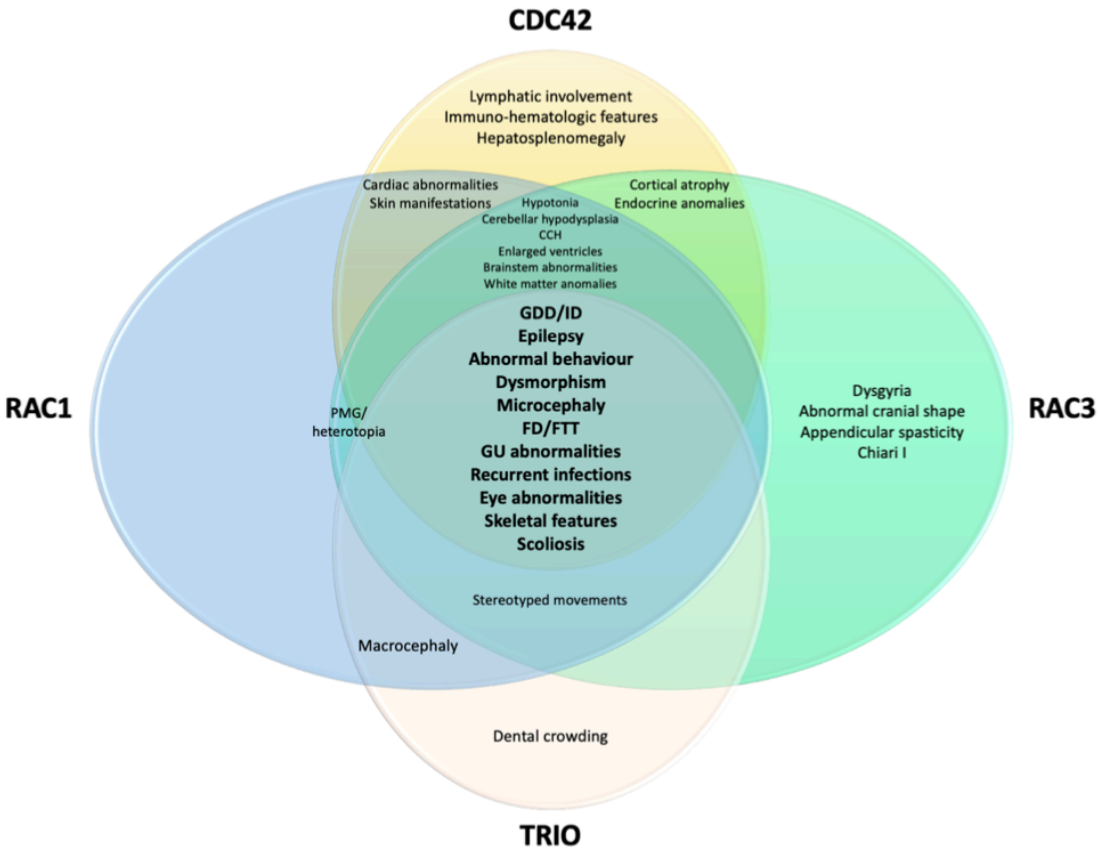


Figure S11. Clinical overlap between Rho GTPase related disorders. Overlapping phenotypic features associated with *RAC1* (RAS-related c3 botulinum toxin substrate 1 - OMIM *602048; mental retardation, autosomal dominant 48 - OMIM #617751), *CDC42* (CELL DIVISION CYCLE 42 - OMIM *116952; TAKENOUCHI-KOSAKI SYNDROME - OMIM #616737), *RAC3* (RAS-related c3 botulinum toxin substrate 3 - OMIM *602050; neurodevelopmental disorder with structural brain anomalies and dysmorphic facies - OMIM #618577), and *TRIO* (Triple functional domain - * 601893; intellectual developmental disorder, autosomal dominant 44, with microcephaly or macrocephaly; OMIM #617061 and #618825). Abbreviations: CCH = corpus callosum hypoplasia; FD = feeding difficulties; FTT = failure to thrive; GDD = global developmental delay; GU = genitourinary; ID = intellectual disability; PMG = polymicrogyria.

Another Rho family protein CDC42 is essential for the control of cell polarity, migration, endocytosis, and cell cycle (*Johnson et al., 1999*). *De novo* missense variations in *CDC42* have been reported in patients with clinically heterogeneous but overlapping phenotypes (Takenouchi-Kosaki syndrome, TKS, OMIM #616737) (*Martinelli et al., 2018; Johnson et al., 1999; Takenouchi et al., 2015*). Biochemical assays revealed that all the variants were activated as in the case of *RAC3* variants (*Martinelli et al., 2018; Hamada et al., 2020*). Considering that CDC42 and RAC proteins are strictly linked in their physiological functions through a variety of shared downstream effectors, it is plausible that the clinical features of *CDC42*-related disorder overlap with those of *RAC1*- and *RAC3*-related conditions. The intersection of the pathogenic mechanisms underlying *RAC1*-, *RAC3*-, *CDC42*-, and *TRIO*-related disorders is reflected in their overlapping clinical phenotypes (Fig. S11).

5. Conclusion

In the present study, we identified six novel *de novo* pathogenic variants in *RAC3* in unrelated individuals with NEDBAF. A variable degree of biochemically activated states was observed for the novel and previously reported RAC3 variants tested. Activation of these variants was also confirmed biologically with primary hippocampal neurons. It is noteworthy that each variant displayed different affinities to a variety of effector molecules *in vitro*, which may dysregulate downstream cellular pathways in variant-specific manners and contribute to the pathogenesis of patient-specific clinical features as well as common ones.

We then performed *in vivo* analyses of four variants (p.D63N, p.E62del, p.Y64C, and p.Q61L) affecting the Switch II region, a variation hot spot. They prevented neuronal migration during corticogenesis, through dysregulation of PAK1-mediated signaling pathway in some cases.

It is notable that the abnormally positioned neurons eventually formed clusters in the VZ/SVZ in developing cerebral cortex, which may be an underlying mechanism of heterotopia, polymicrogyria, and dysgyria. These four variations also caused defective axon development *in vivo,* explaining their involvement in corpus callosum hypoplasia/agenesis. Collectively, we conclude that all the *RAC3* variants identified so far are gain-of-function, and variation-dependent activation of downstream signaling pathways may contribute to the heterogeneous phenotypes observed in individuals with *RAC3*-related disorder.

The results obtained in this study contribute to the understanding of the molecular machinery underlying this complex condition, providing insights for the development of novel targeted drugs aimed at interfering with the disease pathogenesis.

References

Alan J.K., et al. Mutationally activated Rho GTPases in cancer. Small GTPases. 2013;4(3):159-163. doi:10.4161/sgtp.26530

Albertinazzi C., et al. Overexpression of a neural-specific Rho family GTPase, cRac1B, selectively induces enhanced neuritogenesis and neurite branching in primary neurons. Journal of Cell Biology. 1998;142(3):815-825. doi:10.1083/jcb.142.3.815

Amano M., et al. Rho-kinase/ROCK: A key regulator of the cytoskeleton and cell polarity. Cytoskeleton. 2010;67(9):545-554. doi:10.1002/cm.20472

Arrington M.E., et al. The molecular basis for immune dysregulation by the hyperactivated E62K mutant of the GTPase RAC2. Journal of Biological Chemistry. 2020;295(34):12130-12142. doi:10.1074/jbc.ra120.012915

Aspenström P., et al. Rho GTPases have diverse effects on the organization of the actin filament system. Biochemical Journal. 2004;377(2):327-337. doi:10.1042/BJ20031041

Azzarelli R., et al. Regulation of cerebral cortex development by Rho GTPases: Insights from in vivo studies. Frontiers in Cellular Neuroscience. 2015;8(JAN):1-18. doi:10.3389/fncel.2014.00445

Ba W., et al. TRIO loss of function is associated with mild intellectual disability and affects dendritic branching and synapse function. Human Molecular Genetics. 2016;25(5):892-902. doi:10.1093/hmg/ddv618

Barbosa S., et al. Opposite Modulation of RAC1 by Mutations in TRIO Is Associated with Distinct, Domain-Specific Neurodevelopmental Disorders. American Journal of Human Genetics. 2020;106(3). doi:10.1016/j.ajhg.2020.01.018

Bolis A., et al. Differential distribution of Rac1 and Rac3 GTPases in the developing mouse brain: Implications for a role of Rac3 in Purkinje cell differentiation. European Journal of Neuroscience. 2003. doi:10.1046/j.1460-9568.2003.02938.x

Burbelo P.D., et al. A Conserved Binding Motif Defines Numerous Candidate Target Proteins for Both Cdc42 and Rac GTPases. Journal of Biological Chemistry. 1995;270(49):29071-29074. doi:10.1074/jbc.270.49.29071

Causeret F., et al. The p21-activated kinase is required for neuronal migration in the cerebral cortex. Cerebral Cortex. 2009;19(4):861-875. doi:10.1093/cercor/bhn133

Corbetta S., et al. Essential role of Rac1 and Rac3 GTPases in neuronal development. The FASEB Journal. 2009;23(5):1347-1357. doi:10.1096/fj.08-121574

Corbetta S., et al. Generation and Characterization of Rac3 Knockout Mice. Molecular and Cellular Biology. 2005;25(13):5763-5776. doi:10.1128/mcb.25.13.5763-5776.2005

Coso O.A., et al. The small GTP-binding proteins Rac1 and Cdc42regulate the activity of the JNK/SAPK signaling pathway. Cell. 1995;81(7):1137-1146. doi:10.1016/S0092-8674(05)80018-2

Costain G., et al. De novo missense variants in RAC3 cause a novel neurodevelopmental syndrome. Genetics in Medicine. 2019;21(4):1021-1026. doi:10.1038/s41436-018-0323-y

de Curtis I. The Rac3 GTPase in Neuronal Development, Neurodevelopmental Disorders, and Cancer. Cells. 2019;8(9):1063. doi:10.3390/cells8091063

Dvorsky R., et al. Always look on the bright site of Rho: Structural implications for a conserved intermolecular interface. EMBO Reports. 2004;5(12):1130-1136. doi:10.1038/sj.embor.7400293

Guo D., et al. Rho GTPase Regulators and Effectors in Autism Spectrum Disorders: Animal Models and Insights for Therapeutics. Cells. 2020;9(4):835. doi:10.3390/cells9040835

Haataja L., et al. Characterization of RAC3, a novel member of the Rho family. Journal of Biological Chemistry. 1997;272(33):20384-20388. doi:10.1074/jbc.272.33.20384

Hamada N., et al. Role of a heterotrimeric G-protein, Gi2, in the corticogenesis: possible involvement in periventricular nodular heterotopia and intellectual disability. Journal of Neurochemistry. 2017;140(1):82-95. doi:10.1111/jnc.13878

Hamada N., et al. Neuropathophysiological significance of the c.1449T>C/p.(Tyr64Cys) mutation in the CDC42 gene responsible for Takenouchi-Kosaki syndrome. Biochemical and Biophysical Research Communications. 2020;529(4):1033-1037. doi:10.1016/j.bbrc.2020.06.104

Hamada N., et al. De novo PHACTR1 mutations in West syndrome and their pathophysiological effects. Brain. 2018;141(11):3098-3114. doi:10.1093/brain/awy246

Harms F.L., et al. Activating Mutations in PAK1, Encoding p21-Activated Kinase 1, Cause a Neurodevelopmental Disorder. American Journal of Human Genetics. 2018;103(4):579-591. doi:10.1016/j.ajhg.2018.09.005

Hill C.S., et al. The Rho family GTPases RhoA, Racl , and CDC42Hsregulate transcriptional activation by SRF. Cell. 1995;81(7):1159-1170. doi:10.1016/S0092-8674(05)80020-0

Hiraide T., et al. A de novo variant in RAC3 causes severe global developmental delay and a middle interhemispheric variant of holoprosencephaly. Journal of Human Genetics. 2019;64(11):1127-1132. doi:10.1038/s10038-019-0656-7

Horn S, Au M, Basel-Salmon L, et al. De novo variants in PAK1 lead to intellectual disability with macrocephaly and seizures. Brain. 2019;142(11):3351-3359. doi:10.1093/brain/awz264

Ito H., et al. Possible role of Rho/Rhotekin signaling in mammalian septin organization. Oncogene. 2005;24(47):7064-7072. doi:10.1038/sj.onc.1208862

Johnson D.I. Cdc42: An Essential Rho-Type GTPase Controlling Eukaryotic Cell Polarity. Microbiology and Molecular Biology Reviews. 1999;63(1):54-105. doi:10.1128/MMBR.63.1.54-105.1999

Kanie T., et al. Guanine Nucleotide Exchange Assay Using Fluorescent MANT-GDP. BIO-PROTOCOL. 2018;8(7). doi:10.21769/BioProtoc.2795

Kawai S., et al. Effect of three types of mixed anesthetic agents alternate to ketamine in mice. Experimental animals. 2011;60(5):481-487. doi:10.1538/expanim.60.481

Kazanietz M.G., et al. The Rac GTPase in Cancer: From Old Concepts to New Paradigms. Cancer Research. 2017;77(20):5445-5451. doi:10.1158/0008-5472.CAN-17-1456

Krengel U., et al. Three-dimensional structures of H-ras p21 mutants: Molecular basis for their inability to function as signal switch molecules. Cell. 1990;62(3):539-548. doi:10.1016/0092-8674(90)90018-A

Lek M., et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-291. doi:10.1038/nature19057

Lougaris V., et al. RAC2 and primary human immune deficiencies. Journal of Leukocyte Biology. 2020;108(2):687-696. doi:10.1002/JLB.5MR0520-194RR

Lougaris V., et al. A monoallelic activating mutation in RAC2 resulting in a combined immunodeficiency. Journal of Allergy and Clinical Immunology. 2019;143(4):1649-1653.e3. doi:10.1016/j.jaci.2019.01.001

Martinelli S., et al. Functional Dysregulation of CDC42 Causes Diverse Developmental Phenotypes. American Journal of Human Genetics. 2018;102(2):309-320. doi:10.1016/j.ajhg.2017.12.015

Malosio M.L., et al. Differential expression of distinct members of rho family GTP-binding proteins during neuronal development: Identification of RAC1B, a new neural- specific member of the family. Journal of Neuroscience. 1997;17(17):6717-6728. doi:10.1523/jneurosci.17-17-06717.1997

Miki H., et al. IRSp53 is an essential intermediate between Rac and WAVE in the regulation of membrane ruffling. Nature. 2000;408(6813):732-735. doi:10.1038/35047107

Moll J., et al. The murine rac1 gene: cDNA cloning, tissue distribution and regulated expression of rac1 mRNA by disassembly of actin microfilaments. Oncogene. 1991.

Mondal S., et al. A Homogenous Bioluminescent System for Measuring GTPase, GTPase Activating Protein, and Guanine Nucleotide Exchange Factor Activities. Assay and Drug Development Technologies. 2015;13(8):444-455. doi:10.1089/adt.2015.643

Nagata K-i. The MAP kinase kinase kinase MLK2 co-localizes with activated JNK along microtubules and associates with kinesin superfamily motor KIF3. The EMBO Journal. 1998;17(1):149-158. doi:10.1093/emboj/17.1.149

Nakagawa H., et al. IRSp53 is colocalised with WAVE2 at the tips of protruding lamellipodia and filopodia independently of Mena. Journal of Cell Science. 2003;116(12):2577-2583. doi:10.1242/jcs.00462

Nishikawa M., et al. Expression analyses of PLEKHG2, a Rho family-specific guanine nucleotide exchange factor, during mouse brain development. Medical molecular morphology. 2021;54(2):146-155. doi:10.1007/s00795-020-00275-1

Nishikawa M., et al. The interaction between PLEKHG2 and ABL1 suppresses cell growth via the NF-κB signaling pathway in HEK293 cells. Cellular signalling. 2019;61(April):93-107. doi:10.1016/j.cellsig.2019.04.016

Ohori S., et al. A novel PAK1 variant causative of neurodevelopmental disorder with postnatal macrocephaly. Journal of Human Genetics. 2020;65(5):481-485. doi:10.1038/s10038-020-0728-8

Pellegatta M., et al. Rac1 and Rac3 have opposite functions in Schwann cells during developmental myelination. Neuroscience letters. 2021;753:135868. doi:10.1016/j.neulet.2021.135868

Pengelly R.J., et al. Mutations specific to the Rac-GEF domain of TRIO cause intellectual disability and microcephaly. Journal of Medical Genetics. 2016;53(11):735-742. doi:10.1136/jmedgenet-2016-103942

Pennucci R., et al. Rac1 and Rac3 GTPases regulate the development of hilar mossy cells by affecting the migration of their precursors to the hilus. PLoS ONE. 2011;6(9):23-24. doi:10.1371/journal.pone.0024819

Perona R., et al. Activation of the nuclear factor-kappaB by Rho, CDC42, and Rac-1 proteins. Genes & Development. 1997;11(4):463-475. doi:10.1101/gad.11.4.463

Reid T., et al. Rhotekin, a new putative target for Rho bearing homology to a serine/threonine kinase, PKN, and rhophilin in the Rho-binding domain. Journal of Biological Chemistry. 1996;271(23):13556-13560. doi:10.1074/jbc.271.23.13556

Reijnders M.R.F., et al. RAC1 Missense Mutations in Developmental Disorders with Diverse Phenotypes. American Journal of Human Genetics. 2017;101(3):466-477. doi:10.1016/j.ajhg.2017.08.007

Richards S., et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine. 2015;17(5):405-423. doi:10.1038/gim.2015.30

Roberts P.J., et al. Rho family GTPase modification and dependence on CAAX motif-signaled posttranslational modification. Journal of Biological Chemistry. 2008;283(37):25150-25163. doi:10.1074/jbc.M800882200

Shirsat N.V., et al. A member of the ras gene superfamily is expressed specifically in T, B and myeloid hemopoietic cells. Oncogene. 1990.

Sobreira N., et al. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. Human Mutation. 2015;36(10):928-930. doi:10.1002/humu.22844

Soriano-Castell D, et al. ROCK1 is a novel Rac1 effector to regulate tubular endocytic membrane formation during clathrin-independent endocytosis. Scientific Reports. 2017;7(1):1-17. doi:10.1038/s41598-017-07130-x

Stankiewicz T.R., et al. Rho family GTPases: Key players in neuronal development, neuronal survival, and neurodegeneration. Frontiers in Cellular Neuroscience. 2014;8(OCT):1-14. doi:10.3389/fncel.2014.00314

Tabata H., et al. Decoding the molecular mechanisms of neuronal migration using in utero electroporation. Medical Molecular Morphology. 2016;49(2):63-75. doi:10.1007/s00795-015-0127-y

Tabata H., et al. c. Neuroscience. 2001;103(4):865-872. doi:10.1016/s0306-4522(01)00016-1

Tabata H., et al. Multipolar Migration: The Third Mode of Radial Neuronal Migration in the Developing Cerebral Cortex. Journal of Neuroscience. 2003;23(31):9996-10001. doi:10.1523/jneurosci.23-31-09996.2003

Takenouchi T., et al. Macrothrombocytopenia and developmental delay with a de novo CDC42 mutation: Yet another locus for thrombocytopenia and developmental delay. American Journal of Medical Genetics Part A. 2015;167(11):2822-2825. doi:10.1002/ajmg.a.37275

Takenouchi T., et al. Further evidence of a mutation in CDC42 as a cause of a recognizable syndromic form of thrombocytopenia. American Journal of Medical Genetics Part A. 2016;170(4):852-855. doi:10.1002/ajmg.a.37526

Tomasevic N., et al. Differential Regulation of WASP and N-WASP by Cdc42, Rac1, Nck, and PI(4,5)P2. Biochemistry. 2007;46(11):3494-3502. doi:10.1021/bi062152y

Vaghi V., et al. Rac1 and Rac3 GTPases control synergistically the development of cortical and hippocampal GABAergic interneurons. Cerebral Cortex. 2014;24(5):1247-1258. doi:10.1093/cercor/bhs402

Welch B.L. The significnace of the difference between two means when the population variances are unequal. Biometrika. 1938;29(3-4):350-362. doi:10.1093/biomet/29.3-4.350

White J.J., et al. WNT Signaling Perturbations Underlie the Genetic Heterogeneity of Robinow Syndrome. The American Journal of Human Genetics. 2018;102(1):27-43. doi:10.1016/j.ajhg.2017.10.002

Zamboni V., et al. Rho GTPases in Intellectual Disability: From Genetics to Therapeutic Opportunities. International Journal of Molecular Sciences. 2018;19(6):1821. doi:10.3390/ijms19061821

Zhang K., et al. The p21-activated kinases in neural cytoskeletal remodeling and related neurological disorders. Protein & Cell. December 2020. doi:10.1007/s13238-020-00812-9

Zhang K., et al. The p21-activated kinases in neural cytoskeletal remodeling and related neurological disorders. Protein & Cell. December 2020. doi:10.1007/s13238-020-00812-9

B. Gene discovery:

Early-infantile onset epilepsy and developmental delay caused by bi-allelic *GAD1* variants

Abstract

Gamma-aminobutyric acid (GABA) and glutamate are the most abundant amino acid neurotransmitters in the brain. GABA, an inhibitory neurotransmitter, is synthesized by glutamic acid decarboxylase (GAD). Its predominant isoform GAD67, contributes up to 90% of base-level GABA in the CNS, and is encoded by the GAD1 gene. Disruption of GAD1 results in an imbalance of inhibitory and excitatory neurotransmitters, and as Gad1–/– mice die neonatally of severe cleft palate, it has not been possible to determine any potential neurological dysfunction. Furthermore, little is known about the consequence of GAD1 disruption in humans. Here we present six affected individuals from six unrelated families, carrying bi-allelic GAD1 variants, presenting with developmental and epileptic encephalopathy, characterized by early-infantile onset epilepsy and hypotonia with additional variable non-CNS manifestations such as skeletal abnormalities, dysmorphic features and cleft palate. Our findings highlight an important role for GAD1 in seizure induction, neuronal and extra-neuronal development, and introduce GAD1 as a new gene associated with developmental and epileptic encephalopathy.

1. Introduction

The neurotransmitter c-aminobutyric acid (GABA) is one of the main inhibitory neurotransmitters deriving from glutamate (*Cooper et al., 1996*). It plays a critical signaling role

in the nervous system and also in a number of non-neuronal cell types. The enzyme responsible for the conversion of glutamate into GABA is glutamate decarboxylase (GAD), and occurs in two isoforms GAD65 and GAD67, depending on its molecular weight (*Kaufman et al., 1991*). These isoforms are products of two different genes, GAD1 (encoding a 67 kDa molecular weight protein, GAD67) and GAD2 (encoding a 65 kDa molecular weight protein, GAD65). GAD67 is constitutively active and produces 490% of the base level GABA in the CNS, whilst GAD65 is transiently activated (*Asada et al., 1997*).

Animal studies have also shown a distinct role for GAD65 and GAD67. Gd65–/– mice are viable but show a higher susceptibility to seizure induction despite normal GABA levels (*Asada et al., 1996*), whereas Gad67–/– mice are characterized by neonatal death, severe cleft plate and respiratory failure. GAD activity and GABA concentration are also drastically reduced in Gad67–/– mice (*Asada et al., 1997*). The severity of the Gad67–/– phenotype in animal models might suggest severe phenotypical manifestations in humans, yet there have been few reported families with GAD1 mutations (*Lynex et al., 2004; Saito et al., 2010; Curley et al., 2011; Ruzicka et al., 2015; Magri et al., 2018*). Previous reports have described seemingly unparalleled phenotypes, which include schizophrenia, autism spectrum disease and cerebral palsy, and functional studies are missing to confirm the pathogenicity of these reported mutations.

Here we report a series of six affected individuals with dis-tinct phenotypical features from six unrelated families with bi-allelic mutations in the *GAD1* gene (three carrying homozygous missense mutations, one carrying a homozygous frameshift variant, two compound heterozygous variants, and one harboring a homozygous stop gain variant). All affected individuals presented with seizures, strongly impaired neurocognitive development, and reduced

muscle tone of variable severity. Interestingly only one presented with cleft palate, which has been suggested to be one of the key features in the GAD1 animal models.

2. Material and methods

2.1 Patients phenotyping

Six patients from six unrelated families of Persian (Family A), Pakistani (Family B), African American (Family C), Sudanese (Family D), Egyptian (Family E) and Turkish an- cestry (Family F) were identified through GeneMatcher (*Sobreira et al., 2015*) and enrolled in this study. The study was conducted according to the Declaration of Helsinki and with the approval of the institutional review boards of University College of London and participating centres.

2.2 Genetic analysis

Genetic testing through whole exome sequencing (WES) was carried out in different research centres after informed con- sent was obtained from the parents or legal guardians of the studied subjects. Genomic DNA was extracted from peripheral blood obtained from the probands, parents, and un- affected siblings (when available). Exome sequencing and data analysis was performed as follows: Families A, B and D in the according centres as previously described (Monies et al., 2019; Dias et al., 2019), Family C through GeneDx (*Retterer et al., 2016*), and Families E and F at Centogene (*Bauer et al., 2018*). Potential candidate causal variants were subsequently confirmed by independent bi-directional Sanger sequencing.

3. Results

3.1 Clinical manifestations

Four of the affected individuals were born from consanguineous parents (first or second cousins), and all were born at term following a normal pregnancy. A key clinical feature common to all affected individuals was early onset seizures (from 2 to 6 months), predominantly focal motor seizures with and without impaired awareness (two with additional epileptic spasms, four with focal non-motor seizures, and five with bilateral motor seizures). Seizures were pharmaco-logically controlled in three of six affected individuals, with three reported as drug-resistant. Drug regimens differed across the individuals. EEG at seizure onset showed a burst suppression pattern in two individuals, diffuse slowing with multifocal as well as generalized sharp waves in two, and hypsarrhythmia in two.

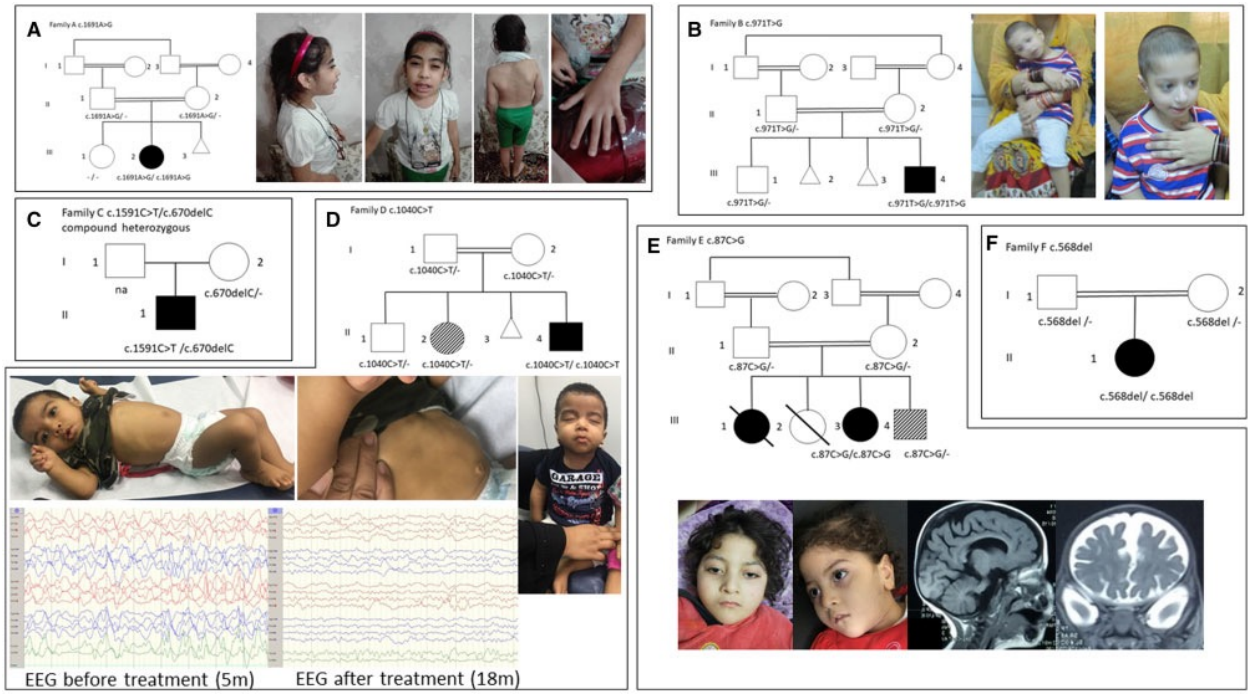


Figure 1 Pedigrees of the reported families and clinical pictures of GAD1 patients. (A) The female patient from Family A (Patient II- 2) carries the homozygous c.1691A4G p.(Asn564Ser) variant and shows dysmorphic features with thick eyebrows, protruding ears, scoliosis, and long fingers with clinodactyly. (B) Patient from Family B (Patient III-4) harbors the c.971T4G p.(Phe324Cys) variant. He has slight dys- morphic features (wide mouth, thin upper lips, bitemporal narrowing and retrognathia). (C) Pedigree showing the segregation of the compound

heterozygous variants c.1591C4T, c.1591C4T p.(Arg531*) and c.670delC p.(Leu224Serfs*5) in Family C. (D) Pedigree of Family D shows the segregation of the c.1040C4T p.(Thr347Met) variant. Patient II-4 carries the variant in homozygous state. He is severely hypotonic and shows severe dysmorphic features (infra-orbital creases, severely depressed nasal bridge, anteverted nares, prominent nasolabial folds). In addition, significant diastasis recti can be observed. His sister (Patient II-2) is heterozygous for the same variant and suffers from a different neurodevelop- mental condition without seizures. (E) Patient III-2 from Family E harbors the c.87C4G (Tyr29*) variant, severely affected with dysmorphic facial features and global atrophy on cardiovascular MRI, one similarly affected sibling passed away without any genetic testing being performed, another sibling passed away only a few hours after birth, no phenotypical or genetic assessment could be carried out, and one sibling is alive with a different phenotype (sensorineural hearing loss, Hirschsprung disease). (F) Patient II-1 from Family F harbors the c.568del (Gln190Serfs*11) variant. His parents are both heterozygous carriers of the same variant. Empty and full symbols represent healthy and affected individuals, re- spectively. The symbol with diagonal lines indicates carrier status/different phenotype. The double line indicates consanguinity.

Follow-up EEGs showed diffuse slowing of background activity or persistent epileptic activity (two of six). Cranial MRI was normal in all but two individuals, one showing slight ventricular enlargement and one moderate global atrophy. The second common clinical feature was severe develop- mental delay. Most patients did not achieve any speech or non-verbal communication, only one was reported to have developed simple speech and basic perceptive language skills. This individual has remained seizure-free under medication. Of the more severely affected individuals, despite remaining seizure-free following pharmacological intervention, they still failed to accelerate in their intellectual development.

The third key feature we observed was a reduced muscle strength (five of six individuals) of varying severity ranging from slight muscle tone (one of five) to limited head control, inability to sit or crawl (four of six) and nasogastric tube dependence, due to dysphagia (two of six). While slight dys- morphic facial features were seen in four of six individuals (Table 1), only one presented with a cleft palate.

Other features that were observed without clear common elements included hirsutism, kidney stones, urogenital mal- formations, diastasis recti, reduced head circumference, clinodactyly, short arms, arthrogryposis of the lower limbs and congenital hip dislocation.

Metabolic workups did not show any abnormalities. A detailed clinical summary of all affected individuals can be found in Table 1 along with images of key features and family pedigrees in Fig. 1.

Table 1. Clinical features of *GAD1* patients

Family	A (Patient III-2)	B (Patient III-4)	C (Patient II-1)	D (Patient II-4)	E (Patient III-3)	F (Patient II-2)
Sex/ ethnic origin	Female/ Persian	Male/ Pakistani	Male/ African American	Male/ Sudanese	Female/ Egyptian	Female/ Turkish
Consanguinity	Yes (double first cousins)	Yes	–	Yes (second cousins)	Yes (first cousins)	Yes (first cousins)
Age at first/last exam	5 y/10 y 3 m	6 m/7 y	2 m/22 m	2 m/18 m	6 m/4 y	1 m/3 y
Development						
Milestones	Delayed in all milestones, simple speech at 4 y, walking delayed, no complex movements	Delayed in all milestones, sitting and crawling	Delayed in all milestones, no head control, no sitting, no speech	Severe delay, poor head control achieved at 18m, no sitting	Severe delay in all milestones, bed ridden	Severe delay in all milestones, no sitting or crawling
ID	Moderate	Severe	Severe	Severe	Severe	Severe
Vision/hearing	High myopia	Normal	NA	Normal	Moderate hearing impairment	High myopia
Dysmorphic facial features	Yes	Yes	–	Yes	Yes	–
Cleft palate	–	–	–	–	Yes (surgical correction)	–
Skeletal abnormalities	Clindodactyly, pes planus, scoliosis	Arthrogryposis of lower limbs	–	Short arms	Congenital hip dislocation and malformation	
Neurological examination	Mild hypotonia	Brisk DTR, stereotypic hand movements, oral automatisms	Mild hypotonia, spasticity in lower extremities, oropharyngeal dysphagia	Severe hypotonia, dysphagia (floppy epiglottis)	Severe hypotonia, hyporeflexia, dysphagia	Severe hypotonia
Epilepsy						
Age at onset	2 m	6 m	2 m	2 m	6 m	2 m
Type(s) of seizure at onset	Focal non/motor with impaired awareness	Focal motor impaired awareness, bilateral tonic-clonic	Focal non/motor with impaired awareness, bilateral tonic clonic	Epileptic spasms	Bilateral tonic clonic	Focal ± to bilateral motor with impaired awareness
Seizure progression (age)	Controlled (10 y), last seizure at age 7 y	Controlled (7 y), last seizure at 5.5 y	Refractory (28 m)	Spasms continue, seizures controlled (18 m)	Refractory (4 y)	Partial control, 1 seizure/week
EEG at onset	Burst suppression	Multifocal and generalized epileptogenic activity	Burst suppression	Hypsarrhythmia	Generalized epileptogenic activity	Hypsarrhythmia
Follow-up EEG (age)	Normal (11 y)	Normal (7 y)	Slowing, multifocal epileptic discharges (28 m)	No epileptic abnormalities (18 m)	NA	Generalized epileptiform activity (4 m)
AEDs trialled	PB, *VPA*, *CBZ*, *CLB*	PB, VPA, *CLZ*	*CBD*, *vigabatrin*, *ketogenic diet*, *CLZ*, PB	steroids, TP, *vigabatrin*	*LEV*, *CLZ*, TP	*PB*, *TP*, *LEV*, CLZ, vigabatrin, steroids
Cardiovascular MRI (age)	Normal (5 y)	Prominent ventricular space (6 m)	Normal (2 m)	Normal (6 m)	Moderate global atrophy (1 y)	Normal (2 m)
Other features	Hydronephrosis, nephrocalcinosis, bilateral kidney stones	–	NG-tube dependent	Diastasis recti	Intermittent NG tube dependence	–

Current antiepileptic drugs (AEDs) are highlighted in italics. CBD = cannabidiol; CBZ = carbamazepine; CLB = clobazam; CLZ = clonazepam; DTR = deep tendon reflexes; ID = intellectual disability; LEV = levetiracetam; NA = not applic- able; NG = nasogastric; OXC = oxcarbazepine; PB = phenobarbital; TP = topiramate; VPA = valproic acid.

3.2 Genetic findings

Variants were prioritized in each family based on allele frequency 50.01%, predicted impact on protein function, and biological consistency. Potentially causal bi-allelic variants in GAD1 were identified in all affected individuals. The segregation of the variants with the clinical phenotype was con- firmed by Sanger sequencing, which showed a recessive mode of inheritance. Detailed

genetic results are provided in Table 2. All affected individuals carried ultrarare GAD1 variants, which were predicted to result in impaired protein function. Homozygous variants were identified in five families (Families A, B, D, E and F), whereas compound hetero- zygous variants were found in Family C (Table 2).

Table 2. Frequency and predicted effect of the reported *GAD1* variants

GAD1 variant [NM_000817.2]	c.87C>G (p.Tyr29Ter) (Family E)	c.568delC (p.Gln190Serfs Ter11) (Family F)	c.670delC p.(Leu224Serfs*5) (Family C)	c.971T>G p.(Phe324Cys) (Family B)	c.1040C>T p.(Thr347Met) (Family D)	c.1591C>T p.(Arg531*) (Family C)	c.1691A>G p.(Asn564Ser) (Family A)
g. (hg19)	g.171678601C>G	g.171693323delC	g.171700586delC	g.171702542T>G	g.171704223C>T	g.171715383C>T	g.171716298A>G
Internal database	–	–	–	–	–	–	–
ExAC/GnomAD	–	–	–	0.00000796 (2 het)	0.00000795 (2 het)	0.00000398 (1 het)	0.00000398 (1 het)
GME	–	–	–	–	–	–	–
Iranome	–	–	–	–	–	–	–
Ensembl	–	–	–	–	–	–	–
SIFT	D- N/A	N/A	N/A	D (0.9125)	D (0.9125)	N/A	D (0.9125)
MutationTaster	DC (1)	DC (1)	DC (1)	DC (0.9768)	DC (1)	DC (1)	DC (1)
PolyPhen-2	N/A	N/A	N/A	PD (0.978)	PD (1)	N/A	PD (1)
GERP score	4.97	5.55	6.17	5.91	5.67	4.67	5.48
CADD score	35	N/A	N/A	28	29.1	43	26.1
ACMG class	5 (PVS1, PM2, PP4)	5 (PVS1, PM2, PP3)	5 (PVS1, PM2, PP3)	3 (PM2, PP3)	3 (PM2, PP3)	5 (PVS1, PM2, PP3)	3 (PM2, PP3)
GeneDx	0	0	0/135 084	0	1/130 874	0/135 084	1/130 874

CADD = Combined Annotation Dependent Depletion; D = damaging; DC = disease causing; GeneDx = variant frequencies from the GeneDx database; GERP = Genomic Evolutionary Rate Profiling; GnomAD = Genome Aggregation Database; GME = Greater Middle East (GME) Variome Project; het = heterozygous; N/A = not applicable; PD = probably damaging; PM2 = Pathogenic Moderate 2; PP3 = Pathogenic Supporting 3; PVS1 = Pathogenic Very Strong 1; SIFT = Sorting Intolerant From Tolerant.

The affected proband from Family A carried a homozygous c.1691A4G, p.(Asn564Ser) variant, which is reported only once in the heterozygous state in gnomAD. This variant has a combined annotation dependent depletion (CADD) score of 26.1 and is predicted to be pathogenic by several bioinformatic prediction tools, including SIFT (score 0.9125), MutationTaster, and PolyPhen-2 (score 1) (Table 2). The proband from Family B harbored a homozygous c.971T4G, p.(Phe324Cys) variant, which causes the substitution of a phenylalanine residue at position 324 with cysteine. This position is not strictly conserved in other species, where leucine is found in place of phenylalanine. However, both phenylalanine and leucine belong to the class of the amino acids with long hydrophobic chains (Fig. 2A) and share several chemical features. This variant has been seen in the heterozygous state in gnomAD with a minor

allele frequency of 0.00000796. It has a CADD score of 28 and is predicted damaging by all the prediction tools used (scores of 0.9125 and 0.978 for SIFT and PolyPhen-2, respectively).

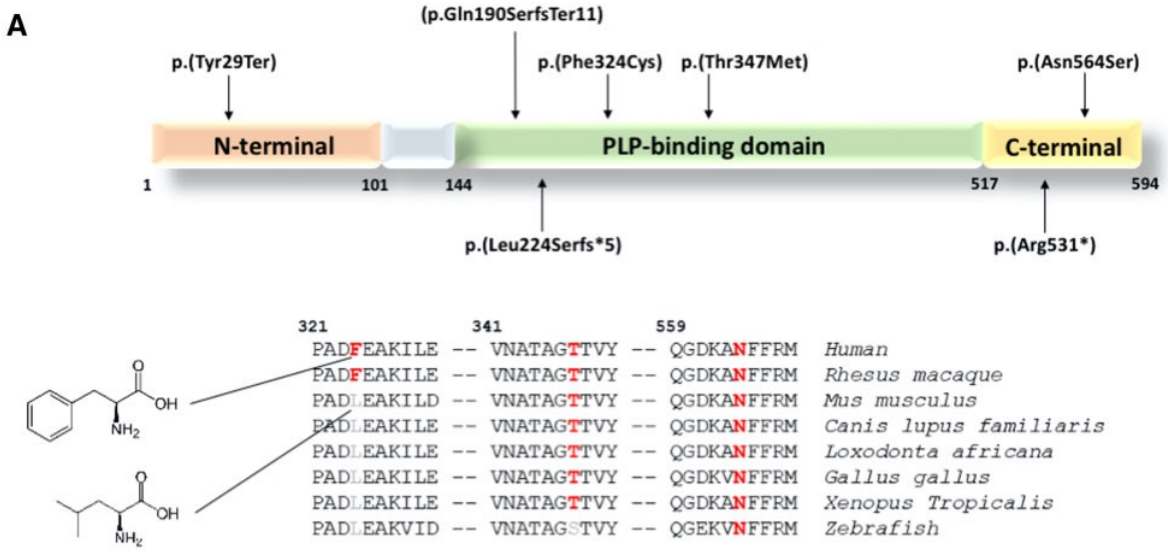


Figure 2. Schematic and cartoon representation of GAD1. (A) Schematic representation of the GAD1 isoform GAD67 (NP_000808.2) with the pathogenic variants identified in this study. Of the six variants, four fall within the PLP-binding domain, a conserved region that is essential for the binding of the crucial cofactor pyridoxal 50-phosphate (PLP). The remaining variants affect the C-terminal domain, which contains the catalytic site of the enzyme. Conservation status among different species is shown for the missense variants.

The proband from Family D carried a homozygous c.1040C4T, p.(Thr347Met) variant, which was reported twice in the heterozygous state in the gnomAD database. It was predicted damaging by both SIFT and PolyPhen-2 with high scores (0.9125 and 1, respectively). The CADD score for this variant was 29.1. Further in silico analysis predicted a reduction in protein stability for p.(Phe324Cys) and p.(Asn564Ser), in association with a break in H-bonds in the pyridoxal 50-phosphate (PLP) binding domain for p.(Thr347Met). Both these changes might result in an impairment of the protein function due to the abnormal degradation or the decreased binding activity to- wards PLP, leading to a likely loss-of-function effect (Fig. 2B and Supplementary Table 1).

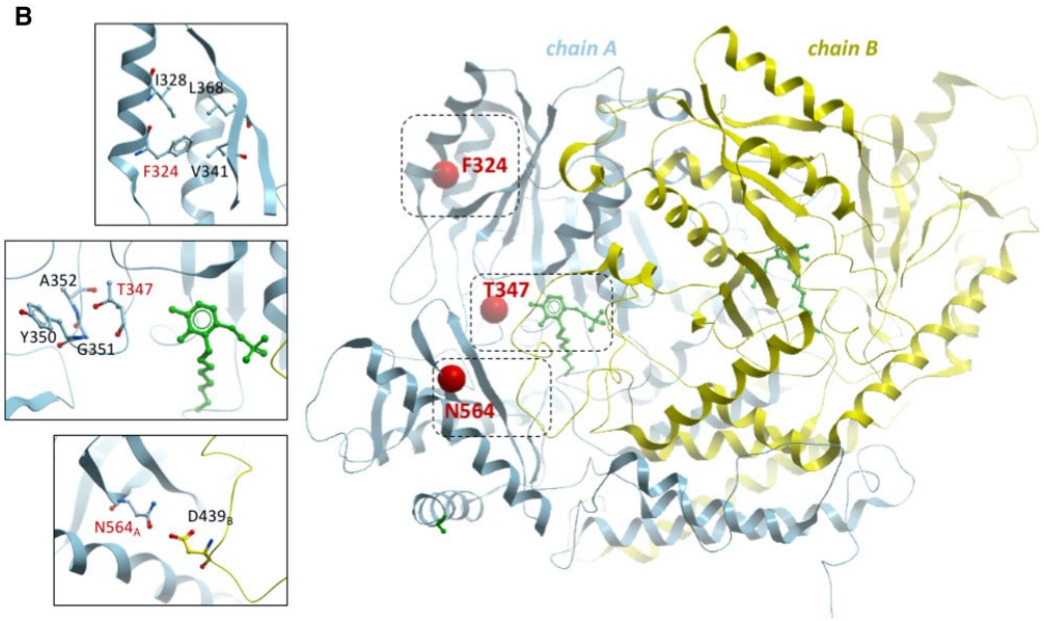


Figure 2. Schematic and cartoon representation of GAD1. (B) Cartoon representation of human GAD1 dimer (PDB: 2okj) with the two subunits in blue and yellow. Sites of the three missense mutations in this study are shown as red spheres, and close-up views of their nearby atomic environment are shown as insets. The PLP cofactor is shown in green.

In Family E, a homozygous stop gain variant c.87C4G, p.(Tyr29Ter) was identified. This variant is absent in all the queried population datasets. It is predicted damaging by SIFT (score not available) and disease-causing by Mutation Taster, with a CADD score of 35. This null variant likely causes a nonsense-mediated mRNA decay (NMD), leading to a complete loss-of-function. The proband from Family F carries a homozygous c.568delC p.(Gln190SerfsTer11) variant, which is absent in gnomAD and other population datasets queried. It is predicted damaging by MutationTaster (score 1) and classified as pathogenic according to the ACMG guidelines. Family C was the only family in which heterozygous variants were found: c.1591C4T, p.(Arg531*) and c.670delC, p.(Leu224Serfs*5). This individual's mother was a carrier for c.670delC; however, the individual's father was not available for genetic testing. The c.1591C4T, p.(Arg531*) variant results in a stop gain with a very high likely impact on protein function, as highlighted by a CADD score of 43. This is only reported in heterozygous state in

the gnomAD database. Similarly, the frameshift variant c.670delC, p.(Leu224Serfs*5) is predicted to be disease-causing by MutationTaster and affects a very conserved residue, with a GERP score of 6.17. The null variants identified in Families B, E, and F may cause a premature termination of the tran- script leading to a truncated protein or, alternatively, affected transcripts might be target of NMD.

4. Discussion

4.1 Genotype spectrum

In humans, GAD1 encodes a 67-kDa protein, GAD67, which is the major contributor to GABA production in the CNS. As the major embryonic GAD isoform, it also plays a pivotal role in synaptogenesis and neuronal development (*Asada et al., 1997; Soghomonian and Martin, 1998; Sgado` et al., 2011*). The 594 amino acid protein GAD67 is composed of a N-terminal domain involved in the generation of GAD65– GAD67 heterodimers and subcellular targeting, a C-terminal domain containing the catalytic site, and a central conserved domain binding PLP (*Bosma et al., 1999; Martin et al., 2000; Lynex et al., 2004*). This pyridoxal-dependent decarboxylase domain is essential for GAD67 function as GAD requires PLP as a cofactor to catalyze the generation of GABA from glutamate (*Lernmark, 1996*).

Four of the seven variants identified in our families affect conserved residues in the PLP-binding domain, likely leading to loss-of-function (Fig. 1A). In particular, the two missense variants c.971T4G, p.(Phe324Cys) and c.1040C4T, p.(Thr347Met) probably cause impaired PLP binding, whereas the null variants c.568delC, p.(Gln190SerfsTer11) and c.670delC p.(Leu224Serfs*5) might result in a truncated protein or NMD. Similarly, the c.87C4G, p.(Tyr29Ter) variant localized to the N-terminal domain results in a pre- mature stop codon, likely leading to NMD. The remaining two variants, c.1591C4T, p.(Arg531*) and c.1691A4G,

p.(Asn564Ser), are localized to the C-terminal domain of the protein and are predicted to result in a complete and partial loss-of-function of the catalytic activity, respectively (Fig. 2B). According to gnomAD, GAD1 shows a moderate intolerance to missense variants (Z-score = 2.32; observed 217 and expected 336.5) and predicted loss-of-function variants (expected 33.1, observed 9). The likely negative impact of the missense variants on PLP-binding and C-terminal catalytic domains, together with the finding of four null pathogenic variants in our case series, supports the idea that the clinical phenotypes observed are likely due to a loss-of- function mechanism.

4.2 Genotype-phenotype correlations

With regard to the possible genotype-phenotype correlation, there have been few clinical case reports of GAD1 mutations, with little overlap in clinical features (*Lynex et al., 2004; Ruzicka et al., 2015; Magri et al., 2018*). The six patients reported here, however, do show a clear phenotypical manifestation. One of the main common features we observed was the occurrence of seizures at a young age (between 2 and 6 months of age). Since GAD1 is centrally involved in the pro- duction of GABA, GAD1 mutations likely lead to an imbalance of GABA in the brain. Previous studies have shown that abnormalities in GABAergic function play an important role in seizure induction (*Olsen and Avoli, 1997*). Dysfunction of the GABAergic system can be caused through either abnormal GABA synthesis (e.g., GAD dysfunction) or abnormal signaling (e.g. GABA receptor malfunction). Animal studies have shown that mutant mice lacking GAD or certain subunits of GABA-A receptors are prone to spontaneous epileptic seizures (*Asada et al., 1996; Kash et al., 1997; DeLorey et al., 1998*). It has also been shown that there is a reduction of GABAergic neurons in the epileptic brain, independent of the seizure's aetiology supporting a

conclusion that seizures themselves further de- crease GABA release within the brain, causing further imbalance (*Wang et al., 2011*). This evidence suggests that there is a clear correlation between abnormalities within the GABAergic system and seizure occurrence. The occurrence of seizures early in life of our patients is therefore not a surprising phenotypical finding. However, previous data suggest that the intellectual development of newborns with epileptic encephalopathies is strongly dependent on seizure control, and seizure freedom usually leads to acceleration of intellectual development (*Bombardieri et al., 2010*).

4.3 Epileptic manifestations

Despite seizure control (seizure freedom achieved in three of six individuals), all patients described here still showed severe intellectual disability. This has been observed previously in other genetic conditions (*Weckhuysen et al., 2012; Berecki et al., 2019*), leading to the assumption that genetic defects themselves are influencing development and cognition, independently from seizure control. This is reflected in the ILAE's terminology of 'developmental and epileptic encephalopathies', though a clear distinction between epileptic and developmental encephalopathy has been advised (*Scheffer et al., 2017; Scheffer and Liao, 2020*). Furthermore, some EEG abnormalities continued to be prominent after achieving seizure freedom. We therefore hypothesize that *GAD1* mutations may also play an important role in intellectual development.

4.4 Extended neurological phenotype

The reduced muscle tone and weakness, causing severe disability in four of six affected individuals was a surprising finding. There is a single case described in the literature of an

individual with cerebral palsy and a GAD1 variant (*Lynex et al., 2004*). However, identification of the variant was based on autozygosity mapping, in which they identified a recessive locus of 5 cM located at 2q24-31.1. The investigators subsequently investigated the most interesting candidate in that region by sequencing the exons of GAD1 and identified a homozygous missense variant c.35C4G (p.Ser12Cys) (NM_013445.3) in the N-terminal domain of the protein (*Lynex et al., 2004*). This variant is rare in gnomAD (exomes allele frequency of 0.0000239) and absent in the homozygous state. However, the predictions on its pathogenicity are conflicting as it is predicted benign by MutationAssessor, DEOGEN2, and MetaLR. Furthermore, the 2q24-31.1 locus encompasses several other possible genes of interest (e.g., DYNC1I2, responsible for a neurodevelopmental disorder characterized by intellectual disability, spasticity, and neuroradiological anomalies), which were not investigated. The N-terminal domain of GAD67 is involved in the generation of GAD65–GAD67 heterodimers and sub- cellular targeting, and whilst we cannot rule out a possible impairment of the interaction of GAD67 with GAD65 or an alteration in GAD67 subcellular targeting as a result of this variant, we suggest that variants affecting the PLP-binding and C-terminal domain cause a more severe deficiency in GAD67 activity.

In addition, we also identified four null variants likely resulting in a loss-of-function. According to these observations and the consistency of the phenotype in our case series, we emphasize that GAD1 pathogenic variants should be considered the cause of a distinctive neurodevelopmental dis- order instead of spastic cerebral palsy (*Lynex et al., 2004*). The clinical observation of muscle weakness is particularly interesting as it has not been described previously. Other GAD1-related diseases, such as antibody-mediated syndromes, have been associated with motor symptoms (*Dayalu and Teener, 2012*). However, these motor phenomena

are usually linked to a hyperexcitability (increased muscle tone leading to rigidity, muscle spasms, and stiff person syndrome), as well as other neurological symptoms (e.g. ,Miller Fisher syndrome, eye movement disorders, cerebellar ataxia, epilepsy, limbic encephalitis, etc.) (*Moersch and Woltman, 1956; Dalakas et al., 2000; Saiz et al., 2008; Tohid, 2016*). Of note, in no case has muscle weakness been linked to GAD1 deficiency.

4.5 Extra-neurological features

While Gad67–/– mice have been reported to die within the first hours of life due to a cleft palate (*Asada et al., 1997; Condie et al., 1997*), only one of our patients (Family E, Patient III-3) was born with a cleft palate (the same patient also showed congenital bilateral hip dislocation with shallow acetabulum, talipes equinovarus and hearing impairment.) Several studies have hypothesized that the development of a cleft palate is linked to reduced tongue movement during embryonic development and therefore secondary to CNS dysfunction (*Iseki et al., 2007; Oh et al., 2010; Saito et al., 2010*). Gad1 expression has also been shown in different non-neural tissues, such as the tail bud, limb mesenchyme, vibrissal placodes, and pharyngeal arches in various stages of embryonic development (*Maddox and Condie, 2001*). This observation has suggested a broader influence of GAD1 and GABA function on non-neural development (*Maddox and Condie, 2001*), supporting a possible primary role of GAD1 impaired function in the pathogenesis of non- neural defects. However, further studies are required to con- firm this in humans.

5. Conclusion

In conclusion, this case series reports distinct phenotypical features caused by GAD1 variants, including early-infantile onset epilepsy, severe developmental delay and muscle weak- ness.

Less consistent features include skeletal abnormalities and dysmorphic facial features, including cleft palate. Functional studies and larger clinical series will be necessary to further assess genotype-phenotype correlations for GAD1 variants.

References

Asada H., et al. Mice lacking the 65 kDa isoform of glutamic acid decarboxylase (GAD65) maintain normal levels of GAD67 and GABA in their brains but are susceptible to seizures. Biochem Biophys Res Commun 1996; 229: 891–895.

Asada H. et al. Cleft palate and decreased brain gamma-aminobutyric acid in mice lacking the 67-kDa isoform of glutamic acid decarboxylase. Proc Natl Acad Sci USA 1997; 94: 6496–9.

Bauer P., et al. Development of an evidence-based algorithm that optimizes sensitivity and specificity in ES-based diagnostics ìof a clin- ically heterogeneous patient population. Genet Med 2018; 21: 53–61.

Berecki G., et al. SCN1A gain of function in early infantile encephalopathy. Ann Neurol 2019; 85: 514–25.

Bombardieri R., et al. Early control of seizures improves long-term outcome in children with tuberous sclerosis complex. Eur J Paediatr Neurol 2010; 14: 146–9.

Bosma P.T., et al. Multiplicity of glutamic acid decarboxylases (GAD) in verte- brates: molecular phylogeny and evidence for a new GAD paralog. Mol Biol Evol 1999; 16: 397–404.

Condie B.G., et al. Cleft palate in mice with a targeted mutation in the gamma-aminobutyric acid-producing enzyme glutamic acid decarboxylase 67. Proc Natl Acad Sci USA 1997; 94: 11451–5.

Cooper J.R., et al. The biochemical basis of neuropharmacology. New York: New York: Oxford University Press; 1996.

Curley A.A., et al. Cortical deficits of glutamic acid decarboxylase 67 expres- sion in schizophrenia: clinical, protein, and cell type-specific features. Am J Psychiatry 2011; 168: 921–9.

Dalakas M.C., et al. The clinical spectrum of anti- GAD antibody-positive patients with stiff-person syndrome. Neurology 2000; 55: 1531–5.

Dayalu P., et al. Stiff Person syndrome and other anti-GAD-associated neurologic disorders. Semin Neurol 2012; 32: 544–9.

DeLorey T.M., et al. Mice lacking the beta3 subunit of the GABAA receptor have the epilepsy phenotype and many of the behavioral characteristics of Angelman syndrome. J Neurosci 1998; 18: 8505–14.

Dias C.M., et al. Homozygous missense variants in NTNG2, encoding a presynap- tic netrin-G2 adhesion protein, lead to a distinct neurodevelopmen- tal disorder. Am J Hum Genet 2019; 105: 1048–56.

Iseki S., et al. Experimental induction of palate shelf elevation in glutamate de- carboxylase 67-deficient mice with cleft palate due to vertically oriented palatal shelf. Birth Defect Res A Clin Mol Teratol 2007; 79: 688–95.

Kash S.F., et al. Epilepsy in mice deficient in the 65-kDa isoform of glutamic acid decarboxylase. Proc Natl Acad Sci USA 1997; 94: 14060–5.

Kaufman D.L., et al. Two forms of the gamma-amino- butyric acid synthetic enzyme glutamate decarboxylase have distinct intraneuronal distributions and cofactor interactions. J Neurochem 1991; 56: 720–3.

Lernmark A. Glutamic acid decarboxylase–gene to antigen to disease. J Intern Med 1996; 240: 259–77.

Lynex C.N., et al. Homozygosity for a missense mutation in the 67 kDa isoform of glutamate decarboxylase in a family with autosomal recessive spastic cerebral palsy: parallels with Stiff-Person Syndrome and other movement disorders. BMC Neurol 2004; 4: 20.

Maddox D.M., et. Dynamic expression of a glutamate decarb- oxylase gene in multiple non-neural tissues during mouse develop- ment. BMC Dev Biol 2001; 1: 1.

Magri C., et al. A novel homozygous mutation in GAD1 gene described in a schizophrenic patient impairs activity and dimerization of GAD67 enzyme. Sci Rep 2018; 8: 15470.

Martin D.L., et al. Structural features and regula- tory properties of the brain glutamate decarboxylases. Neurochem Int 2000; 37: 111–9.

Moersch F.P., et al. Progressive fluctuating muscular rigidity and spasm (“stiff-man” syndrome); report of a case and some obser- vations in 13 other cases. Proc Staff Meet Mayo Clin 1956; 31: 421–7.

Monies D., et al. Lessons learned from large-scale, first-tier clinical exome sequencing in a highly consanguineous population. Am J Hum Genet 2019; 105: 879.

Oh W.-J., et al. Cleft palate is caused by CNS dysfunction in Gad1 and Viaat knockout mice. PloS One 2010; 5: e9758.

Olsen R.W., et al. GABA and epileptogenesis. Epilepsia 1997; 38: 399–407.

Retterer K., et al. Clinical application of whole-exome sequencing across clinical indi- cations. Genet Med 2016; 18: 696–704.

Ruzicka W.B., et al. Circuit- and diagnosis-specific DNA methylation changes at c-aminobutyric acid-related genes in postmortem human hippocampus in schizophrenia and bipolar disorder. JAMA Psychiatry 2015; 72: 541–51.

Saito K., et al. The physiological roles of vesicular GABA transporter during embryonic development: a study using knockout mice. Mol Brain 2010; 3: 40.

Saiz A., et al. Spectrum of neurological syndromes associated with glutamic acid decarboxylase antibodies: diagnostic clues for this association. Brain 2008; 131: 2553–63.

Scheffer I.E., et al. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. Epilepsia 2017; 58: 512–21.

Scheffer I.E., et al. Deciphering the concepts behind “Epileptic encephalopathy” and “Developmental and epileptic encephalopathy”. Eur J Paediatr Neurol 2020; 24: 11–14.

Sgado` P., et al. The role of GABAergic system in neurodevelopmental disorders: a focus on autism and epilepsy. Int J Physiol Pathophysiol Pharmacol 2011; 3: 223–35.

Sobreira N., et al. New tools for Mendelian disease gene identification: phenoDB variant analysis module; and GeneMatcher, a web-based tool for linking investigators with an interest in the same gene. Hum Mut 2015; 36: 425–31.

Soghomonian J.J., et al. Two isoforms of glutamate decarboxyl- ase: why? Trends Pharmacol Sci 1998; 19: 500–5.

Tohid H. Anti-glutamic acid decarboxylase antibody positive neuro- logical syndromes. Neurosciences (Riyadh) 2016; 21: 215–22.

Wang Y., et al. Downregulation of hippocampal GABA after hypoxia-induced seizures in neonatal rats. Neurochem Res 2011; 36: 2409–16.

Weckhuysen S, et al. KCNQ2 encephalopathy: emerging phenotype of a neonatal epileptic encephalopathy. Ann Neurol 2012; 71: 15–25.

C. Case series study: natural history

Clinico-radiological features, molecular spectrum, and identification of prognostic factors in developmental and epileptic encephalopathy due to inosine triphosphate pyrophosphatase (ITPase) deficiency

Abstract

Developmental and epileptic encephalopathy 35 (DEE 35) is a severe neurological condition caused by biallelic variants in *ITPA*, encoding inosine triphosphate pyrophosphatase, an essential enzyme in purine metabolism. We delineate the genotypic and phenotypic spectrum of DEE 35, analysing possible predictors for adverse clinical outcome. We investigated a cohort of 28 new patients and reviewed previously described cases, providing a comprehensive characterization of 40 subjects. Exome sequencing was performed to identify underlying *ITPA* pathogenic variants. Brain MRI scans were systematically analyzed to delineate the neuroradiological spectrum. Survival curves according to the Kaplan–Meier method and Log-Rank test were used to investigate outcome predictors in different subgroups of patients. We identified 18 distinct *ITPA* pathogenic variants, including 14 novel variants, and 2 deletions. All subjects showed profound developmental delay, microcephaly, and refractory epilepsy followed by neurodevelopmental regression. Brain MRI revision revealed a recurrent pattern of delayed myelination and restricted diffusion of early myelinating structures. Congenital microcephaly and cardiac involvement were statistically significant novel clinical predictors of adverse outcome. We refined the molecular, clinical, and neuroradiological characterization of ITPase deficiency, and identified new clinical

predictors which may have a potentially important impact on diagnosis, counselling, and follow-up of affected individuals.

1. Introduction

Developmental and epileptic encephalopathy 35 (DEE 35; MIM# 616647) is a rare neurodegenerative condition characterized by developmental delay (DD), microcephaly, feeding difficulties, early-onset refractory seizures (often within the first 6 months of life) followed by psychomotor stagnation/regression, and lethality in early childhood (*Handley et al., 2019; Kaur et al., 2019; Kevelam et al., 2015*). Cardiac and ocular involvement is frequently observed. White matter involvement is typical and consists of peculiar region-specific abnormalities, predominantly involving early myelinating structures and suggestive of a neuronal degenerative process (*Kevelam et al., 2015*).

Biallelic variants in inosine triphosphate pyrophosphohydrolase (*ITPA*; MIM# 147520) have been first associated with DEE 35 in seven patients from four unrelated families by *Kevelam et al* (*2015*). More recently, a few additional affected individuals have been reported (*Bierau et al., 2007; Burgess et al., 2019; Burgis, 2016; Handley et al., 2019; Kaur et al., 2019; Rochtus et al., 2020; Sakamoto et al., 2020*). Inosine triphosphate pyrophosphatase (ITPase), is an essential enzyme that removes the spontaneously arising noncanonical nucleotides inosine triphosphate (ITP) and deoxy-inosine triphosphate (dITP) from the cellular nucleotide pool, playing a pivotal role in purine metabolism and cell function (*Holmes et al., 1979; Galperin et al., 2006*).

We report 28 unpublished individuals with DEE 35 from 23 unrelated families of different ancestry and review 12 previously reported cases. This clinical and neuroradiological

characterization of a large cohort of 40 individuals allows a refined phenotypic description of ITPase deficiency. We further systematically investigate possible clinical predictors for adverse outcome in this rare condition.

2. Materials and methods

2.1 Ethical considerations

This study adheres to the principles set out in the Declaration of Helsinki and was locally approved by the local Ethics Committees of the involved Institutions: Mayo Clinic, Rochester, 16-004682; King Faisal Specialist Hospital and Research Centre, (RAC: 2121053, 2120022, and 2161245); Technische Universitaet Muenchen (TUM), 5360/12 S; University College London (UCL), project ID: 07/N018, REC Ref: 07/Q0512/26). No IRB approval was necessary for retrospective data analysis of a single patient for the following Institutions: Alberta Children's Hospital, Calgary, Canada; Al-Jawhara Centre for Molecular Medicine, Kingdom of Bahrain; Center for Neurogenomics and Cognitive Research, VU University, The Netherlands; Children's Hospital, Cantonal Hospital Lucerne, Switzerland; Dr. Sami Ulus Training and Research Hospital for Maternity and Children; Emma Children's Hospital, Amsterdam Leukodystrophy Center, The Netherlands; The Children's Memorial Health Institute, Poland. The authors obtained and archived written informed consents from parents or legal guardians of the enrolled subjects to publish genetic and clinical data, including brain magnetic resonance imaging (MRI) images (P1 and P3).

2.2 Patient enrolment

We ascertained the genotype and phenotype information for 28 novel subjects with severe epileptic encephalopathy. Patients were recruited through international collaboration, also using Genematcher (*Sobreira et al., 2015*), from several clinical and research centres in Europe, Africa, Middle East, North America and New Zealand (for details see the Supplementary Material). Written informed consent was obtained from the parents or legal guardians of all enrolled subjects. Phenotypes of two of these individuals (P9 and P14), who were partially described previously, have been extensively reported and updated (*Bierau et al., 2007; Muthusamy et al., 2021*).

2.3 Previously reported cases assessment

All articles indexed in PubMed (https://pubmed.ncbi.nlm.nih.gov/?term=itpa) between October 2015, when *ITPA* variants were first associated with DEE 35 by *Kevelam et al.* (*2015*), and March 2021 were retrieved using the terms “ITPA”, “ITPase deficiency”, and “epileptic encephalopathy 35”. All the articles were thoroughly reviewed concerning the molecular, clinical and neuroradiological spectrum associated with DEE 35. Inclusion criteria for previously published patients were: availability of clinical data (with a focus on epilepsy, developmental, neuro-, cardio- and ophthalmological findings), identification of (likely) pathogenic *ITPA* variants, lack of duplication from other previous reports. Exclusion criteria were: ambiguous clinical presentation not consistent with DEE 35 and inconclusive genetic testing.

2.4 Variant identification and analysis

Next-Generation Sequencing (NGS) panel for epileptic encephalopathies (P5 and P9) or exome sequencing (ES) (P1-4, P6-8, P10-28) was performed on genomic DNA extracted from peripheral blood leukocytes (P1-15 and P17-28) or ORAcollect buccal swab (OCR-100; DNA Genotek, Ottawa, CA) (P16) using standard local protocols. Chromosomal microarray analysis (CMA) was performed in P2, P3, P26, and P27 according to standard methods (Shaw-Smith et al., 2004). The identified variants were filtered according to minor allele frequency ≤ 0.001 in genomic databases (Genome Aggregation Database – gnomAD (*Lek et al., 2016*), Iranome, in-house database of 16,000 control exomes, the Munich in house database (https://github.com/mri-ihg/EVAdb), Great Middle Eastern Variome Project – GME), conservation (Genomic Evolutionary Rate Profiling – GERP, http://mendel.stanford.edu/SidowLab/downloads/gerp/), and predicted effect on protein structure and function. *In silico* prediction tools were used for the interpretation of candidate variants, including Combined Annotation Dependent Depletion (CADD, https://cadd.gs.washington.edu), Mutation Taster (http://www.mutationtaster.org), Sorting Intolerant From Tolerant (SIFT, https://sift.bii.a-star.edu.sg), and Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/). Candidate variants were eventually classified according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines (Richards et al., 2015). Sanger sequencing was performed for validation and segregation analysis. All *ITPA* variants are reported according to RefSeq NM_033453.3, GenBank NC_000020.11. The change in protein stability was calculated with PyRosetta for the variants presented herein and the presumed neutral variants present in gnomAD (https://gnomad.broadinstitute.org/) (Chaudhury et al., 2010; Karczewski et al., 2020). All novel variants reported were deposited in the Leiden Open Variation Database (LOVD,

https://www.lovd.nl) with the following accession numbers: #0000831814, #0000831817, #0000831819, #0000831820, #0000831821, #0000831822, #0000831823, #0000831824, #0000831825, #0000831826, #0000831827, #0000831828, #0000831829, and #0000831830.

2.5 Neuroimaging analysis

Brain MRI scans were locally performed during routine patient care. Of these, 28 scans of 19 individuals were collected and analysed in detail at the Amsterdam Leukodystrophy Center (The Netherlands). MRI scans of adequate quality, at least comprising T1-weighted and transverse T2-weighted images were systematically scored according to a previously published protocol by two independent authors (MDS and MSvdK) (*Van der Knaap et al., 1999*). Additional sequences, such as diffusion-weighted imaging (DWI), magnetic resonance spectroscopy (MRS), and contrast-enhanced images were also evaluated, when available. MRIs were divided into 4 age groups (≤ 2 months; 2 - ≤ 4 months; 4 - ≤ 8 months; > 8 months) and neuroimaging features were systematically analysed in each group of patients.

2.6 Statistical analysis

Descriptive statistics were performed first. Categorical variables were reported as absolute frequencies and percentages, and quantitative variables as median values and first and third quartiles (1^{st} – 3^{rd} q). For comparison of frequencies (e.g., frequency of deaths among males vs. females), the chi-square test or Fisher's Exact test (in case of expected frequencies < 5) was used. Survival curves according to the Kaplan–Meier method were drawn for socio-demographic (sex, age at presentation) and clinical variables (e.g., number of presenting signs, congenital microcephaly, cardiac involvement). Death was considered the event of interest. The Log-Rank

test was used to compare different survival curves. Incidence rates of events were calculated for each category defined by demographic and clinical variables and reported with their 95% confidence intervals. All statistical tests were two-sided and a p-value < 0.05 was considered statistically significant. Statistica (release 9.1, StatSoft Corporation, Tulsa, OK, USA) was used for all the bivariate analyses. Stata (release 11.0, College Station, TX, USA) was used for the Fisher's exact test and to calculate incidence rates and their 95% confidence intervals.

3. Results

We identified 28 new patients harboring biallelic *ITPA* variants and reviewed 12 previously reported subjects with DEE 35 from 4 peer-reviewed articles (*Bierau et al., 2007; Burgess et al., 2019; Burgis, 2016; Handley et al., 2019; Kaur et al., 2019; Kevelam et al., 2015; Rochtus et al., 2020; Sakamoto et al., 2020*), for a total cohort of 40 affected individuals.

3.1 *ITPA* Variants

Eighteen pathogenic or likely pathogenic variants in *ITPA* were detected in the studied cohort (Table 1). In addition to previously reported variants, 14 novel *ITPA* variants were detected in our cohort, including 6 missense (Figure 1a, https://michelanglo.sgc.ox.ac.uk/r/itpa) and 8 loss-of-function (LoF) variants.

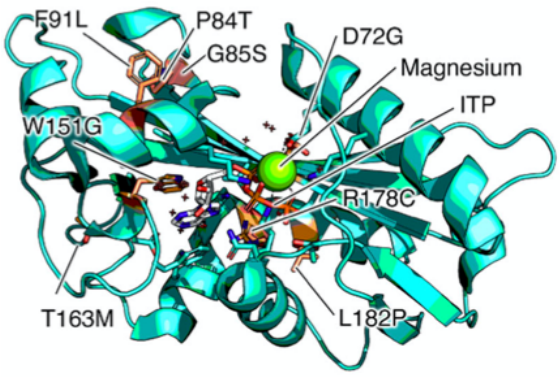


Figure 1. Variant localization. (a) Structure model of human ITPA protein showing the localization of the residues affected by ITPA missense variants in relation to the ITP-binding cleft and Mg2+ binding

All these variants were absent in homozygous state from gnomAD, had a low allele frequency in heterozygous state (ranging from 0 to 0.00003551), affected conserved residues, and were predicted to be damaging by several *in silico* tools (Table 1).

In particular, a significant structural destabilization could be predicted for the tested missense variants identified in our cohort as compared to variants frequently found in the healthy population in gnomAD v3.1 (Figure S1).

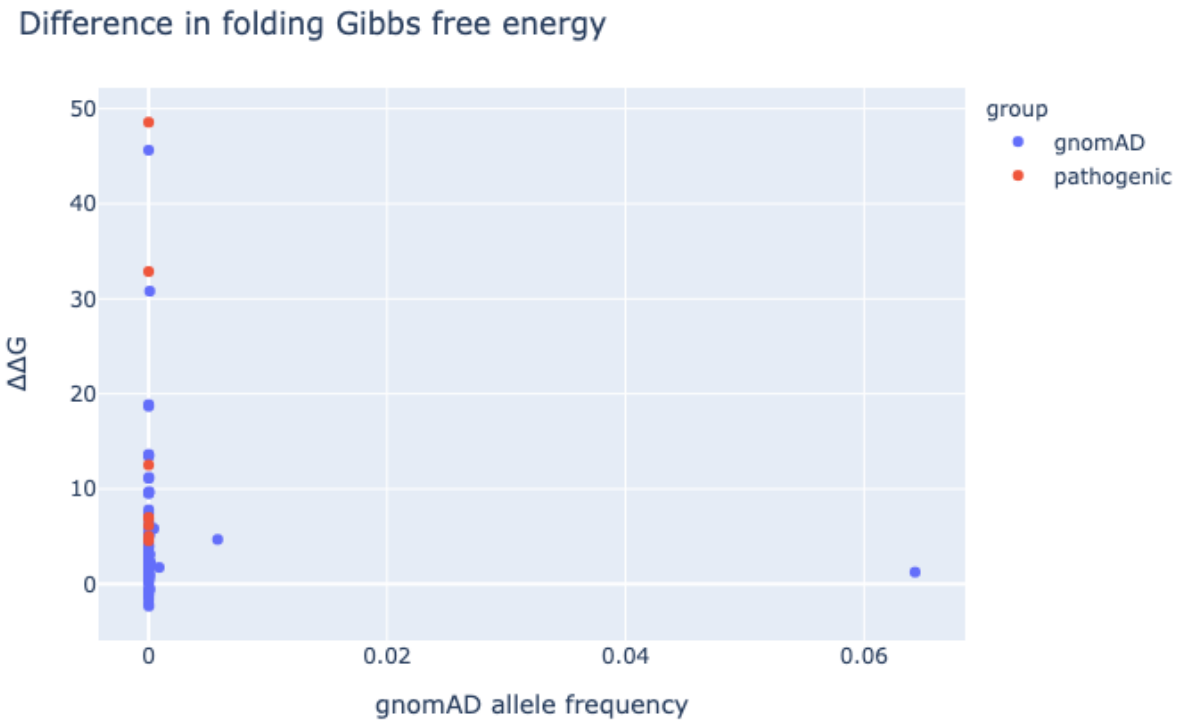


Figure S1. Impact of ITPA missense variants on protein folding. The graph illustrates the difference in folding Gibbs free energy for pathogenic ITPA missense variants in comparison to variants present in healthy individuals (gnomAD, https://gnomad.broadinstitute.org/).

P2 and P3 had a large heterozygous 1.1-Mb deletion, encompassing *ITPA* (hg19, chr20: 2,816,108 – 3,955,033), whereas an intragenic deletion, encompassing exon 1-5 (hg19, chr20: 3,189,364 – 3,196,608) was detected in P27. Seventeen subjects carried homozygous *ITPA* variants, whereas compound heterozygous variants were found in the remaining individuals.

Sanger sequencing confirmed a carrier status for all the parents and showed concordant segregation of the variants with the clinical phenotype.

Table 1. Molecular spectrum of *ITPA* variants.

ITPA variant [NM_033453.3]	Source	ExAC/ gnomAD	ClinVar (ID)/dbSNP/ PMID	SIFT	Mutation Taster	HSF	GERP score	CADD score	ACMG/AMP classification	Min. distance	ΔΔG [kcal/mol]
c.67-1G>A; p.?	This study	-	-	N/A	DC (1)	Broken WT Acceptor Site†	5.49	34	P (PVS1, PM2, PP3)	-	-
c.124+1G>A; p.?	This study	-	LP (VCV000646228.2)	N/A	DC (1)	Broken WT Donor Site	5.45	33	P (PVS1, PM2, PP3, PP5)	-	-
c.124+1G>C; p.?	This study	0.000007952 (2 het)	rs376142053	N/A	DC (1)	Broken WT Donor Site	5.45	33	P (PVS1, PM2, PP3)	-	-
c.142delG; p.(Glu48Serfs*41)	This study	-	-	N/A	DC (1)	N/A	5.68	N/A	P (PVS1, PM2, PP3)	-	-
c.215A>G; p.(Asp72Gly)	This study	-	-	D (0)	DC (1)	N/A	5.32	33	LP (PS4, PM2, PP1, PP3, PP4)	3.0‡	+4.5
c.250C>A; p.Pro84Thr	This study	-	-	D (0)	DC (1)	N/A	5.32	26.8	VUS (PM2, PP3, PP4)	7.2	+6.9
c.253G>A; p.Gly85Ser	This study	0.000003977 (1 het)	rs1343080275	D (0)	DC (1)	N/A	5.32	28.5	VUS (PM2, PP3, PP4)	5.2	+48 (approx.)
c.264-1G>A p.(Ile88Metfs*59)	Sakamoto et al. (2020); This study	0.000007954 (2 het)	rs781254071	N/A	DC (1)	Broken WT Acceptor Site	5.77	34	P (PVS1, PM2, PP3)	-	-
c.264-607_295+1267del p.?	Kevelam et al. (2015)	-	P (VCV000218088)/ 26224535	N/A	N/A	Splice junction loss	N/A	N/A	P (PVS1, PM2, PP3)	-	-
c.271T>C p.(Phe91Leu)	This study	-	LP (VCV000807614)	D (0)	DC (1)	N/A	5.77	28.9	LP (PM2, PM3, PP3, PP5)	7.3	+7.0
c.359_366dup TCAGCACC p.(Gly123Serfs*104)	Kevelam et al. (2015); This study	0.000035551 (10 het)	LP (SCV001168653)/ rs946985349/26224535	N/A	N/A	N/A	N/A	N/A	P (PVS1, PM2, PP3, PP5)	-	-
c.451T>G p.(Trp151Gly)	Kaur et al. (2019)	-	-	D (0)	DC (1)	N/A	5.32	32	LP (PM1, PM2, PP3, PP4, PP5)	2.3	+12.5
c.452G>A p.(Trp151*)	Kevelam et al.; This study (2015)	0.000053009 (15 het)	P (VCV000218089)/ rs200086262/ 26224535, 30856165, 31618474	D due to stop	DC (1)	N/A	5.32	46	P (PVS1, PM2, PP3)	-	-
c.456_488+7del p.?	Handley et al., (2019)	-	-	N/A	N/A	Splice junction loss	N/A	N/A	P (PVS1, PM2, PP3)	-	-
c.488C>T p.(Thr163Met)	This study	0.000007970 (2 het)	VUS (VCV000431714)/ rs758706191	D (0)	DC (1)	N/A	5.32	34	LP (PM1, PM2, PP3, PP4)	8.1	+6.1
c.488+5_488+6delGG p.?	This study	-	-	N/A	N/A	Broken WT Donor Site	5.32	N/A	LP (PM2, PP3, PP4)	-	-

c.489-2A>G; p.?	This study	-	-	N/A	DC (1)	Broken WT Acceptor Site	5.82	34	P (PVS1, PM2, PP3)	-	-
c.489-1 G>A; p.?	Sakamoto et al. (2020)	-	-	N/A	DC (1)	Broken WT Acceptor Site	5.82	34	P (PVS1, PM2, PP3)	-	-
c.489-1 G>T ; p.?	This study	-	-	N/A	DC (1)	Broken WT Acceptor Site	5.82	34	P (PVS1, PM2, PP3)	-	-
c.519delC ; p.(Asn173 Lysfs*51)	This study	0.00001595 (4 het)	rs748042110	N/A	DC (1)	N/A	5.82	N/A	P (PVS1, PM2, PP4)	-	-
c.532C>T ; p.Arg178 Cys	Kevelam et al. (2015); This study	0.000003988 (1 het)	P (VCV000218090)/ rs746930990/26224535	D (0)	DC (1)	N/A	5.82	32	LP (PM1, PM2, PP3, PP4, PP5)	1.9	+5.0
c.545T>C ; p.Leu182 Pro	This study	-	VUS (VCV000452647)	D (0)	DC (1)	N/A	5.82	26.4	VUS (PM2, PP3, PP4)	9.4	33 (approx.)

ACMG/AMP American College of Medical Genetics and Genomics and the Association for Molecular Pathology, D damaging; DC disease causing, ΔΔG difference in relative Gibbs free energy of folding, GME Greater Middle East Variome Project, HSF Human Splice Finder, IR, Iranome; LP Likely pathogenic, Min. distance closest distance of a residue atom to either ITP or Mg2+ in the model, N/A Not applicable, P Pathogenic, VUS Variant of Unknown Significance, WT wildtype. † Possible additional activation of an intronic cryptic acceptor site. ‡ Probable catalytic residue. Variants reported according to RefSeq NM_033453.3, GenBank NC_000020.11.

3.2 Clinical Delineation of DEE 35

The phenotype observed in the studied cohort was consistent with severe DEE (Figure 1b). The age range (current age) was 1-72 months, with a median age at presentation of 3 months, a median age at last follow-up of 42 months, and a male to female sex ratio of 0.69. Severe DD in the first few months of life was diagnosed in 19/40 (47.5%) patients, whereas developmental stagnation/arrest after seizure onset was present in all subjects (Table 2).

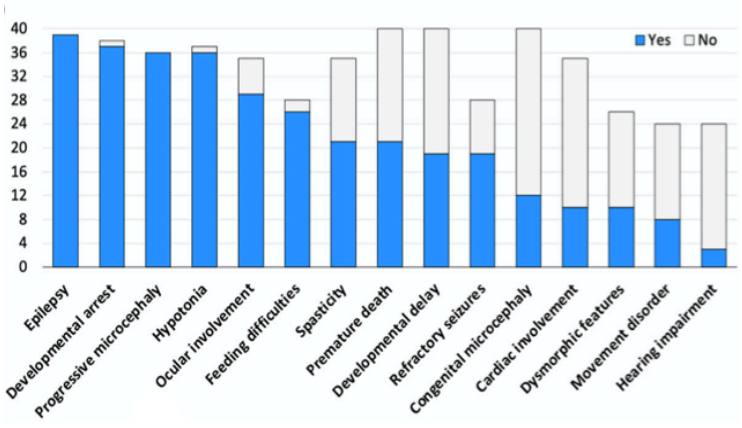


Figure 1. Clinical phenotype. (b) Bar graph illustrating the distribution of core clinical features of DEE 35, from the most to the less common. Blue bars indicate the number of patients in whom a specific feature is present whereas grey bars indicate the number of subjects in whom that feature was ascertained but it was absent. Ocular involvement includes cataract, visual impairment, optic atrophy, and retinal cone dysplasia. Cardiac involvement consists of dilated cardiomyopathy and rhythm disturbances. Movement disorders include tremors, dystonia, choreoathetoid movements, and dyskinesia. Dysmorphic features were observed in absence of a distinctive facial gestalt.

Table 2. Clinical features of DEE 35 patients.

Characteristics at clinical presentation	**n/N (%)**
Sex: Male	12/38 (31.6 %)
Female	26/38 (68.4 %)
Age at presentation (months), *median [1st- 3rd q]*	3 [1 - 4]
Age at first seizure (months), *median [1st- 3rd q]*	4 [2 - 5]
N. of presenting signs†: 1	15/40 (37.5 %)
2	20/40 (50.0 %)
3	3/40 (7.5 %)
4	2/40 (5.0 %)
Congenital microcephaly	12/40 (30.0 %)
Developmental delay	19/40 (47.5 %)
Seizures	27/40 (67.5 %)
Small for gestational age	7/34 (20.6 %)
Clinical features	**n/N (%)**
Progressive microcephaly	36/36 (100 %)
Epilepsy	39/39 (100 %)
Refractory seizures	19/28 (67.9 %)
Developmental arrest after seizure onset	37/38 (97.4 %)
Feeding difficulties	26/28 (92.9 %)
Progressive hypotonia	36/37 (97.3 %)
Spasticity	21/35 (60.0 %)
Movement disorder	8/24 (33.3 %)
Ocular involvement	29/35 (82.9 %)
Cataract	16/28 (57.1 %)
Visual impairment	5/28 (17.9 %)
Optic atrophy	2/28 (7.1 %)
Retinal cone dysplasia	2/28 (7.1 %)
Other‡	3/28 (10.7 %)
Cardiac involvement	10/35 (28.6 %)
Dilated cardiomyopathy	5/10 (50.0 %)
Rhythm disturbances	4/10 (40.0 %)
Both	1/10 (10.0 %)
Dysmorphic features	10/26 (38.5 %)
Hearing impairment	3/24 (12.5 %)
Life status: Alive	19/40 (47.5 %)
Death	21/40 (52.5 %)

N. number. † Including microcephaly, psychomotor delay, seizures, hypotonia, movement disorder. ‡ Strabismus and refractive errors.

A variable number (1-4) of presenting signs was observed (Table 2). In particular, congenital microcephaly was diagnosed in 30% of cases. Interestingly, poor neonatal adaptation was only occasionally observed (P2 and P3). Approximately one-fifth of subjects were born

small for gestational age (SGA) and significant swallowing difficulties were very common from birth onwards (92.9%), leading to failure to thrive in all cases.

The age at first seizure ranged from 2 days to 7 months, with a median age of 4 months. Febrile seizures were only observed in one case (P15), at the age of 4 months. All affected individuals had epilepsy, with refractory seizures in 68% of patients (Figure 1c). Clonic/myoclonic, tonic, and tonic-clonic seizures were observed, occasionally leading to status epilepticus in two patients (P4 and P9). Electroencephalographic abnormalities were variable and included focal, multifocal, and diffuse/generalized discharges often within a slow and disorganized background, consistent with the underlying encephalopathy.

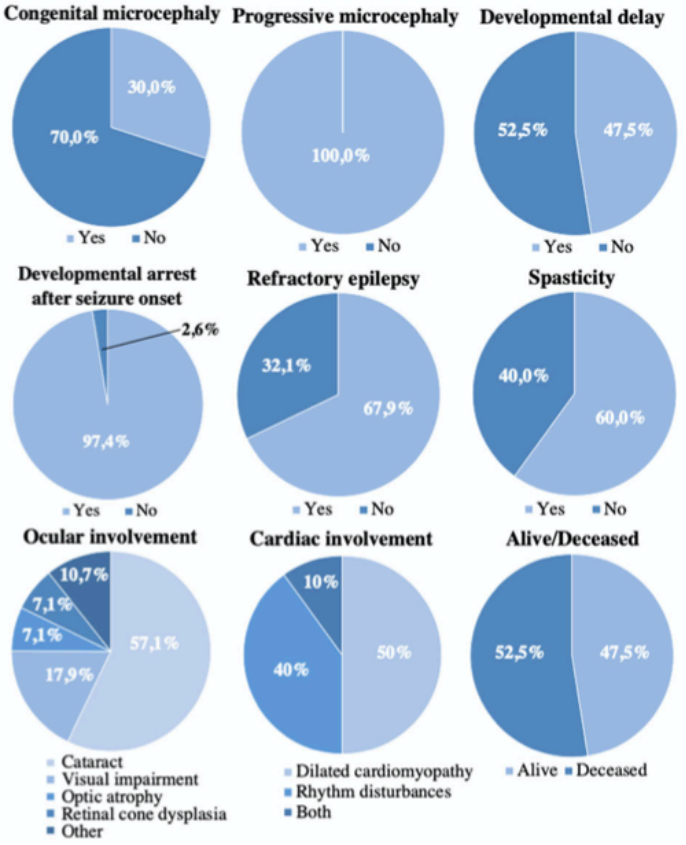


Figure 1. Clinical spectrum. (c) Pie charts illustrating the percent distribution of specific neurological and extra-neurological manifestations of DEE 35. Rhythm disturbances include tachycardia and long QT syndrome.

Progressive microcephaly and axial hypotonia were common, occurring in 100% and 97.3% of patients, respectively. Additional relevant neurological findings included appendicular spasticity with hyperreflexia (60%) and hyperkinetic movement disorders (33%) (e.g., tremors,

choreoathetoid movements, dystonia, and dyskinesia). Extra-neurological manifestations were frequent and consisted of variable ocular disorders in 83% of cases and cardiomyopathy and/or rhythm disturbances in 29% of patients. Mild dysmorphic features were present in 39 % of subjects, but a definite facial gestalt is absent. Early death occurred in 53% of cases, with a median age at death of 24 months. Causes of death included cardiac dysfunction in subjects with heart disease and seizure-related apnoea (P2 and P24) or aspiration pneumonia (P17 and P18) in those without cardiac involvement.

3.3 Neuroradiological phenotype

Twenty brain MRI scans (of 13 patients) were of sufficient quality to be analysed in detail (Figure 1d). MRI characteristics per age group are reported in Table 3.

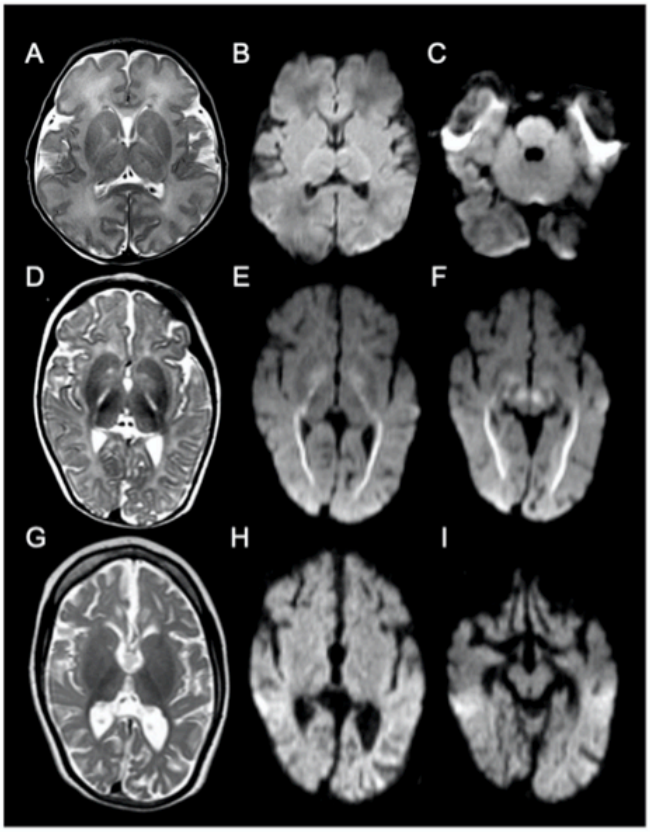


Figure 1. Neuroimaging. (d) MRI findings. MRI of P3 at age 6 days (A, B, C). T2-weighted image (A) shows no atrophy and no signal abnormalities. There is no restricted diffusion (B,C) on diffusion-weighted imaging (DWI, apparent-diffusion coefficient maps not shown). MRI of P1 at age 6 months (D, E, F) shows no atrophy, but moderately delayed myelination and T2-hyperintensity of the posterior limb of the internal capsule (PLIC; D). Restricted diffusion is seen in the optic radiation, PLIC (E), and decussation of the superior cerebellar peduncles (F). Mild diffusion restriction is seen in the globus pallidus (E). MRI of P1 at age 2 years and 8 months (G, H, I) shows seriously deficient myelination and severe cerebral atrophy (G). Restricted diffusion is no longer present (H, I). DEE 35, developmental and epileptic

Table 3. Brain MRI characteristics per age.

Age at MRI	≤ 2 months	2 - ≤ 4 months	4 - ≤ 8 months	> 8 months
Number of patients/scans	3 patients / 3 scans	3 patients / 4 scans	7 patients / 7 scans	4 patients / 6 scans
Myelination: delayed	0/3	1/4 slightly delayed	2/7 slightly delayed, 2/7 mildly delayed, 2/7 delayed	4/6 delayed, 2/6 severely delayed
Cerebral cortex: abn. T2 signal	0/3	0/4	0/7	0/6
Cerebral hemispheric WM: abn. T2 signal	0/3	0/4	0/7	0/6
Basal nuclei/thalami: abn. T2 signal	0/3	0/4	1/7, globus pallidus	1/6, globus pallidus
ALIC: abn. T2 signal	0/3	0/4	1/7	0/6
PLIC: abn. T2 signal	1/3 no low signal	2/4	6/7	3/6
Midbrain: abn. T2 signal	0/3	0/4	1/7, decussation SCP and left crus cerebri	1/6, WM around red nucleus
Pons: abn. T2 signal	0/3	1/4, CTT	0/7	1/6, CTT
Medulla: abn. T2 signal	0/3	0/4	0/7	1/6, everything except inf. olivary n.
SCP: abn. T2 signal	0/3	0/4	0/7	0/6
MCP: abn. T2 signal	0/3	0/4	2/7	1/6
Cerebellar cortex: abn. T2 signal	0/3	0/4	0/7	0/6
Cerebellar WM: abn. T2 signal	0/3	1/4	1/7	1/6, peridentate WM
Hilus dentate n.: abn. T2 signal	0/3	0/4	1/7	0/6
Cerebral atrophy	0/3	3/4 slight, 1/4 mild	3/7 mild, 2/7 moderate	1/6 mild, 1/6 moderate, 3/6 moderate-severe, 1/6 severe
Corpus callosum: thin	0/3	0/4	4/7	6/6
Thalamus: atrophy	0/3	0/4	0/7	2/6
Cerebellar atrophy	0/3	1/4 slight	0/7	1/6 slight, 2/6 mild, 1/6 moderate
Diffusion restriction	1/2	2/3	5/5	3/5
Diffusion restriction specified	globus pallidus, PLIC, left crus cerebri, decussation SCP, SCP, hilus dentate n., cerebellar WM, MCP, ICP	OR 1/2, globus pallidus 1/2, PLIC 2/2, crus cerebri 1/2, brachium inf. colliculus 1/2, CTT in midbrain, pons and medulla 1/2, decussation SCP 1/2, SCP 1/2, hilus dentate n. 1/2	OR 4/5, globus pallidus 4/5, ALIC 1/5, PLIC 5/5, crus cerebri 4/5, brachium inf. colliculus 2/5, CTT in pons 2/5, decussation SCP 5/5, SCP 4/5, hilus dentate n. 1/5, cerebellar WM 2/5, ICP 2/5, pyramids 1/5	cerebral hemispheric WM 2/3, corpus callosum 1/3, OR 2/3, ALIC 3/3, PLIC 3/3, crus cerebri 1/3, WM around red nucleus 1/3, brachium inf. colliculus 1/3, CTT in pons 2/3, decussation SCP 2/3, SCP 1/3, cerebellar WM 1/5, MCP 1/3, ICP 2/3
Contrast abnormalities	0/0	0/0	0/3	0/1
MR spectroscopy: lactate elevation	0/1	0/2	1/1	0/2
Extra features	1/3 dilated inferior horns, 1/3 rarefaction of temporal poles	0/4	2/7 perivascular spaces, 1/7 PLIC and left crus cerebri rarefied	1/6 rarefaction of temporal poles

Abn. abnormal, WM white matter, OR optic radiation, ALIC anterior limb of the internal capsule, PLIC posterior limb of the internal capsule, CTT central tegmental tracts, SCP superior cerebellar peduncle, MCP middle cerebellar peduncle, ICP inferior cerebellar peduncle; n. nucleus.

In patients younger than 2 months (3 scans in 3 patients), brain MRI revealed very few abnormalities. In particular, myelination was normal and no cerebral or cerebellar atrophy or other brain lesions were observed. The posterior limb of the internal capsule (PLIC) did not show

the normal low T2 signal in one scan, but no focal lesion was present. DWI was available in two MRI scans and one showed restricted diffusion of specific structures (Table S2).

Table S1. MRI characteristics, Early MRI scans (≤ 4 months).

Patient / Family	P2 / F2	P2 / F2	P3 / F2	P12 / F11	P24 / F20	P26 / F20	P27 / F21
Age at MRI	2.5 months	4 months	6 days	3 weeks	3 months	1.5 months	3 months
Myelination	normal	normal	normal	normal	normal	normal	slightly delayed
Cerebral cortex	-	-	-	-	-	-	-
Basal nuclei/thalami	-	-	-	-	-	-	-
Cerebral hemispheric WM	-	-	-	-	-	-	-
Corpus callosum	normal for age	normal for age	normal for age	normal for age	normal for age	normal for age	normal for age
ALIC	-	-	-	-	-	-	-
PLIC	-	-	-	no low signal	+	-	+
Midbrain	-	-	-	-	-	-	-
Pons	-	+, CTT	-	-	-	-	-
Medulla	-	-	-	-	-	-	-
SCP	-	-	-	-	-	-	-
MCP	-	-	-	-	-	-	-
Cerebellar cortex	-	-	-	-	-	-	-
Cerebellar WM	-	-	-	-	+	-	-
Hilus dentate n.	-	-	-	-	-	-	-
Cerebral atrophy	slight	mild	no	no	slight	no	slight
Cerebellar atrophy	no	no	no	no	no	no	slight
Diffusion restriction	no	globus pallidus, PLIC, brachium inf. colliculus, CTT in midbrain, pons and medulla, decussation SCP, SCP, hilus dentate n.	no	globus pallidus, PLIC, left crus cerebri, decussation SCP, SCP, hilus dentate n., cerebellar WM, MCP, ICP	not done	Not available	OR, PLIC, crus cerebri
Contrast abnormalities	not administered	not administered	not administered	not administered	not administered	not administered	not administered
MR spectroscopy	not done	no lactate elevation	no lactate elevation	not done	not done	not done	no lactate elevation
Extra features				dilated inferior horns, rarefaction of temporal poles			

+ indicates abnormal T2 signal; - indicates no T2 signal abnormalities; OR, optic radiation; ALIC, anterior limb of the internal capsule; CTT, central tegmental tracts; ICP, inferior cerebellar peduncle; MCP, middle cerebellar peduncle; n., nucleus; PLIC, posterior limb of the internal capsule; SCP, superior cerebellar peduncle; WM, white matter

In patients aged between 2 and 4 months (4 scans in 3 patients), only one scan showed slightly delayed myelination. The PLIC contained a T2-hyperintense focal lesion in two scans.

Initial cerebral atrophy could be observed in all scans and cerebellar atrophy was present in one scan. DWI was available in three MRI scans and restricted diffusion of specific structures was present in two (Table S1). In patients aged between 4 and 8 months (7 scans in 7 patients), variably delayed myelination was present in six scans. The PLIC contained a T2-hyperintense lesion in six scans. Mild-to-moderate cerebral atrophy was seen in five scans, typically associated with a thin corpus callosum, while no cerebellar atrophy was observed. In all scans, there was restricted diffusion of specific structures (Table S2). In patients older than 8 months (6 scans in 4 patients), all scans showed moderately to severely delayed myelination (Table S3). The PLIC contained a T2-hyperintense lesion in three scans. The thalamus was atrophic in two scans. There was mild-to-severe cerebral atrophy with a thin corpus callosum in all, and mild-to-moderate cerebellar atrophy in four scans. Restricted diffusion of specific structures was present in three patients (Table S3).

Diffusion restriction was separately reviewed. Commonly involved structures were globus pallidus, PLIC, pyramidal tracts in the brain stem, cerebellar white matter, hilus of the dentate nucleus, superior cerebellar peduncles, decussation of the superior cerebellar peduncles, middle cerebellar peduncles, optic radiation, brachium of the inferior colliculus, and central tegmental tracts in the midbrain, pons, and medulla. In older patients, restricted diffusion could also be observed in the anterior limb of the internal capsule, corpus callosum, and cerebral hemispheric white matter.

Two patients had sequential MRI scans. In one case, restricted diffusion was absent at 2.5 months but present in specific structures at 4 months. The other patient had four MRI scans (at 6 months, 1 year, 1.7 years, and 2.8 years). Restricted diffusion of specific structures and a T2-

hyperintense lesion of the PLIC were present at 6 months and 1 year, but were not observed at 1.7 and 2.8 years. Cerebral and cerebellar atrophy increased over time.

Table S2. MRI characteristics, Intermediate MRI scans (4 - ≤ 8 months).

Patient / Family	P1 / F1	P5 / F4	P6 / F5	P8 / F7	P9 / F8	P14 / F13	P28 / F23
Age at MRI	6 months	4.5 months	4.5 months	6 months	4.5 months	5 months	8 months
Myelination	moderately delayed	normal	slightly delayed	mildly delayed	mildly delayed	slightly delayed	moderately delayed
Cerebral cortex	-	-	-	-	-	-	-
Basal nuclei/thalami	-	-	+, globus pallidus	-	-	-	-
Cerebral hemispheric WM	-	-	-	-	-	-	-
Corpus callosum	normal	normal	normal	thin	thin	thin	thin
ALIC	-	-	+	-	-	-	-
PLIC	+	+	+	+	+	+	-
Midbrain	+, decussation SCP, left crus cerebri	-	-	-	-	-	-
Pons	-	-	-	-	-	-	-
Medulla	-	-	-	-	-	-	-
SCP	-	-	-	-	-	-	-
MCP	+	-	-	-	-	+	-
Cerebellar cortex	-	-	-	-	-	-	-
Cerebellar WM	-	-	-	-	-	+	-
Hilus dentate n.	+	-	-	-	-	-	-
Cerebral atrophy	no	no	mild	mild	moderate	mild	moderate
Cerebellar atrophy	no	no	no	no	no	no	no
Diffusion restriction	OR, PLIC, decussation SCP	not done	globus pallidus, PLIC, crus cerebri, decussation SCP, SCP, hilus dentate n., pyramids	OR, globus pallidus, PLIC, crus cerebri, CTT in pons, decussation SCP, SCP, hilus dentate n., ICP	OR, globus pallidus, PLIC, crus cerebri, brachium inf. colliculus, CTT in pons, decussation SCP, SCP, hilus dentate n., cerebellar WM, ICP	OR, globus pallidus, ALIC, PLIC, crus cerebri, brachium inf. colliculus, decussation SCP, SCP, hilus dentate n., cerebellar WM	not done
Contrast abnormalities	no	not administered	not administered	not administered	no	no	not administered
MR spectroscopy	not done	not done	lactate elevation	not done	not done	not available	not done
Extra features	PLIC and left cerebral peduncle rarefied		several Virchow Robin-spaces		several Virchow Robin-spaces		

+ indicates abnormal T2 signal; - indicates no T2 signal abnormalities; ALIC, anterior limb of the internal capsule; CTT, central tegmental tracts; ICP, inferior cerebellar peduncle; MCP, middle cerebellar peduncle; n., nucleus; OR, optic radiation; PLIC, posterior limb of the internal capsule; SCP, superior cerebellar peduncle; WM, white matter.

Table S3. MRI characteristics, Late MRI scans (> 8 months).

Patient / Family	P1 / F1	P1 / F1	P1 / F1	P2 / F2	P4 / F3	P10 / F9
Age at MRI	1 year	1.7 years	2.8 years	1.7 years	1.4 years	4.5 years
Myelination	moderately delayed	moderately delayed	moderately delayed	severely delayed	severely delayed	moderately delayed
Cerebral cortex	-	-	-	-	-	-
Basal nuclei/thalami	+, globus pallidus	-	-, thalamus atrophic	-, thalamus atrophic	-	-
Cerebral hemispheric WM	-	-	-	-	-	-
Corpus callosum	thin	thin	thin	thin	thin	thin
ALIC	-	-	-	-	-	-
PLIC	+	-	-	+	-	+
Midbrain	-	-	-	+, WM around red nucleus	-	-
Pons	-	-	-	+, CTT	-	-
Medulla	-	-	-	+, everything except inferior olivary n.	-	-
SCP	-	-	-	-	-	-
MCP	+	-	-	-	-	-
Cerebellar cortex	-	-	-	-	-	-
Cerebellar WM	-	-	-	+, peridentate WM	-	-
Hilus dentate n.	-	-	-	-	-	-
Cerebral atrophy	mild	moderate	moderate-severe	severe	moderate-severe	moderate-severe
Cerebellar atrophy	slight	mild	mild	moderate	no	no
Diffusion restriction	pericentral and parietal WM, OR, ALIC, PLIC, decussation SCP, MCP	no	no	cerebral hemispheric WM, corpus callosum, ALIC, PLIC, crus cerebri, WM around red nucleus, brachium inf. colliculus, CTT in pons, decussation SCP, SCP, cerebellar WM, ICP	OR, ALIC, PLIC, CTT in pons, ICP	not done
Contrast abnormalities	not administered	no	not administered	not administered	not administered	not administered
MR spectroscopy	not done	no lactate elevation	not done	not done	no lactate elevation	not done
Extra features				tram-track T2-hyperintensity PLIC	rarefaction of temporal poles	

+ indicates abnormal T2 signal; - indicates no T2 signal abnormalities; ALIC, anterior limb of the internal capsule; CTT, central tegmental tracts; ICP, inferior cerebellar peduncle; MCP, middle cerebellar peduncle; n., nucleus; OR, optic radiation; PLIC, posterior limb of the internal capsule; SCP, superior cerebellar peduncle; WM, white matter.

3.4 Predictors of early mortality in DEE 35

Epidemiologic and clinical parameters were considered for the investigation of outcome predictors through the analysis of number/percentage and incidence of death events (Table 4). The studied categories included sex, age at presentation, age at first seizure, number of

presenting signs, congenital microcephaly, DD, seizures, hypotonia, SGA status, spasticity, ocular involvement, and cardiac involvement.

Table 4. Number and percentage of events, and Incidence Rates of deaths by different clinical categories.

	N. of deaths / N. of patients	P	Incidence Rate × 100 persons-month (95% CI)	P Log-Rank test
All patients	21/38 (55.3 %)		1.895 (1.236 - 2.907)	
Sex: Male	4/12 (33.3 %)	0.16†	1.031 (0.387 - 2.747)	0.27
Female	15/26 (57.7 %)		2.137 (1.288 - 3.544)	
Age at presentation: < 3 months	6/17 (35.3 %)	0.10†	1.382 (0.621 - 3.077)	0.38
≥ 3 months	12/19 (63.2 %)		2.19 (1.244 - 3.856)	
Age at first seizures: < 4 months	7/15 (46.7 %)	0.85†	1.877 (0.895 - 3.937)	0.54
≥ 4 months	8/16 (50 %)		1.509 (0.755 - 3.018)	
N. of presenting signs: 1	6/15 (40 %)	0.38‡	1.307 (0.587 - 2.909)	0.19
2	11/20 (55 %)		2 (1.108 - 3.611)	
3-4	4/5 (80 %)		4.04 (0.516 - 0.765)	
Congenital microcephaly: yes	10/12 (83.3 %)	**0.011**†	4.032 (2.17 - 7.494)	**0.004**
no	11/28 (39.3 %)		1.279 (0.708 - 2.3)	
Developmental delay: yes	12/19 (63.2 %)	0.20†	2.065 (1.173 - 3.637)	0.71
no	9/21 (42.9 %)		1.708 (0.889 - 3.282)	
Seizures: yes	14/27 (51.9 %)	1.00‡	1.889 (1.119 - 3.19)	0.95
no	7/13 (53.8 %)		1.907 (0.909 - 4)	
Hypotonia: yes	2/6 (33.3 %)	0.40‡	2.439 (0.61 - 9.752)	0.60
no	19/34 (55.9 %)		1.851 (1.181 - 2.903)	
Small for gestational age: yes	3/7 (42.9 %)	0.68‡	1.714 (0.553 - 5.315)	0.95
no	15/27 (55.6 %)		1.923 (1.159 - 3.19)	
Spasticity: yes	12/21 (57.1 %)	0.10†	2.194 (1.246 - 3.863)	0.13
no	4/14 (28.6 %)		0.926 (0.348 - 2.467)	
Ocular involvement: yes	17/29 (58.6 %)	0.63‡	1.959 (1.218 - 3.151)	0.96
no	2/6 (33.3 %)		1.835 (0.459 - 7.336)	
Cardiac involvement: yes	10/10 (100 %)	**0.001**‡	4.049 (2.178 - 7.525)	**0.004**
no	9/25 (36 %)		1.155 (0.601 - 2.22)	

†*P* Fisher's Exact test, ‡*P* Pearson's Chi-square test, *95% CI* 95% Confidence Interval of the Incidence Rate.

Among these, congenital microcephaly and cardiac disorders were significantly associated with poor disease outcome (p = 0.004) (Figure 2). In fact, 10/12 (83%) patients with congenital microcephaly prematurely deceased versus 11/28 (39.3%) subjects with normal occipitofrontal circumference (OFC) at birth. The Incidence Rate (IR) of subjects with congenital microcephaly was 4.032 (95% CI = 2.17 - 7.494) per 100 person-months, while it was 1.279 (95% CI = 0.708 - 2.31) in those with normal OFC at birth. The Hazard Ratio (HR) was 3.427 (95% CI = 1.402 - 8.373). This supports a positive relationship (p = 0.004) between the presence of this clinical

feature and premature death. Similarly, early lethality was observed in 10/10 (100%) subjects with some type of cardiac involvement versus 9/25 (36 %) lacking cardiac abnormalities.

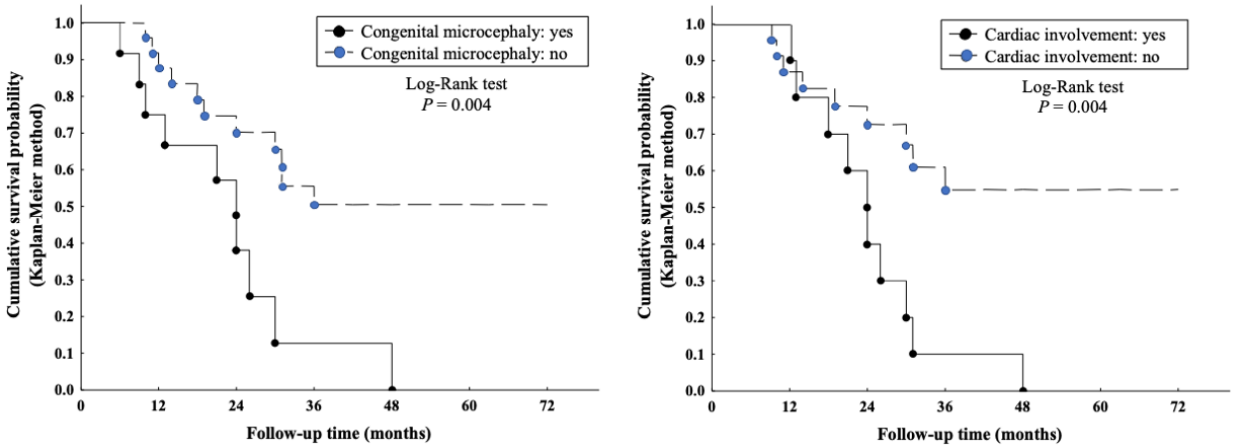


Figure 2. Outcome predictors in developmental and epileptic encephalopathy 35. Survival curves according to the Kaplan–Meier method to the presence/absence of congenital microcephaly and cardiac involvement. Congenital microcephaly and cardiac involvement are independent clinical prognostic factors of poor outcome (p = .004)

The IR among subjects with cardiac manifestations was 4.049; (95% CI = 2.178 - 7.525) versus 1.155 (95% CI = 0.601 - 2.220) in those lacking cardiac involvement. The HR was 3.509 (95% CI = 1.394 -8.834). These findings are suggestive of a higher mortality rate among affected individuals with cardiac disorders (p = 0.004). No other statistically significant associations were detected for the remaining variables (Table 4).

4. Discussion

This is the largest study involving subjects with DEE 35 harboring (likely) pathogenic variants in *ITPA*, allowing to delineate the molecular spectrum, determine the clinical phenotype, identify the neuroimaging patterns, and eventually establish prognostic factors in this rare and severe neurological condition.

4.1 ITPase deficiency

ITPase is a pyrophosphohydrolase catalyzing the conversion of noncanonical purines (NCPs) into the corresponding nucleoside monophosphate (*Simone et al., 2013*). The 45-kDa enzyme has a homodimeric structure composed of two globular 194-aminoacid α/β structural elements supported by a central elongated mixed β-sheet (*Behmanesh et al., 2009; Holmes et al., 1979*). The ITP-binding cleft is located between the dimerization and the N-terminal lobes, whereas the Mg^{2+}-dependent and pH-sensitive catalytic activity lie in a region of the N-terminal lobe close to the dimer interface (*Behmanesh et al., 2009; Holmes et al., 1979*). NCPs may spontaneously originate from the purine biosynthesis pathway or the deamination of nucleosides and nucleotides containing adenine or guanine (*Abolhassani et al., 2010; Galperin et al., 2006*). The increased sensitivity to NCPs caused by ITPase deficiency may lead to delayed cell cycle progression, increased mutation rate, and DNA damage, supporting a pivotal role of ITPase-mediated genomic instability limitation in cellular homeostasis (*Abolhassani et al., 2010; Holmes et al., 1979*). An involvement of ITPase in immunity and drug metabolism has also been reported (*Nakauchi et al., 2016; Shipkova et al., 2011*).

ITPA is highly expressed in the central nervous system (especially in neurons) and the heart (*Holmes et al., 1979*). The accumulation of NCPs resulting from ITPase deficiency can cause direct cellular toxicity, eventually leading to neuronal apoptosis (*Kevelam et al., 2015*). Additionally, the NPCs excess may negatively affect the function of enzymes utilizing adenosine triphosphate (ATP) or guanosine triphosphate (GTP), representing an indirect mechanism of neuronal toxicity. Indeed, the alteration of G-protein signal transduction leads to the inappropriate regulation of critical neuronal processes (e.g., neurotransmitter release, neuronal plasticity, and glucose metabolism) (*Kevelam et al., 2015*). Accordingly, ITPase deficiency

resulting from biallelic *ITPA* pathogenic variants cause a severe DEE with progressive disease course and neuroimaging abnormalities.

4.2 Spectrum of *ITPA* variants

Hitherto, 24 distinct *ITPA* variants are known to be associated with DEE 35, including 19 single nucleotide variants (SNVs), 3 intragenic deletions, 1 intragenic duplication, and 1 whole gene deletion within a larger chromosomal rearrangement. In our cohort, 4 out of the 8 previously reported pathogenic variants were detected: c.264-1G>A; p.(Ile88Metfs*59) (*Sakamoto et al., 2020*), c.359_366dup; p.(Gly123Serfs*104) (*Kevelam et al., 2015*), c.452G>A;p.Trp151* (*Kevelam et al., 2015*), c.489-1G>A (*Sakamoto et al., 2020*). We additionally identified 14 novel SNVs, all rare (allele frequency < 0.001), affecting conserved residues (GERP score range 5.32 – 5.82), and predicted damaging by *in silico* tools (CADD score range 26.4 – 46, ΔΔG all greater than 5 kcal/mol) (Table 1). Splicing and frameshift variants likely lead to truncated transcripts or nonsense-mediated decay (NMD). Missense variants are predicted to alter the structure of the ITP-binding cleft, interfere with the Mg^{2+}-dependent catalytic activity, and/or impair dimerization, eventually leading to a deficient enzymatic function. A loss of function mechanism is also expected in patients harboring partial or whole gene deletions. This is in line with what is usually observed in several epileptic disorders due to underlying metabolic deficiency and offers the possibility of an etiology-specific treatment (*Assi et al., 2017; Rahman et al., 2013; Sharma et al., 2017*).

We detected a multiexon deletion (P27) involving exons 1-5 of *ITPA* and a whole gene deletion (P2 and P3) in the context of a 1.1-Mb deletion, additionally involving *AVP*, *DDRGK1*, *PANK2*, and *SLC4A11*. Biallelic variants in *DDRGK1*, *PANK2*, and *SLC4A11* cause variable

clinical conditions, whereas *AVP* (MIM# 192340) and *SLC4A11* (MIM# 610206) haploinsufficiency is associated with autosomal dominant neuro-hypophyseal diabetes insipidus (MIM# 125700) and corneal dystrophy (Fuchs endothelial, type 4; MIM# 613268) (*Christensen et al., 2004; Vithana et al., 2008*). However, the two subjects harboring the deletion did not display features suggestive of these conditions. In line with these observations, all the detected *ITPA* are predicted to result in ITPase deficiency, supporting an underlying loss-of-function pathogenic model in DEE 35.

4.3 Phenotypic spectrum of DEE 35

Affected individuals present with a severe DEE, in which the underlying metabolic defect is responsible for the absence of development and the uncontrolled epileptic activity additionally contributes to worsen the cognitive impairment (*McTague et al., 2016; Kevelam et al., 2015*). Although only a portion of patients present with DD prior to seizure onset (48%), microcephaly (congenital, 30%), and seizures (67.5%) in the earlies stages of the disorder (Table 2), most will develop the neurological hallmarks of DEE as the disease progresses. The cardinal clinical features of DEE 35, reported in > 90% of cases, include progressive microcephaly (100%), epilepsy (100%), developmental stagnation after seizure onset (97.4%), progressive hypotonia (97.3%), and spasticity (60%). Interestingly, the perinatal period and birth weight are normal in most subjects, whereas neurological involvement and failure to thrive become evident in the first few months of life. Extra-neurological manifestations are particularly relevant in, as emerged from the systematic analysis of new cases and the review of previously published patients. A significant subset of subjects presents with ocular and cardiac involvement, which should be considered in all respects as part of the core clinical phenotype and assessed in all cases.

Ophthalmic manifestations are present in a large number of affected individuals (83%), including cataract, visual impairment, optic atrophy, and retinal cone dysplasia. Although cardiac disorders are less common (29%), when present they suggest an unfavorable outcome (Figure 2) and primarily contribute to the increased likelihood of early lethality (53% of cases). Accordingly, cardiac involvement was absent in the subjects who survived at 48-60 months (P1, P4, P5, P6, P13, P15, P19, and P21), although these patients did not show any peculiar genetic or clinical feature as compared to the rest of the cohort (Table S1).

4.4 Epileptic phenotype

Epilepsy is the cardinal feature of DEE 35, being observed in all affected individuals (Table 2). The epileptic phenotype mainly consists of focal and multifocal clonic/myoclonic seizures, generalized tonic seizures, and generalized tonic-clonic seizures. Patients suffer from both afebrile and, less often, febrile seizures. Status epilepticus may also occur (such as in P4 and P9). EEG usually shows focal or multifocal epileptiform discharges in the context of a slowing and disorganization of the background cerebral activity, consistent with the underlying encephalopathy. Several different ASMs have been employed alone or in combination (e.g., clonazepam, levetiracetam, topiramate, phenytoin, phenobarbitone, vigabatrin, and clobazam), but the response to therapy is usually poor, and refractory seizures occur in more than two-thirds of cases.

Ketogenic diet (KD), a low-carbohydrate dietary regimen reducing neuronal excitability, has proven to be effective in the management of refractory epilepsy in children with DEEs (*Jagadish et al., 2019; Martin et al., 2016; Wells et al., 2020*). More specifically, through several mechanisms working in concert (anti-inflammatory activity, epigenetic function, restoration of

bioenergetics, synaptic dysfunction, and impaired redox homeostasis) KD has been successfully used in many epileptic disorders caused by an underlying metabolic deficiency (*Gavrilovici & Rho, 2021; Lin Lin Lee, et al., 2018*). Among these, KD proved beneficial in metabolic epilepsies involving dysfunctional energy utilization (e.g., glucose transporter type 1 deficiency syndrome (GLUT1-DS) and pyruvate dehydrogenase complex deficiency (PDHCD)) or abnormal neurotransmitter degradation (succinic semialdehyde dehydrogenase deficiency (SSADHD) and non-ketotic hyperglycinemia (NKH)), as well as in mitochondrial epilepsies (e.g., *POLG*-related disorders and Leigh syndrome) (*Gavrilovici & Rho, 2021; Lin Lin Lee, et al., 2018*). In our cohort, KD was administered in four cases (P2, P4, P9, and P12). Although this approach was apparently ineffective in P2, a better seizure control (decreased seizure frequency) was temporarily achieved in P9 and P12, whereas the efficacy is still under investigation in P4. Interestingly, KD may directly increase ATP and adenosine levels, both recognized as crucial modulators of epileptogenic activity (*Boison, 2017; Gavrilovici & Rho, 2021; Masino et al., 2010*). In principle, this effect might prove useful in the biochemical context of ITPase deficiency (*Burgis, 2016; Masino et al., 2010*). However, further dedicated studies are necessary to confirm this hypothesis and investigate whether KD may be considered a potential treatment option in DEE35.

4.5 Neuroimaging spectrum

In the very early stages of the disease, MRI may not reveal abnormalities. The most typical feature of DEE 35, which appears after a few months, is a narrow, short segment of T2 hyperintensity in the PLIC, which disappears after several months (Table 3). Over time, delayed myelination becomes progressively more evident. DWI is most useful, revealing the involvement

of structures that are not T2 hyperintense. The involved structures are those that typically myelinate early. Only in older patients, diffusion restriction in the anterior limb of the internal capsule and cerebral hemispheric white matter is observed, structures that myelinate later than the PLIC, brain stem, cerebellar white matter, and optic radiation. The globus pallidus and thalamus are the gray matter structures with the highest myelin content and they are the only gray matter structures showing abnormalities. Atrophy is a relatively late finding and increases over time. Cerebral atrophy is earlier and more pronounced than cerebellar atrophy. These peculiar aspects are suggestive of a primary neuronal degeneration associated with Wallerian degeneration of white matter tracts, active Wallerian degeneration being characterized by diffusion restriction (*Kevelam et al., 2015*). Since early-myelinating white matter structures are the first to become functionally active, they are precociously involved in the course of the disease (*Kevelam et al., 2015*). Although this sequence of MRI characteristics is characteristic of DEE 35, the first MRI abnormalities may occur after the clinical presentation and therefore MRI may initially not be informative. If present, the typical lesions in the posterior limb of the internal capsule and the pattern of structures with diffusion restriction are suggestive for DEE 35 diagnosis and may help differentiate this condition from *RABGAP2*-related phenotypes (e.g., diffuse brain atrophy, frontotemporal polymicrogyria, corpus callosum hypoplasia) (*Sakamoto et al., 2020*).

4.6 Cardiac involvement in patients with pathogenic *ITPA* variants

The excess of inosine triphosphate (ITP) in cardiac sarcomeres favors the abnormal actomyosin binding of ITP instead of ATP and the accumulation of deoxyinosine monophosphate (dIMP) in nucleic acids (*Behmanesh et al., 2009; Burton et al., 2005*). These events eventually lead to

cardiac toxicity as a result of disorganization of the sarcomeric structure in the developing heart, increased DNA damage, nonfunctional RNAs, delayed cell cycle progression, and impaired cardiac protein function (*Behmanesh et al., 2009; Burton et al., 2005*). Cardiac involvement is of particular interest in DEE 35 individuals. Lethal infantile-onset dilated cardiomyopathy was recently reported in two subjects with Martsolf-like syndrome harboring homozygous null *ITPA* variants (*Handley et al., 2019*). In this study, cardiomyopathy (5/10 subjects), rhythm disturbances (4/10 subjects), or both (1/10 subjects) were observed in 29% of patients (Table 2), making cardiac involvement a key clinical issue in ITPase deficiency. Although there is still limited direct evidence to unveil the potential mechanisms underlying cardiac dysfunction in human subjects with ITPase deficiency, it is tempting to speculate that the restoration of the enzymatic activity might positively impact on cardiac function in these patients (*Burgis, 2016; Handley et al., 2019*).

5. Conclusion

Early diagnosis and timely antiepileptic treatment may favorably impact the management and developmental outcomes of patients with DEEs. Obtaining a genetic diagnosis is especially crucial for parental counselling and beneficial in terms of knowledge on the natural course of the specific disorder. We reported a comprehensive analysis of 40 patients with DEE 35, expanding and refining the molecular and phenotypic spectrum of this severe condition. In addition to a severe, progressive encephalopathy which typically presents during the first months of life, affected individuals show a high prevalence of ocular and cardiac manifestations, and an increased risk of premature death. Congenital microcephaly and cardiac involvement are

independent clinical predictors of poor outcome. Taken together, these findings may have a large impact on diagnosis, counselling, and follow-up of subjects with DEE 35.

References

Abolhassani N., et al. (2010) NUDT16 and ITPA play a dual protective role in maintaining chromosome stability and cell growth by eliminating dIDP/IDP and dITP/ITP from nucleotide pools in mammals. *Nucleic Acids Res, 38,* 2891-2903. https://doi.org/10.1093/nar/gkp1250

Assi L., et al. (2017) Treatable Genetic Metabolic Epilepsies. *Curr Treat Options Neurol, 19,* 30. https://doi.org/10.1007/s11940-017-0467-0

Behmanesh M., et al. (2009) ITPase-deficient mice show growth retardation and die before weaning. *Cell Death Differ, 16,* 1315-1322. https://doi.org/10.1038/cdd.2009.53

Bierau J., et al. (2007) Pharmacogenetic significance of inosine triphosphatase. *Pharmacogenomics, 8,* 1221-1228. https://doi.org/10.2217/14622416.8.9.1221

Boison D. (2017) New insights into the mechanisms of the ketogenic diet. *Curr Opin Neurol, 30,* 187-192. https://doi.org/10.1097/WCO.0000000000000432

Burgess R., et al. (2019) The Genetic Landscape of Epilepsy of Infancy with Migrating Focal Seizures. *Ann Neurol, 86,* 821-831. https://doi.org/10.1002/ana.25619

Burgis N.E. (2016) A disease spectrum for ITPA variation: advances in biochemical and clinical research. *J Biomed Sci, 23,* 73. https://doi.org/10.1186/s12929-016-0291-y

Burton K., et al. (2005) Kinetics of muscle contraction and actomyosin NTP hydrolysis from rabbit using a series of metal-nucleotide substrates. *J Physiol, 563,* 689-711. https://doi.org/10.1113/jphysiol.2004.078907

Chaudhury S., et al. (2010) PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics, 26,* 689-691. https://doi.org/10.1093/bioinformatics/btq007

Christensen J.H., et al. (2004) Six novel mutations in the arginine vasopressin gene in 15 kindreds with autosomal dominant familial neurohypophyseal diabetes insipidus give further insight into the pathogenesis. *Eur J Hum Genet, 12,* 44-51. https://doi.org/10.1038/sj.ejhg.5201086

Galperin M.Y., et al. (2006) House cleaning, a part of good housekeeping. *Mol Microbiol, 59,* 5-19. https://doi.org/10.1111/j.1365-2958.2005.04950.x

Gavrilovici C., et al. (2021) Metabolic epilepsies amenable to ketogenic therapies: Indications, contraindications, and underlying mechanisms. *J Inherit Metab Dis, 44,* 42-53. https://doi.org/10.1002/jimd.12283

Handley M.T., et al. (2019) ITPase deficiency causes a Martsolf-like syndrome with a lethal infantile dilated cardiomyopathy. *PLoS Genet*, 15:e1007605. https://doi.org/10.1371/journal.pgen.1007605

Holmes S.L., et al. (1979) Human inosine triphosphatase: catalytic properties and population studies. *Clin Chim Acta, 97,* 143-153. https://doi.org/10.1016/0009-8981(79)90410-8

Jagadish S., et al. (2019) The Ketogenic and Modified Atkins Diet Therapy for Children With Refractory Epilepsy of Genetic Etiology. *Pediatr Neurol, 94,* 32-37. https://doi.org/10.1016/j.pediatrneurol.2018.12.012

Karczewski K.J., et al. (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature, 581,* 434-443. https://doi.org/10.1038/s41586-020-2308-7

Kaur P., et al. (2019) Identification of a novel homozygous variant confirms ITPA as a developmental and epileptic encephalopathy gene. *Am J Med Genet A, 179,* 857-861. https://doi.org/10.1002/ajmg.a.61103

Kevelam S.H., et al. (2015) Recessive ITPA mutations cause an early infantile encephalopathy. *Ann Neurol, 78,* 649-658. https://doi.org/10.1002/ana.24496

Lek M., et al. Exome Aggregation Consortium. (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536, 285-291. *https://doi.org/10.1038/nature19057*

Lin Lin Lee V., et al. (2018) Treatment, Therapy and Management of Metabolic Epilepsy: A Systematic Review. *Int J Mol Sci, 19,* 871. https://doi.org/10.3390/ijms19030871

Martin K., et al. (2016) Ketogenic diet and other dietary treatments for epilepsy. *Cochrane Database Syst Rev, 2*, CD001903. https://doi.org/10.1002/14651858.CD001903.pub3

Masino S.A., et al. (2010) Purines and the Anti-Epileptic Actions of Ketogenic Diets. *Open Neurosci J, 4,* 58-63. https://doi.org/10.2174/1874082001004010058

McTague A., et al. (2016) The genetic landscape of the epileptic encephalopathies of infancy and childhood. *Lancet Neurol, 15,* 304-316. https://doi.org/10.1016/S1474-4422(15)00250-1

Muthusamy K., et al. (2021) Neuroimaging Findings in Inosine Triphosphate Pyrophosphohydrolase (ITPase) Deficiency. *Neurology, 97(1),* e109-e110. https://doi.org/10.1212/WNL.0000000000011719

Nakauchi A., et al. (2016) Identification of ITPA on chromosome 20 as a susceptibility gene for young-onset tuberculosis. *Hum Genome Var, 3,* 15067. https://doi.org/10.1038/hgv.2015.67

Rahman S., et al. (2013) Inborn errors of metabolism causing epilepsy. *Dev Med Child Neurol, 55,* 23-36. https://doi.org/10.1111/j.1469-8749.2012.04406.x

Richards S., 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,* 405-424. https://doi.org/10.1038/gim.2015.30

Rochtus A., et al. (2020) Genetic diagnoses in epilepsy: The impact of dynamic exome analysis in a pediatric cohort. *Epilepsia, 61,* 249-258. https://doi.org/10.1111/epi.16427

Sakamoto M., et al. (2020) A novel ITPA variant causes epileptic encephalopathy with multiple-organ dysfunction. *J Hum Genet, 65,* 751-757. https://doi.org/10.1038/s10038-020-0765-3

Sharma S., et al. (2017) Inborn Errors of Metabolism and Epilepsy: Current Understanding, Diagnosis, and Treatment Approaches. *Int J Mol Sci, 18,* 1384. https://doi.org/10.3390/ijms18071384

Shaw-Smith C., et al. (2004) Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning

disability/mental retardation and dysmorphic features. *J Med Genet, 41,* 241-248. https://doi.org/10.1136/jmg.2003.017731

Shipkova M., et al. (2011) Association between adverse effects under azathioprine therapy and inosine triphosphate pyrophosphatase activity in patients with chronic inflammatory bowel disease. *Ther Drug Monit, 33,* 321-328. https://doi.org/10.1097/FTD.0b013e31821a7c34

Simone P.D., et al. (2013) ITPA (inosine triphosphate pyrophosphatase): from surveillance of nucleotide pools to human disease and pharmacogenetics. *Mutat Res, 753,* 131-146. https://doi.org/10.1016/j.mrrev.2013.08.001

Sobreira, N., et al. (2015) GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat, 36,* 928-930. https://doi.org/10.1002/humu.22844

Van der Knaap M.S., et al. (1999) Defining and categorizing leukoencephalopathies of unknown origin: MR imaging approach. *Radiology, 213,* 121-133. https://doi.org/10.1148/radiology.213.1.r99se01121

Vithana E.N., et al. (2008) SLC4A11 mutations in Fuchs endothelial corneal dystrophy. *Hum Mol Genet, 17,* 656-666. https://doi.org/10.1093/hmg/ddm337

Wells J., et al. (2020) Efficacy and Safety of a Ketogenic Diet in Children and Adolescents with Refractory Epilepsy-A Review. *Nutrients, 12,* 1809. https://doi.org/10.3390/nu12061809

D. Case series study: genotype-phenotype correlations

De novo truncating *NOVA2* variants affect alternative splicing and lead to heterogeneous neurodevelopmental phenotypes

Abstract

Alternative splicing (AS) is crucial for cell-type-specific gene transcription and plays a critical role in neuronal differentiation and synaptic plasticity. *De novo* frameshift variants in NOVA2, encoding a neuron-specific key splicing factor, have been recently associated with a new neurodevelopmental disorder (NDD) with hypotonia, neurological features, and brain abnormalities. We investigated eight unrelated individuals by exome sequencing (ES) and identified seven novel pathogenic *NOVA2* variants, including two with a novel localization at the KH1 and KH3 domains. In addition to a severe NDD phenotype, novel clinical features included psychomotor regression, attention deficit-hyperactivity disorder (ADHD), dyspraxia, and urogenital and endocrinological manifestations. To test the effect of the variants on splicing regulation, we transfected HeLa cells with wildtype and mutant NOVA2 complementary DNA (cDNA). The novel variants NM_002516.4:c.754_756delCTGinsTT p.(Leu252Phefs*144) and c.1329dup p.(Lys444Glnfs*82) all negatively affected AS events. The distal p.(Lys444Glnfs*82) variant, causing a partial removal of the KH3 domain, had a milder functional effect leading to an intermediate phenotype. Our findings expand the molecular and phenotypic spectrum of NOVA2-related NDD, supporting the pathogenic role of AS disruption by truncating variants and suggesting that this is a heterogeneous condition with variable clinical course.

1. Introduction

Proteins involved in the alternative splicing (AS) of genes encoding ubiquitously-expressed transcriptional regulators have emerged as crucial regulators of cell-type-specific transcription, especially in neurons (*Lipscombe & Lopez Soto, 2019; Porter et al., 2018; Vuong et al., 2016*). Through the inclusion or exclusion of specific alternative exons, AS allows the generation of distinct proteins from a single pre- messenger RNA (mRNA), contributing to cell-restricted transcrip- tional regulation (*Lee & Rio, 2015; Park et al., 2018*). Complex and unique AS patterns occur in neuronal cells, in which AS is essential for every stage of the life cycle and plays a pivotal role in early differentiation, axonal guidance, synapse formation and plasticity, and even programmed cell death (*Lipscombe & Lopez Soto, 2019; Porter et al., 2018; Traunmüller et al., 2016; Vuong et al., 2016; Weyn-Vanhentenryck et al., 2018*). Neuron-specific AS relies on the coordinate actions of multiple brain-specific RNA-binding proteins (RBPs), whose deficient regulation is implicated in the pathogenesis of several neurological disorders (*Porter et al., 2018; Vuong et al., 2016; Wilkinson et al., 2020; Will & Lührmann, 2011*).

Among neuron-specific key splicing factors, Neuro-Oncological Ventral Antigens (NOVAs) dynamically regulate alternative polyadenylation in the brain and the exon content of RNAs encoding crucial proteins for synaptic plasticity (*Licatalosi et al., 2008; Ule et al., 2003, 2006, 2005*). *NOVA1* (OMIM *602157) and *NOVA2* (OMIM *601991) paralogues encode two highly homologous proteins with three K homology (KH)-type RNA-binding domains (KH1-3), through which they bind the YCAY motifs in the mRNA (*Jensen et al., 2000*). NOVA1 is preferentially expressed in the hindbrain and ventral spinal cord, whereas NOVA2 expression is predominant in the forebrain and dorsal spinal cord (*Saito et al., 2016; Vuong et al., 2016; Yang*

et al., 1998). The pivotal role of NOVAs in the development of peripheral and central nervous system is highlighted by the knock-out mouse models, showing motor impairment, neuronal apoptosis, long- term potentiation deficiency, and early lethality (*Huang et al., 2005; Jensen et al., 2000; Ruggiu et al., 2009*).

NOVA2 is specifically implicated in the splicing regulation of genes involved at different levels in brain development and function (axonal guidance and projection, synaptic formation and plasticity, and Purkinje cells function) (*Saito et al., 2019; Vuong et al., 2016*). In humans, NOVA2-related neurodevelopmental disorder (NDD) results from *de novo* frameshift variants clustered between Ala 241 and Val261 and replacing the KH3 domain by the same alternative C-terminal part (*Mattioli et al., 2020*). This condition is characterized by psychomotor delay, cognitive impairment, hypotonia, neurologic features, and brain MRI abnormalities (NDD with or without autistic features and/or structural brain abnormalities—NEDASB, OMIM #618859) (*Mattioli et al., 2020*). Based on the clustering and type of the reported variants, mutational mechanism was suspected to be either a hypomorphic or gain-of-function effect (*Mattioli et al., 2020*).

Here, we report eight new patients harboring novel truncating variants in *NOVA2*. In addition to four frameshifts clustered in the same protein region previously described (p.(Leu252Phefs*144), p.(Leu252Profs*141), p.(Ala263Profs*133), and p.(Leu276-Cysfs*120)), we also identified three novel variants (p.(Gln86*), p.(Leu175Cysfs*6), and p.(Lys444Glnfs*82)) located in KH1, KH2, or KH3 domains. Our data refine the molecular and phenotypic spectrum of *NOVA2* variants and suggest a novel interpretation of disease pathogenicity.

2. Materials and methods

2.1 Ethical considerations

This study adheres to the principles set out in the Declaration of Helsinki. The following Research Ethics Committees approved the study: Gaslini Children's Hospital (Comitato Etico della Regione Liguria (163/2018) and Città della Salute e della Scienza University Hospital (0060884). No institutional review board (IRB) approval was necessary for retrospective data analysis of a single patient for the following Institutions: Center for Pediatrics and Faculty of Medicine, University of Freiburg, Freiburg (Germany); Erasmus Medical Centre, Rotterdam (The Netherlands); Medical University of Gdańsk, Gdańsk (Poland); University of Lakki Marwat, KPK (Pakistan) University of Management and Technology, Lahore, Punjab (Pakistan); Women's and Children's Hospital, Adelaide (Australia). The authors obtained and archived written informed consents from parents or legal guardians of the enrolled subjects to publish genetic and clinical data, including clinical photographs (#2 and #4) and brain magnetic resonance imaging (MRI) images (#2 and #6).

2.2 Subject enrolment and phenotyping

Following the identification through exome sequencing (ES) of a *de novo* truncating variant in NOVA2 in a patient with psychomotor delay, behavioral disturbances, and sleep disorders, we collected data from additional individuals harboring *de novo NOVA2* variants through GeneMatcher (Sobreira et al. 2015). Patients were recruited from several clinical and research centres in Australia (Women's & Children's Hospital, South Australia), Europe (Center for Pediatrics and Faculty of Medicine, University of Freiburg, Germany; Erasmus Medical Centre, The Netherlands; Gaslini Children's Hospital, Italy; Medical University of Gdańsk, Poland;

University of Turin, Italy), and Middle East (King Abdullah International Medical Research Center, Saudi Arabia). Written informed consent was obtained from the parents or legal guardians of all enrolled subjects. Patient data were anonymized before sharing.

Detailed phenotypic information concerning the developmental history, behavioral disturbances, neurological examinations, and electro-clinical findings were provided by the referring physicians. Brain magnetic resonance imaging (MRI) scans were locally per- formed during routine patient care. All articles indexed in PubMed (https://pubmed.ncbi.nlm.nih.gov/?term=itpa) between April 2020, when frameshift NOVA2 variants were first associated with a human neurodevelopmental condition by Mattioli et al., and July 2021 were retrieved using the terms "NOVA2," "Frameshift variants," and "Neurodevelopmental disorders." Previously reported subjects (Mattioli et al., 2020) were critically reviewed in terms of molecular, clinical, and neuroradiological spectrum, and compared with the current cohort.

2.3 Variant identification and analysis

ES was performed on genomic DNA extracted from peripheral blood leukocytes using standard local protocols. The identified variants were filtered according to minor allele frequency ≤ 0.001 in the Genome Aggregation Database (gnomAD; https://gnomad.broadinstitute.org) (*Lek et al., 2016*), presence in ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), conservation (Genomic Evolutionary Rate Profiling—GERP, http://mendel. stanford.edu/SidowLab/downloads/gerp/) (Cooper et al., 2005), and predicted impact on protein structure and function. *In silico* tools were employed to predict the pathogenicity of candidate variants using the Ensembl Variant Effect Predictor (VEP) pipeline (https:// www.ensembl.org/info/docs/tools/vep/index.html), including

Com- bined Annotation Dependent Depletion (CADD, GRCh37-v1.6 version, https://cadd.gs.washington.edu), Sorting Intolerant From Tolerant (SIFT, https://sift.bii.a-star.edu.sg), and Polyphen-2 (http:// genetics.bwh.harvard.edu/pph2/) (*Adzhubei et al., 2010; Ng, 2003; Rentzsch et al., 2019; Schwarz et al., 2014*). Candidate variants were eventually classified according to the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG/AMP) guidelines (*Richards et al., 2015*). Sanger sequencing was performed for the validation of the detected variants and the segregation analysis. All NOVA2 variants are reported according to RefSeq NM_002516.4 (GenBank NC_000019.10), using HGVS recommendations (*den Dunnen et al. 2016*). All variants were submitted to the Leiden Open Variation Database (LOVD, https://www.lovd.nl) with the following accession numbers: #0000797459, #0000797460, #0000797461, #0000797462, #0000797463, #0000 797464, and #0000841619.

2.4 *In vitro* assay to test *NOVA2* variants effect

HeLa cells were transfected as previously described with pcDNA3.1 plasmids containing optimized sequences of human wild-type and mutant NOVA2 cDNA tagged with green fluorescent protein (GFP) and a plasmid containing a FLAG-protein (*Mattioli et al., 2020*). The mutant sequences include one variant reported by Mattioli et al. (p.Val261Glyfs*135, alias Mut1) or the following variants reported in this study: c.754_756delinsTT (p.Leu252Phefs*144) and c.1329dupC (p.Lys444Glnfs*82). Proteins were extracted and separated on a 10% acrylamide gel, visualized using an in-house mouse anti-GFP antibody (1:10,000) and normalized using a FLAG staining (FLAG antibody: 1:1,000; F1804, Sigma-Aldrich). To test the effect of variants on splicing regulation, mRNA was extracted from transfected HeLa cells and

reverse-transcribed as described (Mattioli et al., 2020). PCR was performed with primers used previously to amplify APLP2 Exon 12 (24 cycles), SORBS1 Exon 3 (37 cycles), and SGCE Exon 9 (26 cycles), and the PCR products were analyzed by migration on a 2100 Bioanalyzer instrument (Agilent Technologies).

3. Results

3.1 *NOVA2* variants cluster within or next to the KH domains

ES led to the identification of seven novel truncating variants in NOVA2 in our cohort (Figure 1). Three clustered in proximity to those previously reported and add a similar C-terminal tail: c.787delG p.(Ala263Profs*133) in subject #1, c.754_756delCTGinsTT p.(Leu252Phefs*144) in subject #2, c.826del p.(Leu276Cysfs*120) in subject #4, and c.755_764del p.(Leu252Profs*141) in subjects #7 and #8. We also found three variants located in different protein regions: in subject #3, the distal frameshift c.1329dupC p.(Lys444Glnfs*82) in the KH3 domain; in subject #5, the frameshift variant c.523delC p.(Leu175Cysfs*6) in the KH2 domain; in subject #6, the nonsense variant c.256C>T p.(Gln86*) in the KH1 domain (Figure 1). All these variants are absent in gnomAD database and predicted damaging by *in silico* tools (Table S1).

Table S1. *In silico* analysis of novel *NOVA2* variants.

#CHROM	POS	REF	ALT	annotations	EXON	HGVSc (ENST00000263257)	HGVSp (ENSP00000263257.4)	AF_gnomAD	CADD_PHRED	SpliceAI
19	45940013	G	GG	frameshift	4	c.1329dup	p.Lys444GlnfsTer82	0	NA	0.00
19	45940515	AG	A	frameshift	4	c.826del	p.Leu276CysfsTer120	0	NA	0.00
19	45940554	GC	G	frameshift	4	c.787del	p.Ala263ProfsTer133	0	NA	0.00
19	45940818	AG	A	frameshift	4	c.523del	p.Leu175CysfsTer6	0	NA	0.00
19	45953920	G	A	stop_gained	3	c.256C>T	p.Gln86Ter	0	39	0.00
19	45940585	GCAG	GAA	frameshift	4	c.754_756delinsTT	p.Leu252PhefsTer144	0	NA	0.00

No additional potentially causative variants were identified in the studied subjects. All variants are located in the huge last Exon 4 and are predicted to escape nonsense-mediated mRNA decay (NMD) except c.256C>T p.(Gln86*), located in Exon 3, which is likely to activate NMD leading to haploinsufficiency or to a truncated non- functional protein.

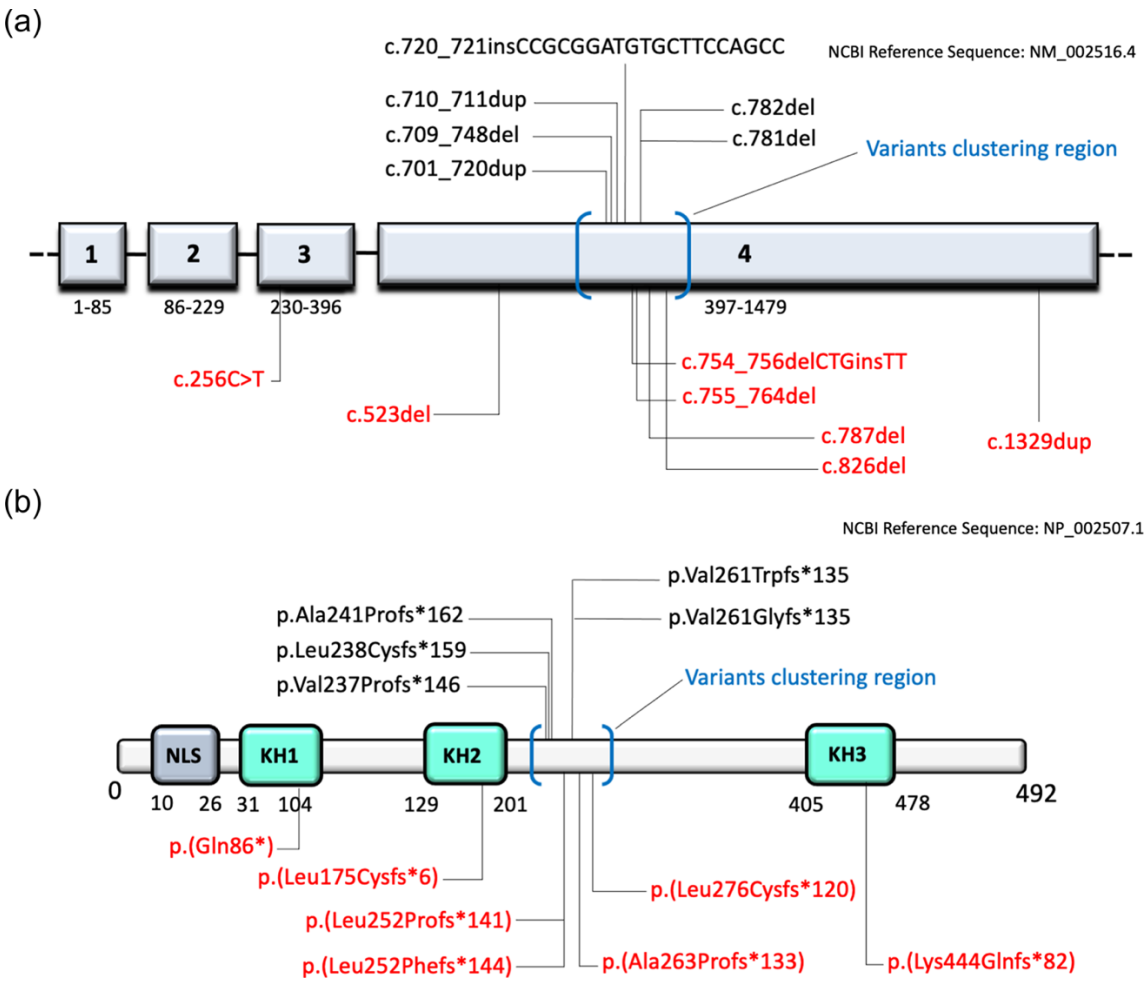


Figure 1 Distribution of *NOVA2* variants. Localization of truncating variants (a) across the exons of *NOVA2* (NCBI Reference Sequence: NM_002516.4; https://www.ncbi.nlm.nih.gov/nuccore/NM_002516.4) and (b) in relation to the K homology (KH)-type RNA-binding domains (KH1-3) domains of NOVA2 protein (NCBI Reference Sequence: NP_002507.1; https://www.ncbi.nlm.nih.gov/protein/NP_002507.1). Previously reported variants are indicated in black above the schematic representation of the gene and the protein, novel variants are reported in red below. The round brackets in blue indicate the potential NOVA2 "variants clustering region" based on the currently available information, with most of the variants falling in the exons 4 (a) and localizing just after the KH2 domain (b). NLS, nuclear localization signal.

3.2 *NOVA2* variants affect AS events

To check the consequences of the identified variants, we introduced them in NOVA2 complementary DNA (cDNA). We included in the analysis the previously reported c.782del p.(Val261Glyfs*135), the c.754_756delinsTT p.(Leu252Phefs*144) variant, which leads to a similar C-terminal tail, and the distal frameshift c.1329dupC p.(Lys444Glnfs*82), predicted to remove only part of the KH3 domain. When overexpressed in HeLa cells, no significant difference could be observed in the level of mutant NOVA2 proteins as compared with wild-type (WT) NOVA2 protein (Figure 2).

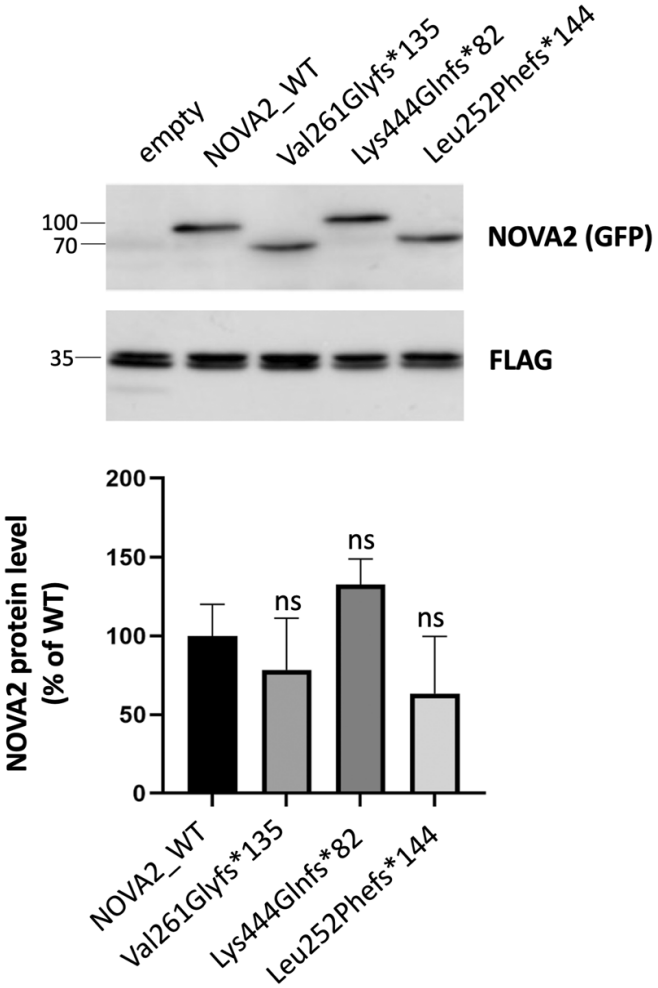


Figure 2. Expression of NOVA2 mutant proteins in HeLa cells. HeLa cells were cotransfected with EGFP-tagged NOVA2 wild-type (WT) or mutant cDNA and a plasmid with a FLAG-tagged protein as a transfection. Proteins were extracted 24 h after transfection, and expression of NOVA2 was analyzed (SDS–PAGE and immunoblotting) using anti-GFP and anti-FLAG antibodies. Quantification of NOVA2 protein level was performed from three independent experiments and normalized on FLAG-protein level. The error bars indicate the standard error mean (SEM). Kruskal–Wallis ANOVA with Dunn's correction for multiple comparisons was performed. ANOVA, analysis of variance; cDNA, complementary DNA; ns, nonsignificant; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis.

As previously described, the overexpression of WT NOVA2 cDNA into HeLa cells leads to significant changes in several AS events: an increase of Exon 9 skipping in SGCE transcripts

and an increase of Exon 3 and Exon 14 inclusion in SORBS1 and APLP2 transcripts, respectively. These changes are not observed when the NOVA2 mutant p.(Val261Glyfs*135) (Mut1) is overexpressed (Figure 3) (*Mattioli et al., 2020*). The NOVA2 mutant protein p.(Leu252Phefs*144) behaves similarly to Mut1 as it fails to regulate these splicing events. However, the overexpression of the NOVA2 p.(Lys444Glnfs*82) protein, carrying the more distal frameshift, leads to an intermediate phenotype between WT and Mut1 NOVA2 proteins. This finding suggests that the p.(Lys444Glnfs*82) variant might have a milder effect than the other frameshift variants located just after the KH2 domain.

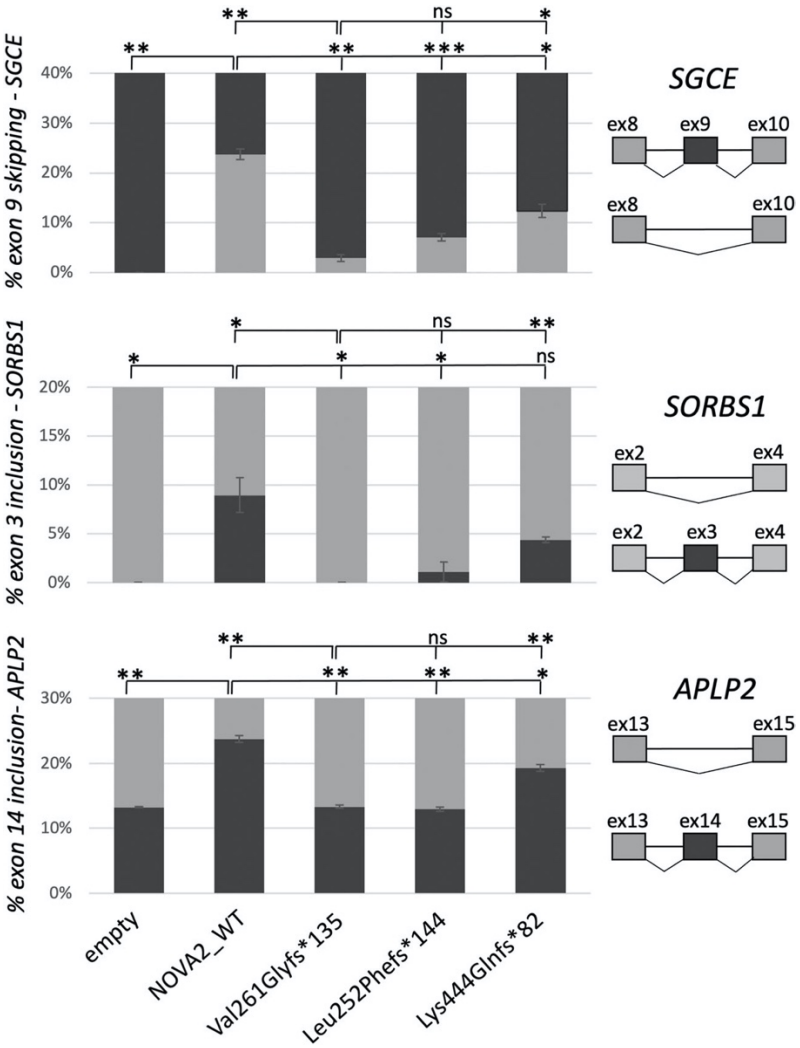


Figure 3. Effect of variants Leu252Phefs*144 and Lys444Glnfs*82 on alternative splicing events regulated by NOVA2. The pcDNA3 eGFP plasmids containing the following NOVA2 cDNA were transfected in HeLa cells: wild-type (WT), variant reported by Mattioli et al. (Val261Glyfs*135, alias Mut1), c.754_756delinsTT, p.Leu252Phefs*144 variant and c.1329dupC, p.Lys444Glnfs*82 variant. Effects of WT and variant NOVA2 overexpression on alternative splicing events (regulating inclusion of SGCE Exon 9, SORBS1 Exon 3, and APLP12 Exon 14) were analyzed by RT-PCR and migration on a 2100 Bioanalyzer instrument (Agilent Technology). Three series of experiments were analyzed. The error bars indicate the standard deviation. One-way ANOVA with Dunnett's multiple comparisons. ANOVA, analysis of variance; ns, nonsignificant; RT-PCR, reverse- transcription polymerase chain reaction.***p < .001, *p < .05.

3.3 *NOVA2* variants cause a heterogeneous NDD with variable clinical severity

The common neurodevelopmental phenotype observed in our cohort consisted of global psychomotor delay, consistently leading to moderate to severe intellectual disability, behavioral abnormalities, sleep disorders, and associated neurological features (Table 1). Neonatal course was unremarkable in most cases, but feeding problems due to swallowing difficulties were common (#1, #2, #6, #7, and #8). However, failure to thrive only occurred in one case (#6). In general, birth parameters were within normal ranges.

Table 1. Summary of genetic, clinical, and neuroimaging features in subjects harboring *de novo* truncating *NOVA2* variants

Subject ID Age, sex Country of origin	**#1** 15.5 y, F Italy	**#2** 7 y, F Australia	**#3** 7 y, M Netherlands	**#4** 10 y, F Poland	**#5** 13 y, M Italy	**#6** 15 y, F Saudi Arabia	**Mattioli et al. 2020†** Mean 6.5 y, 6 cases (M/F = 1) France, USA	**Total**
***NOVA2* variant** (NM_002516.4)	c.787del (p.Ala263Profs*133)	c.754_756delCTGinsTT (p.Leu252Phefs*144)	c.1329dup (p.Lys444Glnfs*82)‡	c.826del (p.Leu276Cysfs*120)	c.523del (p.Leu175Cysfs*6)‡	c.256C>T (p.Gln86*)‡	c.782del (p.Val261Glyfs*135); c.710_711dup (p.Leu238Cysfs*159); c.701_720dup (p.Ala241Profs*162); c.709_748del (p.Val237Profs*146) c.781del (p.Val261Trpfs*135); c.720_721insCCGCGGATGTGC TTCCAGCC (p.Ala241Profs*162)	
Birth parameters Low birth weight Congenital microcephaly	- -	- -	- NA	- -	- -	- -	+ (3), NA (1) NA (1)	3/12 (25%) 0/12 (0%)
Feeding difficulties	+ - -	+ - -	- - -	- - -	- + -	+ - +	+ (5) NA + (3)	8/12 (66.7%)

Poor sucking/swallowing difficulties GE reflux Failure to thrive								1/12 (8.3%) 4/12 (33.3%)
Psychomotor delay Motor delay Speech delay Bowel/urinary incontinence	+ + +	+ + -	+ + +	+ + +	+ + -	+ + -	+ (6) + (6) NA	12/12 (100%) 12/12 (100%) 3/12 (25%)
Intellectual disability (degree)	+ (severe)	+ (NA)	+ (severe)	+ (moderate)	+ (moderate)	+ (moderate)	+ (6)	12/12 (100%)
Psychomotor regression	-	-	-	+	-	+	-	2/12 (16.7%)
Progressive microcephaly	-	+	-	-	-	+	+ (1), NA (1)	3/12 (25%)
Abnormal behavior ASD ADHD Frequent laughter Attraction with water Stereotyped movements	- - - - +	+ - - + +	+ - + + -	+ - + + +	+ + - - +	- + - - -	+ (3) - + (2) + (1) + (5)	7/12 (58.3%) 2/12 (16.7%) 4/12 (33.3%) 4/12 (33.3%) 9/12 (75%)
Neurological features Hypotonia Ataxia/broad-based gait Spasticity Hyperreflexia Movement disorders Apraxia/dyspraxia	- + - + - -	- + + + - -	+ - - - - -	+ + - + + +	- - - + - +	+ + + + + -	+ (3), NA (2) + (3) + (1) + (2) - -	6/12 (50%) 7/12 (58.3%) 3/12 (25%) 7/12 (58.3%) 2/12 (16.7%) 2/12 (16.7%)
Sleep disorders	+	+	+	-	-	-	+ (1)	4/12 (33.3%)
Epilepsy Seizure onset Seizure type EEG Response to AEDs Seizure-free	- - Normal - -	- - Normal - -	- - NA - -	4.5 y MAS Generalized SWA Partial No	- - Normal - -	2 y MS Normal Partial No	+ (2) 2.5-9 y MAS, SS (1); MS (1) NA NA NA	4/12 (33.3%)
Syndromic features Facial dysmorphism Eye abnormalities Urogenital manifestations Endocrinological disorders	- - - +	+ + - -	+ - + -	+ - + +	- - - -	+ - - -	+(4), NA (1) + (1) - -	8/12 (66.7%) 2/12 (16.7%) 2/12 (16.7%) 2/12 (16.7%)
Brain MRI CCH Cortical atrophy WM abnormalities Cerebellar abnormalities Chiari 1	- - - - - -	 + - - - -	NA NA NA NA NA NA	- - - - - -	- - - - + -	+ - + - - -	+ (2) + (1) + (1) + (1) + (1) Pineal cyst (1)	4/12 (33.3%) 1/12 (8.3%) 2/12 (16.7%)

Other	+, prominent CSF spaces	1/12 (8.3%) 2/12 (16.7%) 2/12 (16.7%)

ADHD: attention deficit-hyperactivity disorder; AEDs: antiepileptic drugs; ASD: autism spectrum disorder; CCH: corpus callosum hypoplasia; CSF: cerebrospinal fluid; GE: gastroesophageal; MAS: myoclonic-atonic/astatic seizures; MS: myoclonic seizures; NA: not applicable; SS: staring spells; SWA: spike-and-wave activity; WM: white matter. † PMID: 32197073. ‡ Variants located outside the suggested 'clustering region'.

Progressive microcephaly was instead observed in subjects #2 and #6-8. Dysmorphic facial features were observed in #2, #3, #4, #6, and #7 (Figure 4a). A global developmental delay was noticed in the first year of life in all cases. Patient #7 was never able to walk, while the other subjects could walk, either assisted, or unassisted. Speech impairment was particularly significant in all cases, ranging from nonverbal (#2, #7, and #8) to a few words. Bowel and urinary incontinence were also observed in five cases (#1, #3, #4, #7, and #8). Behavioral abnormalities included autism spectrum disorder (ASD) (#2-5, #7, and #8), attention deficit-hyperactivity disorder (ADHD) (#5-8), frequent laughter (#3, #4, and #8), and attraction with water (#2-4). Most subjects showed variable stereotyped movements of the hands (#1, #2, #4, #5, and #7), especially in association with a state of arousal (Supporting Information: Video 1). These included hand flapping, wringing, and clapping. Additionally, body rocking and head banging were observed in patient #8. Sleep disturbances were also common, ranging from frequent awakenings (#2, #3, #7, and #8) to parasomnias.

In addition to the severe neurodevelopmental phenotype, psychomotor regression was observed in subjects #4 and #6. In particular, patient #4 experienced a loss of motor and verbal skills with increased epileptic activity, whereas no correlation with seizures could be observed in case #6. Interestingly, they also suffered from movement disorders, consisting of tremors (#4 and #6) and choreoathetosis (#4). Epileptic manifestations included refractory myoclonic or myoclonic-atonic seizures with or without postictal state. Seizures were occasionally precipitated

by sound stimuli, with an exaggerated startle response (startle seizures) (Supporting Information: Videos 1-3), and associated with generalized spike- and-wave activity in subject #4 (Figure 4b), who showed a partial response to antiseizure medications. Muscle tone abnormalities consisted of hypotonia (#3, #4, #6-8), spasticity (#2 and #6), or paratonia (#1). Five out of six subjects showed hyperreflexia (#1, #2, and #4-6). Lack of coordination with unsteady gait and, in more severe cases, true ataxia was present (Supporting Information: Video 4). Two patients also presented with motor dyspraxia (#4 and #5).

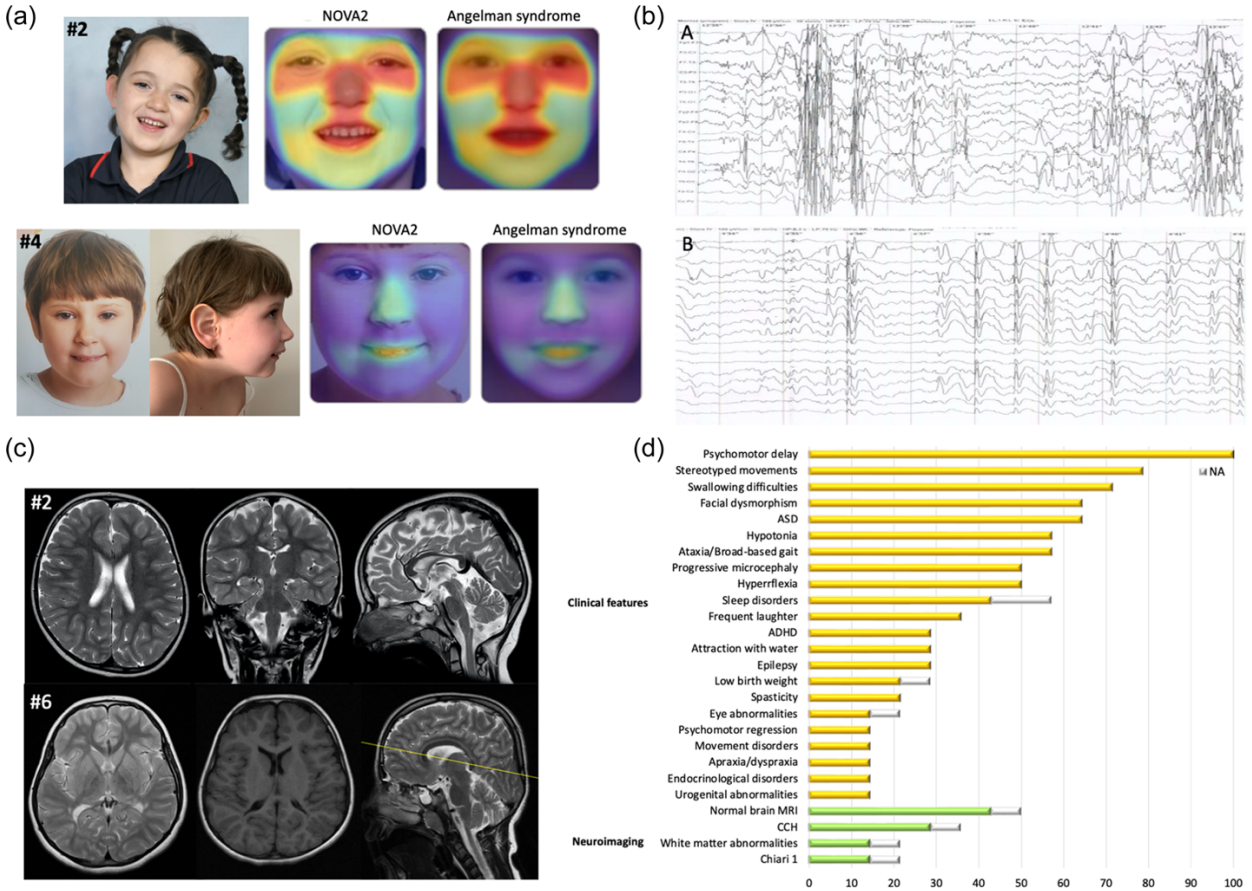


Figure 4. Electroclinical and neuroimaging features of patients harboring *de novo* truncating *NOVA2* variants. (a) Clinical photographs of patients #2 and #4. Patient #2 shows hypertelorism, intermittent esotropia, sunken nasal bridge, thin lips, and simplified protruding ears. Patient #4 shows slightly upslanting palpebral fissures, deep philtrum, large ears, and retrognathia with prominent chin. For each subject, a heat-map comparison between the patient's frontal photograph and a composite picture from subjects with Angelman syndrome (Face2Gene, https:// www.face2gene.com) is reported. (b) Electroencephalographic findings in patient #4. Ictal (A) and interictal (B) EEG showing generalized spike- and-wave activity and nonspecific electrical abnormalities, respectively. (c) Brain MRI (T2-weighted sequences) of patient #2 at 5.7 years shows thinning of the posterior section of the body of the corpus callosum and prominent CSF spaces overlying both frontal lobes. Brain MRI (T1- and T2-weighted

sequences) of patient #6 at 11 years shows bilateral peritrigonal periventricular white matter volume loss with increased signal intensity and extension along the right corona radiata, associated with thinning of the posterior portion of the body of the corpus callosum. (d) Percentage distribution of recurrent clinical manifestations and neuroimaging findings in all reported NOVA2 patients. Yellow and green bars indicate the percentage of patients with the corresponding clinical or neuroimaging feature, respectively. Grey bars indicate the percentage of patients in whom data were not available. ADHD, attention deficit-hyperactivity disorder; ASD, autism spectrum disorder; EEG, electroencephalogram; CCH, corpus callosum hypoplasia; NA, not applicable.

Associated syndromic features included urogenital manifestations (intra-abdominal testis in #3 and glomerulonephritis in #4), and endocrinological abnormalities (precocious puberty in #1 and hypothyroidism in #4). Brain magnetic resonance imaging (MRI) was normal in four cases (#1, #4, #7, and #8) while showed variable abnormalities in the remaining subjects, such as corpus callosum hypoplasia (CCH) (#2 and #6), bilateral peritrigonal periventricular white matter volume loss (#6), Chiari I malformation (#5), and prominent frontal subarachnoid spaces (#1) (Figure 4c).

4. Discussion

The disruption of splicing regulation is involved in complex multi- factorial diseases such as amyotrophic lateral sclerosis (ALS), mendelian disorders such as myotonic dystrophy (*Lee & Rio, 2015; López-Martínez et al., 2020*), and specific NDDs such as PUF60- and PQBP1-related disorders (OMIM #615583 and #309500, respectively) (*El Chehadeh et al., 2016; Kalscheuer et al., 2003; La Cognata et al., 2020*). Variants affecting AS have been associated with an increased risk for psychiatric features and are specifically involved in ASD, bipolar disorder, and schizophrenia (*Cai et al. 2021; Parikshak et al., 2016; Paterson et al., 2017; Reble et al., 2018; Stamova et al., 2013; Weyn-Vanhentenryck et al., 2018*). In particular, the abnormal function of the tissue-specific splicing regulator RNA- binding protein FOX1 (RBFOX1, OMIM *605104), also known as Ataxin-2-Binding Protein 1 (A2BP1), plays a role in the modulation of developing

cerebral cortex architecture and ASD pathophysiology (*Hamada et al., 2015; Lee et al., 2016*). Compelling evidence also implicate AS defects in the pathogenesis of neurodegenerative conditions, including Alzheimer's disease, Huntington's disease, Parkinson disease, and spinocerebellar ataxias (*Apicco et al., 2019; Li et al., 2021*).

NOVA2 is a crucial RBP for the AS regulation of several genes encoding proteins involved in neuronal differentiation and migration during brain development (*Saito et al., 2019*). A model in which a balance of transcript levels is maintained in the brain through a dynamically regulated NOVA-dependent AS-coupled NMD and direct interaction with 3' untranslated region binding elements has been recently suggested (*Eom et al., 2013*). This dynamism is well exemplified by the differential AS regulation of ITPR1 (OMIM *147265), encoding the inositol 1,4,5-triphosphate receptor type 1, in selected neuronal populations within different brain regions (*Saito et al., 2019*). NOVA2 not only operates as a trans-acting AS factor to determine exon definition but also acts as a cis-acting element regulating cell-type specific retention of introns, which titrate the binding of other trans-acting splicing factors (*Saito et al., 2019*). It is directly responsible for the AS regulation of several genes associated with NDDs (e.g., AP1S2) and other neurological conditions (e.g., DAB1) (*Mattioli et al., 2020*). Additionally, NOVA2 recently emerged as the first RBP to promote neural circular RNA (circRNA) biogenesis in the developing brain (*Knupp et al., 2021; Patop et al., 2019; Tang et al., 2021*). These molecules regulate gene transcription through microRNAs (miRNAs) repression and interaction with RBPs, modulating neuronal progenitor status maintenance, gene expression, synaptic transmission, and microglia activation (*Gokool et al., 2020; Meng et al., 2019; Patop et al., 2019*). Abnormal circRNA epigenetic modifications (especially 6-methyladenosine, m6A) affect RNA stability and result in neuronal disorders (*Meng et al., 2019; Zhang et al., 2019*).

De novo truncating variants in *NOVA2* cause a severe NDD characterized by global psychomotor delay, behavioral disorders, stereotyped movements, poor motor coordination, feeding difficulties, and associated neurological features (Figure 4d). Dysmorphic features are common and include hypertelorism, deep-set eyes with epicanthal folds and long eyelashes, large ears, pointed chin, and Angelman-like features, such as sunken nasal bridge and long philtrum (Figure 4a). However, a specific facial gestalt does not appear to emerge. In addition to previously reported clinical features, variable novel manifestations were observed in our cohort, mainly consisting of ADHD, dyspraxia/apraxia, urogenital (intra-abdominal testis, glomerulonephritis), and endocrinological (precocious puberty, hypothyroidism) abnormalities. Remarkably, we also observed an evident psychomotor regression in two cases (#4 and #6), correlating with a smallpox-related increase in the epileptic activity in subject #4. These individuals also suffered from movement disorders consisting of tremors with or without choeroathetosis. Another previously unreported finding is the progressive microcephaly observed in four patients (#2 and #6-8). Taken together, the phenotype extension suggests that this condition may have a heterogeneous and occasionally severe clinical course.

Although seizures have been previously reported in the two subjects harboring the p.(Val261Glyfs*135) and p.(Val261Trpfs*135) variants, the electroencephalographic findings and seizure response to medical treatment remained elusive (*Mattioli et al., 2020*). We observed myoclonic or myoclonic-atonic seizures with onset at 2–4.5 years and generalized spike-and-wave activity at the electroencephalogram in #4. Clonazepam alone was administered in subject #4, whereas patient #6 received a combination of topiramate, clobazam, and valproate. In both cases, only a partial clinical response consisting of decreased seizure frequency could be observed. Interestingly, seizures were occasionally precipitated by sound stimuli in subject #4,

who also showed hyperekplexia, suggesting a differential diagnosis with reflex seizures (*Striano et al., 2012*). Cortical hyperexcitability and epilepsy have been observed in the heterozygous knock-out mouse models, supporting an underlying pathogenic link between NOVA2 function and epilepsy (*Eom et al., 2013*). More specifically, NOVA-mediated regulation of NMD splicing controls the levels of many synaptic proteins (e.g., DLG3, PSD95) and ion channels/transporters (e.g., SCN9A, SLC4A10, and SLC4A3), whose abnormal stoichiometry can lead to electrical imbalance and epileptogenesis (*Eom et al., 2013*). Of note, NOVA2 was found to interact with a cis-acting polymorphism in SCN1A (rs3812718) and modulate the proportions of drug-responsive alternative transcripts (*Heinzen et al., 2007*).

Overall, brain MRI is normal in ~40% of NOVA2 patients (6/14 subjects in total), whereas variable abnormalities are present in the remaining individuals (Table 1). Although some alterations could be observed in more than one subject (i.e., CCH, white matter volume loss, and Chiari I malformation), a common neuroimaging pattern cannot be delineated. More in general, the observed neuroimaging abnormalities might be the consequence of abnormal regulation of transcripts encoding crucial proteins for axonal growth and pathfinding resulting from NOVA2 deficiency (*Saito et al., 2016*). For example, several genes associated with CCH syndromes are target of NOVA2- mediated AS regulation (e.g., CASK and DCC) and CCH is present in 33% of patients (*Saito et al., 2016*). However, this feature is nonspecific and the observation of a normal brain MRI in a not insignificant number of patients remains challenging to explain. Furthermore, while the progressive motor discoordination observed in the conditional mouse model with specifical Nova2 inactivation is recapitulated in NOVA2 patients, there is no comparable cerebellar atrophy (*Saito et al., 2019*). The report of an additional cohort will help further delineate the spectrum of NOVA2-related brain abnormalities.

Among the variants identified in our cohort, four (p.(Leu252- Profs*141), p.(Leu252Phefs*144), p.(Ala263Profs*133), and p.(Leu276Cysfs*120)) are located in proximity of the terminal of KH2 domain, in line with the previous report by *Mattioli et al.* (*2020*). The p.(Leu175Cysfs*6) variant is the first variant falling within the KH2 domain to be reported. Similarly, we identified the first NOVA2 variants localizing to the KH1, p.(Gln86*), and KH3 domain, p.(Lys444Glnfs*82). Of note, the c.256C>T p.(Gln86*) variant is the first non-frameshift *NOVA2* variant to be described and the first variant localized in Exon 3, whereas all other reported variants fall in Exon 4 (Figure 1). The four variants localized to the "cluster region" (p.(Leu252Profs*141), p.(Leu252Phefs*144), p.(Ala263Profs*133), and p.(Leu276Cysfs*120)) (Figure 1) lead to the same alternative frame of previously reported changes (*Mattioli et al., 2020*). The remaining variants are instead predicted to cause different functional consequences. While the p.(Leu175Cysfs*6) likely disrupts the KH2 domain and the distal p.(Lys444Glnfs*82) results in a more preserved alternative C-terminal of the protein, the early p.(Gln86*) variant very likely results in NMD. This partially questions the previously assumed parallel between NOVA2-related disorder and Robinow syndrome as conditions caused by specific distal frameshift variants (*Mattioli et al., 2020; Supek et al., 2020*). However, although puzzling, the localization pattern of all known NOVA2 variants and the observation that most of them lead to the same alternative frame still suggests the existence of a possible mutational hotspot region (Figure 1).

Although a possible gain-of-function effect cannot be excluded, *NOVA2* variants are predicted to act through a partial loss of function (hypomorphic) mechanism (*Mattioli et al., 2020*). However, the functional consequences of proximal p.(Gln86*) and distal p.(Lys444Glnfs*82) variants might fall at the opposite ends of the pathophysiological spectrum.

The addition of the C-terminal part produced by the recurrent frameshift variants next to the KH2 domain allows a residual function of the protein (*Mattioli et al., 2020*). An early truncating variant is instead likely to result in a severe impairment of protein function. This is exemplified by the much stronger loss of AS regulation produced by the overexpression of the truncated p.Tyr231* NOVA2 protein variant in HeLa cells as compared with Mut1 (*Mattioli et al., 2020*). In line with this observation, the p.(Gln86*) is expected to result in a complete loss of NOVA2 function and potentially lead to more severe consequences than other variants. The p.(Lys444Glnfs*82) variant might instead lead to more subtle functional effects, sparing the KH3 domain and allowing the production of a more functionally active truncated protein. This is also supported by the milder impact of this variant on AS events, resulting in an intermediate phenotype between WT and Mut1 p.(Val261Glyfs*135). Indeed, the KH3 domain is crucial for the binding to the UCAY sequence in the pre- mRNA, which cannot be duplicated by KH1 and KH2 domains (*Jensen et al., 2000; Lewis et al., 2000*).

The spectrum of *NOVA2* variants might be wider than previously expected and their localization might influence the residual protein function. Although it is tempting to speculate that the neurological phenotype (epilepsy and psychomotor regression) associated with the proximal p.(Gln86*) variant is more severe than that related to the distal p.(Lys444Glnfs*82), similar features were also observed in the subject harboring the p.(Leu276Cysfs*120) variant. Additionally, the clinical manifestations of patients harboring variants localizing outside of the "clustering region" appear to be quite overlapping with the other subjects except for ADHD, which was never reported in patients with NOVA2 variants before. To further complicate the picture, a possible NMD escape cannot be completely excluded for the proximal p.(Gln86*) variant, making it premature to draw any conclusions (*Inácio et al., 2004; Pereira et al., 2015;*

Dyle et al., 2020). In light of these observations, the report of other cohorts expanding the molecular spectrum of *NOVA2*-related disorder will likely play a crucial role in the delineation of potential genotype-phenotype correlations.

5. Conclusion

De novo truncating NOVA2 variants lead to a severe and heterogeneous neurodevelopmental condition with behavioral disturbances, epilepsy, neurological features, and variable brain MRI abnormalities. Our findings confirm that pathogenic NOVA2 variants negatively affect AS events, likely leading to impaired neuronal development, axon guidance, and synaptic plasticity and function. The milder functional impairment observed for the distal p.(Lys444Glnfs*82) variant suggests that variant localization might influence the residual protein function, possibly determining a wider than expected molecular and phenotypic spectrum. However, this intermediate effect detected in vitro does not necessarily predict an intermediate effect in vivo in a physiological context. Moreover, the limited number of reported subjects and the complex mechanisms involved in NMD escape make it difficult to draw conclusions on the pathophysiological link between specific variants and phenotypic manifestations. Further studies will hopefully help clarify these intriguing aspects, possibly laying the foundation for more robust genotype-phenotype correlations in NOVA2 patients.

References

Adzhubei I.A., et al. (2010). A method and server for predicting damaging missense mutations. Nature Methods, 7(4), 248–249. https://doi.org/10.1038/nmeth0410-248

Apicco D.J., et al. (2019). Dysregulation of RNA splicing in tauopathies. Cell Reports, 29(13), 4377–4388. e4. https://doi.org/ 10.1016/j.celrep.2019.11.093

Cai X., et al. (2021). A human- specific schizophrenia risk tandem repeat affects alternative splicing of a human-unique isoform AS3MTd2d3 and mushroom dendritic spine density. Schizophr Bull, 47(1), 219–227. https://doi.org/10.1093/schbul/sbaa098

Cooper G.M., et al. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Research, 15(7), 901–913. https://doi.org/10. 1101/gr.3577405

Combined Annotation Dependent Depletion (CADD): http://cadd.gs. washington.edu

den Dunnen J.T., et al. (2016). HGVS recommendations for the description of sequence variants: 2016 Update. Human Mutation, 37(6), 564–569.https://doi.org/10.1002/humu.22981

Dyle M.C., et al. (2020). How to get away with nonsense: Mechanisms and consequences of escape from nonsense-mediated RNA decay. Wiley Interdiscip Rev RNA, 11(1), e1560. https://doi.org/10.1002/wrna.1560

El Chehadeh S., et al. (2016). Dominant variants in the splicing factor PUF60 cause a recognizable syndrome with intellectual disability, heart defects and short stature. European Journal of Human Genetics, 25(1), 43–51. https://doi.org/10.1038/ejhg.2016.133

Eom T., et al. (2013). NOVA-dependent regulation of cryptic NMD exons controls synaptic protein levels after seizure. eLife, 2, e00178. https://doi. org/10.7554/eLife.00178

Gokool A., et al. (2020). Circular RNAs: The brain transcriptome comes full circle. Trends in Neurosciences, 43(10), 752–766. https://doi.org/10.1016/j.tins.2020.07.007

Hamada N., e al. (2015). Role of the cytoplasmic isoform of RBFOX1/A2BP1 in establishing the architecture of the developing cerebral cortex. Molecular Autism, 6, 56. https://doi.org/10.1186/s13229-015-0049-5

Heinzen E.L., et al. (2007). Nova2 interacts with a cis-acting polymorphism to influence the proportions of drug-responsive splice variants of SCN1A. The American Journal of Human Genetics, 80(5), 876–883. https://doi.org/10.1086/516650

Huang C.S., et al. (2005). Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. Cell, 123(1), 105–118. https://doi.org/10.1016/j.cell.2005.07.033

Inácio A., et al. (2004). Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. Journal of Biological Chemistry, 279(31), 32170–32180. https://doi.org/10.1074/jbc.M405024200

Jensen K.B., et al. (2000). Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. Neuron, 25(2), 359–371. https://doi.org/10.1016/s0896-6273(00)80900-9

Kalscheuer V.M., et al. (2003). Mutations in the polyglutamine binding protein 1 gene cause X- linked mental retardation. Nature Genetics, 35(4), 313–315. https:// doi.org/10.1038/ng1264

Knupp D., et al. (2021). NOVA2 regulates neural circRNA biogenesis. Nucleic Acids Research, 49(12), 6849–6862. https://doi.org/10.1093/nar/gkab523

La Cognata V., et al. (2020). Splicing players are differently expressed in sporadic amyotrophic lateral sclerosis molecular clusters and brain regions. Cells, 9(1), 159. https://doi.org/10.3390/cells9010159

Lee J.A., et al. (2016). Cytoplasmic Rbfox1 regulates the expression of synaptic and autism-related genes. Neuron, 89(1), 113–128. https://doi.org/10. 1016/j.neuron.2015.11.025

Lee Y., et al. (2015). Mechanisms and regulation of alternative pre- mRNA splicing. Annual Review of Biochemistry, 84, 291–323. https:// doi.org/10.1146/annurev-biochem-060614-034316

Lek M., et al., Exome Aggregation Consortium. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536, 285–291. https://doi.org/10.1038/nature19057

Lewis H.A., et al. (2000). Sequence-specific RNA binding by a Nova KH domain: Implications for paraneoplastic disease and the fragile X syndrome. Cell, 100(3), 323–332. https://doi.org/10.1016/s0092-8674(00)80668-6

Li D., e al. (2021). Neurodegenerative diseases: A hotbed for splicing defects and the potential therapies. Translational Neurodegeneration, 10(1), 16. https://doi.org/10.1186/s40035-021-00240-7

Licatalosi D.D., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature, 456(7221), 464–469. https://doi.org/10.1038/nature07488

Lipscombe D., et al. (2019). Alternative splicing of neuronal genes: New mechanisms and new therapies. Current Opinion in Neurobiology, 57, 26–31. https://doi.org/10.1016/j.conb.2018.12.013

López-Martínez A., et al. (2020). An overview of alternative splicing defects implicated in myotonic dystrophy type I. Genes, 11(9), 1109. https://doi.org/10.3390/genes11091109

Mattioli F., et al. (2020). De Novo frameshift variants in the neuronal splicing factor NOVA2 result in a common C-terminal extension and cause a severe form of neurodevelop- mental disorder. American Journal of Human Genetics, 106(4), 438–452. https://doi.org/10.1016/j.ajhg.2020.02.013

Meng S., et al. (2019). Epigenetics in neurodevelopment: Emerging role of circular RNA. Frontiers in Cellular Neuroscience, 13, 327. https://doi.org/10.3389/fncel.2019.00327

Ng P.C. (2003). SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research, 31, 3812–3814. https://doi.org/10.1093/nar/gkg509

Park E., et al. (2018). The expanding landscape of alternative splicing variation in human populations. American Journal of Human Genetics, 102(1), 11–26. https://doi.org/ 10.1016/j.ajhg.2017.11.002

Parikshak N.N., et al. (2016). Genome- wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature, 540(7633), 423–427. https://doi.org/10.1038/nature20612

Paterson C., et al. (2017). Temporal, diagnostic, and tissue-specific regulation of NRG3 isoform expression in human brain development and affective disorders. American Journal of Psychiatry, 174(3), 256–265. https://doi.org/10.1176/appi.ajp.2016.16060721

Patop I.L., et al. (2019). Past, present, and future of circRNAs. EMBO Journal, 38(16), e100836. https://doi.org/10. 15252/embj.2018100836

Pereira F.J., et al. (2015). Resistance of mRNAs with AUG- proximal nonsense mutations to nonsense- mediated decay reflects variables of mRNA structure and translational activity. Nucleic Acids Research, 43, 6528–6544. https://doi.org/10.1093/nar/gkv588

Porter R.S., et al. (2018). Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders. Molecular and Cellular Neuroscience, 87, 35–45. https://doi.org/10.1016/j.mcn.2017.10.006

Reble E., et al. (2018). The contribution of alternative splicing to genetic risk for psychiatric disorders. Genes, Brain and Behavior, 17(3), e12430. https://doi.org/10.1111/gbb.12430

Rentzsch P., et al. (2019). CADD: Predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Research, 47, D886–D894. https://doi. org/10.1093/nar/gky1016

Richards S., et al. (ACMG Laboratory Quality Assurance Committee 2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine, 17, 405–424. https://doi.org/10.1038/gim.2015.30

Ruggiu M., et al. (2009). Rescuing Z+ agrin splicing in Nova null mice restores synapse formation and unmasks a physiologic defect in motor neuron firing. Proceedings of the National Academy of Sciences of the United States of America, 106(9), 3513–3518. https://doi.org/10.1073/pnas.0813112106

Saito Y., et al. (2016). NOVA2-mediated RNA regulation is required for axonal pathfinding during development. eLife, 5, e14371. https://doi.org/10.7554/ eLife.14371

Saito Y., et al. (2019). Differential NOVA2- mediated splicing in excitatory and inhibitory neurons regulates cortical development and cerebellar function. Neuron, 101(4), 707–720. e5. https://doi.org/10.1016/j.neuron.2018.12.019

Schwarz J.M., et al. (2014). Mutation- Taster2: Mutation prediction for the deep-sequencing age. Nature Methods, 11, 361–362. https://doi.org/10.1038/nmeth.2890

Sobreira N., et al. (2015). Gene- matcher: A matching tool for connecting investigators with an interest in the same gene. Human Mutation, 36(10), 928–930. https://doi.org/10.1002/humu.22844

Stamova B.S., Tian. (2013). Evidence for differential alternative splicing in blood of young boys with autism spectrum disorders. Molecular Autism, 4(1), 30. https://doi.org/10.1186/2040-2392-4-30

Striano S., et al. (2012). Reflex seizures and reflex epilepsies: old models for understanding mechanisms of epileptogenesis. Epilepsy Research, 100(1-2), 1–11. https://doi.org/10.1016/j.eplepsyres.2012.01.013

Supek F., et al. (2020). To NMD or not To NMD: Nonsense-mediated mRNA decay in cancer and other genetic diseases. Trends in Genetics, 37(7), 657–668. https://doi.org/10. 1016/j.tig.2020.11.002

Tang X., et al. (2021). Review on circular RNAs and new insights into their roles in cancer. Computational and Structural Biotechnology Journal, 19, 910–928. https://doi.org/10.1016/j.csbj.2021.01.018

Traunmüller L., et al. (2016). Control of neuronal synapse specification by a highly dedicated alternative splicing program. Science, 352(6288), 982–986. https:// doi.org/10.1126/science.aaf2397

Ule J., et al. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science, 302(5648), 1212–1215. https://doi.org/10.1126/science.1090095

Ule J., et al. (2006). An RNA map predicting Nova-dependent splicing regulation. Nature, 444(7119), 580–586. https://doi.org/10.1038/nature05304

Ule J., et al. (2005). Nova regulates brain-specific splicing to shape the synapse. Nature Genetics, 37(8), 844–852. https://doi.org/10.1038/ng1610

Vuong C.K., et al. (2016). The neurogenetics of alternative splicing. Nature Reviews Neuroscience, 17, 265–281. https://doi.org/10.1038/nrn.2016.27

Weyn-Vanhentenryck S.M., et al. (2018). Precise temporal regulation of alternative splicing during neural development. Nature Communications, 9(1), 2189. https://doi.org/10.1038/s41467-018- 04559-0

Will C.L., et al. (2011). Spliceosome structure and function. Cold Spring Harbor Perspectives in Biology, 3(7), a003707. https://doi. org/10.1101/cshperspect.a003707

Wilkinson M.E., et al. (2020). RNA splicing by the spliceosome. Annual Review of Biochemistry, 89, 359–388. https:// doi.org/10.1146/annurev-biochem-091719-064225

Yang Y.Y., et al. (1998). The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. Proceedings of the National Academy of Sciences of the United States of America, 95(22), 13254–13259. https://doi.org/10.1073/pnas.95.22.13254

Zhang S.F., et al. (2019). The role of non-coding RNAs in neurodevelopmental disorders. Frontiers in Genetics, 10, 1033. https://doi.org/10.3389/fgene.2019.01033

Appendix

Web resources

ChopChop2, http://chopchop.cbu.uib.no/

ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/

Combined Annotation Dependent Depletion (CADD), http://cadd.gs.washington.edu

DECIPHER: https://decipher.sanger.ac.uk

EMBL-EBI Expression Atlas, https://www.ebi.ac.uk/gxa/experiments/E-ERAD-475/Results

Ensembl, https://www.ensembl.org/index.html

Ensembl Variant Effect Predictor (VEP) pipeline,

https://www.ensembl.org/info/docs/tools/vep/index.html

Exon-Intron Graphic Maker, http://wormweb.org/exonintron

FastQC, http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc

Gene Cards, http://www.genecards.org

Gene Matcher, http://www.genematcher.org

Genome Aggregation Database (GnomAD), http://gnomad.broadinstitute.org

Genomic Evolutionary Rate Profiling – GERP,

http://mendel.stanford.edu/SidowLab/downloads/gerp/

Human Genome Variation Society, https://varnomen.hgvs.org

Leiden Open Variation Database (LOVD), https://www.lovd.nl

Mutation Assessor, http://mutationassessor.org/r3/

Mutation Taster, http://www.mutationtaster.org

Online Mendelian Inheritance in Man, http://www.ncbi.nlm.nih.gov/Omim

Polyphen-2, http://genetics.bwh.harvard.edu/pph2/

Proteomics DB, https://www.proteomicsdb.org

PubMed, http://www.ncbi.nlm.nih.gov/pubmed

RefSeq, https://www.ncbi.nlm.nih.gov/refseq

SIFT, https://sift.bii.a-star.edu.sg

Splice AI, https://spliceailookup.broadinstitute.org

UniProt, https://www.uniprot.org

UCSC Human Genome Database, http://www.genome.ucsc.edu

Varsome, https://varsome.com