ARTICLE OPEN

Check for updates

De novo protein-coding gene variants in developmental stuttering

Else Eising[1✉], Ivana Dzinovic[2,3], Arianna Vino[1], Lottie Stipdonk[4], Martin Pavlov[2,3], Juliane Winkelmann[2,3,5,6,7], Martin Sommer[8,9], Marie-Christine J. P. Franken[4], Konrad Oexle[2,3,10,12] and Simon E. Fisher[1,11,12]

Developmental stuttering is a common childhood condition characterized by disfluencies in speech, such as blocks, prolongations, and repetitions. While most children who stutter do so only transiently, there are some for whom stuttering persists into adulthood. Rare-variant screens in families including multiple relatives with persistent stuttering have so far identified six genes carrying putative pathogenic variants hypothesized to act in a monogenic fashion. Here, we applied a complementary study design, searching instead for de novo variants in exomes of 85 independent parent-child trios, each with a child with transient or persistent stuttering. Exome sequencing analysis yielded a pathogenic variant in *SPTBN1* as well as likely pathogenic variants in *PRPF8*, *TRIO*, and *ZBTB7A* - four genes previously implicated in neurodevelopmental disorders with or without speech problems. Our results also highlighted two further genes of interest for stuttering: *FLT3* and *IREB2*. We used extensive bioinformatic approaches to investigate overlaps in brain-related processes among the twelve genes associated with monogenic forms of stuttering. Analyses of gene-expression datasets of the developing and adult human brain, and data from a genome-wide association study of human brain structural connectivity, did not find links of monogenic stuttering to specific brain processes. Overall, our results provide the first direct genetic link between stuttering and other neurodevelopmental disorders, including speech delay and aphasia. In addition, we systematically demonstrate a dissimilarity in biological pathways associated with the genes thus far implicated in monogenic forms of stuttering, indicating heterogeneity in the etiological basis of this condition.

Molecular Psychiatry; https://doi.org/10.1038/s41380-025-03170-2

INTRODUCTION

Developmental stuttering is characterized by disfluencies in speech, such as blocks, prolongations and repetitions. It generally starts early in childhood, between 2 and 5 years of age, affecting approximately 8% of children [1]. While the majority of children recover naturally or with speech therapy within a few years, stuttering persists in a subset of individuals, leading to persistent stuttering in approximately 0.8% of the population [2], three to four times as often in men than in women [1, 3]. In adults, moments of speech disfluency are usually accompanied by various cognitive, behavioural and emotional reactions, which often become a central component of stuttering [4]. Through these internal reactions and negative feedback from the environment, stuttering can have a major impact on a person's physical, psychological, and social quality of life [5].

It is well established that genetic factors play a role in the development of stuttering. Large twin studies on stuttering found a heritability of 40–80% [6–8]. The inheritance patterns observed in some large families suggest that stuttering may sometimes occur as a Mendelian (monogenic) trait, involving a single rare gene variant with a large effect size. Rare variant screens using linkage analysis followed by Sanger sequencing, and more recently using next generation sequencing, identified six genes to be associated with persistent stuttering in large families: *GNPTAB* [9], *AP4E1* [10], *IFNAR1* [11], *ARMC3* [12], *ZBTB20* [13] and *PPID* [14]. Hypothesis-driven genetic screens in unrelated people who stutter and controls also suggested an increased burden of rare variants in *GNPTG* and *NAGPA*, two genes that function in the same enzymatic pathway as *GNPTAB* [9, 15]. In addition, three genome-wide association studies identified the first genome-wide significant loci associated with stuttering [16–18], indicating that this is a genetically complex multifactorial trait for at least some of the affected population. The aetiology of stuttering therefore involves a mixture of monogenic causal factors and complex polygenic influences, but the relative contributions of these different types of genetic influence are not yet known.

The genes implicated in monogenic forms of stuttering have so far not pointed to a shared biological mechanism, but instead are

[1]Language and Genetics Department, Max Planck Institute for Psycholinguistics, 6525 XD, Nijmegen, the Netherlands. [2]Institute of Human Genetics, School of Medicine and Health, Technical University of Munich, 81675 Munich, Germany. [3]Institute of Neurogenomics, Helmholtz Zentrum München, 85764 Oberschleißheim, Germany. [4]Department of Otorhinolaryngology and Head and Neck Surgery, Erasmus University Medical Center, 3015 GD, Rotterdam, the Netherlands. [5]Bavarian Genomes Network for Rare Disorders, Munich, Germany. [6]DZPG, Deutsches Zentrum Für Psychische Gesundheit, Munich, Germany. [7]Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany. [8]Department of Neurology, University Medical Center Göttingen, 37075 Göttingen, Germany. [9]Department of Geriatrics, University Medical Center Göttingen, 37075 Göttingen, Germany. [10]Neurogenetic Systems Analysis Group, Institute of Neurogenomics, Helmholtz Munich, 85764 Munich, Germany. [11]Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 EN, Nijmegen, the Netherlands. [12]These authors contributed equally: Konrad Oexle, Simon E. Fisher. ✉email: Else.Eising@mpi.nl

Received: 22 November 2024 Revised: 24 July 2025 Accepted: 13 August 2025
Published online: 20 August 2025

involved in a wide variety of cellular functions. *GNPTAB, GNPTG* and *NAGPA* encode enzymes that synthesize mannose 6-phosphate recognition markers onto lysosomal enzymes [19]. *AP4E1* encodes a subunit of an adaptor protein involved in intracellular trafficking of vesicles of the Golgi, trans-Golgi network and endosomes, and is hypothesized to control autophagy [20]. *IFNAR1* encodes a subunit of the interferon receptor IFNR that can be activated by type I interferons during pathogen infections and autoimmune reactions [21]. *ARMC3* expresses a protein containing armadillo repeats, which produces a distinct structure that facilitates protein-protein interactions [22]. The exact function of ARMC3 is unknown, but related proteins are often involved in signal transduction and cytoskeleton regulation. *ZBTB20* encodes a transcription factor with essential roles in multiple organ systems [23]. And lastly, *PPID* encodes the protein CYP-40, a cyclophilin that acts as one of the chaperone proteins in the steroid aporeceptor complex [24]. Only a few lines of indirect evidence have pointed towards potential overlapping processes, besides the shared enzymatic pathway of *GNTPAB*, *GNPTG* and *NAGPA*. First, NAGPA and AP4E1 have been shown to interact in a yeast-two-hybrid system [10]. Second, variants in *GNPTAB* and *PPID* have been reported to affect white matter features in transgenic knock-in mouse models, as immunohistochemical staining of the Gfap astrocyte marker was decreased in the corpus callosum of a *Gnptab* mouse model [25], and the microstructure of the left corticospinal tract in the *Ppid* mouse model differed in a brain imaging analysis [14]. Other links between the genes implicated in monogenic stuttering are yet to be identified.

Overlap in disease mechanisms is also not evident from examining the other monogenic disorders linked to genes thus far implicated in stuttering. Homozygous loss of function of *GNPTAB* and *GNPTG* is a cause of mucolipidosis, a severe lysosomal storage disorder [26, 27]. Homozygous mutations in *AP4E1* are a cause of hereditary spastic paraplegia, a neurodevelopmental disorder characterized by developmental delay, moderate to severe intellectual disability and neonatal hypotonia that progresses to spasticity [28]. Homozygous loss-of-function mutations in *IFNAR1* cause an immunologic disorder characterized by increased susceptibility to viral infections [29]. Lastly, heterozygous missense variants in *ZBTB20* cause Primrose syndrome, a neurodevelopmental disorder characterized by intellectual disability, macrocephaly, unusual facial features and progressive features such as hearing loss and muscle wasting [30]. Of note, the types of mutations associated with stuttering are different from the types of mutations associated with these other Mendelian disorders: mainly heterozygous missense variants have been associated with stuttering in *GNPTAB* and *GNPTG* [9, 15], *AP4E1* [10] and *IFNAR1* [11], while a homozygous missense variant was associated with stuttering in *ZBTB20* [13]. These other Mendelian disorders therefore do not yet help elucidate important biological processes involved in stuttering. In contrast, monogenic forms of childhood apraxia of speech [31–33] and speech delay [34] are caused by genes often implicated in neurodevelopmental disorders characterized by intellectual disability, autism and epilepsy through the same types of variants, that often have functions involved in gene expression regulation, and that show co-expression during early brain development. Identifying additional genes implicated in monogenic forms of stuttering is hence important for increasing understanding of the underlying biological mechanisms.

The present study aimed to apply a novel strategy to identify genes involved in monogenic forms of stuttering, moving beyond the multiplex family approaches of prior work. We applied whole exome sequencing to 85 parent-offspring trios, each with a child who stutters or stuttered in the past, and two parents who never stuttered, and searched for de novo variants that were present in the DNA of the child but not in the DNA of both parents. This trio design has been highly successfully applied to other neurodevelopmental disorders [35] as well as to childhood apraxia of speech (in screening efforts with more modest sample sizes than in the present study) [32, 33, 36, 37], but genetic research on stuttering has yet to make use of this. Next, we applied several in silico analyses to investigate whether genes associated with monogenic stuttering show overlap in brain-relevant biological functions involving brain development and white matter structure. Our work identified four newly implicated genes with (likely) pathogenic de novo variants and highlighted another two genes of interest for stuttering. In contrast with other neurodevelopmental disorders including childhood apraxia of speech, genes implicated in monogenic forms of stuttering show highly diverse expression patterns in the developing brain and the adult cortex, and do not show enrichment in certain brain-relevant processes.

METHODS

Participants

Participants and their parents were recruited through three distinct routes. A total of 57 children who stutter and their parents were recruited during a follow-up visit for the RESTART-randomized trial [38] in the Netherlands. Participants were included between September 2007 and June 2010 by 24 Speech Language Pathologists in 20 private practice speech clinics throughout the Netherlands. Inclusion criteria were: (1) age between 3;0 and 6;3 years; (2) stuttering was confirmed by a stuttering severity rating on an 8-point scale (at least a score of 2, i.e. 'mild') by the parent (3) and the clinician; (4) stuttering frequency was at least 3% syllables stuttered (% SS); and (5) stuttering had been present six months or longer [2]. Exclusion criteria were: (1) diagnosis of an emotional, behavioural, learning or neurological disorder; and (2) lack of proficiency in Dutch for children or parents. For more details, see De Sonneville-Koedoot and others [38]. The medical ethics committee of the Erasmus Medical Center in Rotterdam approved this study (registration number: MEC-2006-349) and all parents provided informed consent for the participation of their children and themselves.

All children who were seen for a follow-up visit for the RESTART clinical trial, and their parents, were asked to provide saliva for DNA extraction. For a total of 75 trios, DNA was isolated successfully in high enough concentration and quantity for all three family members. Eighteen trios were excluded from the WES analysis if 1) one or both parents mentioned to have stuttered in the past, or stuttered at the intake of the clinical trial or during follow-up, or 2) > 1 second-degree, > 2 third-degree family members, or > 2 second- and third-degree family members were reported to stutter by a parent. The child's stuttering phenotype (persistent, transient and ambiguous stuttering) was based on parent and teacher ratings (same 8-point scale as described above), and trained observer ratings on the Stuttering Severity Instrument (SSI fourth edition) [39]. Stuttering was categorized as persistent if SSI score ≥ 11, or if SSI score was 9 or 10 and parents or clinician reported presence of stuttering. Stuttering was classified as transient if the SSI score ≤ 8, and both parents and clinician reported absence of any observed stuttering. Conflicts between SSI scores and parents or clinician reports led to categorization of stuttering as ambiguous.

A total of 16 children who stutter and their parents were included via the MPI Erasmus Genetics of Stuttering (MEGS) Study [40] (https://www.mpi.nl/genetica-van-stotteren). People who stutter were recruited to participate in the MEGS study through national media campaigns, promotion through newspaper articles, television broadcasts, support organizations and social media, and via invitation through speech therapists. Included children and their parents participated between December 2019 and December 2022. Parents of children who stutter were asked to participate in our genetic analyses if 1) their child was 9–15 years of age, 2) their child stuttered at the time of participation, as determined from answering "yes'" to the question "Did your child stutter in the past 12 months?", 3) their child stuttered for at least four years, based on self-reported age at onset of stuttering, 4) both parents reported to have never stuttered, 5) parents reported maximally one second-degree family member who ever stuttered, and 6) the child did not have a diagnosis for ADHD, anxiety, autism, depression, behavioural issues, intellectual disability or hearing difficulty. All children included via MEGS were considered to stutter persistently. Trios were included in the trio WES analysis if DNA of all three family members was available. The medical ethics committee of the Erasmus Medical Center in Rotterdam approved

this study (registration number: MEC-2019-0491). Informed consent was obtained from all parents for themselves and their children and from minors aged 12–15 years of age.

A total of 12 adults who stutter and their parents were included through the Kassel Stuttering therapy center (KST) in Germany. This is a private practice delivering a highly standardized fluency-shaping based therapy, documenting therapy-related changes by standardized videos taken before and after therapy. In 2016, all previous 1450 participants of KST were invited to participate in genetics research by mail, of whom 203 responded positively, and of whom 180 sent an intact saliva specimen to the cooperating genetic center in Munich. In 2019, the 180 participants were asked to forward parental information leaflets to their parents, encouraging their parents to participate in our genetic study. Positive replies were received from 108 parents, of whom 33 provided an intact saliva specimen to the cooperating genetic center in Munich. This effort succeeded in assembling 33 trios, of which 12 were included in the present study. All adults who stutter included through the KST were considered to have persistent stuttering. The medical ethics committee of the University of Goettingen approved this study (registration number 19/2/15). Informed consent was obtained from all participants in the study. For all three cohorts, all methods were performed in accordance with the relevant guidelines and regulations.

Whole exome sequencing and variant calling

Whole exome sequencing was performed at the NGS Core Facility, Helmholtz Zentrum, Munich, Germany. Previously published protocols were implemented during sequencing data acquisition and processing [41]. In short, exome sequences were enriched using SureSelect60Mbv6 library preparation kit (Agilent Technologies, Santa Clara, CA, USA) and sequenced by the means of 100 bp long paired-end reads, produced by the Illumina NovaSeq6000 sequencer (Illumina, San Diego, CA, USA). In-house developed scripts were used to map the reads to the GRCh37/hg19 reference genome sequence (UCSC Genome Browser build hg19 with masked pseudo-autosomal region PAR1 on chromosome Y and updated GRCh38 mitochondrial sequence) with Burrows-Wheeler Aligner (BWA). Single nucleotide variants (SNVs) and small insertions and deletions (indels) were called with SAMTools. All samples were imported into the variant interpretation platform EVAdb of the Institute of Human Genetics, Technical University of Munich, Munich, Germany (https://github.com/IHG-MRI/EVAdb). The de novo status of prioritized variants was visually confirmed in Integrative Genomics Viewer (IGV). In addition, all variants classified as (likely) pathogenic or as variant of interest were validated with Sanger sequencing.

De novo variant identification, annotation and filtering

De novo variants were defined as variants that differed from the DNA sequence in both parental samples. The analysis included only small variants (SNVs and indels) within the coding genomic regions with a minimum of 20x coverage. Of those, only non-synonymous variants (missense, nonsense/stop-gain, stop-loss, splice, and frameshift) were kept for the downstream analysis. Genes listed in the actionable incidental findings of the American College of Medical Genetics and Genomics (ACMG SF v3.1, https://www.ncbi.nlm.nih.gov/clinvar/docs/acmg/) were removed prior to the data filtering and interpretation. Variants were filtered based on gene intolerance parameters obtained from the Genome Aggregation Database (gnomAD, v2.1.1) [42]: probable loss-of-function (pLoF) variants were included if the probability of being loss-of-function-intolerant (pLI) was > 0.9 and/or if the loss-of-function observed/expected upper bound fraction (LOEUF) was < 0.6; missense variants were included if the z-score for missense constraint was > 2.5. In addition, pLoF variants were excluded if they were not located in a major transcript, based on low exon-specific expression in developmental and adult brain gene expression data from Brainspan (http://www.brainspan.org/) and GTEx [43], or if they were located within 50 base pairs from the end of the transcript, unless they affected a known functional protein domain. Splice variants were included only if they affected the main acceptor and donor sites.

De novo variants that passed these filtering steps were further annotated using ANNOVAR [44] (version 2017-07-17) with information on minor allele frequencies from gnomAD, measures of evolutionary constraint (Genomic Evolutionary Rate Profiling (GERP + +)) and predictions of functional/pathogenic effects used to predict the impact of missense variants on the protein from Mendelian Clinically Applicable Pathogenicity (M-CAP) [45], rare exome variant ensemble learner (REVEL) [46] and PrimateAI [47]. Similar predictions from AlphaMissense [48] were added from https://alphamissense.hegelab.org [49]. M-CAP and REVEL are ensemble methods, based on scores from a combination of often-used tools such as PolyPhen, SIFT and FATHMM, that were shown to outperform these individual tools. PrimateAI classifies variants based on occurrence in other primate species, and AlphaMissense uses a combination of structural context and evolutionary conservation. Together, these four tools use a wide range of evidence to predict the effects of missense variants. Scores from M-CAP > 0.025, REVEL > 0.5, PrimateAI > 0.8 and AlphaMissense > 0.564 were considered to indicate missense variants with damaging effects on protein function, as recommended [45–48]. In addition, for missense variants, conservation estimates of the amino acids carrying a de novo variant were obtained from ConSurf [50]. These conservation estimates are based on evolutionary rates in aligned homolog sequences while considering their phylogenetic relationships. ConSurf scores range from 1 (variable) to 9 (conserved).

Expression levels of the genes carrying de novo variants were assessed in the developmental human RNA-sequencing dataset of Brainspan [51] and the adult brain gene expression data in GTEx [52]. Isoform- and exon-specific expression was considered to make sure that the exons carrying the de novo variants showed expression in the developing and/or adult brain.

Variant classification

First, we assessed whether genes previously implicated in monogenic forms of stuttering in multiplex families (*GNPTAB* [9], *AP4E1* [10], *IFNAR1* [11], *ARMC3* [12], *ZBTB20* [13] and *PPID* [14]) and hypothesis-driven case/control follow-ups (*GNPTG, NAGPA* [9, 15]) carried a de novo variant in any of the probands. Second, we investigated whether any gene, regardless of evidence from prior work, harboured recurrent de novo mutations in our cohort (i.e. multiple probands carrying a de novo mutation in the same gene). Third, we assessed overlaps of genes carrying de novo variants in our cohort with genes previously associated with known monogenic disorders. Our focus here was on neurodevelopmental disorders, given the significant comorbidity of speech and language disorders with neurodevelopmental conditions [35, 53], and given that genes previously identified as causal for speech disorders have often also been associated with monogenic neurodevelopmental disorders [31, 34]. Searches in PubMed, the Online Mendelian Inheritance in Man (OMIM) database, denovo-db (v1.6.1) [54], and VariCarta [55] (assessed in May 2024) were used to identify phenotypes previously linked to similar variants (rare, highly penetrant, and either pLoF or missense) in genes with de novo variants. We then classified the highlighted variants into 1) pathogenic, 2) likely pathogenic, 3) uncertain significance, 4) likely benign, and 5) benign variants, by combining layers of evidence of possible impact of the variant on the protein and the trait according to the commonly accepted five-tier classification system for Mendelian disorders [56]. Fourth, because our approach for interpreting variants has limited power to detect new gene-disease associations, we similarly evaluated evidence of pathogenicity for variants previously not identified as causal for a monogenic neurodevelopmental disorder. This process allowed us to highlight variants of interest in genes of unknown significance.

Gene set evaluation

To investigate whether there are convergent biological mechanisms that may explain the trait, we created a gene set associated with monogenic forms of stuttering and performed enrichment analyses in datasets that may inform about specific brain processes. This stuttering-associated gene set consisted of twelve genes: the six genes previously associated with monogenic stuttering through genetic investigations of multiplex families: *GNPTAB* [9], *AP4E1* [10], *IFNAR1* [11], *ARMC3* [12], *ZBTB20* [13] and *PPID* [14], as well as the six genes with de novo (likely) pathogenic variants and de novo variants of interest newly identified in our trio analyses. *GNPTG* and *NAGPA* were not included here, as these genes had been previously associated with stuttering via a hypothesis-driven approach (i.e. in targeted case/control follow-ups of *GNPTAB* variant findings, based on knowledge of functional pathways), so that we could avoid inappropriately biasing enrichments towards the enzymatic mannose 6-phosphate pathway. For a background gene list, we considered only genes that passed the same filtering criteria that we had used for the de novo variants, to make sure the results of the enrichment analyses did not reflect our filtering procedure. All 7313 genes intolerant to pLoF or missense variants (pLI ≥ 0.9, LOEUF ≤ 0.6 or mis_z ≥ 2.5), not reported as genes with actionable incidental findings by the ACMG, and not in the stuttering-associated gene list, were included in our background gene list. Next, we

Table 1. Overview of participants.

	RESTART	MEGS	KST cohort	Total
N (% males)	57 (77%)	16 (81%)	12 (67%)	85 (75%)
N persistent stuttering (% males)	19 (100%)	16 (81%)	12 (67%)	47 (85%)
N transient stuttering (% males)	30 (57%)	0	0	30 (57%)
N ambiguous stuttering (% males)	8 (88%)	0	0	8 (88%)
Mean age in years (range)[a]	10.6 (8–13)	10.8 (9–14)	33.9 (26–49)	13.2 (8–49)

[a]Age information is missing for three participants of the KST cohort.

assessed enrichment of biological pathways via three complementary approaches, detailed below.

First, we investigated the expression patterns of the stuttering-associated genes in the developing brain. For this, we used regional bulk gene expression data quantified by RNA sequencing from Brainspan (http://www.brainspan.org/) of a total of 224 samples from 23 human brains collected during different developmental periods (8 weeks post conception up to 12 months of age). Gene expression data, measured as fragments per kilobase per million (FPKM), were log-transformed, and then plotted with the package ggplot2 (version 3.4.4) in R (version 4.0.0). A locally estimated scatterplot smoothing (LOESS) curve visualized gene expression change during development. The Brainspan gene expression data were previously clustered into gene expression modules [32]. In short, co-expression similarity of 14,442 genes with high and variable expression was calculated using weighted correlation network analysis (WGCNA) [57]. A total of 16 co-expression modules were detected using the cutreeDynamic hybrid tree cutting function. Module eigengenes were calculated as the first principal component to summarize the expression pattern of the genes in the modules. Enrichment of stuttering-associated genes in the modules, compared to the background gene list, was investigated with two-sided Fisher exact tests. Gene ontology term enrichment analysis to describe the biological processes represented by the modules was performed with GOrilla [58].

Second, we investigated whether the stuttering gene set was enriched in specific cortical layers or in the white matter of the adult human brain. We made use of spatial gene expression data of the cortex, specifically the inferior frontal gyrus and the posterior part of the superior temporal sulcus; two cortical brain regions relevant to speech and language [59]. This dataset contains spatial gene expression of 48 brain sections (collected from three donors * two brain regions * two tissue blocks * four sections per block) that was measured with Visium spatial transcriptomic slides for a total of 140,192 spots that each represent 3–5 cells, in which clustering analysis distinguished twelve data-driven clusters of spots that were related to cortical layers or the white matter. We investigated whether the set of stuttering-associated genes showed differential expression in any of the clusters representing specific cortical layers or white matter, compared to the background gene set. For this, we downloaded the spatial transcriptomics dataset and cluster assignments from [59] and normalized the count data using 50 principal components calculated from the top 2000 most variable genes using BayesSpace (version 1.12.0) and Harmony (version 1.2.0) packages in R (version 4.3.1) as was done previously [59]. Next, we pseudobulked the spot-level data for each gene into cluster-level data by averaging the normalized counts for the respective gene across all spots in a given cluster with the summarizeAssayByGroup function of the scuttle R package (version 1.15.4). Then, we log-transformed the averaged counts, and plotted their distributions using the ggplot2 R package. The spatial brain expression pattern of the stuttering gene set was compared to that of the background gene set across the clusters representing cortical layers.

Third, we investigated whether common variants (single nucleotide polymorphisms; SNPs) in and near the stuttering-associated genes affect the strength of white matter connections in the adult human brain. For this, we made use of results of multivariate genome-wide association studies (GWAS) on white matter connectivity [60]. GWAS results for node- and edge-level measures of white matter connectivity were downloaded from the NHGRI-EBI GWAS Catalog [61] (study ID GCST90165317 and GCST90165318, downloaded June 2024). SNP-level *p*-values were converted to gene-based *p*-values using Multi-marker Analysis of GenoMic Annotation (MAGMA version 1.10) [62]: SNPs were linked to a gene if they were located within the gene body or an 15 kb upstream gene window. Next, a gene-set enrichment analysis was performed that tests whether the gene-based *p*-values of stuttering-associated genes are lower than those of genes in the background gene list, while correcting for gene size, the level of linkage disequilibrium between SNPs in the gene, and the inverse of the minor allele count of the SNPs in the gene.

RESULTS

A total of 85 parent-offspring trios was included to identify potential pathogenic variants that may explain the stuttering in the children (Table 1). For 28 out of 85 (33%) probands, other diagnoses were reported, of which dyslexia ($n = 14$) and ADD/ADHD ($n = 7$) were the most common (Supplementary Table 1). High-quality WES data were generated from 84 trios and one parent-child duo; the latter because for one father high-quality WES data could not be generated from the available DNA sample. We searched for de novo variants in the sequencing data of the 84 complete trios by excluding all variants present in any of the unaffected parents. Across the whole cohort, a total of 383 de novo non-synonymous variants were called. No de novo variants were identified in any of the genes previously associated with monogenic forms of stuttering. In addition, no gene was identified with recurrent de novo variants, regardless of prior evidence about relevance for stuttering.

Pathogenic and likely pathogenic de novo variants identified in probands who stutter

We identified three de novo probable-loss-of-function (pLoF) variants in constrained genes (Table 2). One of the pLoF variants was found in a gene previously linked to a neurodevelopmental disorder: the stop-gain in *SPTBN1* (c.A520T; p.R174X) in proband MEGS_14 with persistent stuttering (Fig. 1). Missense and pLoF variants in *SPTBN1* were recently implicated in a neurodevelopmental disorder [63, 64]; this variant was therefore classified as pathogenic. The two other pLoF variants, identified in proband RESTART_11 with persistent stuttering and proband RESTART_8 with ambiguous stuttering, are located in *FLT3* and *IREB2*, respectively; two genes not associated with a neurodevelopmental disorder. Because pLoF variants in these genes are highly uncommon, we classified these variants as variants of interest, even though additional evidence is required to verify a causal relation between the two genes and stuttering or other neurodevelopmental disorders.

We identified twelve de novo missense variants that passed our filtering criteria (Table 3). A total of six de novo missense variants were located in genes previously identified as causal for a neurodevelopmental disorder, of which three variants in *PRPF8*, *TRIO* and *ZBTB7A* were classified as likely pathogenic (Fig. 1). In addition to passing the filtering criteria (the variants are de novo and are located in a constrained gene), all three variants are (i) absent from or seen only once in GnomAD, (ii) located in a conserved region of the gene as evident by ConSurf and GERP + +, and (iii) considered damaging by all in silico tools used to predict pathogenicity of missense variants. Proband RESTART_15 with persistent stuttering carries a likely pathogenic missense variant in *PRPF8*. Proband RESTART_20 with transient stuttering carries a

Table 2. De novo frameshift and nonsense variants identified in people who stutter.

Proband	St	Gene	Variant type	Isoform	cDNA change	protein change	pLI	LOEUF	MAF	NDD gene	Classification criteria[a]	Classification[a]
RESTART_11	P	*FLT3*	Stop-gain	NM_004119	c.A520T	p.R174X	0.61	0.35	0	No	–	Variant of interest in GUS
MEGS_14	P	*SPTBN1*	Stop-gain	NM_003128	c.G5678A	p.W1893X	1.00	0.09	0	Yes	PVS1, PS2, PM2	Pathogenic
RESTART_8	A	*IREB2*	Frameshift	NM_004136	c.619delC	p.P207QfsTer9	1.00	0.22	0	No	–	Variant of interest in GUS

St stuttering, classified as, *P* persistent and, *A* ambiguous, *pLI* probability of being loss-of-function intolerant, *LOEUF* loss-of-function observed/expected upper bound fraction, *MAF* minor allele frequency, *NDD* neurodevelopmental disorder, *GUS* gene of unknown significance.
[a]Classification criteria and classification based on ACMG guidelines [56].

p.R1507Q *TRIO* missense variant slightly upstream of the RhoGEF domain, in which several causal missense variants are located [65], in the PH domain that assists and regulates the activity of the RhoGEF domain. Proband RESTART_27 with transient stuttering carries a p.H400Q *ZBTB7A* missense variant located in the first zinc finger domain, which is also the location of two previously described missense variants [66, 67]. Yet, none of the individuals previously described with a pathogenic mutation in *PRPF8*, *TRIO*, or *ZBTB7A* have been described to stutter. Similarly, lack of evidence suggests that probands RESTART_15, RESTART_20, and RESTART_27 do not show symptoms of the severe neurodevelopmental problems typically associated with pathogenic mutations in *PRPF8*, *TRIO*, and *ZBTB7A*. Despite cumulative evidence that the missense variants affect the encoded proteins, additional support such as identifying the same variants in other individuals with a neurodevelopmental disorder or functional validation of an effect on the protein would be needed to fully prove that these variants are pathogenic.

The missense variants in *CHD4*, *PLXNA1*, and *NCDN* were classified as variants of unknown significance, because computational evidence suggested a less deleterious or tolerated effect of the variants on the proteins. Even though the p.N826S variant in *CHD4* in proband RESTART_47 with persistent stuttering is located in the ATPase domain, where multiple disease-causing variants are aggregated [68], the asparagine at position 826 is not as highly conserved as the amino acids mutated in patients with CHD4-related syndrome (ConSurf score of 5 [average], compared to 7–9 [conserved]). It is therefore unlikely that this missense variant has a major effect on the functioning of the CHD4 protein. However, in silico prediction tools of effects of variants on proteins currently cannot fully capture true effects. Additional evidence would therefore be required to conclusively classify these variants as pathogenic or benign according to standard criteria.

Gene-set analyses to investigate biological pathways involved in monogenic stuttering

We investigated whether genes associated with monogenic forms of stuttering share roles in brain-relevant cell types and (developmental) processes. To do so, we tested for enrichment of all genes so far linked with monogenic stuttering in relevant datasets. In the stuttering gene set, we included the genes identified in the current trio analysis with (likely) pathogenic de novo variants (*SPTBN1*, *PRPF8*, *TRIO* and *ZBTB7A*) and with variants of interest (*FLT3* and *IREB2*), as well as the six genes associated with stuttering through previous family-based rare variant investigations (*GNPTAB*, *AP4E1*, *IFNAR1*, *ARMC3*, *ZBTB20* and *PPID*). First, we investigated whether these twelve stuttering-associated genes show similar gene expression patterns in the developing brain. Similar expression during brain development is observed for genes implicated in a number of neurodevelopmental disorders including childhood apraxia of speech [32, 36] and autism spectrum disorder [69]. In contrast, these twelve stuttering-associated genes show very dissimilar expression patterns during brain development (Fig. 2A). To investigate whether a particular expression pattern is overrepresented, we made use of gene co-expression modules of the same dataset. Gene expression modules consist of genes co-expressed in certain regions of the brain during development (Supplementary Fig. 1), and are enriched for brain-relevant developmental processes (Supplementary Table 2). Nine genes associated with monogenic forms of stuttering were assigned to a module. They were present in six of the sixteen modules, representing processes including synapse organization, transcription factor activity and chromatin organization. A maximum of two stuttering-associated genes were assigned to the same module, and none of the modules were enriched for stuttering-associated genes, confirming limited co-expression of the genes. *PRPF8* and *TRIO* were present in the module previously found enriched for genes linked to childhood

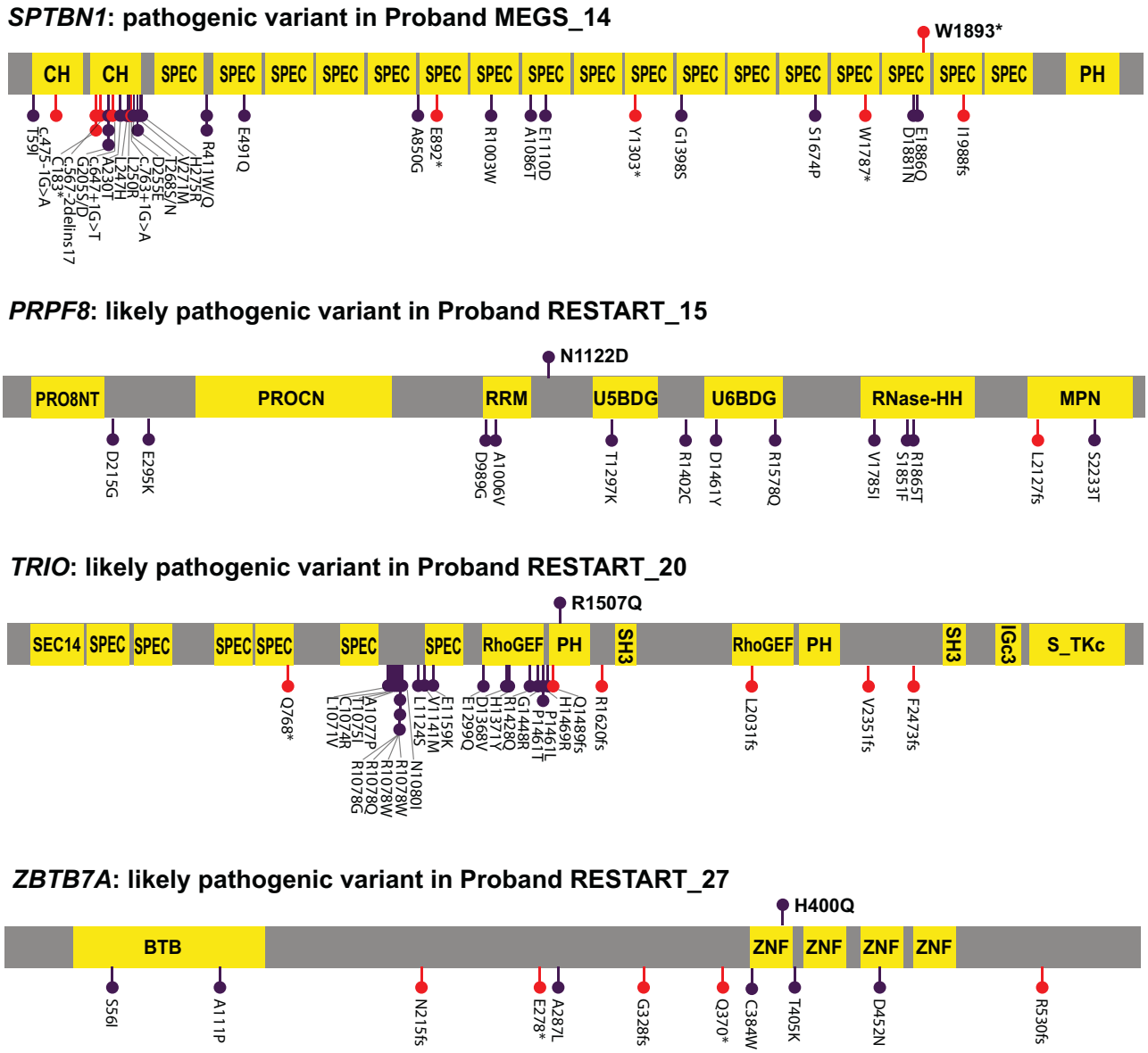

Fig. 1 Locations of identified (likely) pathogenic variants in stuttering and published pathogenic neurodevelopmental disorder variants in the same genes. Pathogenic and likely pathogenic variants identified in this study are visualized above the linear protein schematics. The variants previously published as causal for monogenic neurodevelopmental disorders related to *SPTBN1* [63, 64], *PRPF8* [71], *TRIO* [65, 73] and *ZBTB7A* [66, 67] are visualized below the schematics. Missense variants are indicated in purple and pLoF variants in red. Protein domains are represented with yellow squares: CH calponin homology domain, SPEC spectrin repeats, PH Pleckstrin homology domain, PRO8NT PrP8 N-terminal domain, PROCN central domain in pre-mRNA splicing factors of PRO8 family, RRM RNA recognition motif, U5/6BDG U5/6-snRNA binding site, RNase-HH RNase-H homology domain, SEC14 protein structural domain that binds small lipophilic molecules, RhoGEF guanine nucleotide exchange factor, SH3 Src homology 3, S_TKc Serine/Threonine protein kinases, catalytic domain, BTB Broad-Complex, Tramtrack and Bric a brac, ZNF Zinc finger.

apraxia of speech. Yet, none of the modules showed an enrichment of the stuttering-associated gene set.

Second, we investigated whether genes linked to monogenic forms of stuttering show specific spatial expression patterns in the adult human brain. For this, we made use of spatial transcriptomics data of the human cortex [59]. Spatial transcriptomics is a technique that measures gene expression for many thousands of transcripts in a tissue section, across several thousands of locations (spots), in this case each representing three to five cells. Data-driven clustering of this dataset identified seven clusters that recapture the laminar structure of the cerebral cortex, and five clusters that were located in the white matter. The twelve genes associated with monogenic stuttering show gene expression levels that are very similar to the background gene set in each of the clusters, and do not highlight a certain cortical layer or white matter cluster (Fig. 2B).

Third, we investigated whether the genes so far associated with monogenic forms of stuttering play a role in human brain white matter connectivity. Previously, several lines of evidence indicated a role for reduced white matter connectivity in stuttering. Because our gene expression analyses failed to identify overlapping expression patterns or biological functions between the stuttering-associated genes, we also applied a more direct approach to explore the relation between the stuttering-associated genes and white matter connectivity. For this, we made use of results of a recent multivariate GWAS for white matter connectivity of the human brain, that used brain imaging data of 30,810 individuals of the UK Biobank [60]. The authors of

Table 3. De novo missense variants identified in people who stutter.

Proband	St	Gene	Isoform	cDNA change	protein change	mis_Z	MAF	M-CAP	REVEL	Primate AI	Alpha Missense	GERP++	ConSurf	NDD gene	Classification criteria[a]	Classification[a]
RESTART_15	P	*PRPF8*	NM_006445	c.A3364G	p.N1122D	8.3	0	D	D	D	D	5.31	9	Yes	PS2, PM2, PP2, PP3	Likely pathogenic
RESTART_47	P	*CHD4*	NM_001273	c.A2477G	p.N826S	6.3	8.0×10^{-6}	D	T	D	T	4.9	5	Yes	PS2, PP2, PP3, BP4	VUS
RESTART_56	P	*UNC13A*	NM_001080421	c.T224C	p.V75A	5.6	0	D	D	T	A	4.98	7	No	–	–
KST_2	P	*EIF2AK4*	NM_001013703	c.G3040A	p.V1014M	2.7	0	T	T	T	T	5.82	8	No	–	–
KST_4	P	*GIT1*	NM_001085454	c.G2135T	p.R712L	3.1	0	D	T	T	A	4.65	9	No	–	–
KST_7	P	*ADGRB1 / BAI1*	NM_001702	c.G1441T	p.A481S	4.5	4.2×10^{-6}	T	T	T	T	3.55	4	No	–	–
KST_10	P	*TUT4 / ZCCHC11*	NM_015269	c.A1442T	p.D481V	2.8	0	D	T	T	A	5.51	7	No	–	–
RESTART_18	T	*PLXNA1*	NM_032242	c.G3182C	p.S1061T	3.4	0	D	D	T	A	3.84	7	Yes	PS2, PM2, PP2, PP3, BP4	VUS
RESTART_20	T	*TRIO*	NM_007118	c.G4520A	p.R1507Q	5.3	3.2×10^{-5}	D	D	D	D	5.39	9	Yes	PS2, PP2, PP3	Likely pathogenic
RESTART_22	T	*NCDN*	NM_001014839	c.C2095T	p.R699W	3.8	0	D	T	T	T	5.13	1	Yes	PS2, PM2, PP2, BP4	VUS
RESTART_27	T	*ZBTB7A*	NM_015898	c.C1200G	p.H400Q	4.0	0	D	D	D	D	4	9	Yes	PS2, PM2, PP2, PP3	Likely pathogenic

St stuttering, classified as *P* persistent and *T* transient, *MAF* minor allele frequency, Scores of M-CAP, REVEL, PrimateAI and AlphaMissense were interpreted as, *D* deleterious, *T* tolerated, and *A* ambiguous, *NDD* neurodevelopmental disorder, *GUS* gene of unknown significance, *VUS* variant of unknown significance.
[a]Classification criteria and classification based on ACMG guidelines [56].

that study used fiber tractography of diffusion tensor imaging data, to derive two measures of connectivity, where the nodes captured the sum of the connectivity of each of the 90 brain regions investigated, and the edges captured the connectivity between 947 pairs of brain regions. We converted SNP-based *p*-values from the two multivariate genome-wide analyses into gene-based *p*-values. Three of the twelve stuttering-associated genes showed low gene-based *p*-values for one or both measures of white matter connectivity (Table 4). We next tested whether the set of stuttering-associated genes was enriched for low *p*-values. For both the node-level connectivity (beta = 0.17, se = 0.33, $p = 0.31$) and edge-level connectivity (beta = −0.02, se = 0.35, $p = 0.53$) GWASs, there was no enrichment of low gene-based *p*-values in our stuttering gene set. So, while variation in some stuttering-associated genes may be associated with variability in white matter connectivity, alterations of the latter may not be a common mechanism that is shared across genes implicated in monogenic forms of stuttering.

DISCUSSION

Here, we used whole exome sequencing of 85 children who stutter to identify genes potentially involved in monogenic forms of stuttering. By including parents who had never stuttered, our trio study design enabled us to identify and focus our analyses on de novo variants in the children who stutter. To our knowledge, this is the first time de novo variants have been implicated in stuttering, as previous studies in this area have focused on families in which multiple relatives stutter [9–14]. We identified a de novo stop-gain variant in *SPTBN1* that could be classified as pathogenic, and missense variants in *PRPF8*, *TRIO* and *ZBTB7A* that could be classified as likely pathogenic. In addition, likely damaging de novo variants in genes not previously implicated in neurodevelopmental disorders highlighted two genes of interest for stuttering: *FLT3* and *IREB2*. Our yield of four (likely) pathogenic variants in 84 trios indicates that de novo variants are not a major cause of stuttering, and is notably lower than yields previously found for childhood apraxia of speech (36 in 122 probands across three studies) [31] and speech delay (three in 23 probands) [34]. Still, we show that rare de novo variants might account for a subset of cases and so should not be neglected as a possible cause for stuttering.

Our analyses identified six genes with a possible causal link to stuttering. *SPTBN1* codes for βII-spectrin, a long polypeptide that forms networks associated with the plasma membrane and is essential for plasma membrane domains. Missense and pLoF variants in *SPTBN1* cause a neurodevelopmental disorder characterized by intellectual disability, language and motor delays, autistic features and seizures [63, 64]. In mouse models, homozygous and heterozygous loss of *Sptbn1* in neural progenitors disrupts the development of long-range axons, and leads to lower body size and increased head circumference [63, 70]. *PRPF8* encodes a scaffolding component of a spliceosome complex that splices pre-mRNA into mRNA by removing introns. Recently, missense and LoF variants located throughout the protein have been identified as the cause of a neurodevelopmental condition involving developmental delay and autism [71]. In addition, heterozygous missense variants in the C-terminal MPN-domain cause autosomal dominant retinitis pigmentosa [72]. *TRIO* encodes a Rho guanine nucleotide exchange factor (RhoGEF) disruptions of which have previously been identified as causal for neurodevelopmental disorders, with domain-specific symptoms [65, 73]. Gain-of-function missense variants in and near the spectrin domains are associated with severe developmental delay, speech and language delay, and macrocephaly, while loss-of-function variants and missense variants in the RhoGEF domain show milder developmental delay, speech and language delay, and microcephaly. *ZBTB7A* encodes a transcription factor, variants

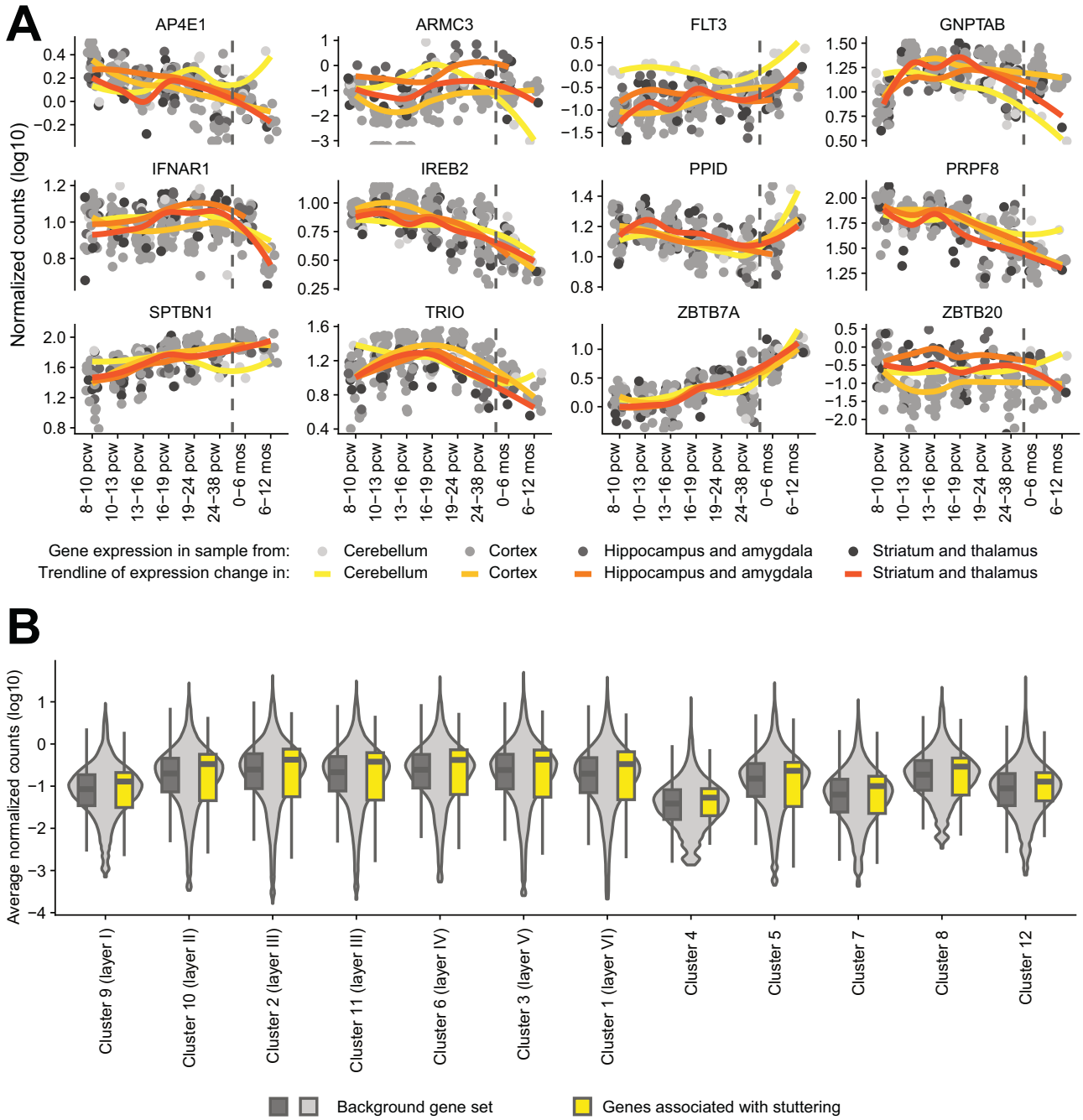

Fig. 2 Neural gene expression patterns of genes associated with monogenic forms of stuttering. A. Developmental brain expression pattern of the twelve stuttering-associated genes across eight developmental periods spanning from eight post conception weeks (pcw) to ten months (mos) of age. Grey circles depict expression levels in individual brain samples collected from the cerebellum, cortex, hippocampus, amygdala, striatum and thalamus. The trendlines in yellow-orange (estimated with locally estimated scatterplot smoothing) visualize the overall pattern of gene expression change over time in the different regions of the brain. The vertical dashed lines represent time of birth. **B.** Gene expression levels of these stuttering-associated genes and the background gene set in spatial gene expression data of the adult human cortex. Spatial gene expression data of 48 human cortex tissue sections were clustered into 12 data-driven clusters, of which seven represent cortical layers and five were located in the white matter [59]. Violin plots and grey box plots show the distribution of gene expression levels of the background gene set in the 12 clusters. Yellow box plots show the gene expression levels of the stuttering-associated genes. Box plots show median and first and third quartiles, with whiskers extending to 1.5 times the interquartile range.

of which cause a neurodevelopmental disorder characterized by intellectual disability, macrocephaly, and overgrowth of adenoid tissue [66]. Genes of interest *FLT3* and *IREB2* have not been associated with a neurodevelopmental disorder. *FLT3* encodes a receptor tyrosine kinase, of which a homozygous knock-out in a mouse model causes immune system deficiencies [74]. *IREB2* encodes an iron-responsive element-binding protein that regulates the transcription of proteins involved in iron metabolism and oxygen sensing. Homozygous knock-out of *Ireb2* in a mouse model leads to iron metabolism dysregulation and a progressive movement disorder [75]. Replication in a person who stutters is required to confirm that stuttering is a feature of the monogenic

Table 4. Gene-based *p*-values of stuttering-associated genes in human brain white matter connectivity GWAS in 30,810 individuals.

Stuttering-associated gene	Number of SNPs assigned to gene	Gene-based *p*-value for node-level connectivity	Gene-based *p*-value for edge-level connectivity
GNPTAB	252	0.012	0.23
AP4E1	307	0.011	0.014
IFNAR1	196	0.69	0.33
ARMC3	338	0.24	0.12
ZBTB20	1641	0.049	**0.0014**
PPID	44	0.027	0.42
SPTBN1	761	$\mathbf{5.49 \times 10^{-4}}$	$\mathbf{5.84 \times 10^{-4}}$
PRPF8	197	0.54	0.60
TRIO	1070	0.042	0.39
ZBTB7A	101	$\mathbf{4.03 \times 10^{-5}}$	$\mathbf{9.38 \times 10^{-5}}$
FLT3	520	0.62	0.52
IREB2	188	0.53	0.60

Gene-based *p*-values were calculated with MAGMA from two multivariate GWAS analyses of 90 node-level and of 851 edge-level connectivity measures of the human brain [60]. *P*-values are marked bold if significant after Bonferroni-correction for 24 tests (*p*-value threshold is 0.0021).

neurodevelopmental disorders associated with mutations in *SPTBN1*, *PRPF8*, *TRIO*, and *ZBTB7A*, and to causally link *FLT3* and *IREB2* to neurodevelopmental disorders.

Our study represents the first rare-variant analysis to include not only persistent cases but also individuals with transient developmental stuttering. To our knowledge, all previous rare-variant investigations of stuttering focused on individuals with persistent developmental stuttering [9–14]. Surprisingly, our yield of (likely) pathogenic variants did not differ between the groups with persistent (two in 47 probands) and transient (two in 30 probands) stuttering. Moreover, beyond these (likely) pathogenic variants, our study highlighted two further genes of interest: in a proband with persistent stuttering, and a proband whose stuttering was classified as ambiguous. Little is known about differences and similarities in the genetic foundations of transient and persistent stuttering. A twin study in 12,892 children that distinguished transient and persistent stuttering showed very similar heritability estimates ($h^2 = 67$ and 60%, respectively), and also identified multiple occurrences of transient and persistent stuttering within a twin pair [7]. In addition, an investigation of inheritance patterns in the extended families of 66 children who stutter found that transient and persistent stuttering have a shared genetic basis, and that persistent stuttering may at least in part be caused by additional (genetic) factors [76].

Another innovative aspect of our findings is the novel evidence of a direct genetic link between stuttering and other neurodevelopmental disorders. Even though the genes *AP4E1* and *ZBTB20* have previously been linked to stuttering [10, 13] and separately to neurodevelopmental disorders, the mode of inheritance (i.e. dominance/recessivity) do not overlap. Heterozygous missense variants in *ZBTB20* cause Primrose syndrome [30], while the gene was associated with stuttering through a recessive mode of inheritance [13]. Similarly, biallelic loss-of-function variants in *AP4E1* cause spastic paraplegia [28], while a haplotype of two missense variants as well as heterozygous variants have been associated with stuttering [10]. According to prior literature on the genes implicated by our de novo analyses, none of the patients with neurodevelopmental disorders caused by mutations in *SPTBN1* [63, 64], *PRPF8* [71], *TRIO* [65, 73], and *ZBTB7A* [66] have been described to stutter. Stuttering may be an uncommon feature of these neurodevelopmental disorders, but an alternative explanation is that stuttering diagnoses may have escaped detection, because stuttering is often not registered well in electronic medical records [53]. The latter is supported by the increased prevalence of developmental conditions including intellectual disability, learning disability, seizures and ADHD in children who stutter [77]. Interestingly, a likely pathogenic missense variant in *SPTBN1* has been described in a proband with speech delay [34], and a missense variant of unknown significance in *TRIO* in a proband with childhood apraxia of speech [32]. In addition, pathogenic variants in *SPTBN1* have been associated with aphasia [53]. Lastly, more general terms for speech difficulties such as delayed speech, expressive and/or receptive language difficulties, and absence of speech have been registered for many patients with neurodevelopmental disorders related to *SPTBN1* [63, 64], *TRIO* [65, 73], and *ZBTB7A* [66], although not for *PRPF8* [71]. Detailed speech and language analysis in people with mutations in neurodevelopmental disorder genes including *KAT6A* [78], *SETBP1* [79], and *BRPF1* [80], that were performed after identification of a (likely) pathogenic variant in these genes in an individual with childhood apraxia of speech, revealed widespread speech and language difficulties. Such phenotypic assessments highlight that speech and language difficulties are usually not systematically investigated. Identification of rare pathogenic variants that cause stuttering and other speech disorders may thus point towards neurodevelopmental disorders in which speech difficulties are a central feature. It is now important to further prove a role for *SPTBN1*, *PRPF8*, *TRIO* and *ZBTB7A*, *FLT3* and *IREB2* in stuttering by identifying recurrent mutations in other people who stutter, or through extensive assessments of the speech phenotypes in people with a mutation in any of these genes.

Several lines of evidence from genetic and brain imaging studies suggest the involvement of altered white matter in stuttering. First, different transgenic mouse models carrying putative pathogenic variants of *GNPTAB* or *PPID* both showed white-matter features that distinguished the knock-in animals from wild-type animals (although the nature of these features differed) [14, 25]. Second, *SPTBN1* (newly implicated in the present study) encodes a cytoskeletal protein important for axonal formation and function [81]. Third, several brain imaging studies in adults and children who stutter have reported decreased white matter integrity, most commonly along parts of the left arcuate fasciculus and/or superior longitudinal fasciculus, white-matter tracts which connect parts of the frontal cortex with cortical areas in the parietal and temporal lobes [82]. We therefore investigated whether the genes thus far linked to monogenic forms of stuttering show enrichment of common variants involved in white-matter connectivity. A few of the stuttering-associated genes: *SPTBN1*, *ZBTB20*, and *ZBTB7A*, showed significant gene-based association with measures of white-matter connectivity. Yet, the full set of the twelve stuttering-associated genes that we investigated here did not show an enrichment of genetic associations with white matter connectivity as derived from GWAS data.

Genes causally implicated in neurodevelopmental disorders with similar features often show overlapping gene expression patterns in the brain and overlaps in functional pathways. For example, the majority of genes thus far implicated in childhood apraxia of speech regulate gene expression through transcription factor activity or chromatin remodeling, and are highly expressed at early stages during brain development [32, 33, 36]. However, the genes thus far linked to monogenic forms of stuttering, through previous family-based investigations and the current trio analysis, do not converge onto one or a few shared processes. The developmental brain expression data and analysis method that previously showed overlaps among genes causal for childhood apraxia of speech [32] and among genes implicated in autism spectrum disorder [69], here found no significant overlaps in expression patterns among genes associated with monogenic

forms of stuttering. A similar lack of convergence was seen when using spatial gene expression data of the human adult brain. Our results may indicate that stuttering can result from differences in a broad range of biological processes and brain regions/cell-types. Alternatively, the current analyses may overlook the biological processes, brain regions, or developmental periods involved, either because they were undersampled or because the bulk and spatial gene expression data did not have the resolution to detect a signal. Lastly, the lack of biological convergence among the genes might indicate that some of the genes in the set were incorrectly associated with stuttering, in our study and/or previous reports. Other datasets or additional implicated genes may be required to identify convergent processes, if these exist.

Our study has several limitations. First, the exploration of de novo variants may overlook inherited causal variants with reduced penetrance or variable expressivity, regulatory variants not located in the exons, and repeat expansions. Even though we selected probands with limited stuttering reported in family members, thereby optimizing our study design for the identification of de novo variants with high effect sizes, variants with low penetrance may explain cases who did report a few family members who stutter, or who failed to report transient stuttering of family members. Variable expressivity, meaning that features of disorders may differ between individuals in severity or type, may have caused us to miss variants inherited from a parent with other features of the neurodevelopmental disorder that were not recorded and used to exclude trios. Second, our filtering to include and exclude variants as likely pathogenic strongly depends on prediction tools that inform about how damaging a variant may be to a protein. Even though we combined evidence from four prediction tools that are based on different types of information and thus may be seen as supplementary layers of evidence, over- or underestimation of the effects of variants may have led us to wrongly include or exclude variants. Recurrent findings or functional testing (in vitro or in vivo) may provide final evidence for a pathogenic or benign role of variants classified as likely pathogenic or VUS. Third, our yield cannot inform about how prevalent monogenic forms of stuttering are, because we only investigated de novo variants and thus do not have information about rare inherited causal variants. Fourth, we currently cannot verify that the (likely) pathogenic variants cause the stuttering in the probands. Even after our careful and strict variant filtering and classification process, we can only judge whether the variants may be (likely) pathogenic for the neurodevelopmental traits previously associated with the gene. Identification of additional pathogenic variants in the same genes in people who stutter, or extensive phenotypic analysis of the speech of people with a mutation in any of these genes, is required to prove that these (likely) pathogenic variants cause stuttering.

In conclusion, by analyzing genome sequencing data of 85 individuals with persistent, transient or ambiguous stuttering and parents who do not stutter, we identified rare de novo variants of which four were classified as (likely) pathogenic and two highlighted genes of interest. We linked stuttering to genes causal for other monogenic neurodevelopmental disorders with and without speech problems. Extensive analysis of two brain gene expression datasets and a neuroimaging GWAS dataset indicates that monogenic forms of stuttering are likely to involve heterogenous biological pathways, rather than a shared mechanistic basis.

DATA AVAILABILITY

Sequencing data of the MEGS cohort, and of the RESTART participants who consented to have their data available for future research, are deposited in The MPI for Psycholinguistics Archive (https://archive.mpi.nl/mpi/), a public data archive hosted by the Max Planck Institute for Psycholinguistics. Data are stored at the archive with the persistent identifier https://hdl.handle.net/1839/a18caaac-bb8f-4bc1-a569-7cf0b24a8f15. Access to the sequencing data from all three cohorts can be granted upon request from the MPI archive and the exome/genome archive of the Institute of Human Genetics of the Technical University of Munich.

CODE AVAILABILITY

Code used to perform the annotation of variants and gene set analyses is available in GitLab (https://gitlab.gwdg.de/else.eising/trio-stuttering-wes.git).

REFERENCES

1. Yairi E, Ambrose N. Epidemiology of stuttering: 21st century advances. J Fluency Disord. 2013;38:66–87.
2. Yairi E, Ambrose NG. Early childhood stuttering I: persistency and recovery rates. J Speech Lang Hear Res. 1999;42:1097–112.
3. Craig A, Tran Y. The epidemiology of stuttering: the need for reliable estimates of prevalence and anxiety levels over the lifespan. Am J Speech Lang Pathol. 2005;7:41–6.
4. Tichenor SE, Yaruss JS. Stuttering as defined by adults who stutter. J Speech Lang Hear Res. 2019;62:4356–69.
5. Connery A, McCurtin A, Robinson K. The lived experience of stuttering: a synthesis of qualitative studies with implications for rehabilitation. Disabil Rehabil. 2020;42:2232–42.
6. Frigerio-Domingues C, Drayna D. Genetic contributions to stuttering: the current evidence. Mol Genet Genomic Med. 2017;5:95–102.
7. Dworzynski K, Remington A, Rijsdijk F, Howell P, Plomin R. Genetic etiology in cases of recovered and persistent stuttering in an unselected, longitudinal sample of young twins. Am J Speech Lang Pathol. 2007;16:169–78.
8. van Beijsterveldt CE, Felsenfeld S, Boomsma DI. Bivariate genetic analyses of stuttering and nonfluency in a large sample of 5-year-old twins. J Speech Lang Hear Res. 2010;53:609–19.
9. Kang C, Riazuddin S, Mundorff J, Krasnewich D, Friedman P, Mullikin JC, et al. Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. N Engl J Med. 2010;362:677–85.
10. Raza MH, Mattera R, Morell R, Sainz E, Rahn R, Gutierrez J, et al. Association between rare variants in AP4E1, a component of intracellular trafficking, and persistent stuttering. Am J Hum Genet. 2015;97:715–25.
11. Sun Y, Gao Y, Zhou Y, Zhou Y, Zhang Y, Wang D, et al. IFNAR1 gene mutation may contribute to developmental stuttering in the Chinese population. Hereditas. 2021;158:46.
12. Rehman AU, Hamid M, Khan SA, Eisa M, Ullah W, Rehman ZU, et al. The expansion of the spectrum in stuttering disorders to a novel ARMC gene family (ARMC3). Genes. 2022;13:2299.
13. Frigerio Domingues CE, Raza MH, Han T-U, Barnes T, Shaw P, Sudre G, et al. Mutations in ZBTB20 in individuals with persistent stuttering. medRxiv. 2022. 2022.2011.2003.22281471:1–23.
14. Morgan AT, Scerri TS, Vogel AP, Reid CA, Quach M, Jackson VE, et al. Stuttering associated with a pathogenic variant in the chaperone protein cyclophilin 40. Brain. 2023;146:5086–97.
15. Raza MH, Domingues CE, Webster R, Sainz E, Paris E, Rahn R, et al. Mucolipidosis types II and III and non-syndromic stuttering are associated with different variants in the same genes. Eur J Hum Genet. 2016;24:529–34.
16. Polikowsky HG, Shaw DM, Petty LE, Chen HH, Pruett DG, Linklater JP, et al. Population-based genetic effects for developmental stuttering. HGG Adv. 2022;3:100073.
17. Shaw DM, Polikowsky HP, Pruett DG, Chen HH, Petty LE, Viljoen KZ, et al. Phenome risk classification enables phenotypic imputation and gene discovery in developmental stuttering. Am J Hum Genet. 2021;108:2271–83.
18. Polikowsky HG, Scartozzi AC, Shaw DM, Pruett DG, Chen HH, Petty LE, et al. Large-scale genome-wide analyses of stuttering. Nat Genet 2025;57:1835–47.
19. Coutinho MF, Prata MJ, Alves S. Mannose-6-phosphate pathway: a review on its role in lysosomal function and dysfunction. Mol Genet Metab. 2012;105:542–50.
20. Shin J, Nile A, Oh JW. Role of adaptin protein complexes in intracellular trafficking and their impact on diseases. Bioengineered. 2021;12:8259–78.
21. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. Nat Rev Immunol. 2014;14:36–49.
22. Coates JC. Armadillo repeat proteins: beyond the animal kingdom. Trends Cell Biol. 2003;13:463–71.
23. Sutherland AP, Zhang H, Zhang Y, Michaud M, Xie Z, Patti ME, et al. Zinc finger protein Zbtb20 is essential for postnatal survival and glucose homeostasis. Mol Cell Biol. 2009;29:2804–15.
24. Freeman BC, Toft DO, Morimoto RI. Molecular chaperone machines: chaperone activities of the cyclophilin Cyp-40 and the steroid aporeceptor-associated protein p23. Science. 1996;274:1718–20.

25. Han TU, Root J, Reyes LD, Huchinson EB, Hoffmann JD, Lee WS, et al. Human GNPTAB stuttering mutations engineered into mice cause vocalization deficits and astrocyte pathology in the corpus callosum. Proc Natl Acad Sci USA. 2019;116:17515–24.
26. Cathey SS, Kudo M, Tiede S, Raas-Rothschild A, Braulke T, Beck M, et al. Molecular order in mucolipidosis II and III nomenclature. Am J Med Genet A. 2008;146A:512–3.
27. Cathey SS, Leroy JG, Wood T, Eaves K, Simensen RJ, Kudo M, et al. Phenotype and genotype in mucolipidoses II and III alpha/beta: a study of 61 probands. J Med Genet. 2010;47:38–48.
28. Ebrahimi-Fakhari D, Teinert J, Behne R, Wimmer M, D'Amore A, Eberhardt K, et al. Defining the clinical, molecular and imaging spectrum of adaptor protein complex 4-associated hereditary spastic paraplegia. Brain. 2020;143:2929–44.
29. Hernandez N, Bucciol G, Moens L, Le Pen J, Shahrooei M, Goudouris E, et al. Inherited IFNAR1 deficiency in otherwise healthy patients with adverse reaction to measles and yellow fever live vaccines. J Exp Med. 2019;216:2057–70.
30. Melis D, Carvalho D, Barbaro-Dieber T, Espay AJ, Gambello MJ, Gener B, et al. Primrose syndrome: characterization of the phenotype in 42 patients. Clin Genet. 2020;97:890–901.
31. Morgan AT, Amor DJ, St John MD, Scheffer IE, Hildebrand MS. Genetic architecture of childhood speech disorder: a review. Mol Psychiatry. 2024;29:1281–92.
32. 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.
33. Kaspi A, Hildebrand MS, Jackson VE, Braden R, van Reyk O, Howell T, et al. Genetic aetiologies for childhood speech disorder: novel pathways co-expressed during brain development. Mol Psychiatry. 2023;28:1647–63.
34. Eising E, Vino A, Mabie HL, Campbell TF, Shriberg LD, Fisher SE. Genome sequencing of idiopathic speech delay. Hum Mutat. 2024;2024:9692863.
35. Deciphering Developmental Disorders S. Prevalence and architecture of de novo mutations in developmental disorders. Nature. 2017;542:433–8.
36. 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–e2167.
37. Formicola D, Podda I, Dirupo E, Andreucci E, Giglio S, Cipriani P, et al. Expanding the molecular landscape of childhood apraxia of speech: evidence from a single-center experience. Front Neurosci. 2024;18:1396240.
38. de Sonneville-Koedoot C, Stolk E, Rietveld T, Franken MC. Direct versus indirect treatment for preschool children who stutter: the RESTART randomized trial. PLoS ONE. 2015;10:e0133758.
39. Riley G. The stuttering severity instrument for adults and children (SSI-4). 4th Edn. Austin. TX: PRO-ED; 2009.
40. Engelen MM, Franken MJP, Stipdonk LW, Horton SE, Jackson VE, Reilly S, et al. The association between stuttering burden and psychosocial aspects of life in adults. J Speech Lang Hear Res. 2024;67:1385–99.
41. Zech M, Jech R, Boesch S, Skorvanek M, Weber S, Wagner M, et al. Monogenic variants in dystonia: an exome-wide sequencing study. Lancet Neurol. 2020;19:908–18.
42. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–43.
43. Consortium GT. The genotype-tissue expression (GTEx) project. Nat Genet. 2013;45:580–5.
44. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164.
45. Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet. 2016;48:1581–6.
46. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. Am J Hum Genet. 2016;99:877–85.
47. Sundaram L, Gao H, Padigepati SR, McRae JF, Li Y, Kosmicki JA, et al. Predicting the clinical impact of human mutation with deep neural networks. Nat Genet. 2018;50:1161–70.
48. Cheng J, Novati G, Pan J, Bycroft C, Zemgulyte A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023;381:eadg7492.
49. Tordai H, Torres O, Csepi M, Padanyi R, Lukacs GL, Hegedus T. Analysis of AlphaMissense data in different protein groups and structural context. Sci Data. 2024;11:495.
50. Yariv B, Yariv E, Kessel A, Masrati G, Chorin AB, Martz E, et al. Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. Protein Sci. 2023;32:e4582.
51. Li M, Santpere G, Imamura Kawasawa Y, Evgrafov OV, Gulden FO, Pochareddy S, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science. 2018;362:eaat7615.
52. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015;348:660–5.
53. Magielski JH, Ruggiero SM, Xian J, Parthasarathy S, Galer PD, Ganesan S, et al. The clinical and genetic spectrum of paediatric speech and language disorders. Brain. 2025;148:663–74.
54. denovo-db, Seattle, WA (denovo-db.gs.washington.edu) [May 2024]
55. Belmadani M, Jacobson M, Holmes N, Phan M, Nguyen T, Pavlidis P, et al. VariCarta: a comprehensive database of harmonized genomic variants found in autism spectrum disorder sequencing studies. Autism Res. 2019;12:1728–36.
56. 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.
57. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
58. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10:48.
59. Wong MMK, Sha Z, Lutje L, Kong XZ, van Heukelum S, van de Berg WDJ, et al. The neocortical infrastructure for language involves region-specific patterns of laminar gene expression. Proc Natl Acad Sci USA. 2024;121:e2401687121.
60. Sha Z, Schijven D, Fisher SE, Francks C. Genetic architecture of the white matter connectome of the human brain. Sci Adv. 2023;9:eadd2870.
61. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023;51:D977–D985.
62. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol. 2015;11:e1004219.
63. Cousin MA, Creighton BA, Breau KA, Spillmann RC, Torti E, Dontu S, et al. Pathogenic SPTBN1 variants cause an autosomal dominant neurodevelopmental syndrome. Nat Genet. 2021;53:1006–21.
64. Rosenfeld JA, Xiao R, Bekheirnia MR, Kanani F, Parker MJ, Koenig MK, et al. Heterozygous variants in SPTBN1 cause intellectual disability and autism. Am J Med Genet A. 2021;185:2037–45.
65. Barbosa S, Greville-Heygate S, Bonnet M, Godwin A, Fagotto-Kaufmann C, Kajava AV, et al. Opposite modulation of RAC1 by mutations in TRIO is associated with distinct, domain-specific neurodevelopmental disorders. Am J Hum Genet. 2020;106:338–55.
66. von der Lippe C, Tveten K, Prescott TE, Holla OL, Busk OL, Burke KB, et al. Heterozygous variants in ZBTB7A cause a neurodevelopmental disorder associated with symptomatic overgrowth of pharyngeal lymphoid tissue, macrocephaly, and elevated fetal hemoglobin. Am J Med Genet A. 2022;188:272–82.
67. Ohishi A, Masunaga Y, Iijima S, Yamoto K, Kato F, Fukami M, et al. De novo ZBTB7A variant in a patient with macrocephaly, intellectual disability, and sleep apnea: implications for the phenotypic development in 19p13.3 microdeletions. J Hum Genet. 2020;65:181–6.
68. Weiss K, Lazar HP, Kurolap A, Martinez AF, Paperna T, Cohen L, et al. The CHD4-related syndrome: a comprehensive investigation of the clinical spectrum, genotype-phenotype correlations, and molecular basis. Genet Med. 2020;22:389–97.
69. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013;155:1008–21.
70. Lorenzo DN, Badea A, Zhou R, Mohler PJ, Zhuang X, Bennett V. betaII-spectrin promotes mouse brain connectivity through stabilizing axonal plasma membranes and enabling axonal organelle transport. Proc Natl Acad Sci USA. 2019;116:15686–95.
71. O'Grady L, Schrier Vergano SA, Hoffman TL, Sarco D, Cherny S, Bryant E, et al. Heterozygous variants in PRPF8 are associated with neurodevelopmental disorders. Am J Med Genet A. 2022;188:2750–9.
72. Pena V, Liu S, Bujnicki JM, Luhrmann R, Wahl MC. Structure of a multipartite protein-protein interaction domain in splicing factor prp8 and its link to retinitis pigmentosa. Mol Cell. 2007;25:615–24.
73. Bonnet M, Roche F, Fagotto-Kaufmann C, Gazdagh G, Truong I, Comunale F, et al. Pathogenic TRIO variants associated with neurodevelopmental disorders perturb the molecular regulation of TRIO and axon pathfinding in vivo. Mol Psychiatry. 2023;28:1527–44.
74. Mackarehtschian K, Hardin JD, Moore KA, Boast S, Goff SP, Lemischka IR. Targeted disruption of the flk2/flt3 gene leads to deficiencies in primitive hematopoietic progenitors. Immunity. 1995;3:147–61.

75. LaVaute T, Smith S, Cooperman S, Iwai K, Land W, Meyron-Holtz E, et al. Targeted deletion of the gene encoding iron regulatory protein-2 causes misregulation of iron metabolism and neurodegenerative disease in mice. Nat Genet. 2001;27:209–14.
76. Ambrose NG, Cox NJ, Yairi E. The genetic basis of persistence and recovery in stuttering. J Speech Lang Hear Res. 1997;40:567–80.
77. Briley PM, Ellis C Jr. The coexistence of disabling conditions in children who stutter: evidence from the national health interview survey. J Speech Lang Hear Res. 2018;61:2895–905.
78. St John M, Amor DJ, Morgan AT. Speech and language development and genotype-phenotype correlation in 49 individuals with KAT6A syndrome. Am J Med Genet A. 2022;188:3389–3400.
79. Morgan A, Braden R, Wong MMK, Colin E, Amor D, Liegeois F, et al. Speech and language deficits are central to SETBP1 haploinsufficiency disorder. Eur J Hum Genet. 2021;29:1216–25.
80. Morison LD, Van Reyk O, Baker E, Ruaud L, Couque N, Verloes A, et al. Beyond 'speech delay': expanding the phenotype of BRPF1-related disorder. Eur J Med Genet. 2024;68:104923.
81. Lorenzo DN, Edwards RJ, Slavutsky AL. Spectrins: molecular organizers and targets of neurological disorders. Nat Rev Neurosci. 2023;24:195–212.
82. Chang SE, Garnett EO, Etchell A, Chow HM. Functional and neuroanatomical bases of developmental stuttering: current insights. Neuroscientist. 2019;25:566–82.

ACKNOWLEDGEMENTS

EE, AV and SEF are financially supported by the Max Planck Society. EE is also supported by a Veni grant of the Dutch Research Council (NWO; VI.Veni.202.072). RESTART study: We gratefully acknowledge all participants, their parents and speech therapists for their participation in the RESTART trial. In addition, we acknowledge the contributions of Caroline de Sonneville-Koedoot, Nina Koemans and Sophie Jacobs to the collection of the follow-up data; of Sarah Graham to the collection of DNA; and of Esther Bunschoten, Chris Struiksma, Conor Jansen and Nienke van den Broek to the rating of stuttering severity from the speech recordings. MEGS: We are grateful to the children and their parents who participated in the MEGS trio study. We also thank Loes van den Heuvel, Cansel Sert, and Anke Verhulst for their assistance regarding data collection. KST cohort: The authors are thankful to the KST and their staff for long term cooperation, to the recently dissolved Department of Clinical Neurophysiology, University Medical Center Göttingen, for administrative support including most postal service charges. They acknowledge receiving a seed grant from the Leibniz Wissenschaftscampus primate cognition to MS and to Prof. Dr. Nivedita Mani, Georg-Elias-Müller Institute for Psychology, Göttingen Sommer/Mani DM 22-606) for assembling a video-archive allowing detailed characterization of the phenotype of 187 participants for which a saliva sample is available.

AUTHOR CONTRIBUTIONS

EE, JW, KO, MS and SEF designed the study. AV, EE, ID, LS, MCF and MS collected the data. EE, ID and MP analyzed and interpreted the data. EE performed the gene set analyses and wrote the initial draft of the manuscript. KO and SEF supervised the study and critically reviewed the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-025-03170-2.

Correspondence and requests for materials should be addressed to Else Eising.