Year : 2019

Identification Of New Genetic Causes Of Syndromic Intellectual Disability

Voisin Norine

Voisin Norine, 2019, Identification Of New Genetic Causes Of Syndromic Intellectual Disability

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_4D5FEE27DD651

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Centre Intégratif de Génomique

Identification Of New Genetic Causes Of Syndromic Intellectual Disability

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Norine VOISIN

Master de Génétique de l'Université de Paris Diderot, France, 2014

Jury

Prof. Bogdan Draganski, Président
Prof. Alexandre Reymond, Directeur de thèse
Prof. Stanislas Lyonnet, expert
Prof. Christel Thauvin-Robinet, expert

Lausanne 2019

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Bogdan	**Draganski**
Directeur·trice de thèse	Monsieur	Prof.	Alexandre	**Reymond**
Expert·e·s	Madame	Prof.	Christel	**Thauvin-Robinet**
	Monsieur	Prof.	Stanislas	**Lyonnet**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Norine Voisin

Master de génétique, Université Paris Diderot - Paris 7, France

intitulée

Identification of new genetic causes of syndromic intellectual disability

Lausanne, le 16 août 2019

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Bogdan Draganski

Acknowledgements

I would like to thank Christel Thauvin-Robinet and Stanislas Lyonnet for accepting to be part of the jury who will evaluate my work and for finding the time in your busy schedule to come all the way from France and be at my PhD defense.

I thank the Faculty of Biology and Medicine for their administrative support and Bogdan Draganski for presiding these important days of my life.

These five years have been an emotional roller-coaster where I went through amazement as well as discouragement, through excitement and deception, and without the support of some people I wouldn't have gone that far.

Of these, I would like to particularly thank Alex, who accepted me in his team and believed in me from the beginning, sending me to represent the lab at various conferences. Thanks to you I discovered many things, professionally as well and personally. You pushed me to think further and question the results of some experiments, but also showed me breath-taking pictures of birds and cats and convinced me to take longer holidays to visit some gems of the Canadian Rockies! Thanks a lot for being such a cool boss!

I would like to thank the people from my lab, who gave me precious feedback and advices when I was stuck and who were always nice. Special thanks to Jacqueline, who shared her office space with me during these 5 years. You were always behind me (literally), supporting and so kind. I don't know what I would have done without you (first I would have never found anything in the lab...). I will miss you a lot!

Some other people already left the lab but will always be close friends and are deeply missed. We will see each other again, in another life... maybe in another country!

Thanks also to our visiting students who filled my batteries with their motivation and energy.

A very special thank you to Iris, my mother from the CIG. Whenever I needed help, advices or just needed to talk to someone, you were there. Your heart is as big as those beautiful mountains and I was very lucky to have you here for the last part of this journey.

Overall, I was extremely lucky to be at the CIG during my PhD where so many talented and kind persons are gathered. I interacted with many of them, who helped me trouble shoot my experiments (Lorène, Pascal), prepare my sample and understand the subtleties of exome sequencing (Floriane, Keith, Johann, Karolina...), who gave me advises to reach my goals (Fred, Vincent), and helped me understand this confusing country (Corinne, Annick). A big thank to Sylvain, Emmanuel and Julien who were patient enough to explain to me the informatic pipelines and answer my dumb questions. And many, many other kind and interesting people I forgot...

During my PhD I was lucky to meet great people outside of the CIG too. People who started as colleagues and collaborators soon became friends and made me love their country and their culture. I was always waiting for the international conferences to see them again! Laima, Egle, Evelina, Ačiū!

I would like to thank my friends, from BSNL and outside, who helped me to have a life after the lab, grow personally and professionally, achieve great projects, climb high mountains and never give up.

Merci à ma famille et à mes amis de France qui ont compris depuis longtemps que j'avais la bougeotte et que j'allais partir, que je n'écris pas souvent même si je pense à eux, et que je les aime fort. Merci de votre soutien et de votre amour. J'espère vous rendre fiers.

Enfin merci à Ludo, pour absolument tout... J'irai où tu iras, mon pays sera toi...

Content

Figures and tables

Figures

Tables

Abstract

Intellectual disability (ID) is a neurodevelopmental disorder that affects 1-3% of the population and represents a major public health issue. ID is highly heterogeneous, regarding the severity of its manifestation, the possible comorbid symptoms, as well as its etiology. While hundreds of genes have already been associated with ID, the genetic causes of many cases still remain elusive. We have decided to use whole exome sequencing (WES) to unravel the genetic causes of unsolved syndromic ID in 23 families.

During my doctorate, I worked in collaboration with clinicians from Lithuania and Italy who collected clinical data of families for which microarrays and Fragile-X tests were negative. While ID was diagnosed for every patient I studied, the presence of additional symptoms made each family a particular case that I had to analyze individually. Candidate variants identified by WES were validated via functional analysis as well as clinical data collection of similar cases through international collaborations.

My work led to the discovery of a new ID syndrome characterized by mesomelic dysplasia, horseshoe kidney and encephalopathic epilepsy, and the description of the underlying molecular mechanism. *De novo* missense variants in the degron motif of *AFF3* are gain-of-function through an accumulation of the encoded protein. This hypothesis is supported by molecular and animal models analysis.

Moreover, I identified a yet unreported *de novo* variant in a known ID gene, *PIK3R2,* in a boy with severe encephalopathy, explaining the molecular mechanism via 3D modeling. I also broadened the phenotypical description of the Aicardi-Goutières syndrome, describing the first inflammatory myopathy case associated with a homozygous missense variant in *TREX1*. I participated in the description of two new syndromes associated with variants in *KIAA1109* and *MAST1*. As exome sequencing of the other families allowed us to identify likely-pathogenic variants or variants of uncertain significance, gathering clinical data of similar cases and functional evidences are warranted.

Overall, my work enabled us to report a genetic diagnosis to multiple patients and their families, providing opportunities to improve their family planning, clinical care and follow-up. Ultimately, a better delineation of the pathological mechanisms involved in neurodevelopmental disorders such as ID will help identifying new therapeutic targets.

Introduction

I. Definition of developmental delay and intellectual disability

Developmental delay (DD) is diagnosed when a child does not acquire the skills associated to the developmental milestones of its age group. These are related to gross and fine motor movements, cognitive function, language/speech, social interaction and personal hygiene. The term DD is preferentially used for children younger than five years of age, as the skills acquisition timing is highly documented, whereas "intellectual disability" (ID) is often used for individuals older than seven years of age when intellectual quotient (IQ) testing is more adequate and dependable[1].

According to the American Association on Intellectual and Developmental Disability (AAIDD), ID, formerly known as "mental retardation", is characterized by significant limitations in intellectual functioning and adaptive behavior, including everyday social, conceptual and practical skills, originating before the age of 18. An extended version of this definition is given in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5[2]) published by the American Psychiatric Association (APA). The APA bases the ID diagnosis on several criteria. Firstly, deficits in intellectual functioning have to be observed such as problem solving, reasoning, abstract thinking or judgment. Mental capabilities can be measured via IQ tests that are standardized and culturally appropriate. Secondly, individuals must show deficits or impairments in adaptive functioning. These are various skills considered as mandatory for a responsible and independent daily living such as communication, social skills, personal independence, adaptive and learning capacities. Lastly, this diagnosis has to be established during the developmental period, namely before the age of 18. Any neurological defect diagnosed after that are called neurocognitive disorder and may result from brain injury due to trauma or infections for instance.

The APA emphasizes the importance of clinical evaluation as well as the use of standardized IQ test. While several IQ test have been established, one of the most commonly used is the Wechsler test which has variations adapted to the age of the person tested: the Wechsler Pre-School & Primary Scale of Intelligence (WPPSI) covers children from 2 years and 6 months to 7 years and 7 months, the Wechsler Intelligence Scale for Children (WISC) covers children from 6 to 16 years old, and the Wechsler Adult Intelligence Scale (WAIS) test covers teenagers from 16 years of age through adulthood.

A distribution of the general population IQ can be established on Wechsler scale, defining categories of intellectual performances (Figure 1). The average IQ for the general population is estimated at 100, although it may vary depending on the social and cultural environment (e.g.

country, social class, access to education). Two standard deviation below the average on the Wechsler scale, corresponding to a score of 70, is considered as significant cognitive impairment.

ID is one of the most prevalent brain disorders as it affects 1-3% of the population[3,4]. As one of the most common causes of referrals to pediatric, neurological and genetic counseling services, it is responsible for 5-10% of health care costs[5,6]. For decades efforts have been made to understand the etiology of ID. Severity level and clinical presentations are broad and only standardized evaluations can lead to comparable clinical reports. However, despite the common guidelines for the evaluation and classification of ID, standard tests are not always conducted as they are long and require a special consulting. Therefore, pediatricians, clinical geneticists and neurologists often diagnose and evaluate the degree of ID according to the global skills of the patient during consultation.

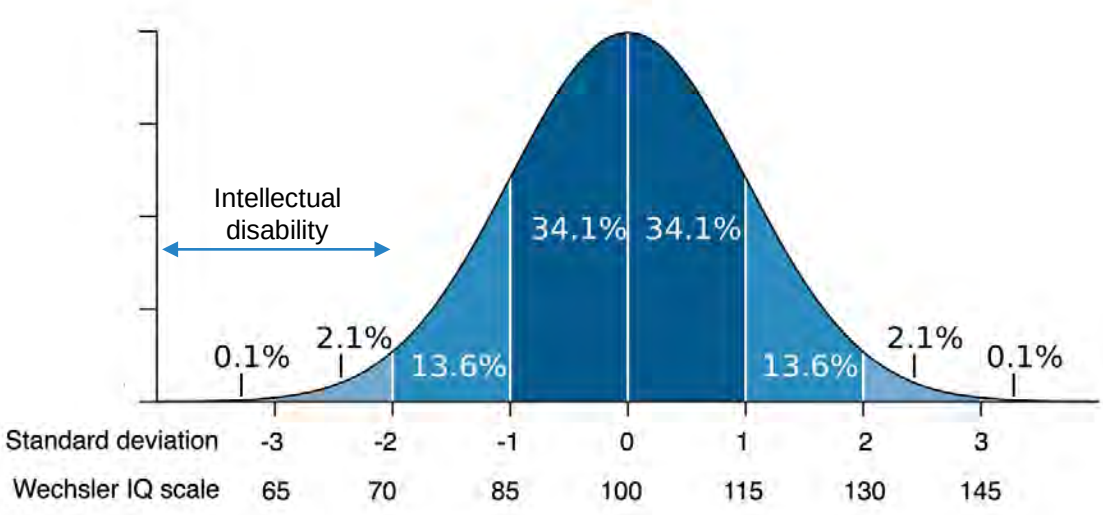


Figure 1: Distribution of the population on Wechsler IQ scale

II. Comorbidities of ID

The spectrum of ID presentation is large by its severity level (IQ score) but it is also broadened by the co-occurrence of additional symptoms. We can thus define two forms of ID: *isolated ID*, when it is diagnosed alone, without any other clinical anomalies; and *syndromic ID*, when it is accompanied by other severe symptoms (Table 1). Isolated forms of ID are rare, as it is difficult to rule out the presence of other more discrete manifestations, while syndromic forms of ID are more frequent. Indeed, comorbidity of neurodevelopment disorders has been established as they originate from developmental defects of the same neurological structures and/or perturbation of common molecular pathways[7]. But ID is also often observed with affected structures outside of the nervous system[8,9](e.g. bones, face, internal organs, etc.). The description of these additional features allows the identification of specific syndromes for which clinical synopses have been established. For instance, particular facial dysmorphisms are associated with distinct syndromic ID and constitute a precious diagnostic tool (e.g. Down, Kabuki, Cornelia de Lange syndromes). Innovative artificial intelligence-based phenotyping softwares (e.g. Face2Gene[10]) have been developed to help prioritize genetic disorders and variants based on the facial features of patients and their similarities with hundreds of known genetic disorders. Additional features can also point towards genes or pathways involved in the affected structures development or function. Consequently, the use of standardized phenotypic description as proposed by the Human Phenotype Ontology[11] (HPO), is highly recommended, to allow comparison of clinical reports and identification of noteworthy similarities and differences[12].

Associated Traits	**Number of genes**
Neurological symptoms	307
Epilepsy	222
Metabolic/mitochondrial anomalies	182
Microcephaly	146
Short stature	143
Eye anomalies	141
Lethality	138
Progression/regression	133
Non-structural anomalies	128
Behavioral anomalies	123

Table 1: Top ten traits associated with 650 ID genes retrieved from OMIM[9]

III. Etiology of ID

While the spectrum of manifestations of ID is wide, the underlying causes are also diverse, and many remain elusive. However, we can distinguish two main origins for ID, that are not mutually exclusive: environmental and genetic factors.

These factors could alter the development of the brain at various stages (Figure 2). Indeed, the fine-tuning and sequential regulation of neuronal proliferation, migration, differentiation, synapses formation and myelination, is crucial for a proper brain development and functioning. As most of neurodevelopmental disorders are associated with neuroanatomical defects, imaging techniques such as Magnetic Resonance Imaging (MRI) are precious tools to pinpoint the affected structures.

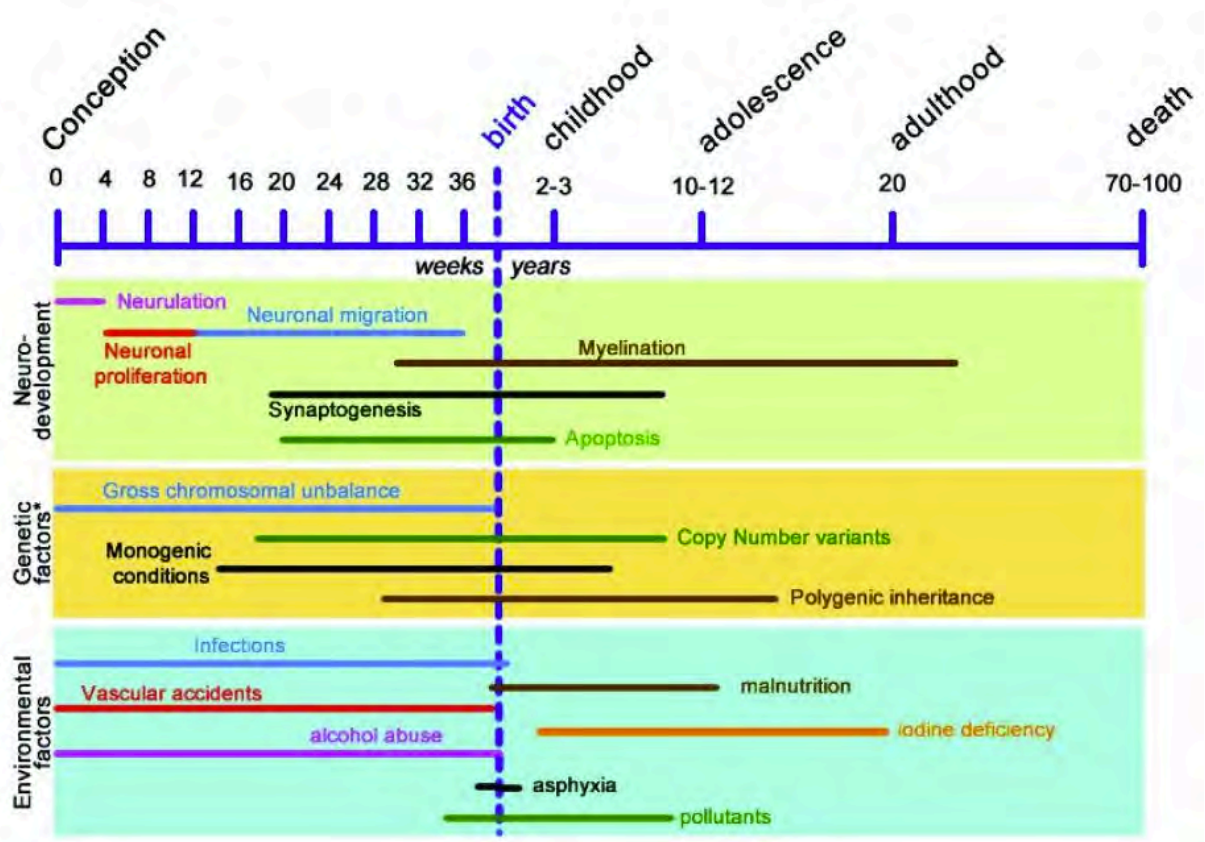


Figure 2: Etiology of ID over neurodevelopmental stages[13]

A. Environmental factors

Many external risks factors are yet to be identified and the prevalence of the known environmental factors for ID is hard to establish as they may vary according to sociocultural background. Moreover, there is no specific diagnostic test and only familial testimony and clinicians' experience prevail.

The most common environmental ID cause according to the World Health Organization (WHO) is the consumption of alcohol by the mother during pregnancy, leading to Fetal Alcohol Syndrome (FAS). Because there is no treatment, prevention campaign efforts raising awareness

have been made these past decades. Other toxic exposures known to be detrimental for neurodevelopment include drugs such as amphetamine or opioids, and endocrine disruptors such as lead or bisphenol A. Similarly, infectious diseases contracted by the mother during early pregnancy like cytomegalovirus infection or rubella, and childhood infections like meningitis, can affect the development. A maternal metabolic disorder such as diabetes, anemia or chronic kidney disease can also lead to developmental defects such as intrauterine growth restriction (IUGR). Finally, vascular accident during pre-natal period or hypoxia at peri-natal stage may be responsible for physical lesions of the brain and traumatic ID. When considering environmental factors, dosage but also timing of exposure is critical as the nervous system is especially sensitive to certain factors at specific time points.

B. Genetic factors

Besides environmental factors, the heritability and the shared characteristics of some ID syndromes pointed toward genetics as the most common ID cause. Historically, the first genetic diagnosis of ID was the identification of trisomy 21 as the cause of Down syndrome[14], since then recognized as the most frequent genetic form of ID[15]. Chromosomal abnormalities such as aneuploidy and large structural variants (over 5Mb) account for up to 15% of genetic ID[16] although chromosomal aneuploidies are often lethal[17]. They were traditionally detected via karyotyping but nowadays, comparative genomic hybridization array (aCGH) is able to detect cytogenetic rearrangements as small as a few kb and has become the first scanning approach for genetic ID diagnosis[18]. Its diagnostic yield is significantly increased when combined with a clinical checklist score[19]. Although aCGH cannot detect balanced rearrangements, inversions and low-level mosaicisms, these are rarely (in <1% of cases) responsible for ID[20]. aCGH improved the discovery and identification of copy number variants (CNVs) that lead to imbalanced genetic dosage. Deletions and duplications of dosage-sensitive regions are associated with many clinical features observed in genomic disorders[21]. Especially, recurrent CNVs explain distinct ID syndromes such as the mirror 16p11.2 deletion and duplication syndromes[22]. However, as CNVs account for 12% of the human genome in the general population[23], it is hard to evaluate their contribution to ID.

The observation of a sex bias in ID patients (1.3-1.4 affected males for 1 female) and the recurrence of large families with only male patients focused an early interest of researchers on the X chromosome. Molecular characterization of X-linked ID started in 1991 with the discovery of expansion of CGG repeats in *FMR1*, also the first microsatellite disorder to be described[24,25]. Extensive study of the X chromosome by international consortiums through familial linkage analysis or translocation studies subsequently followed by the sequencing of its entire coding

regions led to the identification of more than 100 genes implicated in monogenic forms of X-linked ID, explaining up to 10% of ID in males[26].

Research on autosomal causes of ID was initially restricted to the study of CNVs and their dosage-sensitive genes and those disrupted at breakpoints. The limitations of cytogenetic and sequencing technologies as well as the absence of large pedigrees with autosomal ID slowed down the discovery of autosomal ID genes. Autosomal recessive ID (ARID) is traditionally elucidated through the study of linkage and runs-of-homozygosity in large consanguineous families, often coming from North Africa, Middle East and South East Asian countries, were consanguinity is more frequent than in Western countries, China and Japan[27]. Large studies collecting numerous consanguineous families allowed the identification of many novel ID genes[28], highlighting the extreme genetic heterogeneity of ARID. Autosomal dominant ID (ADID) was less investigated as dominant variants were unlikely transmitted from one generation to another, due to the limited fitness of the severely affected individuals, and incomplete penetrance is hard to detect. The idea of spontaneous *de novo* germline mutations as a major cause of ADID emerged from the study of CNVs[19]. The progresses in sequencing technologies[29,30] as well as large-scale projects like the Deciphering Developmental Disorders (DDD) Study recruiting patients with similar developmental disorders[31] allowed the identification of many novel causative genes in rare syndromic ID.

So far, more than 700 genes have been associated with monogenic forms of ID[32], but the discovery curve of new ID genes did not reach a plateau, indicating that there are still ID genes to be identified (Figure 3).

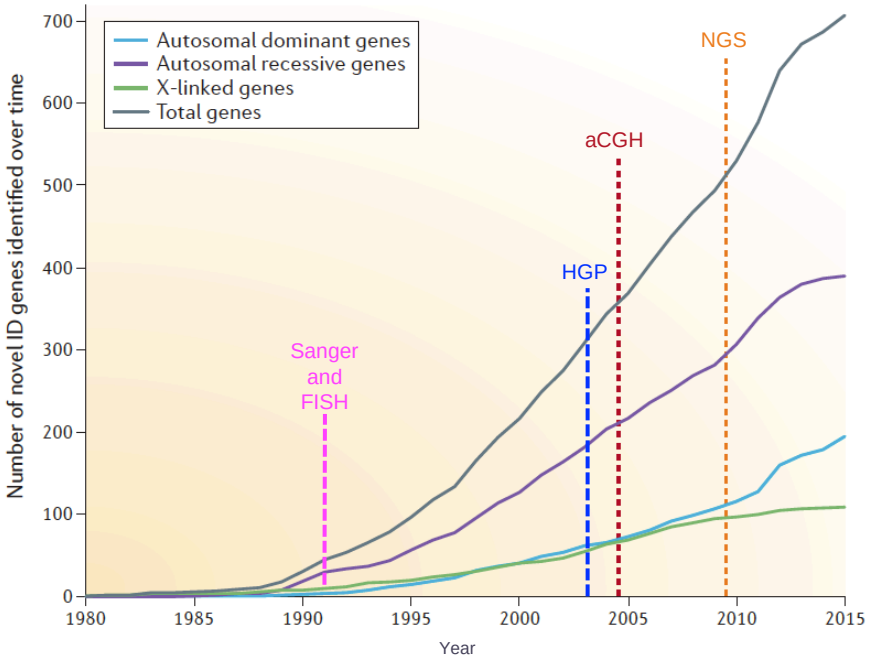


Figure 3: Discovery of novel ID genes over time and technologies development[32]

IV. Next generation sequencing applied to ID

Conventional karyotyping was the routine test in the 70s to detect chromosomal aneuploidy and large rearrangements. It was later implemented with fluorescence in situ hybridization (FISH) and aCGH, allowing the detection of submicroscopic CNVs. Linkage analysis and homozygosity mapping were used for studies involving big cohorts of patients with similar phenotypes, or large families, but they were often defining regions associated with disorders, and hardly single nucleotide variants (SNVs). In the 90s, Sanger sequencing, so far restricted to the targeted sequencing of small fragments and thus the discovery of substitutions and small insertions/deletions in candidate genes or regions already associated with a disorder, was used to decipher the entire human genome. While the Human Genome Project (HGP) took 13 years, it provided a reference sequence to work on, and to compare individuals' sequences. Next generation sequencing (NGS), or high-throughput sequencing, refers to massive parallel sequencing technologies which revolutionized genomic research and in particular the discovery of novel ID genes (Figure 3). NGS technologies detect the incorporated nucleotides as millions of DNA fragments are synthetized at the same time, providing high-quality sequences in record time that then needed to be assembled. As the technologies changed, the bioinformatic field faced new challenges and had to evolve along. Indeed, management, storage and analysis of such unprecedented amount of newly generated data required bioinformatics innovations, development of many dedicated tools, and the emergence of a new era of research.

A major improvement brought by NGS for ID genetic diagnosis is an unbiased approach. The output consists in lists of variations compared to the reference genome, without *a priori* knowledge of ID genes or homozygous region segregating with the disorder in a family. The drawback of this technology lies within its strength: the tremendous amount of data that are then to interpret. While whole-genome sequencing (WGS) refers to the sequencing of nearly the entire human genome and generates sequences that are still hard to interpret like non-coding regions, an alternative was found in the whole-exome sequencing (WES). Cheaper and faster than WGS, it specifically captures approximately 95% of the coding regions of the genome, comprising around 20,000 genes. WES approach is justified by the fact that only 3% of all known disease-causing mutations are located outside of genes[33]. Moreover, various studies showed that it is possible to prioritize pathogenic variants according to their presence in the exomes of patients with overlapping syndromes[34,35] and that trio based (proband and parents) exome sequencing is efficient to identify *de novo* variants in ID and neurodevelopmental disorders[30,36,37]. Pathologies that were previously disregarded given their rarity, clinical and genetic heterogeneity and the absence of multiplex families, have shown high discovery rate using WES. Besides being extremely efficient at discovering new ID genes, WES is also very useful to identify variants in already know ID genes. Using both aCGH and exome sequencing, the DDD Study combined the recruitment of a

large cohort of over 1'000 patients with similar features with a trio-based approach. This meta-analysis provided a diagnosis to 31% of the patients. 28% exhibited variations within previously associated developmental disorder genes and 12 novel developmental-disorder-associated gene were identified[31], proving that a comprehensive strategy is possible and efficient. Although WGS gives a huge amount of (yet) uninterpretable information about the non-coding regions, it has been shown that it also gives a higher quality sequencing of the coding regions compared to WES, with a higher diagnosis rate[30] and a truly unbiased approach.

Over the years, the advancements in sequencing technologies have helped increasing the number of know ID genes (Figure 3). Nowadays, aCGH and targeted sequencing of candidate genes (panels of suspected genes via identification of a known syndrome) are the first approaches to identify the genetic causes of ID[18]. WES is used when the aforementioned tests are inconclusive. The cost of WGS is for now refraining most diagnostic laboratories from using it routinely, restricting its use to genomics research laboratories. Nevertheless, as cost, processing time and data management became more and more affordable, and as our knowledge about non-coding regulatory regions increases, we can predict that WGS will be more used in the future.

V. Variants interpretation

WES detects about 60,000 variants per individuals and identifying the pathogenic one in case of monogenic form of ID can be difficult. To help researchers prioritize some variants, the WES output lists are implemented with information collected on many public databases and prediction software tools (detailed Method p18). For each variant, the coordinates in the human genome has to be retrieved. Official nomenclature following the Human Genome Variation Society (HGVS) guidelines can also be assigned to the variants, facilitating further references. Large-scale sequencing projects in the general population give information about the frequencies of these variations in unaffected individuals and constraint metric for the four different classes of variants (synonymous, missense, loss of function and CNV) can be calculated for each gene, indicating its intolerance to variations[38,39]. The evolutionary conservation of the affected positions and the predicted effects of those changes at the DNA and protein levels can also be scored via various tools[40,41,42,43]. It is crucial to accumulate the maximum number of information possible to try and interpret the variants. In the case of familial or cohort studies, the segregation of the variants with the studied phenotype is also highly informative. According to all these evidences, variants can be classified as possibly pathogenic or benign. When the pathogenicity of a variant is uncertain, it is classified as variant of unknown significance (VUS). When a variant is predicted as pathogenic but localized inside a gene never associated with ID, the gene is considered as a gene of unknown significance (GUS) for this pathology.

Besides computational mapping, frequency and pathogenicity predictors, data mining is required to interpret the shortlisted variants. Firstly, catalogs of genes and their expression profiles, the role of the encoded proteins, their interactants as well as the pathologies associated have to be explored to possibly make a link between some of the shortlisted genes and the phenotypes observed. For instance, genes only expressed in the stomach would not be good candidates for neurodevelopmental disorders while genes expressed in the developing brain and associated with the retinoic-acid pathway would be good candidates for a syndromic form of ID with heart defects. Secondly, collaborative platforms such as GeneMatcher[44] have been created to facilitate the identification of other individuals with similar phenotypes and/or candidate variants within the same genes. These represent important resources to constitute cohorts, compare patients and resolve VUS and GUS. Finally, functional validation can be performed to confirm the pathogenicity of variants. Study of *in-silico* and *in-vivo* models can confirm that a VUS affects the protein function (via knock-in models), and that a GUS can be associated with the observed phenotypes such as brain function and other affected structures in syndromic forms (often via knock-out models).

Aim of this project

Whereas many genetic causes of ID have been discovered, more are still to be unraveled. To identify new genetic causes of syndromic DD/ID, our lab has enrolled around 150 families coming from Lithuania, Italy and Saudi Arabia, with one or two affected individuals. The aCGH screening showed that 12% (17 families) carry probably causative CNVs. We subjected the remaining unsolved cases to exome sequencing. During my doctorate, I have been working in close collaboration with clinicians from Lithuania and Italy. They provided me with clinical reports of 23 families, for which I have performed WES, searching for causative variants. Whereas ID was a shared phenotype among all studied cases, a broad range of other features was associated, making each family a specific syndromic form of ID that needed to be studied independently. My work aimed at giving a genetic diagnosis to the patients and their families, providing opportunities to improve their family planning, clinical care and follow-up. It also aimed to discover new ID genes, describe new syndromes and reach a better understanding of known ones and their underlying molecular mechanisms. Ultimately, a better delineation of the pathological mechanisms involved in neurodevelopmental disorders such as ID will help identifying new therapeutic targets.

Method

I. Patients recruitment

Patients were referred to us by collaborators in Vilnius and Naples upon clinical consultations and diagnosis of syndromic DD/ID of unknown genetic cause. Families where a specific syndrome such as Fragile-X was suspected were Sanger sequenced. When this approach was inconclusive, aCGH was performed. Only the samples for which both aforementioned diagnostics were negative were sent to our laboratory for further investigation. We extracted DNA from peripheral blood samples of the patients and their close relatives (parents and siblings if applicable). During this project, I studied 12 Italian and 11 Lithuanian families with various manifestations of syndromic ID (Table 2).

Family	Number of affected individuals	Number of sequenced exomes	Main features
Italian_1	1	4	ID, seizures, megalencephaly, polymicrogyria, corpus callosum hyperplasia, focal cortical dysplasia
Italian_2	1	4	Hypotonia, short stature, moderate ID, stumbling gait and fine motor skills problems, hyperplasia of corpus callosum and neuronal migration defect
Italian_3	1	5	Epilepsy, encephalopathy, bone deformities, horseshoe kidney, hypertrichosis, facial dysmorphism
Italian_4	1	4	Hypotonia, DD, distal amyotrophy, ataxia, spasticity
Italian_5	1	4	Corpus callosum hyperplasia, speech delay, moderate ID, behavioral problem, hepatomegaly, splenomegaly
Italian_6	1	3	Isolated agenesis of corpus callosum, mild speech delay
Italian_7	1	0	X-linked frontonasal dysplasia like
Italian_8	1	3	DD, speech delay, ASD, corpus callosum hyperplasia
Italian_9	4	5	Cornelia de Lange-like, short stature, maternal inheritance
Italian_10	1	4	Prader-Willy-like, macrocephaly, scaphocephaly, hyperphagia, DD, ASD
Italian_11	1	5	ID, behavioral problems, facial dysmorphism, cryptorchidism, Type 1 Chiary malformation, hepatosplenomegaly
Italian_12	1	3	ID, corpus callosum hyperplasia

Family	Number of affected individuals	Number of sequenced exomes	Main features
Lithuanian_1 (006)	2	4	Mild ID, delayed speech, behavioral problems, wide-spaced teeth, micropenis, obesity, pes planus, hypothyroidism
Lithuanian_2 (009)	1	4	Severe DD, failure to thrive, feeding problems, microcephaly, dysmorphic facial features, cryptorchidism, finger joint hypermotility, atrial septal defect, dextrocardia, chronical heart failure, hemivertebrea, dysplastic ribs, scoliosis, enlarged thymus, tantrums
Lithuanian_3 (040)	2	4	ASD, sleep disorder, no speech, retrognathia, tantrums, hyperkinetic disorder
Lithuanian_4 (057)	1	5	Severe ID, DD, cognitive decline, progressive loss of vision, motor impairment, cerebral atrophy, progressive microcephaly, spasticity, inflammatory myopathy
Lithuanian_5 (093)	2	4	Mild ID, DD, optic nerves and chiasm atrophy, long face, high arched palate, gingival hypertrophy, spaced teeth, clinodactyly, synophrys, scoliosis
Lithuanian_6 (094)	2	4	Pre-eclampsia, speech delay, behavioral problems, hypermetropia, stereotypic movements, additional teeth
Lithuanian_7 (107)	2	4	Achromatopsia, hyperactivity, ID
Lithuanian_8 (112)	1	4	ID, joint hypermotility, cubitus valgus, anxiety, walking on tip-toes
Lithuanian_9 (119)	1	5	ID, no speech, ocular disorder, coordination instability
Lithuanian_10 (048)	2	4	Severe ID, DD, plagiocephaly, epilepsy, hypotonia, arthrogryposis, hypermetropia, brain atrophy, hydrocephaly
Lithuanian_11 (LitX)	3	4	1 stillbirth and 2 terminated pregnancies, severe internal hydrocephalus

Table 2 (above and previous page): Studied families during this project

II. DNA sequencing

A. Exome sequencing and variants annotation

In total, for this project, exomes of 90 individuals were captured using the Agilent SureSelect Human All Exon V5 enrichment kit and multiplex sequenced (6-plex) on an Illumina HiSeq 2500 platform to reach 80 to 120-fold coverage on average. Purity-filtered reads were adapters and quality trimmed with FastqMcf[45], and aligned to the reference human genome (GRCh37/hg19 and decoy sequence) using BWA-MEM[46]. The following steps were performed on each sample separately: (1) duplicate marking using Picard tools,(2) Indel realignment with GATK, (3) base quality score recalibration, (4) variants calling with GATK HaplotypeCaller in gVCF mode[47,48]. Samples were then pooled for genotyping using GATK GenotypeGVCFs and variants were filtered using GATK Variant Quality Score Recalibration[48] (Figure 4).

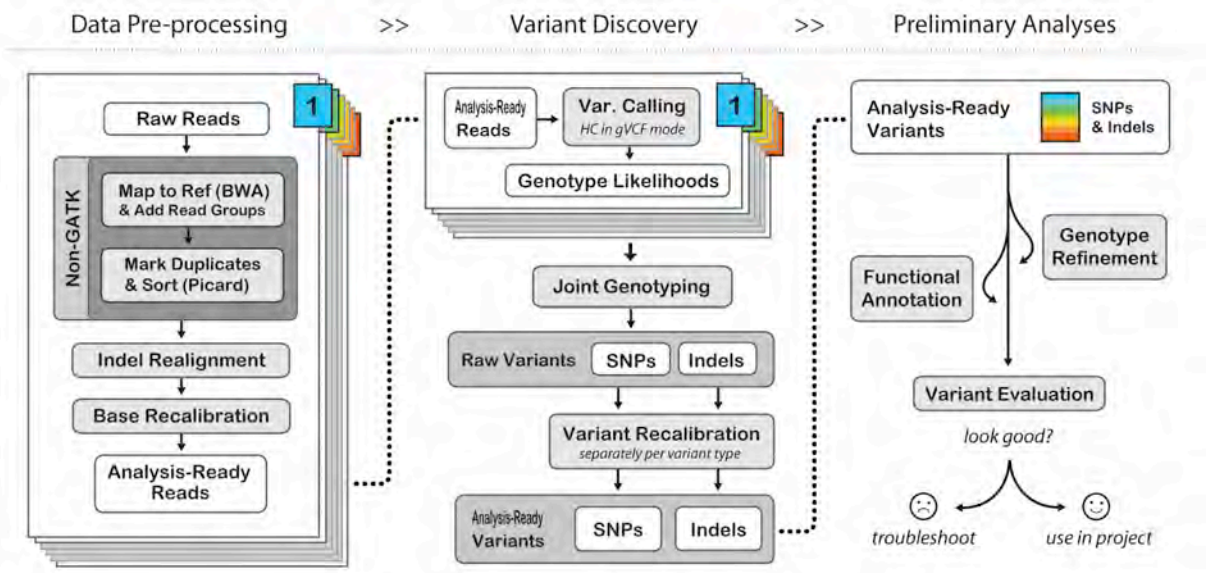


Figure 4: Command-line tools for processing biological sequencing data[45]

B. Variants filtering

Analysis-ready variants in variant call format (VCF) files were imported on Varapp software[49] for further annotation and filtering. Varapp is a tool developed in-house and assessed with our data for software implementation and troubleshooting. Calling several databases, Varapp provides an accurate nomenclature of the variants based on HGVS and output frequency information as well as pathogenicity predictions from various sources and calculators. Variants were filtered based on inheritance patterns including autosomal recessive, X-linked and *de novo*/autosomal dominant. Lists of variants for each transmission mode were filtered automatically based on:

- the quality of the variants: Quality filter PASS keeps the variants above the threshold defined by the VQSLOD value above which 99.9% of the variants are included; Quality score ≥50 for the assertion made in ALT from VCF QUAL field; Fisher Strand bias ≤20 to reduce reads number bias between the strands;
- their frequency in the population: ExAC[38]/GnomAD[39] frequency ≤2%;
- their predicted pathogenicity: High (exon deleted, frameshift, splice acceptor, splice donor, start loss, stop gain, stop loss, nonsynonymous start, transcript codon change (CC), rare amino-acid, chromosomal large deletion); Medium (nonsynonymous coding, inframe codon gain, inframe codon loss, inframe CC, CC deletion, CC insertion, 5'UTR deletion, 3'UTR deletion, splice region, mature miRNA, regulatory region (RR), RR ablation, RR amplification, transcription factor binding site (TFBS), TFBS ablation, TFBS amplification); CADD[50] scaled ≥10 (a Phred-like scoring not applicable to indels; top 10% of all deleterious variants have a CADD scaled score of 10 or greater).

The filters used here are less stringent than usual practice in order to reduce the number of false negative (variants filtered out despite being real and potentially pathogenic). The number of false positive is limited by the manual annotation that follows and during which more stringent thresholds (e.g. predicted deleteriousness) are applied for variants ranking, but without loss of information.

C. Variants interpretation and validation

To interpret the variants after automated filtering, we considered all the evidences supporting or invalidating the pathogenicity of a variant in a specific gene and its link with the observed phenotypes (Figure 5). Manual curation through databases mining included:

- predicted deleteriousness of the variant (SIFT[40], PolyPhen-2[41], GERP[51]);
- gene association with disorders (OMIM catalog of all human genes associated with monogenic disorders, their phenotypic spectra and molecular bases, Decipher[31]);
- gene association with ID (cross reference with lists of known ID genes[32] of various sources including SFARI[52] and sysID[9]);
- function of the encoded proteins (GeneOntology[53], UniProt[54], Human Protein Atlas[55]);
- expression in the affected tissues (GTex);
- animal models (IMPC[56], ZFIN[57], PubMed).

Sanger sequencing was used to confirm the anticipated segregation of the short-listed variants. We considered a variant causative, if it segregated with the disorder in the family, was previously associated with ID or is identified in a known ID gene and is predicted to be deleterious. This genetic diagnosis was transmitted to the clinicians who was responsible to relay it to patients and their families. GUS and VUS were uploaded on GeneMatcher[44] aiming at collaborations with

laboratories having patients presenting similar phenotypes and variants in the same genes. If no match was found at the moment of the submission, the genes will continue to be queried by new entries, allowing long-term data collection. In the meantime, functional analyses were carried out, especially when the affected gene was not previously associated with ID (GUS).

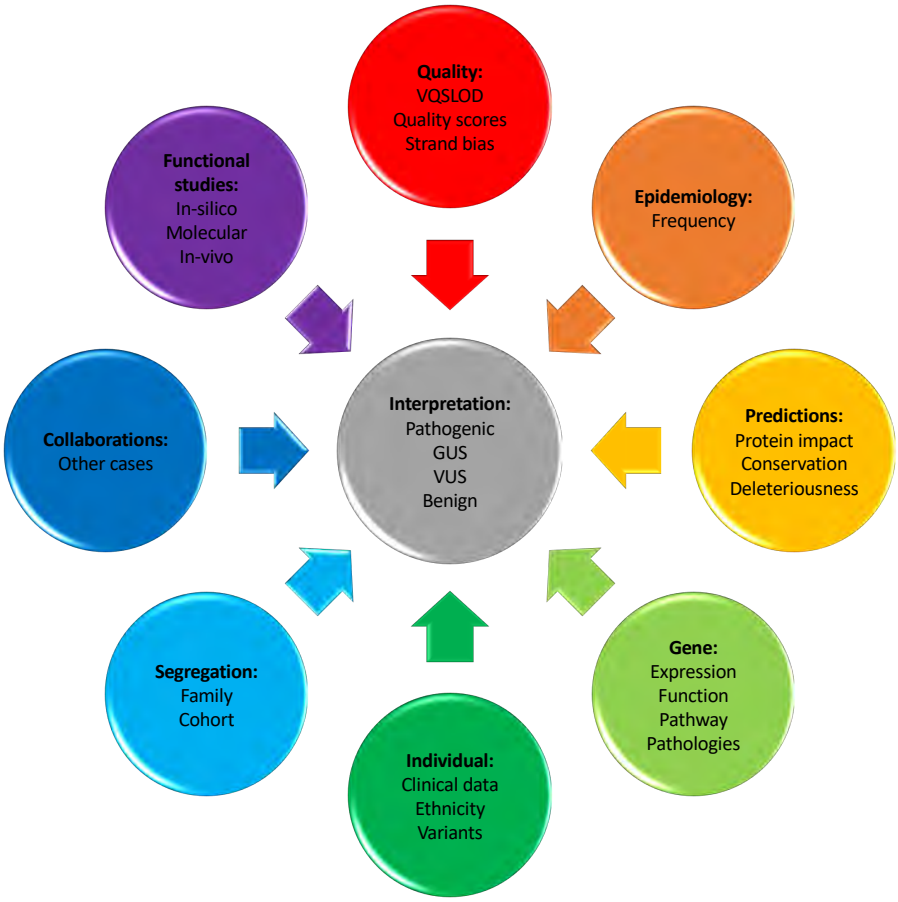


Figure 5: Data collection for variants filtering and interpretation

Results

During my doctorate, I was able to reach a conclusive genetic diagnosis for seven families out of the 23 I studied (30%) (Table 3), ranging within the usual discovery rates of WES for ID and neurodevelopmental disorders (25-40%)[58]. In the following chapters I describe the three projects I led during my PhD.

I also participated to two other projects: *KIAA1109* project led by my colleague, Lucie Gueneau, and *MAST1* project led by Austrian collaborators from David Keays' laboratory. For the *KIAA1109* project I was in charge of the collection of the clinical data, coordination of the collaboration that unite seven teams worldwide and participated in writing of the article. I also presented this work at the American Society of Human Genetics 2016 meeting in Vancouver, Canada as a semifinalist for the Charles J. Epstein Trainee Awards for Excellence in Human Genetics Research. Our common results were published in AJHG with me as co-first author[59]. As for the *MAST1* project, our collaborators collected patients' data and performed functional analyses. I identified one of these patients within our collection and one other through my contacts with clinical geneticists. The results of this project were published in Neuron[60]. Copies of both articles are attached as Appendices 3-4.

Please note that while we analyzed the exomes of Italian family 4 and suspected the pathogenicity of a homozygous recessive variant in *TBCE*, another diagnostic laboratory that was following the same family without our prior knowledge reached the same conclusion and provided evidence for the implication of *TBCE* in neurodegeneration. This work was published independently[61].

Finally, the *POMK* patient is being written as a case study, while we are collecting clinical and genetic data of individuals with *SATB1* variants and overlapping features through international collaborations. Functional validations of the implication of *SATB1* variants in syndromic ID are being coordinated with another team we partnered with.

Family	Results
Pathogenic variants identified (7/23)	
Italian_3	*AFF3, de novo* (Voisin et al, revision to be submitted to AJHG); see p23
Italian_1	*PIK3R2, de novo* (Terrone et al, EJHG, 2016)[62] ; see p59
Lithuanian_4 (057)	*TREX1*, homozygous recessive (Tumiene et al, EJMG, 2017)[63] ; see p64
Italian_4	*TBCE*, homozygous recessive (Sferra et al, AJHG, 2016)[61]
Lithuanian_10 (048)	*KIAA1109,* compound heterozygous (Gueneau et al, AJHG, 2018)[59]
Italian_2	*MAST1, de novo* (Tripathy et al, Neuron, 2018)[60]
Lithuanian_11 (LitX)	*POMK,* homozygous recessive (Preiksaitiene et al, in progress)
Likely-pathogenic variants identified, GUS (3/23)	
Lithuanian_5 (093)	*DCHS2*, compound heterozygous
Lithuanian_7 (107)	*CNGB3,* homozygous recessive
Lithuanian_8 (112)	*SATB1, de novo*
Pathogenic variants to be identified, VUS (13/23)	
Italian_5, Italian_6, Italian_7, Italian_8, Italian_9, Italian_10, Italian_11, Italian_12, Lithuanian_1 (006), Lithuanian_2 (009), Lithuanian_3 (040), Lithuanian_6 (094), Lithuanian_9 (119)	VUS

Table 3: Results of the present study

I. Variants in the degron motif of AFF3 cause a multi-system disorder with skeletal dysplasia and severe neurologic involvement

This project started with an Italian patient exhibiting a multi-system disorder with skeletal dysplasia and neurodevelopmental defects, in the genome of which we identified an *AFF3 de novo* missense variants. Thanks to worldwide collaborations, we gathered a total of 10 individuals with *de novo* missense variants of *AFF3*. They all map within codons encoding the core degron motif, a signal for protein degradation by the ubiquitin-degradation system[64,65]. We describe a new syndrome characterized by a recognizable pattern of anomalies including global developmental and epileptic encephalopathies, horseshoe kidney, dysmorphic facial features, a mesomelic form of skeletal dysplasia resembling Nievergelt/Savarirayan type and other skeletal features. Immunoblotting, as well as zebrafish and mice models, support a gain-of-function mechanism through an increased amount of AFF3 protein product in affected individuals. Transcriptomic analyses[66] and the partial phenotypic overlap between this new syndrome and the *AFF4* variants-associated CHOPS syndrome[66] suggest specialized functions of these AFFs transcription factors.

As first author of this paper, I was in charge of the writing, literature mining and figures assembly. I performed the functional analyses including zebrafish modeling, mouse knock-in modeling, transcriptomic analyses and immunoblotting. I coordinated data sharing between all researchers and clinicians involved in this collaboration. I personally presented this work at the European Society of Human Genetics 2017 congress in Copenhagen, Denmark and was awarded one of the six Young Investigator Awards for best platform presentation. This work is now under revision by our collaborators and will soon be resubmitted for publication (see following pages).

Variants in the degron of *AFF3* cause a multi-system disorder with mesomelic dysplasia, horseshoe kidney and developmental and epileptic encephalopathy

Authors:

Norine Voisin[1], Rhonda E. Schnur[2,3], Sofia Douzgou[4,5], Susan M. Hiatt[6], Cecilie F. Rustad[7], Natasha J. Brown[8,9,10], Dawn L. Earl[11], Boris Keren[12], Olga Levchenko[13], Sinje Geuer[14,15,34], David Amor[9,10], Alfredo Brusco[16,17], E. Martina Bebin[18], Gerarda Cappuccio[19], Joel Charrow[20], Nicolas Chatron[1,21], Gregory M. Cooper[6], Elena Dadali[13], Julien Delafontaine[22], Ennio Del Giudice[19], Ganka Douglas[2], Tara Funari[2], Giuliana Giannuzzi[1], Nicolas Guex[1,22], Delphine Heron[12], Øystein L. Holla[23], Anna C.E. Hurst[24], Jane Juusola[2], David Kronn[25], Alexander Lavrov[13], Crystle Lee[8], Else Merckoll[26], Anna Mikhaleva[1], Jennifer Norman[27], Sylvain Pradervand[1,22], Victoria Sanders[20], Fabio Sirchia[28], Toshiki Takenouchi[29], Akemi J. Tanaka[30,31], Heidi Taska-Tench[20], Elin Tønne[7], Kristian Tveten[23], Giuseppina Vitiello[19], Tomoko Uehara[29], Caroline Nava[12], Binnaz Yalcin[1,32], Kenjiro Kosaki[29], Dian Donnai[4,5], Stefan Mundlos[14,15], Nicola Brunetti-Pierri[19,33], Wendy K. Chung[30,31], Alexandre Reymond[1]

[1]Center for Integrative Genomics, University of Lausanne, Lausanne, CH-1015, Switzerland

[2]GeneDx, Gaithersburg, MD 20877, USA

[3]Cooper Medical School of Rowan University, Division of Genetics, Camden, NJ, 08103 USA

[4]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester M13 9WL, UK

[5]Division of Evolution and Genomic Sciences, School of Biological Sciences, University of Manchester, M13 9NT, UK

[6]HudsonAlpha Institute for Biotechnology, Huntsville AL, 35806 USA

[7]Department of Medical Genetics, Oslo University Hospital, 0424 Oslo, Norway

[8]Victorian Clinical Genetics Services, Flemington Road, Parkville, Victoria 3052, Australia

[9]Murdoch Children's Research Institute, Flemington Road, Parkville, Victoria 3052, Australia

[10]Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia

[11]Seattle Children's, Seattle, WA 98105, USA

[12]Department of Genetics, Pitié-Salpêtrière Hospital, Assistance Publique - Hôpitaux de Paris, Groupe de Recherche Clinique Déficience Intellectuelle et Autisme UPMC, Paris 75013, France
[13]Research Centre for Medical Genetics, Moscow, 115522, Russia
[14]Max Planck Institute for Molecular Genetics, Berlin, 14195, Germany
[15]Institute for Medical and Human Genetics, Charité Universitätsmedizin Berlin, Berlin, 10117, Germany
[16]Department of Medical Sciences, University of Torino, Torino, 10126 Italy
[17]Medical Genetics Unit, Città della Salute e della Scienza University Hospital, Torino, 10126, Italy
[18]Department of Neurology, University of Alabama at Birmingham, Birmingham, AL 35294 USA
[19]Department of Translational Medicine, Section of Pediatrics, Federico II University, Naples, 80131, Italy
[20]Division of Genetics, Birth Defects & Metabolism, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL 60611, USA
[21]Genetics Department, Lyon University Hospital, Lyon, 69007, France
[22]Swiss Institute of Bioinformatics (SIB), Lausanne, CH-1015, Switzerland
[23]Department of Medical Genetics, Telemark Hospital Trust, 3710 Skien, Norway
[24]Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, 35233, USA
[25]New York Medical College, Valhalla, NY 10595, USA
[26]Department of Radiology, Oslo University Hospital, 0424 Oslo, Norway
[27]Integris Pediatric Neurology, Oklahoma City, OK 73112, USA
[28]Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste, 34100, Italy
[29]Center for Medical Genetics, Department of Pediatrics, Keio University School of Medicine, Tokyo, 1608582, Japan
[30]Department of Pediatrics, Columbia University, New York, NY 10032, USA
[31]Department of Medicine, Columbia University, New York, NY 10032, USA
[32]Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, 67404, France
[33]Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Naples, 80078, Italy

Current address:
[34]Center for Human Genetics, Bioscientia, Ingelheim, 55218, Germany

Correspondence should be addressed to:
Alexandre Reymond, alexandre.reymond@unil.ch

Running title: AFF3 degron variants

Keywords: mesomelic dysplasia, horseshoe kidney, intellectual disability, AFF3, AFF4

Abstract

The ALF transcription factor paralogs, *AFF1, AFF2, AFF3* and *AFF4*, are components of the transcriptional super elongation complex that regulates expression of genes involved in neurogenesis and development. We describe a new autosomal dominant disorder associated with *de novo* missense variants in the degron of AFF3, a nine amino acid sequence important for its degradation. Consistent with a causative role of *AFF3* variants, the mutated AFF3 proteins show reduced clearance. Ten affected individuals were identified, and present with a recognizable pattern of anomalies, which we named KINSSHIP syndrome (KI for horseshoe KIdney, NS for Nievergelt/Savarirayan type of mesomelic dysplasia, S for Seizures, H for Hypertrichosis, I for Intellectual disability and P for Pulmonary involvement), partially overlapping the *AFF4* associated CHOPS syndrome. An eleventh individual with a microdeletion encompassing only the transactivation domain and degron motif of *AFF3* exhibited overlapping clinical features. A zebrafish overexpression model that shows body axis anomalies provides further support for the pathological effect of increased amount of AFF3 protein.

Whereas homozygous *Aff3* knockout mice display skeletal anomalies, kidney defects, brain malformation and neurological anomalies, knock-in animals modeling the microdeletion and the missense variants identified in affected individuals presented with lower mesomelic limb deformities and early lethality, respectively.

Transcriptome analyses as well as the partial phenotypic overlap of syndromes associated with *AFF3* and *AFF4* variants suggest that ALF transcription factors are not redundant in contrast to what was previously suggested

Introduction

The *AFF1* (AF4/FMR2 family member 1, a.k.a AF4), *AFF2* (a.k.a *FMR2*), *AFF3* (a.k.a *LAF4*) and *AFF4* genes encode members of the ALF (AF4/LAF4/FMR2) family. These transcription factors share five highly conserved domains starting from the amino terminus: (i) an N-terminal homology domain (NHD); (ii) the hallmark ALF domain, which interacts with Seven In Absentia Homolog (SIAH) ubiquitin ligases through the [xPxAxVxPx] degron motif[1,2] and thus regulates protein degradation mediated by the proteasome pathway; (iii) a serine-rich transactivation domain[3]; (iv) a bipartite nuclear localization sequence (NLS); and (v) a C-terminal homology domain (CHD)[4,5]. *AFF1*, *AFF3*, and *AFF4* have each been identified as fusion partners of the mixed-lineage leukemia (*MLL*) gene involved in acute pediatric leukemias[3]. They are part of the super elongation complex[6] implicated in transcription of a set of genes, among them histones, retinoid signaling and *HOX* genes involved in neurogenesis and several other developmental processes (e.g. *Hoxa1*, *Cdx11* and *Cyp26a1*[6,7]). Mutations of the fruit fly ALF orthologous gene *lilliputian* (*lilli)* were shown to prevent neuronal differentiation and to decrease cell growth and size[8,9]. Silencing of *AFF2* by CGG repeat expansion is associated with FRAXE intellectual disability syndrome[10] (OMIM #309548), whereas hypermethylation of a mosaic CGG repeat expansion in the promoter of *AFF3*, which leads to its silencing in the central nervous system, was associated with a cytogenetic fragile site (FRA2A) and intellectual disability in three families[11]. AFF3 is also known for regulating the expression of imprinted genes[12,13] such as *XIST* through binding to differentially methylated regions[14]. An individual carrying a 500kb microdeletion within the *AFF3* locus and presenting with skeletal dysplasia and encephalopathy was described[15].

Six *de novo* missense variants in *AFF4* were recently linked with CHOPS (Cognitive impairment and coarse facies, Heart defects, Obesity, Pulmonary problems, Short stature and skeletal dysplasia) syndrome[16,17] (OMIM#616368). They were suggested to act through reduced clearance of AFF4 by SIAH, a hypothesis supported by the fact that surviving adult *Aff4* null mice have only azoospermia and no features of CHOPS syndrome. However, a majority of *Aff4*$^{-/-}$ embryos died *in utero* with severely shrunken alveoli of the lung[18]. Upregulation of *AFF4* resulted in dysregulation of genes involved in skeletal development and anterior/posterior pattern formation such as *MYC*, *JUN*, *TMEM100*, *ZNF711* and *FAM13C*[16]. These molecular changes were proposed to impair complex function and lead to

cohesinopathies associated with the clinical phenotypes seen in the eleven reported individuals with CHOPS and in Cornelia de Lange syndrome (CdLS; OMIM #122470)[16,17].

Here we describe 10 individuals with *de novo* missense variants in the *AFF3* gene and a recognizable pattern of anomalies including developmental delay, intellectual disability, seizures, dysmorphic facial features, mesomelic dysplasia, and failure to thrive. Although there is some overlap, the clinical presentation of this autosomal dominant disorder appears to be distinct from CHOPS syndrome.

Material and Methods

Enrollment

Participants were enrolled after written informed consent was obtained from parents or legal guardians according to ethical review boards policies. The clinical evaluation included medical history interviews, physical examinations and review of medical records. The Deciphering Developmental Disorders (DDD)[19] identifier of proband 4 is DDD276869.

Exome/Genome sequencing and analysis

Affected individuals were selected for sequencing to establish a diagnosis.

Proband 1: Trio exome analysis was performed on a NextSeq 500 Sequencing System (Illumina, San Diego, CA) after a 12-plex enrichment with SeqCap EZ MedExome kit (Roche, Basel, Switzerland), according to manufacturer's specifications. Sequence quality was assessed with FastQC 0.11.5, reads were mapped using BWA-MEM (v 0.7.13), sorted and indexed in a bam file (samtools 1.4.1), duplicates were flagged (sambamba 0.6.6), coverage was calculated (picard-tools 2.10.10). Variant calling was done with GATK 3.7 Haplotype Caller. Variants were then annotated with SnpEff 4.3, dbNSFP 2.9.3, gnomAD, ClinVar, HGMD, and an internal database. Coverage for these samples was 93% at a 20x depth threshold.

Probands 2 and 10: Exomes were captured using the IDT xGen Exome Research Panel v1.0 for proband 2 and her parents and SureSelect Human All Exon V4 (50 Mb) for proband 10 and his parents. Sequencing and analyses were performed as previously described[20]. The general assertion criteria for variant classification are publicly available on the GeneDx ClinVar submission page.

Proband 3: The exomes of proband 3, his parents and two healthy siblings were captured and sequenced as described[21]. Variants were called and filtered using the Varapp software[22]. Sanger sequencing confirmed the anticipated segregation of the potentially causative variants.

Proband 4: Exome capture and sequencing was performed as previously described[19].

Proband 5: Exome sequencing of the proband was performed as previously described[23]. Sanger sequencing of samples from parents revealed *de novo* segregation of the variant.

Proband 6: Trio genome analysis was performed as previously described[24]. Sanger sequencing confirmed the *de novo* variant reported here.

Proband 7: Trio exome analysis was performed as previously described[25].

Proband 8: Sample preparation and enrichment was performed using TruSeq DNA Exome kit (Illumina) and sequencing was performed using NextSeq 500 (Illumina) with mean region coverage 83x. Variant were called using VarAft software. Variant analysis was performed according to standards and guidelines for the interpretation of sequence variants[26]. Sanger sequencing confirmed the *de novo* origin of variant.

Proband 9: Trio exome analysis was performed with Agilent SureSelect CRE exome capture, Illumina NextSeq 500 sequencer and a mean coverage of 100x. Data were processed using Cpipe[27] and variant filtering and prioritization were phenotype driven (gene lists: intellectual disability, Mendeliome). Variant classification followed ACMG guidelines.

Protein alignment

Alignments of ALF family members were made using Clustal Omega[28] (v1.2.4) and imported on Jalview[29] for visualization.

Interaction modeling

3D modeling for AFF3 (UniProt entry P51826) and SIAH1 (Q8IUQ4) interaction[30] was obtained on Swiss-PdbViewer-DeepView[31] v4.1. As no structural model for human SIAH1 ubiquitin-ligase was available, we used mouse ubiquitin ligase structure (pdb 2AN6) 100% conserved with human sequence in the binding region[32].

Mouse models

Brain neuroanatomical studies were performed on three 16-week-old male mice in C57BL/6N background with homozygous knock-out of the *Aff3* (a.k.a. *Laf4*) gene[33]. Seventy-eight brain parameters were measured across three coronal sections as described[34] and data were analyzed using a mixed model and comparing to more than 100 wild-type males using a

false discovery rate of 1%. Other metabolic and anatomical phenotypes were assessed by the Welcome Trust Sanger Institute through phenotyping of 6 to 13 homozygous and 7 to 14 heterozygous mice and are available on the International Mouse Phenotyping Consortium website. Engineering of *Aff3*^{del} mice model carrying a 353 kb deletion homologous to the one harbored by an affected individual[15] was previously published[35]. E18.5 animals were processed and stained as described[36]. With Taconic Biosciences GmbH, Cologne, Germany, we engineered a constitutive *Aff3*^{A233T} knock-in through CRISPR/Cas9-mediated gene editing using TGGTGGATGCACGCCGGTTA as guide (NM_001290814.1, NP_001277743.1). This allowed the insertion of an additional silent mutation that creates an *Ale*I restriction site for analytical purposes.

Zebrafish overexpression model

Human wild-type ORFs (AFF3, NM_002285.2 and AFF4, NM_014423.4) cloned into the pEZ-M13 vector were transcribed using the mMessage mMachine kit (Ambion) as prescribed. We injected 1-2 nL of diluted RNA (100-300 ng) inside the yolk, below the cell of wild-type zebrafish embryos at the 1- to 2-cell stage. Phenol red dye with distilled water was injected as vehicle control in similar volume. Injected embryos were raised at 28°C and fixed in 4% PFA for 2 hrs at 4-5 days post fertilization (dpf) and stored in PBS at 4°C. Pictures of the embryos were taken after embedding in glycerol. Counts were compared by Fisher exact test.

Protein accumulation assay

Tagged human wild-type mRNAs cloned into a CMV promoted expression vector were obtained from GeneCopoeia. The ORFs of *AFF3* and *AFF4* were inserted in pEZ-M13 vector with a C-terminal FLAG tag, while the ORF of SIAH1 (NM_001006610) was inserted in pEZ-M07 vector with a C-terminal 3xHA tag. The *AFF3* NM_002285.2:c.697G>A, c.704T>G, and *AFF4* NM_014423.4:c.772C>T mutations were engineered using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) following the manufacturer's instructions. HEK293T cells cultured in complete medium (DMEM containing 10% FBS and 1% penicillin-streptomycin) were transiently transfected with wild type and mutated plasmids using calcium phosphate. 24 hrs after transfection, medium was changed to fresh complete medium. Total protein extracts were obtained after 48 hrs using RIPA buffer with protease and phosphatase inhibitor cocktail. Denatured protein extracts were immunoblotted with anti-FLAG (F3165), -HA (12CA5) and -β-actin (A2066) antibodies from Sigma-Aldrich.

Results

We identified ten unrelated affected individuals (probands 1-10) with *de novo* missense variants in the ALF domain of AFF3 (**Figure 1A and Table 1**) through trio-based exome sequencing and data aggregation of multiple laboratories and clinical centers via GeneMatcher[37]. The four different identified variants (**Table 1**) (i) are not present in the Genome Aggregation Database (gnomAD[38] v2.1.1); (ii) are predicted to be deleterious by SIFT[39], PROVEAN[40], PolyPhen2[41] and MutationTaster2[42]; (iii) are part of the top 1% of all deleterious variants with CADD scores over 20; and (iv) modify highly conserved amino acids (**Figure 1B-C**). Nine of the probands present variants affecting the same codon of exon 6, c.772G>T p.(A258S) (probands 1-2), c.772G>A p.(A258T) (probands 3-8), c.773C>T p.(A258V) (proband 9), whereas proband 10 carries a variant perturbing a neighboring codon c.779T>G p.(V260G) (NM_001025108.1, NP_001020279.1; **Table 1**). An eleventh individual (deletion proband) carrying a 500kb microdeletion and an overlapping phenotype (see below) was previously described[15]. This deletion removes exons 4 to 13 of *AFF3*, which encode its N-terminal region, including the ALF and its degron and part of the transactivation domains and was proposed to act as a dominant negative[35] (**Figure 1A**).

All *AFF3* variants described here and CHOPS syndrome-associated *AFF4 de novo* missense previously published[16,17] map within the degron motif of the ALF domain. This highly conserved 9 amino acid sequence [xPxAxVxPx] (**Figure 1A-B**) mediates interaction with the SIAH E3 ubiquitin ligase and regulates their degradation[1]. According to pathogenic variant enriched regions (PER)[43], the degron is predicted to be constrained within the ALF family. Pathogenicity of the four *de novo AFF3* identified variants is further supported by the three-dimensional representation of part of the encoded peptide (**Figure 1D**). The mutated residues are located within the degron motif (KPT**$\mathbf{A_{258}}$**Y**$\mathbf{V_{260}}$**RPM), which adopts a beta-strand conformation directly contacting the SIAH ubiquitin ligase binding groove[30]. The side chains of Alanine 258 and Valine 260 are embedded into the hydrophobic core of the beta-sandwich where the binding pockets are too small to accommodate larger side chains[32]. Thus, the variants p.(A258T), p.(A258S), p.(A258V) and p.(V260G) are likely to weaken or prevent binding to the ubiquitin ligase. Hence, all these *de novo* variants, as well as the 500kb deletion previously reported[15] that encompasses the degron, could result in hindered degradation and thus accumulation of AFF3. Consistent with this hypothesis, transiently transfected FLAG-tagged AFF3^{A258S} and AFF3^{V260G} proteins were more stable than wild-type FLAG-tagged AFF3

(**Figure 1E**). The previously reported AFF4 *de novo* variants p.(P253R), p.(T254A), p.(T254S), p.(A255T), p.(R258W) and p.(M260T) that also affect the degron motif (K**P**$_{253}$**T**$_{254}$**A**$_{255}$YV**R**$_{258}$P**M**$_{260}$) (**Figure 1A**) were similarly shown to reduce clearance of the ALF transcription factor by SIAH[16,17].

We compared the phenotypes of the ten individuals with *de novo* variants in *AFF3* described here and that of the previously reported case carrying *AFF3* partial deletion[15] (**Table S1** for detailed phenotypes). They exhibit severe developmental epileptic encephalopathy (10 probands out of 11), along with mesomelic dysplasia resembling Nievergelt/Savarirayan mesomelic skeletal dysplasia (NSMSD) (10/11) and failure to thrive (10/11). These three features are often associated with microcephaly (7/11), global brain atrophy and/or ventriculomegaly (7/9) (**Figure S1**), fibular hypoplasia (9/11), horseshoe kidney (8/11), abnormalities of muscle tone (9/10), gastroesophageal reflux disease (5/10) and other gastrointestinal symptoms (10/10). They also share common dysmorphic facial features such as a bulbous nasal tip (6/9), a wide mouth with square upper lip (7/10), abnormalities of the teeth and gums (9/10) and hypertrichosis (8/9) (**Figure 2-3**). Respiratory difficulties/pulmonary involvement were observed in about half of the probands with *de novo* variants (6/11). Whereas respiratory arrest led to the death of proband 3 at 21 years, the deletion proband died at four months after recurrent apneic episodes (**Table S1**).

This constellation of features recalls some features of CHOPS-affected individuals. The three originally described probands[16], along with the eight recently identified[17], presented with distinctive facial dysmorphic features reminiscent of CdLS, short stature with obesity (11/11), developmental delay/intellectual disability (DD/ID) (11/11) and microcephaly (6/11) without epilepsy. They showed gastrointestinal abnormalities (8/11), accompanied by abnormal feeding behavior (6/6), hearing loss (8/11), cardiac (8/11) and pulmonary defects (8/11) and rarely horseshoe kidney (2/11). Whereas they present with vertebral abnormalities (5/11) and brachydactyly (8/11), mesomelic dysplasia is never observed and hypoplastic fibula rarely (1/11).

Although phenotypes of *AFF3* and *AFF4* missense carriers are overlapping, they are not identical. We thus suggest naming the distinct autosomal dominant AFF3-associated disorder KINSSHIP syndrome (KIdney anomalies, Nievergelt/Savarirayan mesomelic dysplasia, Seizures, Hypertrichosis and Intellectual disability with Pulmonary involvement, MIM

#XXXX) to evoke both some of its cardinal characteristics, as well as its similarity (common mode of action and inheritance and overlapping phenotypes) with CHOPS syndrome.

To better understand the functional effects of *AFF3* variation, we investigated both knock-out and knock-in mouse models (**Table 2**). We first studied the knock-out mouse line engineered by the International Mouse Phenotyping Consortium[33] (IMPC). The IMPC routinely measures an extensive series of parameters and evaluate if those are significantly different from wild-type mice[44] ($p \leq 10^{-04}$). *Aff3*$^{+/-}$ and *Aff3*$^{-/-}$ mice exhibit skeletal defects including fusion of vertebral arches, vertebral transformation and decreased caudal vertebrae number. Homozygous knock-out mice also show an abnormal skull shape with a small, deviated snout and malocclusion as well as decreased serum fructosamine and albumin levels that could reflect kidney defects and/or metabolic dysregulation. Neurological dysfunctions were also noted with an increased or absent threshold for auditory brainstem response (signs of hearing impairment) and diminished grip strength. As *Aff3* is expressed in progenitor neurons[45] and required for neuronal migration in the cerebral cortex[46], we further assessed the consequences of *Aff3* disruption on brain development by measuring a standardized set of 78 parameters across 22 brain regions[34]. Compared with wild type males, homozygous *Aff3*$^{-/-}$, but not heterozygous *Aff3*$^{+/-}$ males, exhibited significantly enlarged lateral ventricles ($p = 1.24 \times 10^{-04}$) and decreased corpus callosum size ($p = 3.02 \times 10^{-06}$; **Figure 4**), similar to the phenotypes observed in proband 2, 3 and 6 and in the previously reported deletion proband (**Table S1, Figure S1**)[15]. These features are in stark contrast with results obtained with another engineered *Aff3*$^{-/-}$ line that showed no phenotypic perturbations possibly because of genetic background differences, i.e. C57BL/6N versus CD1, and/or focusing on limb morphology only[35].

We then reassessed mouse models mimicking the deletion identified in the previously described proband, which were previously engineered to assess an aggregation method for the rapid generation of structural variants[35]. Consistent with the phenotype of the deletion proband, homozygous animals chimeric for a 353kb deletion syntenic to the 500kb human deletion exhibited mesomelic dysplasia, triangular tibia, severe hypoplastic fibula and polydactyly of the feet[35] (**Table 2**). Reexamination of these *Aff3*$^{del/del}$ (a.k.a. *Laf4*$^{del/del}$) mice showed that they also presented with reduced body size, craniofacial dysmorphisms with delayed ossification of skull bones, hypoplastic pelvis, intestinal prolapse and neurological dysfunction (**Figure 5A-C**). Chimeric *Aff3*$^{del/+}$ heterozygotes presented with variable features ranging from unaffected to homozygous deletion-like phenotypes. Whereas *Aff3*$^{del/+}$ animals with low chimerism were

fertile they produced no heterozygous offspring suggesting lethality of the 353kb deletion (**Table 2**). While these results support a causative role for the deletion in the deletion proband, they do not allow differentiating between gain-of-function and haploinsufficiency.

To further assess the underlying mutational mechanism in our missense probands, we engineered a knock-in mouse model carrying the *Aff3*^{A233T} mutation that is the equivalent of the most commonly observed *de novo* variant, identified in probands 3 to 8 [p.(A258T)]. The microinjection of a total of 410 C57BL/6NTac zygotes and transfers into 14 recipient females to allow CRISPR/Cas9 editing resulted in only 13 pups at weaning. Genotyping showed that most of them were either wild type (8 individuals) or carried CRISPR/Cas9-mediated mutations (4) although reduced gRNA activity was used for microinjection. A single female F0 founder animal showed the targeted A233T knock-in but with a very low mosaicism rate of 16.7% in an ear biopsy. Genotyping showed that none of its offspring from four consecutive pregnancies were heterozygous for the mutation. These results suggest that the *Aff3*^{A233T} mutation is lethal with high mosaicism (homozygous *Aff3*$^{A233T/A233T}$ and heterozygous *Aff3*$^{+/A233T}$ chimeras), in gametes or during the fetal period (heterozygous *Aff3*$^{+/A233T}$; **Table 2**). The success statistics of similar CRISPR/Cas9 knock-in projects performed by Taconic Biosciences GmbH (Cologne, Germany) through the years further support this hypothesis. Out of 92 attempted knock-in constructs 98% were successful with only 2% failing to generate F0 animals. For most projects, positive F1 animals were also generated.

To lend further support to the model centered on a pathological increase of AFF3 protein product in affected individuals, we assessed its accumulation in zebrafish. Whereas the genome of these teleosts encodes four ALF transcription factors orthologous to the mammalian *AFF1* to *AFF4*, these genes do not harbor a [xPxAxVxPx] degron motif suggesting that their degradation is regulated differently in fish. Therefore, we modeled accumulation by independently overexpressing increasing amounts of unmutated human AFF3 and AFF4 mRNA in zebrafish embryos. We observed a dose-dependent increase in the fraction of 4 dpf embryos with morphological defects upon overexpression of AFF3. The observed phenotypes included bent body axis, yolk sac edema and generalized body development defects at higher doses (**Figure 5D-E**). A similar albeit less pronounced dose-dependent increase in zebrafish embryos with morphological defects was seen upon overexpression of AFF4 (**Figure 5E**).

To further assess the redundancy of ALF transcription factors, we took advantage of published knockdown experiments[47]. Luo and colleagues established and profiled the transcriptome of stable HEK293T cell lines independently knocked down for *AFF2*, *AFF3* and *AFF4* expression by specific shRNAs. Reanalysis of these data confirmed that ALF transcription factors have mostly different target genes, as 55% (125 out of 226), 62% (261/423) and 87% (966/1116) of the genes are specifically perturbed by the knock down of *AFF2*, *AFF3* and *AFF4*, respectively (**Figure 6A**). Intriguingly, the subset of common targets is similarly influenced by decreased expression of *AFF2* and *AFF3* (**Figure 6C,D**), whereas knocking down *AFF3* and *AFF4* had opposite effect (**Figure 6B,C**). 95% (119 out of 125) of common targets are decreased upon reduction of *AFF3* and increased upon reduction of *AFF4* expression suggesting that these two transcription factors act as positive and negative regulators of common pathways. Within the genes perturbed by both *AFF3* and *AFF4*, we observed a significant overrepresentation of genes implicated in the gastrin hormone pathway (CCKR signaling map, P06959) and a proton pump complex (vacuolar proton-transporting V-type ATPase complex, GO:0016471) possibly associated with the gastroesophageal reflux disease observed in both KINSSHIP and CHOPS individuals. Genes linked to the gonadotropin-releasing hormone receptor pathway are similarly enriched (P06664). This observation could be related to cryptorchidism of KINSSHIP proband 1 and small genitalia/cryptorchidism in three out of five males affected by CHOPS syndrome[16,17], as well as the erratic menstrual cycle of proband 4 (most probands being too young to predict any pubertal anomaly) and popliteal pterygium in proband 8 (**Table S1**).

Discussion

All eleven individuals with an *AFF3* variant we identified have a complex but overlapping clinical presentation, which we named KINSSHIP syndrome. One of the cardinal characteristics of this rare autosomal dominant syndrome is mesomelic dysplasia with short forearms, radial head dislocation/subluxation, triangular and/or short tibia, fibular hemimelia, hip dislocation, tarsal and/or metatarsal synostosis resembling NSMSD (**Figure 3**). NSMSD is a sporadic or rare autosomal dominant condition[48,49] associated with neurodevelopmental and often urogenital abnormalities[50,51]. KINSSHIP affected individuals similarly present with vertebral and bone mineralization defects, scoliosis, epilepsy, severe global DD/ID sometimes associated with structural brain abnormalities, significant feeding difficulties, horseshoe kidney, hypertrichosis and recognizable facial features. Multiple probands showed coarsening facial features with age, including a large nose with bulbous nasal tip, a prominent columella and a wide mouth with square upper lip (**Figure 2-3, Table S1, Figure S1**).

AFF3 is one of the targets of the Wnt/β-catenin pathway, an important contributor to pathways involved in bone development and homeostasis[52,53]. Variants in *WNT* genes cause a diverse range of skeletal dysplasias including mesomelic defects (*WNT5A*; Robinow syndrome, dominant type, OMIM#180700), decreased bone density (*WNT1*; Osteogenesis imperfecta, type XV, OMIM#615220) and limb hypoplasia–reduction defects including fibular a/hypoplasia (*WNT3* and *WNT7A*; Tetra-amelia OMIM#273395 and Fuhrmann syndrome OMIM#228930, respectively). Of note individuals with Robinow rhizo/mesomelic dysplasia also present with developmental kidney abnormalities[54], whereas perturbations of the Wnt/b-catenin pathway have been associated with the development of ectodermal appendages like hair and teeth[55]. Nine out of ten KINSSHIP probands show dental anomalies. While widespread hypertrichosis may have been partially caused by multi-drug, antiepileptic treatment in probands 3 and 4, its presence in the only non-epileptic *AFF3* individual (proband 9) and the much younger proband 5 seems to confirm the association of this feature with *AFF3* genetic variants. It is possible that the complex clinical presentation of the cases described here (**Table S1**) may represent the effects of impaired *AFF3* function on a number of downstream targets within the Wnt/β-catenin pathway. In-depth transcriptome analysis of affected individuals and/or animal models is warranted to confirm this hypothesis.

Despite the limited number of individuals for both conditions, similarities and differences are notable between individuals with KINSSHIP and CHOPS[16,17] syndrome. Individuals with variants in *AFF3* and *AFF4* share features that include respiratory difficulties and vertebral abnormalities, as well as less specific clinical findings such as microcephaly, DD/ID and gastroesophageal reflux disease. Although skeletal abnormalities are reported in both CHOPS and KINSSHIP syndrome, KINSSHIP individuals present with mesomelic dysplasia, whereas CHOPS individuals show brachydactyly. Seizures and failure to thrive are specific to KINSSHIP and obesity with short stature to CHOPS. Congenital heart defects and hearing loss are typically observed in CHOPS, while horseshoe kidney and hypoplastic fibula are predominantly present in KINSSHIP (80% versus 18% and 80% versus 9% of affected individuals, respectively). Despite having thick hair and coarse facies in common, CHOPS probands differ from KINSSHIP probands by their round face and dysmorphic features resembling those of CdLS individuals[16,17].

Although proteins encoded by *AFF2*, *AFF3* and *AFF4* were reported to be functionally redundant, at least in regulating splicing and transcription during normal brain development[56], the clinically distinct phenotypes of individuals carrying *de novo* variants in the degron of AFF3 and AFF4 and our zebrafish results suggest that the encoded proteins are not fully redundant. Further support for this hypothesis is provided by the intolerance to LoF variant of *AFF1* (pLI=0.8), *AFF2* (pLI=1), *AFF3* (pLI=1) and *AFF4* (pLI=1) reported by GnomAD. Whereas homozygous *Aff3*$^{-/-}$ knockout mice display features comparable to those presented by KINSSHIP individuals such as skeletal anomalies, kidney defects, brain malformations and neurological anomalies, these animals do not recapitulate the characteristic mesomelia contrary to *Aff3*$^{del/del}$ mice model. This result and the aforementioned intolerance to LoF suggest that *AFF3* could be associated with two different syndromes, the one described here caused by missense degron variants and a hemizygous deletion of the degron, as well as a second one associated with LoF variants for which affected humans remain to be identified. Although this hypothesis warrants further investigation, we have identified by exome sequencing an individual with features partially overlapping those of KINSSHIP. He is compound heterozygote for a truncating mutation and a predicted to be deleterious missense variant outside of the degron.

In conclusion we describe a new pathology that we propose to name KINSSHIP syndrome. It is associated with variants in the degron of *AFF3* that affect the degradation of the encoded protein. This syndrome shows similarities with the *AFF4*-associated CHOPS syndrome, in particular its gain of protein stability and affected tissues. However, specific KINSSHIP features such as mesomelic dysplasia combined with horseshoe kidney allow a differential diagnosis.

Supplemental Data

Supplemental data include 1 figure and 1 table.

Figure S1: Brain MRI of proband 7 carrying a de novo variant in AFF3

Table S1: Phenotype of individuals with AFF3 variants

Conflicts of Interests

Tara Funari, Ganka Douglas, Jane Juusola, and Rhonda E. Schnur and Wendy K. Chung are employees and former employees of GeneDx, respectively. Sinje Geuer is employed at Bioscientia. The remaining authors declare that they have no competing interests.

Acknowledgments

We thank the probands and their families for their participation in this study. We are grateful to Jacques S. Beckmann, Giedre Grigelioniene and the Genomic Technologies Facility of the University of Lausanne. This work was supported by grants from the Swiss National Science Foundation (31003A_182632 to AR), the Simons Foundation (SFARI274424 to AR and SFARI337701 to WKC) and NHGRI (grant UM1HG007301 to SMH, EMB, GMC). The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003], a parallel funding partnership between Wellcome and the Department of Health, and the Wellcome Sanger Institute [grant number WT098051]. The views expressed in this publication are those of the author(s) and not necessarily those of Wellcome or the Department of Health. The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network. This study makes use of DECIPHER which is funded by the Wellcome. We acknowledge the Sanger Mouse Genetics Project for providing mouse samples, funded by the Wellcome Trust grant number 098051. The Australian National Health & Medical Research Council (NHMRC) provided funding for sequencing proband 9 under the Australian Genomics Health Alliance (GNT1113531); the contents are solely the responsibility of the individual authors and do not reflect the views of the NHMRC. The research conducted at the Murdoch Children’s Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Web Resources

ClustalOmega: http://www.clustal.org/omega/

DDD: http://www.ddduk.org/

GeneMatcher: https://genematcher.org/

GeneDx ClinVar submission page: http://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/

GnomAD: https://gnomad.broadinstitute.org/about.

IMPC: http://www.mousephenotype.org/

MutationTaster2: http://www.mutationtaster.org/

PANTHER: http://www.pantherdb.org

PER viewer: http://per.broadinstitute.org/

PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/index.shtml

PROVEAN: http://provean.jcvi.org/index.php

SIFT: http://sift.jcvi.org/

Varaft: https://varaft.eu

Varapp: https://varapp-demo.vital-it.ch

References

1. House, C.M., Frew, I.J., Huang, H.L., Wiche, G., Traficante, N., Nice, E., Catimel, B., and Bowtell, D.D. (2003). A binding motif for Siah ubiquitin ligase. Proc Natl Acad Sci U S A 100, 3101-3106.
2. Bitoun, E., and Davies, K.E. (2005). The robotic mouse: unravelling the function of AF4 in the cerebellum. Cerebellum 4, 250-260.
3. Meyer, C., Hofmann, J., Burmeister, T., Groger, D., Park, T.S., Emerenciano, M., Pombo de Oliveira, M., Renneville, A., Villarese, P., Macintyre, E., et al. (2013). The MLL recombinome of acute leukemias in 2013. Leukemia 27, 2165-2176.
4. Nilson, I., Reichel, M., Ennas, M.G., Greim, R., Knorr, C., Siegler, G., Greil, J., Fey, G.H., and Marschalek, R. (1997). Exon/intron structure of the human AF-4 gene, a member of the AF-4/LAF-4/FMR-2 gene family coding for a nuclear protein with structural alterations in acute leukaemia. Br J Haematol 98, 157-169.
5. Oliver, P.L., Bitoun, E., Clark, J., Jones, E.L., and Davies, K.E. (2004). Mediation of Af4 protein function in the cerebellum by Siah proteins. Proc Natl Acad Sci U S A 101, 14901-14906.
6. Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. Nat Rev Mol Cell Biol 13, 543-547.
7. Jonkers, I., and Lis, J.T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol 16, 167-177.
8. Tang, A.H., Neufeld, T.P., Rubin, G.M., and Muller, H.A. (2001). Transcriptional regulation of cytoskeletal functions and segmentation by a novel maternal pair-rule gene, lilliputian. Development 128, 801-813.
9. Wittwer, F., van der Straten, A., Keleman, K., Dickson, B.J., and Hafen, E. (2001). Lilliputian: an AF4/FMR2-related protein that controls cell identity and cell growth. Development 128, 791-800.
10. Gecz, J., Gedeon, A.K., Sutherland, G.R., and Mulley, J.C. (1996). Identification of the gene FMR2, associated with FRAXE mental retardation. Nat Genet 13, 105-108.
11. Metsu, S., Rooms, L., Rainger, J., Taylor, M.S., Bengani, H., Wilson, D.I., Chilamakuri, C.S., Morrison, H., Vandeweyer, G., Reyniers, E., et al. (2014). FRA2A is a CGG repeat expansion associated with silencing of AFF3. PLoS Genet 10, e1004242.
12. Luo, Z., Lin, C., Woodfin, A.R., Bartom, E.T., Gao, X., Smith, E.R., and Shilatifard, A. (2016). Regulation of the imprinted Dlk1-Dio3 locus by allele-specific enhancer activity. Genes Dev 30, 92-101.
13. Wang, Y., Shen, Y., Dai, Q., Yang, Q., Zhang, Y., Wang, X., Xie, W., Luo, Z., and Lin, C. (2017). A permissive chromatin state regulated by ZFP281-AFF3 in controlling the imprinted Meg3 polycistron. Nucleic Acids Res 45, 1177-1185.
14. Zhang, Y., Wang, C., Liu, X., Yang, Q., Ji, H., Yang, M., Xu, M., Zhou, Y., Xie, W., Luo, Z., et al. (2018). AFF3-DNA methylation interplay in maintaining the mono-allelic expression pattern of XIST in terminally differentiated cells. J Mol Cell Biol.
15. Steichen-Gersdorf, E., Gassner, I., Superti-Furga, A., Ullmann, R., Stricker, S., Klopocki, E., and Mundlos, S. (2008). Triangular tibia with fibular aplasia associated with a microdeletion on 2q11.2 encompassing LAF4. Clin Genet 74, 560-565.
16. Izumi, K., Nakato, R., Zhang, Z., Edmondson, A.C., Noon, S., Dulik, M.C., Rajagopalan, R., Venditti, C.P., Gripp, K., Samanich, J., et al. (2015). Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. Nat Genet 47, 338-344.

17. Raible, S.E., Mehta, D., Bettale, C., Fiordaliso, S., Kaur, M., Medne, L., Rio, M., Haan, E., White, S.M., Cusmano-Ozog, K., et al. (2019). Clinical and molecular spectrum of CHOPS syndrome. Am J Med Genet A.
18. Urano, A., Endoh, M., Wada, T., Morikawa, Y., Itoh, M., Kataoka, Y., Taki, T., Akazawa, H., Nakajima, H., Komuro, I., et al. (2005). Infertility with defective spermiogenesis in mice lacking AF5q31, the target of chromosomal translocation in human infant leukemia. Mol Cell Biol 25, 6834-6845.
19. Deciphering Developmental Disorders, S. (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223-228.
20. Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K.G., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. Genet Med 18, 696-704.
21. Alfaiz, A.A., Micale, L., Mandriani, B., Augello, B., Pellico, M.T., Chrast, J., Xenarios, I., Zelante, L., Merla, G., and Reymond, A. (2014). TBC1D7 mutations are associated with intellectual disability, macrocrania, patellar dislocation, and celiac disease. Hum Mutat 35, 447-451.
22. Delafontaine J., M.A., Liechti R., Kuznetsov D., Xenarios I., Pradervand S. (2016). Varapp: A reactive web-application for variants filtering. bioRxiv preprint.
23. Holla, O.L., Bock, G., Busk, O.L., and Isfoss, B.L. (2014). Familial visceral myopathy diagnosed by exome sequencing of a patient with chronic intestinal pseudo-obstruction. Endoscopy 46, 533-537.
24. Bowling, K.M., Thompson, M.L., Amaral, M.D., Finnila, C.R., Hiatt, S.M., Engel, K.L., Cochran, J.N., Brothers, K.B., East, K.M., Gray, D.E., et al. (2017). Genomic diagnosis for children with intellectual disability and/or developmental delay. Genome Med 9, 43.
25. Takenouchi, T., Yamaguchi, Y., Tanikawa, A., Kosaki, R., Okano, H., and Kosaki, K. (2015). Novel overgrowth syndrome phenotype due to recurrent de novo PDGFRB mutation. J Pediatr 166, 483-486.
26. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17, 405-424.
27. Sadedin, S.P., Dashnow, H., James, P.A., Bahlo, M., Bauer, D.C., Lonie, A., Lunke, S., Macciocca, I., Ross, J.P., Siemering, K.R., et al. (2015). Cpipe: a shared variant detection pipeline designed for diagnostic settings. Genome Med 7, 68.
28. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7, 539.
29. Waterhouse, A.M., Martin, D.M.A., Barton, G.J., Procter, J.B., and Clamp, M. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189-1191.
30. Santelli, E., Leone, M., Li, C., Fukushima, T., Preece, N.E., Olson, A.J., Ely, K.R., Reed, J.C., Pellecchia, M., Liddington, R.C., et al. (2005). Structural analysis of Siah1-Siah-interacting protein interactions and insights into the assembly of an E3 ligase multiprotein complex. J Biol Chem 280, 34278-34287.
31. Johansson, M.U., Zoete, V., Michielin, O., and Guex, N. (2012). Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinformatics 13, 173.

32. House, C.M., Hancock, N.C., Moller, A., Cromer, B.A., Fedorov, V., Bowtell, D.D., Parker, M.W., and Polekhina, G. (2006). Elucidation of the substrate binding site of Siah ubiquitin ligase. Structure 14, 695-701.
33. Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T., et al. (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature 474, 337-342.
34. Mikhaleva, A., Kannan, M., Wagner, C., and Yalcin, B. (2016). Histomorphological Phenotyping of the Adult Mouse Brain. Curr Protoc Mouse Biol 6, 307-332.
35. Kraft, K., Geuer, S., Will, A.J., Chan, W.L., Paliou, C., Borschiwer, M., Harabula, I., Wittler, L., Franke, M., Ibrahim, D.M., et al. (2015). Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. Cell Rep 10, 833-839.
36. Mundlos, S. (2000). Skeletal morphogenesis. Methods Mol Biol 136, 61-70.
37. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36, 928-930.
38. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv, 531210.
39. Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4, 1073-1081.
40. Choi, Y., and Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745-2747.
41. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-249.
42. Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nature Methods 11, 361.
43. Pérez-Palma, E., May, P., Iqbal, S., Niestroj, L.-M., Du, J., Heyne, H., Castrillon, J., O'Donnell-Luria, A., Nürnberg, P., Palotie, A., et al. (2019). Identification of pathogenic variant enriched regions across genes and gene families. bioRxiv, 641043.
44. Brown, S.D., and Moore, M.W. (2012). The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping. Mamm Genome 23, 632-640.
45. Fietz, S.A., Lachmann, R., Brandl, H., Kircher, M., Samusik, N., Schroder, R., Lakshmanaperumal, N., Henry, I., Vogt, J., Riehn, A., et al. (2012). Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. Proc Natl Acad Sci U S A 109, 11836-11841.
46. Moore, J.M., Oliver, P.L., Finelli, M.J., Lee, S., Lickiss, T., Molnar, Z., and Davies, K.E. (2014). Laf4/Aff3, a gene involved in intellectual disability, is required for cellular migration in the mouse cerebral cortex. PLoS One 9, e105933.
47. Luo, Z., Lin, C., Guest, E., Garrett, A.S., Mohaghegh, N., Swanson, S., Marshall, S., Florens, L., Washburn, M.P., and Shilatifard, A. (2012). The super elongation complex family of RNA polymerase II elongation factors: gene target specificity and transcriptional output. Mol Cell Biol 32, 2608-2617.
48. Nakamura, M., Matsuda, Y., Higo, M., and Nishimura, G. (2007). A family with an autosomal dominant mesomelic dysplasia resembling mesomelic dysplasia Savarirayan and Nievergelt types. Am J Med Genet A 143A, 2079-2081.

49. Bonafe, L., Cormier-Daire, V., Hall, C., Lachman, R., Mortier, G., Mundlos, S., Nishimura, G., Sangiorgi, L., Savarirayan, R., Sillence, D., et al. (2015). Nosology and classification of genetic skeletal disorders: 2015 revision. Am J Med Genet A 167A, 2869-2892.
50. Tuysuz, B., Zeybek, C., Zorer, G., Sipahi, O., and Ungur, S. (2002). Patient with the mesomelic dysplasia, Nievergelt syndrome, and cerebellovermian agenesis and cataracts. Am J Med Genet 109, 206-210.
51. Savarirayan, R., Cormier-Daire, V., Curry, C.J., Nashelsky, M.B., Rappaport, V., Rimoin, D.L., and Lachman, R.S. (2000). New mesomelic dysplasia with absent fibulae and triangular tibiae. Am J Med Genet 94, 59-63.
52. Lefevre, L., Omeiri, H., Drougat, L., Hantel, C., Giraud, M., Val, P., Rodriguez, S., Perlemoine, K., Blugeon, C., Beuschlein, F., et al. (2015). Combined transcriptome studies identify AFF3 as a mediator of the oncogenic effects of beta-catenin in adrenocortical carcinoma. Oncogenesis 4, e161.
53. Zhong, Z., Ethen, N.J., and Williams, B.O. (2014). WNT signaling in bone development and homeostasis. Wiley Interdiscip Rev Dev Biol 3, 489-500.
54. Tufan, F., Cefle, K., Turkmen, S., Turkmen, A., Zorba, U., Dursun, M., Ozturk, S., Palanduz, S., Ecder, T., Mundlos, S., et al. (2005). Clinical and molecular characterization of two adults with autosomal recessive Robinow syndrome. Am J Med Genet A 136, 185-189.
55. Kimura, R., Watanabe, C., Kawaguchi, A., Kim, Y.I., Park, S.B., Maki, K., Ishida, H., and Yamaguchi, T. (2015). Common polymorphisms in WNT10A affect tooth morphology as well as hair shape. Hum Mol Genet 24, 2673-2680.
56. Melko, M., Douguet, D., Bensaid, M., Zongaro, S., Verheggen, C., Gecz, J., and Bardoni, B. (2011). Functional characterization of the AFF (AF4/FMR2) family of RNA-binding proteins: insights into the molecular pathology of FRAXE intellectual disability. Hum Mol Genet 20, 1873-1885.

Figure Titles and Legends

Figure 1: AFF3 and AFF4 degron motif variants

(**A)** Schematic protein structure of ALF proteins with from the amino terminus a N-terminal homology domain (NHD), the AF4/LAF4/FMR2 homology domain (ALF)[21] containing the SIAH-binding degron motif, a serine-rich transactivation domain (TAD)[3], a bipartite nuclear/nucleolar localization sequence (NLS), and a C-terminal homology domain (CHD). The sequences of the degron motif of AFF3 and AFF4 are shown above. The residues modified in the KINSSHIP probands described in this manuscript and individuals affected by CHOPS[16,17] are highlighted in bold and numbered. The extent of the 500kb deletion previously described[35] is indicated here. A red arrow pinpoints the position of the degron motif.

(B) Amino acid sequences alignment of human AFF1, AFF2, AFF3 and AFF4 proteins (ENSP00000305689, ENSP00000359489, ENSP00000317421 and ENSP00000265343, respectively) showing the highly conserved degron motif (red rectangle) of the ALF homology domain that provides the binding moiety to the SIAH ubiquitin-ligase. Sequences alignment was performed using Clustal Omega and edited using Jalview. Shading is proportional to conservation among sequences.

(C) Amino acid sequences alignment of different AFF3 vertebrate orthologs showing the conservation of the degron motif (red rectangle). Accession numbers are ENSP00000317421 (human), ENSMUSP00000092637 (mouse), ENSFCAP00000024603 (cat), ENSLAFP00000010776 (elephant), ENSPSIP00000007060 (chinese turtle), ENSACAP00000008035 (anole lizard) and ENSPMAP00000008605 (lamprey).

(D) 3D modeling of the binding of human AFF3 degron to the mouse Siah ubiquitin ligase. PDB entry 2AN6[32] was loaded in Swiss-PdbViewer[31] and used as a template to align the human SIAH ubiquitin ligase (uniprot entry Q8IUQ4)[30]. With respect to the mouse crystal structure, the only difference is the presence of an aspartic acid residue instead of a glutamic acid at position 116. The region of AFF3 containing the degron motif (LRPVAMVRPTV) was then aligned onto the Siah-interacting protein[50] peptide present in the crystal structure (QKPTAYVRPMD) to highlight the position of the variants reported in this study. For clarity, only side-chains of the core degron motif (P256, A258, V260 and P262) are shown, with yellow highlights on the KINSSHIP mutated residues. A zoom in is displayed on the right. The core degron motif adopts a beta-strand conformation directly contacting the ubiquitin ligase-binding groove. The sidechains of A258 and V260 are embedded into binding pockets too small to accommodate larger side chains[32]. They are in direct proximity of Siah residues T156 (pink)

and M180 (cyan), identified as key binding residues in a series of pull-down assays[32]. The longer side-chains of A258T, A258S, A258V variants and the smaller V260G are likely to weaken or prevent the interaction with the ligase.

(E) Immunoblot showing the accumulation of mutated forms of AFF3 and AFF4 proteins compared to wild type (WT). Protein extracts of HEK293T cells independently expressing FLAG-tagged $AFF3^{WT}$, $AFF3^{A258S}$, $AFF3^{V260G}$, $AFF4^{WT}$ and $AFF3^{R258W}$proteins were immunoblotted with an anti-FLAG antibody (upper portion) and an anti-ß-actin antibody for loading control (bottom portion). The positions of FLAG-AFF3, FLAG-AFF4 and ß-actin proteins are indicated on the right; they are 133kD, 127kD and 42kD respectively. Signal intensity is measured and normalized on corresponding loading control.

Figure 2: Photographs of KINSSHIP individuals with *AFF3 de novo* missense variants

(A) Proband 2 at 2 years 6 months old;

(B, I) Proband 3 at 18 years old;

(C) Proband 4 at 9 months and **(D, J)** 21 years old;

(E) Proband 5 at 1 year 7 months and **(N-O)** 16 days old;

(F, K-M) Proband 6 at 9 years old;

(G) Proband 7 at 8 years old;

(H, P-R) Proband 8 at 7 years 9 months old;

(S) Proband 10 at 11 years old.

Note the synophrys and micrognathia, protruding ears, large nose with prominent nasal tip and prominent teeth in proband 3 **(B)**, 4 **(D),** 6 **(F)** and 8 **(H),** as well as their hypertrichosis of the limbs **(I, J, M, P)**. Together with probands 5 and 7, they exhibit thick hair, long eyelashes and a wide mouth **(E, G).** Facial features coarsen with age as shown by pictures of proband 4 at different ages **(C-D),** explaining the more delicate features of younger probands **(A, E).**

AFF3 de novo missense variant carriers also have hypoplastic talipes and abnormalities of toes **(I, J, M-S)**. Proband 6 also shows clinodactyly and soft tissue syndactyly of both hands **(K, L).**

Figure 3: X-rays of KINSSHIP individuals with *de novo* missense variants in *AFF3*

(A-D) Proband 3 at 18 years old;

(E-I) Proband 4 at 21 years old;

(J-L) Proband 5 at 10 months old;

(M) Proband 7 at 8 years old;

(N-P) Proband 10 at 10 years old.

(A) Severe scoliosis and fusion of C2-C3 vertebral bodies and L5-S1 vertebral cleft **(B)** Dorsal and radial bowing of the radius and "V-shaped" proximal carpal bones as seen in Madelung deformity, **(C)** metaphyseal widening and hypoplastic fibula and **(D)** hypoplastic talipes. **(E)** Static scoliosis, **(F)** Short ulna and radius and bilateral dislocation/subluxation of radial heads. Note erratic articulation of the styloid process of the ulna on the radius rather than on the carpal bones, **(G)** Congenital fusion of the bases of the second and third right metatarsals, **(H**) Hypoplastic bowing femora and **(I)** short tibias with enlarged metaphyses. **(J)** Right foot with 4^{th} and 5^{th} metatarsals synostosis **(K)** and left foot missing the lateral ray, **(L)** Extremely short rectangular fibula and bowed tibia. **(M)** Hypoplastic fibula. **(N)** Scoliosis and cervical ribs, **(O)** bowed radius and ulna, **(P)** bowed tibia, severely hypoplastic fibula and oligodactyly.

Figure 4: Neuroanatomical defects in *Aff3*$^{-/-}$ mice

Merged double-stained sections in *Aff3*$^{-/-}$ mice (right of dashed lines) and their matched controls (WT: wild type, left of dashed lines) at the striatum **(A)** and at the hippocampus **(B)** levels with schematic representation of the affected areas **(C,D).** Histograms showing the percentage of increase or decrease of parameters in measured areas as compared to the controls, for striatum **(E)** and hippocampus **(F)** sections. Red shading is proportional to the stringency of the significance threshold. Numbers indicate studied areas: 1=total brain area, 2=lateral ventricles, 3=cingulate cortex (section 1) and retrosplenial cortex (section2), 4=corpus callosum, 5=caudate putamen (section 1) and hippocampus (section 2), 6=anterior commissure (section 1) and amygdala (section 2), 7=piriform cortex, 8=motor cortex, 9=somatosensory cortex, 10=mammilo-thalamic tract, 11=internal capsule, 12=optic tract, 13=fimbria, 14=habenular, 15=hypothalamus, and 16=third ventricle. Results demonstrate an enlargement of lateral ventricles (**LV**; p=1.24E-04 on section 1, p=4.64E-02 on section 2) and a smaller genu of the corpus callosum (**gcc**; decreased corpus callosum size p=6.35E-02 indicated by the black dash and double arrow, decreased bottom width of the corpus callosum p=3.02E-06 and decreased height of the corpus callosum p=4.96E-02). Other phenotypes of lesser stringency can be observed such as atrophy of the anterior commissure (**aca**; p=1.02E-02) and smaller hippocampus (p=4.02E-02).

Figure 5: Animal models

(**A)** Schematic representation of the deletion generated in mice ES cells with the CRISPR/Cas9 system, which models the mutation observed in the deletion proband[15,35].

(B) Skeletal staining of E18.5 mouse embryos show mesomelic dysplasia with triangular tibia and hypoplastic fibula (see Figure 3 in reference[35]), as well as a hypoplastic pelvis in *Aff3*$^{del/del}$ mice, especially noticeable in the iliac wing (black arrows) and acetabulum (orange arrows), perturbations also observed in the deletion proband.

(C) Delayed ossification of flat bones in the skull of *Aff3*$^{del/del}$ mice.

(D) Lateral (top line) and dorsal (bottom line) views of the observed phenotypes of 4 dpf AB-WT zebrafish embryos injected with human AFF3 mRNA (hAFF3). hAFF3-injected zebrafish embryos exhibit severe developmental defects including a bent body axis and yolk sac edema (D3-6), as well as extreme malformations with absence of body axis, tail and fins and cyclopia (D7-8). Embryos with normal development are displayed for comparison (D1-2).

(E) Proportions of normal and developmentally defective 4 dpf AB-WT zebrafish embryos upon injection of increasing doses of hAFF3 (left panel) and hAFF4 (right panel) mRNA. Dark and light colors indicate developmentally defective and normal animals, respectively. Control injections with Phenol Red show no significant (ns) differences with WT in both AFF3 and AFF4 experiments (Fisher exact test, p=0.09 and p=0.12 respectively). hAFF3 mRNA injection significantly increases the number of zebrafish with developmental defects when compared to controls starting from 150ng (*; p=0.03) and reinforced at 300ng (***; p=3.2E-5). AFF4 injections do not have a significant impact on zebrafish development compared to WT, even at the same dose (300ng, p=0.29).

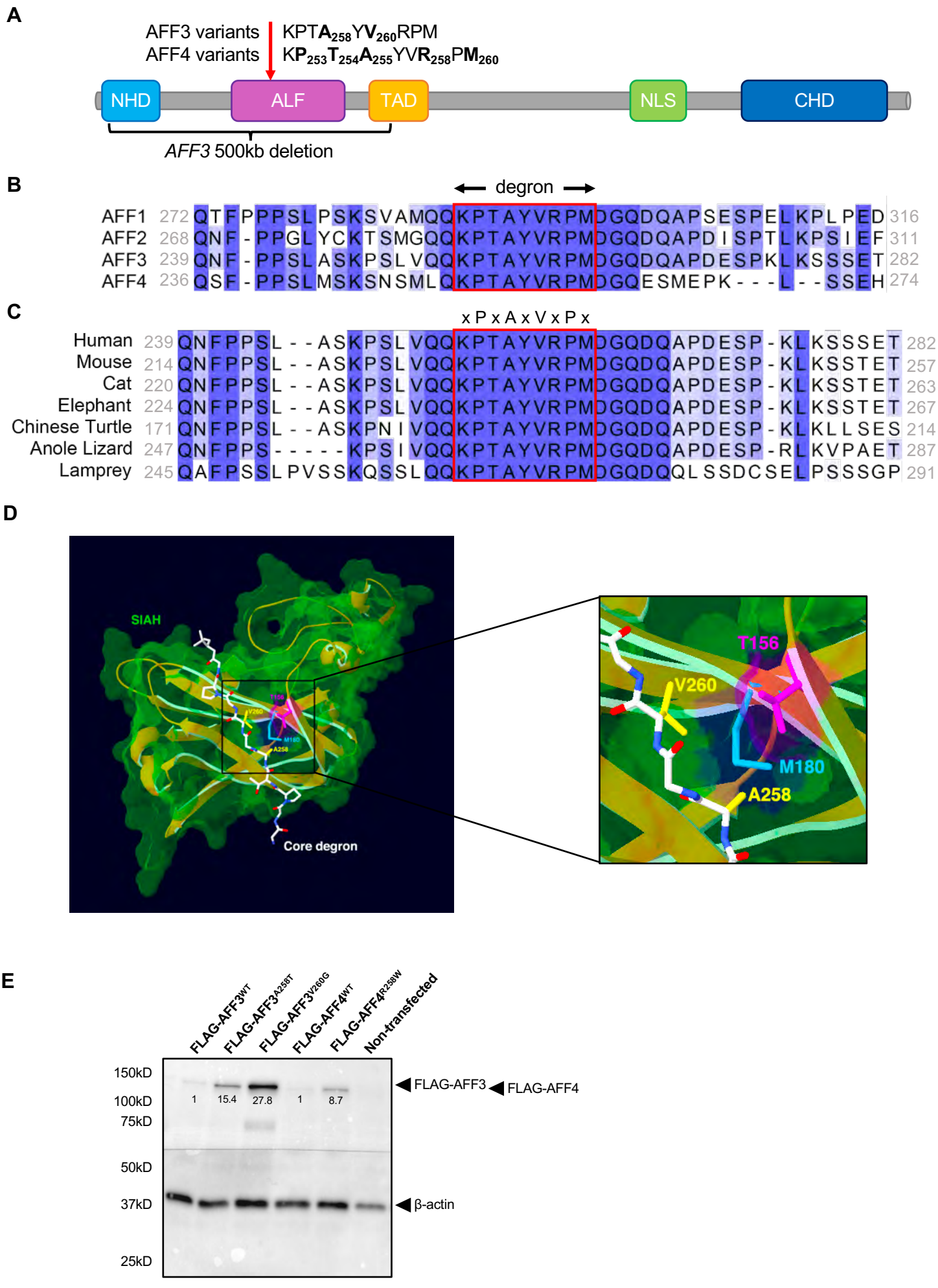
Figure 6: *AFF2, AFF3* and *AFF4* targets

(A) Venn diagram of the differentially expressed genes from independent RNAi of *AFF2, AFF3*, and *AFF4* (adapted from Luo et al[47]) showing that ALF transcription factors have different targets. The knocked-down gene color code is indicated on the right.

The expression modifications of common targets are presented in panels **B-E**. Common differentially expressed genes show adverse regulation between knockdown of *AFF3* and *AFF4* **(B)**, *AFF2* and *AFF4* **(E)** and *AFF2/AFF3* and *AFF4* **(C).** Whereas RNAi of *AFF2* and *AFF3* similarly influence their 76 common targets **(D)**, knockdowns of *AFF2* or *AFF3* have opposite effects to that of *AFF4* on their common targets [54/64=84% **(E)** and 119/125=95% **(B)** of common targets with opposite perturbation of their expression levels, respectively].

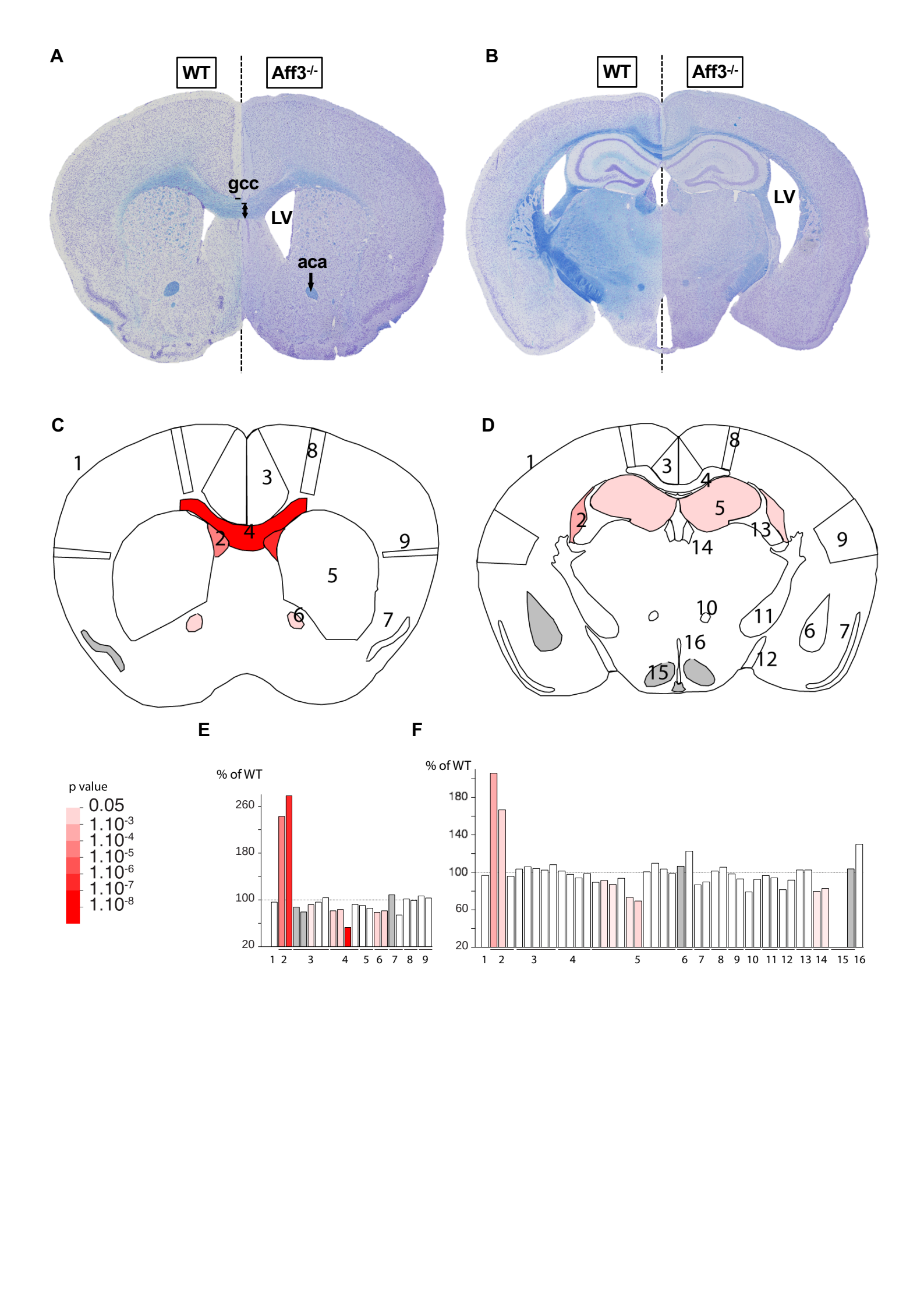
Table 1. Predicted pathogenicity and allele frequencies of *AFF3 de novo* variants											
Gene	Individual	Chromosome coordinates (GRCh37/hg19)	Nucleotide change	Amino acid change	dbSNP (v152)	GnomAD allele frequency (v2.1.1)	Deleteriousness prediction (score)				
							CADD_PHRED (GRCh37-v1.4)	SIFT (v4.0.3)	PROVEAN (v1.1)	PolyPhen2 (v2.2.2)	Mutation Taster2 (09.01.19)
AFF3 NM_001025108.1 NP_001020279.1	1	Chr2:100623270	c.772G>T	A258S	-	0	23.8	Damaging (0.000)	Neutral (-2.48)	Probably damaging (1.000)	Disease causing
	2										
	3	Chr2:100623270	c. 772G>A	A258T	rs1131692272	0	24.3	Damaging (0.000)	Deleterious (-3.30)	Probably damaging (1.000)	Disease causing
	4										
	5										
	6										
	7										
	8										
	9	Chr2:100623269	c.773C>T	A258V	-	0	24.2	Damaging (0.000)	Deleterious (-3.30)	Probably damaging (1.000)	Disease causing
	10	Chr2:100623263	c. 779T>G	V260G	-	0	24.4	Damaging (0.000)	Deleterious (-5.85)	Probably damaging (1.000)	Disease causing
Footnote: SIFT cutoff=0.05, PROVEAN cutoff=-2.5, PolyPhen2 cutoff=0.5.											

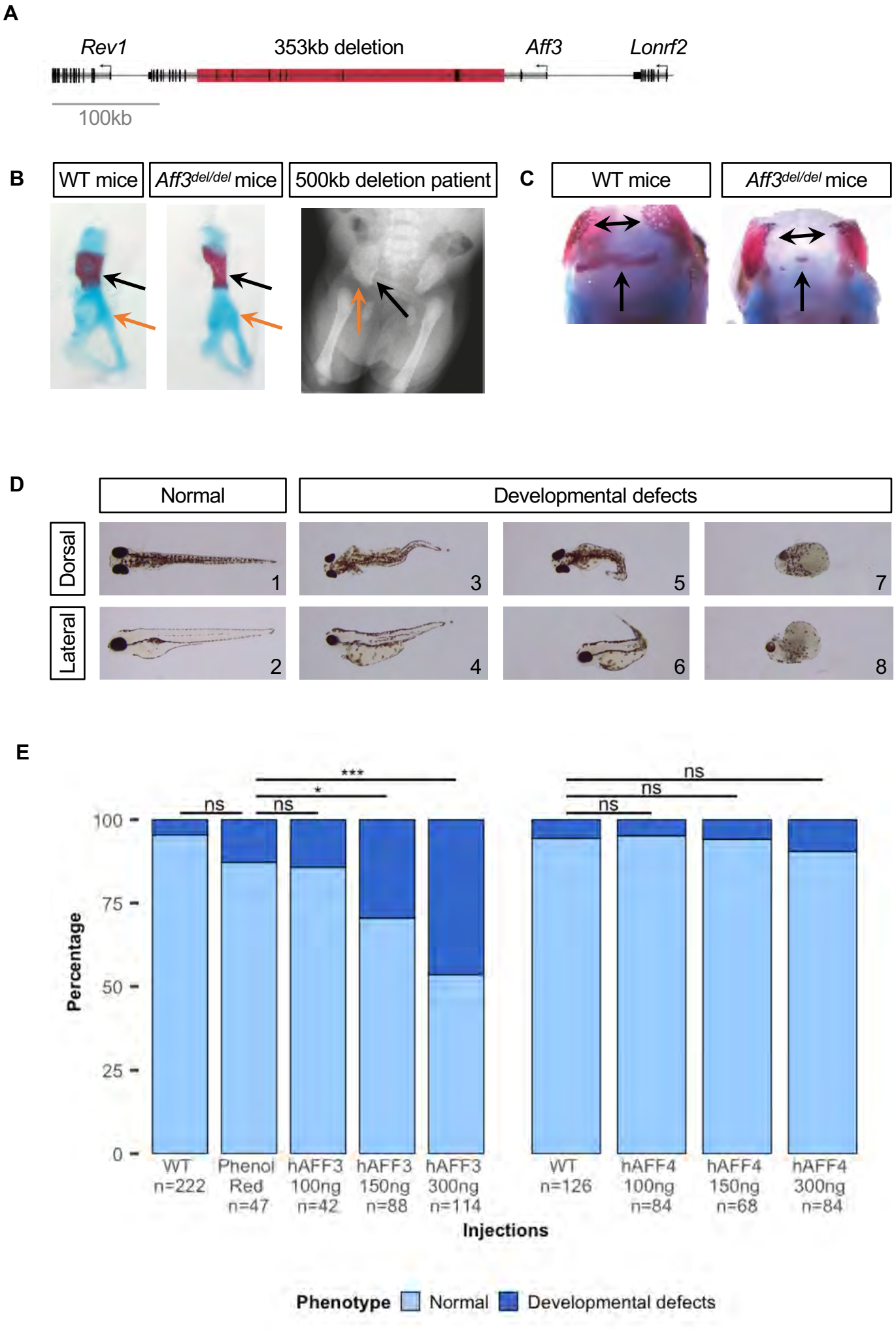
Table 2: AFF3 mice models and their phenotypes							
Producer		IMPC[33]		Kraft et. al.[35]	This work		
Background		*C57BL/6N*		CD1	CD1		*C57BL/6N*
Genotype		*Aff3*$^{+/-}$	*Aff3*$^{-/-}$	*Aff3*$^{-/-}$	*Aff3*$^{del/+}$	*Aff3*$^{del/del}$	*Aff3*^{A233T}
Phenotype	Craniofacial anomalies	-	+	NA	Variable features, from non-affected to homozygous-like phenotype	+	NA
	Vertebral malformations	+	+	NA		NA	
	Mesomelic dysplasia	-	-	-		+ *	
	Polydactyly	-	-	-		+	
	Kidney malfunction	-	+	NA		NA	
	Intestinal prolapse	-	-	NA		+	
	Neurological dysfunctions	-	+	NA		+	
	Neuroanatomical defects	NA	+	NA		NA	
	Reduced body size	-	-	NA		+	
	Lethality	-	-	NA	Low chimerism, no heterozygous offspring, suggesting lethality	Postnatal lethality	No heterozygous offspring, suggesting lethality
NA: Not available; IMPC significance threshold $p \leq 10^{-04}$, *Kraft et. al[35]							


A
AFF3 variants
AFF4 variants
KPTA258YV260RPM
KP253T254A255YVR258PM260
NHD
ALF
TAD
NLS
CHD
AFF3 500kb deletion
B
degron
AFF1 272 QTFPPPSLPSKSVAMQQKPTAYVRPMDGQDQAPSESPELKPLPED 316
AFF2 268 QNF-PPGLYCKTSMGQQKPTAYVRPMDGQDQAPDISPTLKPSIEF 311
AFF3 239 QNF-PPSLASKPSLVQQKPTAYVRPMDGQDQAPDESPKLKSSSET 282
AFF4 236 QSF-PPSLMSKSNSMLQKPTAYVRPMDGQESMEPK---L--SSEH 274
C
x P x A x V x P x
Human 239 QNFPPSL--ASKPSLVQQKPTAYVRPMDGQDQAPDESP-KLKSSSET 282
Mouse 214 QNFPPSL--ASKPSLVQQKPTAYVRPMDGQDQAPDESP-KLKSSTET 257
Cat 220 QNFPPSL--ASKPSLVQQKPTAYVRPMDGQDQAPDESP-KLKSSTET 263
Elephant 224 QNFPPSL--ASKPSLVQQKPTAYVRPMDGQDQAPDESP-KLKSSTET 267
Chinese Turtle 171 QNFPPSL--ASKPNIVQQKPTAYVRPMDGQDQAPDESP-KLKLLSES 214
Anole Lizard 247 QNFPPS-----KPSIVQQKPTAYVRPMDGQDQAPDESP-RLKVPAET 287
Lamprey 245 QAFPSSLPVSSKQSSLQQKPTAYVRPMDGQDQQLSSDCSELPSSSGP 291
D
SIAH
Core degron
T156
V260
M180
A258
E
FLAG-AFF3WT
FLAG-AFF3A258T
FLAG-AFF3V260G
FLAG-AFF4WT
FLAG-AFF4R258W
Non-transfected
150kD
100kD
75kD
50kD
37kD
25kD
1
15.4
27.8
1
8.7
FLAG-AFF3
FLAG-AFF4
β-actin


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P


A
WT
Aff3-/-
gcc
LV
aca
B
WT
Aff3-/-
LV
C
1
2
3
4
5
6
7
8
9
D
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
E
p value
0.05
1.10-3
1.10-4
1.10-5
1.10-6
1.10-7
1.10-8
% of WT
260
180
100
20
1 2 3 4 5 6 7 8 9
F
% of WT
180
140
100
60
20
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16


A
Rev1
353kb deletion
Aff3
Lonrf2
100kb
B
WT mice
Aff3del/del mice
500kb deletion patient
C
WT mice
Aff3del/del mice
D
Normal
Developmental defects
Dorsal
Lateral
1
2
3
4
5
6
7
8
E
***
*
ns
ns
ns
ns
ns
100
75
50
25
0
Percentage
WT n=222
Phenol Red n=47
hAFF3 100ng n=42
hAFF3 150ng n=88
hAFF3 300ng n=114
WT n=126
hAFF4 100ng n=84
hAFF4 150ng n=68
hAFF4 300ng n=84
Injections
Phenotype
Normal
Developmental defects

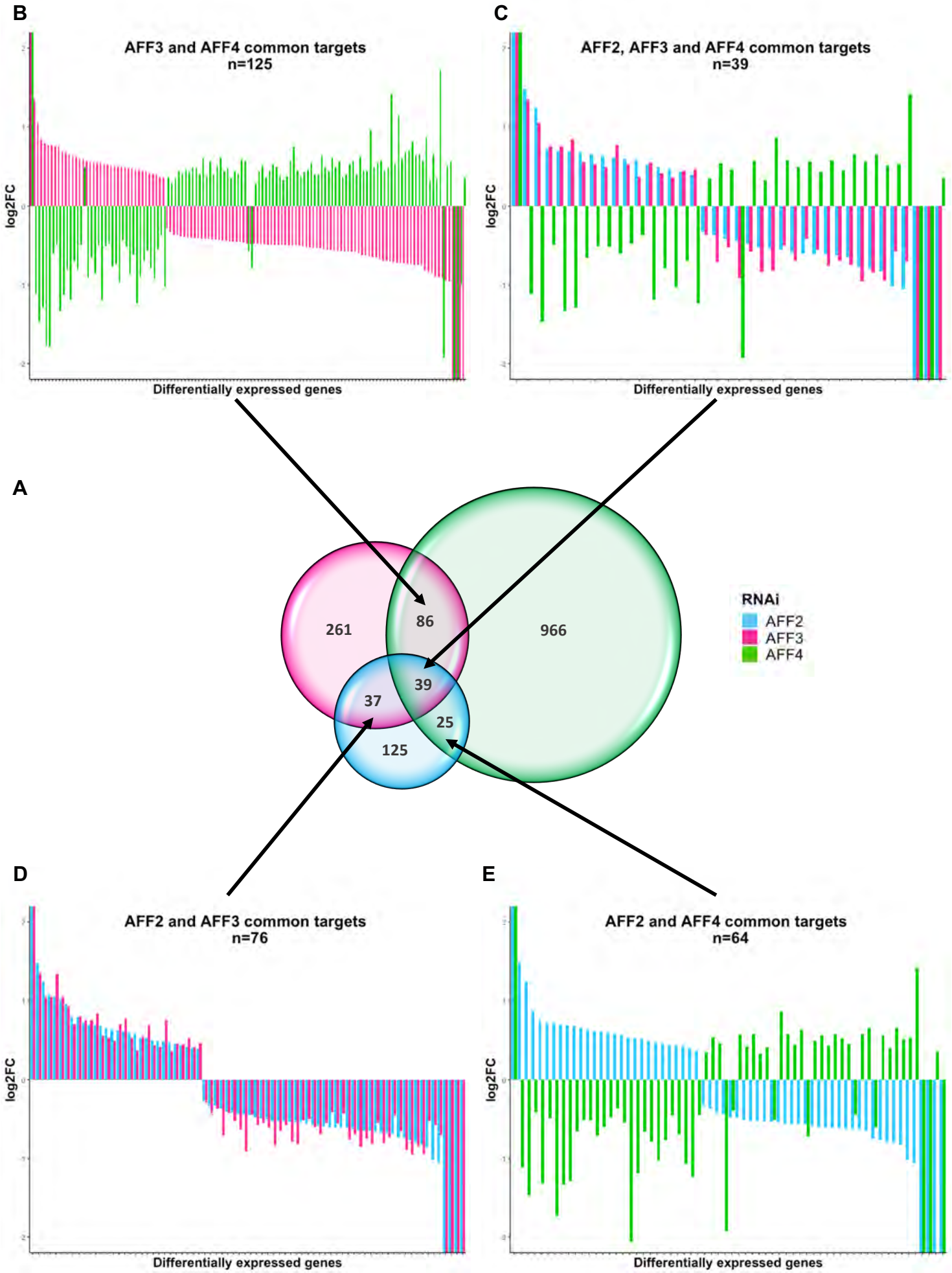

B
AFF3 and AFF4 common targets
n=125
log2FC
Differentially expressed genes
C
AFF2, AFF3 and AFF4 common targets
n=39
log2FC
Differentially expressed genes
A
261
86
966
39
37
25
125
RNAi
AFF2
AFF3
AFF4
D
AFF2 and AFF3 common targets
n=76
log2FC
Differentially expressed genes
E
AFF2 and AFF4 common targets
n=64
log2FC
Differentially expressed genes

Table S1. Phenotype of individuals with *AFF3* variants

Individual	Proband 1	Proband 2	Proband 3	Proband 4	Proband 5	Proband 6	Proband 7	Proband 8	Proband 9	Proband 10	Deletion proband*
Age [years]	17	3 years	18	21	5	9	8	8	8	11	4 months
Sex	M	F	M	F	F	F	F	F	M	M	F
AFF3 variant	p.(A258S)	p.(A258S)	p.(A258T)	p.(A258T)	p.(A258T)	p.(A258T)	p.(A258T)	p.(A258T)	p.(A258V)	p.(V260G)	500kb deletion
Inheritance	*De novo*	*De novo*	*De novo*	*De novo*	*De novo*	*De novo*	*De novo*	*De novo*	*De novo*	*De novo*	*De novo*
Neurodevelopmental anomalies											
Severe DD/ID	+	+	+	+	+	+	+	+	+	+	+
Epilepsy	Generalized tonic-clonic seizures (onset at 5 years), nocturnal, treatment with Keppra and Micropakine	Multifocal epileptiform discharges in bilateral posterior quadrant, no clinical seizures.	Generalized, tonic-clonic seizures (onset at 5 months), initially controlled with phenobarbital, developed drug-resistance, partially controlled with carbamazepine, clobazam, nitrazepam and phenobarbital	Generalized, tonic-clonic seizures	Generalized, tonic-clonic seizures	Complex partial with secondary generalization	Generalized, tonic-clonic seizures	Generalized (onset at 3 years), tonic-clonic seizures, inefficient treatments but remission since 6 years 9 months	-	Generalized, tonic-clonic seizures	Myoclonic jerks, convulsions
Muscle tone	Limb hypertonia, spastic tetraparesis	Generalized hypotonia	Axial hypotonia, peripheric hypertonia	Limb hypertonia	Hypotonia	Axial hypotonia, peripheric hypertonia	Hypotonia	Hypotonia	Normal	Hypotonia	N/A
Vision and hearing	Normal	Cortical visual impairment, hyperopic refractive error and small angle intermittent strabismus	Normal	Myopia, strabismus	Strabismus	Central vision loss due to occipital impairment, central progressive hearing loss	Normal	Strabismus inconstant	Conductive hearing loss (resolved), vision not yet formally assessed	Strabismus	N/A
Brain MRI	Enlargement of the ventricular system and peri-cerebral spaces, thin and irregular appearance of the corpus callosum	Partial agenesis of the corpus callosum, forshortened and undersulcated frontal lobes, small cerebellar vermis with mega cisterna magna and wide sylvian fissures	Pachygyria of frontal lobes, abnormal opercularization of insulae with thicker insular cortices, wide Sylvian fissures, cerebral atrophy, brainstem hypoplasia, increased volume of the trigones and occipital horns of the lateral ventricles	N/A	Cerebral atrophy with possible brainstem hypoplasia	Prominence of CSF spaces	Cerebral atrophy	Cerebral atrophy, pachygyria of frontal lobes	Normal	N/A	Brain atrophy, ventriculomegaly
Other	Stereotypic movements	-	-	-	-	-	-	-	Autism and ADHD	-	-

Craniofacial features											
Microcephaly	+ (plagiocephaly)	-	+	-	+	- (brachycephaly)	+	+	+	+	-(dolicocephaly)
Nose	Small nose, anteverted nares	Prominent columella	Large nose with bulbous nasal tip and low hanging columella	Large nose with bulbous nasal tip and low hanging columella	Small nasal tip	Bulbous nasal tip, low hanging columella with low insertion	Large nose with bulbous nasal tip and low hanging columella	Large nose with bulbous nasal tip and low hanging columella	Normal	N/A	N/A
Short philtrum	-	-	+	+	-	+ (smooth)	+	+	Smooth philtrum	N/A	N/A
Mouth	Wide mouth with square upper lip	Wide mouth, downturned corners, thin upper lip	Wide mouth with square upper lip	Wide mouth with square upper lip	Wide mouth with square upper lip	Wide mouth with downturned corners, thick lower lip vermillion	Prominent upper lip	Wide mouth	Thin upper vermillion, mild ankyloglossia (snipped), high arched palate	Ankyloglossia	N/A
Teeth and gum abnormalities	Small, gingival hypertrophy	Small, widely spaced, bruxism	+	+	Small, widely spaced	+	-	Widely spaced	Hypomineralisation	+	N/A
Chin	Prognathism	Normal	Micrognathia	Triangular chin	Normal	Micrognathia	Mild micrognathia	Prognathism	Micrognathia	N/A	N/A
Synophrys and hypertrichosis	+	-	+	+	+	+	+	+	+	N/A	N/A
Other	Hypertelorism, short neck	Full cheeks, mild facial asymmetry	-	-	-	Long palpebral fissures, low-set and posteriorly rotated large ears with a simple helix, facial asymmetry	Long palpebral fissures	Long palpebral fissures, low set ears, mild facial asymmetry, gingival hyperplasia diabetes	Slight metopic prominence	-	Short palpebral fissures, low set ears, short neck
Skeletal abnormalities											
Mesomelic dysplasia	-	Lower limbs	4 limbs	4 limbs	4 limbs	4 limbs	4 limbs	4 limbs	Mild lower limbs	Lower limbs	4 limbs
Arms and legs	Bilateral elbow dislocation	Bilateral fibular agenesis, short and curved tibia, bilateral Syme amputations with resection of cartilaginous fibular anlage and bilateral tibial osteotomies for angular deformity correction, fitted with bilateral lower extremity prosthetics at 2 years 3 months.	Madelung deformity, slender limb bones, fibular hypoplasia/agenesis	Short ulna and radius, radial head dislocation/subluxation, styloid process of ulna on radius, carpal coalition, hypoplastic femora, short and curved tibia with metaphyseal flaring, mid tibial dimples, deviated knees, hypoplastic and gracile fibula	Short and thick ulna, slightly shortened radius with convex distal end bilaterally, dislocation of right radial head, short and curved tibia, extremely short rectangular fibula	Short fibula, discoid meniscus, limited knee extension	Limited supination, radial head dislocation/subluxation, hypoplastic fibula	Short humerus, hypoplastic short fibula	Mild lateral bowing of both radii, normal bone age	Bowed radii, unilateral bowed ulna, shortened ulna, abnormal radial diaphysis, bowed and angulated tibias, hypoplastic fibula	Radial head dislocation/subluxation, slightly short radius and ulna, short and dysplastic triangular tibias, fibular agenesis
Hands and feet	Limited pronosupination, bilateral camptodactyly, edema of back-hands and -feet, bilateral simian creases, tapered fingers, dorsum pedis edema, small toes, hypoplasia of	Right single transverse and left bridged palmar crease, bilateral hypoplastic 4th metatarsals, absence of the 5th ray and phalanges of lateral toes, 4 splayed toes	Hypoplastic talipes, fusion of tarsal bones	Carpal coalition, small feet, hypoplastic left 5th, metatarsal synostosis	Talus ossified in hindfoot, one ossified bone in midfoot (cuneiform), missing one lateral ray in left foot, 4th-5th right metatarsal synostosis	Soft tissue syndactyly of fingers 3rd-4th, small feet, pes planus, 2nd toe overlapping hallux bilaterally	Limited supination, pes planus, broad toe tips	Transverse palmar crease, limited pronosupination, proximal deviation of thumbs, small feet, absent calcanei, broad 1th toes, polydactyly, cutaneous	Very broad feet. Bilateral hind foot varus deformity, mild metatarsus adductus. misshaped and large talus, cuboid and calcaneum, mild shortening and Y	Wide distal radial metaphyses, oligodactyly: 2 tarsal bones on each foot, absent/hypoplastic calcanei, 3 metatarsals, 3 associated phalanges, 1 phalanx not	Small equinovalgus feet, oligodactyly: 4 toes on 1 foot, 5 on the other, abnormally spaced

	distal phalanges, ungual hypoplasia, neonatal arthrogryposis							process on the side of the 5th finger and cutaneous syndactyly 3th-6th toes on left foot, four metatarsals and partial syndactyly 3th-4th toes on right foot	shaped fusion of the left 4th and 5th metatarsals, hyperplastic/dysplastic toenails, hand Xrays normal	associated with a metatarsal bone	
Ribs and spine	Scoliosis	13 rib-bearing thoracic-type vertebrae and 5 lumbar type vertebrae, hypoplastic L1 with focal kyphosis	Severe scoliosis, C2-C3 vertebral fusion, L5-S1 vertebral cleft	Scoliosis	Bilateral cervical ribs	Scoliosis, incomplete coronal cleft of T9 and T12 vertebrae, low lying spinal cord, termination of conus medullaris at upper border of L3	Pectus excavatum	Scoliosis, fusion 1th-2th ribs, sacral sinus	6 lumbar vertebrae, 13 pairs of ribs, sacral dimple	Scoliosis, cervical ribs, anterior superior vertebral notching, tethered cord	Sacral sinus
Hips and pelvis	Bilateral coxa valga, dislocation of the hips	Normal	Normal	Bilateral coxa valga with hypoplastic ilia, hip dislocation	Coxa valga	Bilateral coxa valga	Hip dislocation	Coxa valga, hip dysplasia	Normal	Coxa valga, unilateral hip dysplasia	N/A
Osteopenia	N/A	+	+ (osteoporosis)	+	-	+	+ (osteoporosis)	-	-	+	N/A
Additional features											
Horseshoe kidney	-	+	-	+	+	+	-	+	+	+	+
Gastro-intestinal symptoms	N/A	GERD, dysphagia, gastrostomy tube dependent, concerns for esophageal dysmotility ± abnormal gastric accommodation, abnormal gastric emptying with no evidence of small intestinal dysmotility	GERD, constipation	GERD, constipation	Constipation, swallowing difficulties, percutaneous endoscopic gastrostomy	GERD, gastrojejunostomy tube dependent, chronic constipation, hiatal hernia, pancreatitis	Constipation	Constipation, anal dystopia	Constipation, feeding difficulties due to floppy larynx and adenotonsillar hypertrophy, frequent obstruction	GERD, constipation	Colon malrotation
Failure to thrive	+	-	+	+	+	+	+	+	+	+	+
Respiratory problems	Apnea	Multicompartmental respiratory disease (upper airway obstruction, lower airway obstruction, ineffective mucociliary clearance, restrictive lung disease, aspiration and pneumonia), moderate to	Neonatal respiratory distress, recurrent pneumonia, frequent hiccups improved with carbamazepine, respiratory arrest leading to death at 21 years	-	Nightly desaturations treated with CPAP from 3 years of age	-	-	-	Severe obstructive sleep apnea, adenotonsillectomy at 3 years	-	Recurrent apnea, respiratory arrest leading to death at 4 months

		severe mixed sleep apnea, severe laryngomalacia status post supraglottoplasty at 18 months, cough assist and inhaled steroid and bronchodilator and supplemental oxygen with sleep, tonsillectomy and adenoidectomy planned			
Other	Bilateral cryptorchidism, bicuspid aortic valve	History of bilateral vesicoureteral reflux, grade II	Menstrual cycle perturbations	Popliteal pterygium	Short stature

Footnote: AFF3 variants are described according to RefSeq NP_002276.2. *__Steichen-Gersdorf *et al* (2008)__[15]
Abbreviations: DD= developmental delay, ID= intellectual disability, N/A = not available, ADHD= Attention deficit hyperactivity disorder, GERD= gastroesophageal reflux disease

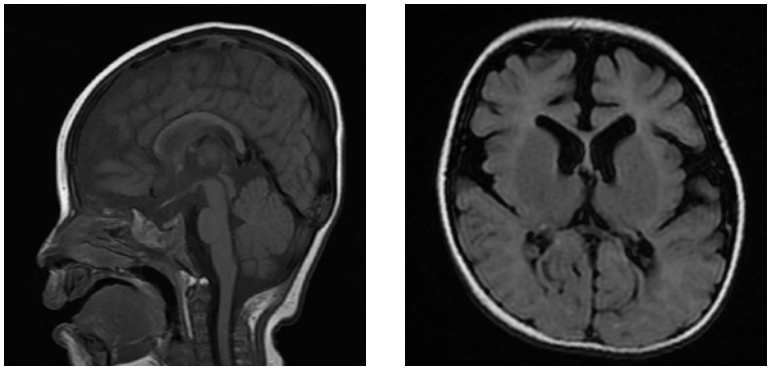


Figure S1: Brain MRI of proband 7 carrying a *de novo* variant in *AFF3* FLAIR (right) and T1 (left) at 9 months old

II. De novo PIK3R2 variant causes polymicrogyria, corpus callosum hyperplasia and focal cortical dysplasia.

This study of an 8-year-old Italian patient led to the discovery of a yet-unreported *de novo* missense variant in *PIK3R2* that was recently associated with megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) syndrome and bilateral perisylvian polymicrogyria (BPP)[67,68]. Besides expanding the number of patients presenting with this phenotypic spectrum and helping at its delineation, we described a variant outside of the SH2 domain whose variants were the only ones previously associated with this disorder. We modeled the effect of this residue change in its interface domain with the PhosphatidylInositol 4,5 bisphosphate 3 kinase catalytic subunit alpha (PIK3CA) using 3D structure.

I was in charge of the genetics analysis by exome sequencing, variants filtering and Sanger validation. I also took part of the writing process. A manuscript describing these results was recently published in the European Journal of Human Genetics[62] (see following pages).

European Journal of Human Genetics (2016) 24, 1359–1362

www.nature.com/ejhg

SHORT REPORT

De novo PIK3R2 variant causes polymicrogyria, corpus callosum hyperplasia and focal cortical dysplasia

Gaetano Terrone[1,9,10], Norine Voisin[2,9,10], Ali Abdullah Alfaiz[2,3,4,5], Gerarda Cappuccio[1], Giuseppina Vitiello[1], Nicolas Guex[2,3], Alessandra D'Amico[6], A James Barkovich[7], Nicola Brunetti-Pierri[1,8], Ennio Del Giudice[1,11] and Alexandre Reymond*[,2,11]

We report an 8-year-old boy with a complex cerebral malformation, intellectual disability, and complex partial seizures. Whole-exome sequencing revealed a yet unreported *de novo* variant in the *PIK3R2* gene that was recently associated with megalencephaly–polymicrogyria–polydactyly–hydrocephalus (MPPH) syndrome and bilateral perisylvian polymicrogyria (BPP). Our patient showed cerebral abnormalities (megalencephaly, perisylvian polymicrogyria, and mega corpus callosum) that were consistent with these conditions. Imaging also showed right temporal anomalies suggestive of cortical dysplasia. Until now, only three variants (c.1117G>A (p.(G373R)), c.1126A>G (p.(K376E)) and c.1202T>C (p.(L401P))) affecting the SH2 domain of the PIK3R2 protein have been reported in MPPH and BPP syndromes. In contrast to the variants reported so far, the patient described herein exhibits the c.1669G>C (p.(D557H)) variant that affects a highly conserved residue at the interface with the PI3K catalytic subunit α. The phenotypic spectrum associated with variants in this gene and its pathway are likely to continue to expand as more cases are identified.
European Journal of Human Genetics (2016) **24,** 1359–1362; doi:10.1038/ejhg.2016.7; published online 10 February 2016

INTRODUCTION

Megalencephaly-related (MEG) syndromes are overgrowth disorders presenting with increased brain volume, cortical malformations, developmental vascular anomalies, benign mesenchymal masses, and distal limb malformations.[1] Somatic-activating variants in the PI3K–AKT–mTOR pathway have been recently recognized as causative of these overgrowth disorders.[2] Postzygotic *de novo* variants in *PIK3CA* are responsible for the PIK3CA-related overgrowth spectrum (PROS), including the megalencephaly-capillary malformation syndrome (MCAP) and hemimegalencephaly (HMEG) syndrome.[3,4] *De novo* germline variants in *PIK3R2* and *AKT3* genes, which are upstream components of the mTOR pathway, have been found in patients with megalencephaly–polymicrogyria–polydactyly–hydrocephalus syndrome (MPPH).[2,5] More recently, *de novo* variants in *Cyclin D2* (*CCND2*), a gene downstream of the PI3K–AKT pathway were reported in patients with MPPH,[6] whereas germline and somatic variants in *PIK3R2* were reported from patients affected by bilateral perisylvian polymicrogyria (BPP).[5] Although the phenotypes of these syndromes largely overlap, they can be distinguished based on somatic features because, in contrast to MCAP, MPPH lacks skin vascular malformations, somatic overgrowth, connective tissue dysplasia, and syndactyly.[2] Somatic *PIK3CA* variants are also a common cause of isolated lymphatic malformations in patients without brain involvement.[7] The only somatic features reported in MPPH include postaxial polydactyly, found in ~41% of patients, and mild facial dysmorphic features such as prominent forehead, low nasal bridge, and hypertelorism, that are likely secondary to the increased brain volume.[1] Additional brain abnormalities reported in these patients include cerebellar tonsillar ectopia (Chiari 1 malformation), whereas increased thickness of corpus callosum was observed in ~7% of cases.[1]

Here, we report a patient with MPPH syndrome harboring a previously unreported *de novo* variant in *PIK3R2* gene detected by whole-exome sequencing.

SUBJECT AND METHODS

Patient

The reported patient of Italian origin was diagnosed at the Department of Translational Medicine, Section of Pediatrics and the Department of Diagnostic Imaging, Neuroradiology Unit, Federico II University, Naples, Italy. His parents gave written informed consent for this study.

Methods

To uncover genetic variants associated with the abnormalities shown by the patient, we performed whole-exome sequencing of DNA extracted from blood of the proband, both his parents and his unaffected brother (Supplementary Figure S1) as previously described.[8,9] Briefly, exomes were captured using the Agilent SureSelect Human All Exon V4 enrichment kit (Agilent, Santa Clara, CA, USA) and sequenced on an Illumina HiSeq platform. Variants were filtered based on adherence to either an autosomal recessive inheritance pattern or as being *de novo* in the proband. Further filtering was based on functional candidates among the 33 remaining genes using GeneOntology and OMIM databases. Variants with MAF<0.1 in control populations of European descent and predicted to be deleterious by SIFT[10] and/or PolyPhen-2[11] were prioritized.

[1]Department of Translational Medicine, Section of Pediatrics, Federico II University, Naples, Italy; [2]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; [3]Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland; [4]Bioinformatics Section, King Abdullah International Medical Research Center (KAIMRC), Riyadh, Kingdom of Saudi Arabia; [5]King Saud Bin Abdulaziz University for Health Sciences, Riyadh, Kingdom of Saudi Arabia; [6]Department of Diagnostic Imaging, Neuroradiology Unit, Federico II University, Naples, Italy; [7]Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA, USA; [8]Telethon Institute of Genetics and Medicine, Pozzuoli, Naples, Italy
[9]These authors contributed equally to this work.
[10]Co-first authors.
[11]Co-senior authors.
*Correspondence: Dr A Reymond, Center for Integrative Genomics, University of Lausanne, Genopode building, Lausanne CH-1015, Switzerland. Tel: +41 21 692 39 60; Fax: +41 21 692 39 65; E-mail: alexandre.reymond@unil.ch
Received 29 September 2015; revised 8 December 2015; accepted 16 December 2015; published online 10 February 2016

PCR and Sanger sequencing were used to validate their proper segregation in the family. The model of PIK3R2/PIK3CA interaction was obtained using UNIPROT entry O00459 and Swiss-PdbViewer-DeepView v4.1. All *PIK3R2* variants mentioned in the text are described according to transcript NM_005027.3 and protein NP_005018.1. The newly identified *PIK3R2* variant was deposited in the Leiden Open Variation Database.

RESULTS

Clinical findings

The patient presented with reduced mobility of his left upper arm, developmental delay, and macrocephaly (occipitofrontal circumference (OFC) 48 cm; >98th centile; Supplementary Figure S2A) at 8 months. At 2 years, he showed a left spastic hemiplegia and dysmorphic features (Supplementary Figure S2B), including synophrys, depressed nasal bridge, anteverted nares, *pectus excavatum*, broad thumb and hallux. A brain MRI, at age 3, revealed hyperplasia of corpus callosum (Figure 1) and asymmetrical bilateral polymicrogyria (PMG; Figure 1a and Supplementary Figure S3A). Reduced myelination of the underlying white matter of the anterior right temporal lobe was suggestive of a focal cortical dysplasia (FCD) type I (Figure 1b). A thick cortical infolding and blurring of a single sulcus of the anterior right insula suggested FCD type IIa; lateral to this lesion, a cyst was identified in the frontal right operculum cortex (Figure 1c and Supplementary Figure S3C). It was unclear how (if) the cystic area was related to the dysplasia. The right thalamus and right cerebral peduncle/brain stem were reduced in size (Supplementary Figure S3B). The left hemisphere had coarse PMG involving perisylvian and suprasylvian cortex (Supplementary Figure S3D). By 4 years of age, the patient developed complex partial seizures characterized by leftward gaze deviation, reduction of muscle tone and loss of consciousness with falls during prolonged epileptic discharges. Antiepileptic therapy with valproic acid, topiramate and clobazam resulted in good seizure control. Interictal-EEG revealed bursts of slow spike/polyspike-and-wave complexes worsening during sleep, especially in right temporal–occipital regions (Supplementary Figure S4).

Genetic findings

Exome sequencing resulted in five variants that complied with the above filtering (Supplementary Table S1). While we identified in the proband homozygous variants in *PSPN* and *TSEN54*, and compound heterozygous variants in *CCDC41* that complied with an autosomal recessive inheritance, we also uncovered a *de novo* variant in the *PIK3R2* gene, a gene previously found in patients with MCAP and MPPH.[2] The identified variant is a G to C substitution in exon 13 of the *PIK3R2* gene (chr19:18,278,049 [hg19]), which modifies a highly conserved aspartic acid (Supplementary Figure S5) into a histidine residue at the amino acid position 557 [NM_005027.3: c.1669G>C; NP_005018.1: p.(D557H)]. Although somatic variants in components of the PI3K–AKT3–mTOR pathway were previously shown to cause hemimegalencephaly,[4] the almost 50:50 proportion of G and C nucleotides at position chr19:g.18,278,049 in both high throughput and Sanger sequencing is consistent with a *bona fide* germline variant. Specifically, we obtained 41 (57%) and 31 (43%) Illumina sequencing reads corresponding to the reference and the alternative (*de novo*) allele, respectively, whereas the Sanger chromatogram shows equal size peaks (Supplementary Figure S1).

Structural modeling

The PI3K kinase complex is composed of a catalytic moiety, encoded by the *PIK3CA, PIK3CB* or *PIK3CD* gene, and one of five different regulatory subunits, such as PIK3R2 (Supplementary Figure S6A). In contrast to the already reported *PIK3R2* variants associated with this disorder, (c.1117G>A (p.(G373R)), c.1126A>G (p.(K376E)) and c.1202T>C (p.(L401P)))[2,5,12,13] which are positioned within the region encoding the first Src homology 2 (SH2) domain of the protein, the c.1669G>C (p.(D557H)) lies within another interface between the two subunits of the kinase (Figure 1e; Supplementary Figure S6A and B). Both the D557 and the neighboring N561 residues of PIK3R2 generate H-bonds with the side-chain nitrogen atom of the N345 residue of PIK3CA (Figure 1f and Supplementary Figure S6C). It is expected that the c.1669G>C (p.(D557H)) variant would fit nicely in this environment (Supplementary Figure S6D). We hypothesize that the lack of a negative charge does not perturb the interaction as there are no positively charged residues in this portion of the α-catalytic unit. It appears, however, that an ~311 cubic angstroms (Å) groove that could accommodate a small compound is present between the two subunits of the kinase (Figure 1g). Consistent with this hypothesis, in the 4l2y crystal structure, a molecule of glycerol is located at the entry of this groove. The direct contact by the tip of the D557 side chain with the surface of the groove (Figure 1g), as well as the negative charge of this residue could be essential in positioning an adaptor molecule that acts in between the two kinase units *in vivo*. We hypothesize that a slightly larger and neutral or positively charged histidine variant at this position would result into a smaller groove, as well as altered hydrogen bonds with an adaptor molecule.

DISCUSSION

Although normocephalic patients with *PIK3R2* variants were reported,[5] our patient shares MEG, bilateral perisylvian PMG, mega corpus callosum, intellectual disability, abnormal muscle tone, spasticity, and epilepsy with the 12 previously described MPPH syndrome patients with *PIK3R2* variants.[2,12] Cortical brain malformations were also reported in PROS, particularly in MCAP syndrome with megalencephaly its most consistent feature (OFC>2.5–10 SD).[1] Our patient showed a mild asymmetric brain overgrowth, a feature more frequently observed in PROS than MPPH.

The most common pattern of PMG reported in MCAP and MPPH patients is bilateral perisylvian PMG that often extends beyond the perisylvian region.[14] Other patterns occur such as focal PMG or PMG involving bilateral frontal lobes.[1,14] Although the extent of PMG involvement is bilateral, many patients show a mildly asymmetric presentation, as observed in our case.[5] The proband's brain MRI revealed PMG with broad cerebral involvement extending to the frontoparietal cortex; in addition a sulcus with cortical-white matter blurring in the anterior right insula and abnormal white matter signal in right anterolateral temporal lobe, and an infolding of thick cortex with cortical blurring in the right frontal operculum are consistent with FCD,[15] a malformation not yet described in MPPH. Nevertheless, FCD was observed in PROS, highlighting the overlap between these clinical entities.[3] Brain-mosaic-activating variants of PI3K–AKT–mTOR were identified in patients with isolated FCD type II, HMEG, MEG and intractable epilepsy without MRI-identifiable lesions.[16,17] The patient brain MRI also revealed a small area with cyst-like appearance, adjacent to the region of dysplastic cortex in the frontal right operculum that could represent either a small dysembryoplastic neuroepithelial tumor or other cyst, a feature not reported in previous cases. Considering the risk of brain cancer,[2] an ongoing surveillance should be considered.

Epilepsy is another distinctive phenotypic characteristic of MPPH syndrome. A wide spectrum of seizures was reported ranging from complex partial seizures to infantile spasms.[5] The

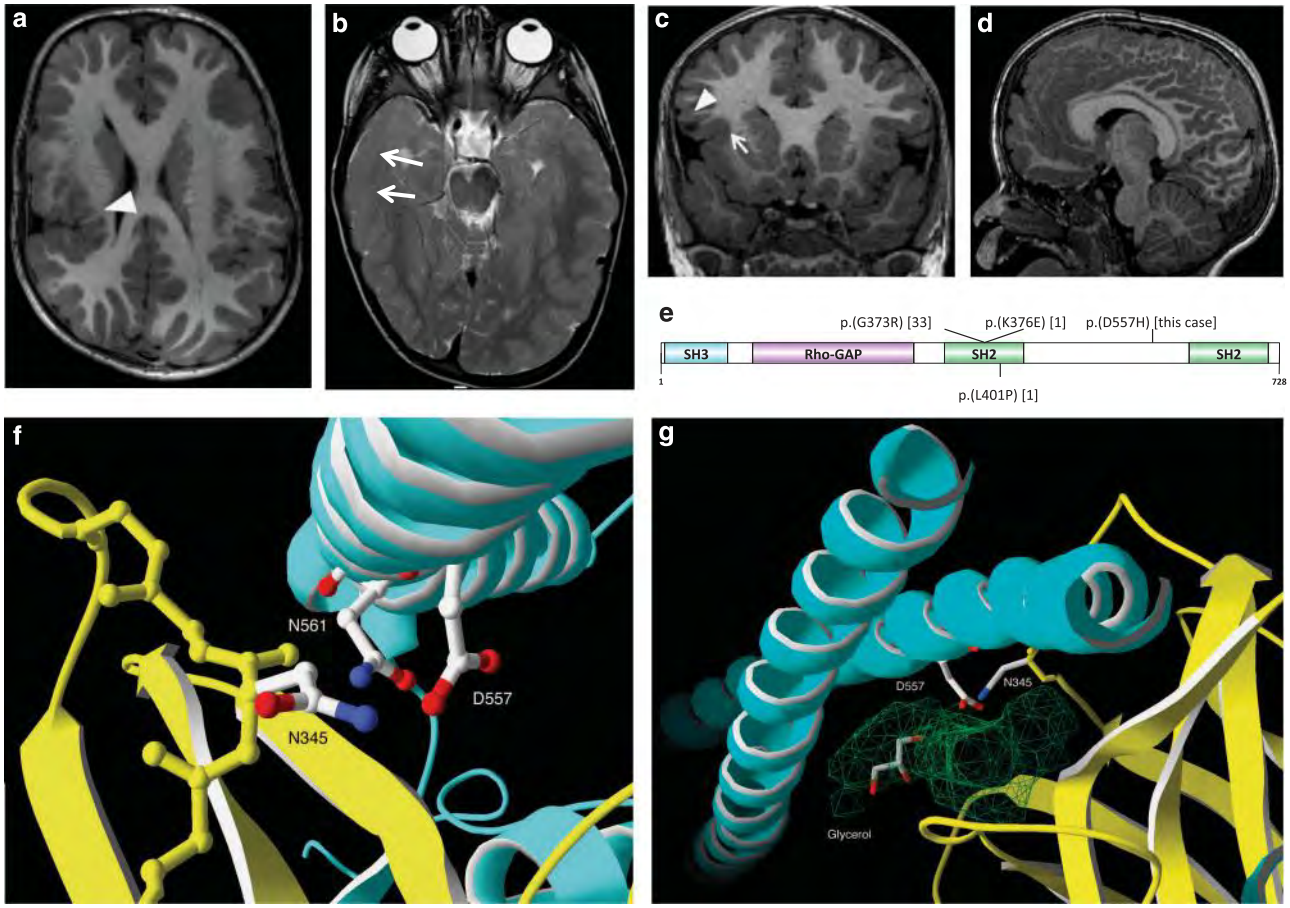


Figure 1 Brain MRI, localization and modeling of PIK3R2-mutated residue. (**a-d**) Brain MRI of the patient. (**a**) T1-weighted axial image: bilateral frontoparietal polymicrogyria and incomplete perisylvian opercularization of the right hemisphere (white arrowhead). (**b**) T2-weighted axial image: thickness and blurring of the cortex-white matter junction in the right temporal cortex (white arrows) with diminished myelination of the underlying white matter and hypoplasia of right cortico-spinal tract. (**c**) T1-weighted coronal image shows thick cortical infolding with subtle cortical-white matter blurring in the anterior right insula (white arrow) and a cystic lesion in the right frontal operculum (white arrowhead). (**d**) T1-weighted mid-sagittal image: increased thickness of corpus callosum. Additional brain MRI images are shown in Supplementary Figure S3. (**e**) Schematic representation of the PIK3R2 protein domains and positions of the variants reported in literature. Thirty-three cases with c.1117G>A (p.(G373R)) *de novo* and mosaic variants, one case with a *de novo* c.1126A>G (p.(K376E)) variant and one case with a p.(L401P) *de novo* variant[2,5,12,13] were reported beside the c.1669G>C (p.(D557H)) patient described here. SH3, Src homology 3 domain; Rho-GAP, Rho GTPase-activating protein domain; SH2, Src homology 2 domain. (**f**, **g**) Context of the c.1669G>C (p.(D557H)) variant. The PIK3CA (uniprot P42336) and PIK3R2 (uniprot O00459) are shown in yellow and cyan ribbon diagrams according to the PDB entry 4l2y model, respectively. (**f**) Detail view of the interaction showing PIK3R2 D557 and PIK3CA N345 forming an H-bond. (**g**) Groove of ~311 Å^3 (in green) with a glycerol molecule present in the pdb entry complex showing PIK3R2 D557 residue in direct contact with its surface. A general view of the interaction between PIK3R2 and PIK3CA and positioning of H557 is shown in Supplementary Figure S6.

seizure presentation is predominantly focal, even if no clear recurrent epilepsy pattern is described. In fact, atypical absences, myoclonic jerks, generalized tonic–clonic or complex febrile seizures are reported in these individuals.[5,14] Our patient showed focal seizures characterized by dyscognitive features with a critical onset in the right hemisphere. As described in asymmetrical bilateral PMG, our case had epilepsy laterality with more severe and difficult-to-treat seizures.[18]

In absence of consistent dysmorphic features in MPPH syndrome, it is worthwhile pinpointing phenotypic characteristics, including synophrys, broad-looking thumbs, large great toes, shared by our patient with two cases recently described.[5]

Although the PIK3R2 *de novo* variant identified in the proband does not lie within the SH2 domain, whose variants were previously associated with this disorder, it is the most likely causative variant as it affects a highly conserved residue at the interface with PIK3CA. This variant putatively modifies the size and shape of the groove separating the two kinase subunits.

This case extends the spectrum of MPPH syndrome and highlights the role of the interface domain of PIK3R2.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Authors thank the family for its contribution and the members of the Lausanne Genomic Technologies Facility. AAA is recipient of a scholarship from the Saudi Arabian National Guard Health Affairs. This work was supported by a grant of the Swiss National Science Foundation 31003A_160203 and the Lithuanian-Swiss Cooperation Programme (AR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

NOTE ADDED IN PROOF

During the revision of this manuscript, Mirzaa *et al.*[5] reported *PIK3R2* variants in BPP, we modified our manuscript to discuss our results in view of these recently published data to help potential readers getting a complete view of current knowledge.

1 Mirzaa GM, Riviere JB, Dobyns WB: Megalencephaly syndromes and activating mutations in the PI3K-AKT pathway: MPPH and MCAP. *Am J Med Genet C Semin Med Genet* 2013; **163C**: 122–130.
2 Riviere JB, Mirzaa GM, O'Roak BJ *et al*: De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. *Nat Genet* 2012; **44**: 934–940.
3 Keppler-Noreuil KM, Rios JJ, Parker VE *et al*: PIK3CA-related overgrowth spectrum (PROS): diagnostic and testing eligibility criteria, differential diagnosis, and evaluation. *Am J Med Genet A* 2015; **167A**: 287–295.
4 Lee JH, Huynh M, Silhavy JL *et al*: De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. *Nat Genet* 2012; **44**: 941–945.
5 Mirzaa GM, Conti V, Timms AE *et al*: Characterisation of mutations of the phosphoinositide-3-kinase regulatory subunit, PIK3R2, in perisylvian polymicrogyria: a next-generation sequencing study. *Lancet Neurol* 2015; **14**: 1182–1195.
6 Mirzaa GM, Parry DA, Fry AE *et al*: De novo CCND2 mutations leading to stabilization of cyclin D2 cause megalencephaly-polymicrogyria-polydactyly-hydrocephalus syndrome. *Nat Genet* 2014; **46**: 510–515.
7 Luks VL, Kamitaki N, Vivero MP *et al*: Lymphatic and other vascular malformative/overgrowth disorders are caused by somatic mutations in PIK3CA. *J Pediatr* 2015; **166**: 1048–1054, e1041-1045.
8 Alfaiz AA, Micale L, Mandriani B *et al*: TBC1D7 mutations are associated with intellectual disability, macrocrania, patellar dislocation, and celiac disease. *Hum Mutat* 2014; **35**: 447–451.
9 Borck G, Hog F, Dentici ML *et al*: BRF1 mutations alter RNA polymerase III-dependent transcription and cause neurodevelopmental anomalies. *Genome Res* 2015; **25**: 609.
10 Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009; **4**: 1073–1081.
11 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
12 Nakamura K, Kato M, Tohyama J *et al*: AKT3 and PIK3R2 mutations in two patients with megalencephaly-related syndromes: MCAP and MPPH. *Clin Genet* 2014; **85**: 396–398.
13 Tapper WJ, Foulds N, Cross NC *et al*: Megalencephaly syndromes: exome pipeline strategies for detecting low-level mosaic mutations. *PLoS ONE* 2014; **9**: e86940.
14 Mirzaa GM, Conway RL, Gripp KW *et al*: Megalencephaly-capillary malformation (MCAP) and megalencephaly-polydactyly-polymicrogyria-hydrocephalus (MPPH) syndromes: two closely related disorders of brain overgrowth and abnormal brain and body morphogenesis. *Am J Med Genet A* 2012; **158A**: 269–291.
15 Blumcke I, Thom M, Aronica E *et al*: The clinicopathologic spectrum of focal cortical dysplasias: a consensus classification proposed by an ad hoc Task Force of the ILAE Diagnostic Methods Commission. *Epilepsia* 2011; **52**: 158–174.
16 Jansen LA, Mirzaa GM, Ishak GE *et al*: PI3K/AKT pathway mutations cause a spectrum of brain malformations from megalencephaly to focal cortical dysplasia. *Brain* 2015; **138**: 1613–1628.
17 Lim JS, Kim WI, Kang HC *et al*: Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy. *Nat Med* 2015; **21**: 395–400.
18 Shain C, Ramgopal S, Fallil Z *et al*: Polymicrogyria-associated epilepsy: a multicenter phenotypic study from the Epilepsy Phenome/Genome Project. *Epilepsia* 2013; **54**: 1368–1375.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)

III. Inflammatory Myopathy in a Patient with Aicardi-Goutières Syndrome

This study focused on an 11-year-old female Lithuanian patient with symptoms reminiscent of CK syndrome[69,70] (OMIM 300831). As this X-linked recessive disorder is associated with *NSDHL* variants[71], and targeted sequencing were negative, we performed exome sequencing of the proband, her parents and her two healthy siblings. This allowed us to diagnose type I interferonopathy Aicardi–Goutières syndrome 1 (AGS; OMIM 225750), caused by a previously described homozygous missense mutation in *TREX1* gene[72]. Histological analysis showed that our patient exhibits inflammatory myopathy, expanding the range of AGS clinical symptoms and suggesting a possible mitochondrial damage triggering the autoimmune response in this syndrome.

I was in charge of the genetic diagnosis of AGS, analyzing exomes and identifying the causal variant in accordance to known syndromes and clinical features of the patient. A manuscript describing these results was recently published in the European Journal of Medical Genetics[63] (see following pages).

European Journal of Medical Genetics 60 (2017) 154–158

Contents lists available at ScienceDirect

European Journal of Medical Genetics

journal homepage: http://www.elsevier.com/locate/ejmg

Inflammatory myopathy in a patient with Aicardi-Goutières syndrome

Birutė Tumienė [a, *], Norine Voisin [b], Eglė Preikšaitienė [a], Donatas Petroška [c, d], Jurgita Grikinienė [e], Rūta Samaitienė [e], Algirdas Utkus [a], Alexandre Reymond [b], Vaidutis Kučinskas [a]

[a] Department of Human and Medical Genetics, Centre for Medical Genetics, Vilnius University, Vilnius, Lithuania
[b] Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland
[c] Department of Pathology, Forensic Medicine and Pharmacology, Faculty of Medicine, Vilnius University, Vilnius, Lithuania
[d] National Centre of Pathology, Vilnius University Hospital Santariškių Klinikos, Vilnius, Lithuania
[e] Clinic of Children's Diseases, Faculty of Medicine, Vilnius University, Vilnius, Lithuania

ARTICLE INFO

Article history:
Received 8 August 2016
Received in revised form
28 October 2016
Accepted 10 December 2016
Available online 9 January 2017

Keywords:
Inflammatory myopathy
Aicardi-Goutières syndrome
Type I interferonopathy
Autoimmune myositis
Exome sequencing

ABSTRACT

Aicardi–Goutières syndrome (AGS) is an inflammatory disorder belonging to the recently characterized group of type I interferonopathies. The most consistently affected tissues in AGS are the central nervous system and skin, but various organ systems and tissues have been reported to be affected, pointing to the systemic nature of the disease. Here we describe a patient with AGS due to a homozygous p.Arg114His mutation in the *TREX1* gene. The histologically proven inflammatory myopathy in our patient expands the range of clinical features of AGS. Histological signs of muscle biopsies in the proband, and in two other AGS patients described earlier, are similar to those seen in various autoimmune myositises and could be ascribed to inappropriate IFN I activation. In view of signs of possible mitochondrial damage in AGS, we propose that mitochondrial DNA could be a trigger of autoimmune responses in AGS.

1. Introduction

Aicardi–Goutières syndrome (AGS) is an inflammatory disorder caused by mutations in any of seven genes involved in nucleic acid metabolism/signaling and belonging to a recently characterized group of inborn errors of immunity termed type I interferonopathies (Crow et al., 2015). The hallmark of these diseases is an inappropriate up-regulation of type I interferon (IFN I) with further induction of IFN I-mediated innate and adaptive immune responses. Increased expression of IFN I-stimulated genes in serum and cerebrospinal fluid, so-called IFN I signature, was observed in 90% (74 of 82) of patients with mutation-proven AGS (Rice et al., 2013). Four human diseases were associated with *TREX1* mutations, including Aicardi-Goutières syndrome 1 (OMIM #225750), familial chilblain lupus (FCL, OMIM #610448), systemic lupus erythematosus (SLE, OMIM #152700), and retinal vasculopathy with cerebral leukodystrophy (OMIM #192315) (Rice et al., 2015). The most severe of these nosologies, AGS1, can encompass the features of both FCL and SLE (Crow et al., 2015; Ellyard et al., 2014; Ramantani et al., 2010). Correspondingly, families segregating both FCL and AGS through several generations have been described (Abe et al., 2013). SLE was the first human disease consistently associated with increased levels of IFN I and IFN I signature (Kalliolias and Kirou, 2012). Moreover, activation of the IFN I pathway in the blood and affected tissues has been found in a vast variety of systemic autoimmune diseases, including systemic sclerosis, Sjögren's syndrome, dermatomyositis, polymyositis, and rheumatoid arthritis (Kalliolias and Kirou, 2012; Greenberg, 2010; Greenberg et al., 2012; Hayashi et al., 2015). All these disorders can present considerable clinical overlap and affect almost any human organ or tissue (Kalliolias and Kirou, 2012).

The most consistently affected tissues in AGS are the central nervous system and skin, but a vast range of clinical features in various organ systems and tissues have been reported, pointing to the systemic nature of the disease (Crow et al., 2015; Ramantani et al., 2010). Here we describe a patient with histologically proven inflammatory myopathy and AGS due to a homozygous p.(Arg114His) missense variant in the *TREX1* gene. Correlation with earlier published cases of muscle biopsy investigations in AGS and

* Corresponding author. Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Santariškių 2, LT-08661, Vilnius, Lithuania.
E-mail addresses: tumbir@gmail.com, birute.tumiene@santa.lt (B. Tumienė).

http://dx.doi.org/10.1016/j.ejmg.2016.12.004
1769-7212/

with histopathological signs of autoimmune myositis is provided with suggestions concerning possible common pathogenetic mechanisms in all these human pathologies.

2. Case report

The proband, an 11-year-old female at the time of diagnosis, was the youngest child of healthy unrelated Lithuanian parents. Although she had two older healthy male and female siblings, her oldest brother died at the age of 2.5 months with a presumable, though not laboratory-confirmed, perinatal infection, neonatal seizures, and focal brain lesions in neurosonoscopy. She was born at term with birth weight and birth length in the 50th and the 75th centile respectively and occipitofrontal head circumference (OFC) in the 3rd centile (32 cm). She had an uneventful postnatal period and was considered healthy until the age of three months, but retarded early psychomotor development milestones such as lack of a smile and head lag were reported retrospectively. At the age of three months, she became extremely irritable, with inconsolable crying, pyrexia, feeding difficulties, infantile spasms, and marked startle reactions provoked by acoustic, tactile and other stimuli. Multiple brain calcifications were observed in a head CT already at that age. Developmental regress with progressive microcephaly and failure to thrive became evident. She subsequently developed spastic tetraparesis with truncal hypotonia, blindness with pale optic nerve discs, and cardiac arrhythmia, with episodes of paroxysmal supraventricular tachycardia, multifocal extrasystoles, and atrioventricular/sinoatrial heart block with normal findings on echocardiography. Right eye and subsequently left eye glaucoma developed at the age of nine and ten years. Sensorineural hearing loss was diagnosed at the age of two years. Episodes of sterile pyrexia lasting one to two days were observed three to eight times per year, with increasing frequency at the age of twelve to thirteen years of age. The patient suffered several episodes of pneumonia during her life. Chilblain lesions on her toes and feet appeared at the age of 11 years and increased in frequency at the age of twelve years. She also developed severe kyphoscoliosis and contractures of her knees, ankles and elbows and generalized muscle atrophy, as well as being of short stature and hypotrophic. The patient had severe intelectual disability with absent psychosocial and motor functions. Epileptic spasms and startle reactions provoked by acoustic stimuli were observed several times per day. At the age of 6 years, electroencephalographic records showed a slow disorganized background with intermittent spikes and sharp waves, more pronounced frontotemporally on the right. At the age of 11 years, groups of spikes over the right parietooccipital, deformed complexes of spikes and waves, and high *theta* waves over the right parietotemporal regions were registered in an EEG. Multiple confluent calcifications periventriculary, subcortically, in the basal ganglia, and in the cerebellar white matter and marked diffuse cerebral and cerebellar atrophy were observed in a head CT at the age of 11 years. She had never had any endocrinological abnormalities. The patient developed severe dysphagia necessitating gastrostoma, respiratory insufficiency, increasingly frequent pyrexias, and chilblain lesions and passed away at the age of thirteen years (Fig. 1).

Genetic counseling with subsequent genetic testing were provided to the proband at the age of 11 years. Blood and urine tests for inborn errors of metabolism and karyotyping did not demonstrate any diagnostic abnormalities. Muscle biopsy, except from several isolated cytochrome oxidase (COX)-negative fibers, failed to reveal signs suggestive of primary mitochondrial diseases. Marked perivascular inflammatory infiltrates with lymphocytes and isolated plasmocytoid dendritic cells, not extending into the endomysium, were however observed (Fig. 2). There was ubiquitous up-regulation of MHC-I expression in otherwise structurally normal muscular fibers. To uncover genetic variants associated with the abnormalities shown by the proband, we performed whole exome sequencing of DNA extracted from her blood, her parents, and her two healthy siblings as previously described (Alfaiz et al., 2014, 2016). Briefly, exomes were captured using the Agilent SureSelect Human All Exon V5 enrichment kit and multiplex sequenced on an Illumina HiSeq 2500 platform. Variants were filtered according to the quality of the calling, their frequency in control populations, their adherence to either being *de novo* in the proband or with an autosomal recessive inheritance pattern, and their predictive impact on the function of the protein (see Supplementary methods, Supplementary Table S1). Five variants in four genes, all confirmed by Sanger Sequencing, complied with these filtering steps (Supplementary Table S2). We uncovered in particular that the proband was homozygote for a known pathogenic variant in *TREX1* (NM_033629.4:c.341G > A; NP_338599.1:(p.Arg114His)). This variant has an allele frequency of 0.00016 in ExAC Version 0.3 (http://exac.broadinstitute.org/). It is the most common of the known pathogenic *TREX1* variants, being found in 35 out of 70 (50%) of parents of Northern European ancestry with AGS due to *TREX1* mutations (Crow et al., 2015). These findings allow us to hypothesize that the deceased elder brother was homozygous for the same variant, as his clinical features matched AGS.

3. Discussion

Type I interferons play a pivotal role in the immune response to infection by influencing development of innate and adaptive immune responses. The outcomes of IFN I responses can however vary from clearing out the pathogens to chronic infection and autoimmune diseases (Ivashkiv and Donlin, 2014). One of the causes of recently ascribed to type I interferonopathies Aicardi-Goutières syndrome is a genetic alteration in the *TREX1* gene encoding the major mammalian and ubiquitous *3′-5′* DNA exonuclease (Crow et al., 2006, 2015). Though AGS is mostly considered to be a neurological disorder with onset in the neonatal period or early childhood and mostly stable disease course after the first episode of acute or subacute illness, it is known to affect multiple tissues and organs, including the skin, endocrinological and gastrointestinal systems, and heart. However, persistent throughout the whole life of patients and progressive disease process was suggested because of the persistently found up-regulation of IFN-I induced genes (IFN I signature) and recurrent episodes of fever and chilblains (Crow et al., 2015; Crow, 2015). Features or histological signs of inflammatory myopathy were not mentioned in earlier reports of AGS patients, but it is difficult to ascertain skeletal muscle damage in a profoundly disabled child with absent or severely disturbed motor functions. Of note, signs of myopathy in electromyography were observed in a child with AGS (Haaxma et al., 2010). Also, 3.3% of AGS cases from a large cohort of 374 patients had infantile-onset hypertrophic cardiomyopathy (Crow et al., 2015). *Trex1*$^{-/-}$ mice exhibit reduced survival and develop inflammatory myocarditis leading to progressive cardiomyopathy. Focal lymphoid aggregates were also seen in murine liver, lungs, and other tissues (Morita et al., 2004).

There are several reports of muscle biopsies performed in AGS patients. A patient with a *de novo* heterozygous p.(Asp18Asn) mutation in the *TREX1* gene and a mild presentation of AGS syndrome was found to have isolated ragged red fibers, COX-negative fibers, and decreased overall energy (ATP and creatine phosphate) production in the presence of normal activities of the individual respiratory chain complexes (Haaxma et al., 2010). Southern blot analysis of muscle tissues from an affected child and two affected fetuses demonstrated multiple mtDNA deletions in a family with AGS5 (OMIM#612952) due to a large homozygous deletion of the

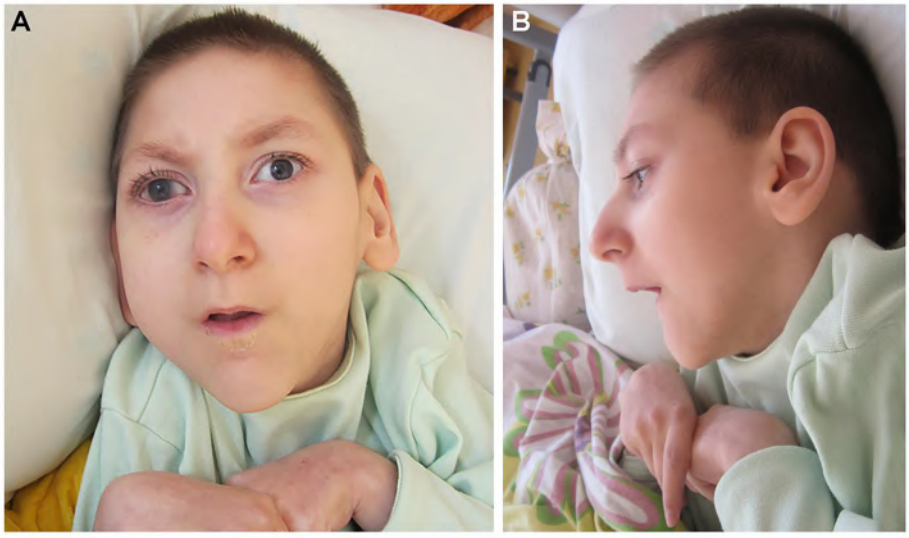


Fig. 1. (A and B). Phenotype of the patient at the age of 11 years: marked atrophy of facial and dorsal interosseous muscles, microcephaly, convex nasal bridge, and retrognathia.

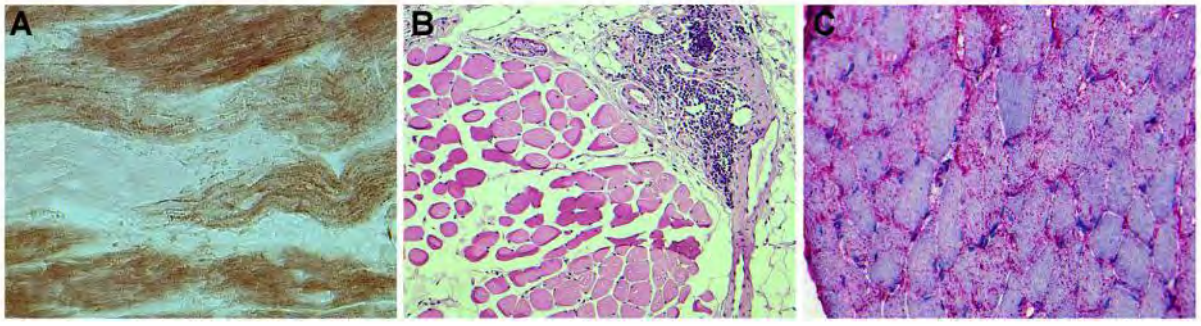


Fig. 2. Quadriceps muscle biopsy with isolated COX-negative fibers (**A**), perivascular inflammatory infiltrates with lymphocytes and isolated plasmocytoid dendritic cells, not extending into the endomysium (**B**), and homogenous ubiquitous up-regulation of MHC-I expression in otherwise structurally normal muscular fibers (**C**).

SAMHD1 gene (Leshinsky-Silver et al., 2011). There is no mention of any inflammatory signs in muscle biopsies in these cases. Interestingly, ragged red fibers, COX-negative fibers, and multiple mitochondrial DNA deletions are frequent findings in inclusion body myositis, a progressive autoimmune disorder with an insufficiently evaluated etiopathogenesis (Dalakas, 2011). Autoimmune myositis with inflammatory infiltrate in muscle biopsy is a constant feature of other diseases such as dermatomyositis and polymyositis and occurs with increased frequency in patients with other systemic autoimmune diseases, including systemic lupus erythematosus (SLE) (Dalakas, 2011). Perivascular and perimysial inflammatory infiltrates were recently found in 46% of biopsied patients (7 of 15) with SLE (Jakati et al., 2015). Interestingly, no correlation with clinical features or serological signs was observed. Perivascular inflammatory infiltrates of lymphocytes and dendritic plasmocytes is a constant histological feature of dermatomyositis (Greenberg, 2010; Dalakas, 2011). Plasmacytoid dendritic cells are the main producers of IFN I, and all these disorders were shown to be associated with an IFN I signature (Kalliolias and Kirou, 2012; Greenberg, 2010; Greenberg et al., 2012; Higgs et al., 2011).

It is well known that the accumulation of cytosolic nucleic acids activates the IFN I response (Crow et al., 2006; Atianand and Fitzgerald, 2013). The precise trigger of the IFN I response in AGS is unknown, and there are hypotheses implicating metabolites of DNA replication and repair (Yang et al., 2007) or fossil genome retroelements in IFN I activation (Crow et al., 2006; Stetson et al., 2008). Almost half of the human genome consists of retroelements and many of them are still active. The origins of the disease in a substantial proportion of patients have a prenatal onset (22.8%) in the biggest cohort to date of 374 AGS patients (Crow et al., 2015) when confrontation with exogenous sources of nucleic acids is limited. Interestingly, treatment with three reverse transcriptase inhibitors in doses comparable to those given to patients with HIV substantially reduced mortality and inflammatory myocarditis in *Trex1*$^{-/-}$ mice, further pointing to the role of genome retroelements in the activation of autoimmunity (Beck-Engeser et al., 2011). However, since mitochondria are evolutionary endosymbionts containing DNA similar to bacterial DNA, there is a possibility that mitochondrial DNA could be an additional trigger of autoimmunity. Indeed, as was recently shown, mitochondrial DNA that escapes from autophagy cells autonomously leads to Toll-like receptor (TLR) 9 mediated inflammatory responses in cardiomyocytes and is capable of inducing myocarditis and dilated cardiomyopathy (Oka et al., 2012). There is a possibility that a secondary mitochondrial damage as evidenced by ragged red fibers, COX-negative fibers, and multiple DNA deletions in skeletal muscle biosies of AGS patients (Haaxma et al., 2010; Leshinsky-Silver et al., 2011) and accumulating mitochondrial DNA are further activators of autoimmune responses in AGS.

In conclusion, histologically proven inflammatory myopathy in

our patient expands further the range of clinical features in AGS syndrome. Histological signs of muscle biopsies in our patient and in two other genetically confirmed AGS patients described earlier are similar to those seen in various autoimmune myosites and could be ascribed to a common pathogenetic mechanism of inappropriate IFN I activation. Fossil genome retroelements and other substrates of TREX1 such as metabolites of DNA replication and repair have been suggested as a possible trigger of autoimmune responses in AGS; in view of signs of possible mitochondrial damage in AGS and a recent publication linking accumulated mitochondrial DNA with autoimmune myocarditis, we propose that mitochondrial DNA could be another trigger of autoimmune responses in AGS. However, the latter hypothesis warrants further evaluation.

Ethical approval

The Regional Research Bioethics Committee approved this study. All study participants provided written informed consent.

Funding

This work was supported by grants from the Lithuanian–Swiss cooperation program to reduce economic and social disparities within the enlarged European Union (No. CH-3-SMM-0104, Unigene project) to AR and VK and the Swiss National Science Foundation (31003A_160203) to AR. The providers of the grants had no role in the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

Declaration of conflicting interests

The authors declare no conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements

We sincerely thank the patient's family for co-operation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejmg.2016.12.004.

References

Abe, J., Izawa, K., Nishikomori, R., Awaya, T., Kawai, T., Yasumi, T., Hiragi, N., Hiragi, T., Ohshima, Y., Heike, T., 2013. Heterozygous TREX1 p.Asp18Asn mutation can cause variable neurological symptoms in a family with Aicardi-Goutieres syndrome/familial chilblain lupus. Rheumatol. Oxf. 52 (2), 406–408.

Alfaiz, A.A., Micale, L., Mandriani, B., Augello, B., Pellico, M.T., Chrast, J., Xenarios, I., Zelante, L., Merla, G., Reymond, A., 2014. TBC1D7 mutations are associated with intellectual disability, macrocrania, patellar dislocation, and celiac disease. Hum. Mutat. 35 (4), 447–451.

Alfaiz, A.A., Müller, V., Boutry-Kryza, N., Ville, D., Guex, N., de Bellescize, J., Rivier, C., Labalme, A., des Portes, V., Edery, P., Till, M., Xenarios, I., Sanlaville, D., Herrmann, J.M., Lesca, G., Reymond, A., 2016. West syndrome caused by homozygous variant in the evolutionary conserved gene encoding the mitochondrial elongation factor GUF1. Eur. J. Hum. Genet. 24 (7), 1001–1008.

Atianand, M.K., Fitzgerald, K.A., 2013. Molecular basis of DNA recognition in the immune system. J. Immunol. 190 (5), 1911–1918.

Beck-Engeser, G.B., Eilat, D., Wabl, M., 2011. An autoimmune disease prevented by anti-retroviral drugs. Retrovirology 8 (8), 91.

Crow, Y.J., 2015. Type I interferonopathies: mendelian type I interferon up-regulation. Curr. Opin. Immunol. 32, 7–12.

Crow, Y.J., Hayward, B.E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D.N., van Bokhoven, H., Brunner, H.G., Hamel, B.C., Corry, P.C., Cowan, F.M., Frints, S.G., Klepper, J., Livingston, J.H., Lynch, S.A., Massey, R.F., Meritet, J.F., Michaud, J.L., Ponsot, G., Voit, T., Lebon, P., Bonthron, D.T., Jackson, A.P., Barnes, D.E., Lindahl, T., 2006. Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutières syndrome at the AGS1 locus. Nat. Genet. 38 (8), 917–920.

Crow, Y.J., Chase, D.S., Lowenstein Schmidt, J., Szynkiewicz, M., Forte, G.M., Gornall, H.L., Oojageer, A., Anderson, B., Pizzino, A., Helman, G., Abdel-Hamid, M.S., Abdel-Salam, G.M., Ackroyd, S., Aeby, A., Agosta, G., Albin, C., Allon-Shalev, S., Arellano, M., Ariaudo, G., Aswani, V., Babul-Hirji, R., Baildam, E.M., Bahi-Buisson, N., Bailey, K.M., Barnerias, C., Barth, M., Battini, R., Beresford, M.W., Bernard, G., Bianchi, M., Billette de Villemeur, T., Blair, E.M., Bloom, M., Burlina, A.B., Carpanelli, M.L., Carvalho, D.R., Castro-Gago, M., Cavallini, A., Cereda, C., Chandler, K.E., Chitayat, D.A., Collins, A.E., Sierra Corcoles, C., Cordeiro, N.J., Crichiutti, G., Dabydeen, L., Dale, R.C., D'Arrigo, S., De Goede, C.G., De Laet, C., De Waele, L.M., Denzler, I., Desguerre, I., Devriendt, K., Di Rocco, M., Fahey, M.C., Fazzi, E., Ferrie, C.D., Figueiredo, A., Gener, B., Goizet, C., Gowrinathan, N.R., Gowrishankar, K., Hanrahan, D., Isidor, B., Kara, B., Khan, N., King, M.D., Kirk, E.P., Kumar, R., Lagae, L., Landrieu, P., Lauffer, H., Laugel, V., La Piana, R., Lim, M.J., Lin, J.P., Linnankivi, T., Mackay, M.T., Marom, D.R., Marques Lourenço, C., McKee, S.A., Moroni, I., Morton, J.E., Moutard, M.L., Murray, K., Nabbout, R., Nampoothiri, S., Nunez-Enamorado, N., Oades, P.J., Olivieri, I., Ostergaard, J.R., Pérez-Dueñas, B., Prendiville, J.S., Ramesh, V., Rasmussen, M., Régal, L., Ricci, F., Rio, M., Rodriguez, D., Roubertie, A., Salvatici, E., Segers, K.A., Sinha, G.P., Soler, D., Spiegel, R., Stödberg, T.I., Straussberg, R., Swoboda, K.J., Suri, M., Tacke, U., Tan, T.Y., te Water Naude, J., Wee Teik, K., Thomas, M.M., Till, M., Tonduti, D., Valente, E.M., Van Coster, R.N., van der Knaap, M.S., Vassallo, G., Vijzelaar, R., Vogt, J., Wallace, G.B., Wassmer, E., Webb, H.J., Whitehouse, W.P., Whitney, R.N., Zaki, M.S., Zuberi, S.M., Livingston, J.H., Rozenberg, F., Lebon, P., Vanderver, A., Orcesi, S., Rice, G.I., 2015. Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. Am. J. Med. Genet. A 167A (2), 296–312.

Dalakas, M.C., 2011. Review: an update on inflammatory and autoimmune myopathies. Neuropathol. Appl. Neurobiol. 37 (3), 226–242.

Ellyard, J.I., Jerjen, R., Martin, J.L., Lee, A.Y., Field, M.A., Jiang, S.H., Cappello, J., Naumann, S.K., Andrews, T.D., Scott, H.S., Casarotto, M.G., Goodnow, C.C., Chaitow, J., Pascual, V., Hertzog, P., Alexander, S.I., Cook, M.C., Vinuesa, C.G., 2014. Identification of a pathogenic variant in TREX1 in early-onset cerebral systemic lupus erythematosus by Whole-exome sequencing. Arthritis Rheumatol. 66 (12), 3382–3386.

Greenberg, S.A., 2010. Dermatomyositis and type 1 interferons. Curr. Rheumatol. Rep. 12 (3), 198–203.

Greenberg, S.A., Higgs, B.W., Morehouse, C., Walsh, R.J., Kong, S.W., Brohawn, P., Zhu, W., Amato, A., Salajegheh, M., White, B., Kiener, P.A., Jallal, B., Yao, Y., 2012. Relationship between disease disease activity and type 1 interferon- and other cytokine-inducible gene expression in blood in dermatomyositis and polymyositis. Genes Immun. 13 (3), 207–213.

Haaxma, C.A., Crow, Y.J., van Steensel, M.A., Lammens, M.M., Rice, G.I., Verbeek, M.M., Willemsen, M.A., 2010. A de novo p.Asp18Asn mutation in TREX1 in a patient with Aicardi-Goutières syndrome. Am. J. Med. Genet. A 152A (10), 2612–2617.

Hayashi, K., Sasai, M., Iwasaki, A., 2015. Toll-like receptor 9 trafficking and signaling for type I interferons requires PIKfyve activity. Int. Immunol. 27 (9), 435–445.

Higgs, B.W., Liu, Z., White, B., Zhu, W., White, W.I., Morehouse, C., Brohawn, P., Kiener, P.A., Richman, L., Fiorentino, D., Greenberg, S.A., Jallal, B., Yao, Y., 2011. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. Ann. Rheum. Dis. 70 (11), 2029–2036.

Ivashkiv, L.B., Donlin, L.T., 2014. Regulation of type I interferon responses. Nat. Rev. Immunol. 14 (1), 36–49.

Jakati, S., Rajasekhar, L., Uppin, M., Challa, S., 2015. SLE myopathy: a clinicopathological study. Int. J. Rheum. Dis. 18 (8), 886–891.

Kalliolias, G.D., Kirou, K.A., 2012. Type I interferons as biomarkers in autoimmune diseases. Biomark. Med. 6 (2), 137–140.

Leshinsky-Silver, E., Malinger, G., Ben-Sira, L., Kidron, D., Cohen, S., Inbar, S., Bezaleli, T., Levine, A., Vinkler, C., Lev, D., Lerman-Sagie, T., 2011. A large homozygous deletion in the SAMHD1 gene causes atypical Aicardi-Goutiéres syndrome associated with mtDNA deletions. Eur. J. Hum. Genet. 19 (3), 287–292.

Morita, M., Stamp, G., Robins, P., Dulic, A., Rosewell, I., Hrivnak, G., Daly, G., Lindahl, T., Barnes, D.E., 2004. Gene-targeted mice lacking the Trex1 (DNase III) 3'–>5' DNA exonuclease develop inflammatory myocarditis. Mol. Cell Biol. 24 (15), 6719–6727.

Oka, T., Hikoso, S., Yamaguchi, O., Taneike, M., Takeda, T., Tamai, T., Oyabu, J., Murakawa, T., Nakayama, H., Nishida, K., Akira, S., Yamamoto, A., Komuro, I., Otsu, K., 2012. Mitochondrial DNA that escapes from autophagy causes inflammation and heart failure, 10 Nature 485 (7397), 251–255.

Ramantani, G., Kohlhase, J., Hertzberg, C., Innes, A.M., Engel, K., Hunger, S., Borozdin, W., Mah, J.K., Ungerath, K., Walkenhorst, H., Richardt, H.H., Buckard, J., Bevot, A., Siegel, C., von Stülpnagel, C., Ikonomidou, C., Thomas, K., Proud, V., Niemann, F., Wieczorek, D., Häusler, M., Niggemann, P., Baltaci, V., Conrad, K., Lebon, P., Lee-Kirsch, M.A., 2010. Expanding the phenotypic spectrum of lupus erythematosus in Aicardi-Goutières syndrome. Arthritis Rheum. 62 (5), 1469–1477.

Rice, G.I., Forte, G.M., Szynkiewicz, M., Chase, D.S., Aeby, A., Abdel-Hamid, M.S., Ackroyd, S., Allcock, R., Bailey, K.M., Balottin, U., Barnerias, C., Bernard, G., Bodemer, C., Botella, M.P., Cereda, C., Chandler, K.E., Dabydeen, L., Dale, R.C., De Laet, C., De Goede, C.G., Del Toro, M., Effat, L., Enamorado, N.N., Fazzi, E.,

Gener, B., Haldre, M., Lin, J.P., Livingston, J.H., Lourenco, C.M., Marques Jr., W., Oades, P., Peterson, P., Rasmussen, M., Roubertie, A., Schmidt, J.L., Shalev, S.A., Simon, R., Spiegel, R., Swoboda, K.J., Temtamy, S.A., Vassallo, G., Vilain, C.N., Vogt, J., Wermenbol, V., Whitehouse, W.P., Soler, D., Olivieri, I., Orcesi, S., Aglan, M.S., Zaki, M.S., Abdel-Salam, G.M., Vanderver, A., Kisand, K., Rozenberg, F., Lebon, P., Crow, Y.J., 2013. Assessment of interferon-related biomarkers in Aicardi-Goutières syndrome associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR: a case-control study. Lancet Neurol. 12 (12), 1159–1169.

Rice, G.I., Rodero, M.P., Crow, Y.J., 2015. Human disease phenotypes associated with mutations in TREX1. J. Clin. Immunol. 35 (3), 235–343.

Stetson, D.B., Ko, J.S., Heidmann, T., Medzhitov, R., 2008. Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell 134 (4), 587–598.

Yang, Y.G., Lindahl, T., Barnes, D.E., 2007 Nov 30. Trex1 exonuclease degrades ssDNA to prevent chronic checkpoint activation and autoimmune disease. Cell 131 (5), 873–886.

Discussion and perspectives

Before NGS, diagnostic yield was hampered by the technologies available, as well as the prior knowledge of some genetic causes of ID. Indeed, fragile-X testing, aCGH and targeted sequencing of candidate genes were limited strategies, especially to find new ID genes. WES overcame these limitations and greatly improved the diagnosis rate, restricting the previous tests as first-tier genetic tests, followed by WES when inconclusive. As prices decreased, WES replaced fragile-X testing, aCGH and targeted sequencing in some laboratories, as they give the same information, and much more. Researchers provided evidences that WES is more efficient than panel sequencing and WGS as it covers all the genes tested by the panel but is not as expensive as WGS. However, some argued in return that, with today's WGS cost, this technology should be the first-approach as the coverage is greater than WES for all coding parts of the human genome, and it may capture interpretable variants in introns, UTR and/or promoter regions of known ID genes.

Whereas WES has been successfully used to identify new ID genes and provide genetic diagnosis in 25-40% of patients with ID and neurodevelopmental disorders, a high coverage is needed to limit false positive and false negative results, increasing the cost. Also, WES is much more efficient in family- or cohort-based analysis than in singleton study. Sequencing of families allows us to filter according to the segregation of the variants with the condition. Thus, families with patient's siblings, healthy or affected, are usually easier to study. A cohort approach is more limited in the case of neurodevelopmental disorders, as the heterogeneity of phenotypes is broad and can be reflected by the heterogeneity of the genetic causes. In our projects, patients have various phenotypes, too divergent to be considered as a cohort to search for shared variants as the genetic causes are likely to be as different. We used a family-based approach, with patients and their parents sequenced, and encouraged collection of samples from siblings, although it increases the cost of the procedure. We had only 3 trio cases (Italian_6, Italian_8, Italian_12) and they are still undiagnosed after WES as too many VUS remain after filtering. Continuous implementation of databases with acquired knowledge on genes function, residues conservation and variants frequency should help us reduce those lists of GUS and VUS. Regular exome results re-assessment and variants annotation is needed to consider new findings. For example, at the beginning of this project we were collecting variants frequency from the ExAC database. In 2017, the GnomAD implementation increased the population data from 60,706 exomes to 125,748 exomes and 15,708 genomes. Frequencies estimated have consequently changed, and so has our interpretation of some variants.

Validation of GUS and VUS is eased by the increasing number of patients that are sequenced as well as the presence of new platforms of data exchange like GeneMatcher that allow laboratories to share clinical data and assemble groups of patients with variants within the same genes and similar phenotypes. This helped us in particular to collect evidences for new ID-genes identification (*AFF3, MAST1, KIAA1109*). However, although suspected genes remain available on the platform and are queried following new entries, the overlap of clinical data may be limited to what was investigate in a particular laboratory or may not completely correspond to one's features. Consequently, a conclusion on the pathogenicity of a variant may not be reached until further functional analyses are performed.

Performing such analyses for all the GUS would have represented a tremendous work for only one research laboratory. Collaborations allowed us to split the tasks and have 3 new ID-genes identified, as several laboratories took the lead of different experiments. But research studies are usually long and fastidious, not compatible with the need of a fast diagnosis for the families who may wait for personalized care. An ethical issue is raised with the disclosing of uncertain findings (GUS and VUS) for which analyses are warranted to conclude on their pathogenicity as well as findings unrelated to the primary diagnosis purpose, so called "secondary findings". It is nowadays up to the clinicians and the ethical board of local hospitals to decide on whether to communicate to the families the inconclusive results and the secondary findings, or not. Noteworthy, to promote standardized reporting, the American College of Medical Genetics and Genomics (ACMG) published a list of 59 medically actionable genes recommended for report of secondary findings. In my humble opinion, results should always be communicated to the families, as they willingly took part in these studies and are waiting for feedback. But caution is needed when explaining the sequencing outcomes as they consist of suppositions, added evidences, but not definitive proves. That is why I personally think that clinicians and researchers should collaborate on the material that should be provided to the patients and their family, explaining the process of research, the techniques used, the time needed to perform the validation steps and how to interpret the results. This work of transmission of our knowledge is a duty, as those families are the tax payers who found our research projects and have the right to know what is done with their money, and their samples. Moreover, involving them in the process by giving them a clear idea and follow-up on our experiments and results may convince them of the importance of clinical re-examinations and sample collections. This could help us collect pictures or biological samples that are sometimes hard to get, due to the uncertainty of use of these data. Giving them all these details would, to me, lead to a true enlightened consent for experiments, data collections and data sharing, crucial in our field. By helping them understand our work, they will help us do it.

Besides, as research work can take a long time between the identification of the pathogenic variants and the collection of functional evidences, I think it is necessary to not wait for publication to communicate the findings to the families although we have to stay cautious. Indeed, sequencing results can be given by genetic councilors and followed by clinical consulting which can suggest treatments in the light of the collected data. For instance, the *AFF3* project took us 3 years to collect clinical data of 10 patients, as well as performing functional analyses. In the meantime, the observation of older probands 3 and 4 and reports on the in/efficacy of treatments enabled the clinicians of other probands to prioritize drugs. Also, the observation of specific features that may not yet be present in some of the proband raised the attention of the clinicians who adapted their follow-up and warned the parents about the possibility of development of these difficulties. In addition to this, we were able to put in contact families of the *AFF3* project at their request. Some created a Facebook support group where they shared their daily life and difficulties, treatments and medical interventions. Their feedback is precious for us, to have up-to-date clinical status without waiting for consultations' reports, but also for the other families to share their experience, their struggle and their hope. To my mind, at any step of our research, the dialog between researchers, clinicians and families should be encouraged, being beneficial and rewarding for each side.

Out of the 7 families we elucidated, 4 had variants within genes already identified in neurodevelopmental disorders (*PIK3R4, TREX1, POMK, TBCE*), as interpretation of the variant may be facilitated. However, restricting WES to the research of variants within known ID-genes would limit its power. Indications for new ID-genes identification may come from the recurrence of pathways associated with syndromic ID and neurodevelopmental defects. Indeed, a lot of the ID genes are involved in metabolism or gene expression regulation. It is admitted that a fine-tuning of gene expression during neurodevelopment is crucial for complete and proper functioning, but it is intriguing that variants in such ubiquitous proteins, targeting numerous other genes, lead to specific defects in brain development and synaptic functions. This might come from the fact that neural cells are more sensitive the gene-dosage than other cell types, as suggested by the prominence of haploinsufficiency mechanisms in neurodevelopmental disorders associated with chromatin-related genes. As functional validation, RNA sequencing can be a valuable complement of WES to understand possible dysregulation of genes. However, it is usually limited to peripheral blood RNA quantification and thus may not reflect what happened in the affected tissues at a certain time point of their development, those being hardly accessible.

RNA sequencing and WGS would allow us to identify mutations in introns or promoters, affecting the expression (splicing, expression level) of corresponding mRNA. WGS allows detection of all variants, independently of their coding or non-coding function. Besides, its

uniform coverage improves the detection of balanced and unbalanced structural variants. However, besides the tremendous amount of data produced by WGS that need to be stored, its limitation resides also in the annotation of all these variants detected in non-coding regions. Indeed, there is a lack of interpretation tools that could help annotate these non-coding variants, and thus to prioritize them. Now that WGS is more and more used, international consortia start to accumulate data, for instance on the frequency of these variants. Also, projects like ENCODE strive at annotating the human genome, including regulatory elements and topologically associated domains (TADs) that are known to be important for the action of regulatory elements over distal elements they contact through 3D folding. As our understanding of the non-coding regions increases, we also understand their importance in gene regulation and their possible implication in genetic disorders.

At the release of this manuscript, 16 cases are yet unsolved. For 3 families we have potential candidates and proves have to be made that these GUS can be involved in the phenotype apparition. For instance, in the family Lithuanian_7, 2 boys were suffering of photophobia, nystagmus, and ID. WES helped us to identify a variant in *CNGB3*, a gene already associated with achromatopsia, but never with ID. After a full review of the literature and contacting the original submitters of the *CNGB3/*Achromatopsia study, we could not determine if other achromatopsia cases suffered from neurodevelopmental disorders. Hence, we could not confirm if the observed *CNGB3* variant was the cause of both achromatopsia and ID in our patients, or if the ID phenotype originated elsewhere. For those GUS, only description of other patients and functional analyses will allow us to give a genetic diagnosis. We are currently starting a collaboration for the deeper study of *SATB1* that we suspect at the origin of the syndromic ID in family Lithuanian_8. We collected clinical data of 7 patients with variants in *SATB1* and another laboratory is investigating the molecular impact of the identified amino-acid changes.

For the remaining VUS, their impact is unclear as they are not necessarily considered as pathogenic by all predictors and are not fully conserved, and the corresponding genes are not associated with functions or disorders reminiscent of the observed phenotypes. This absence of "best candidate" incite cautiousness and the diagnosis rely on functional analyses and collaborations. Besides, our approach may be efficient to identify monogenic causes of syndromic ID but we cannot rule out a more complex etiology for the unsolved cases. Multigenic forms of neurodevelopmental disorders cannot easily be identified given the current method used and filtering strategy. Only studies of large cohort and genetic interaction studies may help us delineate the contribution of variants in more complex scenario such as combination of multiple genetic variants. Besides, we hypothesized a full penetrance transmission of the disorders and filtered out variants present in unaffected relatives of the patients whereas there could be

pathogenic variants in one individual, responsible for no or milder phenotype in other, depending on the genetic background. In the Lithuanian_8 family, the original clinical evaluations reported 2 affected brothers. Later reevaluations informed us that the younger brother was borderline for ID but had none of the other co-morbid traits of his brother. We changed the filtering strategy according to this new information, but clinical data suggest that a more complex transmission is at play.

To conclude, for the past decade NGS technologies, and in particular WES, allowed us to identify many new genetic causes of ID. Although this work is far from being completed, as some cases remain unsolved and new ones come regularly, we proved that our family-based exome sequencing approach is a successful method to establish a genetic diagnosis and that world-wide collaborations are required to better delineate syndromic ID, gather clinical data and perform the warranted functional experiments. As our understanding of the human genome structure and function is expanding, we can guess that some of the remaining unsolved cases will obtain a molecular diagnosis in the near future. However, deciphering the etiology of more complex cases, with incomplete penetrance or multifactorial origin, will require the development of new methods.

References

1. Sherr, E.H., Michelson, D.J., Shevell, M.I., Moeschler, J.B., Gropman, A.L., and Ashwal, S. (2013). Neurodevelopmental disorders and genetic testing: Current approaches and future advances. Annals of Neurology 74, 164-170.
2. American Psychiatric, A., American Psychiatric, A., and Force, D.S.M.T. (2017). Diagnostic and statistical manual of mental disorders : DSM-5.(Arlington, VA: American Psychiatric Association).
3. Harris, J.C. (2006). Intellectual disability: Understanding its development, causes, classification, evaluation, and treatment.(New York: Oxford University Press).
4. Maulik, P.K., Mascarenhas, M.N., Mathers, C.D., Dua, T., and Saxena, S. (2011). Prevalence of intellectual disability: a meta-analysis of population-based studies. Res Dev Disabil 32, 419-436.
5. Centers for Disease Control and Prevention. (2004). Economic costs associated with mental retardation, cerebral palsy, hearing loss, and vision impairment-United States, 2003. MMWR Morb Mortal Wkly Rep 53, 57-59.
6. Polder, J.J., Meerding, W.J., Bonneux, L., and van der Maas, P.J. (2002). Healthcare costs of intellectual disability in the Netherlands: a cost-of-illness perspective. J Intellect Disabil Res 46, 168-178.
7. Guilmatre, A., Dubourg, C., Mosca, A.L., Legallic, S., Goldenberg, A., Drouin-Garraud, V., Layet, V., Rosier, A., Briault, S., Bonnet-Brilhault, F., et al. (2009). Recurrent rearrangements in synaptic and neurodevelopmental genes and shared biologic pathways in schizophrenia, autism, and mental retardation. Arch Gen Psychiatry 66, 947-956.
8. Petterson, B., Bourke, J., Leonard, H., Jacoby, P., and Bower, C. (2007). Co-occurrence of birth defects and intellectual disability. Paediatric and Perinatal Epidemiology 21, 65-75.
9. Kochinke, K., Zweier, C., Nijhof, B., Fenckova, M., Cizek, P., Honti, F., Keerthikumar, S., Oortveld, M.A., Kleefstra, T., Kramer, J.M., et al. (2016). Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am J Hum Genet 98, 149-164.
10. Gurovich, Y., Hanani, Y., Bar, O., Nadav, G., Fleischer, N., Gelbman, D., Basel-Salmon, L., Krawitz, P.M., Kamphausen, S.B., Zenker, M., et al. (2019). Identifying facial phenotypes of genetic disorders using deep learning. Nature Medicine 25, 60-64.
11. Kohler, S., Vasilevsky, N.A., Engelstad, M., Foster, E., McMurry, J., Ayme, S., Baynam, G., Bello, S.M., Boerkoel, C.F., Boycott, K.M., et al. (2017). The Human Phenotype Ontology in 2017. Nucleic Acids Res 45, D865-D876.
12. Tomar, S., Sethi, R., and Lai, P.S. (2019). Specific phenotype semantics facilitate gene prioritization in clinical exome sequencing. European Journal of Human Genetics.
13. Chiurazzi, P., and Pirozzi, F. (2016). Advances in understanding - genetic basis of intellectual disability. F1000Res 5.
14. Lejeune, J., Turpin, R., and Gautier, M. (1959). [Chromosomic diagnosis of mongolism]. Arch Fr Pediatr 16, 962-963.
15. Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., Zenker, M., Huffmeier, U., Thiel, C., Ruschendorf, F., et al. (2006). Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. American Journal of Medical Genetics Part A 140a, 2063-2074.
16. Michelson, D.J., Shevell, M.I., Sherr, E.H., Moeschler, J.B., Gropman, A.L., and Ashwal, S. (2011). Evidence report: Genetic and metabolic testing on children with global developmental delay: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Practice Committee of the Child Neurology Society. Neurology 77, 1629-1635.
17. Regan, R., and Willatt, L. (2010). Mental Retardation: Definition, Classification and Etiology. Monogr Hum Genet 18, 16-30.

18. Miller, D.T., Adam, M.P., Aradhya, S., Biesecker, L.G., Brothman, A.R., Carter, N.P., Church, D.M., Crolla, J.A., Eichler, E.E., Epstein, C.J., et al. (2010). Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. American Journal of Human Genetics 86, 749-764.
19. de Vries, B.B., Pfundt, R., Leisink, M., Koolen, D.A., Vissers, L.E., Janssen, I.M., Reijmersdal, S., Nillesen, W.M., Huys, E.H., Leeuw, N., et al. (2005). Diagnostic genome profiling in mental retardation. Am J Hum Genet 77, 606-616.
20. Bernardini, L., Alesi, V., Loddo, S., Novelli, A., Bottillo, I., Battaglia, A., Digilio, M.C., Zampino, G., Ertel, A., Fortina, P., et al. (2010). High-resolution SNP arrays in mental retardation diagnostics: how much do we gain? Eur J Hum Genet 18, 178-185.
21. Lupski, J.R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 14, 417-422.
22. Jacquemont, S., Reymond, A., Zufferey, F., Harewood, L., Walters, R.G., Kutalik, Z., Martinet, D., Shen, Y., Valsesia, A., Beckmann, N.D., et al. (2011). Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature 478, 97-102.
23. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006). Global variation in copy number in the human genome. Nature 444, 444-454.
24. Pieretti, M., Zhang, F.P., Fu, Y.H., Warren, S.T., Oostra, B.A., Caskey, C.T., and Nelson, D.L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. Cell 66, 817-822.
25. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P., et al. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 65, 905-914.
26. Lubs, H.A., Stevenson, R.E., and Schwartz, C.E. (2012). Fragile X and X-linked intellectual disability: four decades of discovery. Am J Hum Genet 90, 579-590.
27. Musante, L., and Ropers, H.H. (2014). Genetics of recessive cognitive disorders. Trends Genet 30, 32-39.
28. Najmabadi, H., Hu, H., Garshasbi, M., Zemojtel, T., Abedini, S.S., Chen, W., Hosseini, M., Behjati, F., Haas, S., Jamali, P., et al. (2011). Deep sequencing reveals 50 novel genes for recessive cognitive disorders. Nature 478, 57-63.
29. Vissers, L.E., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., van Lier, B., Arts, P., Wieskamp, N., del Rosario, M., et al. (2010). A de novo paradigm for mental retardation. Nat Genet 42, 1109-1112.
30. Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., et al. (2014). Genome sequencing identifies major causes of severe intellectual disability. Nature 511, 344-347.
31. Deciphering Developmental Disorders, S. (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature 519, 223-228.
32. Vissers, L.E., Gilissen, C., and Veltman, J.A. (2016). Genetic studies in intellectual disability and related disorders. Nat Rev Genet 17, 9-18.
33. Cooper, D.N., Chen, J.M., Ball, E.V., Howells, K., Mort, M., Phillips, A.D., Chuzhanova, N., Krawczak, M., Kehrer-Sawatzki, H., and Stenson, P.D. (2010). Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. Hum Mutat 31, 631-655.
34. Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., et al. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 42, 790-793.
35. Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., et al. (2010). Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42, 30-35.
36. Veltman, J.A., and Brunner, H.G. (2012). De novo mutations in human genetic disease. Nat Rev Genet 13, 565-575.

37. O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485, 246-250.
38. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.
39. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv, 531210.
40. Kumar, P., Henikoff, S., and Ng, P.C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4, 1073-1081.
41. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-249.
42. Choi, Y., and Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745-2747.
43. Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nature Methods 11, 361.
44. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum Mutat 36, 928-930.
45. Aronesty, E. (2011). ea-utils:"Command-line tools for processing biological sequencing data". In. (https://github.com/ExpressionAnalysis/ea-utils)
46. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.
47. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303.
48. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43, 491-498.
49. Delafontaine J., M.A., Liechti R., Kuznetsov D., Xenarios I., Pradervand S. (2016). Varapp: A reactive web-application for variants filtering. bioRxiv preprint.
50. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res 47, D886-D894.
51. Cooper, G.M., Stone, E.A., Asimenos, G., Program, N.C.S., Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 15, 901-913.
52. Abrahams, B.S., Arking, D.E., Campbell, D.B., Mefford, H.C., Morrow, E.M., Weiss, L.A., Menashe, I., Wadkins, T., Banerjee-Basu, S., and Packer, A. (2013). SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism 4, 36.
53. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25-29.
54. UniProt, C. (2019). UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47, D506-D515.
55. Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., et al. (2015). Proteomics. Tissue-based map of the human proteome. Science 347, 1260419.
56. Dickinson, M.E., Flenniken, A.M., Ji, X., Teboul, L., Wong, M.D., White, J.K., Meehan, T.F., Weninger, W.J., Westerberg, H., Adissu, H., et al. (2016). High-throughput discovery of novel developmental phenotypes. Nature 537, 508-514.

57. Howe, D.G., Bradford, Y.M., Conlin, T., Eagle, A.E., Fashena, D., Frazer, K., Knight, J., Mani, P., Martin, R., Moxon, S.A., et al. (2013). ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. Nucleic Acids Res 41, D854-860.
58. Gieldon, L., Mackenroth, L., Kahlert, A.K., Lemke, J.R., Porrmann, J., Schallner, J., von der Hagen, M., Markus, S., Weidensee, S., Novotna, B., et al. (2018). Diagnostic value of partial exome sequencing in developmental disorders. PLoS One 13, e0201041.
59. Gueneau, L., Fish, R.J., Shamseldin, H.E., Voisin, N., Tran Mau-Them, F., Preiksaitiene, E., Monroe, G.R., Lai, A., Putoux, A., Allias, F., et al. (2018). KIAA1109 Variants Are Associated with a Severe Disorder of Brain Development and Arthrogryposis. Am J Hum Genet 102, 116-132.
60. Tripathy, R., Leca, I., van Dijk, T., Weiss, J., van Bon, B.W., Sergaki, M.C., Gstrein, T., Breuss, M., Tian, G., Bahi-Buisson, N., et al. (2018). Mutations in MAST1 Cause Mega-Corpus-Callosum Syndrome with Cerebellar Hypoplasia and Cortical Malformations. Neuron 100, 1354-1368 e1355.
61. Sferra, A., Baillat, G., Rizza, T., Barresi, S., Flex, E., Tasca, G., D'Amico, A., Bellacchio, E., Ciolfi, A., Caputo, V., et al. (2016). TBCE Mutations Cause Early-Onset Progressive Encephalopathy with Distal Spinal Muscular Atrophy. Am J Hum Genet 99, 974-983.
62. Terrone, G., Voisin, N., Abdullah Alfaiz, A., Cappuccio, G., Vitiello, G., Guex, N., D'Amico, A., James Barkovich, A., Brunetti-Pierri, N., Del Giudice, E., et al. (2016). De novo PIK3R2 variant causes polymicrogyria, corpus callosum hyperplasia and focal cortical dysplasia. Eur J Hum Genet 24, 1359-1362.
63. Tumiene, B., Voisin, N., Preiksaitiene, E., Petroska, D., Grikiniene, J., Samaitiene, R., Utkus, A., Reymond, A., and Kucinskas, V. (2017). Inflammatory myopathy in a patient with Aicardi-Goutieres syndrome. Eur J Med Genet 60, 154-158.
64. Bitoun, E., and Davies, K.E. (2005). The robotic mouse: unravelling the function of AF4 in the cerebellum. Cerebellum 4, 250-260.
65. House, C.M., Frew, I.J., Huang, H.L., Wiche, G., Traficante, N., Nice, E., Catimel, B., and Bowtell, D.D. (2003). A binding motif for Siah ubiquitin ligase. Proc Natl Acad Sci U S A 100, 3101-3106.
66. Izumi, K., Nakato, R., Zhang, Z., Edmondson, A.C., Noon, S., Dulik, M.C., Rajagopalan, R., Venditti, C.P., Gripp, K., Samanich, J., et al. (2015). Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. Nat Genet 47, 338-344.
67. Riviere, J.B., Mirzaa, G.M., O'Roak, B.J., Beddaoui, M., Alcantara, D., Conway, R.L., St-Onge, J., Schwartzentruber, J.A., Gripp, K.W., Nikkel, S.M., et al. (2012). De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. Nat Genet 44, 934-940.
68. Mirzaa, G.M., Conti, V., Timms, A.E., Smyser, C.D., Ahmed, S., Carter, M., Barnett, S., Hufnagel, R.B., Goldstein, A., Narumi-Kishimoto, Y., et al. (2015). Characterisation of mutations of the phosphoinositide-3-kinase regulatory subunit, PIK3R2, in perisylvian polymicrogyria: a next-generation sequencing study. Lancet Neurol 14, 1182-1195.
69. du Souich, C., Chou, A., Yin, J., Oh, T., Nelson, T.N., Hurlburt, J., Arbour, L., Friedlander, R., McGillivray, B.C., Tyshchenko, N., et al. (2009). Characterization of a new X-linked mental retardation syndrome with microcephaly, cortical malformation, and thin habitus. Am J Med Genet A 149A, 2469-2478.
70. Tarpey, P.S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., O'Meara, S., Latimer, C., Dicks, E., Menzies, A., et al. (2009). A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. Nat Genet 41, 535-543.
71. McLarren, K.W., Severson, T.M., du Souich, C., Stockton, D.W., Kratz, L.E., Cunningham, D., Hendson, G., Morin, R.D., Wu, D., Paul, J.E., et al. (2010). Hypomorphic temperature-sensitive alleles of NSDHL cause CK syndrome. Am J Hum Genet 87, 905-914.
72. Crow, Y.J., Chase, D.S., Lowenstein Schmidt, J., Szynkiewicz, M., Forte, G.M., Gornall, H.L., Oojageer, A., Anderson, B., Pizzino, A., Helman, G., et al. (2015). Characterization of human disease phenotypes associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1. Am J Med Genet A 167A, 296-312.

Web Resources

- AAIDD: www.aaidd.org
- "AFF3 Genetic Mutation" Facebook group: https://www.facebook.com/groups/1316717511718324/
- DECIPHER/DDD Study: https://decipher.sanger.ac.uk/
- Erik Aronesty (2011). *ea-utils* : "Command-line tools for processing biological sequencing data"; https://github.com/ExpressionAnalysis/ea-utils
- ExAC: http://exac.broadinstitute.org/
- Face2Gene: https://www.face2gene.com/
- GeneMatcher: https://genematcher.org/
- GeneOntology: http://geneontology.org/
- GnomAD: https://gnomad.broadinstitute.org/
- GTEx: https://www.gtexportal.org/home/
- HGVS: http://varnomen.hgvs.org/
- HPO: https://hpo.jax.org/app/
- IMPC: http://www.mousephenotype.org/
- OMIM: https://www.omim.org/
- MutationTaster2: http://www.mutationtaster.org/
- Picards: http://broadinstitute.github.io/picard
- PolyPhen-2: http://genetics.bwh.harvard.edu/pph2/index.shtml
- PubMed: https://www.ncbi.nlm.nih.gov/pubmed/
- The Human Protein Atlas: https://www.proteinatlas.org/
- PROVEAN: http://provean.jcvi.org/index.php
- SFARI Gene: https://gene.sfari.org/
- SIFT: http://sift.jcvi.org/
- SysID: https://sysid.cmbi.umcn.nl/
- UniProt: https://www.uniprot.org/
- Varapp: https://github.com/varapp; https://varapp-demo.vital-it.ch
- WHO: https://www.who.int/home
- ZFIN: https://zfin.org/

Appendices

Résumé de la thèse en français

La déficience intellectuelle (DI) est un trouble neuro-développemental qui affecte 1-3% de la population, représentant un enjeu de santé publique majeur. La DI est extrêmement hétérogène, de par sa sévérité, la possibilité de symptômes associés, ainsi que son étiologie. Alors que plusieurs centaines de gènes ont été associés avec la DI, les causes génétiques de nombreux cas restent à déterminer. Nous avons utilisé le séquençage d'exome afin de découvrir les causes génétiques de DI syndromiques non-résolues pour 23 familles.

Pendant mon doctorat, j'ai travaillé en collaboration avec des cliniciens de Lituanie et d'Italie qui ont collecté les données cliniques de familles pour lesquelles les tests CGH-array et de l'X-fragile étaient négatifs. Bien qu'une DI ait été diagnostiquée chez tous les patients que j'ai étudiés, la présence de symptômes additionnels a fait de chaque famille un cas particulier que je devais analyser individuellement. Les variants candidats identifiés par exome ont été validés par des analyses fonctionnelles ainsi que par la collecte de données cliniques de cas similaires via des collaborations internationales.

Mon travail a mené à la découverte d'une nouvelle DI syndromique, caractérisée par une dysplasie mésomélique, des reins en fers à cheval et une épilepsie encéphalopathique développementale, et la description du mécanisme moléculaire sous-jacent. Des variants faux-sens *de novo* dans le motif degron de *AFF3* sont responsable d'une accumulation de la protéine correspondante, agissant comme un gain-de-fonction. Cette hypothèse est supportée par l'analyse de modèles moléculaires et animaux. De plus, j'ai identifié un variant *de novo* jusqu'alors non reporté dans un gène de DI connu, *PIK3R2,* chez un garçon avec une encéphalopathie sévère, expliquant le mécanisme moléculaire via modélisation 3D. J'ai aussi étendu la description phénotypique du syndrome Aicardi-Goutières en décrivant le premier cas de myopathie inflammatoire associée à un variant faux-sens homozygote dans le gène *TREX1*. Par ailleurs, j'ai participé à la description de deux nouveaux syndromes associés à des variants dans les gènes *KIAA1109* et *MAST1*. Alors que le séquençage d'exome des autres familles a permis d'identifier des variants potentiellement pathogènes ou des listes de variants de significations incertaines, il est impératif de collecter des données cliniques de cas similaires ainsi que des preuves fonctionnelles.

Mon travail nous a permis de donner un diagnostic génétique à certains patients et à leurs familles, fournissant des pistes d'amélioration de leurs prises en charge et de leurs suivis. Finalement, une meilleure définition des mécanismes pathologiques impliqués dans les troubles neuro-développementaux tels que la déficience intellectuelle aidera à identifier des nouvelles cibles thérapeutiques.

Résumé grand public en français

IDENTIFICATION DE NOUVELLES CAUSES GENETIQUES AUX FORMES SYNDROMIQUES DE DEFICIENCES INTELLECTUELLES

Par Norine VOISIN, doctorante dans le laboratoire du Professeur Alexandre Reymond,
Centre Intégratif de Génomique de l'Université de Lausanne

La déficience intellectuelle (DI), autrefois appelée retard mental, se manifeste durant l'enfance (avant 18 ans) par une limitation du développement psychomoteur et des facultés d'interactions sociales et d'adaptation. Elle se traduit lors de tests spécifiques, par un quotient intellectuel (QI) inférieur à 70. Le degré de sévérité de la DI est très variable, allant de difficultés légères (la majorité des cas) à profondes. Aussi, les personnes atteintes de DI peuvent présenter d'autres symptômes tels que des troubles psychiatriques (hyperactivité, autisme, schizophrénie…) ou encore des pathologies dites somatique, qui touchent certains organes (malformations cérébrales, osseuses, cardiaques…). On parle alors de DI syndromique. La grande hétérogénéité des symptômes s'accompagne d'une grande variété des causes possibles. En effet, des facteurs extérieurs intervenant durant le développement de l'enfant tels que l'abus d'alcool de la femme enceinte ou la malnutrition infantile influencent le développement du système nerveux et ainsi les capacités intellectuelles. De plus, de nombreuses causes génétiques ont été identifiées pour les DI telles que des anomalies du nombre de chromosomes (ex : la trisomie 21) ou encore des mutations de certains gènes (ex : syndrome de l'X fragile due à une mutation du gène *FMR1*).

Lors de consultations de familles avec une ou plusieurs personnes atteintes de DI, les cliniciens tentent d'établir un diagnostic sur la base des symptômes présentés et de la connaissance approfondie des nombreux syndromes déjà identifiés. Quand le diagnostic clinique échoue à déterminer une cause claire, les cliniciens font appel à des généticiens qui étudient alors l'ADN des patients et de leurs familles.

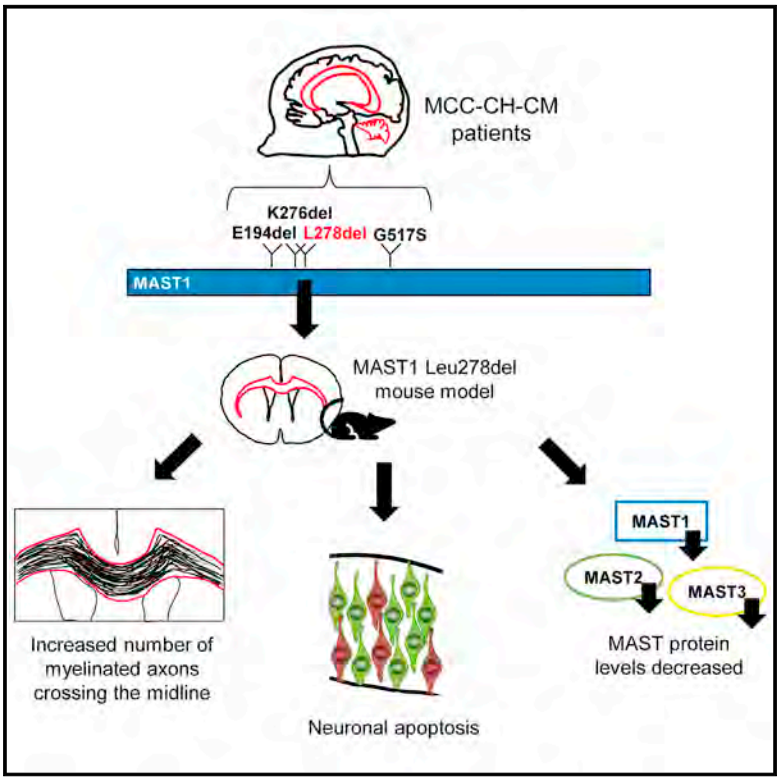
Durant ma thèse, j'ai travaillé en étroite collaboration avec des cliniciens de Lituanie et d'Italie qui m'ont transmis les données cliniques ainsi que les échantillons ADN de 23 familles. Mes travaux ont permis de donner un diagnostic à certaines de ces familles, de les informer des risques de transmission héréditaires ainsi que d'adapter leurs suivis médicaux. Nous avons aussi implémenté la liste de gènes de DI répertoriés et étendu la description de syndromes déjà identifiés, améliorant les capacités de diagnostics de DI. Mon étude a par ailleurs permis d'offrir à une famille lituanienne la possibilité d'avoir recours à une fécondation in vitro (FIV) avec diagnostic génétique pré-implantation (DPI) afin d'avoir un second enfant sain.

Article

Neuron

Mutations in *MAST1* Cause Mega-Corpus-Callosum Syndrome with Cerebellar Hypoplasia and Cortical Malformations

Graphical Abstract



Highlights

- Mast1 associates with the microtubule cytoskeleton in a MAP-dependent manner
- Mast1 is expressed in postmitotic neurons but not neuronal progenitors
- *Mast1* Leu278del mice have an enlarged corpus callosum and smaller cerebellum
- More axons cross the midline in *Mast1* Leu278del mice despite cortical apoptosis

Authors

Ratna Tripathy, Ines Leca, Tessa van Dijk, ..., Jamel Chelly, Nicholas J. Cowan, David Anthony Keays

Correspondence

keays@imp.ac.at

In Brief

Tripathy et al. show that mutations in *MAST1*, a microtubule-associated protein, cause a syndrome characterized by an enlarged corpus callosum. They attribute this thickening to an increase in the number of contralateral projections despite extensive cortical apoptosis.

Tripathy et al., 2018, Neuron *100*, 1354–1368
December 19, 2018
https://doi.org/10.1016/j.neuron.2018.10.044

CellPress

CellPress

Neuron
Article

Mutations in *MAST1* Cause Mega-Corpus-Callosum Syndrome with Cerebellar Hypoplasia and Cortical Malformations

Ratna Tripathy,[1] Ines Leca,[1] Tessa van Dijk,[2] Janneke Weiss,[3] Bregje W. van Bon,[4] Maria Christina Sergaki,[1] Thomas Gstrein,[1] Martin Breuss,[5] Guoling Tian,[6] Nadia Bahi-Buisson,[7] Alexander R. Paciorkowski,[8] Alistair T. Pagnamenta,[9] Andrea Wenninger-Weinzierl,[1] Maria Fernanda Martinez-Reza,[1] Lukas Landler,[1] Stefano Lise,[9] Jenny C. Taylor,[9] Gaetano Terrone,[10] Giuseppina Vitiello,[10] Ennio Del Giudice,[10] Nicola Brunetti-Pierri,[10,12] Alessandra D'Amico,[11] Alexandre Reymond,[13] Norine Voisin,[13] Jonathan A. Bernstein,[14] Ellyn Farrelly,[15] Usha Kini,[16] Thomas A. Leonard,[17] Stéphanie Valence,[18] Lydie Burglen,[18] Linlea Armstrong,[19] Susan M. Hiatt,[20] Gregory M. Cooper,[20] Kimberly A. Aldinger,[21] William B. Dobyns,[21] Ghayda Mirzaa,[21] Tyler Mark Pierson,[22] Frank Baas,[2] Jamel Chelly,[23] Nicholas J. Cowan,[6] and David Anthony Keays[1,24,*]

[1]Research Institute of Molecular Pathology, Campus Vienna Biocenter 1, Vienna Biocenter (VBC), Vienna 1030, Austria
[2]Department of Clinical Genetics, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands
[3]Amsterdam UMC, Vrije Universiteit Amsterdam, Clinical Genetics, De Boelelaan 1117, Amsterdam, the Netherlands
[4]Department of Human Genetics, Radboud University Medical Center, 6525 GA Nijmegen, the Netherlands
[5]Department of Neurosciences, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 92093, USA
[6]Department of Biochemistry & Molecular Pharmacology, NYU Langone Medical Center, New York, NY 10016, USA
[7]Université Paris Descartes, Institut Cochin Hôpital Cochin, 75014 Paris, France
[8]Department of Neurology, University of Rochester Medical Center, Rochester, NY 14642, USA
[9]NIHR Oxford Biomedical Research Centre, Oxford, UK, Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK
[10]Department of Translational Medical Sciences, Section of Pediatrics, Federico II University, 80131 Naples, Italy
[11]Department of Advanced Medical Sciences, University of Naples Federico II, 80131 Naples, Italy
[12]Telethon Institute of Genetics and Medicine, 80078 Pozzuoli, Naples, Italy
[13]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland
[14]Stanford School of Medicine, Stanford, CA 94305, USA
[15]Stanford Children's Health, Palo Alto, CA 94304, USA
[16]Department of Clinical Genetics, Oxford Regional Genetics Service, Churchill Hospital, Oxford OX3 7LJ, UK
[17]Center for Medical Biochemistry, Medical University of Vienna, Max F. Perutz Laboratories, Vienna Biocenter (VBC), Campus Vienna Biocenter 5, 1030 Vienna, Austria
[18]Centre de référence des Malformations et Maladies Congénitales du Cervelet et Département de Génétique et Embryologie Médicale, APHP, Hôpital Trousseau, 75012 Paris, France
[19]Provincial Medical Genetics Programme, BCWH and Department of Medical Genetics, University of British Columbia, Vancouver, BC V6H 3N1, Canada
[20]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA
[21]Seattle Children's Research Institute, Center for Integrative Brain Research, Seattle, WA 98101, USA
[22]Departments of Pediatrics and Neurology & the Board of Governors Regenerative Medicine, Institute Cedars Sinai Medical Center, Los Angeles, CA 90048, USA
[23]Service de Diagnostic Génétique, Hôpital Civil de Strasbourg, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg, France
[24]Lead Contact
*Correspondence: keays@imp.ac.at
https://doi.org/10.1016/j.neuron.2018.10.044

SUMMARY

Corpus callosum malformations are associated with a broad range of neurodevelopmental diseases. We report that *de novo* mutations in *MAST1* cause mega-corpus-callosum syndrome with cerebellar hypoplasia and cortical malformations (MCC-CH-CM) in the absence of megalencephaly. We show that MAST1 is a microtubule-associated protein that is predominantly expressed in post-mitotic neurons and is present in both dendritic and axonal compartments. We further show that *Mast1* null animals are phenotypically normal, whereas the deletion of a single amino acid (L278del) recapitulates the distinct neurological phenotype observed in patients. In animals harboring *Mast1* microdeletions, we find that the PI3K/AKT3/mTOR pathway is unperturbed, whereas Mast2 and Mast3 levels are diminished, indicative of a dominant-negative mode of action. Finally, we report that *de novo MAST1* substitutions are present in patients with autism and microcephaly, raising the prospect that mutations in this gene give rise to a spectrum of neurodevelopmental diseases.

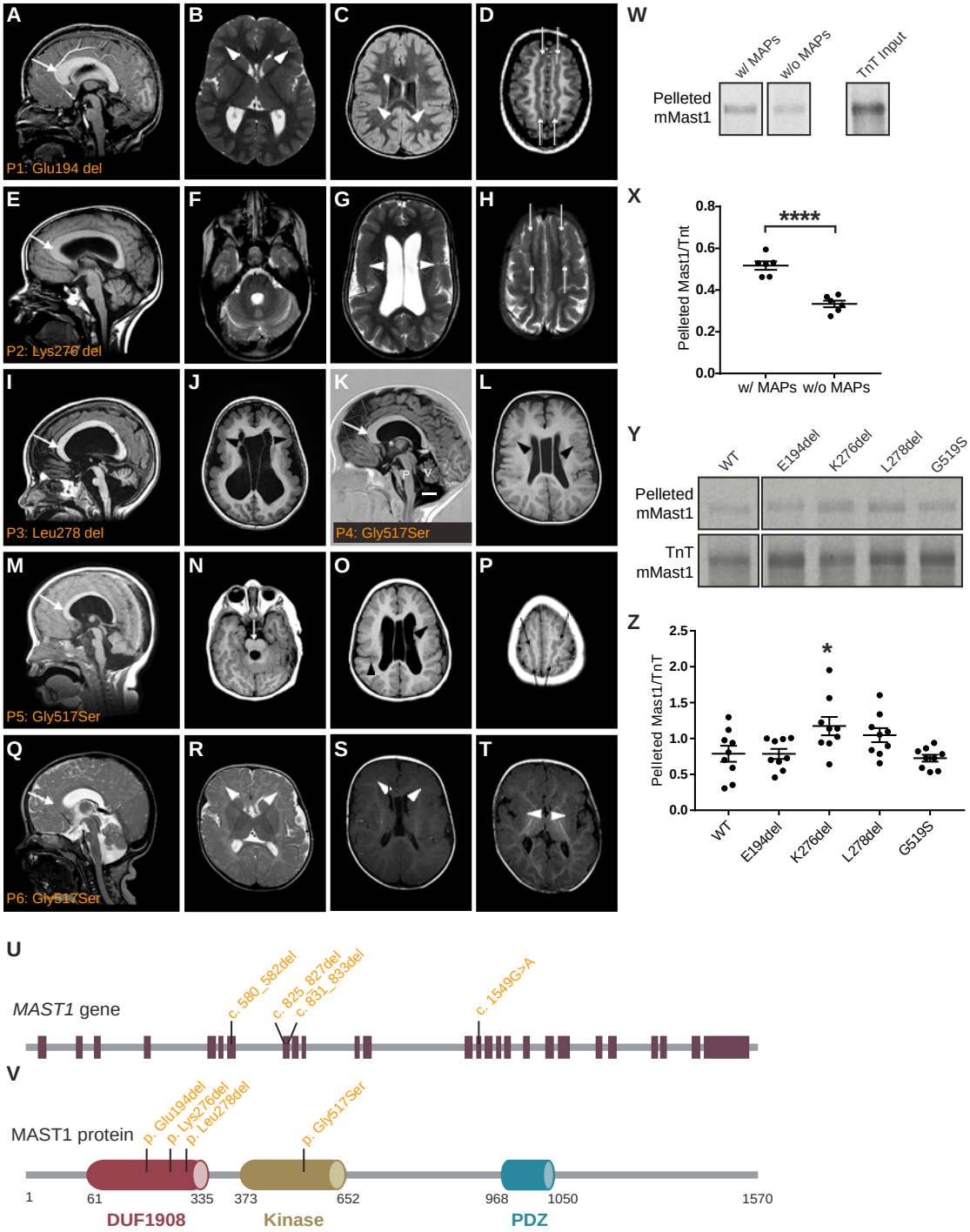


Figure 1. Patients with *MAST1* Mutations

(A–T) Selected magnetic resonance images from patients P1 (A–D), P2 (E–H), P3 (I and J), P4 (K and L), P5 (M–P), and P6 (Q–T) in the midline sagittal plane (A, E, I, K, M), parasagittal plane (Q), and axial planes through the brainstem (F and N), lateral ventricles (B, C, G, J, L, O, R, S, and T), and high convexities (D, H, and P). All patients have a cortical malformation or dysgyria characterized by diffuse undersulcation, shallow sulci (arrowheads in B, C, G, J, L, O, R, S, and T point to selected more obvious areas), and, in the more severely affected, mildly thick cortex (G and J; thus consistent with mild lissencephaly). While diffused, the cortical malformation appears most severe in the posterior frontal and perisylvian regions. The lateral ventricles are mildly to moderately enlarged, and the corpus callosum is abnormally thick (arrows in A, E, I, K, M, and Q), accompanied by mildly thick white matter. The brainstem, especially the pons, is mildly (E) or moderately (A, I, K, M, and Q) small, and in at least one child a prominent ventral midline cleft of the pons is seen (arrow in N). Available axial images through the high convexity of the cerebral hemispheres showed very dysplastic, longitudinally oriented gyral pattern (long arrows in D, H, and P).

(legend continued on next page)

INTRODUCTION

The bilateral integration of sensory, motor, and cognitive inputs is mediated by the corpus callosum, the largest white matter tract of the brain (Paul et al., 2007). A broad array of neurodevelopmental disorders are known to be associated with malformations of this structure. For instance, corpus callosum agenesis has been reported in patients with microcephaly, lissencephaly, and polymicrogyria and has been described in patients with autism (Parrini et al., 2016). Mouse and human genetics studies have provided insight into the molecular machinery that is required for the development of this important anatomical feature (Edwards et al., 2014). It has been shown that the midline crossing of post-mitotic neurons requires transcription factors such as Satb2 (Britanova et al., 2008), cell adhesion molecules such as L1-CAM (Demyanenko et al., 1999), guidance molecules such as Netrin1 and the semaphorins (Niquille et al., 2009; Serafini et al., 1996), and cytoskeletal proteins including Map1b and the β-tubulin Tubb3 (Meixner et al., 2000; Tischfield et al., 2010). Collectively, these molecules specify the fate of neurons destined to traverse the cerebral hemispheres, guiding their leading process to the correct destination.

While thinning of the corpus callosum is relatively common, in rare instances patients present with a thickening of this myelinated structure (Edwards et al., 2014; Marsh et al., 2017). To date, this phenotype has been reported in patients with neurofibromatosis and megalencephaly-polymicrogyria-mega-corpus-callosum syndrome (DiMario et al., 1999; Göhlich-Ratmann et al., 1998). These diseases are associated with a generalized enlargement in brain size, driven by activation of the PI3K/AKT3/mTOR pathway attributable to mutations in *PIK3R2* and *NF1* (Johannessen et al., 2005; Nguyen et al., 2016; Terrone et al., 2016). As mutations in these genes do not account for all cases, it is apparent that our understanding of the molecular pathology that underlies mega-corpus-callosum syndromes is incomplete (Hengst et al., 2010). Here, we present a cohort of patients with an enlarged corpus callosum in the absence of megalencephaly, harboring mutations in the uncharacterized microtubule-associated protein (MAP) *MAST1*.

RESULTS

Identification of *MAST1* Mutations

As part of an ongoing endeavor to identify genetic variants associated with structural brain malformations, we undertook whole-exome sequencing on 7 patient-parent trios where the affected individual presented with a striking enlargement of the corpus callosum. This led to the identification of *de novo* mutations in the gene *MAST1* (*microtubule-associated serine threonine kinase 1*) in 6 of the 7 affected individuals (Figures 1A–1T). Each of these patients presented with a hyperplastic corpus callosum (particularly over the genu and mid-body), cerebellar hypoplasia, ventricular dilation, and impaired motor and verbal performance (Table 1). Four patients presented with gyral simplification (P2 [Pierson et al., 2008], P3, P5, P6), three with dysplastic longitudinal gyri (P1, P5, P6), and one (P4) with periventricular nodules of white matter. We refer to this syndrome as mega-corpus-callosum with cerebellar hypoplasia and cortical malformations (MCC-CH-CM).

Through the Genematcher platform, four additional patients with *de novo* mutations in *MAST1* were identified who presented with either microcephaly accompanied by motor deficits (P7, P8) or autism spectrum disorder (P9, P10) (Figures S1A–S1E, Table S1) (Gilissen et al., 2014; Sobreira et al., 2015). In each case the mutation was verified by Sanger sequencing, was unreported in publicly available genome databases (e.g., dbSNP, 1000 genome, ExAC; see Table S2), and was in a highly conserved residue (Figures S1F–S1L) that was predicted to be deleterious when mutated (Table 1, Table S1, CADD score) (Kircher et al., 2014; Lek et al., 2016). We did not identify any unreported variants in our patient cohort, with the exception of a silent mutation in *TUBGCP5* (T457T) in P2 and a *de novo* mutation in NIMA related kinase 1 (NEK1) in P8 (Table 1, Table S1). It should be noted that mutations in NEK1 have been previously associated with amyotrophic lateral sclerosis, cilial dysfunction, and skeletal diseases; however, patient P8 did not present with symptoms consistent with these disorders (Kenna et al., 2016; Thiel et al., 2011). Three of the MCC-CH-CM-associated variants in MAST1 were single amino acid deletions positioned in the hydrophobic core of a four-helix bundle in the domain of unknown function DUF1908 (P1, p.Glu194del; P2, p.Lys276del; P3, p.Leu278del) (Figure S1M), while the remainder (P4–P6, p.Gly517Ser) harbored a recurrent missense mutation in the kinase domain of the protein (Figures 1U and 1V). These microdeletions were not found in control individuals in the ExAC genome browser; however, a single microdeletion between the kinase and PDZ domains (Glu697del) has been reported (Table S2).

MAST1 Is a Microtubule-Associated Protein

Little is known of the function of the MAST family of proteins (MAST1–4). MAST2 was initially cloned from testes and was

(U) Schematic representation of the *MAST1* genomic locus shows the position of the mutations identified in patients P1–P6.
(V) The MAST1 protein consists of a domain of unknown function (DUF1908, shown in red), a kinase domain (shown in yellow), and a PDZ domain (shown in blue). The amino acid boundaries of each of the domains are shown.
(W–Z) Autoradiograph showing the results of the microtubule binding assay with Mast1. Murine Mast1 was radiolabeled (^{35}S) by *in vitro* transcription and translation (TnT) in rabbit reticulocyte lysate, before incubation with a porcine microtubule extract in the presence or absence of microtubule-associated proteins (MAPs) (W and X). Following microtubule polymerization in the presence of Taxol, pelleted microtubules were analyzed by polyacrylamide gel electrophoresis (PAGE), and the ratio of pelleted radiolabeled-mMast1 to TnT input was determined. This experiment revealed a decrease of binding of Mast1 to microtubules in the absence of MAPs (X; n = 6 technical replicates; two-tailed unpaired t test; $t7 = 7.134$, $p < 0.0001$). Patient mutations were introduced into mMast1, radiolabeled by *in vitro* TnT, and microtubule binding was assessed (Y and Z). The ratio of microtubule-bound Mast1 (in the pelleted fraction) to TnT input was determined. Comparison of pelleted wild-type Mast1 to the K276del mutation shows a significant alteration in microtubule binding and a similar trend for the L278del (Z; n = 6–9 repeated experiments; one-way ANOVA with Dunnett's multiple comparison; WT versus K276del $p < 0.05$).
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. Error bars show mean ± SEM.

Table 1. Clinical Summary of Patients with MCC-CH-CM

Patient	P1	P2	P3	P4	P5	P6
Mutation (chr19)	19:12958677delGAG	chr19:12962798delGAA	chr19:12962804delGTT	19:12975903G>A	19:12975903G>A	19:12975903G>A
Mutation (NM_014975.2)	Glu194del c.580_582del	Lys276del c.825_827del	Leu278del c.831_833del	Gly517Ser c.1549G>A	Gly517Ser, c.1549G>A	Gly517Ser, c.1549G>A
CADD v1.3	16,71	17,35	21,4	27,9	27,9	27,9
Inheritance	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*	*de novo*
Ethnicity	Caucasian	Caucasian	Caucasian	Moroccan	Caucasian	Caucasian
Geographic ancestry	Italian	Hungarian	French	Moroccan	Unknown	French
Sex	Male	Male	Female	Female	Female	Female
Age at last evaluation	9 years	11.5 years	6 years	6 years	10 years	1.5 year
OFC[a] at birth	Unknown	36.5 cm (50^{th}–75^{th} percentile)	Unknown	Unknown	Unknown	35 cm (50^{th} percentile)
OFC[a]	52 cm	53 cm (25^{th}–50^{th} percentile)	51 cm	52 cm (+0.6 SD)	49.2 cm at 2 years, 9 months (50^{th} percentile)	46.5 cm at 18 months
Cortical dysgenesis	Dysplastic longitudinal gyri; subependymal heterotopias	Extensive undersulcated gyral pattern over frontal, temporal, and perisylvian regions; dysplastic longitudinal gyri	Tubulinopathy-like dysgyria with mildly thick cortex and diffuse very shallow sulci	Periventricular lesions	Diffuse subtle dysgyria; radial/shallow sulci with some deep infolds; dysplastic longitudinal gyri	Subtle frontal dysgyria; dysplastic longitudinal gyri
Basal ganglia	Normal	Poorly developed	Normal	Normal	Normal	Normal
Cerebellum	Vermis hypoplasia (++); counterclockwise rotation of cerebellar vermis	Mild diffuse cerebellar hypoplasia (+) with prominent sulci	Hypoplasia (++)	Hypoplasia (++)	Vermis hypoplasia (++) and mild hemispheric hypoplasia	Vermis hypoplasia (++) and mild hemispheric hypoplasia
Brainstem	Hypoplasia (++)	Small pons (+)	Hypoplasia (++)	Pontine hypoplasia (++)	Severe brainstem hypoplasia with a very small pons (++), thin midline cleft, long and mildly enlarged medulla	Severe brainstem hypoplasia (++), small pons with relative sparing of its buldging
Corpus callosum	Hyperplasia, mostly over genu and anterior body (++)	Very thick and dysplastic corpus callosum (mega corpus callosum) (++)	Hyperplasic (++)	Hyperplasia, mostly over genu and anterior body (+)	Thick corpus callosum (+)	Thick corpus callosum (+)
Ventricular dilation	Enlarged 4^{th} ventricle (+)	Enlarged 3^{rd} ventricle (++)	Enlarged 3^{rd} ventricle (++)	Enlarged 4^{th} ventricle (++)	Enlarged 3^{rd} and 4^{th} ventricles (++)	(+)
Cognitive abilities	Moderate cognitive impairment	Intellectual disability	Severe encephalopathy	Severe intellectual disability	Global developmental delay	Global developmental delay
Verbal abilities	Poor	Non-verbal	Non-verbal	Non-verbal	Non-verbal	Only some sounds (vowels)

(*Continued on next page*)

Table 1. *Continued*

Patient	P1	P2	P3	P4	P5	P6
Motor abilities	Clumsy stumbling gait, generalized hypotonia	Generalized hypotonia, mild truncal ataxia	Severe hypotonia	Able to stand with support, hyperreflexia	Can sit up by herself with assistance from her arms; able to roll	Hypotonic, sitting: with cyphosis, and only with support on arms; able to roll
Other	Short stature, paroxysmal EEG abnormalities (spike-wave complexes) with no clinical seizures	Infantile-onset epilepsy, oculomotor apraxia	None	Short stature, strabismus	Seizures (onset 6 years), history of aspiration	Head thrusts suggestive of oculomotor anomality
Other *de novo* or novel variants	None	Inherited a novel silent mutation in TUBGCP5: chr15 (22851109) C>T (T457T)	None	None	None	None

[a]OFC, occipitofrontal circumference

shown to function as a kinase and interact with microtubules via other MAPs (Walden and Cowan, 1993). To ascertain whether MAST1 associates with the microtubule cytoskeleton, we performed an *in vitro* transcription and translation (TnT) reaction employing rabbit reticulocyte lysate with radiolabeled murine (m) Mast1 (which shares 94% sequence identity with human MAST1). The reaction products were then assayed for their ability to bind to microtubules in the absence or presence of MAPs. We found that mMast1 associates with Taxol-stabilized microtubules in a MAP-dependent manner (Figures 1W and 1X; n = 6; $p < 0.0001$). Next, we asked whether the *MAST1* mutations found in our patient cohort influence microtubule binding. We introduced each mutation into our pcDNA 3.1 vector, labeled Mast1 with ^{35}S-methionine, and repeated the aforementioned microtubule binding assay. While translation efficiency was similar for all variants, we found that the K276 deletion significantly enhanced Mast1 binding to microtubules (Figures 1Y and 1Z; n = 9; WT versus K276del $p < 0.05$). Taken together, these data show that Mast1 binds to microtubules in a MAP-dependent manner and that mutations can perturb this interaction.

MAST1 Is Expressed in Post-mitotic Neurons

To gain insight into the role of MAST1 in neurodevelopment, we analyzed its expression in the human and mouse brain during early developmental stages (Figures 2A and 2B). To this end we extracted mRNA from the developing murine brain (E10.5, E12.5, E14.5, E16.5, E18.5, P0, and P6), generated cDNA, and performed qPCR analysis. We observed that Mast1 expression begins at E12.5, peaks at E16.5, and decreases postnatally (Figure 2B). Consistent with this result, qPCR analysis of human fetal brain cDNA showed moderate expression of *MAST1* at gestation weeks 13 and 22 (Figure 2A). To determine which cells express Mast1, we performed immunohistochemistry on the murine cortex at E12.5, E14.5, E16.5, and P0 (n = 3). We used a commercially available Mast1 antibody, having validated its specificity employing a Mast1 knockout (KO) animal (Figures S3B–S3D). We observed staining in the developing cortical plate and intermediate zone at these time points (Figures 2C–2F). Further analysis at E14.5 and E16.5 showed that Mast1 is located in the cytoplasm in Tuj-positive post-mitotic neurons but is largely absent from Sox2-positive progenitors and Tbr2-positive intermediate progenitors (Figures 2S–2AP) (n = 3). This expression pattern is mirrored in human cerebral organoids that were cultured for 64 days, where we again observed co-localization with post-mitotic markers (Figures S2A–S2F) (n = 10 organoids) (Lancaster et al., 2013). Immunostaining on P0 mouse brain sections showed that Mast1 is present in Tuj-positive corpus callosal fibers that cross the midline (n = 3) (Figures S2G–S2L). To gain further insight into the subcellular localization of Mast1, we cultured primary P0 cortical neurons and P7 cerebellar granule neurons and performed immunostaining. We again validated the specificity of our Mast1 antibody for cell culture experiments employing a *Mast1 KO* animal (Figure S2S-T). We performed double staining with antisera specific for the axonal marker Tau or the dendritic marker Map2 (Kosik and Finch, 1987). We observed Mast1 staining throughout the soma, as well as in dendritic and axonal compartments in cortical neurons (Figures 2G–2R, n = 5) and in cerebellar granule neurons

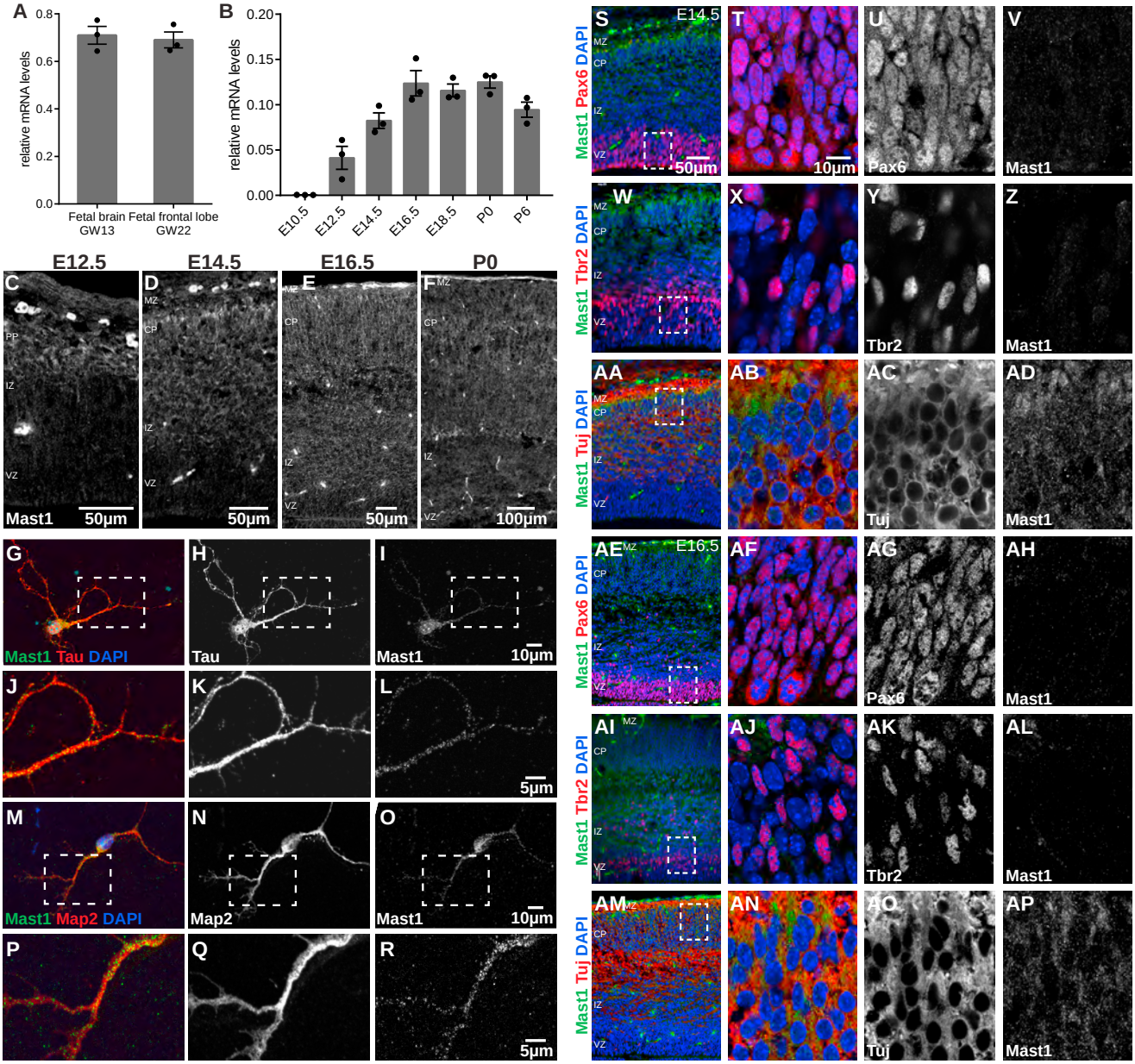


Figure 2. Mast1 Expression in Human and Mouse Embryonic Brain

(A) qPCR analysis reveals that *MAST1* mRNA is expressed in the human developing fetal brain at gestational week (GW) 13 and the fetal frontal lobe at GW22.

(B) qPCR analysis performed on mouse brain cDNA libraries from E10.5 to P6 show that Mast1 expression peaks at E16.5 in mice (n = 3 animals per time point).

(C–F) Immunohistochemistry employing a validated Mast1 antibody (see also Figures S3B-S3D) indicates staining in the post-mitotic cortical plate and intermediate zone from E12.5 to P0 in mice.

(G–R) Maximum-projection images of cultured P0 cortical neurons at 5-DIV staining with the axonal marker Tau (G–L) and the dendritic marker Map2 (M–R) show that Mast1 (I, L, O, and R) is present in both axonal and dendritic compartments. Dashed boxes in (G)–(I) and (M)–(O) are expanded in (J)–(L) and (P)–(R), respectively.

(S–AP) Immunohistochemistry employing the progenitor marker Pax6 (S–U and AE–AG), intermediate progenitor marker Tbr2 (W–Y and AI–AK), and post-mitotic marker Tuj (AA–AC and AM–AO) on E14.5 (S–AD) and E16.5 (AE–AP) murine sections reveals that Mast1 expression is restricted to post-mitotic neurons at these time points (PP, preplate; CP, cortical plate; MZ, marginal zone; IZ, intermediate zone; VZ, ventricular zone).

Error bars show mean ± SEM.

(Figures S2M–S2R, n = 3). This staining had a punctate appearance, suggesting that Mast1 may associate with vesicular structures that are trafficked along the microtubule cytoskeleton. qPCR analysis of Mast1 expression in the adult mouse showed that its expression persists in all brain regions (albeit at much lower levels) and that it is present in the testes, liver, and spleen

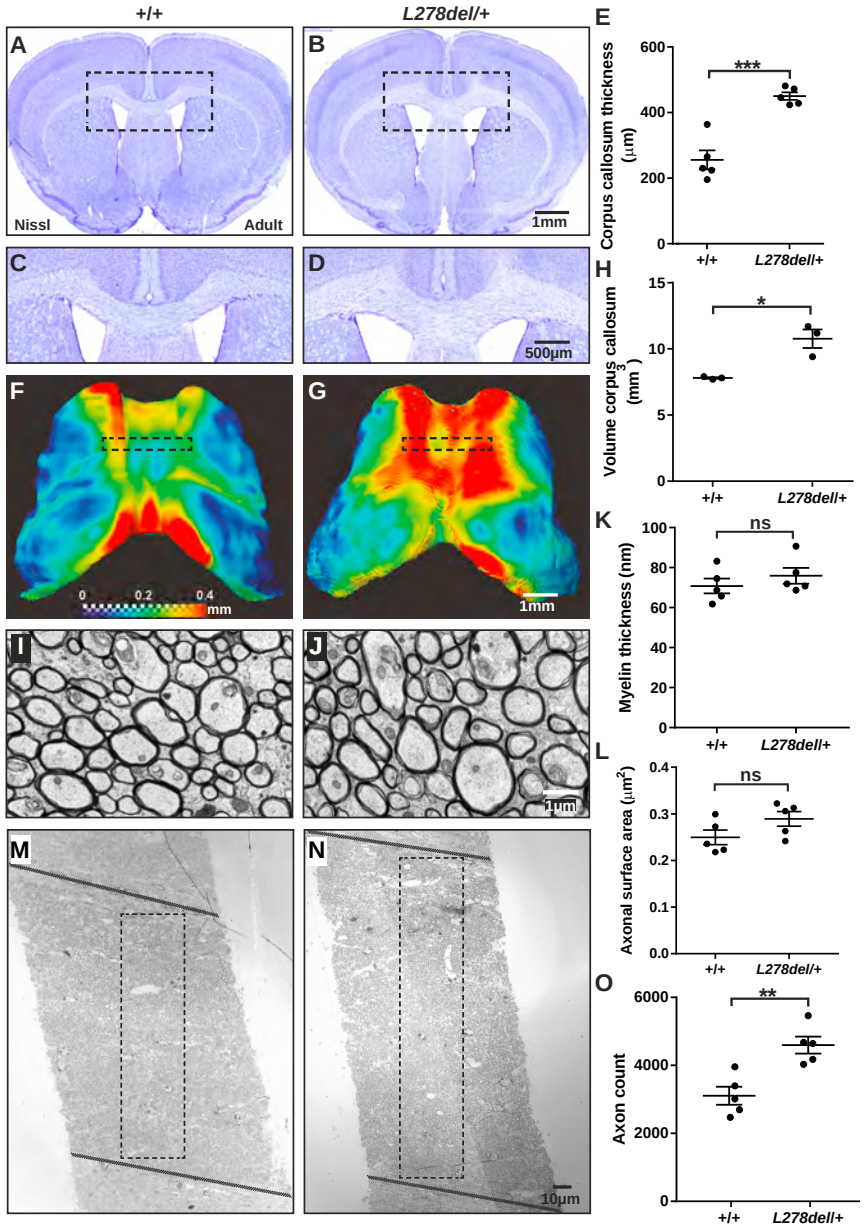


Figure 3. *L278del* Mice Have an Enlarged Corpus Callosum Associated with an Increase in the Number of Callosal Axons

(A–D) Nissl-stained sections of 8-week-old adult brains highlighting the thicker corpus callosum in *L278del*/+ animals compared to wild-type littermates (black boxes in A and B are expanded in C and D).

(E) Quantification of the thickness at the septum reveals a significantly thicker corpus callosum in *L278del*/+ animals (n = 5 animals per genotype; unpaired t test; +/+ versus *L278del*/+; t8 = 6.217, $p < 0.001$).

(F and G) MRI reconstructions of the corpus callosum in wild-type controls (F) and *L278del*/+ mice (G). The heatmap reflects the thickness of the corpus callosum (blue, thinner region; red, thicker region). Note that in *L278del*/+ animals the genu and midbody region are most affected.

(H) MRI volumetric quantification of the corpus callosum shows that this structure is significantly larger in *L278del*/+ animals compared to wild-type littermates (n = 3 animals per genotype; two-tailed unpaired t test; +/+ versus *L278del*/+; t4 = 4.233, $p < 0.05$).

(I and J) Electron microscopy images showing cross-sections of the corpus callosum at the midbody (region depicted with boxes in F and G). Myelinated axons can be clearly visualized.

(K) Quantification of myelin thickness reveals no significant different between *L278del*/+ animals and wild-type littermates (n = 5 animals per genotype, >500 myelinated axons per animal; two-tailed unpaired t test; t8 = 1.001, $p > 0.1$).

(L) Quantification of axonal caliber reveals no significant different between *L278del*/+ animals and wild-type littermates (n = 5 animals per genotype, >1,500 myelinated axons per animal; +/+ versus *L278del*/+; two-tailed unpaired t test; t8 = 1.786, $p > 0.1$; see also Figure S3N).

(M and N) Assessment of the total number of myelinated axons within a 30-μm-wide box extending along the ventro-dorsal axis of the corpus callosum.

(O) Quantification reveals a significant increase in axonal count in the *L278del*/+ animals in comparison to wild-type littermates (n = 5 animals per genotype, 3 images analyzed per animal; two-tailed unpaired t test; t8 = 4.095, $p < 0.01$).

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. Error bars show mean ± the SEM.

at very low levels (n = 3) (Figure S2U). These results are consistent with expression data that are available through online databases (e.g., Allen Brain Atlas, GTEx) (GTEx Consortium, 2013; Morris et al., 2010). Taken together, these data show that Mast1 is predominantly expressed in post-mitotic neurons in the developing nervous system.

An Enlarged Corpus Callosum in a *L278Del Mast1* Mouse Model

To investigate the effect of *Mast1* mutations *in vivo*, we exploited the power of the CRISPR-Cas9 genome editing system to generate a *Mast1 KO* mouse and a *Mast1 L278del* (*L278del*) mouse employing guide RNAs targeted to exon 8 (Figure S3A) (Wang et al., 2013). We decided to recapitulate the mutation identified in patient P3 because of the striking callosal, cortical, and cerebellar phenotypes. As homozygous *L278del/del* animals die shortly after birth, we focused our initial analysis on homozygous *Mast1 KO* and heterozygous *L278del*/+ mice, which are viable. We analyzed the thickness of the genu of the corpus callosum on rostral Nissl-stained coronal sections (bregma 0.86 mm). While *Mast1 KO* animals were indistinguishable from wild-type littermate controls (Figures S3E–S3G; n = 5; +/+ versus *KO/KO*; $p > 0.5$), *L278del*/+ animals presented with a significantly thicker corpus callosum in comparison to wild-type littermate controls (Figures 3A–3E; n = 5; +/+ versus *L278del*/+; $p < 0.001$). High-resolution volumetric magnetic

resonance imaging (MRI) analysis confirmed an increased corpus callosum volume in *L278del*/+ animals (Figure 3H; n = 3; +/+ versus *L278del*/+; $p < 0.05$) and revealed that the thickening was most prominent in the rostral regions encompassing the genu and the mid-body (Figures 3F and 3G). The enlarged corpus callosum was also present in *L278/del* mice aged 10 days (Figures S3K–S3M; n = 5; +/+ versus *L278del*/+; $p < 0.05$). While off-target mutations in Cas9-modified mice are rare (Iyer et al., 2015), we confirmed that this phenotype was not due to a linked mutation by undertaking whole-exome sequencing for our *L278del*/+ mouse line. We identified no other non-synonymous mutations on chromosome 8 in this line (Table S3). Moreover, we generated a second independent *L278del* mouse line and once again observed an enlarged corpus callosum (Figures S3H–S3J; +/+ versus *L278del*/+*_Line#7*; n = 3; $p < 0.01$).

We then asked whether the increase in the thickness of the corpus callosum of the *L278del*/+ mice was due to an increase in the thickness of myelinated axons. In light of our MRI studies, we focused our analysis on the midbody of the corpus callosum, where we observe the most pronounced phenotype in *L278del*/+ mutants (bregma −0.10). The corpus callosum of adult animals was microdissected, samples were stained to enable clear visualization of myelinated neurons crossing the midline, and images captured at high resolution (8,900×) with transmission electron microscopy (Figures 3I and 3J) (Sturrock, 1980; West et al., 2015). We observed no significant difference in either myelin thickness (Figure 3K; n = 5; n > 500 axons per animal; +/+ versus *L278del*/+; $p > 0.1$) or axonal caliber (Figure 3L; n = 5; n > 1500 axons per animal; +/+ versus *L278del*/+; $p > 0.1$) when comparing littermate controls with *L278del*/+ animals. An analysis of the distribution of axon calibers in *L278del*/+ animals again revealed no significant difference in comparison to littermate controls (n = 5; +/+ versus *L278del*/+; $p > 0.1$ for all bins) (Figure S3U). Reflecting this, an assessment of the G-ratio (axonal diameter/total fiber diameter), revealed no significant difference between genotypes (n = 5, n > 300 fibers animal, +/+ versus *L278del*/+; $p > 0.1$) (Figure S3N) (Guy et al., 1989).

Given these results, we asked whether more myelinated axons might cross the midline in *L278del*/+ animals. To assess this, TEM images were acquired at a magnification of 710× along the dorso-ventral axis of the corpus callosum. These images (10–12 per animal) were tiled, resulting in a single image that encompassed the entire cross-section of the corpus callosum. A box 30 μm wide was overlaid onto the structure, and the total number of myelinated axons was counted manually (Figures 3M and 3N). Statistical analysis revealed a significant increase in the number of myelinated axons within that box in *L278del*/+ animals in comparison to littermate controls (Figure 3O; n = 5 animals per genotype, n = 3 images per animal, +/+ versus *L278del*/+; $p < 0.01$). Finally, we counted the number of Olig2-positive oligodendrocytes and GFAP-positive astrocytes in a box of fixed width that was overlaid onto coronal sections of the corpus callosum (Figures S3O–S3R). This revealed that in *L278del*/+ animals there are more oligodendrocytes and astrocytes in comparison to wild-type controls (n = 5; +/+ versus *L278del*/+; Olig2, $p < 0.05$; GFAP, $p < 0.05$) (Figures S3S and S3V). There was, however, no significant difference in the density of these cell types when comparing genotypes (n = 5; +/+ versus *L278del*/+; Olig2, $p > 0.5$; GFAP, $p > 0.05$) (Figures S3T and S3W). Taken together, these data show that our *L278del*/+ mouse model recapitulates the enlarged corpus callosum observed in affected patients, which we attribute to an increase in the number of myelinated axons crossing the midline.

Cortical and Cerebellar Hypoplasia in the *L278del Mast1* Mouse Model

We extended our analysis of the *L278del* and *KO* mouse lines to the cortex and cerebellum, which are consistently affected in our *MAST1* patient cohort. MRI volumetric analysis of adult animals revealed an overall reduction in cortical volume in *L278del* heterozygotes (Figures 4A–4C; n = 3; +/+ versus *L278del*/+; $p < 0.05$). We confirmed this finding by undertaking Nissl and NeuN staining, where we observed a significant reduction in cortical thickness in *L278del* animals that was most pronounced in caudal regions (bregma −1.82 mm; Nissl [Figures 4D–4F]: n = 5, +/+ versus *L278del*/+, $p < 0.001$; NeuN [Figures S4A–S4D]: n = 5; +/+ versus *L278del*/+; $p < 0.0001$). We did observe a reduction in cortical thickness in the rostral motor cortex in *L278del* heterozygotes (bregma 0.86 mm); however, this was not statistically significant (Figures S4E–S4G, n = 5; +/+ versus *L278del*/+; $p > 0.1$). We assessed the thickness of the cortical layers in the affected caudal regions in *L278del* heterozygotes and littermate controls. We observed a significant reduction in the thickness of layer 5, marked by the transcription factor Er81 (Figures 4J–4L; n = 5; +/+ versus *L278del*/+; $p < 0.01$). We did not observe a significant difference in the thickness of Foxp2-positive layer VI neurons (Figures S4H–S4K; n = 5; +/+ versus *L278del*/+; $p > 0.1$) (Ferland et al., 2003) or Cux1-positive granule cells (Figures 4G–4I; n = 5; +/+ versus *L278del*/+; $p > 0.5$) (Cubelos et al., 2015). Analysis of the caudal cortex at P10 by Nissl staining again revealed that it was thinner in *L278del*/+ mutants (Figure S4U; n = 5; +/+ versus *L278del*/+; $p < 0.01$). Consistent with our previous findings, adult *Mast1* KO animals appeared to be phenotypically normal, without morphological or layering defects in the adult *KO* cortex (Figures S4A–S4D and S4H–S4T).

MRI volumetric analysis of the cerebellum revealed a severe reduction in *L278del* heterozygotes (Figures 5A–5C; n = 3; +/+ versus *L278del*/+; $p < 0.001$). Nissl staining as well as NeuN and calbindin immunohistochemistry showed that the foliation and lamination of the cerebellum in *L278del* heterozygotes was comparable to wild types; however, the granule and the molecular cell layer appeared thinner in mutants (Figures 5D–5I). We calculated the density of granule cells as well as the linear density of Purkinje cells, allowing us to estimate the total number of these cell types in midsagittal sections. While the density of both cell types was similar in *L278del* mutants in comparison to littermate controls (Figures S5A and S5B), we observed a significant reduction in the total number of granule cells per section (Figure 5J; n = 5; +/+ versus *L278del*/+; $p < 0.05$). We also observed a reduction in the total number of Purkinje cells per section, but this reduction was not statistically significant (Figure 5K; n = 5; +/+ versus *L278del*/+; $p > 0.5$). Volumetric analysis of the colliculi, putamen, thalamus, and olfactory bulbs revealed no significant difference between wild-type littermates and

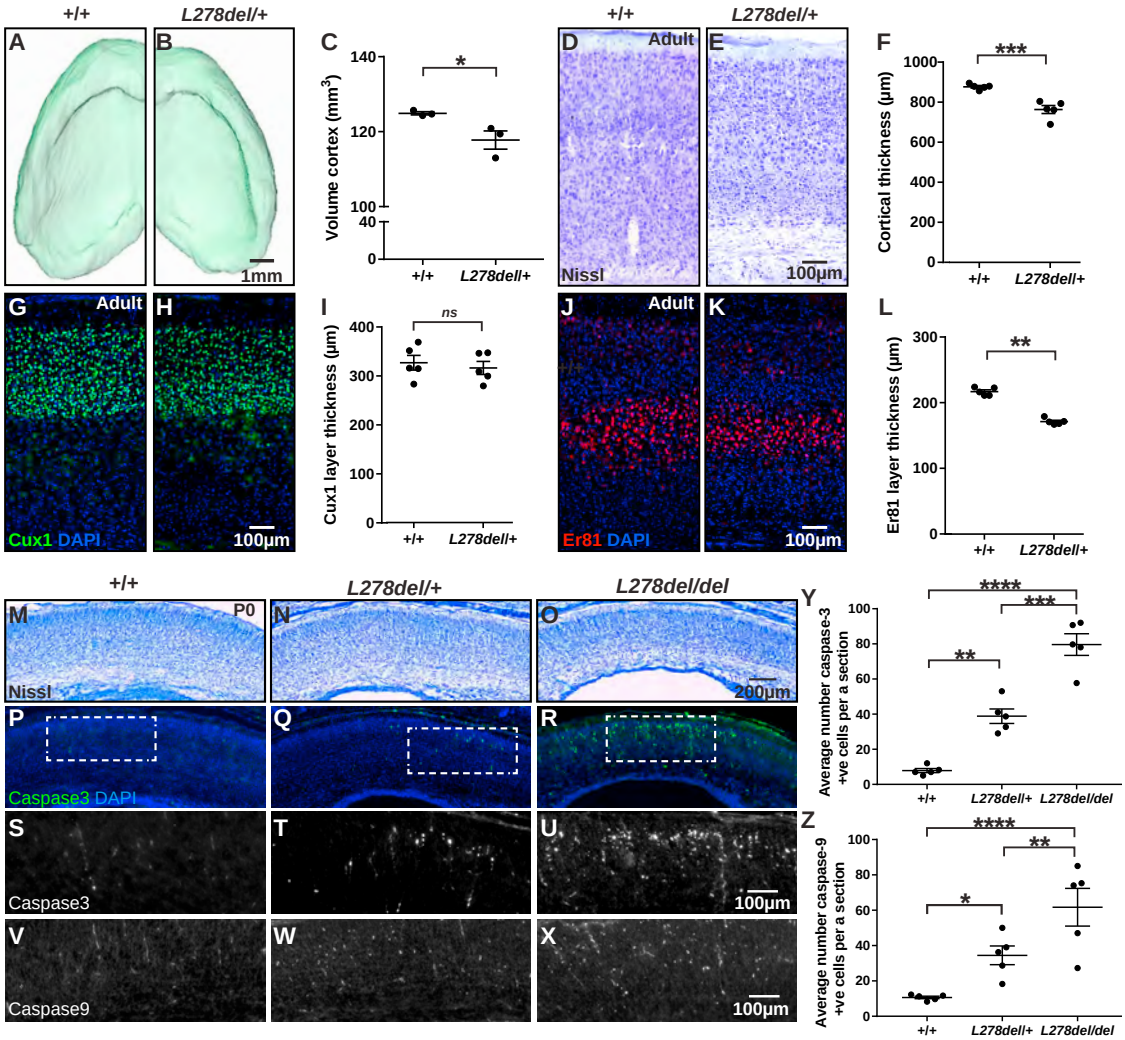


Figure 4. *L278del* Mice Have a Reduction in Cortical Volume Associated with an Increase in Neuronal Apoptosis

(A and B) MRI reconstructions of the cortex in adult animals reveal a reduction in cortical volume in *L278del*/+ mice in comparison to littermate controls.

(C) Quantification of the MRI cortical volume (n = 3 animals per genotype; two-tailed unpaired t test; +/+ versus *L278del*/+; t4 = 2.902, $p < 0.05$).

(D and E) Nissl-stained sections of the adult somatosensory cortex reveal a reduction in cortical thickness in *L278del*/+ adult mice in comparison to littermate controls.

(F) Quantification of caudal cortical thickness (n = 5 animals per genotype; two-tailed unpaired t test; +/+ versus *L278del*/+; t8 = 5.417, $p < 0.001$).

(G–L) Labeling of Cux1-positive layer II-III (G and H) and Er81-positive layer 5 neurons in wild-type and *L278del*/+ adult mice (J and K). Quantification (I) reveals no significant difference in Cux1 layer thickness when comparing genotypes but a significant reduction in the size of the Er81-positive layer (L) (n = 5 animals per genotype; two-way repeated-measures ANOVA with a Bonferroni test for multiple comparisons; +/+ versus *L278del*/+; Cux1 layer $p > 0.1$; Er81 layer $p < 0.01$).

(M–X) Representative Nissl-stained (M–O), activated caspase-3-stained (P–U), and activated caspase-9-stained (V–X) P0 cortical sections of littermate controls (M, P, S, and V), *L278del*/+ (N, Q, T, and W), and *L278del*/*del* (O, R, U, and X) animals. The white boxes in (P), (Q), and (R) are expanded in (S), (T), (U). The number of caspase-3-positive and caspase-9-positive cells in the cortex of P0 animals was counted and averaged (3 sections per animal). There is a dose-dependent increase in apoptotic cells in *L278del*/+ and *L278del*/*del* animals in comparison to littermate controls (see also Figure S4V).

(Y and Z) Quantitation of caspase-3 (Y) and caspase-9 (Z) staining (n = 5 animals per genotype; n = 3 sections per animal; two-way repeated-measures ANOVA with a Bonferroni test for multiple comparison. Caspase-3: +/+ versus *L278del*/+ $p < 0.01$; '+/+ versus *L278del*/*del* $p < 0.0001$; *L278del*/+ versus *L278del*/*del* $p < 0.001$. Caspase-9: +/+ versus *L278del*/+ $p < 0.05$; '+/+ versus *L278del*/*del* $p < 0.0001$; *L278del*/+ versus *L278del*/*del* $p < 0.01$).

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. Error bars show mean ± the SEM.

L278del heterozygotes (Figure S5C; n = 3; +/+ versus *L278del*/+; $p > 0.5$ for all regions). There was a reduction in the volume of the hippocampus in *L278del* heterozygotes, but this difference was not statistically significant (Figure S5C; n = 3; +/+ versus *L278del*/+; $p = 0.34$).

Neuronal Apoptosis in the *L278del Mast1* Mouse Model

To determine if the smaller cortex and cerebellum in our *L278del* mutant mice is a result of neuronal cell death during development, we performed an activated caspase-3 staining. We detected a dose-dependent increase in the number of apoptotic

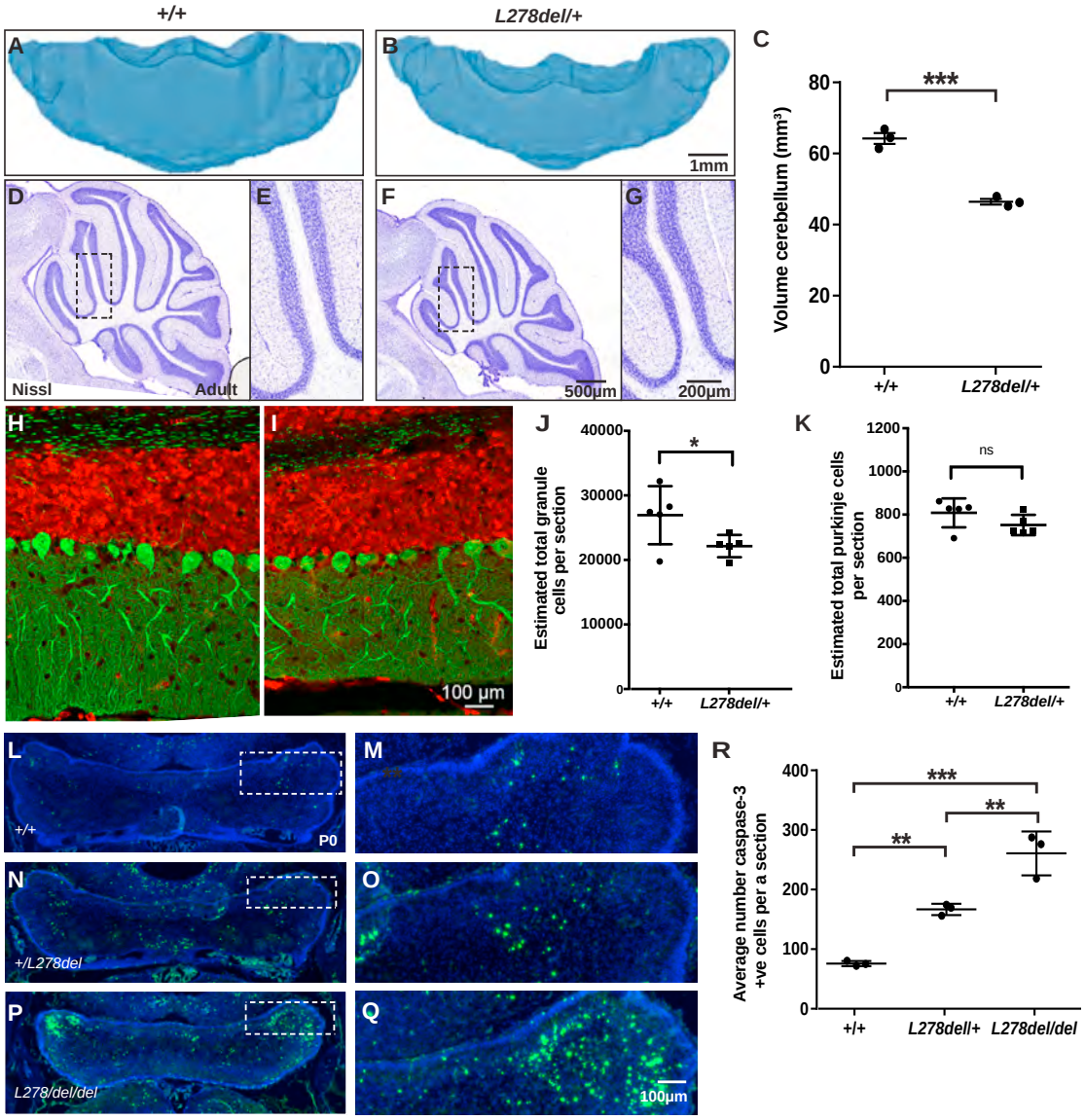


Figure 5. *L278del* Mice Have a Hypoplastic Cerebellum

(A and B) MRI reconstructions of the cerebellum from wild-type littermates (A) and *L278del*/+ (B) adult animals.

(C) Quantification reveals a significant reduction in cerebellar volume (mm^3) in *L278del*/+ animals (n = 3 animals per genotype; two-tailed unpaired t test; +/+ versus *L278del*/+; t4 = 10.16, $p < 0.001$).

(D–G) Nissl-stained sagittal sections of 8-week-old cerebellum confirm the reduction in cerebellar size and indicate that lamination within the cerebellum is normal in *L278del*/+ animals.

(H and I) Immunostaining employing the neuronal marker NeuN (shown in red) and the Purkinje cell marker Calbindin (shown in green) revealed a reduction in the thickness of the granule cell layer and molecular layer in *L278del*/+ animals.

(J and K) Estimates of the total number of granule cells (J) and Purkinje cells (K) per a midsagittal section in *L278del*/+ animals and littermate controls (n = 5 animals per genotype; two-way ANOVA with a Bonferroni correction; +/+ versus *L278del*/+; granule cell counts: $p < 0.05$; Purkinje cell counts: $p > 0.5$; see also Figures S5A and S5B).

(L–Q) Immunohistochemistry employing sera for activated caspase-3 reveals an increase in apoptosis in the *L278del*/+ and *L278del/del* animals in the developing cerebellum at P0. The white boxes in (L), (N), and (P) are expanded in (M), (O), and (Q), respectively.

(R) Quantification of caspase-3 staining (n = 3 animals per genotype; one-way ANOVA with Tukey's multiple-comparisons test; +/+ versus *L278del*/+ $p < 0.01$; +/+ versus *L278del/del* $p < 0.001$; *L278del*/+ versus *L278del/del* $p < 0.01$).

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. Error bars show mean ± the SEM.

cells in the P0 cortex of wild-type versus heterozygous and homozygous *L278del* mice (Figures 4P–4U, 4Y, and S4 AB; n = 5; +/+ versus *L287del*/+ $p < 0.01$; +/+ versus *L278del/del* $p < 0.0001$; *L278del*/+ versus *L278del/del* $p < 0.001$). Consistent with our anatomical analysis of adult animals, we found that apoptosis was more severe in the caudal regions of the P0 cortex in the *L278del*/+ and *L278del/del* animals (Figure S4V; n = 5; rostral versus caudal: *L278del*/+ $p < 0.001$; *L278del/L278del* $p < 0.0001$). At P10 and in 8-week-old adult mice, apoptosis in *L278del*/+ mice was comparable to wild-type controls (Figures S4W–S4Z; n = 5; +/+ versus *L278del*/+; P10, $p > 0.1$; Adult: $p > 0.1$). An analysis of activated caspase-3 staining in the P0 cerebellum revealed a large increase in *L278del*/+ and *L278del/del* animals in comparison to wild-types (Figures 5L–5R; n = 3; +/+ versus *L287del*/+ $p < 0.01$; +/+ versus *L278del/del* $p < 0.001$; *L278del*/+ versus *L278del/del* $p < 0.01$). To ascertain whether this cell death is associated with intrinsic suicide pathways, we stained P0 cortical sections with sera that target activated caspase-9, an upstream regulator of caspase-3 that is associated with mitochondrial mediated apoptosis (Hyman and Yuan, 2012). We observed a dose-dependent increase in the number of activated caspase-9-positive cells in the cortex of heterozygous and homozygous *L278del* mice (Figures 4V–4X and 4Z; n = 5; +/+ versus *278del*/+ $p < 0.05$; +/+ versus *L278del/del* $p < 0.0001$; *L278del*/+ versus *L278del/del* $p < 0.01$). Intriguingly, this was not associated with the upregulation of the tumor suppressor p53 (Figures S4AA–S4AF), which has been associated with a range of stressful cellular events including DNA damage, defects in cell cycle progression, as well as oxidative and nutritional stress (Szybińska and Leśniak, 2017). Taken together, these data show that the *L278del* mutation results in a hypoplastic cortex and cerebellum, a phenotype that is associated with neuronal apoptosis mediated by activated caspases 3 and 9, but not p53 upregulation.

The *L278* Microdeletion Alters Mast1/2/3 Protein Levels but Not the PI3K/AKT3/mTOR Pathway

Next, we exploited these mouse models to investigate the underlying molecular impairment associated with *MAST1* mutations. Western blot analysis of P0 brain lysates showed a complete absence of Mast1 protein in *KO* animals (n = 6), with an unexpected dose-dependent reduction in Mast1 levels in the *L278del* line (Figure 6A-D; n = 4, Mast1: +/+ versus *L278del*/+ $p < 0.0001$; +/+ versus *L278del/Ldel* $p < 0.0001$). This absence of protein in the *Mast1 KO* animals correlates with a dramatic reduction in mRNA levels, but this is not the case in *L278del* mice, where *Mast1* transcript levels are statistically similar in +/+, *L278del*/+, and *L278del/del* animals (Figures S6A and S6D). We assessed whether the reduction in Mast1 protein levels would influence the levels of other MAST family members. Relying on western blot analysis, we observed a striking reduction in the levels of Mast2 and Mast3 in our *L278del* animals (Figures 6A and 6B; n = 4; Mast2: +/+ versus *L278del*/+ $p < 0.0001$; +/+ versus *L278del/del* $p < 0.0001$; Mast3: +/+ versus *L278del*/+ $p < 0.0001$; +/+ versus *L278del/del* $p < 0.0001$). This was not attributable to changes in the mRNA levels of *Mast2* and *Mast3*, which were similar in +/+, *L278del*/+, and *L278del/del* animals (Figures S6B and S6C). In contrast, in our *Mast1 KO* animals we observed a signficant increase in the levels of Mast2 and Mast3 (Mast2: +/+ versus *KO* $p < 0.0001$; Mast3: +/+ versus *KO* $p < 0.001$), again with no change in the mRNA levels (Figures S6E and S6F). These data imply that the *L278del* mutation in Mast1 acts by a dominant-negative mode of action, whereas a form of post-transcriptional compensation occurs in the *Mast1 KO* line.

Finally, as previous studies have implicated activation of the PI3K/AKT3/mTOR pathway in corpus callosum hyperplasia associated with megalencephaly, we assessed the phosphorylation state of AKT and rpS6 in our *L278del* mice (Figure 6E) (Broix et al., 2016; Poduri et al., 2012; Rivière et al., 2012; Terrone et al., 2016). We observed no difference in the levels of p-AKT_{S473} and p-$rpS6_{S240/244}$ when comparing heterozygous and homozygous *L278del* brain lysates with wild-type controls (Figure 6F; n = 5 animals; +/+ versus *L278del*/+ AKT $p > 0.5$; +/+ versus *L278del/del* AKT $p > 0.5$; *L278del*/+ versus *L278del/del* AKT $p > 0.5$; +/+ versus *L278del*/+ rpS6 $p > 0.5$; +/+ versus *L278del/del* rpS6 $p > 0.5$; *L278del*/+ versus *L278del/del* rpS6 $p > 0.5$).

DISCUSSION

Here, we report that *de novo* mutations in *MAST1* cause MCC-CH-CM, a disease characterized by a striking enlargement of the corpus callosum, cerebellar hypoplasia, and cortical malformations. For patients with this constellation of phenotypes, mutations in *MAST1* appear to be the primary genetic cause, as we observed *MAST1* variants in 6 of the 7 patients that we studied. We have shown that MAST1 is expressed predominantly in post-mitotic neurons in the developing nervous system, and that it is present in the soma as well as dendritic and axonal compartments. We demonstrate that a *L278del* mouse line recapitulates the enlarged corpus callosum and cerebellar hypoplasia observed in patients, while the *Mast1 KO* line lacks any morphological defects. We report extensive cortical apoptosis in our *L278del* mice that is mediated by activated caspases 3 and 9, but not p53 upregulation. We report that this apoptosis varies along the rostro-caudal axis of the cortex, being most severe in caudal regions where the enlarged corpus callosum is least pronounced. Our results show that, unlike other syndromes associated with an enlarged corpus callosum, activation of the PI3K/AKT3/mTOR pathway does not appear to be the underlying pathogenic driver (Terrone et al., 2016).

What molecular mechanism underlies MCC-CH-CM? We have shown that MAST1 binds to microtubules in a MAP-dependent manner, but this binding is only altered in the case of the K276del mutation. It therefore seems unlikely that alterations in microtubule affinity cause the spectrum of phenotypes we observe in patients with MCC-CH-CM. Analysis of our *L278del* mouse model has shown a dramatic reduction in the levels of Mast2 and Mast3 protein. This result suggests that pathogenic microdeletions in Mast1 act by a dominant-negative mechanism and that the Mast proteins (like other AGC kinases) may physically interact (Leroux et al., 2018). Notably, three of the deletions detected in patients were located in the domain of unknown function (DUF1608), which adopts a 4-helix bundle structure in solution (PDB: 2M9X). Since the helices pack together to form a hydrophobic core, it is expected that the Glu194del,

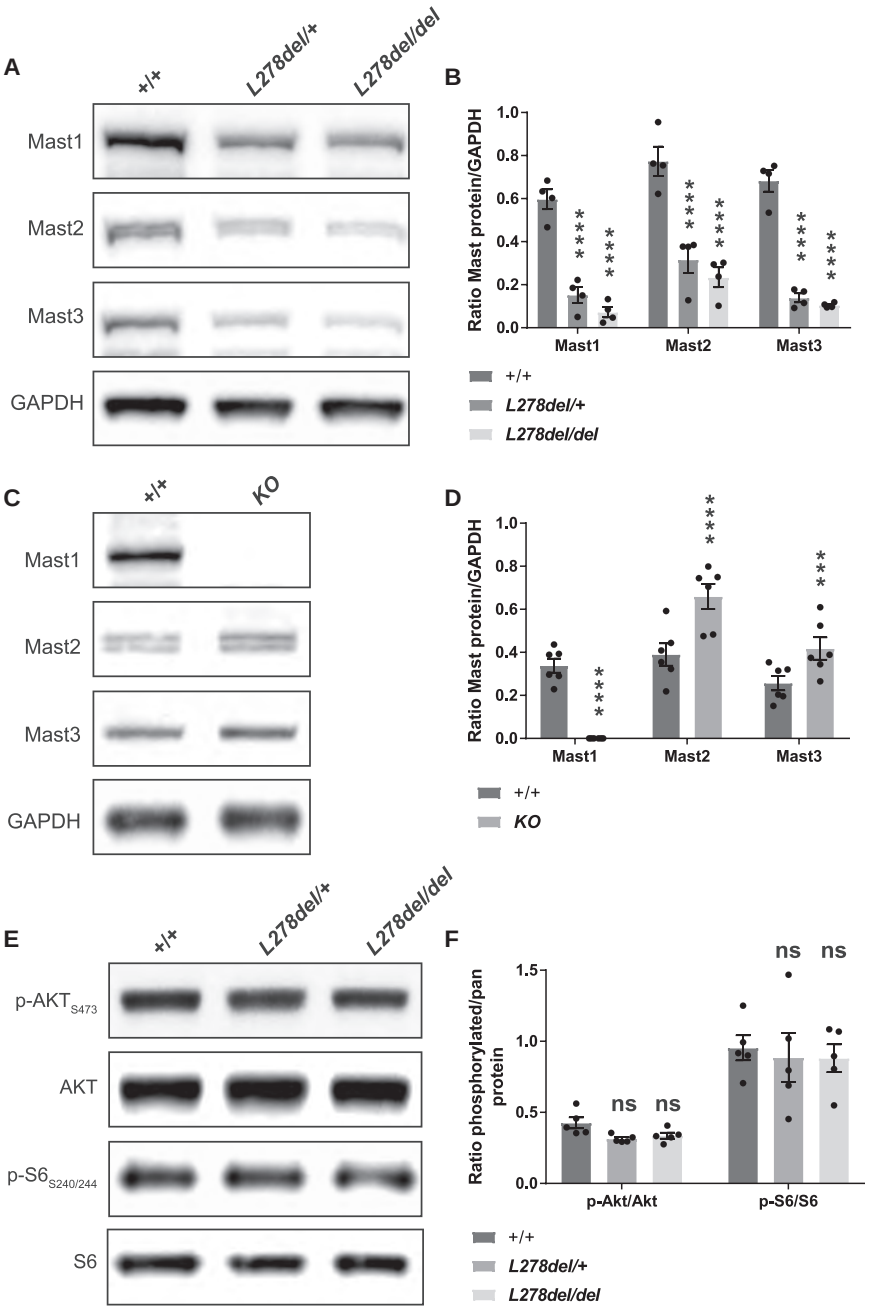


Figure 6. The L278del Mutation Influences Mast1/2/3 Protein Levels but Does Not Activate the PI3K/AKT3/mTOR Pathway

(A and B) Western blot analysis of Mast1, Mast2, and Mast3 on P0 cortical lysates from *L278del* animals. Quantification reveals a dramatic reduction of Mast1, Mast2, and Mast3 protein levels in *L278del* heterozygotes and homozygotes in comparison to littermate controls (n = 4 animals per genotype; two-way repeated-measures ANOVA with a Bonferroni correction). Mast1: +/+ versus *L278del/+* $p < 0.0001$; +/+ versus *L278del/del* $p < 0.0001$. Mast2: +/+ versus *L278del/+* $p < 0.0001$; +/+ versus *L278del/del* $p < 0.0001$. Mast3: +/+ versus *L278del/+* $p < 0.0001$; +/+ versus *L278del/del* $p < 0.0001$. See also Figures S6A–S6C.

(C and D) Western blot analysis of Mast1, Mast2, and Mast3 on brain lysates from *Mast1 KO* animals. While Mast1 is absent, there is a significant increase in the levels of Mast2 and Mast3 when compared to littermate controls (n = 6 animals per genotype; two-way repeated-measures ANOVA with a Bonferroni test for multiple comparison). Mast1: +/+ versus *KO* $p < 0.0001$. Mast2: +/+ versus *KO* $p < 0.0001$. Mast3: +/+ versus *KO* $p < 0.001$. See also Figures S6D–S6F.

(E and F) Levels of phosphorylated AKT and ribosomal S6 proteins, indicators of activation of PI3K/AKT/m-TOR pathway, are not significantly different in wild-type and *L278del* P0 cortex (n = 5 animals per genotype; two-way repeated-measures ANOVA with a Bonferroni test for multiple comparison). +/+ versus *L278del/+* AKT $p > 0.5$; +/+ versus *L278del/del* AKT $p > 0.5$; *L278del/+* versus *L278del/del* AKT $p > 0.5$; +/+ versus *L278del/+* rpS6 $p > 0.5$; +/+ versus *L278del/L278del* rpS6 $p > 0.5$; *L278del/+* versus *L278del/Ldel* rpS6 $p > 0.5$.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$. Error bars show mean ± the SEM.

Lys276del, and Leu278del microdeletions would disrupt this structure by changing the register of the α helices. It is therefore conceivable that these MAST1 microdeletions exert a dominant-negative effect by titrating out functional endogenous MAST proteins and/or targeting them for degradation. The consequence of the reduction in Mast1/2/3 levels is likely to be complex, influencing the phosphorylation of multiple targets. Little is known of the targets of the MAST family; however, Andrade and colleagues have recently shown that in mature dopaminergic neurons in the striatum, MAST3 phosphorylates ARPP-16 at Ser46, which in turn inhibits the serine-threonine phosphatase PP2A (Andrade et al., 2017). Mast2 has been reported to interact with β2-syntrophin, but its targets are otherwise undefined (Lumeng et al., 1999). Future experiments that define and validate the targets and binding partners of the MAST family in the developing brain will provide valuable insight on this front.

Why do our *L278del* mice have a thicker corpus callosum? We report that our *L278del* mice present with an enlarged corpus callosum that is most severe in the midbody and genu and manifests despite an increase in cortical apoptosis. Electron microscopic analysis has revealed that this phenotype is not attributable to an increase in the caliber of myelinated neurons, nor to the thickness of the myelin sheath itself, but is associated with an increase in the number of axons that project to the contralateral hemisphere. Such a phenotype may result from: (1) an increase in the branching of callosal fibers, (2) the formation of multiple axons originating from the soma of projection

neurons, or (3) a change in the fate of differentiating neurons during embryogenesis resulting in an increase in the number of callosal projection neurons at the expense of sub-cortical projection neurons (Baranek et al., 2012; Britanova et al., 2008).

Finally, we report the presence of *de novo MAST1* substitutions in patients with severe autism and microcephaly. Notably, a *de novo* P500L mutation in MAST1 has also been reported in a patient presenting with cerebral palsy (McMichael et al., 2015). This suggests that mutations in *MAST1* give rise to a spectrum of neurodevelopmental diseases. Unlike those mutations that we describe in patients with MCC-CH-CM, these variants are all substitutions, which might account for the different clinical phenotypes. At this stage we are not able to comment on the functional effect of these substitutions, or on the impact of the ~750 coding variants reported in the ExAC control genome database. In light of our work, it is conceivable that mutations in *MAST2* and *MAST3*, both of which are expressed in the developing and adult brain, might also result in neurological disease (Garland et al., 2008). In conclusion, we have defined the genetic lesion that causes MCC-CH-CM and revealed the importance of the MAST family in global brain development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
  - Human studies
  - Cell Lines, primary cultures, microbe strains
- METHOD DETAILS
  - Exome Sequencing
  - qPCR
  - Generation of CRISPR Mice
  - MRI Analysis
  - Electron Microscopy Studies
  - Generation of organoids
  - Nissl Staining and Immunohistochemistry
  - Quantification of Histological Data
  - Immunoblotting
  - Primary neuron culture
  - *In vitro* transcription and translation of Mast1 and microtubule-binding assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABLITITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and four tables and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.10.044.

ACKNOWLEDGMENTS

D.A.K. wishes to thank the Austrian Science Foundation (FWF, P21092 and I2681-B27). We wish to acknowledge the support of Boehringer Ingelheim, who fund basic science at the IMP. We wish to thank the CAUSES study, specifically Shelin Adam, Christele Du Souich, Jane Gillis, Alison Elliott, Anna Lehman, Jill Mwenifumbo, Tanya Nelson, Clara Van Karnebeek, and Jan Friedman. We would also thank Prof. Dr. Bwee Tien Poll and Dr. Saskia Maas for including patient P4 in the study and providing clinical information. T.M.P. is supported by the Diana and Steve Marienhoff Fashion Industries Guild Endowed Fellowship in Pediatric Neuromuscular Diseases. G.M. is generously supported by the National Institute of Neurological Disorders and Stroke (NINDS) (K08NS092898) and Jordan's Guardian Angels. We are indebted to Dr. Vissers for her help with patient P8. J.C.T. and A.T.P. are funded by Oxford NIHR Biomedical Research Centre. N.J.C. acknowledges the support of the NIH (R01GM097376). This work was supported by grants from the Swiss National Science Foundation (31003A, 182632) to A.R. G.M.C. and S.M.H. are grateful for support from the National Human Genome Research Institute (UM1HG007301). T.A.L. is funded by the Austrian Science Fund (FWF P28135), and W.B.D. is supported by NIH (1R01NS058721).

AUTHOR CONTRIBUTIONS

Conceptualization: R.T., D.A.K.; Methodology: R.T.; Validation: R.T., M.F.M.-R.; Formal Analysis: R.T., A.W.-W., I.L., M.C.S., L.L.; Visualization: R.T., I.L., M.C.S; Software: S.L.; Investigation: R.T., I.L., A.W.-W., T.v.D., J.W., B.W.v.B., M.C.S., T.G., M.B., G. Terrone, N.B.-B., A.R.P., A.T.P., G. Tian, G.V., E.D.G., N.B.-P., A.D., N.V., E.F., U.K., T.A.L., S.V., L.A., S.M.H., K.A.A., W.B.D., G.M., M.F.M.-R.; Writing, original draft: R.T., D.A.K; Writing, reviewing, and editing: all authors; Funding acquisition: W.B.D., J.C.T., A.R., J.A.B., L.B., G.M.C., T.M.P., F.B., T.A.L., J.C., N.J.C., D.A.K.; Supervision: J.C.T., A.R., J.A.B., L.B., G.M.C., T.M.P., F.B., J.C., N.J.C., D.A.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 11, 2017
Revised: May 3, 2018
Accepted: October 24, 2018
Published: November 15, 2018

REFERENCES

Andrade, E.C., Musante, V., Horiuchi, A., Matsuzaki, H., Brody, A.H., Wu, T., Greengard, P., Taylor, J.R., and Nairn, A.C. (2017). ARPP-16 Is a Striatal-Enriched Inhibitor of Protein Phosphatase 2A Regulated by Microtubule-Associated Serine/Threonine Kinase 3 (Mast 3 Kinase). J. Neurosci. *37*, 2709–2722.

Baranek, C., Dittrich, M., Parthasarathy, S., Bonnon, C.G., Britanova, O., Lanshakov, D., Boukhtouche, F., Sommer, J.E., Colmenares, C., Tarabykin, V., and Atanasoski, S. (2012). Protooncogene Ski cooperates with the chromatin-remodeling factor Satb2 in specifying callosal neurons. Proc. Natl. Acad. Sci. USA *109*, 3546–3551.

Braun, A., Breuss, M., Salzer, M.C., Flint, J., Cowan, N.J., and Keays, D.A. (2010). Tuba8 is expressed at low levels in the developing mouse and human brain. Am. J. Hum. Genet. *86*, 819–822, author reply 822–823.

Breuss, M., Morandell, J., Nimpf, S., Gstrein, T., Lauwers, M., Hochstoeger, T., Braun, A., Chan, K., Sánchez Guajardo, E.R., Zhang, L., et al. (2015). The expression of Tubb2b undergoes a developmental transition in murine cortical neurons. J. Comp. Neurol. *523*, 2161–2186.

Breuss, M., Fritz, T., Gstrein, T., Chan, K., Ushakova, L., Yu, N., Vonberg, F.W., Werner, B., Elling, U., and Keays, D.A. (2016). Mutations in the murine homologue of TUBB5 cause microcephaly by perturbing cell cycle progression and inducing p53-associated apoptosis. Development *143*, 1126–1133.

Britanova, O., de Juan Romero, C., Cheung, A., Kwan, K.Y., Schwark, M., Gyorgy, A., Vogel, T., Akopov, S., Mitkovski, M., Agoston, D., et al. (2008). Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. Neuron *57*, 378–392.

Broix, L., Jagline, H., Ivanova, E., Schmucker, S., Drouot, N., Clayton-Smith, J., Pagnamenta, A.T., Metcalfe, K.A., Isidor, B., Louvier, U.W., et al.;

Deciphering Developmental Disorders study (2016). Mutations in the HECT domain of NEDD4L lead to AKT-mTOR pathway deregulation and cause periventricular nodular heterotopia. Nat. Genet. *48*, 1349–1358.

Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). 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 (Austin) *6*, 80–92.

GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.

Cubelos, B., Briz, C.G., Esteban-Ortega, G.M., and Nieto, M. (2015). Cux1 and Cux2 selectively target basal and apical dendritic compartments of layer II-III cortical neurons. Dev. Neurobiol. *75*, 163–172.

Demyanenko, G.P., Tsai, A.Y., and Maness, P.F. (1999). Abnormalities in neuronal process extension, hippocampal development, and the ventricular system of L1 knockout mice. J. Neurosci. *19*, 4907–4920.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.

DiMario, F.J., Jr., Ramsby, G.R., and Burleson, J.A. (1999). Brain morphometric analysis in neurofibromatosis 1. Arch. Neurol. *56*, 1343–1346.

Edwards, T.J., Sherr, E.H., Barkovich, A.J., and Richards, L.J. (2014). Clinical, genetic and imaging findings identify new causes for corpus callosum development syndromes. Brain *137*, 1579–1613.

Ferland, R.J., Cherry, T.J., Preware, P.O., Morrisey, E.E., and Walsh, C.A. (2003). Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. J. Comp. Neurol. *460*, 266–279.

Garland, P., Quraishe, S., French, P., and O'Connor, V. (2008). Expression of the MAST family of serine/threonine kinases. Brain Res. *1195*, 12–19.

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

Göhlich-Ratmann, G., Baethmann, M., Lorenz, P., Gärtner, J., Goebel, H.H., Engelbrecht, V., Christen, H.J., Lenard, H.G., and Voit, T. (1998). Megalencephaly, mega corpus callosum, and complete lack of motor development: a previously undescribed syndrome. Am. J. Med. Genet. *79*, 161–167.

Guy, J., Ellis, E.A., Kelley, K., and Hope, G.M. (1989). Spectra of G ratio, myelin sheath thickness, and axon and fiber diameter in the guinea pig optic nerve. J. Comp. Neurol. *287*, 446–454.

Hengst, M., Tücke, J., Zerres, K., Blaum, M., and Häusler, M. (2010). Megalencephaly, mega corpus callosum, and complete lack of motor development: delineation of a rare syndrome. Am. J. Med. Genet. A. *152A*, 2360–2364.

Hyman, B.T., and Yuan, J. (2012). Apoptotic and non-apoptotic roles of caspases in neuronal physiology and pathophysiology. Nat. Rev. Neurosci. *13*, 395–406.

Iyer, V., Shen, B., Zhang, W., Hodgkins, A., Keane, T., Huang, X., and Skarnes, W.C. (2015). Off-target mutations are rare in Cas9-modified mice. Nat. Methods *12*, 479.

Johannessen, C.M., Reczek, E.E., James, M.F., Brems, H., Legius, E., and Cichowski, K. (2005). The NF1 tumor suppressor critically regulates TSC2 and mTOR. Proc. Natl. Acad. Sci. USA *102*, 8573–8578.

Keays, D.A., Tian, G., Poirier, K., Huang, G.J., Siebold, C., Cleak, J., Oliver, P.L., Fray, M., Harvey, R.J., Molnár, Z., et al. (2007). Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans. Cell *128*, 45–57.

Kenna, K.P., van Doormaal, P.T., Dekker, A.M., Ticozzi, N., Kenna, B.J., Diekstra, F.P., van Rheenen, W., van Eijk, K.R., Jones, A.R., Keagle, P., et al.; SLAGEN Consortium (2016). NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat. Genet. *48*, 1037–1042.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

Korogod, N., Petersen, C.C., and Knott, G.W. (2015). Ultrastructural analysis of adult mouse neocortex comparing aldehyde perfusion with cryo fixation. eLife *4*, https://doi.org/10.7554/eLife.05793.

Kosik, K.S., and Finch, E.A. (1987). MAP2 and tau segregate into dendritic and axonal domains after the elaboration of morphologically distinct neurites: an immunocytochemical study of cultured rat cerebrum. J. Neurosci. *7*, 3142–3153.

Lancaster, M.A., and Knoblich, J.A. (2014). Generation of cerebral organoids from human pluripotent stem cells. Nat. Protoc. *9*, 2329–2340.

Lancaster, M.A., Renner, M., Martin, C.A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. Nature *501*, 373–379.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Leroux, A.E., Schulze, J.O., and Biondi, R.M. (2018). AGC kinases, mechanisms of regulation and innovative drug development. Semin. Cancer Biol. *48*, 1–17.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Lumeng, C., Phelps, S., Crawford, G.E., Walden, P.D., Barald, K., and Chamberlain, J.S. (1999). Interactions between beta 2-syntrophin and a family of microtubule-associated serine/threonine kinases. Nat. Neurosci. *2*, 611–617.

Marsh, A.P., Heron, D., Edwards, T.J., Quartier, A., Galea, C., Nava, C., Rastetter, A., Moutard, M.L., Anderson, V., Bitoun, P., et al. (2017). Mutations in DCC cause isolated agenesis of the corpus callosum with incomplete penetrance. Nat. Genet. *49*, 511–514.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.

McMichael, G., Bainbridge, M.N., Haan, E., Corbett, M., Gardner, A., Thompson, S., van Bon, B.W., van Eyk, C.L., Broadbent, J., Reynolds, C., et al. (2015). Whole-exome sequencing points to considerable genetic heterogeneity of cerebral palsy. Mol. Psychiatry *20*, 176–182.

Meixner, A., Haverkamp, S., Wässle, H., Führer, S., Thalhammer, J., Kropf, N., Bittner, R.E., Lassmann, H., Wiche, G., and Propst, F. (2000). MAP1B is required for axon guidance and is involved in the development of the central and peripheral nervous system. J. Cell Biol. *151*, 1169–1178.

Morris, J.A., Royall, J.J., Bertagnolli, D., Boe, A.F., Burnell, J.J., Byrnes, E.J., Copeland, C., Desta, T., Fischer, S.R., Goldy, J., et al. (2010). Divergent and nonuniform gene expression patterns in mouse brain. Proc. Natl. Acad. Sci. USA *107*, 19049–19054.

Nguyen, L.S., Schneider, T., Rio, M., Moutton, S., Siquier-Pernet, K., Verny, F., Boddaert, N., Desguerre, I., Munich, A., Rosa, J.L., et al. (2016). A nonsense variant in HERC1 is associated with intellectual disability, megalencephaly, thick corpus callosum and cerebellar atrophy. Eur. J. Hum. Genet. *24*, 455–458.

Niquille, M., Garel, S., Mann, F., Hornung, J.P., Otsmane, B., Chevalley, S., Parras, C., Guillemot, F., Gaspar, P., Yanagawa, Y., and Lebrand, C. (2009). Transient neuronal populations are required to guide callosal axons: a role for semaphorin 3C. PLoS Biol. *7*, e1000230.

Parrini, E., Conti, V., Dobyns, W.B., and Guerrini, R. (2016). Genetic Basis of Brain Malformations. Mol. Syndromol. *7*, 220–233.

Paul, L.K., Brown, W.S., Adolphs, R., Tyszka, J.M., Richards, L.J., Mukherjee, P., and Sherr, E.H. (2007). Agenesis of the corpus callosum: genetic,

developmental and functional aspects of connectivity. Nat. Rev. Neurosci. *8*, 287–299.

Paxinos, G.H.G., Watson, C., Koutcherov, Y., and Wang, H. (2007). Atlas of the Developing Mouse Brain (Academic Press).

Pierson, T.M., Zimmerman, R.A., Tennekoon, G.I., and Bönnemann, C.G. (2008). Mega-corpus callosum, polymicrogyria, and psychomotor retardation: confirmation of a syndromic entity. Neuropediatrics *39*, 123–127.

Poduri, A., Evrony, G.D., Cai, X., Elhosary, P.C., Beroukhim, R., Lehtinen, M.K., Hills, L.B., Heinzen, E.L., Hill, A., Hill, R.S., et al. (2012). Somatic activation of AKT3 causes hemispheric developmental brain malformations. Neuron *74*, 41–48.

Rivière, J.B., Mirzaa, G.M., O'Roak, B.J., Beddaoui, M., Alcantara, D., Conway, R.L., St-Onge, J., Schwartzentruber, J.A., Gripp, K.W., Nikkel, S.M., et al.; Finding of Rare Disease Genes (FORGE) Canada Consortium (2012). De novo germline and postzygotic mutations in AKT3, PIK3R2 and PIK3CA cause a spectrum of related megalencephaly syndromes. Nat. Genet. *44*, 934–940.

Serafini, T., Colamarino, S.A., Leonardo, E.D., Wang, H., Beddington, R., Skarnes, W.C., and Tessier-Lavigne, M. (1996). Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. Cell *87*, 1001–1014.

Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. Hum. Mutat. *36*, 928–930.

Sturrock, R.R. (1980). Myelination of the mouse corpus callosum. Neuropathol. Appl. Neurobiol. *6*, 415–420.

Szybińska, A., and Leśniak, W. (2017). P53 Dysfunction in Neurodegenerative Diseases - The Cause or Effect of Pathological Changes? Aging Dis. *8*, 506–518.

Terrone, G., Voisin, N., Abdullah Alfaiz, A., Cappuccio, G., Vitiello, G., Guex, N., D'Amico, A., James Barkovich, A., Brunetti-Pierri, N., Del Giudice, E., and Reymond, A. (2016). De novo PIK3R2 variant causes polymicrogyria, corpus callosum hyperplasia and focal cortical dysplasia. Eur. J. Hum. Genet. *24*, 1359–1362.

Thiel, C., Kessler, K., Giessl, A., Dimmler, A., Shalev, S.A., von der Haar, S., Zenker, M., Zahnleiter, D., Stöss, H., Beinder, E., et al. (2011). NEK1 mutations cause short-rib polydactyly syndrome type majewski. Am. J. Hum. Genet. *88*, 106–114.

Tischfield, M.A., Baris, H.N., Wu, C., Rudolph, G., Van Maldergem, L., He, W., Chan, W.M., Andrews, C., Demer, J.L., Robertson, R.L., et al. (2010). Human TUBB3 mutations perturb microtubule dynamics, kinesin interactions, and axon guidance. Cell *140*, 74–87.

Walden, P.D., and Cowan, N.J. (1993). A novel 205-kilodalton testis-specific serine/threonine protein kinase associated with microtubules of the spermatid manchette. Mol. Cell. Biol. *13*, 7625–7635.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell *153*, 910–918.

West, K.L., Kelm, N.D., Carson, R.P., and Does, M.D. (2015). Quantitative analysis of mouse corpus callosum from electron microscopy images. Data Brief *5*, 124–128.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-Mast1	Santa Cruz	Cat#sc-55851; RRID:AB_2141214
anti-Mast1 (G-4)	Santa Cruz	Cat#sc-373845; RRID:AB_10988599
anti-Tbr2	Abcam	Cat#ab23345; RRID:AB_778267
anti-Pax6	Covance	Cat#PRB-278P; RRID:AB_291612
anti-Tuj1	Covance	Cat#MMS-435P; RRID:AB_2313773
anti-Cux1	Santa Cruz	Cat#sc-6327; RRID:AB_2087003
anti-NeuN (A60)	Millipore	Cat#MAB377; RRID:AB_2298772
anti-Er81	Abcam	Cat#ab36788; RRID:AB_732196
anti-Foxp2	Abcam	Cat#ab16046; RRID:AB_2107107
anti-Olig2	Millipore	Cat#AB9610; RRID:AB_570666
anti-GFAP	Dako	Cat#; Z0334; RRID: AB_2314535
anti-Map2	Abcam	Cat#AB24640; RRID:AB_448205
anti-Tau	Abcam	Cat#AB64193; RRID:AB_1143333
anti-cleaved-Caspase-3	Cell Signaling Technology	Cat#9661; RRID:AB_2341188
anti-cleaved-Caspase-9	Cell Signaling Technology	Cat#9509; RRID:AB_2073476
anti-p53	Leica	Cat#P53-CM5P; RRID:AB_2744683
anti-Calbindin	Millipore	Cat#AB1778; RRID:AB_2068336
anti-Mast1	Proteintech	Cat#13305-1-AP; RRID:AB_10639043
anti-Mast2	Santa Cruz	Cat#sc-377198; RRID:AB_2744682
anti-Mast3	Novus	Cat#NBP1-82993; RRID:AB_11017204
anti-p-Akt(Ser473)	Cell Signaling Technology	Cat#4060; RRID:AB_2315049
anti-Akt	Cell Signaling Technology	Cat#9272; RRID:AB_329827
anti-p-S6(Ser240/244)	Cell Signaling Technology	Cat#2215; RRID:AB_2630325
anti-S6	Cell Signaling Technology	Cat#5548; RRID:AB_10707322
anti-GAPDH	Millipore	Cat#MAB374; RRID:AB_2107445
Bacterial and Virus Strains		
E. coli DH5alpha	Invitrogen	Cat#18265017
Experimental Models: Cell Lines		
H9 feeder free hES cells	WiCell, University of Wisconsin	WA09
Experimental Models: Organisms/Strains		
Leu278del	this paper	N/A
Mast1 knockout line5	this paper	N/A
Mast1 knockout line31	this paper	N/A
Tubb5 flx	In house	MGI Cat# 6119480, RRID:MGI:6119480
Nestin Cre	In house	IMSR Cat# JAX:019103, RRID:IMSR_JAX:019103)
Oligonucleotides		
GCAAGGTGTACAGCAGTATGG	sigma	mMAST1_qPCR_F
TGGGTCCCGCTTGCTG	sigma	mMAST1_qPCR_R
GGTGCATCTGGAGGAACAG	sigma	hMAST1_qPCR _F
GATGGTATCGAAGTCATTCTCCC	sigma	hMAST1_qPCR _R
AAGGTTCAGCAAGGTTTACAGTAG	sigma	mMAST2_qPCR_F
AAGCCGTCTGAGTGATCTTC	sigma	mMAST2_qPCR_R

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AAGATGGTCACTTGCTTCC	sigma	mMAST3_qPCR_F
AGGACGAGCTTGACGATAC	sigma	mMAST3_qPCR_R
TAATCATCTCTCGCCCTGCA	sigma	mMast1_ex8_gRNA2
GGAGAGGGTCTTACTTGCTT	sigma	mMast1_ex8_mutscrn_F
TCTTCTGGGTTGAATTCCTA	sigma	mMast1_ex8_mutscrn_R
Recombinant DNA		
pcDNA3.1-5′UTRmMast1-E194del	this paper	N/A
pcDNA3.1-5′UTRmMast1-K276del	this paper	N/A
pcDNA3.1-5′UTRmMast1-L278del	this paper	N/A
pcDNA3.1-5′UTRmMast1-G519S	this paper	N/A
Software and Algorithms		
ImageJ	open souce (at NIH)	N/A
GraphPad prism v.7	GraphPad Software	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. David Keays (david.keays@imp.ac.at).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

Mice were maintained on a 12:12 light dark cycle and food was available *ad libitum*. Mice of ages E10.5, E12.5, E14.5, E16.5, P0, P6, 8-10 weeks were used in this study. Animals were not subject to randomization and allocated to experimental groups according to their genotypes as littermates regardless of the sex. For experiments with mice of ages E10.5, E12.5, E14.5, E16.5, P0, P6 the sex was not determined. For generating CRISPR mouse lines, female BL6/CBA F1 mice were used, whereas all backcrosses were performed using C57/BL6 mice to the 7th generation. All procedures were carried out according to legal requirements and covered by an approved license (M58/006093/2011/14).

Human studies

Informed consent was obtained from all patients included in this study. Details of each patient and the condition is provided in Tables 1 and S1. This study was conducted within approved ethical frameworks University of Oxford (08/MRE09/55); Stanford University Institutional Review Board (28362); University of Alabama at Birmingham (X130201001); and at the Seattle Children's Research Institute (IRB protocol #13291); and by the Commissie Mensgebonden Onderzoek Regio Arnhem-Nijmegen (NL36191.091.11).

Cell Lines, primary cultures, microbe strains

For generating cerebral organoids, feeder free H9 human embryonic stem cells (ESCs) were obtained from Wicell (WA09), having a normal karyotype, female sex, and no contaminants. These cells were maintained using the commercially available mTESR media (Stem cell technologies #05850) at 37°C with 5% CO_2 levels. These cells were not authenticated. Primary cell cultures were generated from P0 and P7 cortex or cerebellum. The sex of these pups was not determined. The primary neurons were cultured in Neurobasal media with B27, Penstrep and L-glutamine supplements at 37°C with 5% CO_2 levels. The *Escherichia coli* DH5alpha strain was used for cloning purposes.

METHOD DETAILS

Exome Sequencing

Patients with structural brain phenotypes and their parents were recruited in accordance with existing ethical frameworks and internal review boards. DNA from blood was extracted using standard methods and subject to exome sequencing. We exploited capture arrays (such as the Agilent Human All exon 50Mb array) to capture exonic genomic DNA, which was then subject to next generation paired end sequencing (Illumina). Bioinformatic filtering employed platforms such as the Genome Analysis Toolkit (GATK), and excluded variants with a population frequency greater than 1%. For this purpose, we relied on available genomic databases including dbSNP and the 1000 Genomes server. Sanger sequencing was employed to confirm any putative *de novo* variants. This study was

conducted within an approved ethical framework (08/MRE09/55). For mouse exome sequencing genomic DNA was extracted from the tails of BL6J, CBA, and L278del mice and then subjected to Illumina library preparation and captured on the Agilent SureSelect mouse (51Mb) array. The samples were sequenced on an Illumina HiSeq2500 with 100-125 nucleotides paired-end reads followed by data quality analysis using FASTQC. Reads were mapped to the mouse genome (GRCm38, mm10) using BWA (Li and Durbin, 2010). The BAM files were used as an input to UnifiedGenotyper module from the GATK-lite toolkit (v2.3). (McKenna et al., 2010) Variants were called and dbSNP variants were used as known sites, according to GATK Best Practices recommendations (DePristo et al., 2011). Finally, Variants were annotated by SnpEff v4.1 (Cingolani et al., 2012). Sanger sequencing was employed to confirm putative *de novo* variants on chromosome 8, potentially linked to the L278del mutation (Table S2).

qPCR

The following tissues were dissected from C57/BL6 mice: embryonic brains at stages E10.5, E12.5, E14.5, E16.5, E18.5 (undetermined sex); brains from early postnatal mice at P0, P6 (undetermined sex); the brain regions (cortex, cerebellum, hippocampus, striatum, midbrain, colliculi, hypothalmus, brain stem, spinal cord, olfactory bulbs) and organs (liver, spleen, heart, muscle, testis, kidney, lung) from male adult littermates (n = 3). To ascertain Mast1, Mast2, and Mast3 mRNA levels in L278del and KO animals, cortices from littermate homozygote, heterozygote, and wild-type P0s were obtained (undetermined sex) (n = 4). Tissue samples were snap frozen before total RNA extraction and cDNA synthesis (SuperScript III First-Strand Synthesis System, Invitrogen, 18080-051). We used SYBR green on a Bio-Rad Cycler together with intro spanning primers to amplify murine Mast1 (mMAST1_qPCR_F/GCAAGGTGTAC AGCAGTATGG, mMAST1_qPCR _R/TGGGTCCCGCTTGCTG). For qPCR on human fetal tissue (hMAST1_qPCR _F/GGTGCATCT GGAGGAACAG, hMAST1_qPCR_R/GATGGTATCGAAGTCATTCTCCC) we performed technical triplicates on two different cDNA libraries (Biochain; C1244051, C1244035) again using intro spanning primers. In addition, we amplified three control genes (Pgk1, Tfrc, and Hprt). For each control gene technical triplicates were performed, and the geometric mean of the Ct values for the three control genes calculated. This geometric mean was then used to obtain the delta Ct value by subtracting it from the average (of triplicate runs) Ct value for Mast1. The relative mRNA levels were then obtained by taking into account the efficiency of the qPCR primers (see Braun et al., 2010).

Generation of CRISPR Mice

Female BL6/CBA F1 mice (3-4 weeks old) were superovulated according to standard protocols. Zygotes were isolated from donor females at the day of the coagulation plug (= E0.5). Surrounding cumulus cells were removed by incubation of the cumulus-zygote-complexes in hyaluronidase solution (at ~0.3 mg/mL). The injection mix (50 ng/μL Cas9 mRNA + 50 ng/μL guide RNA + 200 ng/μL L278del-ssODN repair template) was injected into the cytoplasm of zygotes. Injected zygotes were incubated for at least 15 min in a CO_2 incubator at 37°C. Zygotes that survived microinjection were transferred the same day into the oviducts of pseudopregnant recipient females (= 0.5 days post coitum). Resultant pups were genotyped by PCR amplification of chromosome 8 followed by sequencing using primers GGAGAGGGTCTTACTTGCTT and TCTTCTGGGTTGAATTCCTA, and backcrossed to C57/BL6.

MRI Analysis

MRI analysis was performed as described previously (Breuss et al., 2016). Briefly, 8-week old littermates were perfused with 0.9% NaCl and 4% PFA supplemented with 10% ProHance Solution (Bracco Imaging Group, 4002750). Images were then acquired with a 15.2 T Biospec horizontal bore scanner (Bruker BioSpin, Ettlingen, Germany). Brain regions of interest were manually segmented, blinded to the genotype, using Amira 5.6 (Visualization Science Group), relying on a mouse brain atlas, and the volumes obtained.

Electron Microscopy Studies

Mice aged 8-9 weeks (n = 5) were sacrificed and brain tissue prepared for electron microscopy (Korogod et al., 2015). Briefly, animals were perfused employing a constant flow of 7 mL/min with 2.5% glutaraldehyde and 2% paraformaldehyde in phosphate buffer (0.1 M, pH 7.4). Following dissection of the brain, 80-μm thick coronal slices were prepared on a vibratome (Leica VT 1000S). Matching slices were selected from the medial region of each brain (bregma −0.10), the corpus callosum microdissected, and then post fixed for 40 min in 1% osmium tetroxide in cacodylate buffer (0.1 M, pH 7.4). Samples were then washed twice for 5 min in ddH20 and stained with 1% uranyl acetate for 40 min before being dehydrated in a graded acetone series and embedded in Epon resin. Ultrathin sections were prepared (70nm), mounted on 50 mesh grids and post-stained with uranyl acetate (10 min, 2% in water) and with Reynolds' lead citrate (5 min). To assess differences in myelin thickness and axon, internal diameter images were captured at 8900x using a FEI Morgagni 268D (FEI Company) 100 kV transmission electron microscope operating at 80 kV. A customized script written in Definiens Architect XD (Version: 2.7.0; Build 60765X64) was employed to ascertain axonal surface area and myelin thickness determined using ImageJ. To ascertain the number of axons crossing the midline, images were acquired at 710x along the dorsa-ventral axis. These images (approx 10-12) per animal were then tiled employing iTEM, resulting in a single image. A box measuring 30 μm along the x axis was then overlaid onto the structure, and the total number of myelinated axons counted manually. Three stitched images were prepared per animal and all analysis was performed blind to genotype. To determine the G-ratio of axons crossing the midline we measured the total diameter of fibers (> 300 per animal) and the axonal diameter (without myelin) on high resolution EM images. The G-ratio was then calculated (axon diameter/fiber diameter) (Guy et al., 1989).

Generation of organoids

Human cerebral organoids were generated from H9 human ESCs (Wicell, WA09) as previously described (Lancaster and Knoblich, 2014)_ENREF_14. 64 days-old organoids were then embedded in Neg-50 Medium (Richard-Allan Scientific) and sliced into 20 μm sections in a cryostat. The slides were dried overnight before storing at −20°C and stained as described below.

Nissl Staining and Immunohistochemistry

Brains were removed at E12.5, E14.5, E16.5, E18.5, P0 and drop-fixed in 4% PFA overnight followed by dehydration in 30% sucrose. Brains were embedded in Neg-50 Medium (Richard-Allan Scientific) and sliced into 12 μm sections in a cryostat. Adults brains were recovered from the animals after perfusion with 0.9% NaCl and 4% PFA. These brains were post-fixed, dehydrated, and sliced using a sledge microtome into 40 μm sections. For Nissl staining, sections were washed in PBS, and bathed in Cresyl Violet for 3 min [0.25% Cresyl Violet acetate (Sigma, C5042) dissolved in distilled water with ten drops of glacial acetic acid per 100 mL of solution]. After briefly washing in PBS, slides were dehydrated in an alcohol series (30%, 70%, 96% and 100% ethanol, 2 min each), xylol (twice, 2 min each), mounted with DPX mountant (Fluka, 44581) and left overnight at room temperature. Images were acquired using the transmitted Mirax slide scanner (Zeiss).

Immunofluorescence experiments were performed as previously described (Breuss et al., 2016; Breuss et al., 2015). Where necessary antigen retrieval was performed on sections by slow heating of the slides in antigen retrieval buffer (Vector, H-3301) up to 90°C, followed by gradual cooling at room temperature for 20 min. Primary antibody in 0.1%–0.3% Triton X-100/PBS with 4% donkey serum as the blocking agent, was incubated overnight at 4°C. The following concentrations of primary antibodies were used: 1:300 Mast1 (Santa Cruz, sc-55851), 1:100 Mast1 (Santa Cruz, sc-373845) 1:300 cleaved-Caspase-3 (Cell Signaling, #9661), 1:300 cleaved-Caspase-9 (Cell Signaling #9509), 1:300 Pax6 (Covance, PRB-278P), 1:300 Tbr2 (Abcam, ab23345), 1:1000 Tuj (Covance, MMS-435P), 1:300 NeuN (Millipore, MAB377), 1:100 Cux1 (Santa Cruz, sc-6327), 1:1000 Er81 (Abcam, ab36788), 1:300 FoxP2 (Abcam, ab16046), 1:250 Calbindin (Millipore, AB1778), 1:1000 Olig2 (Millipore, AB9610), GFAP 1:500 (Dako, Z0334), 1:300 Map2 (Abcam, ab24640), 1:300 Tau (Abcam, ab61493), 1:1000 p53 (Leica, P53-CM5P). The next day, sections were washed in PBS and a species-specific secondary antibody (Molecular Probes, A-10037, A-11057, A-10042, A-21206; 1:500) was applied for 1 h in blocking solution at 4°C, followed by Hoechst 33342 staining (1:2000 in PBS) for 5 min. Subsequently the slides were mounted with Fluorescent Mounting Medium (Dako, S302380) and dried overnight at 4°C before imaging. Images were taken using a laser scanning confocal microscope (LSM 780 Zeiss).

Quantification of Histological Data

To determine the thickness of the corpus callosum in Nissl-stained sections the septum in the genu region was measured three times and an average determined (bregma co-ordinates 0.86mm). To determine the cortical thickness six measurements were obtained for rostral (bregma co-ordinates 0.86 mm) and caudal regions (bregma co-ordinates −1.82mm) and averaged for each animal. To assess the thickness of various cortical layers, matched caudal immunostained (Foxp2, Cux1, Er81) sections were measured four times in the region of the somatosensory cortex and averaged for each animal. To assess the levels of apoptosis in the cortex of P0 animals caspase-3 and caspase-9-positive puncta were counted in 3 matched sections (spanning anatomical co-ordinates 3.75mm to 4.0mm [Paxinos et al., 2007]), and the number averaged per animal. To compare rostral and caudal cortical apoptosis at P0 caspase-3 positive punctae were counted in a box measuring 1000 × 300 μm^2, in rostral sections (n = 2 sections per animal, co-ordinates 2.3-2.5mm [Paxinos et al., 2007]) and caudal sections (n = 2 sections per animal, co-ordinates: 3.75-4.0mm [Paxinos et al., 2007]). To assess levels of apoptosis in the P0 cerebellum, caspase-3-positive positive puncta were counted in 2 matched sections per animal (spanning anatomical co-ordinates 7mm to 7.25mm [Paxinos et al., 2007]) and the number averaged per section. To quantify the number and density of oligodendrocytes and astrocytes within the corpus callosum coronal sections from 8 week old animals were stained with Olig2 and GFAP. A box measuring 770 μm wide was placed over the septum of the corpus callosum extending from the ventral to dorsal edges, and positively stained cells were manually counted. To quantitate the number of Purkinje and granule cells in the cerebellum, free-floating, 40 μm, sagittal midline cerebellar sections were stained with anti-sera for Calbindin and NeuN, followed by counterstaining with DAPI (n = 2 sections per animal). Folium III was imaged and all Purkinje cells in this folium were counted allowing determination of the Purkinje cell linear density (cells/mm of Purkinje cell layer). The total number of Purkinje cells per section was then estimated by measuring the entire length of the Purkinje cell layer. To estimate the number of granule cells per section 4 images per animal were captured within the granule cell layer of folium III, and the density of granule cells/mm^2 of granule layer (GL) was determined. The total number of cells per section was then estimated by measuring the entire surface area of the GL and mutilpying it by the average density for that animal. All quantifications were done using ImageJ software and performed blinded to the genotype.

Immunoblotting

Brains were snap frozen in liquid nitrogen and stored at −80°C. Protein lysates were prepared from frozen tissue samples, homogenized in chilled lysis buffer (20 mM TrisHCl pH 7.5, 100 mM NaCl, 10% glycerol, 1% Triton X-100) supplemented with protease and phosphatase inhibitors (Pierce #88668). Using a Tungsten Carbide Bead the tissue was lysed in a QIAGEN Tissue Lyser (2 × 1 min, 20Hz), incubated at 4°C for 1 hr, and then centrifuged twice for 20 min. Protein concentration of the collected supernatant was measured using the Pierce BCA Protein Assay Kit (#23225). Protein lysates were run on NuPAGE protein gels, blotted onto

nitrocellulose membranes and blocked in 5% skimmed milk/TBST. Primary antibodies [1:2000 GAPDH (Millipore, MAB374); 1:1000 Mast1 (Proteintech, 13305-1-AP); 1:300 Mast2 (Santa Cruz, sc-377198); 1:500 Mast3 (Novus; NBP1-82993); 1:1000 pan-Akt (Cell Signaling Technology #9272); 1:1000 Phospho-Akt (Ser473) (Cell Signaling Technology #4060); 1:500 S6 Ribosomal Protein (Cell Signaling Technology #5548); 1:1000 Phospho-S6 Ribosomal Protein (Ser240/244) (Cell Signaling Technology #2215)] were incubated overnight in blocking solution at 4°C. The following day, the blot was incubated with the corresponding HRP conjugated secondary antibody (abcam) for 1 hr at room temperature. The signal was detected on Amersham Hyperfilm ECL film using the ECL Western Blotting Detection Reagents kit (GE healthcare, RPN2106). Western blots were quantified using ImageJ.

Primary neuron culture

Cortical neurons were prepared from brains of C57/BL6 or *Mast1 KO* P0 mice while P7 mice were used for cerebellar neuronal cultures. In short, the brain region of interest was isolated, and the meninges carefully removed. The tissue was treated with 0.25% trypsin solution in a 37°C water bath for 15 mins, followed by three washes in HBSS (20 mM HEPES, 2 mM CaCl2, 5.4 mM KCl, 1 mM MgCl2, 1 mM NaPi, 5.6 mM glucose, pH 7.3). The trypsinised tissue was titurated in plating media (DMEM, 10% Horse serum, 200 mM L-glutamin, 100 mM sodium pyruvate), first using a P1000 pipette and then a P200 pipette. The cells were passed through a 70 μm strainer and the resulting cell suspension counted. Neurons plated at $\sim 1 \times 10^5$ cells per well in a 24-well plate containing poly-L-lysine coated coverslips. After 3-4 hr the medium was changed to Neurobasal containing B27, L-glutamine and PenStrep supplements. After 4-5 days in culture the coverslips were fixed in 4% PFA and stained with anti-Map2, anti-Tau and anti-Mast1 antibodies.

In vitro transcription and translation of Mast1 and microtubule-binding assay

These experiments were performed as previously described (Keays et al., 2007; Walden and Cowan, 1993). Briefly, the full-length murine Mast1 cDNA was cloned into the pcDNA 3.1+ vector (Invitrogen) driven by the T7 promoter. The different patient mutations screened were introduced by site-directed mutagenesis using the QuikChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent #210518) kit. Plasmids (20 ng/μL final concentration) were used to drive expression in rabbit reticulocyte lysate (TnT T7 Coupled Reticulocyte Lysate System, Promega) containing 35S-methionine (specific activity: 1000 Ci/mMol; Perkin Elmer). Reactions were incubated at 30°C for 90 min. Prior to co-polymerization reactions, the TnT reaction product was cleared by centrifugation at 60,000 g for 10 min at 4°C. MAP rich porcine brain tubulin (Cytoskeleton #ML116) was depolymerized in tubulin buffer (0.1 M PIPES buffer pH 6.9, 1 mM MgCl2 and 0.5 mM EGTA) by incubation on ice. Co-polymerization reactions were set up by adding the 35S-labeled Mast1 to MAP rich tubulin (1mg/ml), in tubulin buffer containing 1 mM GTP, 2 mM AMP-PNP and 20 μM taxol. To investigate Mast1 binding to microtubules in either the presence or absence of MAPs, the labeled Mast1 reaction was distributed equally between MAP rich and 99% pure porcine brain tubulin (Cytoskeleton #T240). All co-polymerization reactions were incubated at 30°C for 30 min to promote microtubule polymerization and binding of associated proteins. Reaction products were then loaded onto cushions (0.20 mL) containing 1 M sucrose in tubulin buffer, 0.5 mM GTP, 1 mM AMP-PNP and 10 μM taxol and centrifuged at 60,000 g for 15 min at 30°C. All centrifugations were done in a Beckman TL-100.3 rotor. Proteins in the pellet were solubilized in SDS loading dye and separated in a 7% SDS-PAGE. Radioactivity content of the microtubule pellets obtained was determined following an overnight exposure to a high-sensitive film (KODAK BioMax MR film).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analysis was executed using GraphPad Prism software package (v7.0c). Numbers (n) of animals or replica for experiments are indicated in the text and the statistical test used indicated in individual figure legends. Statistical tests used in this study include the Student's t test, one-way ANOVA with Dunnett's or Tukey's multiple comparisons test, and two-way repeated-measures ANOVA with Bonferroni's multiple comparisons test. Samples and animals were not subject to randomization, but were assigned to experimental groups based on their genotype. Shapiro-Wilk or Kolmogorov-Smirnov tests were applied to assess whether assumptions of normality were met, and corrected for multiple comparisons were necessary. A summary of the statistical tests applied for each experiment, the n numbers, and the results obtained are shown in Table S4.

DATA AND SOFTWARE AVAILABLITITY

In accordance with ethical frameworks and our commitment to patient confidentiality we have not deposited the primary exome sequencing data generated in this study in a community repository.

KIAA1109 Variants Are Associated with a Severe Disorder of Brain Development and Arthrogryposis

Lucie Gueneau,[1,27] Richard J. Fish,[2,27] Hanan E. Shamseldin,[3,27] Norine Voisin,[1,27] Frédéric Tran Mau-Them,[4,5,27] Egle Preiksaitiene,[6,27] Glen R. Monroe,[7,27] Angeline Lai,[8,9] Audrey Putoux,[10,11] Fabienne Allias,[12] Qamariya Ambusaidi,[13] Laima Ambrozaityte,[6] Loreta Cimbalistienė,[6] Julien Delafontaine,[14] Nicolas Guex,[14] Mais Hashem,[3] Wesam Kurdi,[13] Saumya Shekhar Jamuar,[8,15] Lim J. Ying,[8] Carine Bonnard,[16] Tommaso Pippucci,[17] Sylvain Pradervand,[1,14] Bernd Roechert,[14] Peter M. van Hasselt,[7] Michaël Wiederkehr,[1] Caroline F. Wright,[18] DDD Study, Ioannis Xenarios,[1,14] Gijs van Haaften,[7] Charles Shaw-Smith,[19] Erica M. Schindewolf,[20] Marguerite Neerman-Arbez,[2] Damien Sanlaville,[10,11] Gaëtan Lesca,[10,11] Laurent Guibaud,[11,21] Bruno Reversade,[16,22,23,24] Jamel Chelly,[4,5] Vaidutis Kučinskas,[6] Fowzan S. Alkuraya,[3,25,26,28,*] and Alexandre Reymond[1,28,*]

Whole-exome and targeted sequencing of 13 individuals from 10 unrelated families with overlapping clinical manifestations identified loss-of-function and missense variants in *KIAA1109* allowing delineation of an autosomal-recessive multi-system syndrome, which we suggest to name Alkuraya-Kučinskas syndrome (MIM 617822). Shared phenotypic features representing the cardinal characteristics of this syndrome combine brain atrophy with clubfoot and arthrogryposis. Affected individuals present with cerebral parenchymal underdevelopment, ranging from major cerebral parenchymal thinning with lissencephalic aspect to moderate parenchymal rarefaction, severe to mild ventriculomegaly, cerebellar hypoplasia with brainstem dysgenesis, and cardiac and ophthalmologic anomalies, such as microphthalmia and cataract. Severe loss-of-function cases were incompatible with life, whereas those individuals with milder missense variants presented with severe global developmental delay, syndactyly of 2nd and 3rd toes, and severe muscle hypotonia resulting in incapacity to stand without support. Consistent with a causative role for *KIAA1109* loss-of-function/hypomorphic variants in this syndrome, knockdowns of the zebrafish orthologous gene resulted in embryos with hydrocephaly and abnormally curved notochords and overall body shape, whereas published knockouts of the fruit fly and mouse orthologous genes resulted in lethality or severe neurological defects reminiscent of the probands' features.

Introduction

The advent of high-throughput sequencing led to the delineation of multiple syndromes. Neurological genetic diseases are the main class of these Mendelian disorders[1] with, for example, approximately 700 different genes confidently associated with intellectual disability (ID) and developmental delay notwithstanding that about 50% of yet unexplained ID-affected case subjects are predicted to have a genetic basis in genes remaining to be discovered.[2,3] Neurodevelopmental disorders characterized by brain malformations represent an important group among these unexplained conditions and are likely associated with mutations in genes implicated in cortical or

[1]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland; [2]Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland; [3]Department of Genetics, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia; [4]Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM Unité 964, 67404 Illkirch Cedex, France; [5]Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France; [6]Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, 08661 Vilnius, Lithuania; [7]Department of Genetics and Center for Molecular Medicine, University Medical Center Utrecht, 3584 CX Utrecht, the Netherlands; [8]KK Women's and Children's Hospital, Singapore 229899, Singapore; [9]Lee Kong Chian School of Medicine, Nanyang Technological University-Imperial College London, Singapore 639798, Singapore; [10]Service de Génétique, Hospices Civils de Lyon, 69002 Lyon, France; [11]Centre de Recherche en Neurosciences de Lyon, INSERM U1028, UMR CNRS 5292, Université Claude Bernard Lyon 1, 69675 Bron Cedex, France; [12]Département de Pathologie, Hospices Civils de Lyon, 69002 Lyon, France; [13]Department of Obstetrics and Gynecology, King Faisal Specialist Hospital and Research Center, Riyadh 11211, Saudi Arabia; [14]Swiss Institute of Bioinformatics (SIB), 1015 Lausanne, Switzerland; [15]Duke-NUS Medical School, Singapore 169857, Singapore; [16]Institute of Medical Biology, A*STAR, Singapore 138648, Singapore; [17]Sant'Orsola-Malpighi Hospital, Medical Genetics Unit, Pavillon 11, 2nd floor, Via Massarenti 9, 40138 Bologna, Italy; [18]Deciphering Developmental Disorders (DDD) Study, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; [19]Department of Clinical Genetics, Royal Devon and Exeter NHS Foundation Trust, Exeter EX1 2ED, UK; [20]Center for Fetal Diagnosis and Treatment, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [21]Département d'imagerie pédiatrique et fœtale, Centre Pluridisciplinaire de Diagnostic Prénatal, Hôpital Femme Mère Enfant, Université Claude Bernard Lyon 1, 69677 Bron Cedex, France; [22]Institute of Molecular and Cell Biology (IMCB), A*STAR (Agency for Science, Technology and Research), 61 Biopolis Drive, Singapore 138673, Singapore; [23]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore; [24]Amsterdam Reproduction & Development, Academic Medical Centre & VU University Medical Center, 1105 AZ Amsterdam, the Netherlands; [25]Department of Anatomy and Cell Biology, College of Medicine, Alfaisal University, Riyadh 11533, Saudi Arabia; [26]Saudi Human Genome Program, King Abdulaziz City for Science and Technology, Riyadh 12371, Saudi Arabia
[27]These authors contributed equally to this work
[28]These authors contributed equally to this work
*Correspondence: falkuraya@kfshrc.edu.sa (F.S.A.), alexandre.reymond@unil.ch (A.R.)
https://doi.org/10.1016/j.ajhg.2017.12.002.

cerebellar development. They can be classified into four main categories depending on the origin of the defect.[4] First are disorders due to abnormal proliferation of neuronal and glial cells including brain under-growth (microcephaly) or overgrowth (megalencephaly). Second are neuronal migration disorders that include: (1) lissencephaly, i.e., the absence or decrease of gyration responsible for a smooth brain; (2) cobblestone cortical malformations; and (3) neuronal heterotopia, i.e., the abnormal localization of a neuronal population. Third are pathologies characterized by malformations caused by postmigrational abnormal cortical organization, mainly polymicrogyria, i.e., the increase of small gyration. The last category regroups malformations of the mid-hindbrain with early anteroposterior and dorsoventral patterning defects. This phenotypic heterogeneity is paralleled by molecular heterogeneity as more than 100 genes have been implicated to date.[4–7] The causative genes can be arranged into specific biological pathways (for instance synapse structure, cellular growth regulation, apoptosis, cell-fate specification, actin cytoskeleton, and microtubule assembly) that do not necessarily correlate with the type of malformations described above, as emerging evidence suggests that brain disorders are far more heterogeneous than the classification suggests.[4,8]

We report, through description of 19 affected individuals, an autosomal-recessive brain malformation disorder with arthrogryposis caused by variants within *KIAA1109* (MIM: 611565).

Material and Methods

Enrollment

Families were recruited in Lithuania, the United Kingdom, France, Saudi Arabia, the USA, and Singapore. The institutional review boards of the Vilnius University Faculty of Medicine, NHS Foundation Trust, Hôpitaux Universitaires de Strasbourg, King Faisal Specialist Hospital and Research Center, the Children's Hospital of Philadelphia, "Hospices Civils de Lyon," and KK Women's and Children's Hospital approved this study. Participants were enrolled after written informed consent was obtained from parents or legal guardians. The clinical evaluation included medical history interviews, a physical examination, medical imaging as appropriate, and review of medical records.

Exome Sequencing and Analysis

To uncover genetic variants associated with the phenotypes of the two affected members of the Lithuanian (LT) family, we sequenced their exomes and that of their parents, as described.[9] DNA libraries were prepared from leukocytes by standard procedures. Exomes were captured and sequenced using different platforms as specified below to reach 50- to 120-fold coverage on average. Variants were filtered based on inheritance patterns including autosomal recessive, X-linked, and *de novo*/autosomal dominant. Variants with MAF $<$ 0.05% in control cohorts (dbSNP, the 1000 Genome Project, NHLBI GO Exome Sequencing Project, the ExAC, and our in-house databases) and predicted to be deleterious by SIFT,[10] PolyPhen-2,[11] and/or UMD predictor[12] were prioritized. This exome analysis singled out compound heterozygote variants in *KIAA1109* as possibly causative in both affected siblings, prompting us to look for other individuals with overlapping phenotypes and variants in the same gene through GeneMatcher, the DDD portal, and clinical genetics meetings. These searches led to the identification of a total of 17 additional affected individuals.

In the Algerian (AL) family, exome sequencing was performed by the Centre National de Génotypage (CNG, Evry, France), Institut de Génomique, CEA. Exomes were captured with Human All Exon v5; 50 Mb (Agilent Technologies) and sequenced on a HiSeq2500 platform (Illumina) as paired-end 100 bp reads. For the Saudi Arabian (SA1–SA3) families, exome capture and sequencing was performed in conjunction with autozygome analysis as previously described.[13] For the family from Singapore (SG), exome capture, sequencing, and variant calling and analysis were performed as described.[14] For the two families from Tunisia (TU1 and TU2), exome sequencing was performed on a NextSeq500 (Illumina) after SeqCapEZ MedExome Library preparation and analyzed with BWA and GATK HaplotypeCaller. Variants with MAF $<$ 0.1% in ExAC database and predicted to be deleterious by SIFT,[10] PolyPhen-2,[11] and Mutation Taster[15] were prioritized. The UK family's exome capture and sequencing was performed as previously described.[16] For the US family, exome capture, sequencing, and variant calling and analysis were performed as described.[16,17] The breakpoints of the paternally inherited deletion were determined by whole-genome sequencing.

Breakpoint Mapping by Whole-Genome Sequencing

100 ng of genomic DNA were sheared using Covaris with a target fragment size of 500 bp. The sequencing library was prepared using Tru-Seq DNA PCR-free Sample Prep Kit (Illumina) and 100-bp paired-end reads sequenced on a HiSeq 2500 platform (Illumina). The PCR-free kit was used to prepare the library in order to avoid PCR duplicates. Sequence-control, software real-time analysis, and bcl2fastq conversion software v.1.8.4 (Illumina) were used for image analysis, base calling, and demultiplexing. Purity-filtered reads were adapters- and quality-trimmed with FastqM*cf.* v.1.1.2 and aligned to the human_g1k_v37_decoy genome using BWA-MEM (v.0.7.10[18]). PCR duplicates were marked using Picard tools (v.2.2.1). We obtained a sequence yield of 11.4 Gb of aligned bases with a 3.6× mean coverage. Aligned reads within the *KIAA1109* locus were visualized and evaluated using Integrative Genomics Viewer (IGV) in search of chimeric inserts. We identified a single pair of paired-ends reads mapping unequivocally 8,971 bp apart within *KIAA1109* allowing us to map the paternally inherited deletion of the US proband breakpoints within exon 68 and intron 72. The breakpoints were then finely mapped with Sanger sequencing to coordinates chr4:123254885 and chr4:123263438 (hg19) (Figure S1).

Zebrafish Manipulations, CRISPR/Cas9 Editing, and Design of Morpholinos

Zebrafish animal experimentation was approved by the Ethical Committee for Animal Experimentation of the Geneva University Medical School and the Canton of Geneva Animal Experimentation Veterinary authority. Wild-type TU (Tübingen) zebrafish were maintained in standard conditions (26°C–28°C, water conductivity at 500 μS [pH 7.5]). Embryos obtained by natural matings were staged according to morphology/age.

Zebrafish *kiaa1109* mutant lines were developed using CRISPR-Cas9-mediated genome editing. Using the ZiFiT online tool,[19] we

identified three suitable 20-nucleotide sites upstream of protospacer adjacent motifs (PAM) for *S. pyogenes* Cas9 and targeting *kiaa1109* exons 1, 4, and 7 (numbering according to GenBank: NM_001145584.1). Annealed oligonucleotides carrying the 20-nucleotide target sequence were ligated into pDR274 (Addgene plasmid # 42250), and clones verified by Sanger sequencing, linearized, and used for *in vitro* transcription of single-guide RNAs (sgRNAs) using the MEGAshortscript T7 Transcription Kit (ThermoFisher). sgRNAs were mixed with recombinant Cas9 nuclease (PNA Bio), Danieau buffer, and phenol red as a tracer, and approximately 1 nL injected into early zebrafish embryos. Each injection contained 0.25 ng of sgRNA and 0.5 ng of Cas9 per nL. Evidence for genome editing was assessed qualitatively by PCR amplification around the target sites in each exon in injected embryo lysates. Heterogeneous PCR products, consistent with mosaic editing, was seen as smeared bands by gel electrophoresis, compared to uninjected embryos (not shown). Injected fish embryos were raised to adulthood and screened for their ability to transmit mutant *kiaa1109* alleles by out-crossing and PCR genotyping. PCR products were cloned with pCRII TOPO (ThermoFisher) to separate alleles, and colony PCRs were sequenced to detect germline transmission of potential *kiaa1109* frameshift alleles. Out-crossed F1 embryos were raised to adulthood for mutations detected in exon 1, 4, and 7, as separate lines. F1 adult fish were tail-clipped, targeted exons were amplified by PCR, and PCR products were cloned to pCRII TOPO (ThermoFisher) to identify specific *kiaa1109* mutant alleles in heterozygosity by colony PCR and DNA sequencing. Heterozygous F1 fishes carrying the same *kiaa1109* mutation were then in-crossed to assess embryonic survival and phenotype in homozygosity. *kiaa1109* genotyping for embryos from these crosses was made using PCR, amplifying the target exon regions. Products from wild-type, heterozygous, or homozygous mutant amplicons were distinguished by gel electrophoresis. Details of the sgRNA target sites, representative mutant allele sequencing chromatograms, and predicted frameshifts for the three *kiaa1109* mutant lines described are given in Figures S2 and S3.

To knock down *kiaa1109* (GenBank: NM_001145584.1) in zebrafish, we designed two non-overlapping splice-blocking MOs (morpholinos) targeting pre-mRNA: (1) sbE4MO- 5′-TGTTCTGT TTTTGCACTGACCATGT-3′ and (2) sbE2MO- 5′-CAACATTGAGA CAGACTCACCGATG-3′ (Gene Tools) that target the exon 4/intron 4 and exon 2/intron 2 boundaries, respectively. The standard Ctrl-MO (5′-CCTCTTACCTCAGTTACAATTTATA-3′) (Gene Tools) without any targets in the zebrafish genome was used for mock injections. MOs were dissolved in nuclease-free water and their concentrations determined by NanoDrop. The fish were injected at 1- to 2-cell stages (1–2 nL) using phenol red as a tracer in Danieau buffer. The following amounts of MO: 3.35 and 6.7 ng of sbE4MO; 5.6, 11.3, 16.9, and 22.3 ng of sbE2MO and the equivalent of Ctrl-MO for the higher doses were injected into wild-type zebrafish embryos, respectively. Uninjected, standard control MO, and *kiaa1109* MO-injected embryos were collected at 2 dpf and total RNAs were isolated using standard Trizol protocol (Invitrogen). 1 μg of total RNA from each sample was used to synthesize cDNA with the Superscript III kit with Oligo d(T) primers (Invitrogen). Dilutions 1/20 of cDNA were used for standard PCR reactions (JumpStart RED Taq ReadyMix, Sigma-Aldrich). Basic quantifications of agarose gels were performed with ImageQuant TL software (GE Healthcare). We assessed embryos for morphological changes at 2 days post-fertilization. We grouped the embryos into four classes by morphology: normal embryos, embryos with clear midbrain and/or hindbrain ventricle swelling, curved embryos, and embryos with both phenotypes. The degree of hydrocephaly was not measured; hydrocephaly was assessed by clear deviation from the normal embryo morphology (see Results). Curved embryos showed caudal axis curvature. They were clearly distinguishable from the straight anterior-posterior axis of normal 2-day-old embryos (see Results). The most severely affected embryos had a combination of hydrocephaly and caudal axis curvature (see Results).

Results

We first identified compound heterozygous missense variants in *KIAA1109* in a Lithuanian family with two affected siblings (LT.II.1 and LT.II.2) presenting with a constellation of severe global developmental delay, cerebral parenchymal rarefaction and ventriculomegaly (observed at 20 months of age), plagiocephaly, paretic position of hands and feet at birth, early-onset epilepsy, muscle hypotonia, stereotypical movements, hypermetropia, and lack of walking function (Table 1, Figure 1). As a homozygous stop-gain allele in this gene was suspected to cause a syndromic neurological disorder in a fetus (described in more details in this manuscript as fetus SA1.II.1) with cerebellar malformations, hydrocephalus, micrognathia, club feet, arthrogryposis with flexed deformity, pleural effusion, and death 1 hr after birth,[13] we hypothesized that *KIAA1109* variants cause an autosomal-recessive (AR) brain development disorder with arthrogryposis.

Our searches for more case subjects led to the identification of a total of 19 affected individuals from 10 families (including 6 undiagnosed miscarriages) recruited in Algeria (AL), Lithuania (LT), Saudi Arabia (SA1–SA3), Singapore (SG), Tunisia (TU1, TU2), the United Kingdom (UK), and the United States of America (US) (Figure 2A). Genetic variants associated with the complex phenotype of interest were uncovered through exome sequencing of the affected individuals and their healthy parents with the exception of SG.II.4. We found only one gene, *KIAA1109*, compliant with AR Mendelian expectations and bearing two putatively deleterious variants in all affected individuals. GENCODE[20] catalogs in Ensembl 16 isoforms of *KIAA1109*; two encode the full-length 5,005-amino acids protein, six have no coding potentials, and the remaining eight isoforms encode protein of lengths varying from 164 to 1,674 amino acids. All the mutations reported in this manuscript affect the full-length GenBank: NP_056127 protein. Consistent with consanguineous unions, the affected members of families AL, TU1/TU2, SA1, SA2, and SA3 were homozygous for variants c.9149C>A (p.Pro3050His), for c.10153G>C (p.Gly3385Arg), for c.1557T>A (p.Tyr519Ter), for c.11250−1G>A (r.11250_11465del, p.His3751_Arg3822del), and for c.12067G>T (p.Glu4023Ter), respectively, whereas the affected individuals from LT, SG, UK, and US families were heterozygote for c.3986A>G (p.Tyr1329Cys) and c.5599G>A (p.Val1867Met), for c.2902C>T (p.Arg968Cys) and

c.3611delA (p.Asn1204Thrfs*6), for c.4719G>A (p.Met1573Ile) and the *de novo* c.5873G>A (p.Arg1958Gln), and for c.997dupA (p.Ile333Asnfs*5) and the deletion g.123254885_123263438delinsG (c.11567_12352delinsG, p.Lys3856Argfs*44), respectively (nomenclature according to GenBank: NM_015312.3, NP_056127.2; Figures 2A and 2B, Table S1). The fact that families TU1 and TU2 are not known to be related suggests a Tunisian founder effect of variant c.10153G>C (p.Gly3385Arg). Sanger sequencing in each family confirmed the anticipated segregation of the *KIAA1109* variants, with the exception of family TU1 whose parents declined to be assessed. It also confirmed the genetic status of the SG.II.4 affected sibling (Figure 2A). All variants are either absent or encountered (as heterozygous variants) with a frequency lower than 1/10,000 in ExAC (v.0.3.1)[21] (Table S1). The missense variants are predicted to be functionally damaging at least by two of the three PolyPhen-2,[11] Provean,[22] and SIFT[10] predictors with the exception of the UK.II.1 variants predicted to be benign, neutral, and tolerated, respectively (Table S1). They might be under "compensated pathogenic deviation in human, a phenomenon that contributes to an unknown, but potentially large, number of false negatives to the evaluation of functional sites" as demonstrated in Jordan et al.[23] Missense variants and CNVs are underrepresented compared to expectation in ExAC (missense Z score = 4.97; CNV Z score = 0.77) indicating that *KIAA1109* is under constraint. The identification of 50 LoF variants compared to the 176.1 expected, while not significant with a pLI = 0.0, does not contradict this hypothesis. In agreement with a possible contributing role of bi-allelic *KIAA1109* LoF variants to the phenotype of affected individuals SA1.II.1, SA2.II.1, SA3.II.1, and US.II.3, ExAC does not report homozygous LoF variants in *KIAA1109*. The splice variant c.11250−1G>A identified in fetus SA2.II.1 is predicted to abolish the consensus acceptor site of intron 66.[24] A prediction validated by our RT-PCR experiments that showed a partial skipping of 216-nucleotides-long exon 67 in lymphoblastoid cell line of the affected SA2.II.1 fetus (Figure S4). The corresponding transcript would encode a protein lacking 72 amino acids. All missense variants identified in the AL, LT, SG, TU, and UK families affect highly conserved residues within evolutionary conserved region of the encoded protein (Figure S5).

As exemplified by the LT.II.1 and LT.II.2 siblings and the SA1.II.1 proband, the phenotype of the 19 affected individuals ranges from global developmental delay with/without inability to stand to stillbirth. Many of the more severely affected case subjects harbor homozygous or compound heterozygote truncating alleles (families SA1–SA3 and US) (Table 1, Supplemental Note). While the phenotype of proband SA1.II.1 is summarized above, SA2.II.1 and SA3.II.1 stillborn fetuses shared hydrocephalus, cerebellar hypoplasia, arthrogryposis, and skeletal anomalies (Figures 3 and S6). Proband SA2.II.1 had bilateral overlapping fingers, apparent contractures of the hands and feet, and bilateral sandal gaps, along with shortened long bones and nuchal thickening. He also presented with absence of corpus callosum and abnormal kidneys. Proband SA3.II.1 showed skin edema, bilateral talipes, and arthrogryposis (Figure 3; Table 1; Supplemental Note). The US.II.3 stillborn fetus resulted from a 3rd pregnancy attempt of the couple (Figure 2A). The fetus demonstrated major central nervous anomalies including thin cerebral parenchyma with lissencephalic pattern, prominent germinal matrix, ventriculomegaly, brain stem vermian dysgenesis (kinked brain stem and elongated pons), and absence of corpus callosum, as well as closed spinal defect at L4-L5, associated with extra-central nervous anomalies including coarctation of the aorta, small omphalocele, echogenic bowel, hydrops, cystic hygroma, pleural effusion, possible anal atresia, low-set ears, short penis, clinodactyly, talipes, and abnormal posturing of the limbs (Figure 3). The SG family had four pregnancy attempts; two resulted in miscarriages (SG.II.2 and SG.II.3) and two in fetuses who did not pass the first semester (Figures 2A and 4). The SG.II.1 elder brother had minimal respiratory effort at birth and required immediate intubation and mechanical ventilation. He presented with macrocephaly, hypertelorism, posteriorly rotated ears, flattened nasal bridge, congenital cataract, and microphthalmia. He had generalized arthrogryposis and bilateral congenital talipes equinovarus. He also had hypotonia and an ano-rectal malformation with recto-perianal fistula (Figure 4). Brain MRI showed major cerebral parenchymal thinning with lissencephalic aspect, severe ventriculomegaly, absence of corpus callosum, and severe cerebellar and pontine hypoplasia (Figure 4). He passed away at 3 months of age from pneumonia and septic shock. The SG.II.4 younger sibling was remarkably similar to his elder with hypertelorism, bilateral low-set ears, short nose, anteverted nares, bilateral congenital cataract, microphthalmia, webbed neck, bilateral structural congenital talipes equinovarus, generalized arthrogryposis, and hypotonia. Brain MRI showed severe hydrocephalus with marked thinning of the cerebral parenchyma. The corpus callosum was absent, the cerebellum and brainstem were hypoplastic, and there was a pontomesencephalic kink. He remained ventilator dependent from birth and passed away at 1 month of age (Figure 4). The AL.II.1 fetus presented with an equally severe phenotype so the parents elected to terminate the pregnancy. He showed multiple brain malformations including hydrocephalus, vermis fusion, lamination defect of cerebellar cortex, and absence of the corpus callosum, combined with arthrogryposis with flexed deformity and bilateral adductus thumbs, diffuse effusion, and other clinical features (Figures 3 and S6; Table 1; Supplemental Note). Affected individuals and fetuses TU1.II.1, TU1.II.4, and TU2.II.2 had arthrogryposis and the same cerebral malformative pattern, associating cerebellar and brainstem dysgenesis, parenchymal thinning with major lack of gyration, corpus callosum agenesis, and hyperplastic germinal matrix protruding within ventriculomegaly. The severity of features of the AL.II.1

Table 1. Overlapping Clinical Features of Individuals with KIAA1109 Variants

Family #	Individual #	Gender, Age	Ethnicity	Gene Mutations	ID	Mutation Coordinates (GRCh37/hg19)	Cerebral Anomalies (Pre/Post-natal Images) and Pathological Findings	Head and Face
LT	LT.II.1 (brother)	male, 13 yo	Lithuanian	compound heterozygote	severe, global developmental delay, no language, cannot stand or walk without support	Chr4:123160823; c.3986A>C, Chr4:123170727; c.5599G>A	post-natal brain MRI: small posterior fossa arachnoid cyst, discrete vermian atrophy, slight increase of the fluid-filled retro and infra-cerebellar space and mild enlargement of subarachnoid spaces of frontal regions.	plagiocephaly
LT	LT.II.2 (sister)	female, 7 yo	Lithuanian	compound heterozygote	severe, global developmental delay, no language, cannot sit or stand without support	Chr4:123160823; c.3986A>C, Chr4:123170727; c.5599G>A	post-natal brain MRI: discrete parenchymal rarefaction involving the frontal lobes	plagiocephaly
UK	UK.II.1, DDD# 263241	female, 11 yo	British	compound heterozygote with one *de novo* missense mutation	global developmental delay, mild to moderate learning disability	Chr4:123164200; c.4719G>A and Chr4:123171679; c.5873G>A	prenatal imaging (US and MRI): major microcephaly (HC -5 SD) with reduced white matter volume and mild ventriculomegaly	hypertelorism, slightly upslanting palpebral fissures
AL	AL.II.1	male, termination of pregnancy at 21 weeks of amenorrhea	Algerian	homozygous missense mutation	not applicable	Chr4:123207807; c.9149C>A	prenatal US findings: triventricular ventriculomegaly and corpus callosum agenesis; neuropathological findings: absence of cortical lamination and diffuse migration anomalies within a thin parenchymal mantle, ventriculomegaly, and voluminous germinal matrix. Corpus callosum was not identified. Infra-tentorial space: hypoplasia of the pons with absence of the longitudinal and transversal fibers and dysplasia of the cerebellum characterized by lack of foliation and poorly identified vermis; narrowing of the aqueduct	hypertelorism, posteriorly rotated ears

Eyes	Mouth	Joints	Limbs	Gastro-intestinal	Urogenital	Heart	Muscles	Behavior	Other Symptoms
hypermetropia, strabismus, astigmatism	delayed eruption of permanent teeth	mild contractures of large joints	syndactyly of 2nd and 3rd toes, hands and feet paresis at birth, talipes valgus	normal	scrotum hypoplasia	none	muscle hypotonia, atrophy	stereotypic movements, spontaneous paroxysms of laughter	early-onset epilepsy
hypermetropia, strabismus, astigmatism	normal	mild contractures of large joints	paretic position of hands and feet in infancy, talipes valgus	chronic constipation	none	none	muscle hypotonia, atrophy	stereotypic movements	early onset epilepsy, dermatitis, psoriasis
ocular motor apraxia, hypermetropia, strabismus	dental crowding, high palate	mild bilateral talipes managed by physiotherapy only; asymmetry of the thorax	syndactyly of 2nd and 3rd toes, 5th toe clinodacytly, hallux valgus	gastro-esophageal reflux	none	complex congenital heart disease (tetralogy of Fallot with pulmonary atresia)	none	poor concentration, immature behavior with minor self-harm (head-banging) when angry/ frustrated	none
bilateral cataract with crystalline fibers of variable size and orientation	retrognathism, big horizontalized mouth	arthrogryposis (flexed deformity of shoulders, elbow and hips, and bilateral adductus thumbs)	bilateral equinovarus foot	choanal atresia	scrotum hypoplasia	pericardial effusion	not available	not applicable	slight pleural effusion, peritoneal effusion, dilatation of lymph vessels in lung with lympho-hematopoietic elements

(*Continued on next page*)

Table 1. *Continued*

Family #	Individual #	Gender, Age	Ethnicity	Gene Mutations	ID	Mutation Coordinates (GRCh37/hg19)	Cerebral Anomalies (Pre/Post-natal Images) and Pathological Findings	Head and Face
TU1	TU1.II.1	female, died at 3 days of age	Tunisian	homozygous missense mutation	not applicable	Chr4:123230520; c.10153G>C	prenatal imaging (US and MRI): cerebellar hypoplasia and brainstem dysgenesis (flat and elongated pons and slightly kinked brainstem with increased fluid filled retro-cerebellar spaces); severe parenchymal thinning with major lack of gyration (lissencephalic aspect) associated with voluminous germinal matrix protruding within moderate ventriculomegaly and absence of corpus callosum. Cephalic biometry was normal.	hypotelorism
TU1	TU1.II.4	male, termination of pregnancy at 23 weeks	Tunisian	homozygous missense mutation	not applicable	Chr4:123230520; c.10153G>C	prenatal US findings: severe parenchymal thinning with lack of gyration associated with ventriculomegaly and corpus callosum agenesis. Neuropathological findings: complete corpus callosum agenesis, ventricular dilatation, severe cortical malformations with a reduced cortical plate, neuronal depletion, heterotopia within white matter, dysplasia of brainstem and cerebellum.	none
TU2	TU2.II.2	female, died at 12 days of age	Tunisian	homozygous missense mutation	not applicable	Chr4:123230520; c.10153G>C	prenatal imaging (US and MRI): cerebellar hypoplasia and dysgenesis associated to severe brainstem dysgenesis characterized by flat and elongated pons and slightly kinked brainstem with increased fluid filled retro-cerebellar spaces. Corpus callosum was not identified. Supratentorial anomalies include severe parenchymal thinning associated with lissencephalic aspect as well as voluminous germinal matrix protruding within severe ventriculomegaly.	none
SA1	SA1.II.1, 13DG1900	female, death at 1 hr after delivery	Saudi	homozygous nonsense mutation	not applicable	Chr4: 123128323; c.1557T>A	prenatal US findings: severe ventriculomegaly with supratentorial cerebral mantle thinning associated with cerebellar hypoplasia	small eyes, low-set ears

Eyes	Mouth	Joints	Limbs	Gastro-intestinal	Urogenital	Heart	Muscles	Behavior	Other Symptoms
none	deep palate	left club foot	long fingers	none	none	left heart hypoplasia	not available	not applicable	none
none	none	arthrogryposis (hip and shoulder contractures)	clenched hands, camptodactyly, club feet	none	none	none	not applicable	not applicable	none
microphthalmia, blepharophimosis	none	club feet	club feet and hands	none	none	none	hypotonia	not applicable	narrow chest
small eyes	micrognathia	severe arthrogryposis (fixed elbows, fixed bilateral talipes, bilateral overlapping fingers, bilateral clinodactyly)	bilateral club foot	not available	not available	not available	not available	not applicable	pleural effusion

(*Continued on next page*)

Table 1. *Continued*

Family #	Individual #	Gender, Age	Ethnicity	Gene Mutations	ID	Mutation Coordinates (GRCh37/hg19)	Cerebral Anomalies (Pre/Post-natal Images) and Pathological Findings	Head and Face
SA2	SA2.II.1, 15DG0595	female, stillborn	Saudi	homozygous splice mutation	not applicable	Chr4:123252480; c.11250−1G>A	prenatal US findings: hydrocephalus, absent corpus callosum, hypoplastic cerebellum	not available
SA3	SA3.II.1, 15DG1933	female, stillborn	Saudi	homozygous nonsense mutation	not applicable	Chr4:123258092; c.12067G>T	prenatal US findings: hydrocephalus, hypoplastic cerebellum	not available
US	US.II.3	male, termination of pregnancy at 19 weeks	Caucasian	compound heterozygote	not applicable	Chr4:123113479; c.997dupA and Chr4: 123254885_ 123263438del; c.11567_ 12352delinsG	prenatal imaging (US and MRI): severe ventriculomegaly, thin cerebral parenchyma and cortical mantle associated with lissencephalic pattern, prominent germinal matrix, brain stem and vermian dysgenesis (kinked brain stem) and elongated pons; corpus callosum agenesis	low-set ears, webbed neck
SG	SG.II.1	male, died at 3 months of age	Chinese	compound heterozygote	not applicable	Chr4:123147970; c.2902C>T and Chr4:123159280; c.3611delA	post-natal MRI: supratentorial findings include both severe parenchymal (or cerebral mantle) thinning and smooth cortical surface, germinolytic cysts involving voluminous germinal matrix protruding within severe ventriculomegaly without any identification of corpus callosum. Infratentorial findings include severe cerebellar hypoplasia with severe brain-stem dysgenesis characterized by a kinking aspect.	macrocephaly; hypertelorism; posteriorly rotated ears; flattened nasal bridge
SG	SG.II.4	male, died at 1 month of age	Chinese	compound heterozygote	not applicable	Chr4:123147970; c.2902C>T and Chr4:123159280; c.3611delA	post-natal MRI: supratentorial findings include both severe parenchymal (or cerebral mantle) thinning and smooth cortical surface, germinolytic cysts involving voluminous germinal matrix protruding within severe ventriculomegaly without any identification of corpus callosum. Infratentorial findings include severe cerebellar hypoplasia with severe brain-stem dysgenesis characterized by a kinking aspect	macrocephaly; hypertelorism; bilateral low-set ears, short nose; anteverted nares

Eyes	Mouth	Joints	Limbs	Gastro-intestinal	Urogenital	Heart	Muscles	Behavior	Other Symptoms
not available	not available	arthrogryposis multiplex	bilateral overlapping fingers, bilateral cleft feet, bilateral cleft toes, and bilateral sandal gaps	normal	bilaterally abnormal kidneys	not available	not available	not applicable	skeletal shortening, nuchal thickening
not available	not available	arthrogryposis	bilateral talipes	normal	normal	absent fetal heart	not available	not applicable	skin edema
normal	unremarkable	severe arthrogryposis with flexion contractures and pterygia, hyperflexed wrists, bilateral clinodactyly	bilateral talipes	normal	echogenic malrotated bowel without ascites, short penis with bulbous shaft	coarctation of the aorta	muscle atrophy	not applicable	low conus non-immune hydrops with scalp edema, cystic hygroma, anal atresia, bilateral pleural effusion
congenital cataract; microphthalmia	no structural anomalies	arthrogryposis (involving bilateral shoulders, elbows, wrists, hands, knees)	bilateral structural congenital talipes equinovarus (CTEV)	ano-rectal malformation with recto-perianal fistula	no structural anomalies	small atrial septal defect/ patent foramen ovale	hypotonia	not applicable	excess skin folds of neck
congenital cataracts; microphthalmia	no structural anomalies	arthrogryposis (involving bilateral elbows, wrists, hands, knees, hips)	bilateral structural congenital talipes equinovarus (CTEV)	normal	no structural anomalies	small to moderate fenestrated atrial septal defect	hypotonia	not applicable	webbed neck; inverted nipples

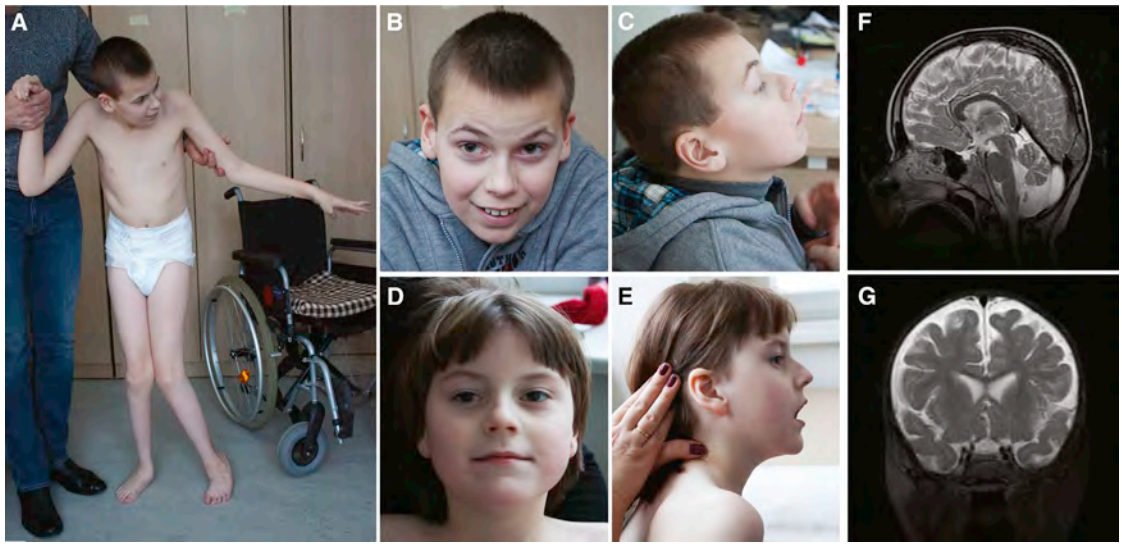


Figure 1. Pictures and Brain MRI from Surviving Individuals
Front and side views of the LT affected brother LT.II.1 (A–C) and sister LT.II.2 (D, E) at the ages of 13 years and 7 years, respectively. Brain MRI images of affected individual LT.II.1 at age of 8 years showed small posterior fossa arachnoid cyst, discrete vermian atrophy, and slight increase of the fluid-filled retro and infra-cerebellar space (F). Brain MRI images of affected individual LT.II.2 at age of 1 year showed discrete parenchymal rarefaction involving mainly the frontal lobes (G).

fetus, the SG.II.1 and SG.II.4 siblings, and the TU1.II.1, TU1.II.4, and TU2.II.2 Tunisian affected individuals and their resemblance with those observed in carriers of truncating variants suggest that the missense variants p.Pro3050His, p.Gly3385Arg, and p.Arg968Cys act as LoF or strong hypomorphs. Consistent with this hypothesis, the C>T transition in exon 24 of the latter variant is predicted to alter an exonic splicing enhancer site and thus proper splicing. More experiments are warranted to further demonstrate these assumptions.

All case subjects compatible with life carry missense variants (Figure 2; Table S1; Supplemental Note). Whereas the two LT.II.1 and LT.II.2 Lithuanian siblings are briefly described above, the UK.II.1 British proband showed global developmental delay, microcephaly, absence of the pulmonary valve, tetralogy of Fallot and ventricular septal defect, ocular motor apraxia, hypermetropia, dental crowding, 5th toe clinodactyly, syndactyly of the 2nd and 3rd toe like the LT.II.1 elder sibling, hallux valgus, and pes planus (Supplemental Note).

In line with the clinical presentation of LoF affected individuals, ablation in fruit flies and mice of the *KIAA1109* orthologs, *tweek* and *Kiaa1109*, respectively, resulted in lethality. Whereas *Kiaa1109*$^{-/-}$ mice engineered and phenotyped by the International Mouse Phenotyping Consortium[25,26] exhibited complete penetrance of preweaning lethality, some rare homozygous *tweek* mutants survive to adulthood.[27] These survivors presented with severe neurological defects such as seizures, inability to stand upright for long periods or walk, suggesting that *tweek* was involved in synaptic function.[27] These results further support a causative role of LoF of *KIAA1109* in the phenotypes observed in families AL, SA1–SA3, and US. Consistent with this hypothesis, *KIAA1109* has higher expression in the pituitary, the cerebellum, and the cerebellar hemispheres according to GTex.[28]

To further assess the consequences of decreased *KIAA1109* activity, we used both CRISPR/Cas9 genome editing and morpholinos (MO) technology in zebrafish. We generated three different stable lines with frameshift variants in exons 1, 4, and 7 of *kiaa1109*. Crosses of each heterozygote line with themselves suggest that these mutations are not lethal. To explain the discrepancy between these results and what was observed in mice and fruit flies, we profiled the transcriptome of homozygotes larvae. While we observed subtle differences between homozygous fish and their wild-type clutchmates, by and large we see no changes in expression of the different *kiaa1109* exons (Table S2). Our results suggest that the expression of *kiaa1109* isoforms containing only downstream exons encode proteins providing all the non-redundant functions of kiaa1109. More work is warranted to assess whether the engineered variants are inducing nonsense-mediated decay and whether there is any maternal contribution. In parallel, we knocked down *kiaa1109* using two different non-overlapping morpholinos (MOs). While we are aware that unspecific effects have been reported when using MOs,[29] we still favored this approach to mimic to a certain degree the situation observed in the LT.II.1 and LT.II.2 siblings and the UK.II.1 affected individual. Injection of early zebrafish embryos with 6.7 ng of sbE4MO resulted in a 50% reduction of the *kiaa1109* transcripts through skipping of 65 nucleotides long exon 4 (Figure S7). 49% of morphants were hydrocephalic or presented with other head defects, whereas only 3% of the mock-injected fish showed such phenotypes (Figures 5A

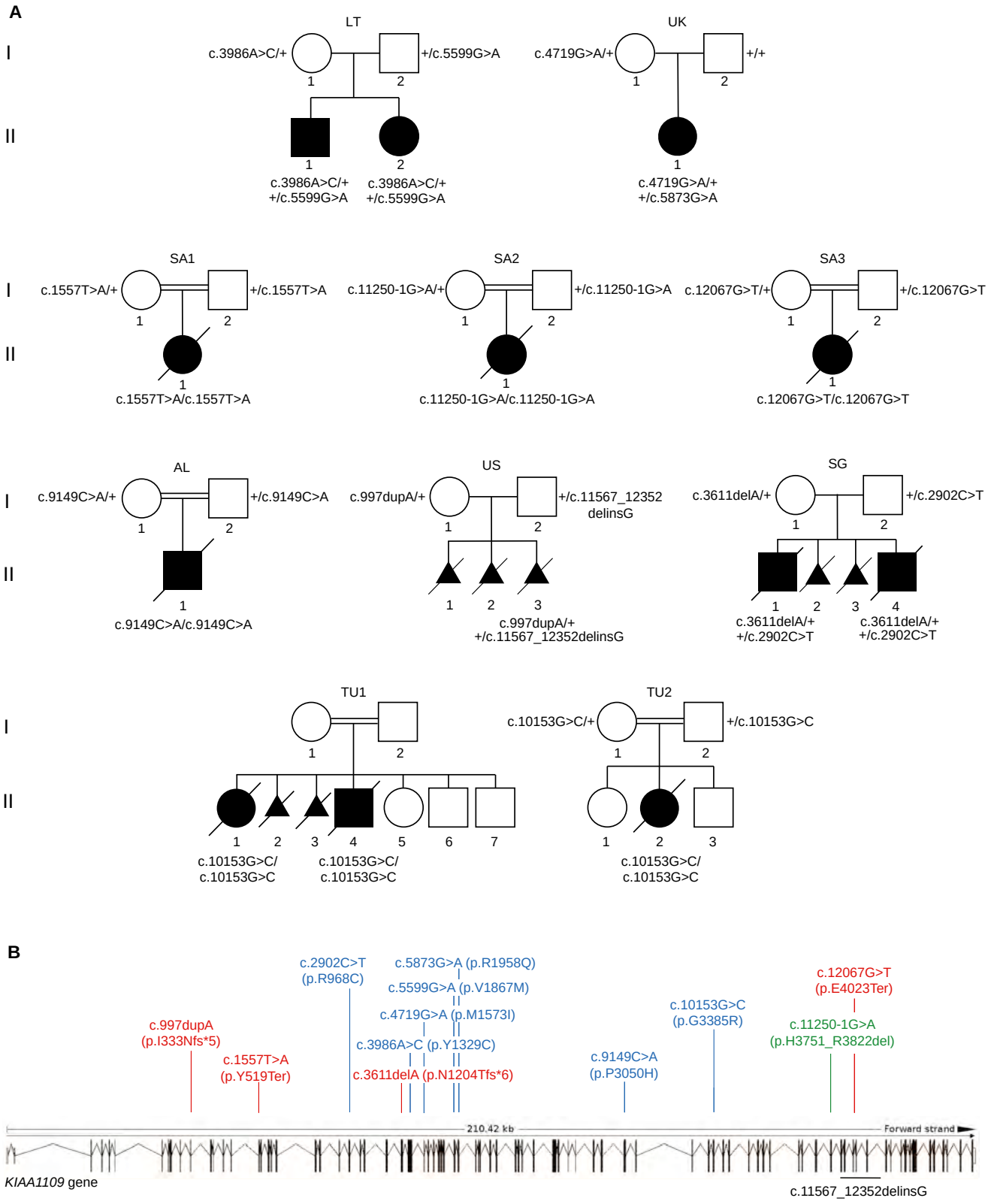


Figure 2. *KIAA1109* Pedigrees and Variants

(A) Pedigrees of the ten families carrying *KIAA1109* variants. The affected individuals of the Lithuanian (LT), Singaporean (SG), British (UK), and American (US) families are compound heterozygotes for rare variants, whereas the probands of the Algerian (AL), Saudi Arabian (SA1–SA3), and Tunisian (TU1, TU2) consanguineous families are homozygous for *KIAA1109* variants.

(B) Distribution of variants along the schematically represented 86 exons of *KIAA1109*. Missense variants are depicted in blue, nonsense in red, and the splice site variant in green. The extent of the deletion identified in the proband of the US family is indicated in black below.

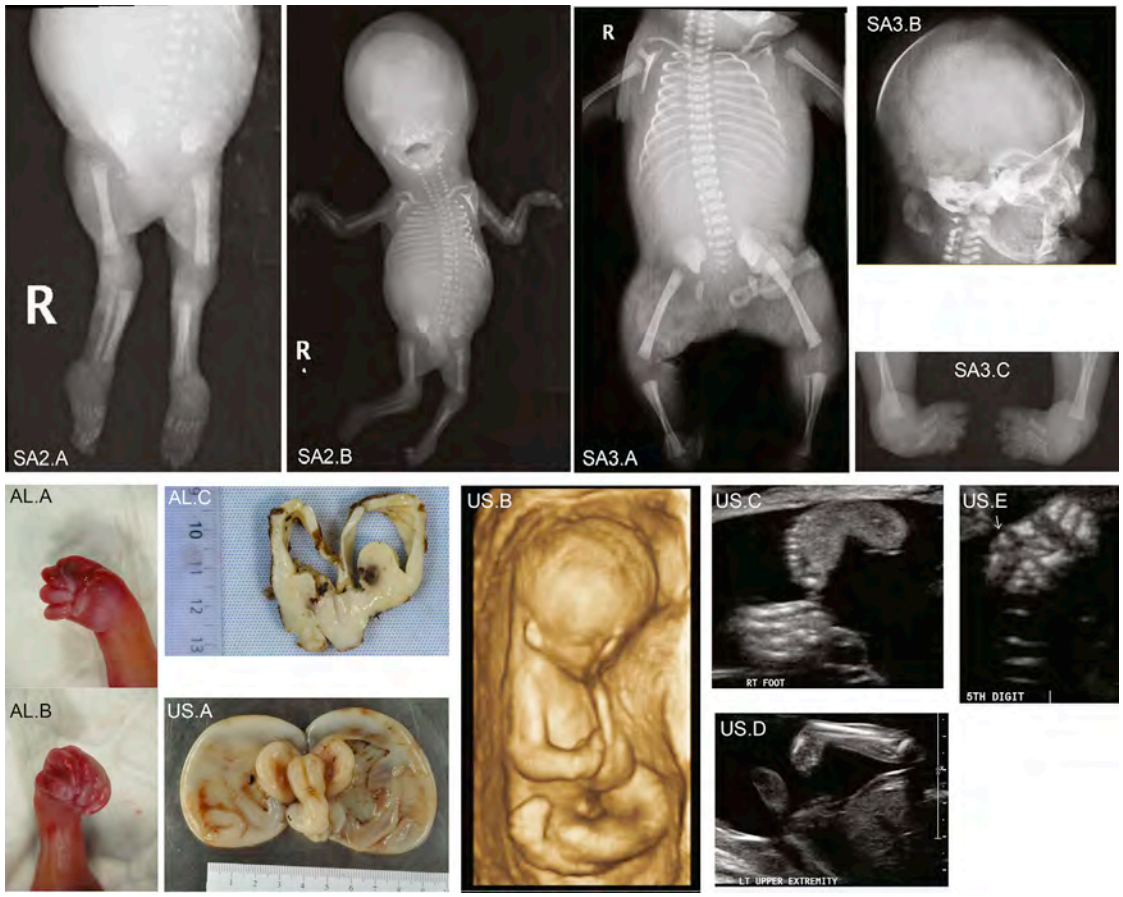


Figure 3. Ultrasound, X-Rays, and Autopsy Images of the SA2.II.1, SA3.II.1, AL.II.1, and US.II.3 Fetuses
X-ray images showing arthrogryposis of SA2.II.1 fetus (SA2.A and SA2.B).
X-ray images showing SA3.II.1 skeleton (SA3.A), head (SA3.B), and club feet (SA3.C).
Autopsy pictures from the AL.II.1 fetus showing right (AL.A) and left (AL.B) adductus thumbs of the fetus, and dilatation of cerebral ventricles with agenesis of corpus callosum (AL.C).
Autopsy image of the brain from US.II.3 fetus showing hydrocephalic brain with diaphanous pallium (US.A). The colliculi appear as single elongated ridges separated by a midline futter and the midline appears angulated on the brainstem, which is small as is the cerebellum. Antenatal ultrasound scan showed general arthrogryposis (US.B), one hyperflexed wrist (US.C), club feet (US.D), and bilateral clinodactyly of one hand (US.E).

and 5B). The second MO, sbE2MO, acts through retention of 88-nucleotide-long intron 2 and decreases *kiaa1109* levels by about 50% after injection of a large dose of 16.9 ng (Figure S7). Under such condition, about two-thirds (66%) of MO-sbE2 morphants were hydrocephalic or presented with other head defects compared to 8% of mock-injected fish affected at such dose (Figures 5A and 5C). Of note, rescue experiments of MO-injected zebrafish could not be performed due to the large size of the *kiaa1109* transcript. In summary, knockdowns of the zebrafish *KIAA1109* ortholog using two different MOs resulted in hydrocephalic animals reminiscent of probands' features.

Discussion

Data aggregation of exome sequencing from multiple laboratories allowed associating homozygous and compound heterozygote variants in *KIAA1109* with a syndrome that we suggest naming Alkuraya-Kučinskas syndrome (AKS), as these clinicians first described affected individuals at the severe and mild ends of the phenotype, respectively. AKS combines severe brain malformations (13 affected individuals out of 13), in particular hydrocephaly/ventriculomegaly (11/13) and corpus callosum agenesis (8/13) with arthrogryposis/contractures (10/13) and/or talipes valgus/talipes equinovarus/club foot (12/13) and heart defects (6/13).

AKS presents multiple overlaps with Aase-Smith syndrome 1 (ASS1 [MIM: 147800]) characterized by arthrogryposis, hydrocephalus, Dandy-Walker malformation, talipes equinovarus, cardiac defects, and risks of stillbirth or premature death. However, the two described families with one father and two children affected[30] and one mother and her affected daughter[31] are suggestive of a dominant rather than a recessive mode of transmission. Consistent with the view that AKS and ASS1 have different

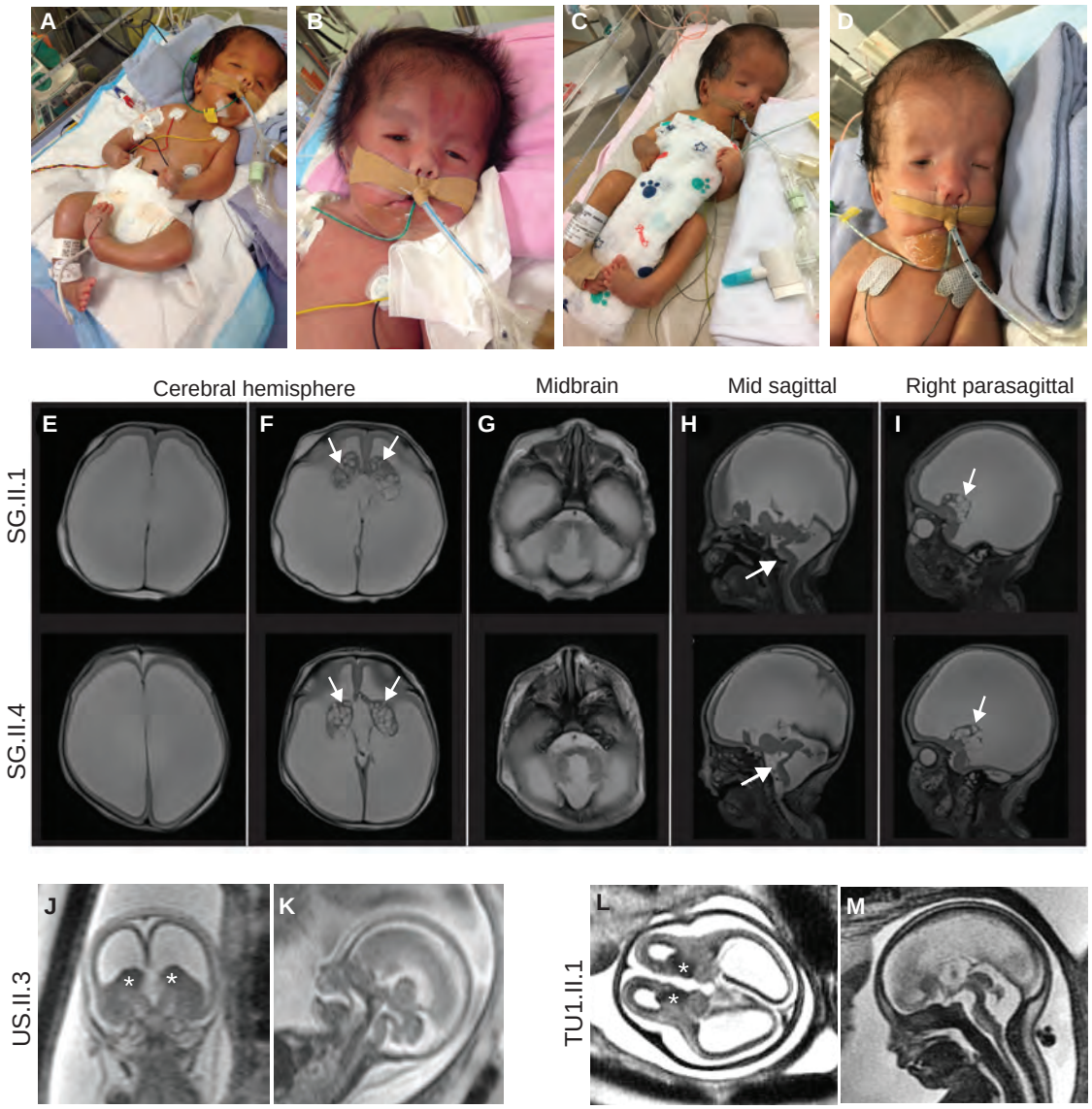


Figure 4. Pictures and Brain MRI Images of the SG.II.1 and SG.II.4 Babies and US.II.3 and TU1.II.1 Fetuses

(A–D) Photographs of SG.II.1 (A and B) and SG.II.4 (C and D) babies showing their whole bodies (A and C) and a close up of their faces (B and D).

(E–I) Brain MRI images of the elder brother SG.II.1 (top) and the younger brother SG.II.4 (bottom). Axial T2 weighted images showed severe ventriculomegaly, associated with severe thinning of the brain parenchyma (E, F). The brain parenchyma showed absence of normal gyral/sulcal pattern with smooth appearance in keeping with lissencephaly (E, F). Corpus callosum appeared to be absent (E, H). Note the prominent germinal matrix with germinolysis cysts (solid arrows) (F, I). The pons and cerebellum appeared hypoplastic with dilatation of the 4th ventricle (G, H) and Z shaped appearance of the brainstem (solid arrows) (H).

(J–M) Coronal (J), axial (L), and midsagittal (K, M) T2-weighed fetal prenatal MRI images of US.II.3 at 18.5 weeks of pregnancy (J and K) and TU1.II.1 at 28 weeks of pregnancy (L and M) demonstrating a similar imaging pattern including thin parenchyma (lissencephalic aspect), prominent germinal matrix marked by an asterisk, ventriculomegaly, and brain stem and vermian dysgenesis (kinked brain stem and elongated pons).

In summary, we observe a similar brain malformation pattern both prenatally—US.II.3 in (J) and (K), TU1.II.1 in (L) and (M), AL.II.1 (see text), TU1.II.4 (see text), and TU2.II.2 (see text)—and postnatally (SG.II.1 [E–I top] and SG.II.4 [E–I bottom]).

etiologies, both ASS1-affected families presented individuals affected with cleft palate, a birth defect not present in the 13 *KIAA1109* individuals described here, including those at the severe end of the phenotypic spectrum. Interestingly, we have identified by exome sequencing an individual with partially overlapping features carrying a single *de novo* variant in *KIAA1109*. Although we cannot exclude that a second variant is present outside of the open reading frame, we might, alternatively, be dealing (1) with a spurious association or (2) with another syndrome related to AKS and associated with single variants in *KIAA1109*, similar to the *de novo* and biallelic variation recently associated with mitochondrial dynamics pathologies.[32] The high missense ExAC Z-score of

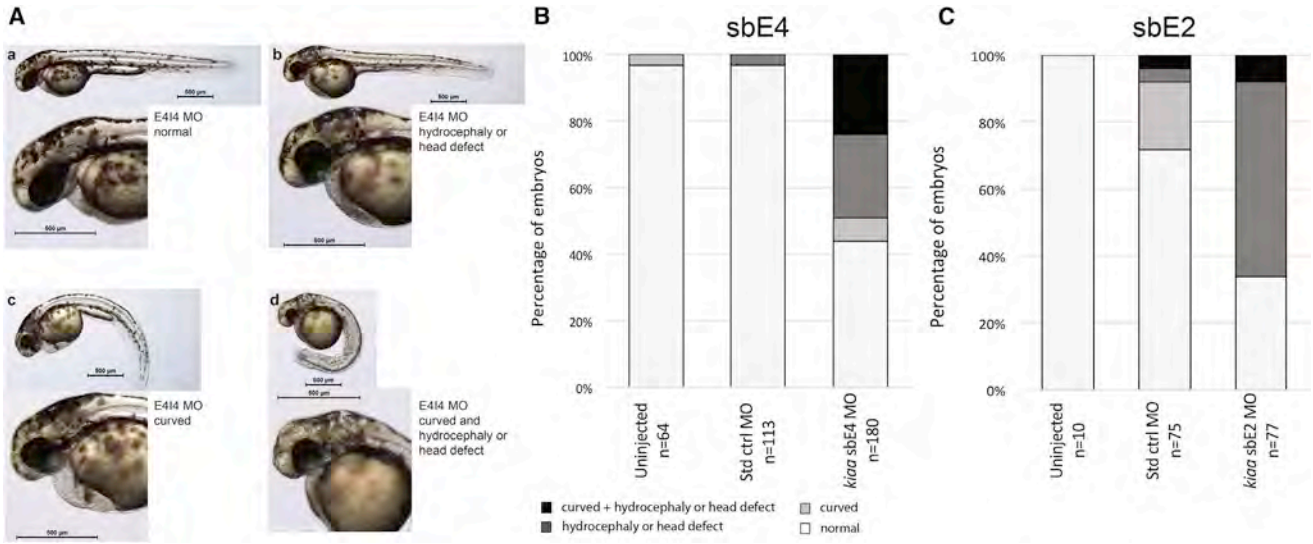


Figure 5. *kiaa1109* Knockdown in Zebrafish Results in Phenotypes Reminiscent of Probands' Clinical Features
(A) Lateral views representing the four classes of observed phenotypes in 2 dpf TU zebrafish embryos injected with sbE4-MO (morpholino) targeting *kiaa1109*: from top left to bottom right, normal, hydrocephalic or other head defects, curved and curved with head defect.
(B) Results for uninjected embryos (left) and those injected with equivalent amounts of standard control MO (center) or *kiaa1109* sbE4-MO (6.7 ng, right). Phenotyping and scoring were performed at 2 dpf in two independent experiments.
(C) Results for uninjected embryos (left) and those injected with equivalent amounts of standard control MO (center) or *kiaa1109* sbE2-MO (16.9 ng, right). Phenotyping and scoring were performed at 2 dpf in two independent experiments.

KIAA1109 is compatible with such hypothesis. The identification of other similarly affected individuals will allow disentangling this conundrum.

The observed combination of intellectual disability, corpus callosum hypoplasia, hydrocephalus, and talipes equinovarus is also reminiscent of the constellation of features seen in the *L1CAM*-associated (neural cell adhesion molecule L1 [MIM: 308840]) HSAS (hydrocephalus due to congenital stenosis of aqueduct of sylvius [MIM: 307000]) and CRASH (corpus callosum hypoplasia, retardation, adducted thumbs, spastic paraplegia, and hydrocephalus [MIM: 303350]) syndromes. Whereas HSAS syndrome leads to neonatal or infant death, *L1CAM* variants survivors are described as affected by CRASH syndrome. Similarly, ten of the herein described AKS-affected case subjects did not survive past infancy (16 if accounting for undiagnosed miscarriages), whereas the living UK.II.1, LT.II.1, and LT.II.2 individuals presented other prominent features reported in CRASH syndrome, such as adducted thumbs, short stature, microcephaly, language impairment, and abnormalities of tone. L1CAM is a cell adhesion molecule that plays critical roles in neuronal migration and differentiation.[33] Congenital joint contractures, limb deformities, hydrocephalus, corpus callosum agenesis, hypoplastic brainstem, cortical thinning, and high proportions of stillborn or neonatal death also are reminiscent of the PVHH (proliferative vasculopathy and hydranencephaly-hydrocephaly [MIM: 225790]) syndrome, a recessive disorder caused by variant in the transmembrane calcium transporter, *FLVCR2* (feline leukemia virus subgroup C receptor 2 [MIM: 610865]).

The *Drosophila* ortholog of *KIAA1109* named *tweek* is widely expressed but enriched in the brain lobes and in the ventral nerve cord.[27] Neuronal phosphatidylinositol-4,5-bisphosphate [PI(4,5)P(2)] levels are critical in restricting synaptic growth via localization and activation of presynaptic Wiscott-Aldrich syndrome protein/WASP, a phenomenon dependent on *tweek* but not on bone morphogenetic protein signaling.[34] The 5,005-amino-acid-long KIAA1109 protein is conserved from nematodes to vertebrates (Figure S8) in spite of a lack of recognizable domains, with the exception of a 22-residue amino-terminal transmembrane segment and a small central coiled-coil of 22 residues. It is described by specialists as an unconstrained peptide thought to adopt a definite conformation upon binding to its interactors.[35] Consistent with this hypothesis, multiple high-throughput protein-protein interactions screens coupling near-endogenous expression levels with quantitative proteomics and mass spectrometry have identified human or mouse KIAA1109 interactors. For example, CTNNB1 (catenin beta-1), a protein associated with a dominant form of intellectual disability (MIM: 615075), interacts with two separate regions of KIAA1109.[36] Another set of experiments showed high-confidence interactions with BUB3, DNAJB1, and PTPA, three proteins implicated in cell division.[37] BUB3 participates to the spindle-assembly checkpoint signaling and the establishment of kinetochore-microtubule attachments. It inhibits the ubiquitin ligase activity of the anaphase-promoting complex (APC/C) by phosphorylating its activator CDC2. PTPA, one of four major Ser/Thr phosphatases, negatively controls cell growth and division. DNAJB1 (a.k.a. HSP40)

interacts with HSP70 and stimulates its ATPase activity and its association with HIP. Interestingly, lower-confidence KIAA1109 protein interactors include BAG2 that competes with HIP for binding to the HSC70/HSP70 ATPase domain, as well as DRC1 and SMAD2. *DRC1* (MIM: 615288) encodes a central component of the nexin-dynein complex that regulates the assembly of ciliary dynein and is associated with primary ciliary dyskinesia (MIM: 615294). *SMAD2* (MIM: 601366) regulates cell proliferation, apoptosis, and differentiation through mediation of TGF-β signaling.

Conclusion

We propose that bi-allelic LoF and missense variants in *KIAA1109* cause an autosomal-recessive brain malformation disorder with cerebral parenchymal underdevelopment ranging from major cerebral parenchymal thinning with lissencephalic aspect to moderate parenchymal rarefaction, severe to mild ventriculomegaly, and cerebellar hypoplasia with brainstem dysgenesis, associated with club foot and arthrogryposis. Severe cases are incompatible with life. Although further studies have to be engaged, our findings suggest that *KIAA1109* is potentially involved in cell cycle control, particularly of the central nervous system.

Accession Numbers

Alkuraya-Kucinskas syndrome as described in this paper has been assigned MIM: 617822.

Supplemental Data

Supplemental Data include Supplemental Note, eight figures, and two tables and can be found with this article online at https://doi.org/10.1016/j.ajhg.2017.12.002.

Acknowledgments

We thank the affected individuals and their families for their contribution to this study. We are grateful to Keith Joung for reagents. This work was supported by grants from the Swiss National Science Foundation (31003A_160203) to A.R., the Lithuanian-Swiss cooperation program to reduce economic and social disparities within the enlarged European Union (CH-3-ŠMM-0104, Unigene project) to V.K. and A.R., and King Abdulaziz City for Science and Technology grant 13BIO-1113-20 to F.S.A. We also acknowledge the support of the Saudi Human Genome Program. The DDD Study presents independent research commissioned by the Health Innovation Challenge Fund (grant number HICF-1009-003), a parallel funding partnership between the Wellcome Trust and the Department of Health, and the Wellcome Trust Sanger Institute (grant number WT098051). The study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC). The research team acknowledges the support of the National Institute for Health Research, through the Comprehensive Clinical Research Network. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Received: December 2, 2016
Accepted: December 4, 2017
Published: December 28, 2017

Web Resources

1000 Genomes, http://browser.1000genomes.org/index.html
Burrows-Wheeler Aligner, http://bio-bwa.sourceforge.net/
dbSNP, https://www.ncbi.nlm.nih.gov/projects/SNP/
Ensembl genome assembly GRCh37, http://grch37.ensembl.org/Homo_sapiens/Info/Index
ExAC Browser, http://exac.broadinstitute.org/
ExomeDepth, https://cran.r-project.org/web/packages/ExomeDepth/index.html
GATK, https://software.broadinstitute.org/gatk/
GenBank, https://www.ncbi.nlm.nih.gov/genbank/
GTEx Portal, https://www.gtexportal.org/home/
NHLBI Exome Sequencing Project (ESP) Exome Variant Server, http://evs.gs.washington.edu/EVS/
OMIM, http://www.omim.org/
Picard, http://broadinstitute.github.io/picard/
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
SIFT, http://sift.bii.a-star.edu.sg/
SnpEff, http://snpeff.sourceforge.net/
Splicing Finder, http://www.umd.be/HSF3/HSF.html
Swiss PDB Viewer, https://spdbv.vital-it.ch/

References

1. Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N. Engl. J. Med. *369*, 1502–1511.
2. Anazi, S., Maddirevula, S., Faqeih, E., Alsedairy, H., Alzahrani, F., Shamseldin, H.E., Patel, N., Hashem, M., Ibrahim, N., Abdulwahab, F., et al. (2017). Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Mol. Psychiatry *22*, 615–624.
3. Vissers, L.E., Gilissen, C., and Veltman, J.A. (2016). Genetic studies in intellectual disability and related disorders. Nat. Rev. Genet. *17*, 9–18.
4. Mirzaa, G.M., and Paciorkowski, A.R. (2014). Introduction: Brain malformations. Am. J. Med. Genet. C. Semin. Med. Genet. *166C*, 117–123.
5. Kato, M. (2015). Genotype-phenotype correlation in neuronal migration disorders and cortical dysplasias. Front. Neurosci. *9*, 181.
6. Spalice, A., Parisi, P., Nicita, F., Pizzardi, G., Del Balzo, F., and Iannetti, P. (2009). Neuronal migration disorders: clinical, neuroradiologic and genetics aspects. Acta Paediatr. *98*, 421–433.
7. Verrotti, A., Spalice, A., Ursitti, F., Papetti, L., Mariani, R., Castronovo, A., Mastrangelo, M., and Iannetti, P. (2010). New trends in neuronal migration disorders. Eur. J. Paediatr. Neurol. *14*, 1–12.
8. Parrini, E., Conti, V., Dobyns, W.B., and Guerrini, R. (2016). Genetic basis of brain malformations. Mol. Syndromol. *7*, 220–233.
9. Graziano, C., Wischmeijer, A., Pippucci, T., Fusco, C., Diquigiovanni, C., Nõukas, M., Sauk, M., Kurg, A., Rivieri, F., Blau, N., et al. (2015). Syndromic intellectual disability: a new phenotype caused by an aromatic amino acid decarboxylase gene (DDC) variant. Gene *559*, 144–148.

10. Ng, P.C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. Genome Res. *11*, 863–874.
11. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat. Methods *7*, 248–249.
12. Salgado, D., Desvignes, J.P., Rai, G., Blanchard, A., Miltgen, M., Pinard, A., Lévy, N., Collod-Béroud, G., and Béroud, C. (2016). UMD-Predictor: a high-throughput sequencing compliant system for pathogenicity prediction of any human cDNA substitution. Hum. Mutat. *37*, 439–446.
13. Alazami, A.M., Patel, N., Shamseldin, H.E., Anazi, S., Al-Dosari, M.S., Alzahrani, F., Hijazi, H., Alshammari, M., Aldahmesh, M.A., Salih, M.A., et al. (2015). Accelerating novel candidate gene discovery in neurogenetic disorders via whole-exome sequencing of prescreened multiplex consanguineous families. Cell Rep. *10*, 148–161.
14. Xue, S., Maluenda, J., Marguet, F., Shboul, M., Quevarec, L., Bonnard, C., Ng, A.Y., Tohari, S., Tan, T.T., Kong, M.K., et al. (2017). Loss-of-function mutations in LGI4, a secreted ligand involved in Schwann cell myelination, are responsible for arthrogryposis multiplex congenita. Am. J. Hum. Genet. *100*, 659–665.
15. Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nat. Methods *11*, 361–362.
16. Deciphering Developmental Disorders Study (2015). Large-scale discovery of novel genetic causes of developmental disorders. Nature *519*, 223–228.
17. Xue, Y., Ankala, A., Wilcox, W.R., and Hegde, M.R. (2015). 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. *17*, 444–451.
18. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
19. Sander, J.D., Maeder, M.L., Reyon, D., Voytas, D.F., Joung, J.K., and Dobbs, D. (2010). ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. Nucleic Acids Res. *38*, W462–W468.
20. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.
21. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.
22. Choi, Y., and Chan, A.P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics *31*, 2745–2747.
23. Jordan, D.M., Frangakis, S.G., Golzio, C., Cassa, C.A., Kurtzberg, J., Davis, E.E., Sunyaev, S.R., Katsanis, N.; and Task Force for Neonatal Genomics (2015). Identification of cis-suppression of human disease mutations by comparative genomics. Nature *524*, 225–229.
24. Desmet, F.O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., and Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. *37*, e67.
25. Brown, S.D., and Moore, M.W. (2012). The International Mouse Phenotyping Consortium: past and future perspectives on mouse phenotyping. Mamm. Genome *23*, 632–640.
26. Skarnes, W.C., Rosen, B., West, A.P., Koutsourakis, M., Bushell, W., Iyer, V., Mujica, A.O., Thomas, M., Harrow, J., Cox, T., et al. (2011). A conditional knockout resource for the genome-wide study of mouse gene function. Nature *474*, 337–342.
27. Verstreken, P., Ohyama, T., Haueter, C., Habets, R.L., Lin, Y.Q., Swan, L.E., Ly, C.V., Venken, K.J., De Camilli, P., and Bellen, H.J. (2009). Tweek, an evolutionarily conserved protein, is required for synaptic vesicle recycling. Neuron *63*, 203–215.
28. Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., et al.; GTEx Consortium (2015). Human genomics. The human transcriptome across tissues and individuals. Science *348*, 660–665.
29. Kok, F.O., Shin, M., Ni, C.W., Gupta, A., Grosse, A.S., van Impel, A., Kirchmaier, B.C., Peterson-Maduro, J., Kourkoulis, G., Male, I., et al. (2015). Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Dev. Cell *32*, 97–108.
30. Aase, J.M., and Smith, D.W. (1968). Dysmorphogenesis of joints, brain, and palate: a new dominantly inherited syndrome. J. Pediatr. *73*, 606–609.
31. Patton, M.A., Sharma, A., and Winter, R.M. (1985). The Aase-Smith syndrome. Clin. Genet. *28*, 521–525.
32. Harel, T., Yoon, W.H., Garone, C., Gu, S., Coban-Akdemir, Z., Eldomery, M.K., Posey, J.E., Jhangiani, S.N., Rosenfeld, J.A., Cho, M.T., et al.; Baylor-Hopkins Center for Mendelian Genomics; and University of Washington Center for Mendelian Genomics (2016). Recurrent de novo and biallelic variation of ATAD3A, encoding a mitochondrial membrane protein, results in distinct neurological syndromes. Am. J. Hum. Genet. *99*, 831–845.
33. Itoh, K., and Fushiki, S. (2015). The role of L1cam in murine corticogenesis, and the pathogenesis of hydrocephalus. Pathol. Int. *65*, 58–66.
34. Khuong, T.M., Habets, R.L., Slabbaert, J.R., and Verstreken, P. (2010). WASP is activated by phosphatidylinositol-4,5-bisphosphate to restrict synapse growth in a pathway parallel to bone morphogenetic protein signaling. Proc. Natl. Acad. Sci. USA *107*, 17379–17384.
35. Toth-Petroczy, A., Palmedo, P., Ingraham, J., Hopf, T.A., Berger, B., Sander, C., and Marks, D.S. (2016). Structured states of disordered proteins from genomic sequences. Cell *167*, 158–170.e12.
36. Miyamoto-Sato, E., Fujimori, S., Ishizaka, M., Hirai, N., Masuoka, K., Saito, R., Ozawa, Y., Hino, K., Washio, T., Tomita, M., et al. (2010). A comprehensive resource of interacting protein regions for refining human transcription factor networks. PLoS ONE *5*, e9289.
37. Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome in three quantitative dimensions organized by stoichiometries and abundances. Cell *163*, 712–723.