A clustering of missense variants in the crucial chromatin modifier WDR5 defines a new neurodevelopmental disorder

Lot Snijders Blok^{1,2,3*}, Jolijn Verseput^{1*}, Dmitrijs Rots^{1,3}, Hanka Venselaar⁴, A. Micheil Innes⁵, Connie Stumpel⁶, Katrin Ounap^{7,8}, Karit Reinson^{7,8}, Eleanor G. Seaby^{9,10}, Shane McKee¹¹, Barbara Burton¹², Katherine Kim¹², Johanna M. van Hagen¹³, Quinten Waisfisz¹³, Pascal Joset¹⁴, Katharina Steindl¹⁴, Anita Rauch¹⁴, Dong Li¹⁵, Elaine Zackai^{16,17}, Sarah Sheppard^{15,16}, Beth Keena¹⁶, Hakon Hakonarson^{15,16,17}, Andreas Roos^{18,19}, Nicolai Kohlschmidt²⁰, Anna Cereda²¹, Maria Iascone²², Erika Rebessi²³, Kristin D. Kernohan²⁴, Philippe M. Campeau^{25,26}, Francisca Millan²⁷, Jesse A. Taylor²⁸, Hanns Lochmüller²⁹, Martin R. Higgs³⁰, Amalia Goula³⁰, Birgitta Bernhard³¹, Simon E. Fisher^{2,32}, Han G. Brunner^{1,3,6}, Tjitske Kleefstra^{1,3}

**These authors contributed equally to this work*

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

Affiliations:

1. Human Genetics Department, Radboud University Medical Center, Nijmegen, the Netherlands.
2. Language & Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands.
3. Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands.
4. Centre for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, 6500HB, the Netherlands.
5. The Department of Medical Genetics and Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada.
6. Department of Clinical Genetics and School for Oncology and Developmental Biology (GROW), Maastricht UMC+, Maastricht, the Netherlands.
7. Department of Clinical Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia.
8. Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia.
9. Translational Genomics Group, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
10. Genomic Informatics Group, University Hospital Southampton, Southampton, United Kingdom.
11. Northern Ireland Regional Genetics Service, Belfast City Hospital, Belfast HSC Trust, Belfast BT9 7AB, United Kingdom.
12. Ann and Robert H. Lurie Children's Hospital and Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.
13. Department of Human Genetics, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.
14. Institute of Medical Genetics, University of Zurich, CH-8952 Schlieren, Switzerland.
15. Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
16. Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA.
17. Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.
18. Department of Neuropediatrics, Developmental Neurology and Social Pediatrics, Centre for Neuromuscular Disorders in Children, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
19. Children's Hospital of Eastern Ontario Research Institute, Ottawa, ON, Canada
20. Institute of Clinical Genetics and Tumor Genetics, Bonn, Germany.
21. Department of Pediatrics, ASST Papa Giovanni XXIII, Bergamo, Italy.
22. Laboratory of Medical Genetics, ASST Papa Giovanni XXIII, Bergamo, Italy.
23. Pediatric Neurological Unit and Epilepsy Center, Fatebenefratelli Hospital, Milan, Italy.

24. Newborn Screening Ontario, Children's Hospital of Eastern Ontario and Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada.
25. CHU Sainte-Justine Research Center, Montreal, QC H3T 1C5, Canada.
26. Sainte-Justine Hospital, University of Montreal, Montreal, QC H3T 1C5, Canada.
27. GeneDx, Gaithersburg, MD 20877, USA.
28. Division of Plastic Surgery, Department of Surgery, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
29. Children's Hospital of Eastern Ontario Research Institute; Division of Neurology, Department of Medicine, the Ottawa Hospital; Brain and Mind Research Institute, University of Ottawa, Ottawa, Canada
30. Lysine Methylation and DNA Damage Laboratory, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK
31. North West Thames Regional Genetics Service, 759 Northwick Park Hospital, Harrow, United Kingdom.
32. Donders Institute for Brain, Cognition & Behaviour, Radboud University, Nijmegen, the Netherlands.

Corresponding authors: Lot Snijders Blok (lot.snijdersblok@radboudumc.nl) & Tjitske Kleefstra (tjitske.kleefstra@radboudumc.nl)

Abstract

WDR5 is a broadly studied, highly conserved protein involved in a wide array of biological functions. Among these functions, WDR5 is a part of several protein complexes that affect gene regulation via post-translational modification of histones. Here, we present data from ten unrelated individuals with six different rare *de novo* missense variants in *WDR5*; one identical variant was found in four individuals, and another variant in two individuals. All ten individuals had neurodevelopmental disorders including speech/language delays (N=10), intellectual disability (N=8), epilepsy (N=6) and autism spectrum disorder (N=4). Additional phenotypic features included abnormal growth parameters (N=6), heart anomalies (N=2) and hearing loss (N=2). All six missense variants occurred in regions of the *WDR5* locus that are known to be extremely intolerant for variation. Three-dimensional structures indicate that all the residues affected by these variants are located at the surface of one side of the WDR5 protein. It is predicted that five out of the six amino-acid substitutions disrupt interactions of WDR5 with RbBP5 and/or KMT2A/C, as part of the COMPASS family complexes. Thus, we define a new neurodevelopmental disorder associated with missense variants in *WDR5* and a broad range of associated features including intellectual disability, speech/language impairments, epilepsy and autism spectrum disorders. This finding highlights the important role of COMPASS family proteins in neurodevelopmental disorders.

Introduction

WDR5 is a small highly conserved protein that is able to interact with a large number of other proteins^{1,2}. As a core constituent of many different chromatin-related protein complexes, it controls crucial regulatory processes during development^{3,4}. The indispensable function of WDR5 is illustrated by its high evolutionary conservation. Even very basic multicellular organisms such as *Trichoplax adhaerens* have a protein with around 90% similarity to the 334 amino acids of the human orthologue^{4,5}. WDR5 is a member of the WD40 repeat family and has seven WD40 domains that each forms a propeller-like wing, resulting in a final barrel-shaped protein⁶. Using two binding sites on opposite sides of the protein, WDR5 can act as an adapter to link different proteins and form protein complexes. Since the protein is highly multifunctional and ubiquitously expressed⁷ (data available from <https://v19.proteinatlas.org/ENSG00000196363-WDR5>), the unavailability of a well-functioning WDR5 could potentially impact myriad downstream processes.

Most of the protein complexes that WDR5 participates in affect gene regulation via post-translational modification of histones. The MLL/SET complexes (also known as COMPASS family complexes) catalyse histone 3 lysine 4 (H3K4) di- and trimethylation^{8,9} and the NSL and ATAC complexes are involved in histone acetylation^{10,11}. WDR5 can also be part of an embryonic stem cell-specific NuRD complex that combines nucleosome sliding capacities with histone deacetylation activity¹². In addition to interactions with other proteins, WDR5 is also able to bind to >1000 different endogenous RNAs (including long non-coding RNAs), and binding to certain long non-coding RNAs can be crucial for WDR5 stability and function in cells¹³. A recent study linked WDR5, as part of the COMPASS complex, to a newly discovered genetic compensation mechanism: nonsense-induced transcriptional compensation¹⁴. In short, this mechanism is triggered by a truncating variant in a gene, and leads to the expression of related genes, thereby compensating for the effects of a deleterious variant¹⁵. One of the most well-studied aspects of WDR5 function is its role in embryonic stem cell (ESC) self-renewal and maintenance of a pluripotent state^{16,17}. Recently, a direct interaction between the p53 protein and WDR5 has been uncovered, in which mouse ESC stem cell fate (the differentiation into neuroectoderm or mesoderm) is regulated in a p53-dependent manner¹⁸. Thus, WDR5 has already been implicated in multiple different molecular pathways and mechanisms, and this list is growing steadily.

While the biological functions of the WDR5 protein have been studied from numerous angles, there is still little known about the impact of rare germline variants in the gene that encodes it. Using data from large-scale sequencing resources, it is clear that *WDR5* is extremely intolerant for loss-of-function variation: in both the gnomAD database¹⁹ (version 2.1.1; containing sequencing data of 141,456 individuals) and the TOPMED database²⁰ (containing sequencing data of 62,784 individuals) not a single truncating variant in *WDR5* is reported. Similarly, *WDR5* is also depleted for missense variants. Therefore, the initial finding of a *de novo* missense variant (p.(Thr208Met)) in *WDR5* in a child with a developmental speech disorder²¹, prompted us to investigate the effects and possible pathogenicity of rare germline variants in this gene. We collected clinical information on this proband and nine additional individuals with rare *de novo* germline variants in *WDR5*, collated from several clinical exome or genome sequencing studies. We used a range of

in silico tools and analyses of variants using three-dimensional structures in order to evaluate the likely consequences of the different variants found.

Methods

Study participants and consent

Individuals with *WDR5* variants were identified via matchmaking using GeneMatcher²², the Dutch genetic diagnostic variant classification database (VKGL database)²³⁻²⁵, ClinVar²⁶ and denovo-db²⁷. Clinical data and details on variants were collected in a Castor EDC database²⁸. Informed consent to share and publish these data was given by all individuals or their legal guardian. Consent for publication of photographs was obtained separately. Genetic testing and research were performed in accordance with protocols approved by the local Institutional Review Boards.

Next generation sequencing and *in silico* variant analyses

Details on next generation sequencing methods used to identify the *WDR5* variants found in all individuals are included in table S1. Variants were analysed using Alamut Visual 2.10. Conservation was studied using a Clustal²⁹ alignment of *WDR5* amino acid sequences extracted from Uniprot (human, mouse and *C. elegans*)³⁰. To assess the likelihood of pathogenicity, the prediction programs SIFT³¹, PolyPhen-2³² and CADD v1.4³³ were used.

Three-dimensional (3D) protein modelling

The effects of the identified variants on the *WDR5* protein and its interaction with other proteins in the COMPASS family complexes were analyzed using YASARA View³⁴ with FoldX v4.0 plugin³⁵. For the *WDR5* structure, PDB file 2GNQ was used. PDB files 6KIV and 6KIW³⁶ were used for the analysis of the core MLL1 and MLL3 complexes, respectively; the 6UH5 file³⁷ of the yeast COMPASS model was used for the comparison with the human MLL1 complex. To optimize the position of amino-acid sidechains, all the PDB files that were used were corrected by the FoldX repair function using default settings. Different protein structures were aligned with SHEBA procedure³⁸, as implemented in YASARA.

Results

Identification and clinical characterization of individuals with *de novo* *WDR5* variants

We report clinical and molecular data on ten unrelated individuals with a *de novo* missense variant in *WDR5*. Six different missense variants were reported in these ten individuals: p.(Ala169Pro), p.(Arg196Cys), p.(Ala201Val), p.(Thr208Met), p.(Asp213Asn) and p.(Lys245Arg). The p.(Thr208Met) variant was found in four unrelated individuals, and the p.(Arg196Cys) variant in two unrelated individuals.

All individuals had neurodevelopmental disorders with a spectrum of overlapping associated features (Figure 1 and table S1). Intellectual disability (ID) was present in eight out of ten individuals, with a severity ranging from moderate ID (IQ 35-50, five individuals) to mild ID (IQ 50-70, three individuals). Of the two remaining individuals, one individual had a borderline level of intellectual functioning (IQ 70-85) and another individual had no intellectual disability. Interestingly, in all ten individuals speech delays were reported. Two individuals were non-verbal, and three other individuals had a developmental language disorder diagnosis (mixed expressive/receptive language disorder in two individuals, and expressive language disorder in one). One of these latter three individuals was also reported with verbal dyspraxia. In addition, five individuals were reported to have nasal speech, and one individual was reported with persistent stuttering.

All but one of the individuals with *WDR5* variants had motor development delays, with the age of first steps ranging from 12-40 months. Hypotonia was reported in six individuals. Concerning the behavioural phenotype, four individuals had an autism spectrum disorder (ASD) diagnosis, and two individuals were diagnosed with Attention Deficit Hyperactivity Disorder (ADHD).

Six out of the ten individuals were diagnosed with epilepsy, with different forms of presentations varying from absence seizures to refractory generalized myoclonic epilepsy. In four of these six individuals, the disorder was only present in childhood and medication could be discontinued at a later age. A brain MRI scan was performed for seven individuals, showing different types of abnormalities in three of these scans: mild ventricular dilatation with thinning of the posterior corpus callosum in one individual, subtle grey matter heterotopias in another individual, and periventricular gliosis in the third individual with an abnormal MRI result.

Individuals with *WDR5* variants showed divergent growth parameters. One individual had macrocephaly (head circumference $\geq +2$ SD) and another had microcephaly (≤ -2 SD). Two individuals had tall stature ($\geq +2$ SD). A weight of $+2$ SD or more (for height) was seen in four individuals in total, including all three adult individuals in our study. One individual had a low weight (≤ -2 SD). No clear correlation between height, weight and head circumference was observed (table S1), with the exception of one individual which showed a striking pattern of progressive overgrowth (height $+3$ SD, weight $+5$ SD, head circumference $+4$ SD at adult age) and another individual with a milder generalized overgrowth phenotype (height $+3$ SD, weight $+1.5$ SD, head circumference $+1.5$ SD).

Different abnormalities of the skeleton and limbs were present in a subset of individuals. Bilateral clubfeet were reported in one individual and another individual had hemihypertrophy of the left leg. One individual was reported with a hemivertebra of L5 and kyphosis (possibly secondary to the hemivertebra), and another individual had scoliosis. Two individuals had single palmar creases. In two individuals, heart abnormalities were reported: cardiac arrhythmias and decompensated heart failure requiring surgery in one individual, and left ventricular non-compaction cardiomyopathy in another individual. Three individuals were reported with frequent infections of the ears and/or airways. One individual had strabismus, another individual had amblyopia and hyperopia with astigmatism.

Significant facial dysmorphisms were noted in only a subset of individuals. When comparing facial features of seven individuals in our cohort, overlapping facial features included a bulbous nasal tip, low-set, posteriorly rotated and/or dysplastic ears, ptosis and thin lip vermilion (Figure 1B). Two individuals in our cohort had distinct facial features compared to the others, with severe micrognathia (requiring tracheostomy in one), a small mouth and prominent down-slanting palpebral fissures. These two individuals both had conductive hearing loss too, a feature not reported in any of the other individuals. Clinical features reported in individuals in our cohort are described in more detail in table S1.

***In silico* analysis of variants**

The missense variants in our cohort are all located within or flanking the fourth and the fifth WD40 domain of WDR5, with each WD40 domain encoding one ‘propeller’ of the three-dimensional WDR5 structure (Figure 2A and 2B). All missense variants in our study were absent from the gnomAD database¹⁹. We used *in silico* prediction programs to evaluate likely pathogenicity for the six different missense variants. The resulting scores are summarised in Table 1. All CADD scores were above 22, while SIFT and Polyphen-2 predicted three out of the six variants to be pathogenic: p.(Ala169Pro), p.(Arg196Cys) and p.(Thr208Met).

Three-dimensional structure analysis

Using three-dimensional structure analysis, we determined that the residues affected by all the identified missense variants cluster together on one side of the WDR5 surface, suggesting that these variants may disrupt specific interactions with other proteins (Figure 2B). Different surfaces of WDR5 are known to mediate interactions with different proteins, thereby linking them together. In previous co-precipitation experiments with WDR5 and short fragments of proteins from complexes in which WDR5 is involved, two distinct binding sites were identified: the ‘WDR5-interacting’ (WIN) site³⁹⁻⁴¹ and the ‘WDR5-binding motif’ (WBM) site^{41,42}, located on opposite sides (often referred to as ‘top’ and ‘bottom’, respectively) of the protein. The variants found in our study are not located in the vicinity of these well-studied binding locations.

However, recently published cryo-electron microscopy 3D structures of the core MLL1 and MLL3 complexes revealed a region located between the WIN and WBM binding sites, that is involved in the interaction with RbBP5 and/or histone-lysine methylase (KMT) enzymes in these complexes³⁶. Five out of the six variants that we identified map within this RbBP5/KMT interaction region. Based on the 3D structure analysis of the MLL1 and MLL3 complexes, p.(Ala169Pro) and p.(Asp213Asn) are predicted to affect the WDR5 interaction with KMT enzymes, and p.(Ala201Val) and p.(Thr208Met) are predicted to affect the interaction with the RbBP5 enzymes, while p.(Arg196Cys) most likely influences the interaction with both enzymes (Figure 2C-F). The effects of the p.(Lys245Arg) variant cannot be predicted using the currently available 3D structures. A detailed description of the predicted effects of all variants, from the perspective of the structural modelling analyses, is provided in Supplementary Note 1.

Discussion

We identified rare *de novo* missense variants in *WDR5*, a gene that encodes a core chromatin regulator and is extremely intolerant for variation in the population. We clarify the molecular and phenotypic consequences of these rare variants which define a novel Mendelian disorder. By using the GeneMatcher database and international collaborations, we collected clinical data on a cohort of ten individuals with a *de novo* variant in *WDR5*. We studied all variants in three-dimensional conformations of *WDR5* in interaction with COMPASS complex family subunits, and showed that the missense variants are predicted to affect important binding sites of *WDR5*. By combining data from these clinical and *in silico* approaches, we can confidently link *WDR5* to a neurodevelopmental disorder with a broad spectrum of associated features, further confirming that proteins in the COMPASS complex family are important contributors to neurodevelopment.

All individuals in our cohort had a heterozygous *de novo* missense variant in *WDR5*: six different missense variants were found, affecting amino acids on one surface of the *WDR5* protein. The recurrence of the p.(Thr208Met) and p.(Arg196Cys) variant in four and two unrelated individuals, respectively, points to the presence of ‘hotspots’ for recurrent variants in *WDR5*. Three different missense variants in our cohort, p.(Ala169Pro), p.(Arg196Cys) and p.(Thr208Met), were found to locate at adjacent amino positions in the three-dimensional structure of *WDR5*. The fact that all identified variants are missense variants, the recurrence of specific variants in unrelated individuals and the spatial clustering of the variants on the protein surface of *WDR5*, all suggest that specific pathogenic mechanisms that might be at play are not just loss-of-function or haploinsufficiency.

To the best of our knowledge, truncating variants (e.g. frameshift or nonsense variants) in *WDR5* have not been identified so far: not in our cohort or any disease cohort in literature, nor in healthy individuals (e.g. in the gnomAD or TOPMED database). According to sequencing data from the gnomAD database, *WDR5* is extremely intolerant for both missense and loss-of-function variation. The gene has a LOEUF score of 0.124, which is well within the first decile of most highly constrained genes against loss-of-function¹⁹. In contrast to the absence of truncating variants, heterozygous chromosomal microdeletions encompassing the whole *WDR5* gene have been reported; the Decipher database lists eleven heterozygous deletions that include *WDR5*⁴³. This means that haploinsufficiency for *WDR5* is compatible with life, but it is unclear how the loss of *WDR5* contributes to specific phenotypes found in individuals with these deletions, as all deletions are larger than 3 Mb and encompass many other genes as well.

Analysis of the missense variants in a three-dimensional structure of *WDR5* in the context of the MLL1 or MLL3 complex, revealed that all but one were located at amino-acid positions that are important for binding of *WDR5* with other proteins of these complexes. Interestingly, analysis of the intolerance landscape of the *WDR5* gene in the three-dimensional structure of the encoded protein shows that *WDR5* is generally intolerant to missense variants, but that residues interacting with other proteins are most intolerant for normal variation (Figure S1). For five out of six missense variants, we predict that the

corresponding amino-acid change disrupts the interaction of WDR5 with the MLL1/MLL3 complex subunits RbBP5 and with KMT2A/C simultaneously or separately. WDR5 is a crucial core protein within the COMPASS complex family; it is essential for complex assembly and activity^{44,45}. Based on 3D analysis of the MLL1 and MLL3 complexes, it seems that differently composed COMPASS complexes make use of different interaction surfaces of WDR5. Some variants might therefore disrupt interactions in only one specific complex. As the core function of the WDR5 protein seems to be to act as an ‘adapter’ protein and form links between different molecules, disruption of protein-protein interactions within the complex might have important effects on complex activity.

However, it is important to take into account that the extensive and detailed three-dimensional structures that we used for these analyses are only available for the MLL1/3 complex, and not for all other complexes and interactions in which WDR5 is involved. Therefore, it remains unclear whether the predicted disruptive effects on WDR5 interactions are specific to those with RbBP5/KMT and MLL1/3, or whether interactions with other molecules might also be disturbed. For the p.(Lys245Arg) variant, we were not able to predict a possible pathogenic mechanism analysing available three-dimensional structures. One hypothesis could be that the variant affects a so far uncharacterized interaction site with RbBP5 or KMT2A/C, as a comparison of available three-dimensional structures between human MLL1 and yeast COMPASS complex suggest even more extensive interaction surfaces between WDR5 and histone methylases (Figure S2). Another hypothesis is that the p.(Lys245Arg) variant affects the interaction with other molecules that are not involved in the MLL1/MLL3 complex. Lastly, although the *de novo* occurrence of the p.(Lys245Arg) and the phenotypic similarity to the rest of the cohort suggest pathogenicity, the possibility that this is a benign variant without any effect on WDR5 function cannot be excluded.

The individuals in our study showed a broad range of clinical features: neurodevelopmental phenotypes with several additional abnormalities. All individuals had developmental delays, with mild or moderate intellectual disability being present in the majority. Speech and language problems were a prominent feature, as was epilepsy. Abnormal height, weight and head circumferences were frequently seen, and two individuals had a remarkable distinct facial phenotype with severe micrognathia, a small mouth and down-slanted palpebral fissures. When comparing the phenotypes corresponding to different variants in our study, no clear genotype-phenotype correlation was established. Even in four individuals with the exact same p.(Thr208Met) variant a different clinical presentation was seen: e.g. borderline vs. moderate intellectual disability, normal growth parameters vs. a generalized overgrowth phenotype, and variability in the presence of additional phenotypic features. Altogether, *WDR5*-associated disorder can be characterized as a neurodevelopmental disorder with variable expressivity of associated features. This is in line with other disorders caused by variants in genes encoding COMPASS complex family subunits, such as Kabuki syndrome (caused by variants in *KMT2D* or *KDM6A*)⁴⁶, Kleeftstra syndrome type 2 (caused by variants in *KMT2C*)⁴⁷ and the neurodevelopmental disorder associated with *SETD1A* loss-of-function variants⁴⁸⁻⁵⁰. In all these COMPASS complex-associated disorders a variable spectrum of associated features can be present in varying degrees of severity, including intellectual disability, speech and language delays, autism spectrum disorders, epilepsy and abnormal growth parameters⁴⁶⁻⁵⁰.

This study represents the first characterization of a Mendelian disorder associated with germline variants in *WDR5*, and was initiated after the report of a *de novo* *WDR5* variant in a child with a speech disorder²¹. It is worth mentioning that one additional *de novo* variant in *WDR5* is reported in the literature: a p.(Lys7Gln) variant, found in a child with a conotruncal heart defect with a right aortic arch⁵¹. This missense variant is located in the N-terminal tail of *WDR5*, an intrinsically disordered region of the protein (not available for three-dimensional protein structure analysis), which is not involved in the beta-propeller structure of *WDR5*, and has been shown to be dispensable for COMPASS complex assembly⁵². A study in *Xenopus tropicalis* shows that this p.(Lys7Gln) variant might interfere with the ability of *WDR5* to localize to the bases of left-right organizer cilia, independent from the H3K4-methylation-related functions of *WDR5*⁵³. As p.(Lys7Gln) is located in a different region of the *WDR5* protein compared to the variants found in our study, and since complete phenotypic details are not available for this individual, it is currently unclear whether this reported individual has the *WDR5*-associated neurodevelopmental disorder presented in this study, or whether this specific variant gives rise to a different disorder with different pathogenic mechanisms.

All in all, with this study *WDR5* can be added to the list of genes robustly associated with autosomal dominant neurodevelopmental disorders. All variants found so far are missense variants, and although they are not always predicted to be pathogenic by the commonly used prediction programs, our three-dimensional protein structure analysis showed that it is very likely that five out of the six found missense variants disturb important protein-protein interactions. Future studies are needed to confirm that these *in silico* observations in 3D structures indeed affect protein-protein interactions between *WDR5* and RbBP5 and KMT enzymes, and what the downstream consequences are on H3K4 methyltransferase activity. But taken together, with our study we implicate *WDR5* variants in an autosomal dominant disorder associated with intellectual disability, speech and language delays, epilepsy and autism spectrum disorder, thereby highlighting the role of COMPASS complex family protein disruptions in neurodevelopmental disorders.

Figures

Figure 1: Clinical features reported in individuals with *WDR5* variants

A

B

Photographs not available in this preprint.

A: Graphical overview of clinical features reported in ten individuals with WDR5 missense variants. Growth parameters are shown as standard deviations to the mean for a certain age. All graphs include data for all ten individuals (N=10). 'EEG abn.' = EEG abnormalities.

B: Facial images of seven individuals with a heterozygous WDR5 variant (photographs not available in this preprint). Several overlapping facial features are seen, such as a bulbous nasal tip, low-set, posteriorly rotated and/or dysplastic ears, ptosis and thin upper lip vermilion. In addition to this, two individuals have severe micrognathia, a small mouth and down-slanting palpebral fissures.

Figure 2

A: Linear structure of WDR5 protein (334 amino acids) with the seven different WD40-domains and all identified missense variants shown; in total six different missense variants were found in ten individuals, as one variant (p.(Thr208Met)) was found in four unrelated individuals and another variant (p.(Arg196Cys)) was found in two unrelated individuals.

B: Three-dimensional visualization of WDR5 (PDB 2GNQ), locations of the amino acids involved in missense variants are shown with magenta balls. Colours of the different domains match with the colours used in panel A.

C: WDR5 (green) is shown as part of the core MLL1 complex, with RbBP5 (yellow), ASH2L (blue), DPY30 (purple) and KMT2A (cyan) (PDB:6KIV). The nucleosome is shown in grey. The locations of the amino acids affected in patients identified in this study are shown with balls (magenta).

D, E, F: WDR5 (green, p.33-332) is shown together with RbBP5 (yellow, p.1-380) and KMT2A (cyan, p.3764-3969), as part of the core MLL1 complex (PDB:6KIV). The locations of the amino acids affected in patients identified in this study are shown with balls (magenta), from three different angles: facing the WIN site (D), the WBM site (E) and a side between WIM and WBM sites (F).

Data availability

Figure 1B and Table S1 are available from the corresponding author on request.

Declaration of interests

F.M. is a full time employee at GeneDx, Inc. The authors declare no other conflicts of interest.

Acknowledgements

We thank all included individuals and their families for their contribution to this research project.

This work was generated within ITHACA: European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability. Funding was provided by the Netherlands Organization for Scientific Research (NWO) Gravitation Grant 24.001.006 to the Language in Interaction Consortium (L.S.B., S.E.F., and H.G.B.), the Max Planck Society (S.E.F.) and the Netherlands Organization for Health Research and Development (ZonMw grant 91718310 to T.K.). The research of A.C., M.I. and E.R. was supported by PROGETTEO GENE (GENE = Genomic analysis Evaluation Network) founded by PROGETTI DI INNOVAZIONE IN AMBITO SANITARIO E SOCIO SANITARIO (BANDO EX DECRETO N. 2713 DEL 28/02/2018). A.R. and N.K. acknowledge funding from the European Regional Development Fund (ERDF). K.Ö. and K.R. were supported by the Estonian Research Council grants PUT355 and PRG471. The Broad Center for Mendelian Genomics (UM1 HG008900) is funded by the National Human Genome Research Institute with supplemental funding provided by the National Heart, Lung, and Blood Institute under the Trans-Omics for Precision Medicine (TOPMed) program and the National Eye Institute.

Individuals 3 and 4 in this study were part of the DDD study cohort. The DDD study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003]. This study makes use of DECIPHER (<http://decipher.sanger.ac.uk>), which is funded by Wellcome. See Nature PMID: 25533962 or www.ddduk.org/access.html for full acknowledgement.

References

1. Trievel RC, Shilatifard A. WDR5, a complexed protein. *Nat Struct Mol Biol.* 2009;16(7):678-680.
2. van Nuland R, Smits AH, Pallaki P, Jansen PW, Vermeulen M, Timmers HT. Quantitative dissection and stoichiometry determination of the human SET1/MLL histone methyltransferase complexes. *Mol Cell Biol.* 2013;33(10):2067-2077.
3. Wysocka J, Swigut T, Milne TA, et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell.* 2005;121(6):859-872.
4. Guarnaccia AD, Tansey WP. Moonlighting with WDR5: A Cellular Multitasker. *J Clin Med.* 2018;7(2).
5. Copley RR. The Unicellular Ancestry of Groucho-Mediated Repression and the Origins of Metazoan Transcription Factors. *Genome Biol Evol.* 2016;8(6):1859-1867.
6. Xu C, Min J. Structure and function of WD40 domain proteins. *Protein Cell.* 2011;2(3):202-214.
7. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science.* 2015;347(6220):1260419.
8. Roguev A, Schaft D, Shevchenko A, et al. The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J.* 2001;20(24):7137-7148.
9. Miller T, Krogan NJ, Dover J, et al. COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A.* 2001;98(23):12902-12907.
10. Cai Y, Jin J, Swanson SK, et al. Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J Biol Chem.* 2010;285(7):4268-4272.
11. Suganuma T, Gutierrez JL, Li B, et al. ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat Struct Mol Biol.* 2008;15(4):364-372.
12. Ee LS, McCannell KN, Tang Y, et al. An Embryonic Stem Cell-Specific NuRD Complex Functions through Interaction with WDR5. *Stem Cell Reports.* 2017;8(6):1488-1496.
13. Yang YW, Flynn RA, Chen Y, et al. Essential role of lncRNA binding for WDR5 maintenance of active chromatin and embryonic stem cell pluripotency. *Elife.* 2014;3:e02046.
14. Ma Z, Zhu P, Shi H, et al. PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components. *Nature.* 2019;568(7751):259-263.
15. Wilkinson MF. Genetic paradox explained by nonsense. *Nature.* 2019;568(7751):179-180.
16. Ang YS, Tsai SY, Lee DF, et al. Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell.* 2011;145(2):183-197.
17. Li X, Li L, Pandey R, et al. The histone acetyltransferase MOF is a key regulator of the embryonic stem cell core transcriptional network. *Cell Stem Cell.* 2012;11(2):163-178.

18. Li Q, Mao F, Zhou B, et al. p53 Integrates Temporal WDR5 Inputs during Neuroectoderm and Mesoderm Differentiation of Mouse Embryonic Stem Cells. *Cell Rep.* 2020;30(2):465-480 e466.
19. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.
20. NHLBI UoMa. The NHLBI Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Program. *BRAVO variant browser.* 2018.
21. Eising E, Carrion-Castillo A, Vino A, et al. A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. *Mol Psychiatry.* 2019;24(7):1065-1078.
22. Sobreira N, Schiettecatte F, Valle D, Hamosh A. GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum Mutat.* 2015;36(10):928-930.
23. van der Velde KJ, Imhann F, Charbon B, et al. MOLGENIS research: advanced bioinformatics data software for non-bioinformaticians. *Bioinformatics.* 2019;35(6):1076-1078.
24. Swertz MA, Dijkstra M, Adamusiak T, et al. The MOLGENIS toolkit: rapid prototyping of biosoftware at the push of a button. *BMC Bioinformatics.* 2010;11 Suppl 12:S12.
25. Swertz MA, Jansen RC. Beyond standardization: dynamic software infrastructures for systems biology. *Nat Rev Genet.* 2007;8(3):235-243.
26. Landrum MJ, Lee JM, Benson M, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062-d1067.
27. Scholte E, Van Duijn E, Dijkxhoorn Y, Noens I, Van Berckelaer-Onnes I. Vineland screener 0–6 years: manual of the Dutch adaptation. *PITS, Leiden.* 2008.
28. Castor EDC. Castor Electronic Data Capture. <https://castoredc.com>. Published 2019. Accessed August 28, 2019.
29. Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7:539. doi:10.1038/msb.2011.75. Accessed 2011.
30. Consortium TU. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2018;47(D1):D506-D515.
31. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc.* 2016;11(1):1-9.
32. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7(4):248-249.
33. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet.* 2014;46(3):310-315.
34. Krieger E, Vriend G. YASARA View—molecular graphics for all devices—from smartphones to workstations. *Bioinformatics.* 2014;30(20):2981-2982.
35. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic acids research.* 2005;33(Web Server issue):W382-W388.
36. Xue H, Yao T, Cao M, et al. Structural basis of nucleosome recognition and modification by MLL methyltransferases. *Nature.* 2019;573(7774):445-449.

37. Hsu PL, Shi H, Leonen C, Kang J, Chatterjee C, Zheng N. Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS. *Mol Cell*. 2019;76(5):712-723.e714.
38. Jung J, Lee B. Protein structure alignment using environmental profiles. *Protein Engineering, Design and Selection*. 2000;13(8):535-543.
39. Patel A, Dharmarajan V, Cosgrove MS. Structure of WDR5 bound to mixed lineage leukemia protein-1 peptide. *J Biol Chem*. 2008;283(47):32158-32161.
40. Patel A, Vought VE, Dharmarajan V, Cosgrove MS. A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J Biol Chem*. 2008;283(47):32162-32175.
41. Dias J, Van Nguyen N, Georgiev P, et al. Structural analysis of the KANSL1/WDR5/KANSL2 complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. *Genes Dev*. 2014;28(9):929-942.
42. Odho Z, Southall SM, Wilson JR. Characterization of a novel WDR5-binding site that recruits RbBP5 through a conserved motif to enhance methylation of histone H3 lysine 4 by mixed lineage leukemia protein-1. *J Biol Chem*. 2010;285(43):32967-32976.
43. Firth HV, Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*. 2009;84(4):524-533.
44. Dou Y, Milne TA, Ruthenburg AJ, et al. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol*. 2006;13(8):713-719.
45. Steward MM, Lee JS, O'Donovan A, Wyatt M, Bernstein BE, Shilatifard A. Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. *Nat Struct Mol Biol*. 2006;13(9):852-854.
46. Adam MP, Hudgins L, Hannibal M. Kabuki Syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews((R))*. Seattle (WA)1993.
47. Koemans TS, Kleefstra T, Chubak MC, et al. Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. *PLoS Genet*. 2017;13(10):e1006864.
48. Yu X, Yang L, Li J, et al. De Novo and Inherited SETD1A Variants in Early-onset Epilepsy. *Neurosci Bull*. 2019;35(6):1045-1057.
49. Singh T, Kurki MI, Curtis D, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. *Nat Neurosci*. 2016;19(4):571-577.
50. Kummeling J, Stremmelaar DE, Raun N, et al. Characterization of SETD1A haploinsufficiency in humans and Drosophila defines a novel neurodevelopmental syndrome. *Mol Psychiatry*. 2020.
51. Zaidi S, Choi M, Wakimoto H, et al. De novo mutations in histone-modifying genes in congenital heart disease. *Nature*. 2013;498(7453):220-223.
52. Schuetz A, Allali-Hassani A, Martin F, et al. Structural basis for molecular recognition and presentation of histone H3 by WDR5. *EMBO J*. 2006;25(18):4245-4252.
53. Kulkarni SS, Khokha MK. WDR5 regulates left-right patterning via chromatin-dependent and -independent functions. *Development*. 2018;145(23).