Ormieres *et al. Molecular Autism* (2025) 16:10
https://doi.org/10.1186/s13229-025-00642-8

Molecular Autism

RESEARCH Open Access

Deciphering the genetic basis of developmental language disorder in children without intellectual disability, autism or apraxia of speech

Clothilde Ormieres[1,2,7*], Marion Lesieur-Sebellin[1,2], Karine Siquier-Pernet[2], Geoffroy Delplancq[1], Marlene Rio[1,2], Mélanie Parisot[3], Patrick Nitschké[4], Cristina Rodriguez-Fontenla[5], Alison Bodineau[2], Lucie Narcy[2,6], Emilie Schlumberger[2,6], Vincent Cantagrel[2†] and Valérie Malan[1,2*†]

Abstract

Background Developmental language disorder (DLD) refers to children who present with language difficulties that are not due to a known biomedical condition or associated with autism spectrum disorder (ASD) or intellectual disability (ID). The clinical heterogeneity of language disorders, the frequent presence of comorbidities, and the inconsistent terminology used over the years have impeded both research and clinical practice. Identifying subgroups of children (i.e. DLD cases without childhood apraxia of speech (CAS)) with language difficulties is essential for elucidating the underlying genetic causes of this condition. DLD presents along a spectrum of severity, ranging from mild speech delays to profound disturbances in oral language structure in otherwise typically intelligent children. The prevalence of DLD is ~7-8% or 2% if severe forms are considered. This study aims to investigate a homogeneous cohort of DLD patients, excluding cases of ASD, ID or CAS, using multiple genomic approaches to better define the molecular basis of the disorder.

Methods Fifteen families, including 27 children with severe DLD, were enrolled. The majority of cases ($n=24$) were included in multiplex families while three cases were sporadic. This resulted in a cohort of 59 individuals for whom chromosomal microarray analysis and exome or genome sequencing were performed.

Results We identified copy number variants (CNVs) predisposing to neurodevelopmental disorders with incomplete penetrance and variable expressivity in two families. These CNVs (i.e., 15q13.3 deletion and proximal 16p11.2 duplication) are interpreted as pathogenic. In one sporadic case, a *de novo* pathogenic variant in the *ZNF292* gene, known to be associated with ID, was detected, broadening the spectrum of this syndrome.

†Vincent Cantagrel and Valérie Malan are co-last authors.

*Correspondence:
Clothilde Ormieres
clothilde.ormieres@hug.ch
Valérie Malan
valerie.malan@aphp.fr

Full list of author information is available at the end of the article

Limitations The strict diagnostic criteria applied by our multidisciplinary team, including speech-language physicians, neuropsychologists, and paediatric neurologists, resulted in a relatively small sample size, which limit the strength of our findings.

Conclusion These findings highlight a common genetic architecture between DLD, ASD and ID, and underline the need for further investigation into overlapping neurodevelopmental pathways.

Trial registration ClinicalTrials.gov Identifier: NCT06660108.

Keywords Developmental language disorder, Neurodevelopmental disorders, Intellectual disability, Autism, ZNF292, 16p11.2 locus, 15q13.3 locus

Background

Language acquisition, the process enabling humans to communicate through language, represents a pivotal stage in child development. Language disorders are highly prevalent in children, with estimated rates ranging from 4 to 10%, varying by age and type of disorder [1–3]. These disorders, in their most severe and long-lasting forms, have a significant impact on academic and professional performance throughout life. They can co-occur with various neurodevelopmental and psychiatric pathologies. They are a heterogeneous entity of varying severity, and confusion over nomenclature has long been an obstacle to understand their origins. To address the lack of consistency in criteria and terminology for children with language difficulties, experts have proposed standardized definitions and nomenclature [4]. The term, developmental language disorder (DLD) refers to children who present with persistent language difficulties that significantly affect social interactions or educational progress and when the defects persist beyond five years of age with poor prognosis. By definition, DLD is not associated with an identified biomedical cause (i.e., brain injury, neurodegenerative condition, sensorineural hearing loss) or autism spectrum disorder (ASD) and intellectual disability (ID). However, it has been acknowledged that DLD can co-occur with other conditions such as attention deficit hyperactivity disorder (ADHD), developmental dyslexia or coordination problems leading to a heterogeneous group of patients that encompasses a wide range of impairments. The prevalence of DLD is ~7-8% or 2% if severe forms are considered and the diagnosis is based on standardized language tests [2, 5]. There is a continuum of severity ranging from speech delay to severe oral language structure disturbances in typically intelligent children. Childhood apraxia of speech (CAS), also known as developmental verbal dyspraxia, is a motor speech disorder, considered as a different clinical entity within the broader category of 'speech sound disorder' [6]. CAS is often associated with other neurodevelopmental disorders (NDD) such as ID, ADHD, ASD and it can also overlap with DLD [7]. CAS belongs with DLD to the large group of 'speech, language and communication disorders'.

While language defects have a multifactorial origin with socio-cultural and educational factors, strong evidences point to the involvement of genetic causes. Indeed, the incidence of DLD is 32% when a family history of language acquisition difficulties is present, compared with only 4% in the general population [8]. Additionally, monozygotic twins exhibit higher concordance rates for DLD compared to dizygotic twins [9]. However, the clinical heterogeneity of language disorders, the presence of co-morbidities and the inconsistent terminology used for many years have hindered research and clinical practice [10]. Distinguishing sub-groups of children with DLD alone (i.e. without children affected by DLD and CAS) is crucial when tackling the underlying genetic causes of this disease. Recently, several studies using high-throughput sequencing have better defined the genetic basis of CAS [10, 11]. Such studies focusing on DLD are limited [12]. The investigation of more homogeneous cohorts of individuals that clearly distinguish DLD cases, from ID and not including children with CAS should improve our understanding of the genetic basis of this disorder. In this study, we aimed to investigate a well-characterized cohort of sporadic and familial severe DLD individuals, distinct from CAS, using comprehensive phenotyping through clinical scales, psychometric tests, and standardized language assessments. Then, genomic analyses were performed using chromosomal microarray analysis (CMA) and trio approaches using whole exome sequencing (WES) or whole genome sequencing (WGS).

Methods

Participants

All the participants were recruited by expert child neurologists specialized in language disorders and learning impairments at Raymond-Poincaré Hospital. Eligible families included at least one child over five years old with a formal diagnosis of severe and isolated DLD according to Phase 2 CATALISE criteria (i.e., without ID, ASD or CAS diagnosis) [4]. Patients have undergone age-appropriate speech, language and reading evaluations by a speech-language physician and cognitive evaluations by a neuropsychologist, as well as evaluation by a paediatric neurologist to identify co-occurring developmental

disorders (e.g., ADHD, ASD) and a medical geneticist for known genetic disorders and genetic testing recommendations. All children included received appropriate speech therapy for at least one year, with a progress report indicating the persistence of language difficulties. However, the profile of these patients is dynamic, as the disorders evolve with age and rehabilitation. Each situation was linked to the school environment to confirm the impact of the disorder on social and school life. Exclusionary criteria were cognitive impairment with nonverbal intellectual quotient (IQ) below 2 SD assessed with the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), or the Wechsler Intelligence Scale for Children (WISC-IV or V) according to the age-appropriateness, ASD, moderate to severe hearing loss, orofacial structural abnormalities, known neurological or genetic disorders at the initial assessment. None of the patients met the diagnostic criteria for CAS according to the American Speech-Language-Hearing Association, 2007 (Childhood apraxia of speech www.asha.org/policy*)*.

Blood samples from affected children and both parents were collected and then stored in the Imagine Institute's biobank. Patients' data were collected and included into a de-identified interactive database created in collaboration with the data science core of the Imagine Institute. Written parental consent and child assent were obtained for participation and data publication. The study received approval from the "Comité de Protection des Personnes", a national committee ensuring ethical patient protection in research. Fifteen families, including 27 children diagnosed with severe DLD, were enrolled after clinical evaluation and speech, language, and cognitive assessments. Pedigree charts are shown in Fig. 1. Twenty-four cases were part of twelve multiplex families, and three cases were sporadic (families DLD-6, DLD-12 and DLD-13). In families DLD-5, DLD-8, DLD-11, one parent was affected. This yielded a set of 59 individuals including 26 affected children, and three affected parents, who were tested by WES (DLD-1 to DLD-6) or WGS (DLD-7 to DLD-15). CMA and WGS or WES were performed on all affected children except for family DLD-8 where the second affected sister (II.2) was investigated only with CMA.

Molecular cytogenetics

Agilent CGH Microarray 60 K (Agilent Technologies, Santa Clara, CA, USA) was used for genomic copy number analyses that was carried out according to manufacturers' recommendations. This microarray is spotted with 60,000 oligonucleotides and the space between two consecutive probes is approximately 60 kb. Agilent CytoGenomics v5.0.2 software was used to analyse and report the data. Aberrations were detected with the ADM2 algorithm and the filtering option using a threshold of three probes. Thus, Copy Number Variants (CNVs) which are approximately 180 kb in size are detected. Genomic positions are relative to human genome Build GRCh37/hg19. Using standard protocols, chromosomal rearrangement characterization and parental testing were performed with fluorescence in situ hybridization using bacterial artificial chromosome clones on chromosome preparations from leukocyte cultures: RP11-1128L19 located on Xp22.12 for family DLD-8, RP11-504I2 located on 16p11.2 (*TBX6* locus) for family DLD-10, CTD-2515C15 located on 16p11.2 (*SH2B1* locus) for family DLD-11, and RP11-265I17 located on 15q13.3 for family DLD-13. The 5p13.2 duplication in family DLD-9 was detected by WGS.

High-throughput sequencing and analyses

WGS and WES have been performed as previously reported [13, 14]. Trio approaches that include at least the proband and both parents were systematically used. Whole genome DNA libraries were constructed using either TruSeq DNA PCR-Free Sample Preparation Kit (Illumina) starting with 2.2 µg of each patient's genomic DNA, or DNA PCR-Free Prep Tagmentation (Illumina) protocol starting with 350 ng of each patient's genomic DNA. An equimolar pool of the libraries was prepared according the manufacturer instructions. The pool of libraries was sequenced on an Illumina NovaSeq6000 (paired-end sequencing 150 + 150 bases, Xp mode). Downstream processing was carried out as described [13]. In the WGS analysis, structural variants were detected using a combination of three different software programs, Wisecondor, Canvas and Manta, as previously described [15].

An in-house software (Polyweb) developed by the Bioinformatics Platform of the Imagine Institute (University Paris Cité) was used to filter the annotated variants. To focus on potentially pathogenic variants, standard filtering criteria are applied. These include limiting the number of gnomAD alleles to less than 1000 (equivalent to a frequency of ~0.7%) and the number of gnomAD homozygotes to less than 10 (gnomAD v2). The system also considers predicted protein impact across all gene transcripts, such as stop gain, stop loss, start loss, frameshift mutations, in-frame deletions or insertions, missense mutations, and predicted splice regions. Our internal variation database "Déjà Vu" applies additional filters, such as a patient allele count below 100 and a homozygote count below 10. This database includes more than 8300 genomes and 23,600 exomes mostly from families with children affected with rare genetic diseases including various neurodevelopmental disorders.

Once potentially pathogenic variations are identified, Polyweb uses an intrinsic scoring system to rank them. This system is based on a number of key criteria:

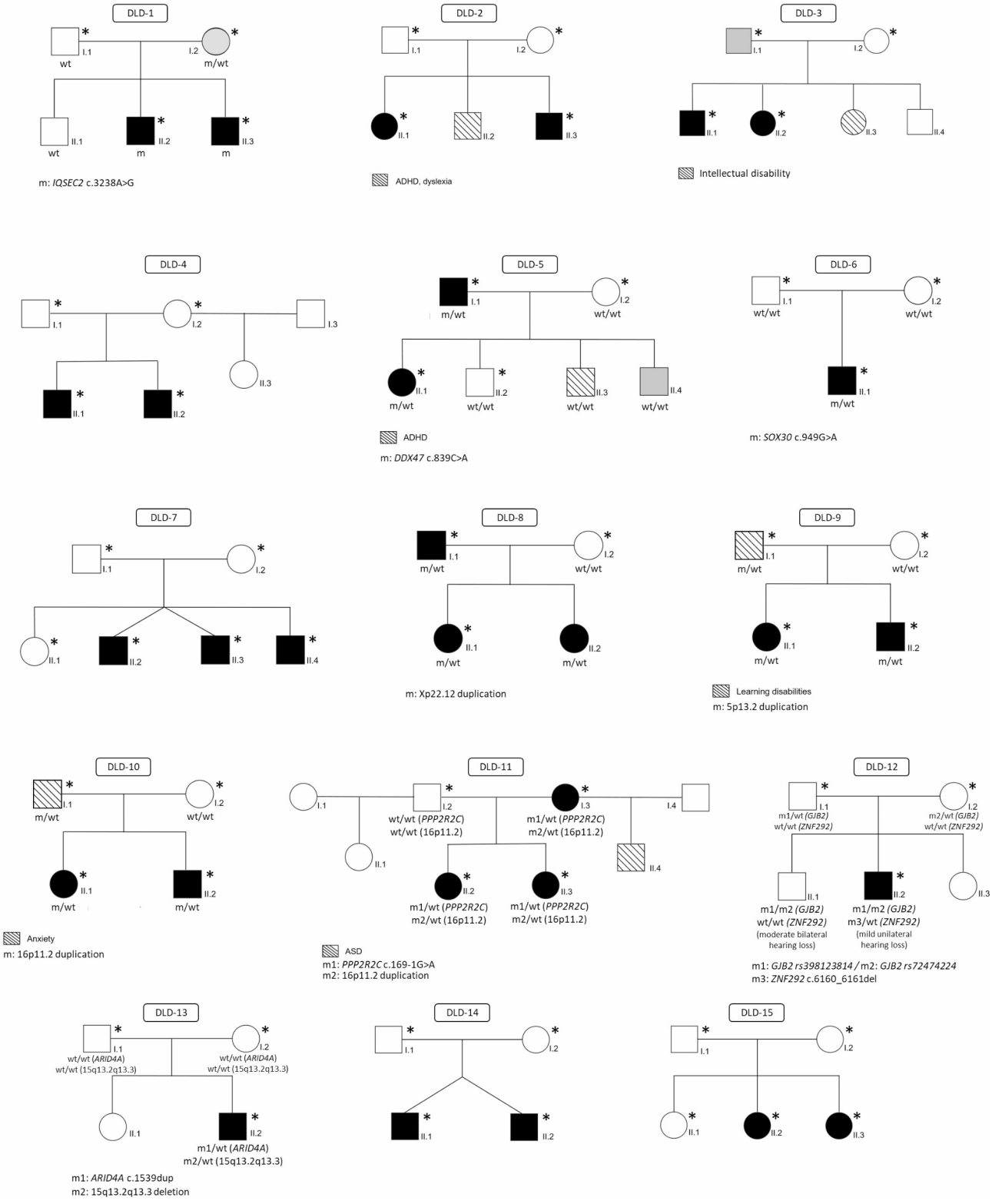

Fig. 1 Pedigrees of the 15 families included in the study. Individuals with language development disorders are depicted in black. Grey indicates transient language delays, or forms considered moderate because they have no impact on daily life, schooling or professional integration. An asterisk (*) denotes those who underwent exome or genome sequencing. Variants of interest, when identified, are marked as "m" beneath the corresponding individual in the pedigree. Wt indicates wild type

- Variation sequence quality: the accuracy and reliability of the sequence data itself.
- Plausibility of all inheritance models (autosomal dominant, autosomal recessive, X-linked dominant, and X-linked recessive). If the variation is *de novo* in the patient, it can indicate a higher likelihood of pathogenicity, especially in cases of severe phenotypes. For genes with autosomal recessive inheritance, the system looks for homozygous or compound heterozygous variations. It also considers X-linked variations in males and cases of uniparental disomy.
- Gene relevance: whether the variation occurs in a gene known to be associated with ID or listed in OMIM (Online Mendelian Inheritance in Man, https://www.omim.org/).
- Predicted effect on protein or splicing: the predicted functional effect of the variation on the gene or protein, including how it might affect splicing processes.
- Known pathogenicity: the presence of the variation in known pathogenic databases such as HGMDpro (https://digitalinsights.qiagen.com/products-overview/clinical-insights-portfolio/human-gene-mutation-database/) or ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), which may add weight to its clinical significance.
- Population frequency: the frequency of the variation in the general population, as reported in gnomAD (https://gnomad.broadinstitute.org/), with rarer variation more likely to be pathogenic. In addition to external databases, our internal variation database is also used to provide a broader context for the rarity and potential significance of variations, especially within our specific patient cohort. These criteria ensure a comprehensive analysis of all variations in known human genes (whether or not they are listed in OMIM) that are predicted to affect proteins. The intrinsic scoring system helps to prioritize variations for further investigation, balancing the likelihood of pathogenicity with the need to minimize false positives. We use the following variant pathogenicity prediction tools to filter and or assess the impact of the variant. Combined Annotation Dependent Depletion (CADD) is a tool that integrates multiple annotations into one metric and can assess multi-nucleotide substitutions and insertion/deletions variants. The Rare Exome Variant Ensemble Learner (REVEL) and the Missense deleteriousness predictor (MISTIC) are dedicated to the evaluation of missense variants. REVEL uses 13 different pathogenicity prediction tools (e.g., PolyPhen-2, SIFT, MutationTaster) and MISTIC is based on the combination of two complementary machine learning algorithms and the integration of 113 missense features. We excluded variants with a PHRED-like scaled CADD score ≤ 20 regarding nonsynonymous substitutions. In familial cases, we considered that the DLD individuals have the same disease and candidate variants should be present in all affected family members. The incomplete penetrance hypothesis was included in the analysis. We focused on variants affecting splice sites or coding regions (nonsynonymous substitutions, insertions, or deletions), or intronic variants with a predicted effect on splicing.

Results

Phenotypic data

A total of 27 affected children (16 males and 11 females) including two dizygotic twins were included in the study (Fig. 1). Of these, 24 children were part of multiplex families and three cases were sporadic (DLD-6, DLD-12 and DLD-13). In three families (DLD-5, DLD-8 and DLD-11), one of the parents (two fathers and one mother) was diagnosed with a DLD. None of the patients had an IQ below 70 or displayed ASD at the time of the assessment. Six families had a family history of a neurodevelopmental or psychiatric disorder (ASD, learning disabilities/ID, ADHD, dyslexia, anxiety, and gaming addiction). Three individuals from families DLD-1, DLD-3, and DLD-5 were deemed to exhibit mild symptoms, as they did not fully meet the previously established clinical criteria. They have a transient or moderate form of language impairment that does not interfere with daily life, school or professional integration.

All paediatric cases presented with severe delays in speech and language development. The majority of affected individuals (24/27) exhibited an impairment of written language. Hearing was normal in all but two children demonstrated mild hearing loss that did not explain the severity of the DLD. Additional clinical characteristics including ADHD ($n = 8$), anxiety ($n = 12$), coordination development disorder ($n = 4$) and behavioural issues ($n = 4$) were noted in 18 children. One patient had microcephaly. Ten probands underwent magnetic resonance imaging (MRI) of the brain, which revealed no abnormalities except for one case in which a nonspecific hypersignal of the white substance was detected. All affected children received or had received speech therapy. The phenotype of the participants is summarized in Table 1; Fig. 2.

Inherited and *de novo* neurodevelopmental CNVs contribute to DLD

Chromosomal microarray analysis was performed in all affected children. One *de novo* heterozygous 15q13.2q13.3 deletion (MIM #612001) was identified in

Table 1 Clinical features of the 27 patients included in our cohort

Family	DLD-1		DLD-2		DLD-3		DLD-4		DLD-5	DLD-6	DLD-7			DLD-8	
Individual	**II.2**	**II.3**	**II.1**	**II.3**	**II.1**	**II.2**	**II.1**	**II.2**	**II.1**	**II.1**	**II.2**	**II.3**	**II.4**	**II.1**	**II.2**
Sex	M	M	F	M	M	F	M	M	F	M	M	M	M	F	F
Affected parents[a]	Mother	Mother	-	-	Father	Father	-	-	Father	-	-	-	-	Father	Father
Motor delay	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Intellectual disability	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Speech and language delay	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Dissociation between PRI-VCI (psychometric test)[b]	17	20	18	15	23	11	72	NA	45	28	20	16	23	38	20
Impairment of written language	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+
Mild hearing loss	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Autistic features	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Behavioural problems	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Adapted schooling	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Anxiety	-	+	-	-	+	+	-	-	-	+	-	-	-	+	-
ADHD	+	+	+	-	-	-	-	-	+	-	-	-	+	-	-
Coordination development disorder	-	-	+	-	-	-	-	-	-	+		-	-	-	-
Microcephaly	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Seizures	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Other clinical features	Short fingers, Bilateral single transverse palmar crease	Left single transverse palmar crease, Epicanthus, Short fingers	-	Laxity	-	-	-	Unilateral ear aplasia	-	Hyperlaxity	-	-	-	-	-
Brain MRI	NAb	NAb	NAb	NA	NA	NA	NA	NA	NAb	NA	NA	NA	NA	NAb	NA
EEG	NA	NA	NAb	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 1 (continued)

Family	**DLD-1**		**DLD-2**		**DLD-3**		**DLD-4**		**DLD-5**	**DLD-6**	**DLD-7**			**DLD-8**	
Individual	**II.2**	**II.3**	**II.1**	**II.3**	**II.1**	**II.2**	**II.1**	**II.2**	**II.1**	**II.1**	**II.2**	**II.3**	**II.4**	**II.1**	**II.2**
Family	**DLD-9**		**DLD-10**		**DLD-11**				**DLD-12**	**DLD-13**	**DLD-14**			**DLD-15**	
Individual	**II.1**	**II.2**	**II.1**	**II.2**	**II.2**		**II.3**		**II.2**	**II.2**	**II.1**	**II.2**		**II.2**	**II.3**
Sex	F	M	F	M	F		F		M	M	M	M		F	F
Affected parents[a]	-	-	Father	Father	Mother		Mother		-	-	-	-		-	-
Motor delay	-	-	-	-	-		+		-	-	-	-		-	-
Intellectual disability	-	-	-	-	-		+		-	-	-	-		-	-
Speech and language delay	+	+	+	+	+		+		+	+	+	+		+	+
Dissociation between PRI-VCI (psychometric test)[b]	9	10	15	NA	22		27		40	20	25	28		NA	NA
Impairment of written language	-	+	+	-	+		+		+	+	+	+		+	+
Mild hearing loss	-	-	-	-	-		-		-	-	+	-		-	-
Autistic features	-	-	-	-	-		-		-	-	-	-		-	-
Behavioural problems	-	+	-	-	-		+		+	-	+	-		-	-
Adapted schooling	+	+	+	+	+		+		+	+	+	+		+	+
Anxiety	-	+	-	-	+		+		+	+	-	+		+	-
ADHD	-	+	-	-	-		-		+	-	-	-		+	-
Coordination development disorder	-	-	-	-	-		-		-	-	-	-		+	+
Microcephaly	-	+	-	-	-		-		-	-	-	-		-	-
Seizures	-	-	-	-	-		-		-	-	-	-		-	-
Other clinical features	Long fingers	-	-	-	-		+		+	-	-	-		-	-
Brain MRI	NAb	Unspecific white matter hyperintensities	NA	NA	NA		NAb		NA	NA	NA	NA		NAb	NAb
EEG	NA	NA	NA	NA	NA		NA		NA	NA	NA	NA		NA	NA

F, female; M, male; ADHD, attention deficit hyperactivity disorder; EEG, electroencephalogram; MRI, magnetic resonance imaging; PRI, perceptual reasoning index; VCI, verbal comprehension index; NA, not available; NAb, no abnormalities

[a] Parents reported as affected (families DLD-5, 7, 11) have not been formally tested but have been seen by specialists during medical consultations. These adults are autonomous in their daily lives, professionally integrated and without psychopathology. However, a language disorder has been present since childhood and still affects their daily life. Mildly affected parent (families DLD-1, 3) have transient or moderate forms of developmental language disorder that do not affect daily life, school or professional integration

[b] We considered a difference of more than 10 points to be significant

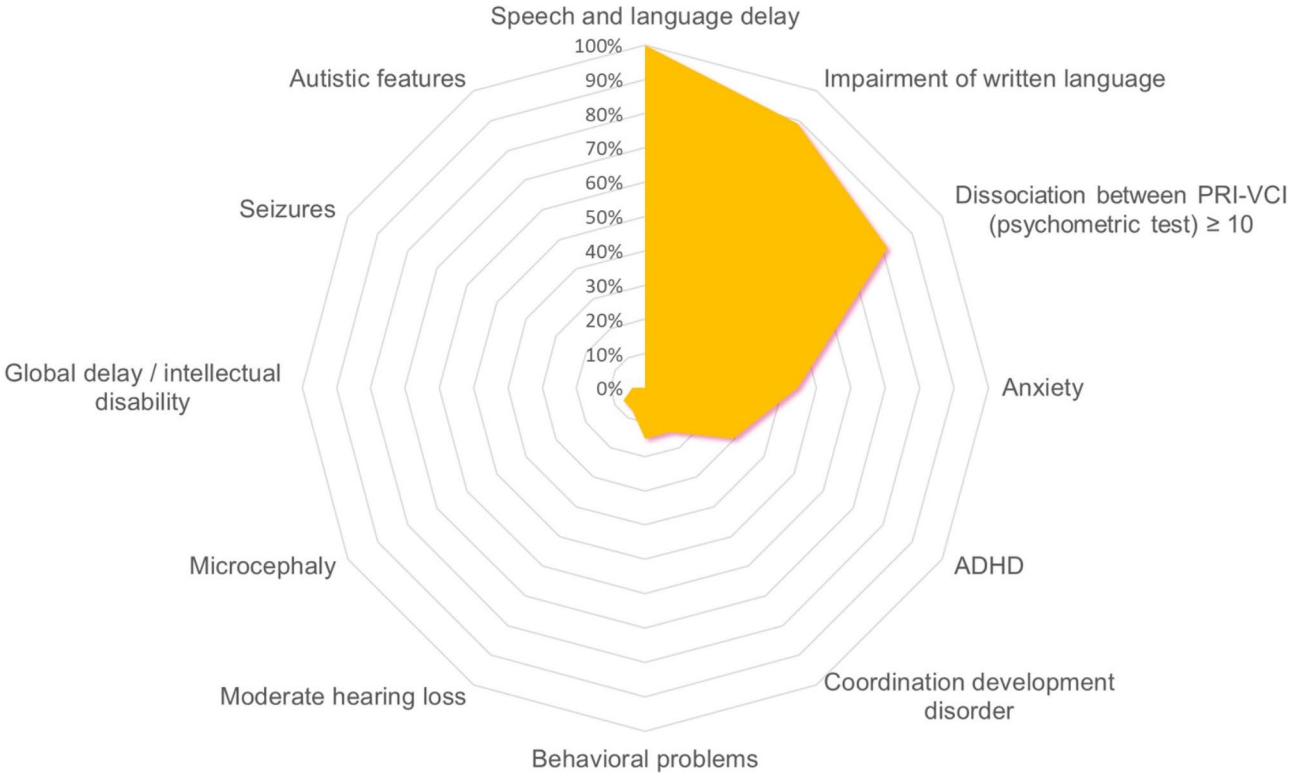

Fig. 2 Phenotypic overlap of the 27 patients included in our cohort. ADHD, attention deficit-hyperactivity disorder, PRI, perceptual reasoning index; VCI, verbal comprehension index

the affected male from family DLD-13. This recurrent deletion contains seven genes, including *CHRNA7* (cholinergic receptor nicotinic alpha 7 subunit; MIM *118511) and *OTUD7A* (OTU deubiquitinase 7 A; MIM *612024). In family DLD-10, the recurrent 16p11.2 duplication (BP4-BP5) was identified in the two affected children and their father. It is noteworthy that the father does not have DLD but has anxiety and a gaming addiction. This duplication encompasses 29 genes, including *KCTD13* (potassium channel tetramerization domain Containing 13; MIM *608947) and *TAOK2* (TAO kinase 2; MIM *613199) that are likely to be involved in the neuropsychiatric phenotype associated with this CNV [16, 17]. The 15q13.3 deletion and the proximal 16p11.2 duplication identified in these two families are established risk factors for NDDs and are interpreted as pathogenic [18]. Finally, three families were found to harbour a duplication classified as variant of uncertain significance (VUS) according to the recommendation of the American College of Medical Genetics and Genomics (ACMG) [19] and ClinGen (https://www.clinicalgenome.org/). In family DLD -11, a recurrent distal 16p11.2 duplication (BP2-BP3) of approximately 200 kb containing the *SH2B1* (SH2B adaptor protein 1; MIM *608937) gene was detected in the two affected children and their affected mother. While the recurrent deletion CNV of this locus is pathogenic and associated with NDDs with incomplete penetrance and variable expressivity, mirror duplication is a VUS in the absence of further evidence. In family DLD-8, the two affected females carry an Xp22.12 duplication involving the *RPS6KA3* gene (ribosomal protein S6 kinase A3; MIM *300075). Loss-of-function (LoF) variants in this gene have been shown to be responsible for Coffin-Lowry syndrome CLS (MIM #303600). The Xp22.12 duplication was inherited from the affected father. Finally, in family DLD-9, WGS identified a 5p13.2 duplication that was missed by CMA. The duplicated segment contains three whole coding genes, *CPLANE1* (ciliogenesis and planar polarity effector complex subunit 1; MIM *614571), *NUP155* (nucleoporin 155; MIM 606694), *WDR70* (WD repeat domain 70; MIM *617233), and exons 21–47 of *NIPBL* (nipped-B-Like; MIM *608667; NM_133433.4). This duplication was found in both affected siblings. The father, who has learning disabilities, also harbours the CNV. Monoallelic variants in the *NIPBL* gene resulting in loss of function are the major cause of Cornelia de Lange syndrome (MIM #122470). Genetic findings related to these structural variants are summarized in Table 2.

Table 2 Copy number variants of interest identified in our cohort

Family	DLD-8	DLD-9		DLD-10		DLD-11		DLD-13
Individual	**II.1**	**II.1**	**II.2**	**II.1**	**II.2**	**II.2**	**II.3**	**II.2**
CNV type	Tandem duplication	Tandem duplication		Duplication (orientation not determined)		Duplication (orientation not determined)		Heterozygous deletion
Nomenclature	seq[GRCh37] dup(X)(p22.12p22.12)pat NC_000023.10:g.19947598_20300709dup	seq[GRCh37] dup(5)(p13.2p13.2)pat NC_000005.9:g.37009768_37625742dup		seq[GRCh37] dup(16)(p11.2p11.2) pat NC_000016.9:g.(?_29446858)_(30301461_?)dup		seq[GRCh37] dup(16)(p11.2p11.2) mat NC_000016.9:g.(?_28775441)_(29061554_?)dup		seq[GRCh37] del(15)(q13.2q13.3)dn NC_000015.9:g.(?_30460631)_(32907245_?)del
Size	353 kb	616 kb		855 kb		286 kb		2.5 Mb
Inheritance	Paternal	Paternal		Paternal		Maternal		*de novo*
Landmark gene(s) [a]	*RPS6KA3*	*NIPBL* (partially), *CPLANE1*, *NUP155*		*TBX6*, *KIF22*, *PRRT2*, *TLCD3B*, *ALDOA*, *CORO1A*		*SH2B1*, *TUFM*, *ATP2A1*, *CD19*, *LAT*		*OTUD7A*, *CHRNA7* *FAN1*, *TRPM1*,
Classification	VUS	VUS		Pathogenic		VUS		Pathogenic

CNV, copy number variant; VUS, variant of uncertain significance; kb, kilobases; Mb, megabases

[a] See supplementary Fig. 1 for full list of genes

Contribution of likely pathogenic sequence variants to DLD

Using high-throughput sequencing, we analysed data from 29 affected individuals (26 children and 3 parents) and 30 healthy or mildly affected individuals. As *de novo* variants are largely involved in NDDs including CAS, we looked for these variants by filtering based on the allele frequency using in-house and gnomAD databases, and CADD score, and prioritized them based on a known or possible role in neurodevelopment. We identified three d*e novo* sequence variants in the families with sporadic cases. In family DLD-6 we identified a missense variant p.(Val317Met) in *SOX30* (SRY-box transcription factor 30; NM_178424.2, c.949G > A; MIM *606698) and in family DLD-13, a frameshift variant p.(Glu514ArgfsTer3) in *ARID4A* (AT-rich interaction domain 4 A; NM_002892.4, c.1539dup; MIM *180201). Both of these genes are not involved in Mendelian disorders yet but are intolerant to LoF variation (Table 3). In family DLD-12, we identified a previously reported truncating variant p.(Glu2054LysfsTer14) in the *ZNF292* gene (zinc finger protein 292; NM_015021.3, c.6160_6161del; MIM *616213), which is involved in ID [20].

Extensive research of variants inherited through a recessive mode failed to identify causative variants besides known variants in the *GJB2* gene (gap junction protein beta 2; MIM *121011), which cause autosomal recessive non syndromic hearing loss with variable expressivity and incomplete penetrance and explain the previously undiagnosed hearing loss in the non-affected brother in family DLD-12 [21]. In addition, we found three other sequence variants segregating through a dominant mode of inheritance, inherited from parents with DLD or moderate language disorder. We identified missense variants in *IQSEC2* (IQ motif and Sect. 7 domain ArfGEF 2; MIM *300522) and *DDX47* (DEAD-box helicase 47; MIM 615428) in families DLD-1 and DLD-5 respectively (Table 3). We also identified a variant at an essential splice acceptor site in the *PPP2R2C* gene (protein phosphatase 2 regulatory subunit B gamma; MIM *605997) that is predicted to be LoF intolerant (pLI = 0.9 and LOEUF = 0.376) in family DLD-11.

All variants were classified as VUS according to ACMG criteria except for the variant in *ZNF292*, which was considered pathogenic [22]. The expression of most of these genes is well detected during human brain development (Supplementary Fig. 1A). Among them, *ZNF292*, *ARID4A* and *DDX47* show an upregulated expression from 8 to 26 post-conception weeks compared to latter stages (Supplementary Fig. 1B), suggesting a role in the early development of the cortex, hippocampus, striatum and cerebellum.

Discussion

The objective of this study was to screen sporadic and multiplex families with children diagnosed with severe DLD selected through strict criteria, in order to better define the underlying genetic factors of this disease. In our cohort, five families were identified as having a CNV of interest. Of these, two were recurrent pathogenic CNVs (including the 15q13.3 microdeletion and the proximal 16p11.2 duplication) which have previously

Table 3 Sequence variants of interest identified in our cohort

Family		DLD-1		DLD-5	DLD-6	DLD-11		DLD-12	DLD-13
Individual		**II:2**	**II:3**	**II:1**	**II:1**	**II:2**	**II:3**	**II:2**	**II:2**
Gene		**IQSEC2**		**DDX47**	**SOX30**	**PPP2R2C**		**ZNF292**	**ARID4A**
Sequence variants identified	Genomic coordinates (GRCh37/hg19)	NC_000023.11: g.53238184T > C		NC_000012.12: g.12,823,958 C > A	NC_000005.10: g.157,651,130 C > T	NC_000004.12: g.6,378,573 C > T		NC_000006.12: g.87259789_87259790del	NC_000014.9: g.58351207dup
	cDNA	NM_001111125.3: c.3238 A > G		NM_016355.4: c.839 C > A	NM_178424.2: c.949G > A	NM_020416.4: c.169-1G > A		NM_015021.3: c.6160-6161del	NM_002892.4: c.1539dup
	Predicted protein	p.(Ile1080Val)		p.(Ala280Asp)	p.(Val317Met)	p.?		p.(Glu2054LysfsTer14)	p.(Glu514Argfs-Ter3)
	Status	Homozygous		Heterozygous	Heterozygous	Heterozygous		Heterozygous	Heterozygous
	Inheritance	Maternal		Paternal	*de novo*	Maternal		*de novo*	*de novo*
	Exon	12		8	1	intron 2		8	16
Frequency in population databases	gnomAD v2 AF	0		0	0	0		0	0
	gnomAD v4 AF	0		0.000001590	0	0		0.000001246	0
	deCAF AF	0		0	0	0		0	0
	All of Us AF	0.000002		0	0	0		0.000004	0
in silico predictions of deleteriousness	PHRED-like scaled CADD score	20		27	25	34		33	NA
	REVEL score	0.063		0.511	0.482	NA		NA	NA
	MISTIC score	0.32		0.32	0.71	NA		NA	NA
Gene constraint scores	gnomAD v2 LOEUF score	0.13		0.98	0.23	0.38		0.14	0.14
	gnomAD v2 missense Z score	5.19		0.39	0.78	3.53		1.41	1.45
ACMG classification		3		NA	NA	NA		**5**	NA

NA: not applicable; AF, allele frequency. A CADD score is a ranking, with higher scores indicating a greater likelihood of being deleterious. The REVEL and MISTIC scores can range from 0 to 1, with higher scores reflecting a greater likelihood that a given missense variant is disease-causing

been associated with NDDs, such as DD/ID, ASD, and psychiatric disorders [18]. These CNVs, which are mediated by non-allelic homologous recombination (NAHR) between segmental duplications (called breakpoints [BP] on chromosomes 15 and 16), can result in a spectrum of clinical phenotypes, ranging from no discernible symptoms to severe NDDs. In the sporadic case from DLD-13, a 1.5 Mb microdeletion of 15q13.3 was detected. This recurrent deletion between BP4-BP5 contains seven genes, including *CHRNA7* and *OTUD7A*. This recurrent CNV predisposes to a wide range of phenotypes, including schizophrenia, ASD and speech delay/language impairment [23]. Interestingly, *Otud7a*-null mice show impaired vocalization among other neurodevelopmental features [23]. In family DLD-10, the proximal 16p11.2 duplication, which is approximately 600 kb in size, was identified in the two affected children and their father. Deletions and reciprocal duplications of the proximal 16p11.2 interval have been associated with DD/ID, ASD and mirror phenotypes with head circumference and body weight affected in opposite ways [24, 25]. Several studies have shown that the proximal 16p11.2 region is involved in CAS (for the deletion carriers only) and a broad spectrum of communication impairment (both deletion and duplication carriers), which frequently occur in conjunction with other neurobehavioral deficits [26–29]. It is noteworthy that in the absence of ASD and cognitive impairment, language impairment represents a prominent clinical feature in individuals with proximal 16p11.2 deletion and duplication [29]. Interestingly, in family DLD-10, the father who carries the proximal 16p11.2 duplication is asymptomatic but displays anxiety and a gaming addiction. This illustrates the incomplete penetrance and possibly the variable expressivity of this recurrent CNV, which is presumably due to additional factors, including common variants, epigenetics, and

environmental factors. Interestingly, Loviglio et al. have demonstrated that the two non-overlapping proximal and distal regions at 16p11.2 are reciprocally involved in complex chromatin looping as well as coordinated expression and regulation of encompassed genes [24]. In the family DLD-11, the mother and her two daughters, carrying the distal 16p11.2 duplication, were diagnosed with DLD. Despite the current lack of evidence for the distal 16p11.2 duplication, which precludes its classification as pathogenic, this CNV may play a role in the observed phenotype. Overall, previous studies and our results show that recurrent pathogenic CNVs (e.g., the 15q13.3 microdeletion and the proximal 16p11.2 duplication) are frequent in language impairment, including DLD cases.

In addition to these pathogenic CNVs, CNVs considered to be VUS were also identified. In the multiplex family DLD-8, a 353 kb duplication of the Xp22.12 region was identified in the affected daughters and the affected father. This segment encompasses the *RPS6KA3* gene. *RPS6KA3* variants resulting in LoF cause either syndromic X-linked ID, known as Coffin-Lowry syndrome or non-syndromic X-linked ID (MIM# 300844). Coffin-Lowry syndrome is characterized by moderate to severe ID, growth retardation, characteristic facial and digital abnormalities, and various skeletal anomalies. Carrier females are more likely to be mildly affected. Small duplications involving this Xp22.12 segment are rare [30, 31]. Patients have mild or borderline ID with few associated clinical features. Among the few other duplicated genes, *RPS6KA3* was considered as the only candidate gene for the phenotype. Interestingly, one patient has been a diagnosed with dyslexia [31]. Lastly, in the multiplex family DLD-9, we identified a 616 kb duplication of the 5p13.2 region involving four coding genes including *NIPBL*, which was partially duplicated. Small duplications of the 5p13 band, encompassing *NIPBL*, have been reported in few patients presenting with hypotonia, DD/ID, variable facial characteristics and minor hand abnormalities (chromosome 5p13 duplication syndrome, MIM# 613174) [32, 33]. *NIPBL* was been suggested to be the major dosage-sensitive gene in this microduplication syndrome, which can have an incomplete penetrance and variable expressivity [32, 33]. Unfortunately, in addition to DD and ID, no description of the language phenotype has been provided for the patients. In total, in five of the fifteen families, we were able to identify two recurrent pathogenic CNVs (i.e., the 15q13.3 microdeletion and the 16p11.2 proximal duplication) which are strongly associated with cognitive impairment and three structural variants (i.e., the 16p11.2 distal duplication, a 5p13 duplication and a Xp22.12 duplication) that may play a role in the phenotype.

With regard to the contribution of rare sequence variants in our cohort, a truncating variant in the *ZNF292* gene was identified in a sporadic case (DLD-12). This variant was classified as pathogenic or likely pathogenic on five occasions in ClinVar (RCV001260794.4, RCV001292573.11, RCV001879995.6, RCV001261752.3, RCV003353266.2). The male patient presents with severe expressive and receptive language disorder, no written language skills and severe inhibition, which subsequently evolved into social withdrawal at the age 11. It should be noted that he does not have ID. Additionally, he presents with a mild unilateral hearing loss, which does not account for the observed DLD. At the initial clinical assessment and at the time of inclusion, which was more than a year after the beginning of rehabilitation, the patient presented with a DLD that was fairly typical of that observed in early childhood. Following the receipt of the genetic results, the clinical picture of the patient had evolved with a relational disorder in the foreground. This highlights the importance of conducting a re-evaluation of the cognitive and behavioural profile of children, both for adjusting support and for exploring the aetiology. LoF variants in *ZNF292* have been associated with a spectrum of neurodevelopmental features including ID and ASD with incomplete penetrance. Other clinical features such as motor delay, ADHD, and nonspecific dysmorphic features may be observed [20]. It is noteworthy that one case reported by Mirzaa et al. [20] did not present with evidence of ID but rather exhibited characteristics of ASD and speech delays at the age of six years. The precise function of this gene remains unclear. However, it is highly expressed during brain development, particularly in the cerebellum (Supplementary Fig. 1). These findings provide compelling evidence that the *de novo* variant in *ZNF292* is responsible for the DLD phenotype of this individual, which is an expansion of the clinical spectrum.

In addition, we identified five VUS in five genes, two of which have been associated with NDDs. In a sporadic case (family DLD-6), we found a *de novo* missense variant in *SOX30*, which has not previously been associated with DLD or NDD. SOX family proteins are characterized by a DNA-binding domain, a high mobility group box that exhibits a high degree of similarity with SRY. Members of this family are conserved during evolution, and they have been shown to play pivotal roles during animal development [34]. Although its role is mostly documented during spermiogenesis, *sox30* is expressed in a specific manner at the midbrain-hindbrain boundary during zebrafish neurogenesis [35, 36]. In a multiplex family (DLD-1), a missense variant in *IQSEC2*, known to be involved in a X-linked intellectual developmental disorder, was detected in the two affected sons and their mother who displayed moderate language disorder. Pathogenic variants in *IQSEC2* cause ID, ASD and epilepsy in males (MIM# 309530). Females are less severely

affected and tend to have learning difficulties or mild intellectual disability. In another multiplex family (DLD-5), a missense variant in the *DDX47* gene was identified in both the affected child and the affected father. DDX47 belongs to the DDX/DHX family and has been proposed as a candidate gene for syndromic NDDs [37, 38]. Mono- and biallelic variants, which were considered potentially pathogenic, have been identified in several patients with variable clinical manifestations. Although the variant meets the criteria for segregation with DLD in our family, it remains only potentially disease-causing.

In family DLD-11, a splice-site variant in *PPP2R2C* was identified in the affected individuals (two children and their mother). This gene, which encodes a subunit of protein phosphatase 2 A with a unique expression pattern in the brain, was considered as an interesting candidate gene for mild ID, epilepsy, and behavioural problems in a family with reciprocal translocation (4;6)(p16.1;q22) [39]. In the affected individuals, the translocation that disrupted the *PPP2R2C* gene was found to segregate with the phenotype. Finally, in family DLD-13, a *de novo* frameshift variant in *ARID4A* was detected in the affected child. Wu et al. showed that *ARID4A* and *ARID4B* are members of epigenetic complexes that regulate genomic imprinting at the Prader-Willi syndrome and Angelman syndrome [40]. Consequently, *ARID4A* and *PPP2R2C* are considered as candidate genes for DLD. It can be hypothesized that the variants in *PP2R2C* and *ARID4A* act as a second hit, in conjunction with the 16p11.2 distal duplication and the 15q13.3 deletion in families DLD-11 and DLD-13. The CNVs could render the individuals susceptible to NDD, while the additional genetic alteration may influence the phenotypic trajectory.

In total, in our three sporadic cases, we identified a truncating pathogenic variant in *ZNF292* (family DLD-12), a recurrent 15q13.3 CNV associated with a VUS sequence variant in *ARID4A* in family DLD-13, and a variant of interest in *SOX30* in family DLD-6. In contrast, in the twelve multiplex families, only a limited number of variants that could play a role in the phenotype were identified. It could be speculated that oligogenic or polygenic mechanisms are involved in these multiplex families, as has been suggested for ASD [41–43]. A combination of inherited rare and common variants with variable weight could contribute to the pathogenicity of DLD, whether through additive or synergistic effects. Our results lend support to the hypothesis that DLD and ASD share a similar genetic architecture, as evidenced by the presence of shared CNVs and sequence variants.

Limitations

In this study, we aimed to investigate a homogeneous cohort of individuals diagnosed with DLD, excluding children with CAS and/or ID. Furthermore, in order to obtain a comprehensive phenotyping, all participants underwent clinical scales, psychometric tests and standardized language assessments. The strict inclusion criteria applied by our multidisciplinary team, including speech pathologists, neuropsychologists and paediatric neurologists, resulted in a relatively small sample size, which limits the strength of our findings.

Conclusion

Our approach, combining CMA and WES/WGS, identified a pathogenic sequence variant in the *ZNF292* gene in which LoF variants have been found to be associated with a spectrum of neurodevelopmental features. In two families, known recurrent pathogenic CNVs implicated in NDD, were detected, resulting in an overall diagnostic yield of 20% (3/15 families). We were also able to identify novel genes and CNVs potentially involved in DLD. Lastly, while likely causative *de novo* events appear to be prevalent in sporadic cases of DLD, the majority of familial cases remain unresolved. DLD is a heritable complex disorder, with compelling evidence indicating that genetic factors are likely to be shared with those involved in ASD and ID.

Abbreviations

ACMG	American College of Medical Genetics and Genomics
ADHD	Attention deficit hyperactivity disorder
ASD	Autism spectrum disorder
CADD	Combined Annotation Dependent Depletion
CAS	Childhood apraxia of speech spectrum disorder
BP	Breakpoints
CMA	Chromosomal microarray analysis
CNVs	Copy Number Variants
DD	Developmental delay
DLD	Developmental language disorder
ID	Intellectual disability
IQ	Intellectual quotient
LoF	Loss-of-function
MISTIC	Missense deleteriousness predictor
NAHR	Non-allelic homologous recombination
OMIM	Online Mendelian Inheritance in Man
REVEL	Rare Exome Variant Ensemble Learner
SLCD	Speech, language and communication disorders
VUS	Variant of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing
WISC	Wechsler Intelligence Scale for Children
WPPSI	Wechsler Preschool and Primary Scale of Intelligence

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13229-025-00642-8.

Supplementary Material 1

Acknowledgements

We are very grateful to the families who agreed to participate in this study.

Author contributions

Clothilde Ormières: conception of the project, clinical assessment of the patients, provide feedback on the manuscript, read and approved the final manuscript. Marion Lesieur-Sebellin: analyze of the data, provide feedback on

the manuscript, read and approved the final manuscript. Karine Siquier-Pernet: analyze of the data, provide feedback on the manuscript, read and approved the final manuscript. Geoffroy Delplancq: analyze of the data, provide feedback on the manuscript, read and approved the final manuscript. Marlène Rio: clinical assessment of the patients, provide feedback on the manuscript, read and approved the final manuscript. Mélanie Parisot: help and support for genomic analysis, provide feedback on the manuscript, read and approved the final manuscript. Patrick Nitschké : help and support for bioinformatics analysis, provide feedback on the manuscript, read and approved the final manuscript. Cristina Rodriguez-Fontenla : analyze of the data, provide feedback on the manuscript, read and approved the final manuscript. Alison Bodineau : analyze of the data, provide feedback on the manuscript, read and approved the final manuscript. Lucie Narcy : clinical assessment of the patients, provide feedback on the manuscript, read and approved the final manuscript. Emilie Schlumberger : clinical assessment of the patients, provide feedback on the manuscript, read and approved the final manuscript. Vincent Cantagrel : conception of the project, analyze of the data, provide feedback on the manuscript, read and approved the final manuscript. Valérie Malan : conception of the project, analyze of the data, writing the manuscript, read and approved the final manuscript.

Funding
We are very grateful to Eric Perrier and the group "Entrepreneurs Amis d'Imagine" for their financial support. This work was supported by State funding from the Agence Nationale de la Recherche under the "Investissements d'avenir" program (ANR-10-IAHU-01) and the Fondation Bettencourt Schueller.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Consent for publication
We have obtained consent to collect and use the data for research and publication purposes.

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
Written informed consent was obtained from all individuals. All studies were carried out in accordance with the declaration of Helsinki and were approved by a national ethics committee (CPP Ile de France, RIPH2G reference DI 24.01180.000212, N°2024-A00519-38, CPP reference 29-2024, promoter reference C23-79; promoter: Inserm). ClinicalTrials.gov Identifier: NCT06660108.

Author details
[1]Service de Médecine Génomique des Maladies Rares, Hôpital Necker-Enfants Malades, APHP, Paris 75015, France
[2]Developmental Brain Disorders Laboratory, Université Paris Cité, INSERM, Imagine Institute, Paris 75015, France
[3]Genomics Core Facility, Université Paris Cité, Imagine Institute, Structure Fédérative de Recherche Necker, INSERM, CNRS, Paris 75015, France
[4]Bioinformatics Core Facility, Université Paris Cité, Imagine Institute, INSERM, Paris 75015, France
[5]Grupo de Medicina Xenómica, Centre for Research in Molecular Medicine and Chronic Diseases, CIBERER, Universidade de Santiago de Compostela, Santiago 15728, Spain
[6]Centre Référent des Troubles du Langage et des Apprentissages, Hôpital Raymond Poincaré, APHP, Garches 92380, France
[7]Département de la femme enfant et adolescent, Service de pédiatrie générale, Centre de coordination interdisciplinaire et de soins des maladies rares et complexes de l'enfant, de l'adolescent ou adolescente (CORAIL), Hôpitaux universitaires de Genève, Genève 1205, Suisse

Received: 14 October 2024 / Accepted: 9 January 2025
Published online: 13 February 2025

References

1. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. J Speech Lang Hear Res. 1997;40:1245–60.
2. Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. J Child Psychol Psychiatry. 2016;57:1247–57.
3. Weindrich D, Jennen-Steinmetz C, Laucht M, Esser G, Schmidt MH. Epidemiology and prognosis of specific disorders of language and scholastic skills. Eur Child Adolesc Psychiatry. 2000;9:186–94.
4. Bishop DVM, Snowling MJ, Thompson PA, Greenhalgh T, 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. 2017;58:1068–80.
5. Law J, Boyle J, Harris F, Harkness A, Nye C. Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. Int J Lang Commun Disord. 2000;35:165–88.
6. den Hoed J, Fisher SE. Genetic pathways involved in human speech disorders. Curr Opin Genet Dev. 2020;65:103–11.
7. Chilosi AM, Podda I, Ricca I, Comparini A, Franchi B, Fiori S, et al. Differences and commonalities in children with Childhood Apraxia of Speech and Comorbid Neurodevelopmental disorders: a multidimensional perspective. J Pers Med. 2022;12:313.
8. Choudhury N, Benasich AA. A family aggregation study: the influence of family history and other risk factors on language development. J Speech Lang Hear Res. 2003;46:261–72.
9. Lewis BA, Thompson LA. A study of developmental speech and language disorders in twins. J Speech Hear Res. 1992;35:1086–94.
10. Eising E, Carrion-Castillo A, Vino A, Strand EA, Jakielski KJ, Scerri TS, et al. A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. Mol Psychiatry. 2019;24:1065–78.
11. Hildebrand MS, Jackson VE, Scerri TS, Van Reyk O, Coleman M, Braden RO, et al. Severe childhood speech disorder: gene discovery highlights transcriptional dysregulation. Neurology. 2020;94:e2148–67.
12. Yahia A, Li D, Lejerkrans S, Rajagopalan S, Kalnak N, Tammimies K. Whole exome sequencing and polygenic assessment of a Swedish cohort with severe developmental language disorder. Hum Genet. 2024;143:169–83.
13. Coolen M, Altin N, Rajamani K, Pereira E, Siquier-Pernet K, Puig Lombardi E, et al. Recessive PRDM13 mutations cause fatal perinatal brainstem dysfunction with cerebellar hypoplasia and disrupt Purkinje cell differentiation. Am J Hum Genet. 2022;109:909–27.
14. Ucuncu E, Rajamani K, Wilson MSC, Medina-Cano D, Altin N, David P, et al. MINPP1 prevents intracellular accumulation of the chelator inositol hexakisphosphate and is mutated in Pontocerebellar Hypoplasia. Nat Commun. 2020;11:6087.
15. Nicolle R, Siquier-Pernet K, Rio M, Guimier A, Ollivier E, Nitschke P, et al. 16p13.11p11.2 triplication syndrome: a new recognizable genomic disorder characterized by optical genome mapping and whole genome sequencing. Eur J Hum Genet. 2022;30:712–20.
16. Golzio C, Willer J, Talkowski ME, Oh EC, Taniguchi Y, Jacquemont S, et al. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature. 2012;485:363–7.
17. Richter M, Murtaza N, Scharrenberg R, White SH, Johanns O, Walker S, et al. Altered TAOK2 activity causes autism-related neurodevelopmental and cognitive abnormalities through RhoA signaling. Mol Psychiatry. 2019;24:1329–50.
18. Mollon J, Almasy L, Jacquemont S, Glahn DC. The contribution of copy number variants to psychiatric symptoms and cognitive ability. Mol Psychiatry. 2023;28:1480–93.
19. Riggs ER, Andersen EF, Cherry AM, Kantarci S, Kearney H, Patel A, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet Med. 2020;22:245–57.
20. Mirzaa GM, Chong JX, Piton A, Popp B, Foss K, Guo H, et al. De novo and inherited variants in ZNF292 underlie a neurodevelopmental disorder with features of autism spectrum disorder. Genet Med. 2020;22:538–46.
21. Shen J, Oza AM, Del Castillo I, Duzkale H, Matsunaga T, Pandya A, et al. Consensus interpretation of the p.Met34Thr and p.Val37Ile variants in GJB2 by the ClinGen hearing loss Expert Panel. Genet Med. 2019;21:2442–52.

22. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17:405–24.
23. Pavone P, Ruggieri M, Marino SD, Corsello G, Pappalardo X, Polizzi A, et al. Chromosome 15q BP3 to BP5 deletion is a likely locus for speech delay and language impairment: report on a four-member family and an unrelated boy. Mol Genet Genomic Med. 2020;8:e1109.
24. Loviglio MN, Leleu M, Männik K, Passeggeri M, Giannuzzi G, van der Werf I, et al. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. Mol Psychiatry. 2017;22:836–49.
25. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. 2008;358:667–75.
26. Taylor CM, Smith R, Lehman C, Mitchel MW, Singer K, Weaver WC, et al. 16p11.2 recurrent deletion. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Beproxian LJ, et al. editors. GeneReviews® [Internet]. Seattle (WA). University of Washington, Seattle; 1993. [cited 2024 Aug 23]. http://www.ncbi.nlm.nih.gov/books/NBK11167/.
27. Hippolyte L, Maillard AM, Rodriguez-Herreros B, Pain A, Martin-Brevet S, Ferrari C, et al. The number of genomic copies at the 16p11.2 Locus modulates Language, Verbal Memory, and inhibition. Biol Psychiatry. 2016;80:129–39.
28. Mei C, Fedorenko E, Amor DJ, Boys A, Hoeflin C, Carew P, et al. Deep phenotyping of speech and language skills in individuals with 16p11.2 deletion. Eur J Hum Genet. 2018;26:676–86.
29. Kim SH, Green-Snyder L, Lord C, Bishop S, Steinman KJ, Bernier R, et al. Language characterization in 16p11.2 deletion and duplication syndromes. Am J Med Genet B Neuropsychiatr Genet. 2020;183:380–91.
30. Uliana V, Bonatti F, Zanatta V, Mozzoni P, Martorana D, Percesepe A. Spectrum of X-linked intellectual disabilities and psychiatric symptoms in a family harbouring a Xp22.12 microduplication encompassing the RPS6KA3 gene. J Genet. 2019;98:10.
31. Madrigal I, Rodríguez-Revenga L, Badenas C, Sánchez A, Martinez F, Fernandez I, et al. MLPA as first screening method for the detection of microduplications and microdeletions in patients with X-linked mental retardation. Genet Med. 2007;9:117–22.
32. Yan J, Zhang F, Brundage E, Scheuerle A, Lanpher B, Erickson RP, et al. Genomic duplication resulting in increased copy number of genes encoding the sister chromatid cohesion complex conveys clinical consequences distinct from Cornelia De Lange. J Med Genet. 2009;46:626–34.
33. Kariminejad A, Ghaderi-Sohi S, Gholami S, Najafi K, Kariminejad R, Hennekam RCM. 5p13 microduplication in a malformed fetus and his unaffected father. Am J Med Genet A. 2023;191:370–7.
34. Osaki E, Nishina Y, Inazawa J, Copeland NG, Gilbert DJ, Jenkins NA, et al. Identification of a novel sry-related gene and its germ cell-specific expression. Nucleic Acids Res. 1999;27:2503–10.
35. Bai S, Fu K, Yin H, Cui Y, Yue Q, Li W, et al. Sox30 initiates transcription of haploid genes during late meiosis and spermiogenesis in mouse testes. Development. 2018;145:dev164855.
36. De Martino SP, Errington F, Ashworth A, Jowett T, Austin CA. sox30: a novel zebrafish sox gene expressed in a restricted manner at the midbrain-hindbrain boundary during neurogenesis. Dev Genes Evol. 1999;209:357–62.
37. Paine I, Posey JE, Grochowski CM, Jhangiani SN, Rosenheck S, Kleyner R, et al. Paralog studies augment Gene Discovery: DDX and DHX genes. Am J Hum Genet. 2019;105:302–16.
38. Järvelä I, Määttä T, Acharya A, Leppälä J, Jhangiani SN, Arvio M, et al. Exome sequencing reveals predominantly de novo variants in disorders with intellectual disability (ID) in the founder population of Finland. Hum Genet. 2021;140:1011–29.
39. Backx L, Vermeesch J, Pijkels E, de Ravel T, Seuntjens E, Van Esch H. PPP2R2C, a gene disrupted in autosomal dominant intellectual disability. Eur J Med Genet. 2010;53:239–43.
40. Wu M-Y, Tsai T-F, Beaudet AL. Deficiency of Rbbp1/Arid4a and Rbbp1l1/Arid4b alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain. Genes Dev. 2006;20:2859–70.
41. Yap CX, Alvares GA, Henders AK, Lin T, Wallace L, Farrelly A, et al. Analysis of common genetic variation and rare CNVs in the Australian Autism Biobank. Mol Autism. 2021;12:12.
42. Antaki D, Guevara J, Maihofer AX, Klein M, Gujral M, Grove J, et al. A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. Nat Genet. 2022;54:1284–92.
43. Cirnigliaro M, Chang TS, Arteaga SA, Pérez-Cano L, Ruzzo EK, Gordon A, et al. The contributions of rare inherited and polygenic risk to ASD in multiplex families. Proc Natl Acad Sci U S A. 2023;120:e2215632120.

Publisher's note