scientific reports

OPEN

Whole exome sequencing identifies *ABHD14A* and *MRNIP* as novel candidate genes for developmental language disorder

Amal Bouzid[1,2]✉, Malek Belcadhi[1,3], Amal Souissi[1], Meryam Chelly[1,11], Fakher Frikha[1,4], Hela Gargouri[1], Crystel Bonnet[5], Fida Jebali[1], Salma Loukil[1], Christine Petit[5,6], Saber Masmoudi[1], Rifat Hamoudi[2,7,8,9,10]✉ & Mariem Ben Said[1]✉

Developmental language disorder (DLD) is a neurodevelopmental disorder involving impaired language abilities. Its genetic etiology is heterogeneous, involving rare variations in multiple susceptibility loci. However, family-based studies on gene mutations are scarce. We performed whole-exome sequencing (WES) of a first-time-described Tunisian-family with DLD. Analyses of segregation patterns with stringent filtering of the exome data identified disease-causing compound heterozygous variants. In the *MRNIP* gene, two variants were detected including a synonymous low-frequency variant c.345G > C and a nonsense rare variant c.112G > A predicted pathogenic. In the *ABHD14A* gene, four variants were identified including a rare missense variant c.689T > G and three splice-site variants c.70-8C > T, c.282-25A > T and c.282-10G > C with low-frequency MAF < 5%. Complementary analyses showed that these variants are predicted pathogenic and the missense variant Leu230Arg significantly affects the stability and structure modelling of the ABHD14A protein. Biological functions and interconnections analyses predicted the potential roles of *ABHD14A* and *MRNIP* in neuronal development pathways. These results suggest *ABHD14A* and *MRNIP*, as putative candidate genes for DLD susceptibility. Our findings reveal the involvement of novel candidate genes in the genetic etiology of DLD and explore the potential future utility of WES in the diagnosis of such complex disorders.

Keywords Developmental language disorder, whole exome sequencing, genetic etiology, candidate gene, *ABHD14A*, *MRNIP*

Abbreviations

DLD	Developmental language disorder
SLI	Specific language impairment
CNV	Copy number variants
WES	Whole-exome sequencing
OAEs	Pure-tone Audiometry and Otoacoustic Emissions
BAER	Brainstem auditory evoked response
MAF	Minor allele frequency
RMSF	Root-Mean-square Fluctuation

[1]Laboratory of Molecular and Cellular Screening Processes, Center of Biotechnology of Sfax, Sfax, Tunisia. [2]Research Institute for Medical and Health Sciences, University of Sharjah, Sharjah, United Arab Emirates. [3]Department of Otorhinolaryngology, Farhat Hached University Hospital, Sousse, Tunisia. [4]Faculty of Sciences of Sfax, Department of Biology, University of Sfax, Sfax, Tunisia. [5]Université Paris Cité, Institut Pasteur, AP-HP, Inserm, Fondation Pour l'Audition, Institut de l'Audition, IHU reConnect, Paris F-75012, France. [6]Collège de France, Paris F-75005, France. [7]College of Medicine, University of Sharjah, Sharjah, United Arab Emirates. [8]Division of Surgery and Interventional Science, University College London, London, UK. [9]ASPIRE Precision Medicine Research Institute Abu Dhabi, University of Sharjah, Sharjah, United Arab Emirates. [10]BIMAI-Lab, Biomedically Informed Artificial Intelligence Laboratory, University of Sharjah, Sharjah, United Arab Emirates. [11]Department of Engineering, University of Messina, C.da Di Dio, I-98166 Messina, Italy. ✉email: abouzid@sharjah.ac.ae; rhamoudi@sharjah.ac.ae; mariem.bensaid@cbs.rnrt.tn

Developmental language disorder (DLD), also referred to a specific language impairment (SLI)[1], is a neurodevelopmental disorder characterized by persistent difficulties in the communication ability that interferes with using language, understanding and learning, although developing normally in all other areas. DLD recurrently persists from early childhood into adulthood[2]. It is associated with emotional, social, behavioral, and literacy difficulties[3]. DLD is a common, but relatively under-investigated and extremely underdiagnosed developmental disorder[4]. Its prevalence ranges from 5 to 8% among preschool children[5,6]. A combination of risk factors including a family history of language disorders, socio-economic difficulties, and late talker status has been considered to increase the risk of developing DLD in children[7,8]. Because specific deficit characteristics of various DLD subtypes persist into adulthood[9], early detection and intervention are crucial to mitigate long-term negative consequences of DLD[10]. In the current work, further to a consensus agreement on terminology, the term DLD was considered to describe children with developmental language difficulties across many modalities of language and communication (including speech) that are not co-occurring with any intellectual disability, other neuro-developmental disorders, or hearing impairment[11,12].

In addition to the considerable clinical heterogeneity of language disabilities, language development is largely affected by environmental and genetic factors[8,12]. Several approaches with distinct evidence have been explored in order to understand the genetic features of language impairments. Linkage analysis studies, which examine genetic markers within family members' cohorts to identify inherited chromosomal loci associated with the disease, underlined several susceptibility loci to DLD including *SLI1* (OMIM*606711), *SLI2* (OMIM*606712), *SLI3* (OMIM*607134), *SLI4* (OMIM*612514), and *SLI5* (OMIM*615432)[13–15]. However, linkage studies showed limitations particularly in mapping genes and causative variants within the identified regions along with in replicating the findings in other populations[14]. With the same assumption that a few common variants are contributing to the disease but with a larger sample size, higher resolution, and greater statistical power, the genome-wide association studies (GWAS), that screen the whole genome for the association of common variants with the trait, identify more robust and replicable findings than linkage studies[16,17]. For instance, GWAS in patients with language impairment identify potential risk variants in *NDST4*[18], *ZNF385D*[18], *COL4A2*[18], and *NOP9*[14,19] genes. Candidate gene association analysis presents another approach that concerns particular genes selected on prior biological knowledge or hypothesized involvement in the DLD or language traits. Although these studies are targeted and cost-effective, they have limitations in finding significant results[20] and may show false-positive results[21]. Moreover, rare variant analyses, conducted in well-characterized family patients or unique populations, identify less common genetic variants but with a larger effect on the trait[22,23], such as *NFXL1*[24] which presents a key example of exploring family-based DLD causative variants identification. Additionally, the recent developments in genomic sequencing using whole-exome sequencing (WES), which enables the simultaneous analysis of all genes in one experiment, have extended the discovery of causative genetic variations and further supported the evidence of candidate genes in DLD including *SETBP1*[25], *GRIN2A*[26], *SRPX2*[26], *ERC1*[26], *ARHGEF39*[27], *BUD13*[28], *TM4SF20*[29], and *NFXL1*[29]. Although the rare variants analysis approach gives strong evidence of causality for the variants segregating with the disorder, only a few of them have been replicated due to the rarity of the variations and the necessity of large sample sizes or similar well-characterized families[30]. On the other hand, chromosomal abnormalities and copy number variants (CNV) have been related to language impairment etiology[31,32], showing that CNV may result in different outcomes. Although these findings are promising as they highlight the growing identification rate of genetic variations associated with DLD and emphasize the significance of its genetic etiology, the replication of specific genetic signatures is essential to deepen our understanding of the biological mechanisms underlying DLD.

Genetic background factors that vary by language, race, and population ethnicity have been suggested to influence the manifestation of DLD[1]. Most previous studies on DLD have predominantly focused on English-speaking families from Western cultures, overlooking genetic variations that might influence language impairments specific to this demographic[6,33]. However, these observations may be limited to be generalized to universal populations. On another hand, some critical studies have emerged from other racial backgrounds of non-English speaking and/or non-Westernized populations involving for instance *TM4SF20* and *NFXL1* as candidate genes of language impairments. Of note, the *TM4SF20* gene was identified through genome-wide analysis in children with behavioral or motor developmental delays from numerous ethnicities, and a complex deletion on a codon exon of *TM4SF20* was reported to co-segregate with childhood communication disorders in Southeast Asian families[34]. Coding variants in the *NFXL1* gene were particularly identified in an isolated founder Chilean population from Robinson Crusoe Island[21]. By studying a non-Western family in our research, we aimed to highlight the importance of examining DLD across a specific demographic of a not previously explored population. This approach not only enriches our understanding of the genetic backgrounds involved in DLD but also emphasizes the need for inclusive research that considers diverse behaviorally and genetically populations.

In the present study, we reported a two-generation Tunisian North African family with a complex genetic background form of DLD. Using WES, we aimed to identify inherited variations and expand our knowledge of the pathophysiology and etiology of developmental language disorders.

Results

Clinical presentation and phenotypes

For the studied two-generation Tunisian family, the parents of the probands gave a detailed history, covering first concerns and early history until the day of investigation, and thus revealing that their siblings interact with them and with unfamiliar persons like teachers and friends at school. The speech-language pathologist confirmed using the "Thiberge" test, the non-word repetition test and the sentence repetition test that the three siblings struggle to find the words to express ideas and have trouble organizing their sentences. The probands are educated but they have difficulty in reading and writing. Their non-word repetition test (NWRT) standardized

scores were ranged between 52 and 57 (Table 1). For the sentence repetition test, the analysis was without scores. The parents' speech-language examinations indicated no subtle signs of language disorder, and their language abilities were found to be in a typical average range of 99–105 for the NWRT (Table 1). The test of nonverbal intelligence, second edition (TONI-2) recorded only in the three siblings showed typical scores ranging between 90 and 106. Additionally, according to a psychiatrist examination, the three siblings did not display repetitive behaviours or restricted and fixed interests suggesting that Autism Spectrum Disorder (ASD) is excluded. Moreover, the results from the parental examination of the ASD screening also showed no indications of ASD. The audiological tests showed no hearing loss indicating that the language disorder in the three probands does not result from sensory hearing impairment. Indeed, the EMG denoted the absence of a genetic peripheral neuropathy associated with hearing loss in the three siblings. In addition, the MRI scans excluded the presence of congenital cochlear malformation of the Mondini type (*Mondini dysplasia*) in the three siblings which consists of an inner ear abnormality characterized by an incomplete cochlea development. Since such hearing perturbations or malformation in the inner ear can significantly affect hearing and language development, their exclusion ensures a homogeneous auditory function among the three siblings. No further clinical features were reported in the three siblings. On the other hand, during the interviews, the parents excluded the fact that other distant family members might have exhibited signs of DLD or related language impairments, although these were not formally diagnosed. Taken together, for the studied two-generation Tunisian family, the clinical and phenotype assessments of DLD relied upon the presence of a serious language disorder in the three siblings which was not caused by any intellectual disability, a sensory deficit, a malformation of the speech organs, or an ASD.

Genetic findings

The WES showed similar qualities among the five family members including the coverage of reads ranging from 25X to 85X. Using WES, and applying the selection criteria described in the methods, we identified compound heterozygous variants segregating within the family, particularly in two candidate genes: *ABHD14A* (Fig. 1a) and *MRNIP* (Fig. 1b). These variants were present in the heterozygous state in the three affected siblings and either in the heterozygous state or in the wild state in the parents. In the *ABHD14A* gene, we identified four variants including one rare missense variant c.689T > G (p.Leu230Arg) with minor allele frequency (MAF) of 0.0021 and three splice-site variants c.70-8C > T, c.282-25A > T and c.282-10G > C at low-frequency 1% ≤ MAF < 5%. While in the *MRNIP* gene, two variants were detected including a synonymous low-frequency variant c.345G > C (p.Ser115Ser), and a nonsense rare variant c.112G > A (p.Gln38Ter) with a MAF of 0.01 and 0.003, respectively. The details of the genetic findings are listed in Table 2. The CNV analysis did not show any significant changes (p-value > 0.05) that segregated with the DLD phenotype among the family members.

Family member		III.1 (Child1)	III.2 (Child2)	III.3 (Child3)	II.1 (Father)	II.2 (Mother)
Sex		Male	Male	Female	Male	Female
Age (years)		13	11	10	47	40
Speech-language pathologist examination	General examination	The "Thiberge" test was undertaken. Patients were examined several times using non-word repetition tests as well as sentence repetition tests. Patients struggle to find the words to express ideas and have trouble organizing their sentences.			Absence of language disorder	
	Non-Word Repetition Test (NWRT)	52	55	57	105	99
Psychiatrist examination	Test of Nonverbal Intelligence (TONI-2)	94	90	106	NR	NR
	Autism Spectrum Disorder (ASD)	Absence of ASD			Absence of ASD	Absence of ASD
Pediatrics examination	Otorhinolaryngology examination	Normal			NR	NR
	Pure-tone Audiometry	Normal			NR	NR
	Otoacoustic Emissions (OAEs)	Present			NR	NR
	Brainstem Auditory Evoked Response (BAER)	Normal			NR	NR
	Electromyography of lower and upper limbs (EMG)	Normal			NR	NR
	Brain Magnetic Resonance Imaging (MRI)	Normal			NR	NR
Education		Educated Difficulty in reading and writing			NR	NR

Table 1. Clinical features of the studied family with a developmental language disorder. *NR* Not Recorded. All measures are standardized with a mean of 100 and a Standard Deviation of 15.

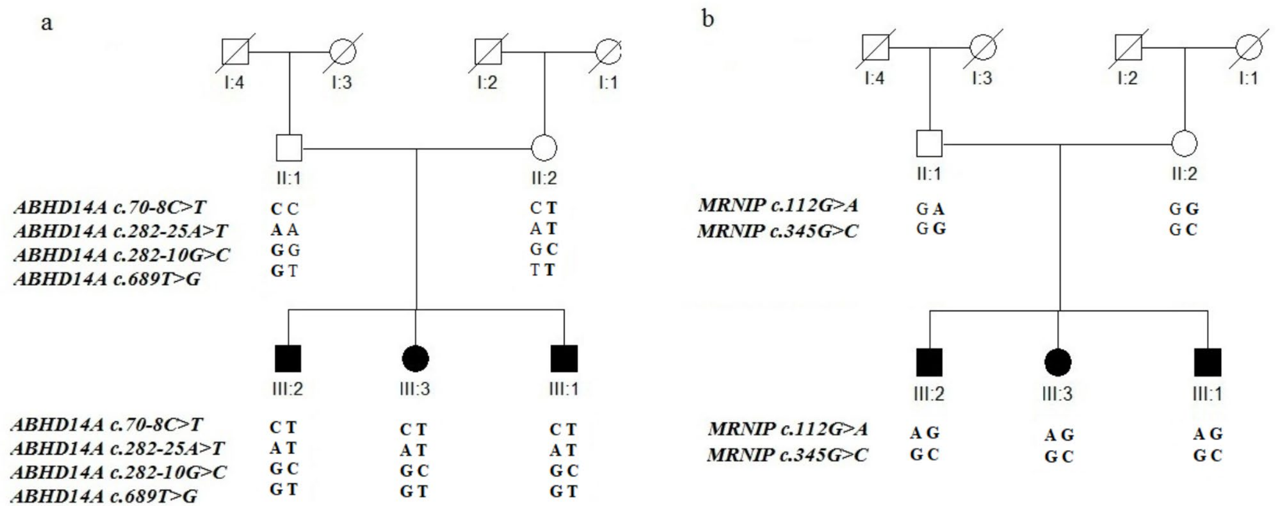

Fig. 1. Pedigree and relevant genetic findings of a Tunisian family with DLD. The filled symbols represent affected members. The patients (III.1, III.2, III.3) have compound heterozygous mutations in (**a**) *ABHD14A* and (**b**) *MRNIP* genes.

Genomic position (hg19)	Gene name	dbSNP_ID	MAF in the general population (ExAC v1.0)	Variation in cDNA	Variation in protein
chr3.52011879 C > C/T	*ABHD14A*	rs411209	0.045	c.70-8C > T	–
chr3.52012265 G > G/C		rs75627697	0.013	c.282-10G > C	–
chr3.52012250 A > A/T		rs380788	0.036	c.282-25A > T	–
chr3.52014907 T > T/G		rs34149506	0.0021	c.689T > G	p.Leu230Arg
chr5.179269011 G > G/C	*MRNIP (C5orf45)*	rs61736195	0.01	c.345G > C	p.Ser115Ser
chr5.179280392 G > G/A		rs112867427	0.0026	c.112G > A	p.Gln38Ter

Table 2. The most relevant variants identified by WES in the studied Tunisian family with developmental language disorder. *MAF* Minor Allele Frequency.

The splice variants have a higher disruptive impact on the *ABHD14A* gene function

Different *in-silico* tools were applied to predict that the splice-site variants identified in the *ABHD14A* gene are considered damaging (Table 3). Noteworthy, the three splice-site variants had a DANN score > 0.5 suggesting a pathogenic effect. Additionally, the Spliceman Ranking scores were > 50% signifying that the point mutations disrupt the splicing. In particular, for chr3.52011879 C > C/T and chr3.52012265 G > G/C, the MaxEntScan scores were positive, about 11.4 and 11.10 respectively, and the NNPLICE scores were close to 1 for both variants suggesting a high potential for a functional splice site. Indeed, using the PhyloP100way score, GERP and PhastCons100way, the genomic positions of the three splice-site variants in *ABHD14A* are significantly conserved among species (Table 4). These findings would seem to suggest that these splice variants have a higher disruptive impact on the *ABHD14A* gene function.

Missense and nonsense variants of higher risk in the DLD family

The theoretical potential pathogenicity of the missense/nonsense variants was also evaluated and scored accordingly by multiple bioinformatics prediction programs as recommended by MacArthur et al.[35]. Damaging variants were defined as those having a deleterious or damaging effect in at least two prediction tools. The nonsense variant c.112G > A (p.Gln38Ter) in the *MRNIP* gene was predicted pathogenic with MutationTaster (disease-causing), CADD (score = 44), EIGEN (pathogenic) and FATHMM-MKL (damaging) prediction programs. While, importantly, the unique missense variant c.689T > G (p.Leu230Arg) in the *ABHD14A* gene was predicted pathogenic with all nine used prediction tools suggesting its probable cause effect in the DLD etiology of the studied family. Noteworthy, these pathogenic variants are predicted evolutionary conserved suggesting their higher risk in the genetic etiology of DLD susceptibility. The variants' information with their representative predictions are summarized in Table 4.

The Leu230Arg variation significantly affects the stability and structure modelling of the ABHD14A protein

Given the potential function that *ABHD14A* could play in DLD and the higher risk of the potential pathogenic impact of the missense variant c.689T > G (Leu230Arg) in the *ABHD14A* gene, we further evaluated the effect

Variant	chr3.52011879 C > C/T	chr3.52012265 G > G/C	chr3.52012250 A > A/T
dbSNP.ID	rs411209	rs75627697	rs380788
Consequence (cDNA)	c.70-8C > T	c.282-10G > C	c.282-25A > T
MAF (ExAC v1.0)	0.045	0.013	0.036
DANN score[(1)]	***0.5453***	***0.5113***	***0.6791***
CADD PHRED score[(2)]	8.769	9.172	3.874
Spliceman Ranking[(3)]	***60%***	***72%***	***53%***
MaxEntScan score[(4)]	***11.40***	***11.10***	–
NNPLICE score[(5)]	***0.98***	***0.94***	–
CRYP-SKIP prediction[(6)] (Probability of cryptic splice site activation)	***0.73***	***0.55***	***0.55***
Splice Rover prediction	–	***A possible donor splice site***	–
PhyloP100way score[(7)]	***1.06***	− 0.333	− 0.41
GERP[(8)]	***5.69***	***5.8299***	***4.09***
PhastCons100way[(9)]	***0.104***	***0.055***	0

Table 3. *In-silico* prediction of the splice site variants effects in the *ABHD14A* gene. All the prediction results in favor of the pathogenicity of the variant are marked in bold and italic. *chr* chromosome, *dbSNP* Single Nucleotide Polymorphism Database, *MAF* Minor Allele Frequency, *DANN* Deleterious annotation of genetic variants, *CADD* Combined annotation-dependent depletion, *PhyloP100way* measure evolutionary conservation at individual alignment sites of 100 vertebrate species, *GERP* Genomic Evolutionary Rate Profiling, *PhastCons100way* measure conservation from genome-wide multiple alignments with 100 vertebrate species. [(1)]Variants are predicted to be pathogenic by a score > 0.5. [(2)]A score greater or equal to 10 indicates that these are predicted to be the 10% most deleterious substitutions that you can do to the human genome. [(3)]The higher the percentile rank, the more likely the point mutation is to disrupt splicing. [(4)]A positive value predicts a functional splice site. [(5)]A higher score implies greater potential for splice site ((Score (0–1)). [(6)]Values are between zero and one, with higher values speaking in favour of cryptic splice-site activation. [(7),(9)]Positive scores measure conservation, which is slower evolution than expected, at sites that are predicted to be conserved. [(8)]Score ranges from − 12.3 to 6.17, with 6.17 being the most conserved.

of the Leu230Arg missense variant on its protein properties, and protein structure modelling. Noteworthy, both the structure models of the ABHD14Aw and ABHD14Am have a globular shape with an α/β architecture. The core structure consists of 8 β-strands β-sheet (β1- β2- β4- β3-β5- β6-β7-β8) surrounded by α-helices αA and αF on one side and α-helices αB–αC-αD–αE on the other side (Fig. 2A). The residue Leu230Arg was located in the αE helix and showed an h-bond with Ala226 and Arg234 through a contact energy of -4.146 kcal/mol and an accessible surface area of 107.4 Å2 (47.35% of exposition). The Arg230 residue displayed an h-bond with Ala226 through a contact energy of -1.041 kcal/mol and an accessible surface area of 165.7 Å2 (61.34% of exposition). Moreover, the molecular dynamics simulation was carried out for 10 ns for both the wild-type ABHD14Awt and mutated ABHD14Amt-Leu230Arg forms. To quantify the dynamic conformational changes throughout the simulations, the Cα-atoms root-mean-square deviation (RMSD) of simulated snapshots relative to the original structure was calculated. The two systems became relatively stable after 2 ns simulations, which implied that the dynamic process reached equilibrium states. During the last 8 ns, the RMSD values for the wild-type and mutated forms were 0.226 ± 0.016 nm and 0.149 ± 0.016 nm, respectively, demonstrating a significant difference (p-value < 0,0001) between the wild-type ABHD14Awt and mutated ABHD14Amt-Leu230Arg forms (Fig. 2B). Indeed, the Root-Mean-square Fluctuation (RMSF) was calculated to reveal the differences in the fluctuation of local regions for the two systems. Importantly, the ABHD14Amt-Leu230Arg mutant form displayed a low RMSF, especially for the residues Leu89-Ala92 and Pro139-Leu153, suggesting that the mutated form was more stable (Fig. 2C). Furthermore, the trajectory of the ABHD14Awt and ABHD14Amt model generated with GROMACS software was then used to perform a DCCM analysis. The latter showed that the Leu230Arg variant resulted in a decrease of correlated motions (correlation and anticorrelation), especially for the residues near Pro139-Leu153 (red box, Fig. 2D). Taken together, *ABHD14A* was suggested as a novel and potential candidate gene for DLD.

Biological functions and interconnections of the candidate genes

To better explore the interconnections between the identified candidate genes including *ABHD14A*, and *MRNIP (C5orf45)*, a functional gene interaction was explored based on physical interactions, co-expression, predicted, co-localization, common pathway, genetic interactions and shared protein domains using the GeneMANIA database (Fig. 3). Furthermore, functional annotations including related pathways and ontologies of biological process and molecular function were performed to assess the functional enrichment of *ABHD14A*, and *MRNIP(C5orf45)*, and their co-operators. The gene interaction networks showed two main subnetworks involving *ABHD14A*, and *MRNIP* with an indirect link between them suggesting that DLD in this family could be linked to different inter-gene interactions and combinations of genes resulting from multi-biological factors. Moreover, *ABHD14A* and *MRNIP* simultaneously appeared in the co-expression network of *SH2B1*, *ECHS1*, *APPB3* and *CTSA* indicating the presence of functional interactions between them. Importantly, *MRNIP* was found physically interconnected with *MRE11*, *RAD50* and *NBN* (*NBS1*) genes which form the *MRE11/RAD50/*

Variant (hg 19)	chr3.52014907 T > T/G	chr5.179269011 G > G/C	chr5.179280392 G > G/A
Gene name	*ABHD14A*	*MRNIP*	*MRNIP*
dbSNP.ID	rs34149506	rs61736195	rs112867427
Variation cDNA	c.689T > G	c.345G > C	c.112G > A
Variation protein	p.Leu230Arg	p.Ser115Ser	p.Gln38Ter
PROVEAN Prediction	***Deleterious***	Neutral	–
SIFT Prediction	***Damaging***	Tolerated	–
PolyPhen2 Prediction	***Probably damaging***	–	–
MutationTaster Prediction	***Disease-causing***	Polymorphism	***Disease-causing***
MutPred2 Prediction(1)	***Score = 0.57***	–	–
CADD Prediction(2)	***Score = 25.9***	Score = 4.106	***Score = 44***
EIGEN Prediction	***Pathogenic***	–	***Pathogenic***
FATHMM-MKL Prediction	***Damaging***	–	***Damaging***
M-CAP Prediction	***Damaging***	–	–
PhastCons100way score(3)	***1.00***	0	***1.000***
PhyloP100way(4)	***4.04***	− 1.092	***4.601***
GERP(5)	***5.40***	***4.28***	***4.03***

Table 4. Predicted effects of the DLD candidate variants using different *in-silico* tools. . (1)A score threshold of 0.50 would suggest pathogenicity. (2)A score greater or equal to 10 suggests that 10% of most deleterious substitutions can be acquired to the human genome. (3),(4)Positive scores measure conservation, which is slower evolution than expected, at sites that are predicted to be conserved. (5) Score ranges from − 12.3 to 6.17, with 6.17 being the most conserved. All the prediction results in favor of the pathogenicity of the variant are marked in bold and italic. chr : chromosome; dbSNP : Single Nucleotide Polymorphism Database; PROVEAN: Protein Variation Effect Analyzer; SIFT: Sorting Intolerant From Tolerant; PolyPhen2: Polymorphism Phenotyping v2; MutPred2 : Predictor of impactful missense variants; CADD: Combined annotation-dependent depletion; EIGEN : A spectral approach to the functional annotation of genetic variants in coding and noncoding regions. FATHMM-MKL: Predicting the functional consequences of both coding and non-coding sequence variants; M-CAP: PhastCons100way: measure conservation from genome-wide multiple alignments with 100 vertebrate species. PhyloP100way : Measure evolutionary conservation at individual alignment sites of 100 vertebrate species; GERP: Genomic Evolutionary Rate Profiling;

NBS1 (MRN) complex responsible for DNA double-strand breaks detection and DNA damage response in the ATM signaling pathway. In addition, the functional annotations mainly identify *MRNIP* to be responsible for DNA damage response (GO:0006974). Taken together, these findings point out that *MRNIP* accounts through its association with the MRN damage-sensing complex in ATM-mediated DNA damage response and cell cycle regulation which could have implications in neuronal development pathways and brain circuitry. On another hand, *ABHD14A* appeared in close interconnections with *GSS*, *NRG2* and *BZW2* which are involved in the nervous system development. The functional annotation showed that the *ABHD14A* gene plays hydrolase activity with a possible role in granule neuron development. These findings further suggest that extensive disruption of ABHD14A, and MRNIP proteins may result in multi-risk factors behind the pathophysiology and etiology of developmental language disorders. Therefore be treated with considerable, these results may be specific to the particular studied family, especially considering the intricate nature of DLD genetics. The detailed results of gene interconnections with their corresponding weight in the gene interaction networks are reported in the Supplementary Table.

Discussion

Because of the high genetic heterogeneity of language disorders and the complexity of inheritance, the genetic factors are poorly explained. Thus, it is of high interest to identify novel DLD candidate genes in non-explored populations which will help to better understand its genetic etiology through stringent familial genotyping–phenotyping correlation studies.

In the present work, we investigated the first Tunisian family with DLD using the WES approach. An autosomal recessive mode of inheritance was suggested from the family's pedigree including healthy parents with affected siblings. Hence, affected offspring that manifest an autosomal recessive disorder should be homozygous for the disease-associated allele while the unaffected parents should be both carriers (heterozygous). However, if the parents are non-consanguineous, which is the case of our study, the most likely explanation of the recessive form is compound heterozygosity for two or more different pathogenic variants[36]. A further important implication in the interpretation of the genetic factors in DLD is that this particular neurodevelopmental disorder is likely considered more genetically heterogeneous compared to other specific developmental language phenotypes[4]. Noteworthy, our parental phenotyping assessment, which included a Speech-Language general examination along with a Non-Word Repetition Test (NWRT), did not reveal troubles in their communications and thus ruled out any subtle signs of DLD. On the other hand, during the interviews, the parents excluded the fact that

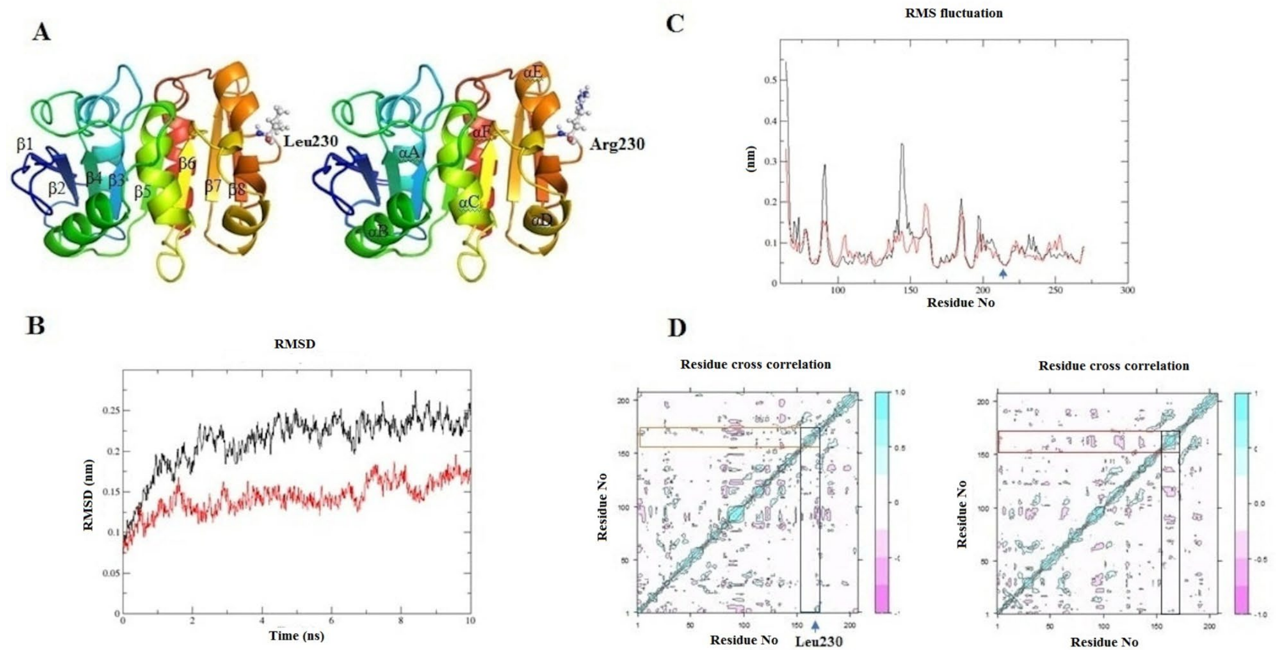

Fig. 2. Protein structure modelling of Leu230Arg variant in ABHD14A. (**A**) Cartoon representation of ABHD14Aw (left panel) and ABHD14Am (right panel) model structures. The structure is colored by amino acid succession, respectively. The L230R is shown as a ball and sticks. The secondary structure element is indicated. (**B**) The RMSF values of the protein's backbone throughout the simulations, where the ordinate is the RMSF (nm), and the abscissa is the residue number. ABHD14Awt is presented in black color while ABHD14Amt is presented in red color. (**C**) RMSD of the protein backbone. ABHD14Awt is presented in black color while ABHD14Amt is presented in red color. (**D**) Dynamic cross-correlation matrix (DCCM) for Cα atom pairs calculated with DCCM function and plotted with plot function of wild type (left panel) and mutated form (right panel). Blue stands showed correlation and pink stands showed anticorrelation.

other distant family members might have exhibited signs of DLD or related language impairments. Given these were not formally diagnosed, it is possible that subtle manifestations of the disorder were missed in distant family members which may be due to the challenges of diagnosing DLD in adulthood[37], suggesting that DLD might have a broader familial impact. This hypothesis highlights the fact that DLD can have variable expressivity and incomplete penetrance, meaning that the genetic predisposition to DLD might not always manifest in the same way across different family members.

In our analysis, we applied a tightly constrained exploration for potential variants, using several strategies. We started by assessing variants occurring within known candidate genes that have previously been identified as susceptibility factors in language disorders however, no potential variants were identified. Next, we characterized low-frequent, rare or novel variants of possible high-risk variants following autosomal recessive or compound heterozygosity inheritance forms. Moreover, in the variants prioritization, we referred to the guidelines for investigating the likely causality of non-synonymous variants, as demonstrated by MacArthur and colleagues[35]. Finally, we observed likely multiple-hit events of variants by searching for more than one potential change across several genes. Our analysis strategy appears to be well substantiated by a previous study showing the identification of novel genes involved in specific language impairment using WES and similar analysis strategies[26].

Using the WES approach, we identified disease-causing compound heterozygous variants that co-segregate with the clinical DLD phenotype in the studied family. It is worthwhile noting that the parents are non-consanguineous, which explains why there were no rare homozygous variants detected. Indeed, compound heterozygous mode of transmission is the condition of having two or more heterozygous variants at a particular locus that can cause genetic disease in a heterozygous state. A variant that is heterozygous in an affected offspring should be particularly heterozygous in one of the parents. Additionally, the variant must be present in at least one of the parents, and it cannot be heterozygous in both parents[36]. Interestingly, these conditions are met by all the variants kept in the two genes including *ABHD14A*, and *MRNIP*. Previous researchers have shown that homozygous or compound-heterozygous gene disruptions explain the disease causality in language disorders[38]. To the best of our knowledge, only one previous study has identified compound heterozygous variants in candidate genes including the *FAT3*, *KMT2D*, *SCN9A*, and *PALB2* genes in unrelated DLD probands[26]. Moreover, Bieder et al. reported compound heterozygous mutations in the *DNAH5* gene in a case study with dyslexia and situs inversus[39]. Other researchers provide additional support for compound heterozygous variants' contribution to the heritability of other communication disorders such as Autism Spectrum Disorder[40,41].

In our variant prioritization analysis strategy, we were less stringent in adopting the minor allele frequency cut-off (MAF < 5% in the general population) that a potential causal variant can have in DLD for many reasons; First,

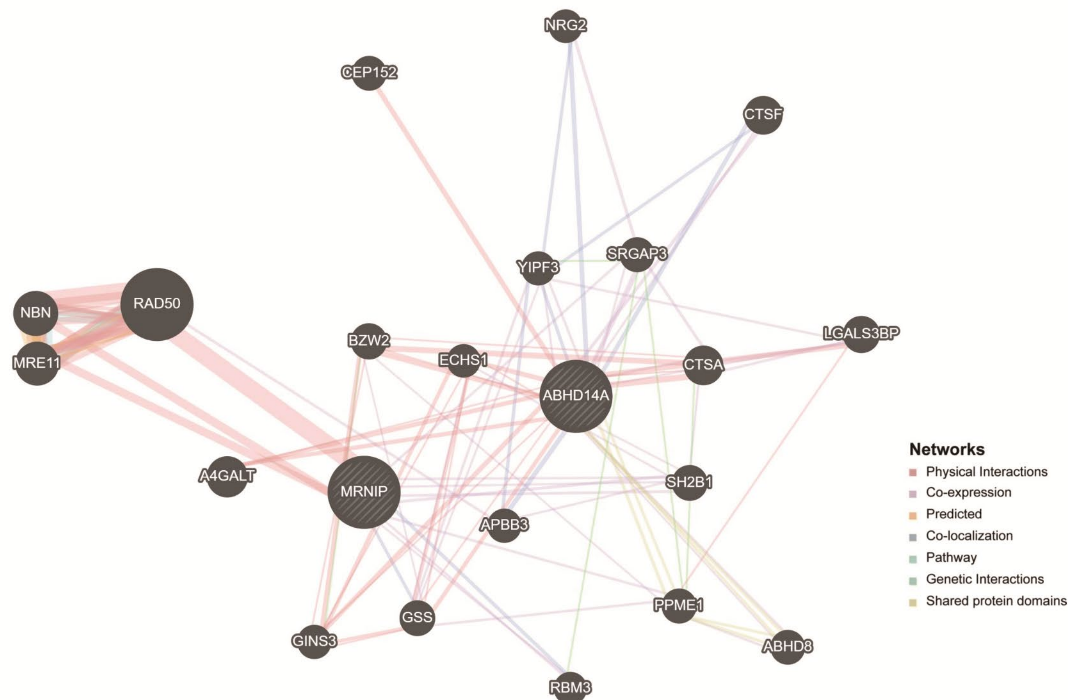

Fig. 3. Gene interaction networks involving *ABHD14A* and *MRNIP (C5orf45)* genes. Gene-gene interaction network created and visualised with Cytoscape, displaying interaction type (colour), interaction strength (edge thickness), multiple edges between nodes, and protein score (node size).

exploring the genetics landscape of DLD through Mendelian inheritance patterns is considered challenging due to its complex modes of inheritance and likely variable penetrance[42]. The hypothesis of our study suggested that the identified variants follow a compound heterozygote inheritance mode. This inheritance pattern supports the pathogenic potential of these variants, even if each variant individually has a higher frequency in the population[26]. Additionally, DLD, like many language disorders, exhibits incomplete penetrance and genetic heterogeneity[33,43]. This means that not all individuals with a causing variant will express the disorder, and there may be multiple genetic contributors. This complexity allows for the possibility that variants with higher population frequencies can still play a role in the disease, particularly when combined with other genetic or environmental factors[30]. On the other hand, our study cohort may have unique genetic backgrounds underrepresented in broader databases. This underrepresentation can result in an overestimation of the allele frequencies of the identified variants for the general population[44]. Given these considerations, along with additional functional and computational data analyses, our study devised a strategy that can help in identifying plausible candidate variants in complex developmental disorders such as DLD.

Our findings showed that low-frequency and rare variants were identified in the two genes including *ABHD14A*, and *MRNIP*, and co-segregated with the DLD phenotype. These results correlate favorably with Graham and al[38] and further support the idea that in order to avoid misattribution of causality to candidate variants, it is crucial to understand the segregation patterns and variant frequencies in the general population. Particularly in our filtering strategy, synonymous variants were not excluded as they have been described in many complex diseases, for which synonym variants could disrupt transcription splicing, co-translational folding, and mRNA stability, and entail a plethora of other functionally relevant changes[45].

Interestingly in the *ABHD14A* gene, four variants were identified including a rare missense variant (p.Leu230Arg) with MAF of 0.0021 and three splice-site low-frequency variants (c.70-8C > T; c.282-25A > T; c.282-10G > C) with (1% ≤ MAF < 5%). *ABHD14A*, Abhydrolase Domain Containing 14 A, is coding for a protein with 271 amino acids which has a conserved alpha/beta hydrolase fold. Even though its function has not yet been fully understood, it was presumed that *ABHD14A* has a possible role in granule neuron development. The mouse *Abhd14a* gene, also called Dorz1, is the human ortholog that shares 82.2% amino acid homology. Earlier studies showed that *Abhd14a* was expressed in mouse cerebellum at different developmental stages from embryonic day 12.5 through 6 weeks of age[46]. In addition, the *Abhd14a* expression was demonstrated to be positively regulated by Zics[46] which encodes a zinc finger protein of cerebellum 1 expressed in the cell lineage of cerebellar granule neurons and controls vertebrate neural development[47]. In a case-report study, it has been demonstrated that gain-of-function mutations in the *ZIC1* gene were associated with learning disability and additional severe characteristic features[48]. Indeed, several studies highlighted that functionally specialized brain regions play crucial roles in language processing[49,50]. Den Hoed and Fisher emphasized that rare variants, as well as emerging cell-based models of human brain development, are intrinsic in terms of a deeper and more thorough understanding of how gene disruptions cause speech disorders (https://www.sciencedirect.com/scie

nce/article/pii/S0959437X20300630). Taken together, it may be assumed that the identified genetic variations are additional variants in the *ABHD14A* gene with more strong evidence for their involvement in language development and processing impairment in this Tunisian family.

Our findings provided further evidence for the possible involvement of the identified *ABHD14A* variants in the DLD etiology. For instance, the Leu230Arg missense variant was predicted to be damaging as well as conserved among species. Furthermore, the functional analysis using RMSD, RMSF and cross-correlation analysis showed that the change of Leucine by Arginine at position 230 may affect the ABHD14A protein dynamics and consequently its function. Moreover, the three splice-site variants in the *ABHD14A* gene demonstrated a possible role in altering splicing phenotypes and disturbing RNA splicing. Using homozygous haplotype mapping, Casey et al. suggested that *ABHD14A* is a candidate gene for Autism Spectrum Disorder[51]. Several propositions indicate that children with Autism Spectrum Disorder are much more similar to those with DLD. However, there are complex and ambiguous cases of figures, particularly when DLD has significant repercussions on the interactions of young children, generating relational disturbances, which make DLD predominantly difficult to distinguish from an autism phenotype[52].

Our biological functions and gene networking analyses demonstrated that *MRNIP* was found physically interconnected with potential players of the ATM signaling pathway including *MRE11*, *RAD50*, and *NBN* (*NBS1*) genes of the MRN complex. It has been reported that the functional disturbing of ATM (ataxia-telangiectasia mutated) gene products results in ataxia-telangiectasia (A-T) which is a rare inherited childhood complex neurological disorder that disturbs the part of the brain controlling motor movement and speech[53]. In this disease, the phenotype is deduced from homozygosity or compound heterozygosity for the causing genes (ATM) leading to the destabilization of the mRNA and/or the protein[54]. Moreover, Volkow and colleagues showed that both ATM homozygotes and heterozygotes patients presented reduced glucose metabolism in diverse brain regions, including the hippocampus, which is the brain area central to memory and learning processes[55]. These observations would seem to conceivably suggest that the compound heterozygote variants identified in the *MRNIP* gene may explain the complex language disability phenotype found in our studied DLD family members compared well with the genetic etiology of A-T disorder. Although the studied DLD family members do not have any motor movement disability, their neurological symptoms including the language disorder could be explained by the destabilization of the ATM signaling pathway through the *MRNIP* gene misfunction. It has been recently demonstrated that upon DNA double-strand breaks formation, MRNIP condensates concentrate the MRN complex inducing the autophosphorylation of ATM and consequently resulting in an accelerated DNA damage response and end resection[56]. It was widely demonstrated that deficiency in DNA repair has crucial pathophysiological implications resulting in cognitive impairments and neurological disorders[57]. Together, these findings highlight how genetic defects in DNA damage repair, in particular, proteins involved in the ATM molecular network, such as the identified *MRNIP* gene can play a fundamental role in cognitive and neurodevelopmental disorders and may improve knowledge about the pathophysiology of DLD.

Although the relatively small sample size of our studied family, this limitation offers an outline to maximize the assumption of finding potential variants in candidate genes. This hypothesizes that the identified variants/genes will reflect a proportion of the trait variance that could be implicated in the DLD etiology. Moreover, further investigations and functional validations with cells and using in *vivo model* systems are required to evaluate the contribution of the identified variants/genes in DLD pathology.

Materials and methods

Ethics statement

This study was approved by the Ethics and Research Committee Farhat Hached University Hospital, Sousse, Tunisia, (OHRP IRB00008931) and followed the guidelines of the Regional Committee of the Protection of Persons, Sfax, Tunisia (CPP SUD reference number 28/2019). The participants and legal guardian of minor participants provided informed consent for participating in the study.

Subjects and clinical evaluations

A two-generation Tunisian family of five members and three siblings (2 males and 1 female; aged 10–13 years) showing DLD with a complex genetic background was subjected to a detailed investigation at Farhat Hached University Hospital of Sousse, Tunisia. The studied family showed an autosomal recessive mode of inheritance of DLD. Parents were interviewed to obtain background information and family history of any language disorders. They affirmed that no consanguinity was recorded in their family lineage. Indeed, parents gave details about the developmental history of each of their offspring and were also asked about other family members who might have exhibited signs of DLD or any related language impairments. For the DLD diagnosis, the probands underwent several phenotype assessments as follows: Initially, a speech-language examination was performed based on the children's performance on the "Thiberge" test which consists of a series of images presented to the child. This test evaluates both comprehension, through the tester's questions and the child's answers, and expression, by assessing terms, structures, vocabulary, and pronunciation. Next, the children were further examined using the non-word repetition test (NWRT) and sentence repetition test. The non-word repetition test was carried out to give an index of phonological short-term memory by requiring the patient to repeat non-words (syllables) that vary in length and complexity. The sentence repetition test was used to assess the syntactic and semantic processing of short-term verbal memory and knowledge of words and grammar by requiring the patient to repeat sentences of varying grammatical structures and lengths. The speech-language examination was performed in the native language of the family members which is the Arabic language. Moreover, a psychiatric evaluation, including an intelligence quotient assessment using the Test of Nonverbal Intelligence, Second Edition (TONI-2), was conducted. The TONI-2 is a non-verbal intelligence test designed to measure problem-solving ability while minimizing the effect of linguistic, motor, and cultural factors. In addition, a screening

for a possible involvement of Autism Spectrum Disorder was carried out within the psychiatric evaluation. Otorhinolaryngology examination, Pure-tone Audiometry and Otoacoustic Emissions (OAEs) were performed to assess hearing loss as a possible cause of the language disorder. Additionally, to check any retro-cochlear disorders, brainstem auditory evoked response (BAER) was evaluated. MRI scans were carried out to assess the presence of any congenital cochlear malformation of the Mondini type (*Mondini dysplasia*). EMG was deployed to diagnose any peripheral neuropathy that could be associated with hearing loss. As well, the parents underwent the Speech-Language general examination in order to assess their language abilities and thus rule out any subtle signs of DLD, in addition to an Autism Spectrum Disorder evaluation to ensure a comprehensive assessment. Peripheral blood samples were obtained from all family members.

Whole exome sequencing

Genomic DNA was extracted from the blood samples of all family subjects according to a standard phenol-chloroform method[58]. The concentration and purity of the extracted DNA were assessed using Nanodrop2000 (ThermoFisher Scientific, USA). The five family members were subjected to WES. Briefly, target genomic DNA was captured using the in-solution enrichment method with biotinylated oligonucleotide probes of the SureSelect XT Clinical Research Exome 54 Mb library according to the manufacturer's instructions (Agilent Technologies, USA). Regions' capture, enrichment and elution were performed according to the manufacturer's instructions without modifications in the protocols with the exception that libraries were prepared using NEBNext Ultra ™ II DNA Library Prep Kit (New England Biolabs, USA). The samples were then sequenced with paired-end 75 base pairs on an Illumina Hiseq4000 platform (Illumina, USA) from the IntegraGen SA genomic platform (Evry, France).

WES data analysis

Image analysis and base calling were conducted using Illumina Real-Time Analysis (version 2.7.3). Adapter sequences were removed using Cutadapt software (version 1.3)[59]. Demultiplexing and alignment were carried out using Illumina's CASAVA pipeline (version 1.8.2). Sequence reads were aligned to the human genome reference build version GRCh37 using the Burrows-Wheeler Aligner software[60]. Variant calling, detection and annotation were performed using the GATK package software (https://software.broadinstitute.org/gatk/). Before filtering, on average 41,431 variants were generated per sample including 38,609 (range 38,182 – 39,241) single nucleotide variants and 2,822 (range 2,731-2,889) indel variants. The given variants were filtered out according to the convenient familial segregation mode; only homozygous or compound heterozygous variants were retained. Subsequently, allele frequencies were extracted from the ExAC v1.0 database; only variants that were not previously reported, rare or less common with a minor allele frequency (MAF) of less than 5% were retained.

In-silico analysis

To evaluate the likely pathogenicity of the missense and splice site variants, several in-silico tools were employed including PROVEAN[61], SIFT[62], PolyPhen2 [63], MutationTaster[64], MutPred2[65], CADD[66], EIGEN[67], FATHMM-MKL[68], M-CAP[69], DANN[70], Spliceman[71], NNPLICE[72], and SpliceRover[73]. Evolutionary conservation was also assessed using the Multiz alignment and scored by the phastCons and phyloP programs[74]. Further evaluation of the candidates' genes was performed for biological significance, and function using PubMed (https://pubmed.ncbi.nlm.nih.gov/), Online Mendelian Inheritance in Man (OMIM) (https://www.omim.org/), Gene Cards (https://www.genecards.org/), ClinVar (https://www.clinicalgenome.org/data-sharing/clinvar/) and Decipher (https://decipher.sanger.ac.uk/). Indeed, in order to explore the broader genetic landscape and understand the potential gene interactions underlying the DLD disorder, functional association and interconnections relationship between the identified candidate genes were explored using networking and gene interaction analysis from GeneMANIA (http://genemania.org/) using the Cytoscape program plugin[75].

Copy number variation analysis

Copy number variations were also assessed for all subjects compared to healthy controls using the DNACopy 1.32.0 R package that detects CNVs from exomes based on a read-depth approach. The CNV detection was calculated based on assuming a Gaussian distribution of the normalized coverage of each region after segmentation, then the standard deviation was considered as a measure of the tolerable deviation for each region. The level of deviation (the change point) was estimated as an increase or decrease based on the Log2Ratio at a significance level of p-value = 0.05.

Protein structure modelling of Leu230Arg variant in ABHD14A

Protein structure modelling was performed in order to study the effect of the Leu230Arg missense variant found in the *ABHD14A* gene on it its protein properties. The Molecular Operating Environment MOE 2019 (MOE) software was used for homology modelling, molecular dynamics and structure visualization. The quality of the model was then validated using PROCHECK[76], ERRAT[77] and VERIFY3D[78] programs which are available at the SAVES server (http://nihserver.mbi.ucla.edu/SAVES). The figures were generated by PyMol program version v2.4.1. Moreover, the molecular dynamics were assessed using the GROMACS V5.1.2 software (www.gromacs.org) while the trajectory analysis of simulation data was evaluated using the Bio3D R package.

Conclusion

In the present study, we explored for the first time a Tunisian-family with DLD using the WES approach. Compound heterozygous variants segregating with DLD were identified in *ABHD14A* and *MRNIP* genes which are suggested as potential candidate genes for DLD susceptibility. However, additional validations are required

through further functional studies and replication of the genetic findings in enlarged sample sizes of familial and sporadic patients with DLD. Elaborating knowledge upon the genetic etiology of DLD and other forms of language disorders will not only be conducive to the early identification of infants at genetic risk but also may contribute to the incorporation of effective preventative measures.

Data availability

The datasets generated during the current study are deposited to a public repository and can be accessed from the accession number DOI: https://doi.org/10.6084/m9.figshare.26766955.

Received: 26 February 2024; Accepted: 11 December 2024
Published online: 02 January 2025

References

1. Leonard, L. B. Specific Language Impairment Across Languages. *Child. Dev. Perspect.* **8**, 1–5 (2014).
2. Elbro, C., Dalby, M. & Maarbjerg, S. Language-learning impairments: a 30-year follow-up of language-impaired children with and without psychiatric, neurological and cognitive difficulties. *Int. J. Lang. Commun. Disord* **46**, 437–448 (2011).
3. Bishop, D. V. M. Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *Int. J. Lang. Commun. Disord* **52**, 671–680 (2017).
4. Nudel, R. et al. Developmental language disorder – a comprehensive study of more than 46,000 individuals. *Psychiatry Res.* **323**, 115171 (2023).
5. Peterson, R. L. & Pennington, B. F. Developmental Dyslexia. *Annu. Rev. Clin. Psychol.* **11**, 283–307 (2015).
6. Tomblin, J. B. et al. Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res. JSLHR* **40**, 1245–1260 (1997).
7. Christensen, D., Taylor, C. L. & Zubrick, S. R. Patterns of Multiple Risk Exposures for Low Receptive Vocabulary Growth 4–8 Years in the Longitudinal Study of Australian Children. *PLOS ONE* **12**, e0168804 (2017).
8. Rudolph, J. M. Case History Risk Factors for Specific Language Impairment: A Systematic Review and Meta-Analysis. *Am. J. Speech Lang. Pathol.* **26**, 991–1010 (2017).
9. Whitehouse, A. J. O., Line, E. A., Watt, H. J. & Bishop, D. V. M. Qualitative aspects of developmental language impairment relate to language and literacy outcome in adulthood. *Int. J. Lang. Commun. Disord* **44**, 489–510 (2009).
10. Law, J., Garrett, Z. & Nye, C. The efficacy of treatment for children with developmental speech and language delay/disorder: a meta-analysis. *J. Speech Lang. Hear. Res. JSLHR* **47**, 924–943 (2004).
11. Bishop, D. V. M., Snowling, M. J., Thompson, P. A. & Greenhalgh, T. consortium. CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PloS One* **11**, e0158753 (2016).
12. Bishop, D. V. M., Snowling, M. J., Thompson, P. A. & Greenhalgh, T. & and the CATALISE-2 consortium. Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *J. Child. Psychol. Psychiatry* **58**, 1068–1080 (2017).
13. Smith, S. D. Genes, language development, and language disorders. *Ment Retard. Dev. Disabil. Res. Rev.* **13**, 96–105 (2007).
14. Reader, R. H., Covill, L. E., Nudel, R. & Newbury, D. F. Genome-Wide Studies of Specific Language Impairment. *Curr. Behav. Neurosci. Rep.* **1**, 242–250 (2014).
15. Andres, E. M., Earnest, K. K., Smith, S. D., Rice, M. L. & Raza, M. H. Pedigree-Based Gene Mapping Supports Previous Loci and Reveals Novel Suggestive Loci in Specific Language Impairment. *J. Speech Lang. Hear. Res.* **63**, 4046–4061 (2020).
16. Villanueva, P. et al. Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. *Eur. J. Hum. Genet.* **19**, 687–695 (2011).
17. Andres, E. M. et al. A genome-wide analysis in consanguineous families reveals new chromosomal loci in specific language impairment (SLI). *Eur. J. Hum. Genet. EJHG* **27**, 1274–1285 (2019).
18. Eicher, J. D. et al. Genome-wide association study of shared components of reading disability and language impairment. *Genes Brain Behav.* **12**, 792–801 (2013).
19. Nudel, R. et al. Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. *Genes Brain Behav.* **13**, 418–429 (2014).
20. Mountford, H. S. et al. Candidate gene variant effects on language disorders in Robinson Crusoe Island. *Ann. Hum. Biol.* **46**, 109–119 (2019).
21. Eising, E. et al. Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proc. Natl. Acad. Sci.* 119, e2202764119 (2022).
22. Rakhlin, N. et al. The language phenotype of a small geographically isolated Russian-speaking population: Implications for genetic and clinical studies of developmental language disorder. *Appl. Psycholinguist.* **34**, 971–1003 (2013).
23. Villanueva, P., de Barbieri, Z., Palomino, H. M. & Palomino, H. High prevalence of specific language impairment in Robinson Crusoe Island. A possible founder effect. *Rev. Médica Chile* **136**, 186–192 (2008).
24. Villanueva, P. et al. Exome Sequencing in an Admixed Isolated Population Indicates NFXL1 Variants Confer a Risk for Specific Language Impairment. *PLOS Genet.* **11**, e1004925 (2015).
25. Kornilov, S. A. et al. Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population. *Pediatrics* **137**, e20152469 (2016).
26. Chen, X. S. et al. Next-generation DNA sequencing identifies novel gene variants and pathways involved in specific language impairment. *Sci. Rep.* **7**, 46105 (2017).
27. Devanna, P. et al. Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol. Psychiatry* **23**, 1375–1384 (2018).
28. Andres, E. M., Earnest, K. K., Zhong, C., Rice, M. L. & Raza, M. H. Family-Based Whole-Exome Analysis of Specific Language Impairment (SLI) Identifies Rare Variants in BUD13, a Component of the Retention and Splicing (RES) Complex. *Brain Sci.* **12**, 47 (2021).
29. Andres, E. M. et al. Study of rare genetic variants in TM4SF20, NFXL1, CNTNAP2, and ATP2C2 in Pakistani probands and families with language impairment. *Meta Gene* **30**, 100966 (2021).
30. Mountford, H. S., Braden, R., Newbury, D. F. & Morgan, A. T. The Genetic and Molecular Basis of Developmental Language Disorder: A Review. *Children* **9**, 586 (2022).
31. Pettigrew, K. A. et al. Copy Number Variation Screen Identifies a Rare De Novo Deletion at Chromosome 15q13.1-13.3 in a Child with Language Impairment. *PLoS ONE* **10**, e0134997 (2015).
32. Kalnak, N. et al. Enrichment of rare copy number variation in children with developmental language disorder. *Clin. Genet.* **94**, 313–320 (2018).
33. van Wijngaarden, V. et al. Genetic outcomes in children with developmental language disorder: a systematic review. *Front. Pediatr.* **12**, (2024).

34. Wiszniewski, W. et al. TM4SF20 ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities. *Am. J. Hum. Genet.* **93**, 197–210 (2013).
35. MacArthur, D. G. et al. Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469–476 (2014).
36. Kamphans, T. et al. Filtering for compound heterozygous sequence variants in non-consanguineous pedigrees. *PloS One* **8**, e70151 (2013).
37. Dubois, P., St, P. M. C., Desmarais, C. & Guay, F. Young Adults With Developmental Language Disorder: A Systematic Review of Education, Employment, and Independent Living Outcomes. *J. Speech Lang. Hear. Res.* **63**, 3786–3800 (2020).
38. Graham, S. A. & Fisher, S. E. Understanding Language from a Genomic Perspective. *Annu. Rev. Genet.* **49**, 131–160 (2015).
39. Bieder, A. et al. Rare variants in dynein heavy chain genes in two individuals with situs inversus and developmental dyslexia: a case report. *BMC Med. Genet.* **21**, 87 (2020).
40. Peter, B. et al. Exome Sequencing of Two Siblings with Sporadic Autism Spectrum Disorder and Severe Speech Sound Disorder Suggests Pleiotropic and Complex Effects. *Behav. Genet.* **49**, 399–414 (2019).
41. Lim, E. T. et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* **77**, 235–242 (2013).
42. Mountford, H. S. & Newbury, D. F. The genomic landscape of language: Insights into evolution. *J. Lang. Evol.* **3**, 49–58 (2018).
43. Newbury, D. F., Monaco, A. P. & Paracchini, S. Reading and Language Disorders: The Importance of Both Quantity and Quality. *Genes* **5**, 285–309 (2014).
44. Ateia, H. et al. Population Genome Programs across the Middle East and North Africa: Successes, Challenges, and Future Directions. *Biomed. Hub* **8**, 60–71 (2023).
45. Zeng, Z. & Bromberg, Y. Predicting Functional Effects of Synonymous Variants: A Systematic Review and Perspectives. *Front. Genet.* **10**, (2019).
46. Hoshino, J., Aruga, J., Ishiguro, A. & Mikoshiba, K. Dorz1, a novel gene expressed in differentiating cerebellar granule neurons, is down-regulated in Zic1-deficient mouse. *Brain Res. Mol. Brain Res.* **120**, 57–64 (2003).
47. Aruga, J. et al. Identification and characterization of Zic4, a new member of the mouse Zic gene family. *Gene* **172**, 291–294 (1996).
48. Twigg, S. R. F. et al. Gain-of-Function Mutations in ZIC1 Are Associated with Coronal Craniosynostosis and Learning Disability. *Am. J. Hum. Genet.* **97**, 378–388 (2015).
49. Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage* **62**, 816–847 (2012).
50. Simonyan, K. & Horwitz, B. Laryngeal motor cortex and control of speech in humans. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **17**, 197–208 (2011).
51. Casey, J. P. et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum. Genet.* **131**, 565–579 (2012).
52. Conti-Ramsden, G., Simkin, Z. & Botting, N. The prevalence of autistic spectrum disorders in adolescents with a history of specific language impairment (SLI). *J. Child. Psychol. Psychiatry* **47**, 621–628 (2006).
53. McKinnon, P. J. ATM and ataxia telangiectasia. *EMBO Rep.* **5**, 772–776 (2004).
54. Biton, S., Barzilai, A. & Shiloh, Y. The neurological phenotype of ataxia-telangiectasia: solving a persistent puzzle. *DNA Repair.* **7**, 1028–1038 (2008).
55. Volkow, N. D. et al. Brain glucose metabolism in adults with ataxia-telangiectasia and their asymptomatic relatives. *Brain J. Neurol.* **137**, 1753–1761 (2014).
56. Wang, Y. L. et al. MRNIP condensates promote DNA double-strand break sensing and end resection. *Nat. Commun.* **13**, 2638 (2022).
57. Madabhushi, R., Pan, L. & Tsai, L. H. DNA damage and its links to neurodegeneration. *Neuron* **83**, 266–282 (2014).
58. Sambrook, J. & Russell, D. W. Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc.* pdb.prot4455 (2006). (2006).
59. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
60. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
61. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinforma Oxf. Engl.* **31**, 2745–2747 (2015).
62. Ng, P. C. & Henikoff, S. S. I. F. T. Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
63. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet. Chapter* **7**, (2013). Unit7.20.
64. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
65. Pejaver, V. et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **11**, 5918 (2020).
66. Rentzsch, P., Schubach, M., Shendure, J. & Kircher, M. CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Med.* **13**, 31 (2021).
67. Ionita-Laza, I., McCallum, K., Xu, B. & Buxbaum, J. D. A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* **48**, 214–220 (2016).
68. Shihab, H. A. et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65 (2013).
69. Jagadeesh, K. A. et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586 (2016).
70. Quang, D., Chen, Y. & Xie, X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinforma Oxf. Engl.* **31**, 761–763 (2015).
71. Lim, K. H. & Fairbrother, W. G. Spliceman–a computational web server that predicts sequence variations in pre-mRNA splicing. *Bioinforma Oxf. Engl.* **28**, 1031–1032 (2012).
72. Reese, M. G., Eeckman, F. H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J. Comput. Biol. J. Comput. Mol. Cell. Biol.* **4**, 311–323 (1997).
73. Zuallaert, J. et al. SpliceRover: interpretable convolutional neural networks for improved splice site prediction. *Bioinforma Oxf. Engl.* **34**, 4180–4188 (2018).
74. Siepel, A. et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
75. Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinforma Oxf. Engl.* **27**, 431–432 (2011).
76. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
77. Colovos, C. & Yeates, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci. Publ Protein Soc.* **2**, 1511–1519 (1993).
78. Eisenberg, D., Lüthy, R. & Bowie, J. U. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **277**, 396–404 (1997).

Acknowledgements
We gratefully thank the family members in this study for their collaboration. We gratefully thank Mr. Abdelkarim Zalfeni for his valuable contribution to the clinical data collection.

Author contributions
AB and MBS: conceived the study, performed the analysis, interpreted the data and wrote the manuscript. MB: ascertained patients and obtained clinical data. AS, MC, HG, FJ, and SL collected and validated the data. CB and CP: designed, generated and supervised the whole exome sequencing data. FF: performed the protein modelling analysis. RH: supervised the bioinformatics analysis. SM and MBS designed the research and aquired the fundings. AB, SM, RH and MBS reviewed and revised the manuscript. All authors reviewed the manuscript.

Funding
This work was supported by the Ministry of Higher Education, Scientific Research of Tunisia (LR15-CBS-07) and by a grant from Fondation Pour l'Audition (FPA IDA05).

Declarations

Competing interests
The authors declare no competing interests.

Ethical approval
All procedures performed in studies involving human participants were in accordance with the guidelines of the Regional Committee of the Protection of Persons, Sfax, Tunisia (CPP SUD reference number 28/2019).

Consent statement
The participants and legal guardian of minor participants provided informed consent for participating in the study.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-83115-x.

Correspondence and requests for materials should be addressed to A.B., R.H. or M.B.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024