Published Ahead of Print on April 28, 2020 as 10.1212/WNL.0000000000009441

ARTICLE

Severe childhood speech disorder

Gene discovery highlights transcriptional dysregulation

Michael S. Hildebrand, PhD,* Victoria E. Jackson, PhD,* Thomas S. Scerri, PhD,* Olivia Van Reyk, MSpPath, Matthew Coleman, BSc (Hons), Ruth O. Braden, MSpPath, Samantha Turner, PhD, Kristin A. Rigbye, BSc (Hons), Amber Boys, PhD, Sarah Barton, DPsych, Richard Webster, MD, Michael Fahey, MD, PhD, Kerryn Saunders, MD, Bronwyn Parry-Fielder, BAppSci, Georgia Paxton, MD, Michael Hayman, MD, David Coman, MD, Himanshu Goel, MD, Anne Baxter, MD, Alan Ma, MD, Noni Davis, MD, Sheena Reilly, PhD, Martin Delatycki, MBBS, PhD, Frederique J. Liégeois, PhD, Alan Connelly, PhD, Jozef Gecz, PhD, Simon E. Fisher, DPhil, David J. Amor, MBBS, PhD, Ingrid E. Scheffer, MBBS, PhD,* Melanie Bahlo, PhD,* and Angela T. Morgan, PhD*

Correspondence
Dr. Hildebrand
michael.hildebrand@unimelb.edu.au
or Dr. Morgan
angela.morgan@mcri.edu.au

Neurology® 2020;94:1-e20. doi:10.1212/WNL.0000000000009441

Abstract

Objective

Determining the genetic basis of speech disorders provides insight into the neurobiology of human communication. Despite intensive investigation over the past 2 decades, the etiology of most speech disorders in children remains unexplained. To test the hypothesis that speech disorders have a genetic etiology, we performed genetic analysis of children with severe speech disorder, specifically childhood apraxia of speech (CAS).

Methods

Precise phenotyping together with research genome or exome analysis were performed on children referred with a primary diagnosis of CAS. Gene coexpression and gene set enrichment analyses were conducted on high-confidence gene candidates.

Results

Thirty-four probands ascertained for CAS were studied. In 11/34 (32%) probands, we identified highly plausible pathogenic single nucleotide (n = 10; *CDK13*, *EBF3*, *GNAO1*, *GNB1*, *DDX3X*, *MEIS2*, *POGZ*, *SETBP1*, *UPF2*, *ZNF142*) or copy number (n = 1; 5q14.3q21.1 locus) variants in novel genes or loci for CAS. Testing of parental DNA was available for 9 probands and confirmed that the variants had arisen de novo. Eight genes encode proteins critical for regulation of gene transcription, and analyses of transcriptomic data found CAS-implicated genes were highly coexpressed in the developing human brain.

Conclusion

We identify the likely genetic etiology in 11 patients with CAS and implicate 9 genes for the first time. We find that CAS is often a sporadic monogenic disorder, and highly genetically heterogeneous. Highly penetrant variants implicate shared pathways in broad transcriptional regulation, highlighting the key role of transcriptional regulation in normal speech development. CAS is a distinctive, socially debilitating clinical disorder, and understanding its molecular basis is the first step towards identifying precision medicine approaches.

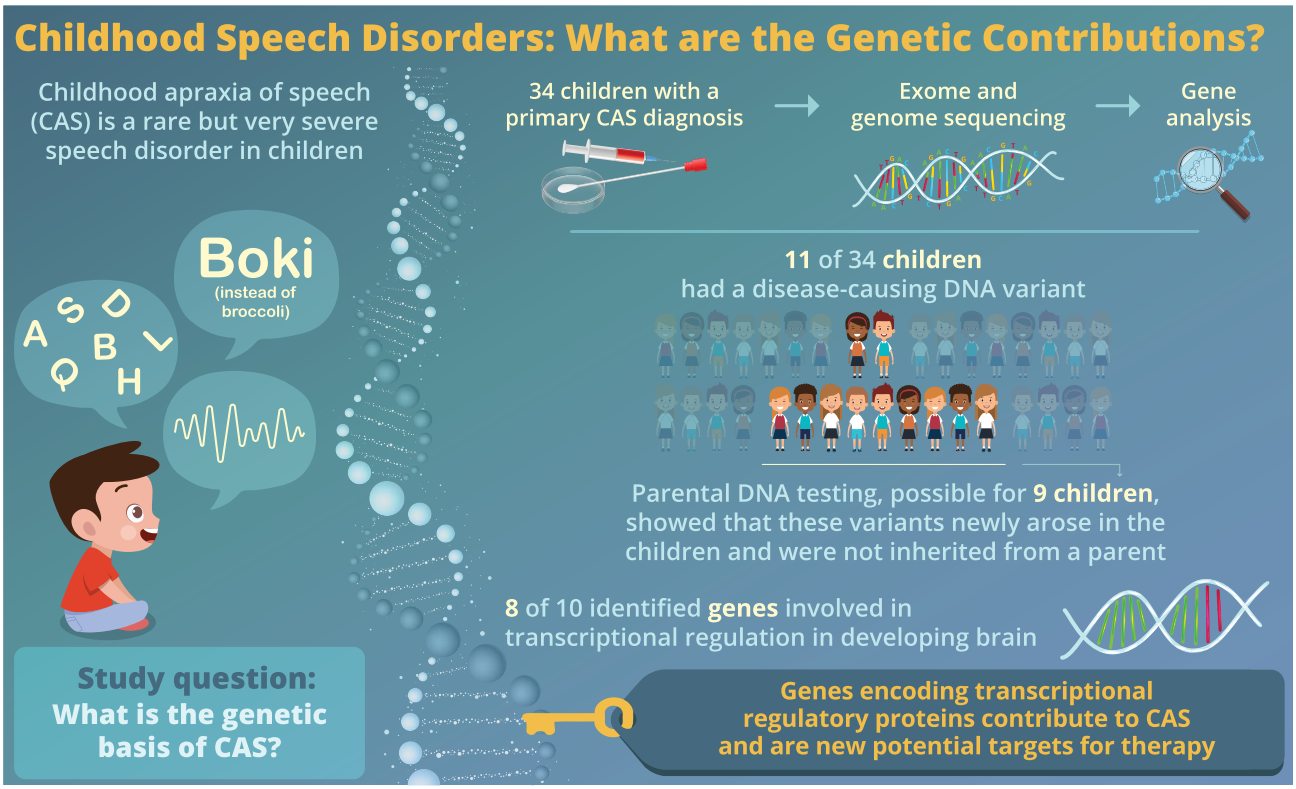
*These authors contributed equally to this work.

From the Department of Medicine (M.S.H., M.C., K.A.R., I.E.S.), The University of Melbourne, Austin Health, Heidelberg; Population Health and Immunity Division (V.E.J., T.S.S., M.B.), The Walter and Eliza Hall Institute of Medical Research; Departments of Medical Biology (V.E.J., T.S.S., M.B.) and Audiology and Speech Pathology (R.O.B., A.T.M.) and Department of Paediatrics, The Royal Children's Hospital (B.P.-F., G.P., M.H., D.J.A., I.E.S.), The University of Melbourne; Speech and Language (O.V.R., R.O.B., S.T., S.B., S.R., A.T.M.), Murdoch Children's Research Institute (M.S.H., D.J.A., I.E.S.); Victorian Clinical Genetics Services (A. Boys, M.D.), Parkville, Victoria; Department of Neurology (R.W.) and Clinical Genetics (A.M.), The Children's Hospital Westmead; Department of Paediatrics (M.F., K.S.), Monash University; Monash Children's Hospital (K.S.), Clayton, Victoria; The Wesley Hospital (D.C.), Auchenflower, Queensland; Hunter Genetics (H.G., A. Baxter), John Hunter Hospital, New Lambton Heights; Melbourne Children's Clinic (N.D.), Victoria; Griffith University (S.R.), Mount Gravatt, Queensland, Australia; UCL Great Ormond Street Institute of Child Health (F.J.L.), London, UK; Florey Institute of Neuroscience and Mental Health (A.C., I.E.S.), Parkville, Victoria; South Australian Health and Medical Research Institute (J.G.), Robinson Research Institute and Adelaide Medical School, University of Adelaide, South Australia; Language and Genetics Department (S.E.F.), Max Planck Institute for Psycholinguistics; and Donders Institute for Brain, Cognition and Behaviour (S.E.F.), Radboud University, Nijmegen, the Netherlands.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Glossary

ACMG = American College of Medical Genetics; **ADHD** = attention-deficit/hyperactivity disorder; **CADD** = combined annotation dependent depletion; **CAS** = childhood apraxia of speech; **DCD** = developmental coordination disorder; **eCDF** = empirical cumulative distribution function; **FS** = Fisher strand; **FSIQ** = full-scale IQ; **GATK** = Genome Analysis Toolkit; **ID** = intellectual disability; **LoF** = loss of function; **MTR** = missense tolerance ratio; **QD** = quality by depth; **RPKM** = reads per kilobase of exon model per million mapped reads; **SIFT** = sorting intolerant from tolerant; **SNV** = single nucleotide variant; **SOR** = strand odds ratio; **STR** = short tandem repeat; **VEP** = variant effect predictor; **VQSR** = variant quality score recalibration; **WES** = whole exome sequencing; **WGS** = whole genome sequencing.



doi:10.1212/WNL.0000000000009441

Neurology®

Childhood speech disorders are common, affecting 1 in 20 preschool children in the general population.[1] The majority of children present with mild articulation (e.g., lisp) or phonological errors (e.g., "f" for "th") that typically resolve with or without intervention.[2] By contrast, approximately 1 in 1,000 patients present with persistent and intractable speech disorders such as childhood apraxia of speech (CAS).[3] These individuals typically have abnormal speech development from infancy, with a history of poor feeding, limited babbling, delayed onset of first words, and highly unintelligible speech into the preschool years, when a diagnosis is usually made.[3] Three core symptoms support a CAS diagnosis, in accordance with consensus-based criteria set by the American Speech-Language-Hearing Association: (1) inconsistent errors on consonants and vowels; (2) lengthened and disrupted coarticulatory transitions between sounds and syllables; and (3) inappropriate prosody. Lifelong impairment is seen, with psychosocial impact, literacy deficits, and restricted educational and employment outcomes.[1]

CAS was not shown to have a genetic basis until 2001, with the seminal discovery that pathogenic variants in *FOXP2* (MIM: 605317), a transcriptional repressor, cause rare cases of CAS (reviewed in reference 4). Later, downstream target *FOXP2* genes such as *CNTNAP2* (MIM: 604569) and closely related family member *FOXP1* (MIM: 605515) were also implicated in speech and language dysfunction.[4] Since then, disruptions of single genes (e.g., *GRIN2A* [MIM: 138253][5]), microdeletions (e.g., 2p16.1, 12p13.33, and 17q21.31 implicating *BCL11A* [MIM: 606557], *ERC1* [MIM: 607127], and *KANSL1* [MIM: 612452][6]), and larger deletions (e.g., 16p11.2 deletion, encompassing >25 genes)[7] have been associated with CAS. A recent genome sequencing study of 19 predominantly US probands with CAS uncovered causal

variants in 8/19 (42%) cases,[8] informing diagnosis and genetic counseling for families.[9] Here, we sought to understand the genetic architecture of CAS by detailed molecular studies of a larger cohort of 34 patients with CAS. We investigated gene coexpression of identified variants with previously published CAS genes.

Methods

Standard protocol approvals, registrations, and patient consents

The Human Research Ethics Committee of The Royal Children's Hospital, Melbourne, Australia, approved this study (Project 37353). Written informed consent was obtained from living participants or their parents or legal guardians in the case of minors or those with intellectual disability.

Phenotyping

Inclusion criteria for probands included a primary clinical diagnosis of severe and persistent speech disorder in childhood (<18 years); that is, not occurring in the setting of severe intellectual disability and where parents and clinicians reported the current primary clinical concern as speech production. Participants were recruited via medical and speech pathology clinicians, online parent support groups for apraxia, or direct parent referral. The medical and developmental history of each proband and participating sibling was taken, with strenuous attempts to obtain all medical, speech, and neuropsychological assessments to identify additional secondary comorbidities, including hearing impairment, motor deficits, epilepsy, attention-deficit/hyperactivity disorder (ADHD), and autism spectrum disorder (tables 1 and 2). Brain MRI results were obtained.

CAS was diagnosed when children met 3 operationally defined American Speech-Language Hearing Association diagnostic criteria[7] scored based on single word transcriptions of the Diagnostic Evaluation of Articulation and Phonology,[10] a polysyllable word test,[11] and a 5-minute conversational speech sample. Dysarthria was diagnosed in the presence of oral tone or coordination disturbance using an oral motor assessment and dysarthric features identified during conversation using the Mayo Clinic Dysarthria rating scale.[6,7] Language, literacy, and cognition were also assessed (table e-1a, doi.org/10.5061/dryad.zkh189363). Parents were assessed with an age-appropriate battery complementary to the child version (table e-1b, doi.org/10.5061/dryad.zkh189363).

Genetic testing

Genomic DNA was extracted from blood using a Qiagen (Valencia, CA) QIAamp DNA Maxi Kit according to the manufacturer's instructions. Only saliva samples were available for some patients, and DNA was extracted using a prepIT•L2P kit (DNA Genotek Inc., Ontario, Canada) according to the manufacturer's instructions. Probands underwent chromosomal microarray testing on Illumina (San Diego, CA) platforms, with the reportable effective resolution of arrays being 200 Kb. Results were analyzed with Karyostudio software version 1.3 or 1.4 (Illumina), using genome reference sequence NCBI36/hg18 (v1.3, pre-2013) or GRCh37/hg19 (v1.4, 2013 onwards).

Variant discovery for the majority of probands was performed using trio or parent–child pair (where 1 parent was unavailable for testing) designs. There were 3 exceptions to this: proband 25, whose monozygotic twin was also sequenced (quad design, twin also affected); proband 26, whose mother, maternal grandmother, and sister were sequenced; and proband 9, who was analyzed as a singleton, as no parental DNAs were available.

Whole exome sequencing (WES) was performed on 64 individuals from 23 families: 24 probands (includes the monozygotic twin pair), 38 parents, and the sister and grandmother of proband 26. Genomic DNA was sonicated to approximately 200 base pair fragments and adaptor-ligated to make a library for paired-end sequencing. Following amplification and barcoding, the libraries were hybridized to biotinylated complementary RNA oligonucleotide baits from the Agilent SureSelect XT Human All Exon + UTR v5 (75 Mb) (Agilent Technologies, Santa Clara, CA) and purified using streptavidin-bound magnetic beads. Amplification was performed prior to sequencing on the Illumina HiSeq 2000 system to average 50-fold depth. Exome sequencing was run on a research basis at the Australian Genome Research Facility, Victorian Comprehensive Cancer Centre, Melbourne.

Whole genome sequencing (WGS) was conducted on 24 individuals from 10 families: 10 probands and 14 parents. Illumina TruSeq DNA Nano genome preparation was completed according to the manufacturer's instructions prior to sequencing on the Illumina X Ten to average 30-fold depth. Genome sequencing was run on a research basis at the Kinghorn Centre for Clinical Genomics, Garvin Institute of Medical Research, Sydney.

The total number of individuals (both unaffected and affected) who had WES or WGS in this study was 88. In the follow-up of candidate variants, targeted Sanger sequencing including on additional family members who had not undergone WES/WGS was carried out to allow further segregation analysis.

Variant analysis and validation

We searched for loss of function (LoF) and predicted damaging variants exome- or genome-wide. Read pairs were mapped to the hg19 reference genome using Burrow-Wheeler Aligner (BWA-MEM, bwa v0.7.15).[12] Reads were sorted using SAMtools (v 1.7) and duplicates marked using Genome Analysis Toolkit (GATK) v4.0.11.0.[13] Base quality score recalibration was performed and variants called using HaplotypeCaller, on a per-sample basis, as implemented by GATK. Genotype calling and quality filtering were performed

Table 1 Medical and neurodevelopmental features of childhood apraxia of speech (CAS) cohort

Case	Age, y; mo	Sex	Core speech phenotype	Gross motor delays	Fine motor delays	Vision impaired	Hearing loss	MRI findings	Seizures	Other NDD	Toileting delays	Dysmorphic features	Other medical
1	8;11	F	CAS, dysarthria	Y	Y	N	N	N	Febrile	N	Y	N	NR
2	11;5	M	Severe phonological	N	N	Glasses	N	NA	N	N	Y	Clinodactyly 5th fingers	Asthma, eczema
3	5;0	F	CAS, phonological delay and disorder	Y	Y	N	N	Mild thinning posterior CC, reduced WM	N	Attention deficits	N	N	NR
4	6;7	F	CAS, phonological delay, articulation disorder	Y	Y	N	N	Nonspecific frontal gliosis	Bilateral temporal discharges at 6 years	Attention deficits	N	Retrognathia	NR
5	4;8	M	CAS, dysarthria	Y	Y	N	N	NA	N	Behavioral problems due to speech frustration	Y	N	Ataxia
6	8;9	F	CAS, phonological delay, articulation disorder	Y	Y	N	N	N	N	NA	Y	Narrow palpebral fissures, arched eyebrows, low columnella, hypoplastic alar nasae	NR
7	11;3	F	CAS	Y	Y	N	N	NA	N	Learning deficits	N	High nasal root, prominent nose, thin upper lip	Atrial septal defect
8	5;1	F	CAS, phonological delay and disorder	Y	Y	N	N	NA	N	Learning deficits	Y	Brachycephaly, flat midface, anteverted nares, cupid bow upper lip	NR
9	16; 10	F	CAS	Y	Y	Glasses	N	NA	N	Mild ASD, auditory processing deficits	Y	Arched eyebrows, sparse laterally, cleft lip and palate repair	NR
10	9;1	F	CAS, dysarthria, phonological delay	Y	Y	Glasses	N	N	NR	Mild ASD[b]	Y	Brachycephaly, small mouth, thin upper lip	Mastocytosis, L hemiplegia
11	4y	M	CAS	Y	Y	N	N	NA	N	Mild ASD	Y	Cupid bow upper lip, hypoplastic columnella	Cystoscopy + retrograde pyelogram, L pelvic kidney without significant reflux
12	8	M	CAS	Y	Y	N	N	N	N	Mild ASD, ADHD	Y	NR	NR
13	6;9	M	CAS, phonological delay and disorder	Y	Y	N	N	NA	N	ADHD, Tourette	Y	NR	NR

Continued

Table 1 Medical and neurodevelopmental features of childhood apraxia of speech (CAS) cohort *(continued)*

Case	Age, y; mo	Sex	Core speech phenotype	Gross motor delays	Fine motor delays	Vision impaired	Hearing loss	MRI findings	Seizures	Other NDD	Toileting delays	Dysmorphic features	Other medical
14	6;11	M	CAS, phonological delay	N	N	N	N	N	N	NA	N	N	Celiac HLA DQ8 haplotype
15	7;9	M	CAS, phonological delay	N	N	N	N	NA	N	N	N	Triangular face, anteverted ears, broad nasal root	NR
16	4;4	M	CAS, phonological delay and disorder	Y	Y	N	N	NA	N	Attention deficits	Y	N	NR
17	11;1	M	CAS, phonological delay	Y	Y	N	N	Multiple foci hyperintensity subcortical WM	N	Mild ASD, ADD, anxiety, depression	Y	N	NR
18	14;1	F	CAS, dysarthria, articulation disorder	Y	Y	N	N	WM hyperintensity below R MFG	N	Attentional and emotional deficits, anxiety, depression	Y	N	Overbite, braces
19	2;9	M	CAS	N	N	N	N	NA	N	N	Y	N	NR
20	11; 11	F	CAS, dysarthria, phonological delay and disorder, articulation disorder	Y	Y	N	N	Delayed frontal lobe myelination	N	Motor dyspraxia	Y	N	NR
21	6;8	M	CAS	Y	Y	N	R low-frequency SNHL	N	N; discharges in sleep[a]	ID	N	Broad forehead, mild hypertelorism	NR
22	3;11	M	CAS	Y	N	N	N	N	4 Febrile	N	Y	N	NR
23	5;9	F	CAS, phonological delay and disorder, articulation disorder	N	Y	N	N	NA	N	N	Y	N	NR
24	5	M	CAS	Y	Y	N	N	N	N	DCD, behavioral deficits	Y	NR	Peanut allergy
25 (a)	4	M	CAS	N	Y	N	N	NA	N	NA	Y	NA	Tongue tie
25 (b)	4	M	CAS	N	Y	N	N	NA	N	NA	Y	NA	Tongue tie
26	5	M	CAS	Y	Y	N	N	NA	N	NA	N	NR	NR

Continued

Table 1 Medical and neurodevelopmental features of childhood apraxia of speech (CAS) cohort *(continued)*

Case	Age, y; mo	Sex	Core speech phenotype	Gross motor delays	Fine motor delays	Vision impaired	Hearing loss	MRI findings	Seizures	Other NDD	Toileting delays	Dysmorphic features	Other medical
27	4;8	M	CAS, phonological delay and disorder, articulation disorder	N	N	N	N	NA	N	N	N	N	NR
28	8;0	F	CAS, phonological delay and disorder	Y	Y	N	N	NA	N	Attention deficits	Y	Large upturned earlobes, brachydactyly, 2,3 toe syndactyly, metacarpal and metatarsal shortening	Central obesity, insulin resistance
29	6;5	F	CAS	Y	Y	N	N	N	N	N	N	N	NR
30	7;8	F	CAS, dysarthria, phonological delay	N	Y	N	N	1 Small focus subcortical hyperintensity	2 Normal EEGs	Mild ASD, migraine, behavioral deficits	N	N	Obesity; sleep issues
31	4;0	M	phonological delay, phonological disorder	Y	Y	N	N	NA	Jerking, 2 normal EEGs	N	Y	Glabellar flame nevus, full nasal root and tip, prominent tongue	NA
32	5;3	M	CAS	N	Y	N	N	NA	N	Learning deficits	Y	N	NR
33	4;10	F	CAS, phonological delay	Y	N	N	N	NA	N	N	N	N	Gluten intolerant

Abbreviations: ADD = attention-deficit disorder; ADHD = attention-deficit/hyperactivity disorder; ASD = autism spectrum disorder; CC = corpus callosum; DCD = developmental coordination disorder; HLA = human leukocyte antigen; MFG = medial frontal gyrus; NA = not assessed; NDD = neurodevelopmental disorder; NR = not reported; SNHL = sensorineural hearing loss; WM = white matter.
[a] Not sufficient to cause epilepsy aphasia syndrome.
[b] Diagnosis reported to be "debatable" by parent.

Table 2 Extended linguistic phenotype and educational outcomes of childhood apraxia of speech cohort

Case	Oral motor impairment	History of feeding issues	Language: receptive	Language: expressive	Reading deficits	Spelling deficits	Speech pathology	IQ[a]	Education setting
1	Y	Y	Severe	Severe	Y	Y	Y	BDLN (FSIQ)	Specialist
2	N	Y	Mild	Severe	Low	Below average	Y	Low average (FSIQ)	Mainstream
3	Y	N	Mild	Mild	Y	NA	Y	BDLN (FSIQ), low average (verbal IQ), low average (NV IQ)	Mainstream
4	Y	Y	Severe	Severe	NA	NA	Y	Extremely low (FSIQ), extremely low (verbal IQ), extremely low (performance score)[c]	School for deaf (because child was signing, but is not deaf)
5	Y	Y	Above average	NA: speech too severe to test	NA	NA	Y	NA	Not yet at school
6	Y	Y	Average	Severe	NA	NA	Y	Unable to calculate FSIQ (clinician concluded moderate impairment)	Mainstream then specialist
7	Y	N	Mild	Average	Y	Y	Y	Low average (FSIQ)	Mainstream
8	Y	N	Mild	Severe	NA	NA	Y	BDLN (FSIQ)	Mainstream kindergarten
9	Y	Y	Moderate –severe	NA: speech too severe to test	Lower extreme	Y	Y	NA	Specialist
10	Y	Y	Severe	Severe	Y	Y	Y	BDLN (FSIQ), BDLN (verbal scale), extremely low (performance scale), BDLN (processing speed)	Mainstream
11	Y	N	Moderate	Severe	Y	Y	Y	NA	Not yet at school
12	Y	Y	NA	NA	NA	NA	Y	NA	Mainstream
13	NA	N	Average	Severe	High average	High average	Y	Low average (FSIQ), average (processing speed), BDLN (working memory), average (perceptual reasoning), low average (verbal comprehension)	Mainstream
14	Y	Y	Moderate	Severe	Average	Average	Y	NA	Mainstream
15	Y	N	Average	Moderate	Y	Y	Y	NA	Mainstream
16	Y	Y	Severe	Severe	NA	NA	Y	NA	Mainstream
17	N	N	Moderate	Moderate	Y	Y	Y	Extremely low (FSIQ), extremely low (verbal), extremely low (processing speed), BDLN (NV)	Mainstream
18	Y	N	Mild	Severe	Y	Y	Y	Unable to calculate FSIQ; low average (verbal comprehension), extremely low (perceptual reasoning), extremely low (processing speed), BDLN (working memory)	Mainstream
19	Y	Y	Above average	NA: speech too severe to test	NA: too young	NA: too young	Y	NA	Not yet at school
20	Y	Y	Severe	NA: speech too severe to test	Y	Y	Y	Extremely low (FSIQ)	Mainstream
21	Y	N	Severe	Severe	NA	Y	Y	Extremely low (FSIQ)	Mainstream
22	Y	N	Average	NA: speech too severe to test	NA	NA	Y	NA	Not yet in school

Continued

Table 2 Extended linguistic phenotype and educational outcomes of childhood apraxia of speech cohort *(continued)*

Case	Oral motor impairment	History of feeding issues	Language: receptive	Language: expressive	Reading deficits	Spelling deficits	Speech pathology	IQ[a]	Education setting
23	Y	N	Average	Severe	NA	NA	Y	Average (verbal), superior (NV), average (processing speed)	Mainstream
24	NA	N	Moderate	Moderate	NA	NA	Y	Unable to calculate FSIQ; BDLN (verbal), low average–average (NV)	Mainstream
25[b]	Y	Y	Mild	Moderate	NA	NA	Y	NA (PPVT WNL)	Mainstream kindergarten
25[c]	Y	Y	Mild	Moderate	NA	NA	Y	NA (PPVT WNL)	Mainstream kindergarten
26	NA	Y	Average	Average	NA	NA	Y	NA	Mainstream kindergarten (repeating kindergarten due to speech)
27	N	N	Average	Mild	NA	NA	Y	NA	Mainstream kindergarten
28	Y	N	Moderate	Severe	Y	Y	Y	BDLN (FSIQ)	Mainstream
29	NA	N	Severe	Average	NA	NA	Y	NA	Mainstream
30	Y	N	Average	Mild	Y	NA	Y	BDLN (verbal comprehension), low average (perceptual reasoning), low average (working memory), average (processing speed)	Mainstream
31	N	Y	Average	NA: speech to severe to test	NA	NA	Y	Average	Mainstream
32	Y	N	Moderate	Severe	NA	NA	Y	Extremely low (FSIQ), extremely low (verbal comprehension), low average (visual spatial, fluid reasoning, working memory), BDLN (processing speed)	Mainstream kindergarten
33	Y	N	Average	Mild	NA	NA	Y	NA	Mainstream kindergarten

Abbreviations: BDLN = borderline; FSIQ = full-scale IQ; NA = not assessed; NV = nonverbal; PPVT = Peabody Picture Vocabulary Test; WNL = within normal limits.
Average = 90–109; low average = 80–89; borderline = 70–79; extremely low = 69 and below.
[a] IQ performance severity descriptors were converted to the same synonymous terms across tools for ease of comparison.
[b] Wide discrepancy in performance in nonverbal subtests and unable to complete verbal subtests due to severe speech impairment; PPVT used as limited proxy for NV IQ.
[c] Results from 3 years prior were less severe: i.e., borderline (FSIQ 76), low average (verbal IQ), borderline (performance score).

separately in the exome and genome sequencing batches, as follows. Per-sample gvcf files were merged and genotypes were jointly called across all samples using GATK's GenotypeGVCFs tool. Variants with excess heterozygosity (Z score >4.5) were removed, then variant quality score recalibration (VQSR) was carried out for single nucleotide variants (SNVs) and indels separately, and a truth sensitivity filter of 99.7 was used to flag variants for exclusion. SNVs were filtered to exclude those flagged by VQSR or any of the following hard filters: low quality by depth (QD <2); evidence of strand bias (Fisher strand [FS] >60 or strand odds ratio [SOR] >3); and evidence of differences between alternate and reference alleles for read mapping qualities (rank sum <−12.6) or position bias (ReadPosRankSum <−8). Indels were filtered to exclude any of the following: those flagged by VQSR; QD <2; FS >200; SOR >10; or ReadPosRankSum <−20.

Analysis was restricted to variants (1) either not present in gnomAD or present with a mean allelic frequency <0.05%, (2) not present in unaffected family members from our sequenced cohort and (3) potentially de novo, or consistent with an appropriate inheritance model. Only variants with read depth >10 and genotype quality >20 in the proband and their sequenced family members were considered. Identified variants were annotated using variant effect predictor (VEP v93.3) using assembly version GRCh37.p13 and categorized based on the following series of annotations.

Predicted LoF candidates

Predicted LoF candidates were defined using VEP annotations meeting 3 criteria: (1) annotated as splice acceptor variant, splice donor variant, frameshift variant, stop lost, stop gained, or start lost; (2) in a gene intolerant to LoF variation (ExACpLI ≥0.9 or LoFtool <0.1); (3) at least one of the following: (a) predicted to be damaging by combined annotation dependent depletion (CADD) Phred score ≥20 or (b) predicted to affect splicing (AdaBoost score ≥0.6 or random forest score ≥0.6 using the dbscSNV VEP plugin). For frameshift variants, the variant was only required to be in an LoF intolerant gene (i.e., criterion 3 was not required).

Predicted damaging candidates

Missense variants had to meet at least 3 criteria: (1) predicted "probably damaging" or "possibly damaging" by PolyPhen-2; (2) predicted "deleterious" or "deleterious low confidence" by sorting intolerant from tolerant (SIFT); (3) predicted damaging with CADD Phred score ≥20; (4) missense tolerance ratio (MTR) significantly different from 1 (false discovery rate <0.05); (5) predicted to affect splicing (AdaBoost score ≥0.6 or random forest score ≥0.6).

Other notable candidates

Missense variants that did not meet the above criteria, but were in genes with biological relevance to speech based on the literature, were also identified as candidates. All candidates were inspected visually in the Integrative Genome Viewer (IGV1.3).

Criteria for reporting rare or novel variants

We report a set of high-confidence candidate variants, categorized as either predicted LoF or damaging candidates, and classified as pathogenic according to the American College of Medical Genetics (ACMG) guidelines.[14] For probands without a high-confidence variant, we report low-confidence candidate variants; these comprise all identified LoF candidates classified as likely pathogenic or of uncertain significance (ACMG guidelines), and a subset of missense variants, in genes of biological relevance to speech based on the literature. ACMG guidelines strictly only apply to known disorder-causing genes.[14]

Rare variant validation

Variants of interest were validated using PCR and Sanger sequencing. Gene variants were amplified using gene-specific primers (oligonucleotide sequences available on request) designed to the reference human gene transcripts (NCBI Gene). Amplification reactions were cycled using a standard protocol on a Veriti Thermal Cycler (Applied Biosystems, Carlsbad, CA) at 60°C annealing temperature for 1 minute. Bidirectional sequencing of all exons and flanking regions was completed with a BigDye v3.1 Terminator Cycle Sequencing Kit (Applied Biosystems), according to the manufacturer's instructions. Sequencing products were resolved using a 3730xl DNA Analyzer (Applied Biosystems). All sequencing chromatograms were compared to the published cDNA sequence; nucleotide changes were detected using Codon Code Aligner (CodonCode Corporation, Dedham, MA).

Interrogation of short tandem repeats

We also examined whether any proband had expanded short tandem repeats (STRs) at any known pathogenic locus (table e-2, doi.org/10.5061/dryad.zkh189363). Genome and exome sequenced samples were examined separately using 2 short tandem repeat detection tools, Expansion Hunter v.2.5.5 and exSTRa. For each locus, we looked for evidence of outlying samples in terms of STR length by inspecting plots of estimated STR size (ExpansionHunter), and empirical cumulative distribution function (eCDF) plots of the number of repeated bases observed for each sample.

Gene coexpression networks and gene set enrichment analyses

Normalized brain expression values (reads per kilobase of exon model per million mapped reads [RPKM]) from the BrainSpan Developmental Transcriptome dataset[15] (Gencode v10 summarized to genes) were used for the gene coexpression analyses. Samples were restricted to include those from all available brain regions, from fetal and infancy periods only (8 postconception weeks to 10 months after birth; data for included samples, tables e-3 and e-4, doi.org/10.5061/dryad.zkh189363). Following sample restriction, genes were removed if they had expression values missing from >50% of samples, expression values of 0 RKPM for ≥50% of samples, or variance of expression across all samples

<0.5. A total of 15,392 genes across 280 samples from 24 individuals remained in the filtered data set. Finally, expression values were $\log_2$ transformed.

Using the log-transformed expression values, a matrix of weighted correlations was generated, with weights determined as $1/\sqrt{n}$, where n is the number of samples contributed by the respective individual. Correlation plots were visualized using the corrplot R package (version 0.84, available at github.com/taiyun/corrplot), with genes ordered by hierarchical clustering, using the median linkage method. Networks of the most highly coexpressed genes were constructed using the qgraph R package.[16] Using the distribution of pairwise correlations of all 15,392 genes in the dataset, a threshold of $|\rho| > 0.647$ was determined, corresponding to the absolute correlation value that the 5% most highly correlated genes exceeded. Networks were then constructed with edges drawn between genes with absolute pairwise correlations above this threshold.

Finally, we determined whether these genes were more highly coexpressed than would be expected for a random set of genes. Given the very large number of combinations of gene sets possible, selected from the full set of 15,392, we used a Monte Carlo sampling approach to approximate the distribution of the median $|\rho|$ for all sets of genes. To this end, we randomly sampled 5,000 sets of genes, the same size as our high-confidence set, and calculated the median $|\rho|$ for each random gene set. We derived an eCDF based on these medians, to which we compared the observed median $|\rho|$ of our high-confidence candidates. Replication of all coexpression analyses was undertaken using independent samples (e-Methods, doi.org/10.5061/dryad.zkh189363).

Gene set enrichment analyses were undertaken using g:Profiler[17] and Gene Ontology molecular function, cellular component, and biological processes databases and KEGG and Reactome pathways.[18,19] A Bonferroni-corrected p value <0.05 was used to determine significant overrepresentation of our candidate genes in a pathway.

Data availability

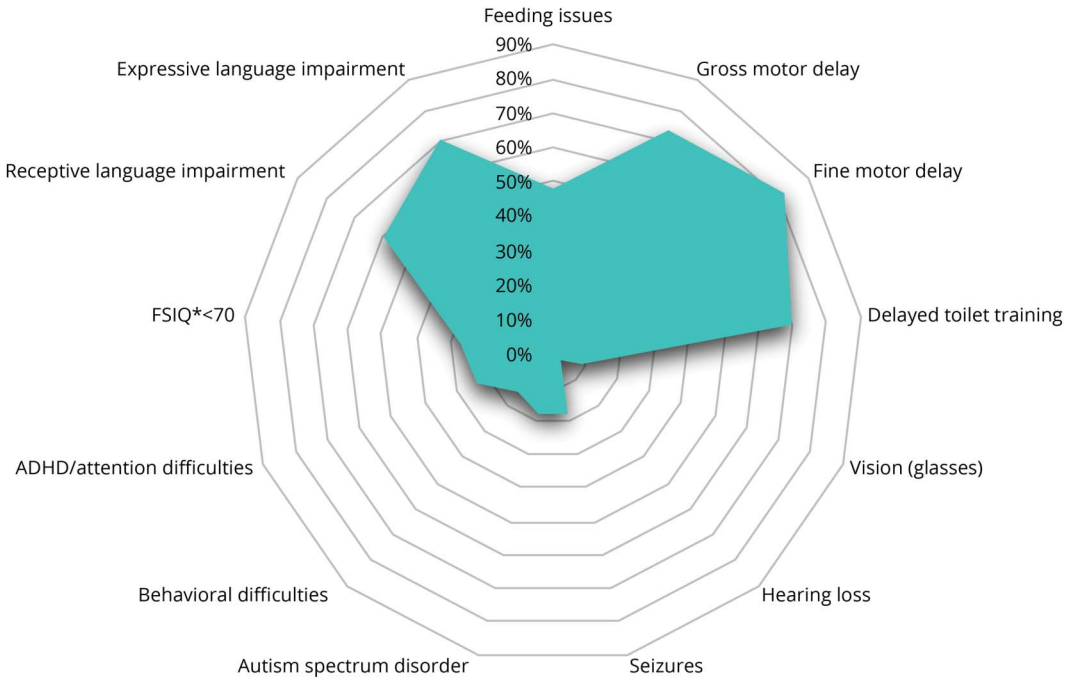
Data not available in this article are available at doi.org/10.5061/dryad.zkh189363.

Results

Phenotypic data

Thirty-four probands (16 male), with a median age of 8 years (range 2 years 9 months to 16 years 10 months), including 1 monozygotic twin pair, were studied (table 1 and figures 1 and 2). Feeding difficulties during infancy or during transition to solids were reported in 16 individuals. Early speech milestones were delayed in 33/34 individuals. Thirty-two children had CAS, either in isolation (n = 13) or co-occurring with other speech disorders of dysarthria (n = 6), phonological delay or disorder (n = 18), or articulation disorder (n = 4) (table 1). Two children (2, 31) ascertained for CAS had phonological disorders on testing, rather

Figure 1 Phenotypic overlap in childhood apraxia of speech cohort



ADHD = attention-deficit/hyperactivity disorder; FSIQ = full-scale IQ.

than CAS. Oral motor coordination and range of movement deficits occurred in 26. Poor performance during single nonspeech oromotor movements reflected impaired lingual movements (e.g., reduced tongue elevation and lateralization), labial–facial movements (e.g., poor lip rounding), and mandibular control (e.g., reduced jaw excursion and stability). Impaired double nonspeech oromotor movements (e.g., "smile and kiss") were also seen, typified by impaired transition, precision of movements, and groping (overt struggle, effort, or excessive excursion of the articulators) (table 2). In 7 children, expressive language could not be evaluated due to poor compliance (n = 1) or severity of verbal impairment (n = 6).

Hearing was normal in all except one child, who wore a hearing aid for unilateral low-frequency sensorineural hearing loss. Two children had a history of severe recurrent otitis media necessitating grommet insertion. A total of 10/34 (29%) patients had dysmorphic features (table 1). Nineteen children had an IQ assessment showing average (n = 2), low average (n = 3), borderline (n = 5), and extremely low average (n = 5) full-scale IQ (FSIQ) (table 2). All but 2 children were attending mainstream schools. For 5 children, an FSIQ could not be calculated because of significant variable performance across verbal and nonverbal subscales. The remainder of the cohort did not have IQ testing, largely because of young age (under age 5 years) or the family declined. Other features included mild autism spectrum disorder (n = 6), ADHD/ADD (n = 3), difficulties with attention (n = 6), Tourette syndrome (n = 1), behavioral problems (n = 5), and anxiety and mood-related symptoms (n = 2). Gross motor (n = 24) and fine motor delays (n = 26) were common, with a slower trajectory in learning to ride a bike, balance appropriately, draw, write, and cut compared to typical peers. Body praxis or developmental coordination disorder (DCD) diagnoses were reported in just 2 children. One 16-year-old adolescent with a repaired cleft lip and palate had severe CAS with unintelligible speech not attributable to the cleft. Several children had a history of seizures: 2 had epilepsy, with 1 on valproate, 2 had febrile seizures, and a further 2 had unconfirmed seizures. Six probands had MRI brain abnormalities including mild thinning of the corpus callosum (case 3), nonspecific frontal gliosis (case 4), foci of white matter hyperintensity in bilateral parietal and posterior fossa (case 17) or right medial frontal gyrus (case 18), 1 small focus of subcortical hyperintensity (case 30), and delayed frontal lobe myelination (case 20). A total of 23/34 children had delayed independent toileting. All cases were receiving or had received speech therapy.

Copy number analysis and short tandem repeats

Chromosomal microarray testing was performed in all patients. Only 1 proband (patient 6) had a significant finding with a de novo mosaic deletion of approximately 9.2 megabases on chromosome 5q14.3q21.1 in about 75% of cells (genomic coordinates GRCh37/Hg19 chr5: 90,779,680-99,959,810) (figure 3 and table 3). We also searched for evidence of expansions of known pathogenic STRs. Most disorders caused by expanded STRs affect the nervous system and often include speech problems such as dysarthria. We found no evidence for an expanded STR in any patient.

Exome and genome sequence analysis

We identified candidate variants in 21/34 (62%) patients (table 3 and figures 2 and 3). We found 12 high-confidence variants: 5 missense, 3 frameshift, and 3 nonsense (stop gain) in 10 genes (*CDK13* [MIM: 603309], *EBF3* [MIM: 607407], *GNAO1* [MIM: 139311], *GNB1* [MIM: 139380], *DDX3X* [MIM: 300160], *MEIS2* [MIM: 601740], *POGZ* [MIM: 614787], *SETBP1* [MIM: 611060], *UPF2* [MIM: 605529], *ZNF142* [MIM: 604083]), and a large mosaic deletion (5q14.3q21.1) by chromosomal microarray. Nine high-confidence variants were confirmed de novo dominant: 1 pair was recessively inherited (compound heterozygous) and, for 1, inheritance could not be assessed by segregation analysis as the proband was adopted. All variants were novel, except for 1 of the compound heterozygous variants, according to the gnomAD database (table 3.a and figure 2). The 6 nonsense or frameshift variants were all in genes intolerant to LoF variation (*DDX3X*, *EBF3*, *GNB1*, *MEIS2*, *SETBP1*, *UPF2*), according to ExACpLI or LoFtool scores. The 5 missense variants were all predicted to be damaging by 3 in silico tools (SIFT, PolyPhen, and CADD). All 12 variants were classified as pathogenic according to ACMG guidelines.[14]

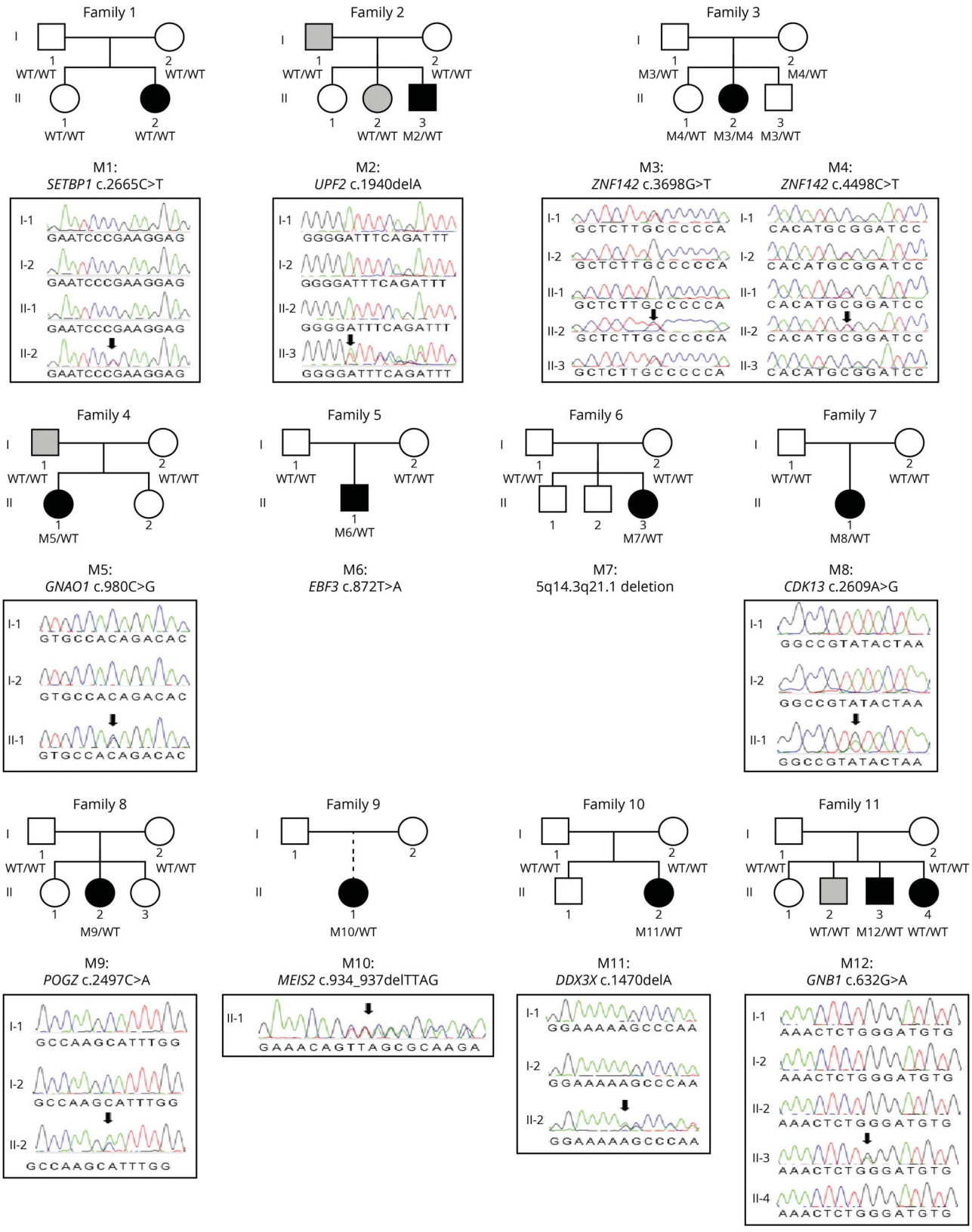
In 9/34 (26%) probands, we found very rare (<0.05%) missense variants predicted to be damaging by multiple in silico tools (table 3) (full list of predicted damaging candidates are available in table e-5, doi.org/10.5061/dryad.zkh189363). This list included variants in *BRWD3* (MIM: 300553), *UBA6* (MIM: 611361), *PTBP2* (MIM: 608449), *ZKSCAN1* (MIM: 601260), *TENM4* (MIM: 610084), and *ASTN2* (MIM: 612856) (table 3.b). We also identified rare variants in *GRIN2A* (MIM: 138253), implicated in epilepsy–aphasia syndromes,[5] and *KIRREL3* (MIM: 607761) in nonsyndromic intellectual disability (*KIRREL3*), but these variants did not meet our strict criteria for predicted damaging candidates.

In a further 4 probands, we identified 5 novel or very rare LoF variants in genes predicted to be intolerant to variation, which were classified as of uncertain significance for CAS (table 3.c). These variants are all predicted to be among the most damaging in these probands; however, none of these genes has been implicated in CAS or neurodevelopmental disorders to date.

Gene coexpression during brain development

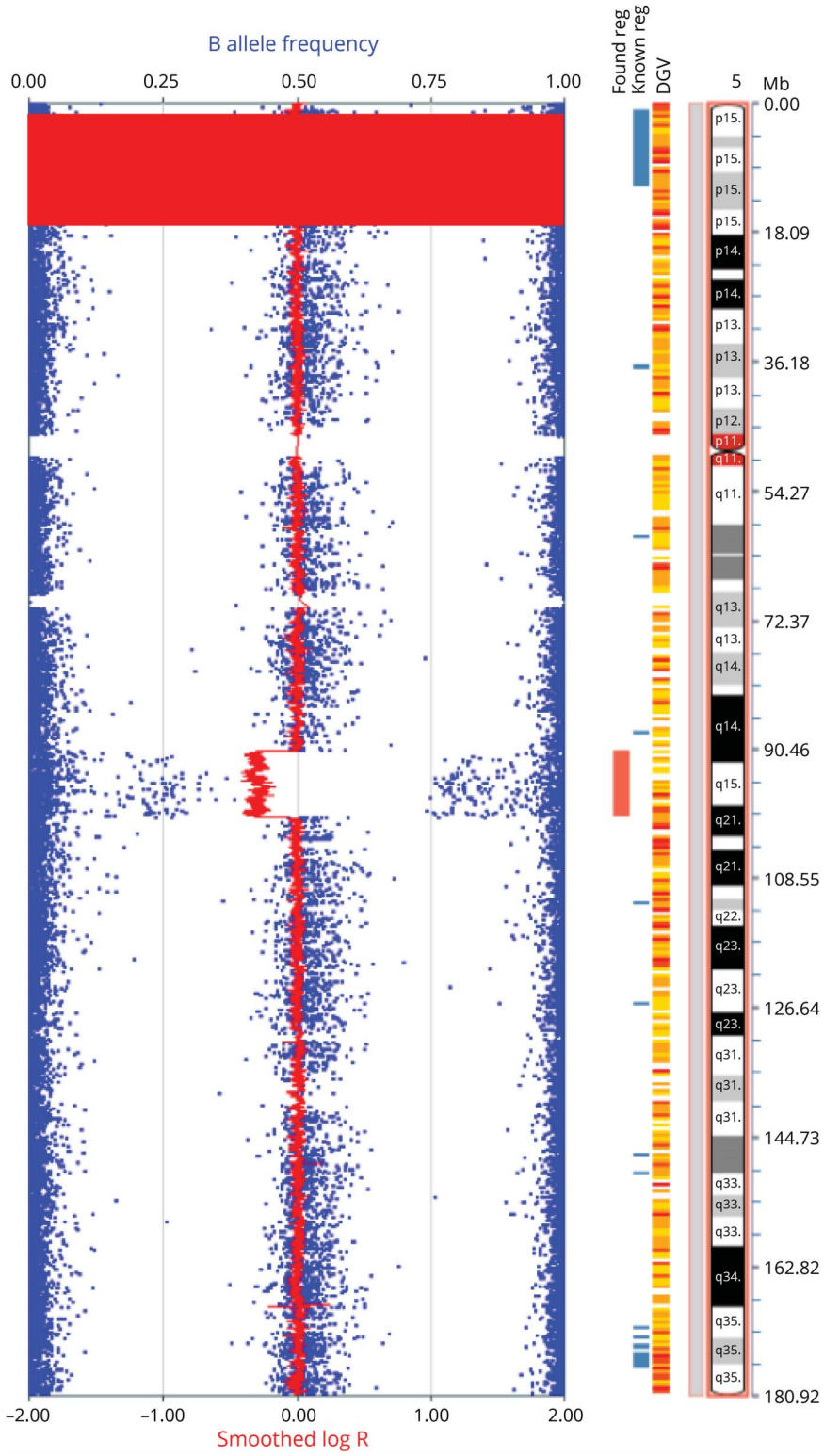
Using brain expression data (RNA-seq) from BrainSpan, we examined coexpression of our 10 high-confidence candidate genes (figure 4A). The median absolute correlation between our 10 high-confidence candidate genes was $|\rho| = 0.463$, and 10 out of the 45 pairwise correlations were among the top 5% most highly correlated gene pairs genomewide ($|\rho| > 0.647$, figure 4B). Using a Monte Carlo sampling approach, we found evidence that this set of genes was more highly coexpressed than expected by chance

Figure 2 Families with high-confidence variants



Families 1–6 analyzed by whole exome sequencing. Pedigrees from 6 families show segregation of 7 high-confidence variants. Sequence chromatograms show confirmed de novo variants in the probands of families 1, 2, and 4, and confirmed compound heterozygous variants in the proband of family 3. Sanger sequencing was not performed for the variant in family 5, and the proband in family 6 had a large deletion, as shown in figure 3. Families 7–11 analyzed by whole genome sequencing. Pedigrees from 5 families show 5 high-confidence variants. Sequence chromatograms show confirmed de novo variants in the probands of families 7, 8, 10, and 11. The proband in family 9 was adopted and her biological parents were unavailable for testing.

Figure 3 Large mosaic deletion in family 6



Illumina Karyostudio image shows the Illumina Infinium Global Screening Array-24v1.0 single nucleotide polymorphism data for chromosome 5. The Smoothed Log R (representing copy number) is depicted as a red line, and the B allele frequency (representing genotyping) is depicted as blue dots. The mosaic 9.2 Mb deletion of chromosome 5q14.3q21.1 is observed as a negative shift in the smoothed log R and a change in the genotyping at 5q14.3 to q21.1. The deletion is present in approximately 75% of cells.

(p = 0.006; figure e-1, doi.org/10.5061/dryad.zkh189363). This suggests that these genes form part of a common pathway affected in CAS, empirically captured by our results. When expanding the coexpression analyses to include the 8 candidate genes for CAS in Eising et al.,[8] we found strong overlap in coexpression patterns between these genes and our 10 high-confidence candidates (figure 4C; figure e-2, doi.org/10.5061/dryad.zkh189363). This set of 18 genes had a median correlation that was significantly

Table 3 Gene variants in childhood apraxia of speech cohort

	Sex	Method	Chr:Pos	Gene	DNA variant	Protein change	Effect	In silico predictions[a]	gnomAD count[b]	Inheritance	ACMG score	Reference
High-confidence variants: pathogenic variants according to ACMG guidelines												
1	F	WES	18: 42531970	*SETBP1*	c.2665C>T	p.R889*	Nonsense	ExACpLI = 1; LoFtool = 0.0297; CADD = 38	0	de novo	PP3, PP4, PM2, PM4, PS2, PS3, PVS1, class 5 pathogenic	Eising et al.[8,c]
2	M	WES	10: 12021068	*UPF2*	c.1940delA	p.F648Sfs*23	Frameshift	ExACpLI = 1	0	de novo	PP4, PM2, PM4, PS2, PS3?, PVS1, class 5 pathogenic	Johnson et al.[28,c]
3	F	WES	10: 219507541, 10: 219505483	*ZNF142*	c.3698G>T, c.4498C>T	p.C1233F, p.R1500W	Missense, Missense	SIFT = Del(0)/Del(0); PolyPhen = Dam (0.998)/Dam (0.998); CADD = 31/26	0 1	Compound heterozygous	PP3, PP4, PM3, PS3?, PVS1, class 5 pathogenic	Khan et al.[40,c]
4	F	WES	16: 56388880	*GNAO1*	c.980C>G	p.T327R	Missense	SIFT = Del(0); PolyPhen = Dam (1); CADD = 28.3	0	de novo	PP3, PP4, PM2, PS2, PVS1, class 5 pathogenic	—
5	M	WES	10: 131666059	*EBF3*	c.872T>A	p.L291*	Nonsense	ExACpLI = 0.999; LoFtool = 0.0389; CADD = 39	0	de novo	PP3, PP4, PM2, PM4, PS2, PVS1, class 5 pathogenic	—
6	F	WES, CMA		5q14.3q21.1 deletion	NA		LOH	NA	0	de novo mosaic	PP4, PM2, PS2, PVS1, class 5 pathogenic	—
7	F	WGS	7: 40102433	*CDK13*	c.2609A>G	p.Y870C	Missense	SIFT = Del(0); PolyPhen = Dam (0.996); CADD = 32; MTR FDR = 0.031	0	de novo	PP3, PP4, PS2, PM2, PVS1, class 5 pathogenic	—
8	F	WGS	1: 151379435	*POGZ*	c.2497C>A	p.H833N	Missense	SIFT = Del(0); PolyPhen = Dam (0.968); CADD = 28.2	0	de novo	PP3, PP4, PS2, PM2, PVS1, class 5 pathogenic	—
9	F	WGS	15: 37242564	*MEIS2*	c.934_937delTTAG	p.L312Rfs*11	Frameshift	ExACpLI = 0.99; LoFtool = 0.091	0	Parents unavailable	PP3, PP4, PM2, PM4, PVS1, class 5 pathogenic	—
10	F	WGS	X: 41205635	*DDX3X*	c.1470delA	p.S492Afs*4	Frameshift	ExACpLI = 1; LoFtool = 0.0555	0	de novo	PP3, PP4, PM2, PM4, PS2, PS3?, PVS1, class 5 pathogenic	Beal et al.[21,c]
11	M	WGS	1:1721901	*GNB1*	c.632G>A	p.W211*	Nonsense	ExACpLI = 1; CADD = 40	0	de novo	PP3, PP4, PM2, PS2, PVS1, class 5 pathogenic	—

Continued

Table 3 Gene variants in childhood apraxia of speech cohort *(continued)*

	Sex	Method	Chr:Pos	Gene	DNA variant	Protein change	Effect	In silico predictions[a]	gnomAD count[b]	Inheritance	ACMG score	Reference
Predicted damaging variants classified as likely pathogenic, or with uncertain significance (ACMG guidelines)												
12	M	WES	1: 97216982	*PTBP2*	c.74G>C	p.R25T	Missense & splice region	SIFT = Del(0); PolyPhen = PosDam (0.641); CADD = 32; MTR FDR = 0.043; Ada = 0.981; RF = 0.886	0	de novo	PP3, PM2, PS2, class 4 likely pathogenic	—
14	M	WES	16: 9858387	*GRIN2A*	c.3014A>G	p.K1005R	Missense	CADD = 21.8	0	Inherited from affected father	PP1, PP4, PM2, class 3 uncertain significance	—
15	M	WES	11: 126294626	*KIRREL3*	c.2186G>T	p.S729I	Missense	CADD = 23.7	1	Unconfirmed, father unavailable	PP1, PP4, class 3 uncertain significance	—
16	M	WES	11: 78614398, 11: 78574177	*TENM4*	c.664G>A, c.1085C>T	p.G222R, p.A362V	Missense, Missense & splice region	PolyPhen = PosDam (0.877)/Dam (0.977); CADD = 24/32; Ada = NA/0.997; RF = NA/0.956	19 5	Compound heterozygous	PP3, PM3, class 3 uncertain significance	—
17	M	WES	X: 79958990	*BRWD3*	c.2824A>G	p.M942V	Missense	SIFT = Del(0.01); CADD = 23.5; MTR FDR = 0.034	0	X-linked hemizygous	PP3, PM2, class 3 uncertain significance	—
18	F	WES	9: 119204816	*ASTN2*	c.3361G>A	p.V1121M	Missense	SIFT = Del(0); PolyPhen = Dam (0.961); CADD = 33	0	Unconfirmed, father unavailable	PP3, PM2, class 3 uncertain significance	—
19	M	WGS	3: 67571051 4: 68501247	*SUCLG2* *UBA6*	c.425T>C c.1766T>C	p.V142A p.L589S	Missense Missense	SIFT = Del(0); PolyPhen = PosDam (0.733); CADD = 27.1 SIFT = Del(0); PolyPhen = Dam (0.979); CADD = 27.6	1 0	Unconfirmed, father unavailable Unconfirmed, father unavailable	PP3, PP4, class 3 uncertain significance PP3, PM2, class 3 uncertain significance	—
20	F	WGS	7: 99627930	*ZKSCAN1*	c.731A>G	p.Q244R	Missense	PolyPhen = PosDam (0.877); CADD = 24	0	de novo	PP3, PS2, PM2, class 4 likely pathogenic	—

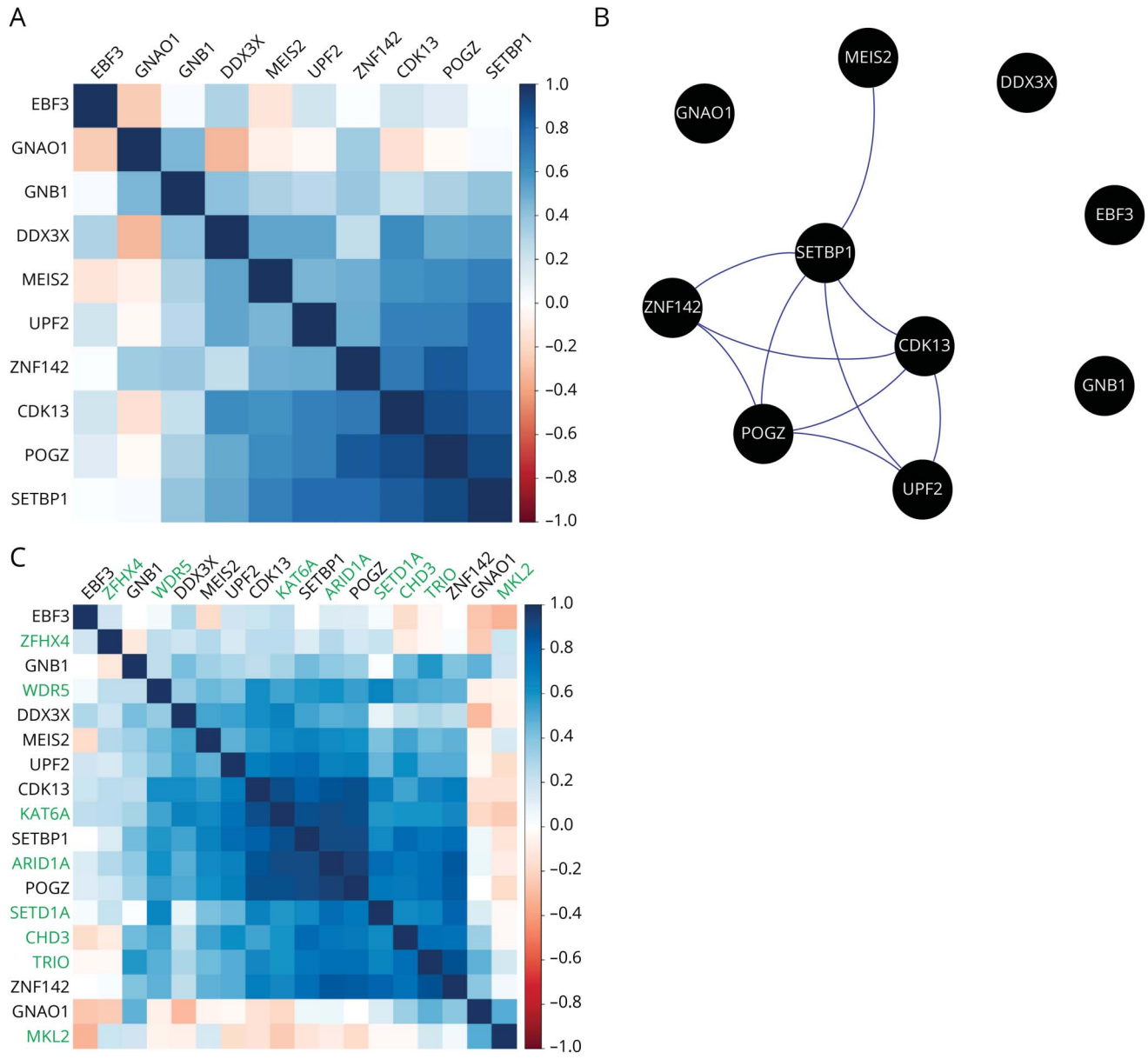
Continued

Table 3 Gene variants in childhood apraxia of speech cohort *(continued)*

	Sex	Method	Chr:Pos	Gene	DNA variant	Protein change	Effect	In silico predictions[a]	gnomAD count[b]	Inheritance	ACMG score	Reference
Predicted LoF variants classified as likely pathogenic, or with uncertain significance (ACMG guidelines)												
12	M	WES	21: 46309189	*ITGB2*	c.1877+2T>C	NA	splice donor	LoFtool = 0.0333; CADD = 25.6; Ada = 0.999; RF = 0.652	0	Inherited from affected father	PP1, PM2, class 3 uncertain significance	—
13	M	WES	2: 69734646	*AAK1*	c.2071G>T	p.E691*	Nonsense	ExACpLI = 1; CADD = 38	0	de novo	PS2, PM2, PM4, class 4 likely pathogenic	—
14	M	WES	13: 52532497	*ATP7B*	c.2304dupG	p.M769Hfs*26	frameshift	LoFtool = 0.034; CADD = 34	32	Inherited from affected father	PP1, PM4, class 3 uncertain significance	—
			10: 121602918	*MCMBP*	c.847delG	p.D283Ifs*21	frameshift	ExACpLI = 1	0	Inherited from affected father	PP1, PM2, PM4, class 3 uncertain significance	—
22	M	WES	X: 71855117	*PHKA1*	c.1601delT	p.L534Rfs*5	frameshift	LoFtool = 0.0318	0	X-linked hemizygous	PM2, PM4, class 3 uncertain significance	—

Abbreviations: ACMG = American College of Medical Genetics; Ada = AdaBoost; CADD = combined annotation dependent depletion; Chr = chromosome; CMA = chromosomal microarray; Dam = damaging; Del = deleterious; ExACpLI = Exome Aggregation Consortium probability of intolerance to loss of function; FDR = false discovery rate; LoF = loss of function; LOH = loss of heterozygosity; MTR = missense tolerance ratio; NA = not applicable; Pos = position; PosDam = possibly damaging; RF = random forest; SIFT = sorting intolerant from tolerant; WES = whole exome sequencing; WGS = whole genome sequencing.
All coordinates correspond to the *Homo sapiens* (human) genome assembly GRCh37 (hg19) from Genome Reference Consortium. All variants were confirmed by Sanger sequencing. Only 22 reported here as no other variants met criterion for remaining probands in cohort.
[a] In silico pathogenicity predictions reported only if in support of pathogenicity: for SIFT, score <0.05 reported; PolyPhen-2, score >0.15 reported; CADD Phred-scaled score, ≥20 reported; MTR FDR, <0.05 reported; Ada (AdaBoost prediction for effect on splicing), score ≥0.6 reported; RF (random forest algorithm for effect on splicing), score ≥0.6 reported; ExACpLI, score >0.9 reported; LoFTool, score <0.1 reported.
[b] Number of alleles for variant from gnomAD.
[c] Published with additional families described by collaborators.

Figure 4 Gene regulation network for speech development



(A) Gene coexpression matrix for the 10 high-confidence candidate genes. Pairwise Spearman correlations between genes shown, based on 280 samples from 24 individuals (8 weeks postconception to 10 months after birth) from the BrainSpan resource. Genes ordered by hierarchical clustering, using the median linkage method. (B) Network of gene coexpression. Nodes represent genes; edges represent gene–pair correlations that exceed the threshold for the top 5% most highly correlated gene pairs genome-wide ($|\rho| > 0.64$). (C) Gene coexpression matrix for the 10 high-confidence candidate genes (black) and the Eising et al.[8] genes (green).

higher than expected (median $|\rho| = 0.463$, $p = 2 \times 10^{-4}$; figure e-3, doi.org/10.5061/dryad.zkh189363), giving evidence of even better capture of our hypothesized biological network/pathway, and providing the first evidence of validation of the Eising et al. results.[8]

Gene set enrichment analyses of our 10 novel genes highlighted that there was an overrepresentation of genes (*CDK13*, *DDX3X*, *EBF3*, *MEIS2*, *POGZ*, *SETBP1*, *UPF2*, and *ZNF142*) involved in DNA binding (GO:0003677; table e-6 and figure e-4a, doi.org/10.5061/dryad.zkh189363). The remaining 2 genes (*GNAO1* and *GNB1*) are part of the heterotrimeric G-protein complex (GO:0005834; table e-6 and figure e-4b, doi.org/10.5061/dryad.zkh189363).

Discussion

We describe the molecular genetic architecture of CAS, a rare and debilitating disorder, in the largest cohort of children studied to date. We identified pathogenic variants in one third (11/34) of the cohort, newly implicating 9 genes (*CDK13*, *EBF3*, *GNAO1*, *GNB1*, *DDX3X*, *MEIS2*, *POGZ*, *UPF2*, *ZNF142*) and providing the first confirmation of the 10th (*SETBP1*).[8] We expand the phenotypic spectra for these genes, to include speech difficulties in the absence of, or with mild,

intellectual disability. All except *ZNF142* have been previously reported with more severe phenotypes of syndromic or nonsyndromic intellectual disability (*CDK13*,[20] *DDX3X*,[21] *EBF3*,[22] *GNB1*[23], *GNAO1*,[24] *MEIS2*,[25] *POGZ*,[26] *SETBP1*,[27] *UPF2*[28]). Broad speech and language deficits were noted, but not precisely phenotyped, in these single gene studies. A further 2 genes (*CHD1*, *NR2F1*), located within a contiguous gene deletion at 5q14.3-21.1 that includes 18 genes, are also potential candidates. *CHD1* has been linked to CAS in a previous report, and is part of a gene family of chromatin remodelers linked to neurodevelopmental disorders (e.g., *CHD2*, *CHD3*, and *CHD8*),[29] while *NR2F1* is associated with an optic atrophy and intellectual disability syndrome for which a variety of speech and language phenotypes (e.g., speech delay, expressive language deficits) have been described.[30]

Our gene set enrichment analyses show that 8 of these 10 genes code for DNA binding proteins and play a role in transcriptional regulation. Using RNA-seq data from the brain, we empirically determined that these same 8 genes are also strongly coexpressed in the developing brain, across multiple brain regions. Furthermore, we found evidence of coexpression between the candidate genes reported here and genes previously implicated in CAS by the Eising et al.[8] study. These findings suggest there is at least 1 distinct network of coexpressed genes emerging from molecular screening of CAS, characterized by similar function and patterns of expression in the brain. Similar observations of gene coexpression networks have been made for other disorders, such as the epileptic encephalopathies,[31] leading to identification and then validation of candidate genes. This approach may also be productive to identify molecular determinants for CAS in future studies. Understanding why and how mutations of genes in this network result in CAS requires in vitro and in vivo functional studies.

Beyond our 10 high-confidence candidate genes, variants of unknown significance were identified in a further 10 genes (table 3, b and c). *ASTN2* (MIM:612856), *BRWD3* (MIM: 300553), *GRIN2A* (MIM:138253), *KIRREL3* (MIM:607761), and *PTBP2* (MIM:608449) have been implicated in neurodevelopmental disorders.[5,32–35] Our remaining variants of unknown significance occur in genes associated with brain development and dysfunction. The protein encoded by *TENM4* (MIM:610084) plays a role in establishing neuronal connectivity during development, and mutations cause essential tremor.[36] *ZKSCAN1* (MIM:601260) encodes a transcription factor that regulates expression of the $GABA_A$ receptor GABRB3 subunit essential for fast inhibitory neurotransmission in brain. *AAK1* (MIM: 616405) has established roles in dendritic arborization and spine development. *PHKA1* (MIM: 311870) causes glycogen storage disease type IX (MIM: 300559), an X-linked recessive metabolic disorder characterized by exercise-induced muscle weakness. Homozygous mutations in *ATP7B* (MIM: 606882) cause Wilson disease (MIM: 277900), a disorder characterized by excess storage of intracellular hepatic copper and neurologic abnormalities; however, these patients usually present in adolescence or later.

These disparate protein functions highlight the challenges associated with determining the significance of gene variants discovered in genome-wide screens of large cohorts, particularly for neurodevelopmental speech and language disorders,[8] as is it well known that benign variants will also be found. Many were missense variants; definitively determining the pathogenicity of this variant class is often challenging. In interpreting their significance, we applied the convention of using the ACMG guidelines[14]; however, these guidelines are more difficult to apply to genes for a novel phenotype that has not yet been studied extensively with next-generation sequencing, and they may be too conservative. Ongoing observations of phenotype–genotype correlations will be critical to determining the relevance of each variant, together with large curated databases of clinical and molecular information.

In this comprehensively phenotyped cohort of children with CAS, we describe a range of co-occurring neurodevelopmental features (figure 1 and tables 1 and 2). Feeding challenges were common in the early years and the trajectory of speech development was delayed and aberrant, consistent with previous reports.[9] Our data support the concept that CAS is often part of a more wide-ranging neurodevelopmental disorder, rather than isolated speech impairment.[3,8] All probands had additional deficits that could involve a range of domains, including motor skills, cognition, attention, behavior, emotional regulation, toileting, and social skills. There were no obvious differences between the phenotypes of children with solved molecular genetic diagnoses compared with those with uncertain or no genetic findings.

A novel finding was the high rate of co-occurrence of delays in fine and gross motor skills in our CAS cohort. Children had challenges with learning specific motor skills beyond speech, such as riding a bicycle or learning to write. Gross and fine motor skills resolved earlier than the persisting speech deficits, and only 2 children had formal diagnoses of motor dyspraxia or DCD. Deficits in implicit motor learning (procedural learning) have long been proposed as a potential root cause for CAS[37] and other specific speech or language deficits.[38] In CAS, the procedural deficit hypothesis proposes that children fail to automatize the ability to sequence sounds into words and words into phrases with little cognitive effort.[37] Further to motor planning and programming deficits, co-occurring neuromuscular tone involvement was seen in some children, and ataxia in one, suggesting additional cerebellar or other common motor pathway deficits for at least one subgroup. Whereas there is increasing evidence linking motor ability with speech outcomes,[39] whether motor skills are causative for or simply correlate with speech outcomes is yet to be elucidated. Attention issues were also noted in 8 probands and 1 child had Tourette syndrome; these conditions have also been linked to the procedural learning hypothesis. A number of children had cognitive involvement, with more generalized learning deficits beyond implicit learning. As acknowledged earlier, many of the genes identified here have been linked to intellectual disability (ID) or other health and medical conditions, including epilepsy and autism, and as such, these comorbidities could play a role in the etiology of CAS.[40]

Not all children with epilepsy, ID, autism, ADHD, or DCD present with CAS, but we posit that there are several neurobiological subtypes of CAS that are more closely correlated with some neurodevelopmental conditions than others.

We provide novel insights into the etiology of CAS. We show that CAS is highly genetically heterogeneous, often occurring as a sporadic monogenic disorder. Inheritance is most frequently de novo dominant, although recessive and mosaic variants can also arise. One-third of patients have pathogenic variants, implicating shared pathways in transcriptional regulation. These findings highlight the key role of transcriptional regulation in normal speech development.

Acknowledgment

The authors thank the families for participating in this study and Tim Green (Epilepsy Research Centre) for performing DNA extractions.

Study funding

M.S.H., F.J.L., S.E.F., A.C., S.R., D.J.A., I.E.S., M.B., and A.T.M. are funded by a National Health and Medical Research Council (NHMRC) Centre of Research Excellence Grant (1116976). M.S.H., F.J.L., A.C., I.E.S., M.B., and A.T.M are supported by an Australian Research Council (ARC) Discovery Project (DP120100285). M.S.H., F.J.L., S.E.F., S.R., D.J.A., and A.T.M. are funded by an NHMRC Project Grant (1127144). M.S.H., A.T.M., I.E.S., and M.B. are supported by the March of Dimes Grant Scheme. M.S.H. is funded by an NHMRC Career Development Fellowship (ID: 1063799). S.E.F. is funded by the Max Planck Society. I.E.S. is funded by an NHMRC Development Grant (1153614) and a Practitioner Fellowship (1006110). M.B. is funded by an NHMRC Senior Research Fellowship (ID: 1102971). A.T.M. is funded by an NHMRC Development Grant (1153614) and a Practitioner Fellowship (1105008).

Disclosure

M.S. Hildebrand, V.E. Jackson, T.S. Scerri, O. Van Reyk, M. Coleman, R.O. Braden, S. Turner, K.A. Rigbye, A. Boys, S. Barton, R. Webster, M. Fahey, K. Saunders, B. Parry-Fielder, G. Paxton, M. Hayman, D. Coman, H. Goel, A. Baxter, A. Ma, N. Davis, S. Reilly, M. Delatycki, F.J. Liegeois, A. Connelly, J. Gecz, S.E. Fisher, D.J. Amor, M. Bahlo, and A.T. Morgan report no relevant disclosures. I.E. Scheffer has served on scientific advisory boards for UCB, Eisai, GlaxoSmithKline, BioMarin, Nutricia, and Xenon Pharmaceuticals; editorial boards of the *Annals of Neurology*, *Neurology®*, and *Epileptic Disorders*; may accrue future revenue on pending patent WO61/010176 (filed 2008): Therapeutic Compound; has received speaker honoraria from GlaxoSmithKline, Athena Diagnostics, UCB, BioMarin, and Eisai; and has received funding for travel from Athena Diagnostics, UCB, Biocodex, GlaxoSmithKline, Biomarin, and Eisai. Go to Neurology.org/N for full disclosures.

Publication history

Received by *Neurology* July 30, 2019. Accepted in final form December 13, 2019.

Appendix Authors

Name	Location	Contribution
Michael S. Hildebrand, PhD	The University of Melbourne, Australia	Designed and conceptualized study, directed project, generated data, interpreted data, analyzed data, wrote manuscript
Victoria E. Jackson, PhD	The Walter and Eliza Hall Institute of Medical Research, Australia	Generated data, analyzed data, interpreted data, wrote manuscript
Thomas S. Scerri, PhD	The Walter and Eliza Hall Institute of Medical Research, Australia	Generated data, analyzed data, interpreted data, wrote manuscript
Olivia Van Reyk, MSpPath	Murdoch Children's Research Institute, Australia	Generated data, analyzed data
Matthew Coleman, BSc (Hons)	The University of Melbourne, Australia	Generated data, analyzed data
Ruth O. Braden, MSpPath	Murdoch Children's Research Institute, Australia	Generated data, analyzed data
Samantha Turner, PhD	Murdoch Children's Research Institute, Australia	Generated data, analyzed data
Kristin A. Rigbye, BSc (Hons)	The University of Melbourne, Australia	Generated data, analyzed data
Amber Boys, BSc (Hons)	Victorian Clinical Genetics Services, Australia	Generated data, analyzed data
Sarah Barton, DPsych	Murdoch Children's Research Institute, Australia	Generated data, analyzed data
Richard Webster, MD	The Children's Hospital Westmead, Australia	Generated data, analyzed data
Michael Fahey, MD, PhD	Monash University, Australia	Generated data, analyzed data
Kerryn Saunders, MD	Monash University, Australia	Generated data, analyzed data
Bronwyn Parry-Fielder, BAppSci	The Royal Children's Hospital, Australia	Generated data, analyzed data
Georgia Paxton, MD	The Royal Children's Hospital, Australia	Generated data, analyzed data
Michael Hayman, MD	The Royal Children's Hospital, Australia	Generated data, analyzed data
David Coman, MD	The Wesley Hospital, Australia	Generated data, analyzed data
Himanshu Goel, MD	John Hunter Hospital, Australia	Generated data, analyzed data
Anne Baxter, MD	John Hunter Hospital, Australia	Generated data, analyzed data
Alan Ma, MD	The Children's Hospital Westmead, Australia	Generated data, analyzed data

Appendix *(continued)*

Name	Location	Contribution
Noni Davis, MD	Melbourne Children's Clinic, Australia	Generated data, analyzed data
Sheena Reilly, PhD	Griffith University, Australia	Generated data, analyzed data
Martin Delatycki, MBBS, PhD	Victorian Clinical Genetics Services, Australia	Generated data, analyzed data
Frederique J. Liégeois, PhD	UCL Great Ormond Street Institute of Child Health, UK	Generated data, analyzed data
Alan Connelly, PhD	Florey Institute of Neuroscience and Mental Health, Australia	Generated data, analyzed data
Jozef Gecz, PhD	University of Adelaide, Australia	Generated data, analyzed data
Simon E. Fisher, PhD	Max Planck Institute for Psycholinguistics, the Netherlands	Generated data, analyzed data
David J. Amor, MBBS, PhD	Murdoch Children's Research Institute, Australia	Generated data, analyzed data
Ingrid E. Scheffer, MBBS, PhD	The University of Melbourne, Australia	Designed and conceptualized study, directed project, wrote manuscript
Melanie Bahlo, PhD	The Walter and Eliza Hall Institute of Medical Research, Australia	Designed and conceptualized study, directed project, wrote manuscript
Angela T. Morgan, PhD	Murdoch Children's Research Institute, Australia	Designed and conceptualized study, directed project, generated data, analyzed data, interpreted data, wrote manuscript

References

1. Reilly S, McKean C, Morgan A, Wake M. Identifying and managing common childhood language and speech impairments. BMJ 2015;350:h2318.
2. Morgan A, Ttofari Eecen K, Pezic A, et al. Who to refer for speech therapy at 4 years of age versus who to "watch and wait"? J Pediatr 2017;185:200–204.
3. Morgan AT, Webster R. Aetiology of childhood apraxia of speech: a clinical practice update for paediatricians. J Paediatr Child Health 2018;54:1090–1095.
4. Graham SA, Fisher SE. Understanding language from a genomic perspective. Annu Rev Genet 2015;49:131–160.
5. Carvill GL, Regan BM, Yendle SC, et al. GRIN2A mutations cause epilepsy-aphasia spectrum disorders. Nat Genet 2013;45:1073–1076.
6. Morgan AT, van Haaften L, van Hulst K, et al. Early speech development in Koolen de Vries syndrome limited by oral praxis and hypotonia. Eur J Hum Genet 2018;26:75–84.
7. Mei C, Fedorenko E, Amor DJ, et al. Deep phenotyping of speech and language skills in individuals with 16p11.2 deletion. Eur J Hum Genet 2018;26:676–686.
8. 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;24:1065–1078.
9. Morgan AT, Fisher SE, Scheffer IE, Hildebrand MS. FOXP2-related speech and language disorders. In: Pagon RA, Adam MP, Ardinger HH, et al, eds. GeneReviews. Seattle: University of Washington; 2017.
10. Dodd B, Hua Z, Crosbie S, Holm A, Ozanne A. Diagnostic Evaluation of Articulation & Phonology (DEAP). London: Pearson Assessment; 2002.
11. Gozzard H, Baker E, McCabe P. Requests for clarification and children's speech responses: changing "pasghetti" to "spaghetti." Child Lang Teach Ther 2008;24:249–263.
12. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–1760.
13. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303.
14. Richards S, Aziz N, Bale S, 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;17:405–423.
15. Miller JA, Ding SL, Sunkin SM, et al. Transcriptional landscape of the prenatal human brain. Nature 2014;508:199.
16. Epskamp S, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network visualizations of relationships in psychometric data. J Stat Softw 2012;48:1–18.
17. Reimand J, Isserlin R, Voisin V, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc 2019;14:482–517.
18. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.
19. Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. Nucleic Acids Res 2017;46:D649–D655.
20. Sifrim A, Hitz MP, Wilsdon A, et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet 2016;48:1060–1065.
21. Beal B, Hayes I, McGaughran J, et al. Expansion of phenotype of DDX3X syndrome: six new cases. Clin Dysmorphol 2019;28:169–174.
22. Harms FL, Girisha KM, Hardigan AA, et al. Mutations in EBF3 disturb transcriptional profiles and cause intellectual disability, ataxia, and facial dysmorphism. Am J Hum Genet 2017;100:117–127.
23. Petrovski S, Kury S, Myers CT, et al. Germline de novo mutations in GNB1 cause severe neurodevelopmental disability, hypotonia, and seizures. Am J Hum Genet 2016;98:1001–1010.
24. Nakamura K, Kodera H, Akita T, et al. De Novo mutations in GNAO1, encoding a Galphao subunit of heterotrimeric G proteins, cause epileptic encephalopathy. Am J Hum Genet 2013;93:496–505.
25. Verheije R, Kupchik GS, Isidor B, et al. Heterozygous loss-of-function variants of MEIS2 cause a triad of palatal defects, congenital heart defects, and intellectual disability. Eur J Hum Genet 2019;27:278–290.
26. Stessman HAF, Willemsen MH, Fenckova M, et al. Disruption of POGZ is associated with intellectual disability and autism spectrum disorders. Am J Hum Genet 2016;98:541–552.
27. Hoischen A, van Bon BW, Gilissen C, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. Nat Genet 2010;42:483–485.
28. Johnson JL, Stoica L, Liu Y, et al. Inhibition of Upf2-dependent nonsense-mediated decay leads to behavioral and neurophysiological abnormalities by activating the immune response. Neuron 2019;104:665–679.e8.
29. Lamar KMJ, Carvill GL. Chromatin remodeling proteins in epilepsy: lessons from CHD2-associated epilepsy. Front Mol Neurosci 2018;11:208.
30. Bosch DG, Boonstra FN, Gonzaga-Jauregui C, et al. NR2F1 mutations cause optic atrophy with intellectual disability. Am J Hum Genet 2014;94:303–309.
31. Oliver KL, Lukic V, Thorne NP, Berkovic SF, Scheffer IE, Bahlo M. Harnessing gene expression networks to prioritize candidate epileptic encephalopathy genes. PLoS One 2014;9:1–11.
32. Doan RN, Bae BI, Cubelos B, et al. Mutations in human accelerated regions disrupt cognition and social behavior. Cell 2016;167:341–354.
33. Field M, Tarpey PS, Smith R, et al. Mutations in the BRWD3 gene cause X-linked mental retardation associated with macrocephaly. Am J Hum Genet 2007;81:367–374.
34. Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 2009;459:569–573.
35. Quintela I, Barros F, Fernandez-Prieto M, et al. Interstitial microdeletions including the chromosome band 4q13.2 and the UBA6 gene as possible causes of intellectual disability and behavior disorder. Am J Med Genet A 2015;167A:3113–3120.
36. Hor H, Francescatto L, Bartesaghi L, et al. Missense mutations in TENM4, a regulator of axon guidance and central myelination, cause essential tremor. Hum Mol Genet 2015;24:5677–5686.
37. Vargha-Khadem F, Gadian DG, Copp A, Mishkin M. FOXP2 and the neuroanatomy of speech and language. Nat Rev Neurosci 2005;6:131–138.
38. Ullman MT, Pierpont EI. Specific language impairment is not specific to language: the procedural deficit hypothesis. Cortex 2005;41:399–433.
39. Visscher C, Houwen S, Scherder EJ, Moolenaar B, Hartman E. Motor profile of children with developmental speech and language disorders. Pediatrics 2007;120:e158–e163.
40. Khan K, Zech M, Morgan AT, et al. Recessive variants in ZNF142 cause a complex neurodevelopmental disorder with intellectual disability, speech impairment, seizures, and dystonia. Genet Med 2019;21:2532–2542.

Severe childhood speech disorder: Gene discovery highlights transcriptional dysregulation

Michael S. Hildebrand, Victoria E. Jackson, Thomas S. Scerri, et al.

Neurology published online April 28, 2020
DOI 10.1212/WNL.0000000000009441