UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Master's Project in Medicine No 4362

The Genetics of Intellectual Disability: whole exome sequencing to find causative variants in severe cases

Student
Winteler Florence

Supervisor
Prof. Reymond Alexandre, Ph.D.
Center for Integrative Genomics, UNIL

Co-Supervisor
Gueneau Lucie, Ph.D.
Center for Integrative Genomics, UNIL

Expert
Prof. Draganski Bogdan, Dr. méd
Département des neurosciences cliniques, CHUV

Lausanne, 25.11.2017

Abstract

Intellectual disability (ID) affects 1-3% of the population. A genetic origin is estimated to account for about half of the currently undiagnosed cases, and despite recent successes in identifying some of the genes, it has been suggested that hundreds more genes remain to be identified. ID can be isolated or part of a more complex clinical picture –indeed other symptoms are often found in patients with severe genetic ID, such as developmental delay, organ malformations or seizures.

In this project, we used whole exome sequencing (WES) to analyse the coding regions of the genes (exons) of patients with undiagnosed ID and that of their families. The variants called by our algorithm were then grossly sorted out using criteria such as frequency in the general population and predicted pathogenicity. A second round of selection was made by looking at the relevant literature about the function of the underlying genes and pathways involved. The selected variants were then Sanger-sequenced for confirmation. This strategy allowed us to find the causative variant and give a diagnosis to the first family we analysed, as the patient was carrying a mutation in the Methyl-CpG binding protein 2 gene (*MECP2*), already known to cause Rett syndrome. For the second family we could not elucidate the cause of ID with certainty. We have a suspicion that a variant in the Synaptotagmin-Like 2 gene (*SYTL2*) may be the cause of the symptoms, but we could not prove it as no other patients harbouring this variant and similar clinical traits have been found up to now. However *SYTL2* remains our strongest candidate as it is involved in a pathway known to cause ID and because *SYTL3* was recently uncovered as a cause of ID in Pr Reymond's lab (unpublished data). A first step forward would be to perform functional studies to confirm the protein-disrupting nature of the variant.

Keywords: intellectual disability, whole exome sequencing, neurological disorders

Table of contents

Introduction

Intellectual disability (ID) affects 1-3% of the population at birth (1). It is defined by significant limitations both in intellectual functioning and adaptive behaviour, originating before the age of 18 years old (2), and resulting in an IQ of less than 70. This condition can be caused by environmental factors such as alcohol abuse or infections during pregnancy, preterm birth, neonatal hypoxia, as well as genetic factors. In fact, it is estimated that between 15% and 50% of ID cases are due to genetic causes (1). To this day, over 700 genes responsible for ID have been uncovered (3), and it has been suggested that a total of about 1,000 may be implicated (4).

Our understanding and knowledge of genetically-caused ID is technology-limited. In the early 20th century, cytogenetic analysis started to be applied to the human genome, successfully leading to the identification, in 1959, of the first (and most prevalent) cause of genetic ID: trisomy 21, namely Down syndrome (15% of ID cases). New technologies kept on arising: Fluorescence in situ hybridisation (FISH) in 1977 (5), Sanger sequencing in 1977 (6) and Array comparative genomic hybridisation (aCGH) in 1992 (7). Thanks to these techniques, new types of genetic abnormalities became accessible to discovery: point mutations, copy-number variants (CNVs), microdeletions and duplications. This allowed for the genetic characterisation of previously idiopathic diseases: Prader-Willi (OMIM[1] #176270) and Angelman (OMIM #105830) syndromes (using FISH), Fragile X syndrome (OMIM #300624, *FMR1* gene), Williams syndrome (OMIM #194050, multiple genes involved). However, these techniques lack the power to explore the entire genome and could only elucidate a fraction of ID cases.

Today with the emergence of next-generation sequencing (NGS) tools and the big data approach, we are able to scrutinise the genome base-by-base. This set of techniques allows us to conduct an unbiased tactic, meaning we collect all the genetic information without having to make prior guesses as to what we are looking for. In the field of ID of genetic origin, these strategies are very welcome because they let investigators collect and analyse the DNA of large cohorts of ID patients and their relatives, making it possible to compare and find the differences on a genome-wide level.

However, there are also some limitations linked to these sequencing techniques. First of all the assembly of the sequences is not always accurate when it comes to repetitive or homologous sequences (8) and mutations in these regions can therefore be missed. Moreover, other modifications like epigenetic changes and chromatin states are not accessible to these techniques. In the case of whole exome sequencing, coverage is also an issue as it is uneven and especially low in repetitive regions and GC-rich regions (9) and leads to difficulties in variant calling. Although over 95% of exons are targeted with good coverage, some exons are still missed and this can be clinically significant, leading to a high rate of false negatives (10, 11). There are other shortcomings that are specific to WES: non-coding regions are completely dismissed and certain mutation types are not

[1] OMIM Online Mendelian Inheritance in Man: An Online Catalog of Human Genes and Genetic Disorders

detected, such as large rearrangements, large copy number variations[2], mutations in genes with a highly homologous region and mitochondrial gene mutations, which are often not targeted (although they can be retrieved using bioinformatics tools (12, 13)). Additionally, the clinical significance of the findings can be difficult to establish as in most cases the variant's impact on protein function and patient phenotype is unknown.

Our goal for this project is to find the causative mutations in unsolved cases of severe ID suspected to have a genetic origin that could not be uncovered by classical karyotype and aCGH. We hope to give clues to the families about the cause of ID and other symptoms, giving a basis for genetic counselling and treatment. On a wider scope we also hope to uncover new mutations or mutations in new genes responsible for ID and describe their impact in detail. Finding new genes responsible for ID would also help to elucidate their function and that of their matching proteins.

[2] Large rearrangements and copy number variations can be detected by FISH or aCGH. These techniques are often performed before a WES is undertaken, so this shortcoming of WES is usually not an issue.

Materials, Patients and Methods

Families with one or several children suffering from severe ID but without a specific diagnosis were selected by our clinician collaborators in the Center for Medical Genetics at the Vilnius University Hospital Santariskiu Klinikos in Lithuania. The patient's observed symptoms and clinical details were sent to us as well as DNA material from the patient and all available first degree family members.

In this project we analysed two different Lithuanian families (See next section for a detailed description of the probands).

For each family we performed whole exome sequencing on the DNA of affected and unaffected family members. Exomes were captured using the Agilent SureSelect Human All Exon V5 enrichment kit and multiplex sequenced on an Illumina HiSeq 2500 platform. Purity-filtered reads were adapted and quality trimmed with FastqMcf (14), and aligned to the reference human genome (GRCh37 and decoy sequence) using BWA-MEM(15). The following steps were performed on each sample separately: [1] duplicate marking using PICARD tools (http://picard.sourceforge.net); [2] Indel realignment with GATK; [3] base quality score recalibration; [4] variants calling with GATK HaplotypeCaller in gVCF mode (16, 17). All samples were then genotyped together with GATK GenotypeGVCFs and variants were filtered using GATK Variant Quality Score Recalibration (16).

The variants[3] called were then analysed and selected using VarApp (18), an application developed by the Swiss Institute of Bioinformatics (SIB) for variant sorting under various scenarios of heredity and with modifiable constraints for variant frequency, read quality and predicted pathogenicity of the variants (See Table 1). The threshold for frequency of the variant in the population was set depending on the scenario of heredity under investigation. For autosomal dominant and *de novo* variants (where only one copy of the variant is found in the patient) and for recessive variants (where two copies of the variant are found in the patient) we set the maximum population frequency at 1% under the assumption that these variants would be rare in the population because of their damaging potential. For compound heterozygous heredity (where two distinct heterozygous variants are found in the same gene, each inherited from one of the parents) we set the maximum population frequency of each variant at 5% because we assume these variants cause no harm individually and would not be strongly selected against at the population level. These variants underwent a second round of selection in which the candidate variants were hand-picked depending on the likelihood of their involvement in the patient's symptoms. In addition to literature reviewing, the following online tools were used to assess the phenotype that might be expected from the detected variants: Ensembl (19), NCBI (20), OMIM (21), GeneCards (22), MalaCards (23), GTEx (24), GIANT (25), Decipher (26), Orphanet (27) and Genematcher (28). We compared these predictions with the documented symptoms of the patients along with the expression and known function of the candidate genes to select the final candidate variants. These have to be validated using Sanger

[3] Please note that in this project we used the GRCh37 assembly of the human genome.

sequencing because NGS is error-prone and generates false positive variants due to assembly misalignments (29).

Primers for selective amplification of the candidate genes were designed with the help of Primer-BLAST (30) and produced by SIGMA (See designed primers in Table 3 and Table 5). PCR amplification of the candidate genes was then first performed on control DNA using the JumpStart REDTaq ReadyMix Reaction Mix from SIGMA and following the manufacturer's protocol. 80ng of DNA was used for each amplification in a total volume of 50µl. The TouchDown 65-55°C PCR program was used as follows: initial denaturation (5min at 94°C) / 10 cycles denaturation-annealing-elongation (30s at 94°C, 30s at 65°C, 45s at 72°C, decreasing the annealing temperature by one degree at each cycle, from 65°C to 56°C) / 25 cycles denaturation-annealing-elongation (30s at 94°C, 30s at 55°C, 45s at 72°C) / final elongation 5min at 72°C (inspired from (31)). To check if the reaction had worked, a 5µl sample of each reaction product was then run on a 1.5% agarose gel stained with the GelRed Nucleic Acid Gel Stain (from Biotium) for 30-40min at 150V. We also ran 3µl of either a 50bp or a 100bp DNA ladder (TrackIt form Thermofisher) to assess the size of our reaction products. If the results from the gel were unsatisfactory (no band, multiple bands, wrong size), we tried changing the PCR conditions. Troubleshooting included: changing the annealing temperature in the PCR program either to TouchDown 70-60°C or to 60-50°C, adding 5% DMSO in the reaction mix or using a different DNA polymerase: the PFU from Promega or the GC-Rich PCR System from Roche. When the result was satisfactory the same protocol was applied to the DNA extracted from the family members and the resulting PCR products were then purified using the MiniElute PCR Purification Kit from QIAGEN following the manufacturer's protocol and sent for sequencing to GATC Biotech. The sequences received were downloaded and then analysed using the Sequencher software.

The variants confirmed by Sanger sequencing were re-analysed more thoroughly: we looked for published mutations, syndromes that were compatible with the clinical picture and asked the clinician for more details about the patient's clinical condition.

Table 1: Generic criteria for variant selection using VarApp

Frequency of occurrence (MAF)	Quality filter	Prediction of pathogenicity	Impact of variant on protein structure
The population frequency of the variant, as determined by the ExAC browser (32), was set at maximum 1% for dominant, *de novo* and recessive inheritance, and 5% for compound heterozygous inheritance.	VQSR (Variant quality score recalibration) was used to calculate a quality score, and then allow the variants to PASS the filter only if they were above a threshold	Scaled CADD score (33) of at least 10 (CADD: Combined Annotation Dependent Depletion)	Only variants expected to have a high impact were selected (frameshift variant, splice acceptor variant, splice donor variant, start lost, stop gained, stop lost, inframe deletion, inframe insertion, missense variant, protein altering variant, splice region variant)

Results

Family 1

Clinical description

The patient is a girl with ID of unknown origin (Picture and family tree in Figure 1). She and her dizygotic twin were born in 2007, into a family with two other siblings. The patient is the only family member affected by ID. She was born at 37 weeks of amenorrhea, with a birth weight of 2800g (between the 10th and 50th percentiles[4]). During pregnancy, twin-twin transfusion syndrome was diagnosed, with no further consequences. She had severe hypotonia at birth but developed normally until she was 6-7 months old, at which point psychomotor and developmental delay started to show. The patient's abilities started to regress, and around the age of 10-11 months she lost fine motor skills. She was only able to learn one word (at 9 months old), and could never walk. In all, she exhibits a severe global developmental delay, with severe ID and an estimated IQ under 20.

She also presents with other notable signs and symptoms such as hand stereotypies (hand mouthing and tapping), small and cold hands and feet, staring episodes, bruxism, scoliosis, kyphosis, hydrocephalus from 10 months of age, generalized epilepsy from age 2 and dystonia from 3-4 years of age. A postnatal deceleration of her head growth was also observed: the patient started between percentiles 3 and 10 (p3-p10), was at p3 at 5 years old and then went below p3.

Her phenotype in 2012 was as follows: height at p3 and weight below p3. She had a broad face, reddish cheeks, a mild synophrys, hypertelorism, bright dense eyelashes, slightly retracted lateral palpebral commissures, low-set ears, a small nose, hypoplastic alae nasi, a short philtrum and a narrow upper lip.

Karyotype, aCGH and metabolic tests were all normal in this patient. Brain MRIs in 2008 and 2010 revealed a myelination disorder, hydrocephalus, and hypoplasia of the corpus callosum.

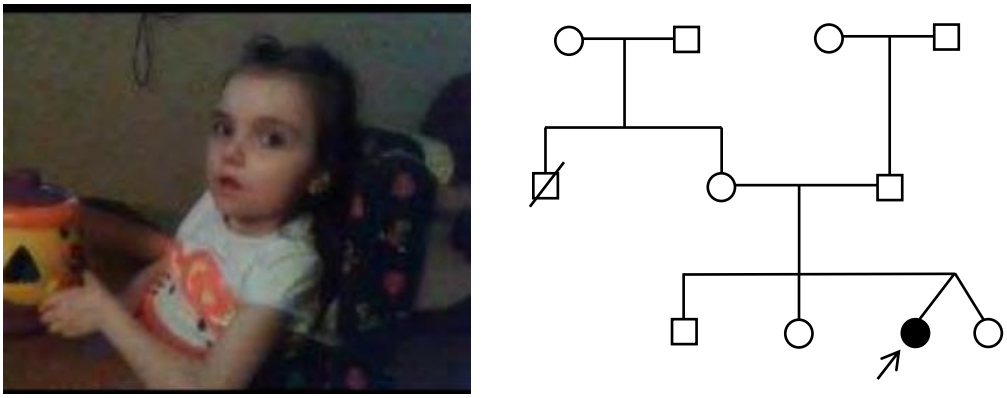


Figure 1: Picture of the proband from Family 1 (left) and family tree (right)

[4] A percentile is a measure used in statistics indicating the value below which a given percentage of observations in a group of observations fall. For example, the 20th percentile is the value (or score) below which 20% of the observations may be found. (Source: Wikipedia)

Whole exome sequencing: variants and their analysis

We applied the whole exome sequencing pipeline to all family members. Without applying any filters, we found 131'149 different variants among the exomes of all family members. By using the criteria described in Table 1 we reduced the number of variants to be analysed manually to 6 *de novo*, 2 recessive and 4 compound heterozygous (See Supplementary Table 1 for details about all the variants).

Based on the expected neurological phenotype the candidate variants could cause, we chose to Sanger sequence only four of them: *AMPD2*, *COPB1*, *MECP2* and *HERC1* (See Table 2*)*. The primers we designed and used for sequencing are listed in Table 3.

Table 2: Variants selected for Sanger sequencing confirmation for family 1 (selection based on expected neurological phenotypes and other symptoms resembling the patient's caused by the variants; underlined)

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity(34, 35) Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
***DE NOVO* variants**			
AMPD2 *adenosine monophosphate deaminase 2*	Missense variant	Autosomal recessive spastic paraplegia (OMIM #615686) Pontocerebellar hypoplasia type 9 (OMIM #615809) (both recessive)	AMPD2 is necessary for neurogenesis; known mutations exist but not that of our patient. Pontocerebellar hypoplasia: microcephaly, optic atrophy, seizures, spasticity, global developmental delay.
	Position: chr1: 110167952 ENST00000256578.3: c.281C>T ENSP00000256578.3: p.Pro94Leu dbSNP: rs748496694		
	SIFT: 0.01 (deleterious low conf.) PolyPhen-2: 0.003 (benign)		
	Allele frequency: 0.00002538		
	Expressed in brain		
COPB1 *coatomer protein complex, subunit beta 1*	Splice acceptor variant	Oculocerebrorenal syndrome (OMIM #309000)	Oculocerebrorenal Syndrome (Lowe syndrome) has symptoms including renal insufficiency, abnormality of the renal tubule and proteinuria, congenital bilateral cataract, often with glaucoma, strabismus and cornea lesions. The kidneys suffer from Fanconi syndrome. Brain development is abnormal, with an important delay in psychomotor development including muscular hypotonia, abnormal behaviour, epilepsy and growth retardation.
	Position: chr11: 14501261 ENST00000439561.2: c.1213-1C>T		
	SIFT: N/A PolyPhen-2: N/A		
	Allele frequency: 0.00009077		

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity(34, 35) Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
	Expressed in brain		COPB1 is involved in Notochord development (zebrafish).
MECP2 *methyl CpG binding protein 2*	Missense variant Position: chrX: 153296806 ENST00000303391.6: c.473C>T ENSP00000301948.6: p.Thr158Met dbSNP: rs28934906 SIFT: 0 (deleterious) PolyPhen-2: 1 (probably damaging) Allele frequency: unknown Expressed in brain	Rett syndrome (OMIM #312750); X-linked syndromic ID; Severe neonatal encephalopathy; Autism; Atypical Rett syndrome; Angelmann syndrome; PPM-X syndrome; Bruxism	MECP2 binds to methylated DNA and can repress transcription from methylated gene promoters. *MECP2* is X-linked and subject to X inactivation. *MECP2* is essential for embryonic development. *MECP2* gene mutations are the cause of most cases of Rett syndrome, a progressive neurologic developmental disorder and one of the most common causes of mental retardation in females.
COMPOUND HETEROZYGOUS variants			
HERC1 *HECT and RLD domain containing E3 ubiquitin protein ligase family member 1*	Both variants are missense variants Position variant 1: chr15: 63950887 ENST00000443617.2: c.9455G>T ENSP00000390158.2:p.Ser3152Phe dbSNP: rs2228513 (paternally inherited) Position variant 2: chr15: 64041905 ENST00000443617: c.1988C>G ENSP00000390158.2: p.Ala663Gly dbSNP: rs137926425 (maternally inherited) SIFT variant 1: 0 (deleterious) PolyPhen-2 variant 1: 0.603 (possibly damaging) SIFT variant 2: 0.29 (tolerated) Polyphen-2 variant 2: 0.071 (possibly damaging) Allele 1 frequency: 0.04 Allele 2 frequency: 0.005754 Expressed in fetal brain and adult frontal cortex	Macrocephaly, dysmorphic facies, and psychomotor retardation MDFPMR (OMIM *605109)	Megalencephaly with thick corpus callosum, cerebellar atrophy, and intellectual disability. Compound heterozygous mutations in the *HERC1* gene were found in two siblings with MDFPMR (36). A homozygous truncating mutation in the *HERC1* gene was found in a 18-year-old man with MDFPMR, born of consanguineous parents (37).

Sanger sequencing allowed us to verify and confirm all the variants except *COPB1*. *COPB1* could not be sequenced properly as there is a Poly-T stretch (fourteen T's) just before the variant which made sequencing with the forward primer impossible. We did not attempt resequencing because after confirming the other variants we were able to give a diagnosis to the patient, making further studies unnecessary. After a more thorough literature search we came to the conclusion that the two most promising candidates were *AMPD2* and *MECP2*, respectively the cause of pontocerebellar hypoplasia (OMIM #615809) and Rett syndrome (OMIM #312750).

Two things go against the hypothesis of pontocerebellar hypoplasia in our patient: the observed variant is *de novo* whereas pontocerebellar hypoplasia is recessive, and the patient has no visual impairments, making this condition quite improbable.

On the other hand, Rett syndrome is a progressive neurologic developmental disorder and one of the most common causes of mental retardation in females. The criteria required for Rett syndrome are precise, and include a period of normal development followed by regression, and then stabilisation or recovery. The main criteria are loss of acquired hand skills, loss of acquired language, gait abnormalities and stereotypic hand movements. Based on the diagnostic criteria published by Neul et al. (38), we were able to come to the conclusion that our patient fits the criteria for atypical Rett syndrome (See Supplementary Table 2 for the diagnostic criteria of Rett syndrome). The mutation we found in this patient (c.473C>T, Thr158Met) is the most common one found in Rett syndrome patients, accounting for 8.9% of cases according to RettBASE (39), and it is usually a *de novo* mutation, just like in our patient. Hydrocephalus does not fit with this diagnosis, we therefore hypothesise that it is an independent phenotype in the patient.

Table 3: Primers designed for the candidate genes in family 1

Candidate genes	Forward Primer 5'-3'	Reverse Primer 5'-3'
AMPD2	CTGCGTTAGAGGTGAGGGTGAG	TGAGGATCAGGAGACAGCCAAG
COPB1	GATTGTGCCACTATTCTCTAGCCT	AATGAGAGTGGAGTTAGCGAGATG
MECP2	CTGCCCTATCTCTGACATTGCTAT	GACTTTTCTCCAGGACCCTTTTCA
HERC1	AGCCAACAATCTAATCATCTCAGGA (exon 9) AACCAAGTTCCAGATGAGCAAAAG (exon 48)	CCACATTGCCCCATTGAGTTATTG (exon 9) ATTCTATCTCACCCTCCCAAACAC (exon 48)

Family 2

Clinical description

The second patient is a boy with ID of unknown origin (Picture and family tree in Figure 2). He was born by caesarean section in 2006 and has a younger asymptomatic brother. He was born at term, with a birth weight of 3.5 kg (p50-p70) and a size of 52 cm (p75-p90). At the age of three months, complete blindness and optic nerve atrophy were diagnosed. As he grew up global developmental delay was observed: he could sit independently only when he was one year old and he could walk and say individual words from the age of 2.5. At the age of 9 he could still not produce sentences. He also has severe autism as well as behavioural and psychiatric problems: temper tantrums, labile emotions, echolalia, self-mutilation habits, multiple phobias and a food picking behaviour. To summarise, this patient's most remarkable traits are congenital blindness, severe ID and autism.

His phenotype in 2010 was as follows: height at the p3, weight at the p97 (obesity) and head circumference at the p35. He has small hands and feet with tapered fingers, is hypotonic and has polyneuropathy.

Karyotype, aCGH and biochemical genetic tests were all normal. A brain MRI only revealed corpus callosum hypoplasia.

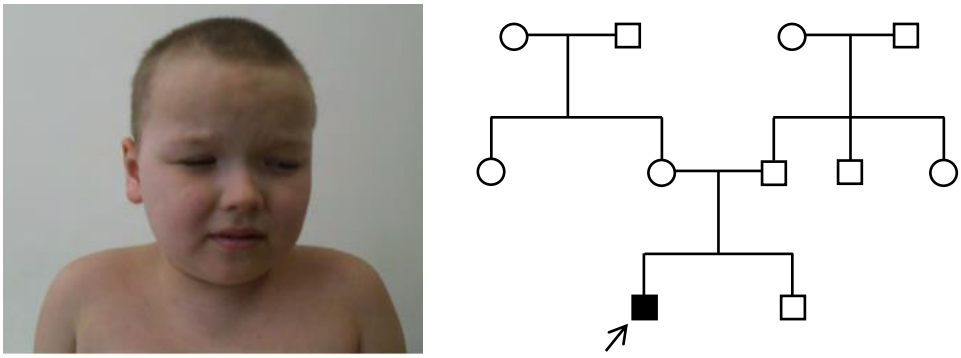


Figure 2: Picture of the proband from Family 2 (left) and family tree (right)

Whole exome sequencing: variants and their analysis

We applied the whole exome sequencing pipeline to all family members. Without applying any filters, we found 129'756 variants among the exomes of all family members. When we use the criteria described in Table 1 we reduced the number of variants to be analysed manually to 4 *de novo*, 2 recessive and 8 compound heterozygous variants (See Supplementary Table 3 for details about all the variants).

Based on their expected influence on neurological traits, we chose to Sanger sequence only three of our candidate variants: *FADS6*, *GRID2IP* and *SYTL2* (Table 4). The primers we designed and used for sequencing are listed in Table 5.

Table 4: Variants selected for Sanger sequencing confirmation for family 2 (selection based on expected neurological phenotypes and other symptoms resembling the patient's caused by the variants; underlined)

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity (34, 35); Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
***DE NOVO* variants**			
FADS6 *fatty acid desaturase 6*	Inframe insertion Position: chr17: 72889649; ENST00000310226.6: c.44_45insTACGGAGCCCATGGAACCTACGGAGCCCATGGAACCTACGGAGCCCATGGAACC; ENSP00000307821.6: p.Pro15_Ala16insThrGluProMetGluProThrGluProMetGluProThrGluProMetGluPro SIFT: N/A PolyPhen-2: N/A Allele frequency: unknown Expressed in brain and other tissues	17p11.2 deletion causes Smith-Magenis Syndrome (OMIM #182290)	The major features of this condition include mild to moderate intellectual disability, delayed speech and language skills, distinctive facial features, sleep disturbances, and behavioural problems.
GRID2IP *Glutamate Receptor, Ionotropic, Delta 2 (Grid2) Interacting Protein*	Missense variant Position: chr7: 6547908 ENST00000457091.2: c.2252C>A ENSP00000397351.2: p.Pro751Gln dbSNP: rs184043502 SIFT: 0.02 (deleterious) PolyPhen-2: 0.403 (benign) Allele frequency: 0.012 Highly expressed in cerebellum	-	Glutamate receptor delta-2 (GRID2) is predominantly expressed at parallel fiber-Purkinje cell synapses and plays crucial roles in synaptogenesis and synaptic plasticity. GRID2IP1 interacts with GRID2 and may control GRID2 signalling in Purkinje cells. Interesting paralog: FMN2 (Formin 2) Diseases associated with *FMN2* include mental retardation (autosomal recessive) and autosomal recessive non-syndromic intellectual disability.

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity (34, 35); Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
COMPOUND HETEROZYGOUS variants			
SYTL2 *synaptotagmin like 2* *other names of the gene: SLP2, EXOPHILIN4, KIAA1597*	Both variants are missense variants	-	The SLP homology domain (SHD) of this protein has been shown to specifically bind the GTP-bound form of Ras-related protein Rab-27A (RAB27A). This protein plays a role in RAB27A-dependent vesicle trafficking and controls melanosome distribution in the cell periphery. Griscelli syndrome type 1 (GS1) (OMIM #214450): hypomelanosis with a primary neurologic deficit and without immunologic impairment or hemophagocytic syndrome is caused by mutations in the *MYO5A* gene. Griscelli syndrome type 2 (GS2) (OMIM #607624), with immune impairment, is caused by mutation in the *RAB27A* gene. Griscelli syndrome type 3 (GS3) (OMIM #609227), characterised by hypomelanosis with no immunologic or neurologic manifestations, can be caused by mutations in the Melanophilin (*MLPH*) or sometimes also *MYO5A* genes.
	Position variant 1: chr11: 85420401 ENST00000359152.5: c.4511C>T ENSP00000352065.5: p.Ala1504Val (G37) ENSP00000489269.1: p.Ala1947Val (G38) dbSNP: rs61740616 (maternally inherited) Position variant 2: chr11: 85438710 ENST00000359152.5: c.362A>G ENSP00000352065.5: p.Asp121Gly (G37) ENSP00000489269.1: p.Asp564Gly (G38) dbSNP: rs61908775 (paternally inherited)		
	SIFT variant 1: 0.1 (tolerated) PolyPhen-2 variant 1: 0.4 (benign) SIFT variant 2: 0 (deleterious low confidence) PolyPhen-2 variant 2: 0.774 (possibly damaging)		
	Allele 1 frequency: 0.0247 Allele 2 frequency: 0.0251		
	Expressed in some brain regions (cortex, amygdala)		

Table 5: Primers designed for the candidate gene in family 2

Candidate genes	Forward Primer 5'-3'	Reverse Primer 5'-3'
FADS6	CCTACTACCTGGGCTGGGT	GCAAGGCGAAGAGGCTGA
GRID2IP	GAGGCTAACCCGTGAGATGG	TCATGTGGCTGGTCTCACTG
SYTL2	GTGTTGAGCCTGAGCCATCT (exon 1) TCCATGATCCTTGTGCTAGGT (exon 6)	CTTTGGGGTGCCTTTTTGGG (exon 1) CCCCCAACTTTTCAAACATTCCT (exon 6)

Only the *SYTL2* variants were confirmed by Sanger sequencing and the other ones were discarded as false positives. *SYTL2* was not a very strong candidate at this point, as the variants have an allele

frequency of around 2% and the pathogenicity prediction was rather mild, but we had selected it because our lab is currently confirming *SYTL3* as the cause of ID in three families (unpublished and confidential data). We did some additional literature reviewing to see if we could strengthen this candidate by finding out more about its function and the pathways in which it is involved.

SYTL2 is a synaptotagmin-like protein (SLP), and it belongs to the family of C2-domain containing proteins (40). It owes its name to its similarity with synaptotagmins, which are C2-domain proteins that regulate membrane traffic and are involved in vesicle fusion (41). They have been shown to regulate the Ca^{2+}-triggered exocytosis of neurotransmitters via their conserved C2 motifs (C2A and C2B) which serve as Ca^{2+} or effector binding domains and Ca^{2+}-sensing modules (42). SYTL2 itself contains two C2 domains. It also contains a SLP homology domain (SHD) at the N-terminus which allows it to specifically bind to the GTP-bound form of Ras-related protein Rab-27A (RAB27A), of which SYTL2 is an effector protein (43). RAB27A plays a versatile role in regulating exocytosis and processes involving lysosome-related organelles (for example melanosomes). It has been shown to be critical in melanocytes, at the immunological synapse and in pancreatic β cells (44). The function of SYTL2 has only been partially understood, but it is known to be linked with vesicle trafficking and exocytosis through its interaction with RAB27A.

Furthermore, RAB27A is known to cause one subtype of Griscelli syndrome, a condition that displays neurologic impairments in some cases. Griscelli syndrome comes in three subtypes (OMIM #214450, #607624, #609227), and may be caused by mutations in at least three different genes. The signs and symptoms vary depending on which gene is causing the disease, but what they all have in common is a hypopigmented phenotype, pigment distribution anomalies in the hair shaft and an excess of melanosomes in melanocytes (see Supplementary Table 4 for the details of Griscelli syndrome). In the type 1 syndrome, caused by mutations in *MYO5A* (OMIM #214450), neurologic manifestations are predominant, and are accompanied by ocular anomalies. In the type 2 syndrome, caused by mutations in *RAB27A* (OMIM #607624), immunological manifestations are the most prominent, and there may be a neurological phenotype. The type 3 syndrome (OMIM #609227) presents with a hypopigmented phenotype only and is caused by mutations in *MLPH*. All three subtypes display an alteration of melanosome transport in melanocytes, which is explained by the fact that all three proteins involved are involved in melanosome transport. Indeed, they have been found to form a tripartite protein complex, with MLPH acting as a linker between RAB27A and MYO5A (45). Melanophilin (*MLPH*, also named *Slac2-a,* SLP homologue lacking C2 domains-c) belongs to the same protein family as SYTL2 and like SYTL2 and SYTL3, it is an effector protein of RAB27A and binds to it via its SHD domain.

The next step is to check whether our patient's symptoms match those caused by Griscelli syndrome. Upon asking further questions to our colleagues in Lithuania we learned that the patient does not seem to have pigmentation anomalies (however the patient is of a light complexion and no specific testing has been undertaken) and did not suffer from recurrent infections or other episodes that would suggest an immunological problem (there were no acute episodes of fever, no cytopenias, no hepato- or splenomegaly, no hemophagocytic episodes), thus virtually ruling out a type 2 *Griscelli-like* syndrome hypothesis. The type 1 Griscelli syndrome, with a strong neurologic involvement

(hypotonia, seizures, ID) and ocular anomalies comes the closest to explaining our patient's condition. Griscelli syndrome type 1 is caused by mutations in Myosin-Va, a protein that is involved together with RAB27A in vesicle trafficking, especially through their mutual interaction with MLPH (45). As SYTL2 binds to RAB27A as well, we can thus hypothesise that SYTL2 could also be involved in that pathway and that its disruption could result in a *Griscelli-like* syndrome.

To go one step further in exploring this idea, it would be interesting to formally test if the other features of the classical Griscelli syndrome are also present in our patient, for example by checking if pigment distribution in the patient's hair is similar to that observed in patients with Griscelli syndrome.

Another important step is to confirm the protein-disruptive nature of the variants we found in *SYTL2*. To do this, we need to predict and analyse the structural alterations they cause in the protein structure. This can be done by using crystallographic data from the protein of interest to model the protein in 3D and make an educated guess at the structural changes triggered. It is also crucial to know what and where the conserved domains of the protein are. SYTL2 indeed contains an N-terminus SHD domain, required for binding to RAB27A, and two C2 domains, located between amino acids 1490-1580 and 1639-1727 (46).

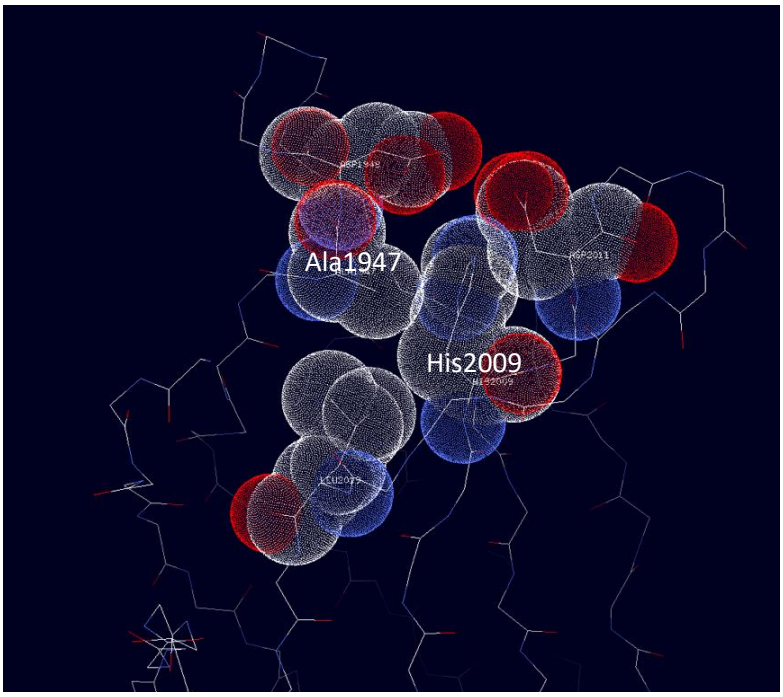
C2 domains are involved in targeting proteins to cell membranes. The C2 domains are also putative Ca^{2+} binding-sites (45). SYTL2 being involved in exocytosis, the C2 domains may be important for this function. The first variant, p.Ala1504Val, is located in a C2 domain and it induces the change of an Alanine into a Valine. Both amino acids have a hydrophobic side chain, but Valine is significantly bulkier than Alanine. The 3D protein structure prediction (See Figure 3 and Figure 4, kindly provided by Nicolas Guex from the SIB) was done by using the C2-domain structure of SYTL4 (PDB ID: 3FDW)(47) which is the closest data available. It shows that having a Valine instead of an Alanine at position 1947 (position 1504 in GRCh37[5]) could force the Histidine in position 2009 to be pushed away from it, causing a change in the C2 domain and possibly disrupting protein function. Another encouraging sign is that this variant is located in the same region as the mutations our lab has previously uncovered to be pathogenic in *SYLT3*.

The second variant, p.Asp121Gly, brings about the change of an Aspartate into a Glycine residue. While Glycine is not charged and consequently hydrophobic, Aspartate is negatively charged and rather bulky. We can imagine that this could be the basis for a change in protein structure, but we cannot conclude anything because there is no crystallographic data available on this region. However, this amino acid is highly conserved among mammals (See Figure 5), suggesting that it may be of importance.

The protein modelling being inconclusive and in itself not a full proof of pathogenicity, the next step would be to set up a functional study to analyse the disrupted protein's ability to interfere with the vesicle trafficking pathway itself.

[5] This project was done using the GRCh37 genome assembly, but the protein structure prediction was done using the GRCh38.

Figure 3: (left) 3D Model of the protein structure of wild-type SYTL2, highlighting Alanine #1947, the amino acid which is altered by the first variant; (right) chemical structure of Alanine



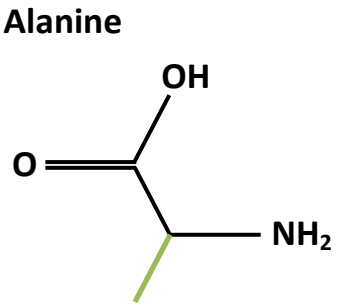
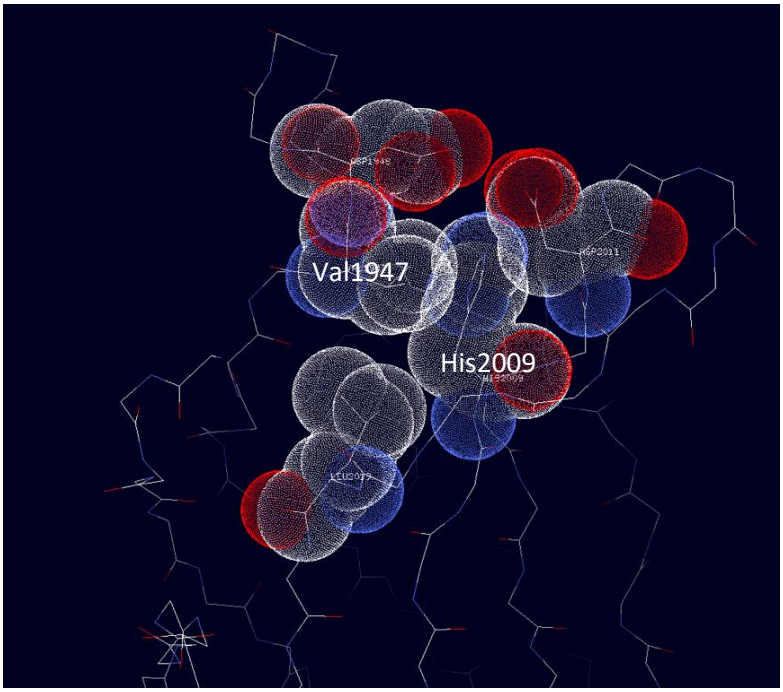


Figure 4: (left) 3D Model of the p.Ala1947Val variant of SYTL2, highlighting the bulkier Valine #1947 and its 3D relationship with Histidine #2009, whose position could be affected by the change in amino acid, thus altering protein function; (right) chemical structure of Valine



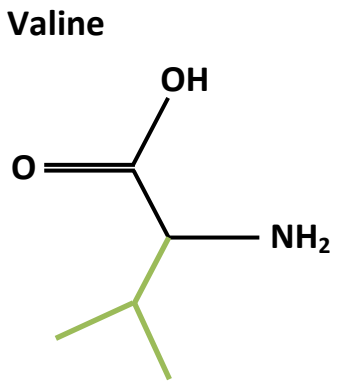


Figure 5: Amino acid alignment of SYTL2 showing that Aspartate #121 (highlighted) is conserved among mammals

To prove that it is a defect in *SYTL2* that is causing ID, it would be interesting to find other patients with symptoms similar to those exhibited by our patient as well as variants in *SYTL2*. To try to find these, we used an online tool for patient-data sharing called Genematcher (28). We made contact with three researchers/clinicians who had studied patients with *SYTL2* variants suspected to be involved in the observed condition of their patient. Here is a short description of the cases:

- Patient A had congenital hypertonia of the limbs, respiratory distress at birth requiring neonatal resuscitation and died aged 3 months. A left retrocerebellar arachnoid cyst was also discovered. A homozygous variant in *SYTL2* (NM_001162951.1:c.2000C>T; p.Pro867Leu; not described in ExAC) was suspected to be involved in the symptoms. (Contact person: Mirna Assoum, University of Burgundy, France)

- Patient B is an adult case with a chiefly immunological phenotype featuring an impairment of lymphocyte function: impaired NK cell function, impaired CD107a mobilization, increased CD163a and hyperferritinemia with splenomegaly and thrombocytopenia. The patient also exhibits hyperpigmentation and minimal neurologic features: anxiety and learning disability. The variants they found follow a hypothesis of compound heterozygosity and are the following: c.458G>A (p.Gly153Glu; MAF=0.001470) and c.285C>A (p.Phe95Leu; not described). (Contact Person: Andre Mattman, Vancouver General Hospital, Canada)

- Patient C has an immune deficiency, with no further description. (Contact person: Megan Cho, GeneDx, USA)

None of the patients described here presents a phenotype that is very similar to that of our patient from family 2. Nevertheless, there are features in patients B and possibly C that go in the direction of a Griscelli-like syndrome, which encourages us to think –as we hypothesised– that SYTL2, via its interaction with RAB27A, can modify the vesicle trafficking pathway and could therefore cause the observed symptoms. However, we cannot so far conclude anything definite from these findings.

Conclusion and Discussion

Whole exome sequencing is a powerful technique, as it allows almost all the exons of the genome in an unbiased manner. As such, it has the potential to greatly increase the diagnostic yield for patients with ID. Based on a recent study by Hamdan et al. (48), the diagnostic yield for *de novo* variants in genes known to cause ID should be around 29%. Moreover, in this study they found possibly pathogenic *de novo* variants in an extra 29% of cases, in genes not previously causally linked to ID. In our project, WES allowed us to find the cause of ID with certainty in one out of two patients, and it was a *de novo* variant.

Indeed for the first family, the variant in the *MECP2* gene that was revealed by WES can be established as causing ID in the patient because it has already been found in others patients with the same syndrome. As such, an expert clinician might have been able to recognise the clinical traits and correctly guess at the gene involved. In this case, a simple Sanger sequencing of the patient's variant of the gene would have been a cheaper alternative.

In the second family, we were able to find a candidate gene that could be the cause of ID, but we were unable to confirm it. Moreover, as our selected variants have a rather high frequency in the general population (about 2% in both cases) and as there are people in the ExAC database who have been found to be homozygous for the variants, it diminishes the chances of their being pathogenic. We lowered the cut-off for the CADD score to 5 to try to find additional variants in this gene of interest. We found one variant that is maternally inherited and has a much lower allele frequency and for which no homozygotes have been found: rs78415563, c.1206C>A, p.Asp402Glu[6], allele frequency 0.008301. It still needs to be confirmed by Sanger sequencing.

In the case of the second family the limitations of WES are very clear: the technique allows us to suspect a gene but not to confirm its pathogenicity. There are several ways of overcoming this weakness. The best way to confirm our suspicion that this variant is the cause of ID would be to find a larger cohort of patients all harbouring the same phenotype as well as variants in the same gene. This is a difficult and demanding step, and one way of going around it would be to perform a functional study of the mutated gene to further investigate and validate the mechanism of pathogenicity. The functional study could be performed on the patient's biological material (cells or tissues) or using cellular or animal models (mice, zebrafish). These steps are however out of the scope of this Master's project.

[6] ENST00000359152.5: c.1206C>A, ENSP00000352065.5: p.Asp402Glu

Acknowledgements

I would like to thank Professor Reymond for welcoming me in his lab for this Master's project, and for the interesting discussions we had about the project and research in general. Many thanks to Lucie Gueneau who guided me through this project and read this master's thesis thrice! My acknowledgements also go to the SIB team who designed VarApp, the application I used throughout the project, and Nicolas Guex, who kindly lent us a hand with the protein structure predictions. Thanks also to all the members of the Reymond lab, who answered my questions and gave their help with a smile. I also thank my family and especially my mother, who did the painstaking job of reading this memoire many times even though her major was mathematics and not life sciences.

Bibliography

1. Karam SM, Riegel M, Segal SL, Félix TM, Barros AJD, Santos IS, et al. Genetic causes of intellectual disability in a birth cohort: A population-based study. American Journal of Medical Genetics Part a. 2015;167(6):1204-14.

2. Definition of intellectual disability from the American Association on Intellectual and Developmental Disabilities. Available from: https://aaidd.org/intellectual-disability/definition.

3. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. Nat Rev Genet. 2016;17(1):9-18.

4. Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. Trends in neurosciences. 2014;37(2):95-105.

5. Rudkin GT, Stollar BD. High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. Nature. 1977;265(5593):472-3.

6. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America. 1977;74(12):5463-7.

7. Kallioniemi A, Kallioniemi O, Sudar D, Rutovitz D, Gray J, Waldman F, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science. 1992;258(5083):818-21.

8. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. Nat Meth. 2011;8(1):61-5.

9. Wang Q, Shashikant CS, Jensen M, Altman NS, Girirajan S. Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. Scientific Reports. 2017;7(1):885.

10. Performance limitations of clinical exome sequencing. CLN Stat. 2014.

11. Park JY, Clark P, Londin E, Sponziello M, Kricka LJ, Fortina P. Clinical Exome Performance for Reporting Secondary Genetic Findings. Clinical chemistry. 2015;61(1):213-20.

12. Griffin HR, Pyle A, Blakely EL, Alston CL, Duff J, Hudson G, et al. Accurate mitochondrial DNA sequencing using off-target reads provides a single test to identify pathogenic point mutations. Genet Med. 2014;16(12):962-71.

13. Diroma MA, Calabrese C, Simone D, Santorsola M, Calabrese FM, Gasparre G, et al. Extraction and annotation of human mitochondrial genomes from 1000 Genomes Whole Exome Sequencing data. BMC Genomics. 2014;15(3):S2.

14. Aronesty E. ea-utils : "Command-line tools for processing biological sequencing data" 2011. Available from: https://github.com/ExpressionAnalysis/ea-utils.

15. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754-60.

16. DePristo MA, Banks E, Poplin RE, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43(5):491-8.

17. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010;20(9):1297-303.

18. Delafontaine J, Masselot A, Liechti R, Kuznetsov D, Xenarios I, Pradervand S. Varapp: A reactive web-application for variants filtering. bioRxiv. 2016.

19. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, et al. Ensembl 2016. Nucleic Acids Research. 2016;44(D1):D710-D6.

20. Coordinators NR. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research. 2016;44(Database issue):D7-D19.

21. Online Mendelian Inheritance in Man, OMIM®. Available from: World Wide Web URL: https://omim.org/.

22. Gene cards. Human gene Database. Available from: www.genecards.org.

23. MalaCards. Available from: http://www.malacards.org/.

24. The Genotype-Tissue Expression (GTEx) Project. Available from: https://gtexportal.org/home/.

25. Greene CS, Krishnan A, Wong AK, Ricciotti E, Zelaya RA, Himmelstein DS, et al. Understanding multicellular function and disease with human tissue-specific networks. Nature genetics. 2015;47(6):569-76.

26. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. The American Journal of Human Genetics. 2009;84(4):524-33.

27. Orphanet: an online database of rare diseases and orphan drugs. Copyright, INSERM 1997. Available at http://www.orpha.net.

28. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. Human Mutation. 2015;36(10):928-30.

29. Fajardo KVF, Adams D, Program NCS, Mason CE, Sincan M, Tifft C, et al. Detecting false positive signals in exome sequencing. Human Mutation. 2012;33(4):609-13.

30. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134-.

31. Korbie DJ, Mattick JS. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protocols. 2008;3(9):1452-6.

32. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.

33. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics. 2014;46(3):310-5.

34. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protocols. 2009;4(8):1073-81.

35. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7.

36. Ortega-Recalde O, Beltrán OI, Gálvez JM, Palma-Montero A, Restrepo CM, Mateus HE, et al. Biallelic HERC1 mutations in a syndromic form of overgrowth and intellectual disability. Clinical Genetics. 2015;88(4):e1-e3.

37. Nguyen LS, Schneider T, Rio M, Moutton S, Siquier-Pernet K, Verny F, et al. A nonsense variant in HERC1 is associated with intellectual disability, megalencephaly, thick corpus callosum and cerebellar atrophy. European Journal of Human Genetics. 2016;24(3):455-8.

38. Neul JL, Kaufmann WE, Glaze DG, Christodoulou J, Clarke AJ, Bahi-Buisson N, et al. Rett Syndrome: Revised Diagnostic Criteria and Nomenclature. Annals of neurology. 2010;68(6):944-50.

39. Christodoulou, J., Grimm, A., Maher, T. and Bennetts, B. (2003), RettBASE: The IRSA MECP2 variation database—a new mutation database in evolution. Hum. Mutat., 21: 466–472. doi:10.1002/humu.10194. Available from: http://mecp2.chw.edu.au/.

40. NCBI website. Available from: https://www.ncbi.nlm.nih.gov/gene/54843.

41. Südhof TC, Rizo J. Synaptotagmins: C₂-Domain Proteins That Regulate Membrane Traffic. Neuron.17(3):379-88.

42. Bai J, Chapman ER. The C2 domains of synaptotagmin – partners in exocytosis. Trends in Biochemical Sciences.29(3):143-51.

43. Kuroda TS, Fukuda M, Ariga H, Mikoshiba K. The Slp Homology Domain of Synaptotagmin-like Proteins 1–4 and Slac2 Functions as a Novel Rab27A Binding Domain. Journal of Biological Chemistry. 2002;277(11):9212-8.

44. Yu M, Kasai K, Nagashima K, Torii S, Yokota-Hashimoto H, Okamoto K, et al. Exophilin4/Slp2-a Targets Glucagon Granules to the Plasma Membrane through Unique Ca(2+)-inhibitory Phospholipid-binding Activity of the C2A Domain. Molecular Biology of the Cell. 2007;18(2):688-96.

45. Fukuda M, Kuroda TS, Mikoshiba K. Slac2-a/Melanophilin, the Missing Link between Rab27 and Myosin Va: IMPLICATIONS OF A TRIPARTITE PROTEIN COMPLEX FOR MELANOSOME TRANSPORT. Journal of Biological Chemistry. 2002;277(14):12432-6.

46. Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing developments at PROSITE. Nucleic Acids Research. 2013;41(Database issue):D344-D7.

47. Bonanno JB, Rutter, M., Bain, K.T., Miller, S., Romero, R., Wasserman, S., Sauder, J.M., Burley, S.K., Almo, S.C. . Crystal structure of a C2 domain from human synaptotagmin-like protein 4. To be published.

48. Hamdan FF, Srour M, Capo-Chichi J-M, Daoud H, Nassif C, Patry L, et al. De Novo Mutations in Moderate or Severe Intellectual Disability. PLoS Genetics. 2014;10(10):e1004772.

Supplementary materials

Supplementary Table 1: List of all the variants for family 1, chosen using the criteria listed in Table 1. *AMPD2, COPB1, MECP2* and *HERC1* were chosen for further analysis, based on the genes' link to neurological phenotypes and other symptoms resembling the patient's (underlined)

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity; Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
***DE NOVO* variants**			
AMPD2 *adenosine monophosphate deaminase 2*	Missense variant	Autosomal recessive spastic paraplegia (OMIM #615686) Pontocerebellar hypoplasia type 9 (OMIM #615809) (both recessive)	AMPD2 is necessary for neurogenesis; known mutations exist but not ours. Pontocerebellar hypoplasia: microcephaly, optic atrophy, seizures, spasticity, global developmental delay.
	Position: chr1: 110167952 ENST00000256578.3: c.281C>T ENSP00000256578.3: p.Pro94Leu dbSNP: rs748496694		
	SIFT: 0.01 (deleterious low conf.) PolypPhen-2: 0.003 (benign)		
	Allele frequency: 0.00002538		
	Expressed in brain		
COPB1 *coatomer protein complex, subunit beta 1*	Splice acceptor variant	Oculocerebrorenal syndrome (OMIM #309000)	Oculocerebrorenal Syndrome (Lowe syndrome) has symptoms including renal insufficiency, abnormality of the renal tubule and proteinuria, congenital bilateral cataract, often with glaucoma, strabismus and cornea lesions. The kidneys suffer from Fanconi syndrome. Brain development is abnormal, with an important delay in psychomotor development including muscular hypotonia, abnormal behaviour, epilepsy and growth retardation; COPB1 is involved in Notochord development (zebrafish).
	Position: chr11: 14501261 ENST00000439561.2: c.1213-1C>T		
	SIFT: N/A PolyPhen-2: N/A		
	Allele frequency: 0.00009077		
	Expressed in brain		
CASKIN1 *CASK Interacting Protein 1*	Expressed in brain	Linked to height	-

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity; Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
GORASP1 *golgi reassembly stacking protein 1*	Not expressed in brain	Social communication problems Linked to cancer	-
CCZ1 *CCZ1 homolog, vacuolar protein trafficking and biogenesis associate*	Not expressed in brain	-	-
MECP2 *methyl CpG binding protein 2*	Missense variant Position: chrX: 153296806 ENST00000303391.6: c.473C>T ENSP00000301948.6: p.Thr158Met dbSNP: rs28934906 SIFT: 0 (deleterious) PolyPhen-2: 1 (probably damaging) Allele frequency: unknown Expressed in brain	Rett syndrome (OMIM #312750); X-linked syndromic ID; Severe neonatal encephalopathy; Autism; Atypical Rett syndrome; Angelmann syndrome; PPM-X syndrome; Bruxism	MECP2 binds to methylated DNA and can repress transcription from methylated gene promoters. *MECP2* is X-linked and subject to X inactivation. MECP2 is essential for embryonic development. *MECP2* gene mutations are the cause of most cases of Rett syndrome, a progressive neurologic developmental disorder and one of the most common causes of mental retardation in females.
RECESSIVE variants			
ZSWIM6 *zinc finger SWIM-type containing 6*	Expressed in brain	Acromelic frontonasal dysostosis (OMIM #603671)	-
BAIAP2L1 *BAI1-associated protein 2-like 1*	Not expressed in brain	Leprechaunism (Donohue syndrome) (OMIM #246200)	Rare disorder characterized by severe insulin resistance, a condition in which the body's tissues and organs do not respond properly to the hormone insulin; associated with diabetes mellitus.

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity; Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
COMPOUND HETEROZYGOUS variants			
HERC1 *HECT and RLD domain containing E3 ubiquitin protein ligase family member 1*	Both variants are missense variants Position variant 1: chr15: 63950887 ENST00000443617.2: c.9455G>T ENSP00000390158.2: p.Ser3152Phe dbSNP: rs2228513 (paternally inherited) Position variant 2: chr15: 64041905 ENST00000443617: c.1988C>G ENSP00000390158.2: p.Ala663Gly dbSNP: rs137926425 (maternally inherited) SIFT variant 1: 0 (deleterious) PolyPhen-2 variant 1: 0.603 (possibly damaging) SIFT variant 2: 0.29 (tolerated) Polyphen-2 variant 2: 0.071 (possibly damaging) Allele 1 frequency: 0.04 Allele 2 frequency: 0.005754 Expressed in fetal brain and adult frontal cortex	Macrocephaly, dysmorphic facies, and psychomotor retardation MDFPMR (OMIM *605109)	Megalencephaly with thick corpus callosum, cerebellar atrophy, and intellectual disability. A compound heterozygous mutation in the *HERC1* gene was found in two siblings with MDFPMR (36). A homozygous truncating mutation in the *HERC1* gene was found in a 18-year-old man with MDFPMR, born of consanguineous parents (37).
HSPG2 *Heparan Sulfate Proteoglycan 2*	Not expressed in brain	Schwartz-jampel syndrome, type 1 (OMIM #255800); Dyssegmental Dysplasia, Silverman-Handmaker Type (OMIM #224410); Hyperglobulinemic purpura	Schwartz jampel syndrome is a disease of bone and muscle. Signs and symptoms may include muscle weakness and stiffness, abnormal bone development, joint contractures, short stature, small, fixed facial features, and eye abnormalities.
CLYBL *Citrate Lyase Beta Like*	Basally expressed in brain	-	-

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity; Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
TGFβ1 *Transforming Growth Factor, Beta 1*	Not expressed in brain	Camurati-Engelmann Disease (OMIM #131300); Cystic fibrosis (OMIM #219700); Keloid scars	Camurati-Engelmann disease is a condition that mainly affects the bones: increased bone density, which can lead to pain in the arms and legs, a waddling walk, muscle weakness, and extreme tiredness. An increase in the density of the skull results in increased pressure on the brain and can cause a variety of neurological problems, including headaches, hearing loss, vision problems, vertigo, tinnitus, and facial paralysis. The added pressure that thickened bones put on the muscular and skeletal systems can cause scoliosis, joint contractures, knock knees, and flat feet.

Supplementary Table 2: Rett syndrome diagnostic criteria

Rett syndrome diagnostic criteria, adapted from (38) **(criteria matching the patient are in bold)**
Required for typical or classic RTT 1 A period of regression followed by recovery or stabilization 2 All main criteria and all exclusion criteria 3 Supportive criteria are not required, although often present in typical RTT
Required for atypical or variant RTT **1 A period of regression followed by recovery or stabilization** **2 At least 2 out of the 4 main criteria** **3 5 out of 11 supportive criteria**
Main Criteria **1 Partial or complete loss of acquired purposeful hand skills** 2 Partial or complete loss of acquired spoken language **3 Gait abnormalities: Impaired (dyspraxic) or absence of ability** **4 Stereotypic hand movements** such as hand wringing/squeezing, clapping/**tapping, mouthing** and washing/rubbing automatisms
Exclusion Criteria for typical RTT 1 Brain injury secondary to trauma (peri- or postnatally), neurometabolic disease, or severe infection that causes neurological problems 2 Grossly abnormal psychomotor development in first 6 months of life
Supportive Criteria for atypical RTT **1 Breathing disturbances when awake** **2 Bruxism when awake** 3 Impaired sleep pattern **4 Abnormal muscle tone** 5 Peripheral vasomotor disturbances **6 Scoliosis/kyphosis** **7 Growth retardation** **8 Small cold hands and feet** 9 Inappropriate laughing/screaming spells 10 Diminished response to pain **11 Intense eye communication - "eye pointing"**

Supplementary Table 3: List of all the variants for family 2, chosen using the criteria listed in Table 1. FADS6, GRID2IP and SYTL2 were chosen for further analysis, based on the genes' link to neurological phenotypes and other symptoms resembling the patient's (underlined)

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity; Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
***DE NOVO* variants**			
FADS6 *fatty acid desaturase 6*	Inframe insertion Position: chr17: 72889649; ENST00000310226.6: c.44_45insTACGGAGCCCATGGAACCTACG GAGCCCATGGAACCTACGGAGCCCATGGAA CC; ENSP00000307821.6: p.Pro15_Ala16insThrGluProMetGluProThr GluProMetGluProThrGluProMetGluPro SIFT: N/A PolyPhen-2: N/A Allele frequency: unknown Expressed in brain and other tissues	17p11.2 deletion causes Smith-Magenis Syndrome (OMIM #182290)	The major features of this condition include mild to moderate intellectual disability, delayed speech and language skills, distinctive facial features, sleep disturbances, and behavioural problems.
COL5A3 *collagen, type V, alpha 3*	Expressed in brain and other tissues	Mutations in this gene are thought to be responsible for the symptoms of a subset of patients with Ehlers-Danlos syndrome type III (OMIM % 130020)	The Ehlers-Danlos syndromes (EDS) are a group of heritable connective tissue disorders that share the common features of skin hyperextensibility, articular hypermobility, and tissue fragility.
CAPNS1 *calpain, small subunit 1*	Expressed in brain less than in other tissues; pseudogene expressed in brain	Involved in cellular functions including apoptosis, proliferation, migration, adhesion, and autophagy. Calpains have been implicated in neurodegenerative processes, such as myotonic dystrophy.	Myotonia, muscular dystrophy, cataracts, hypogonadism, and ECG changes.

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity; Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
GRID2IP *Glutamate Receptor, Ionotropic, Delta 2 (Interacting Protein*	Missense variant Position: chr7: 6547908 ENST00000435185.5: c.1700C>A ENSP00000408364.1: p.Pro567Gln dbSNP: rs184043502 SIFT: 0.03 (deleterious) PolyPhen-2: 0.994 (probably damaging) Allele frequency: 0.012 Highly expressed in cerebellum	-	Glutamate receptor delta-2 (GRID2) is predominantly expressed at parallel fiber-Purkinje cell synapses and plays crucial roles in synaptogenesis and synaptic plasticity. GRID2IP1 interacts with GRID2 and may control GRID2 signalling in Purkinje cells. Interesting paralog: FMN2 (Formin 2) Diseases associated with FMN2 include mental retardation (autosomal recessive) and autosomal recessive non-syndromic intellectual disability.
RECESSIVE variants			
TMEM57 *transmembrane protein 57*	Expressed in brain, upregulated in cerebellum	Acrofacial Dysostosis 1, Nager Type, also known as Nager syndrome (OMIM #154400)	Related to dysostosis and tetralogy of fallot, and has symptoms including micrognathia, hearing impairment and downslanted palpebral fissures. Affiliated tissues include bone, eye and heart.
HLA-DRB1 *Human leukocyte antigen DRB1*	Not expressed in brain	Sarcoidosis; Rheumatoid arthritis; Multiple sclerosis	-
COMPOUND HETEROZYGOUS variants			
LTBP2 *latent transforming growth factor beta binding protein 2*	Very lowly expressed in brain	Primary congenital glaucoma (OMIM #613086); Weill-marchesani syndrome 3 (OMIM #614819).	Weill-marchesani syndrome is an inherited connective tissue disorder that mainly affects the bones and eyes. people with this condition have short stature; short fingers; and limited joint movement, especially of the hands. weill-marchesani syndrome also causes abnormalities of the lens of the eye that lead to severe nearsightedness.

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity; Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
OTOF *otoferlin*	Overexpressed in some brain regions: nucleus accumbens, putamen, caudate nuclei (basal ganglia)	Neurosensory nonsyndromic recessive deafness (OMIM #601071)	-
HEG1 *heart development protein with EGF like domains 1*	Very lowly expressed in brain	-	Crucial regulator of heart and vessel formation and integrity. The zebrafish N-ethyl-N-nitrosourea-induced mutation 'heart of glass' (heg) is characterized by embryonic lethality with a massively enlarged heart.
FAM65B *family with sequence similarity 65 member B*	Very lowly expressed in brain	Autosomal recessive non-syndromic sensorineural deafness type dfnb. (OMIM #616515)	Required for hearing. Involved in skeletal muscle development.
SYTL2 *synaptotagmin like 2* *other names of the gene: SLP2, EXOPHILIN4, KIAA1597*	Both variants are missense variants Position variant 1: chr11: 85420401 ENST00000359152.5: c.4511C>T ENSP00000352065.5: p.Ala1504Val (G37) ENSP00000489269.1: p.Ala1947Val (G38) dbSNP: rs61740616 (maternally inherited) Position variant 2: chr11 : 85438710 ENST00000359152.5: c.362A>G ENSP00000352065.5: p.Asp121Gly (G37) ENSP00000489269.1: p.Asp564Gly (G38) dbSNP: rs61908775 (paternally inherited) SIFT variant 1: 0.1 (tolerated) PolyPhen-2 variant 1: 0.4 (benign) SIFT variant 2: 0 (deleterious low conf.) PolyPhen-2 variant 2: 0.774 (possibly damaging) Allele 1 frequency: 0.0247 Allele 2 frequency: 0.0251	-	The SLP homology domain (SHD) of this protein has been shown to specifically bind the GTP-bound form of Ras-related protein Rab-27A (RAB27A). This protein plays a role in RAB27A-dependent vesicle trafficking and controls melanosome distribution in the cell periphery. Griscelli syndrome type 1 (GS1) (OMIM #214450): hypomelanosis with a primary neurologic deficit and without immunologic impairment or hemophagocytic syndrome is caused by mutations in the *MYO5a* gene. Griscelli syndrome type 2 (GS2) (OMIM #607624), with immune impairment, is caused by mutation in the *RAB27A* gene. Griscelli syndrome type 3 (GS3) (OMIM #609227), characterised by hypomelanosis with no immunologic or neurologic manifestations, can be caused by mutation in the Melanophilin (*MLPH*)

Gene involved (full name in italic)	Impact of variant; Position of the variant; Predicted pathogenicity; Allele frequency (ExAC); Expression in brain (GTEX);	Known associated diseases	Signs and symptoms expected, supplementary information
	Expressed in some brain regions (cortex, amygdala)		or sometimes also *MYO5A* genes.
KLHDC7B *kelch domain containing 7B*	Not expressed in brain	-	-
LRIT3 *leucine-rich repeat, immunoglobulin-like and transmembrane domains 3*	Expressed in brain	Congenital stationary night blindness (OMIM #615058)	Impaired night vision, decreased visual acuity, nystagmus, myopia, and strabismus; The complete form is characterized by the complete absence of rod pathway function, whereas the incomplete form is due to impaired rod and cone pathway function.
TJP2 *tight junction protein 2*	Expressed in brain	Progressive Familial intrahepatic Cholestasis (OMIM #615878); Familial Hypercholanemia (OMIM #607748)	-

Supplementary Table 4: Griscelli syndromes' constellation

Griscelli syndrome type 1 (also known as Elejalde syndrome)	**Griscelli syndrome type 2**	**Griscelli syndrome type 3**
All three subtypes of Griscelli syndromes have in common: hypopigmentation, large granules of melanin in hair shafts (abnormal distribution of pigment), accumulation of melanosomes in melanocytes (all due to alteration of vesicle trafficking)		
Mutation in Myosin-Va (***MYO5A***), on chromosome 15q21. *MYO5A is involved in cytoplasmic vesicle transport. It is abundant in melanocytes and nerve cells.* **Neurologic manifestations** Clinical description: Silvery hair and bronze skin on sun-exposed areas. Large granules of melanin unevenly distributed in the hair shaft. Profound neurologic dysfunction: severe muscular hypotonia, seizures, severe mental retardation. Large spectrum of ocular anomalies. No immune impairment.	Mutation in Ras-related protein Rab-27 (***RAB27A***) on chromosome 15q21. *RAB27A plays a role in cytotoxic granule exocytosis in lymphocytes* **Immunological impairments with or without neurologic manifestations** Clinical description: Partial albinism, large clumps of pigment in the hair shafts and an accumulation of melanosomes in melanocytes. Frequent pyogenic infections, acute episodes of fever, neutropenia and thrombocytopenia. Uncontrolled T lymphocyte and macrophage activation syndrome, often associated with hemophagocytic syndrome Neurologic symptoms such as convulsions during hemophagocytic syndrome could be due to leukocyte infiltration of the brain.	Mutation in Melanophilin (***MLPH***) on chromosome 2 (and sometimes mutations in *MYO5A*) *MLPH is involved in melanosome transport. It serves as a link between melanosome-bound RAB27A and the motor protein MYO5A.* **Hypopigmentation without any immunologic or neurologic manifestations** Clinical description: Pigmentary dilution of the skin and hair, presence of large clumps of pigment in hair shafts, and accumulation of melanosomes in melanocytes, without other clinical manifestations.