

4-27-2016

# Genetic Candidate Variants in Two Multigenerational Families with Childhood Apraxia of Speech

Peter Beate

Ellen M. Wijsman

Alejandro Q. Nato Jr.

Marshall University, nato@marshall.edu

University of Washington Center for Mendelian Genomics

Mark M. Matsushita

*See next page for additional authors*

Follow this and additional works at: [https://mds.marshall.edu/sm\\_bm](https://mds.marshall.edu/sm_bm)

 Part of the [Genetic Phenomena Commons](#), [Genetic Processes Commons](#), [Genetic Structures Commons](#), [Medical Genetics Commons](#), and the [Speech Pathology and Audiology Commons](#)

---

## Recommended Citation

Peter B, Wijsman EM, Nato AQ Jr, University of Washington Center for Mendelian Genomics, Matsushita MM, Chapman KL, et al. (2016) Genetic Candidate Variants in Two Multigenerational Families with Childhood Apraxia of Speech. PLoS ONE 11(4): e0153864. <https://doi.org/10.1371/journal.pone.0153864>

This Article is brought to you for free and open access by the Faculty Research at Marshall Digital Scholar. It has been accepted for inclusion in Biochemistry and Microbiology by an authorized administrator of Marshall Digital Scholar. For more information, please contact [zhangj@marshall.edu](mailto:zhangj@marshall.edu), [martj@marshall.edu](mailto:martj@marshall.edu).

---

**Authors**

Peter Beate, Ellen M. Wijsman, Alejandro Q. Nato Jr., University of Washington Center for Mendelian Genomics, Mark M. Matsushita, Kathy L. Chapman, Ian B. Stanaway, John Wolff, Kaori Oda, Virginia B. Gabo, and Wendy H. Raskind

RESEARCH ARTICLE

# Genetic Candidate Variants in Two Multigenerational Families with Childhood Apraxia of Speech

Beate Peter<sup>1,2,3\*</sup>, Ellen M. Wijsman<sup>4,5</sup>, Alejandro Q. Nato, Jr.<sup>4</sup>, University of Washington Center for Mendelian Genomics<sup>1</sup>, Mark M. Matsushita<sup>4</sup>, Kathy L. Chapman<sup>6</sup>, Ian B. Stanaway<sup>7</sup>, John Wolff<sup>4</sup>, Kaori Oda<sup>4</sup>, Virginia B. Gabo<sup>4</sup>, Wendy H. Raskind<sup>4,8</sup>

**1** Dpt. of Speech and Hearing Sciences, Arizona State University, Tempe, Arizona, United States of America, **2** Dpt. of Communication Sciences and Disorders, Saint Louis University, Saint Louis, Missouri, United States of America, **3** Dpt. of Speech and Hearing Sciences, University of Washington, Seattle, Washington, United States of America, **4** Div. of Medical Genetics, Dpt. of Medicine, University of Washington, Seattle, Washington, United States of America, **5** Dpt. of Biostatistics, University of Washington, Seattle, Washington, United States of America, **6** Dpt. of Communication Sciences and Disorders, University of Utah, Salt Lake City, Utah, United States of America, **7** Dpt. of Environmental and Occupational Health, Environmental Toxicology, University of Washington, Seattle, Washington, United States of America, **8** Dpt. of Psychiatry and Behavioral Sciences, University of Washington, Seattle, Washington, United States of America

¶ Membership of the University of Washington Center for Mendelian Genomics is provided in the Acknowledgments.

\* [Beate.Peter@asu.edu](mailto:Beate.Peter@asu.edu)



**OPEN ACCESS**

**Citation:** Peter B, Wijsman EM, Nato AQ, Jr., University of Washington Center for Mendelian Genomics, Matsushita MM, Chapman KL, et al. (2016) Genetic Candidate Variants in Two Multigenerational Families with Childhood Apraxia of Speech. PLoS ONE 11(4): e0153864. doi:10.1371/journal.pone.0153864

**Editor:** Tao Cai, NIDCR/NIH, UNITED STATES

**Received:** November 19, 2015

**Accepted:** April 5, 2016

**Published:** April 27, 2016

**Copyright:** © 2016 Peter et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The genomic data were obtained from the University of Washington Centers for Mendelian Genomics and are in the process of being submitted to dbGaP. The accession number is NHGRI University of Washington Center for Mendelian Genomics Study phs000693.

**Funding:** The following funding sources are gratefully acknowledged: American Speech-Language-Hearing Foundation New Century Scholars Research Grant (B. Peter), NIDCD T32DC00033 (B. Peter), NIDCD R03DC010886 (B. Peter), R01HD054562 (W. H. Raskind), and

## Abstract

Childhood apraxia of speech (CAS) is a severe and socially debilitating form of speech sound disorder with suspected genetic involvement, but the genetic etiology is not yet well understood. Very few known or putative causal genes have been identified to date, e.g., *FOXP2* and *BCL11A*. Building a knowledge base of the genetic etiology of CAS will make it possible to identify infants at genetic risk and motivate the development of effective very early intervention programs. We investigated the genetic etiology of CAS in two large multigenerational families with familial CAS. Complementary genomic methods included Markov chain Monte Carlo linkage analysis, copy-number analysis, identity-by-descent sharing, and exome sequencing with variant filtering. No overlaps in regions with positive evidence of linkage between the two families were found. In one family, linkage analysis detected two chromosomal regions of interest, 5p15.1-p14.1, and 17p13.1-q11.1, inherited separately from the two founders. Single-point linkage analysis of selected variants identified *CDH18* as a primary gene of interest and additionally, *MYO10*, *NIPBL*, *GLP2R*, *NCOR1*, *FLCN*, *SMCR8*, *NEK8*, and *ANKRD12*, possibly with additive effects. Linkage analysis in the second family detected five regions with LOD scores approaching the highest values possible in the family. A gene of interest was *C4orf21* (*ZGRF1*) on 4q25-q28.2. Evidence for previously described causal copy-number variations and validated or suspected genes was not found. Results are consistent with a heterogeneous CAS etiology, as is expected in many neurogenic disorders. Future studies will investigate genome variants in these and other families with CAS.

R01MH094293 (E. Wijsman). SNP genotyping and exome sequencing were provided by the University of Washington Center for Mendelian Genomics (UW CMG) and was funded by the National Human Genome Research Institute and NIH grant 1U54HG006493 to the three UW CMG members, Drs. M. Bamshad, D. Nickerson and J. Shendure. I. Stanaway was supported under grants from the Environmental Health Sciences of the National Institutes of Health (T32ES015459 and 5P01 ES009601) and the EPA (RD-83451401). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Introduction

Children with speech sound disorder (SSD) fall behind their typically developing peers in acquiring speech that is easily understood by others. As extensively reviewed in the literature, signs and symptoms of SSD include distortions, substitutions, omissions, insertions, errors on the syllable or word level, and prosodic errors affecting rhythm and intonation [1–3]. Children with disordered speech have difficulty expressing their thoughts in ways that are easily understood by others [4] and experience negative perceptions on the part of their peers because of their speech differences [5]. Several SSD subtypes have been proposed. One of these is childhood apraxia of speech (CAS), defined as a motor planning or programming disorder affecting the speech production system. The American Speech-Language-Hearing Association (ASHA) issued a position statement regarding CAS (<http://www.asha.org/docs/html/PS2007-00277.html>) with the following phenotype definition, implicating the central nervous system as the most likely locus of impairment:

Childhood apraxia of speech (CAS) is a neurological childhood (pediatric) speech sound disorder in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits (e.g., abnormal reflexes, abnormal tone). CAS may occur as a result of known neurological impairment, in association with complex neurobehavioral disorders of known or unknown origin, or as an idiopathic neurogenic speech sound disorder. The core impairment in planning and/or programming spatiotemporal parameters of movement sequences results in errors in speech sound production and prosody.

The speech of children with CAS may be characterized by some errors commonly seen in children with other forms of SSD but additionally, by unusual errors such as vowel distortions, difficulty initiating or transitioning between articulatory gestures, lack of differentiation between stressed and unstressed syllables or mis-stressing syllables, distorted substitutions, syllable segregation (resulting in a staccato-like rhythm), schwa insertions, voicing errors, slow rate, slow diadochokinetic rates, and/or increased difficulty with multisyllabic words [6]. Compared to other subtypes of SSD, CAS is considered to be more severe, requiring intense and specialized treatment [7, 8]. Children with CAS are at increased risk for reading/spelling disorders [9–11]. According to one estimate, CAS is diagnosed in .01% to .02% of children in the United States [12].

Disordered speech consistent with CAS can be part of syndromes of genetic etiology. In one large multigenerational family referred to as the KE family, disruptions in the *FOXP2* gene (OMIM #605317) on chromosome (chr) 7 caused a severe speech disorder in the presence of nonverbal oral dyspraxia and disordered language [13–15]. Structural and functional brain changes were observed as well, characterized by reduced grey matter density in the caudate nucleus, cerebellum, and inferior frontal gyrus [16] and reduced activation during a nonword repetition task in the premotor, supplementary, and primary motor cortices and in the cerebellum and basal ganglia [17]. A functionally related gene, *CNTNAP2*, plays a role in language [18] and reading [19] ability. Approximately 18% of children with galactosemia (OMIM #230400, OMIM #606999), a metabolic disease caused by mutations in the *GALT* gene (OMIM #606999) on 9p13.3 [20], exhibit signs of CAS [21]. Variants in the *ELP4* (OMIM #606985) and *PAX6* (OMIM #607108) genes on 11p13 have been associated with Rolandic epilepsy, which is frequently accompanied by disordered speech consistent with CAS [22]. Duplications of a region on 7q11.23 are associated with developmental delays, characteristic facial anomalies, social anxieties, and severe delays in language and speech abilities, the latter consistent with CAS [23, 24]. In five of nine individuals with subtelomeric or interstitial 12p13.33 deletions and speech delays, the speech phenotype was consistent with CAS [25]. In a child with a severe speech disorder characterized by apraxic traits as well as muscle weakness, we found a

*de novo* heterozygous deletion of the *BCL11A* gene (B-cell CLL/lymphoma 11A, OMIM #606557) on chr 2 [26], located within a larger microdeletion region associated with global deficits in motor development and muscle tone as well as growth retardation, intellectual disability, absence of verbal communication, and/or craniofacial and skeletal dysmorphic features [27–34]. Our case study suggests that *BCL11A* plays a role in aspects of motor planning/programming and muscle tone required for speech.

There is evidence that in some cases, idiopathic CAS has a genetic etiology, but causal genes have not yet been validated. In studies of three individuals with CAS, evidence from duplicated or deleted DNA regions pointed to 16p11.2 as a candidate region [35, 36]. In a study of 24 unrelated children with CAS, 12 had copy-number variations (CNVs) on ten different chromosomes; findings included one 16p11.2 deletion [37]. One of the children had a *FOXP2* mutation, and three of the CNV regions contained other candidate genes. In an exome variant study in 10 unrelated individuals with CAS, variants of interest were found on chrs 3, 6, 7, 9, and 17, where some participants had more than one of the variants [38]. Potentially deleterious variants were reported in genes suspected to cause CAS (*FOXP1*, OMIM #605515, *CNTNAP2*, OMIM #604569) and genes associated with phenotypes frequently co-occurring with CAS (*ATP13A4*, OMIM #606693, *CNTNAP1*, OMIM #602346, *KIAA0319*, OMIM #609269, and *SETX*, OMIM #608465).

In two multigenerational families with familial nonsyndromic CAS, we showed that the speech phenotype was associated with oral and hand motor deficits, especially when the tasks required temporal integration of alternating-sequential movements. In the oral motor domain, diadochokinetic (DDK) rapid repetition of multisyllabic tokens (/pata/, /taka/, and /pataka/) was used to assess alternating-sequential functioning and the analogous hand task was rapidly tapping two computer keys using two fingers in an alternating fashion. Less impaired was performance on tasks requiring repetitive movement sequences in the oral domain (/pa/, /ta/, and /ka/) and single key tapping in the hand domain [39]. For a parametric genome-wide linkage analysis in one of these participating families, we used a measure of alternating-sequential DDK ability obtained by subtracting standard scores of performance on monosyllabic syllable repetition from standard scores of performance on multisyllabic syllable repetition as the input variable, with a maximum possible LOD score of 1.78. Two new regions of interest were found, one on 7q36.1-q36.3 (bp 143,723,666–159,138,663; LOD = 1.35) and one on 6p21.2-p12.3 (bp 36,632,927–64,590,642; LOD = 1.10) [40]. The 6p region overlapped with a recently identified region of interest for dyslexia [41].

Together, these findings are consistent with a heterogeneous CAS etiology. Here, we posit that speech development is complex and can be influenced by several genetic and environmental factors with varying levels of impact in the same individuals. Discovery of genetic risk factors of high impact may be more successful in families than in unrelated individuals because these factors are likely to be shared by affected members of the same family.

We recently described phenotypic aspects in a multigenerational family, here referred to as “Family A”, with familial CAS [42]. Most of the individuals with current or past CAS did not produce their first word until age 3 years whereas first words typically emerge around the first birthday. Their speech was difficult to understand by others until they reached age 5 to 7 years, a milestone typically reached by age 4 years. Most of the affected family members required a minimum of three years of speech therapy to acquire intelligible speech. Performance on tasks with high sequential processing loads including multisyllabic DDK testing [43, 44], nonword imitation [45–47], rapid automatic naming [48], nonword decoding [49, 50], and spelling [51] differentiated between family members with and without a history of CAS, whereas there were no group differences in tasks with low sequential processing loads. A qualitative analysis of errors during real word and nonword imitations showed that the adults with a history of disordered

speech produced more phoneme sequencing errors, compared to those without such a history. These findings were interpreted as consistent with a deficit in sequential processing that was not limited to motor programming but also manifested in linguistic and cognitive tasks. Results were replicated in adults from five other families with familial SSD including CAS [52].

The affection pattern in Family A is consistent with a genetic etiology of CAS. The purpose of the present study, hence, was to investigate this hypothesis using a set of complementary methods. Similar methods were used to investigate the same hypothesis in a second multigenerational family with familial CAS, here referred to as Family B, and results were compared.

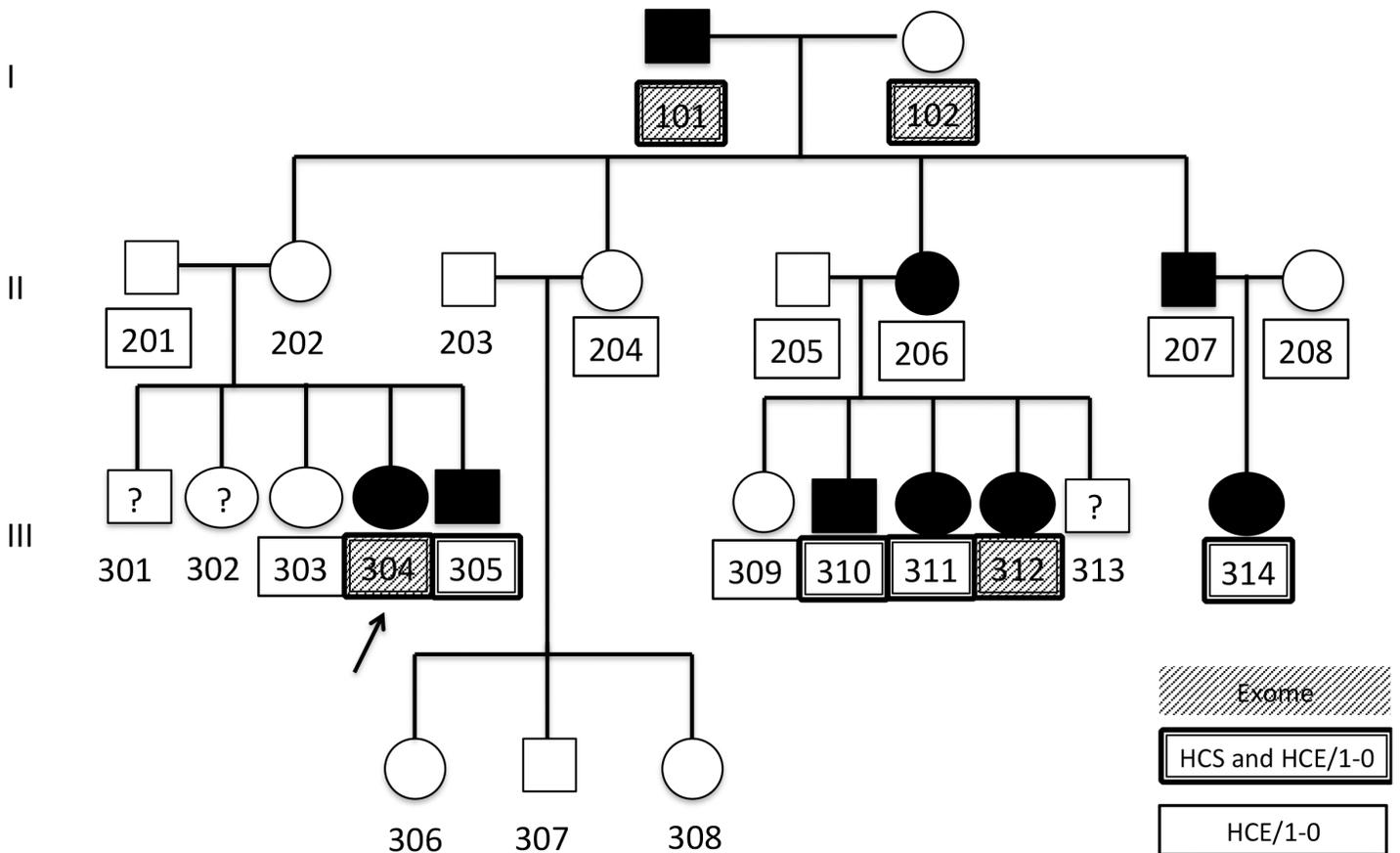
## Materials and Methods

### Participants and Behavioral Measures

This study was conducted with the approval of the University of Washington's institutional review board. Adults gave written consent, parents gave written permission for their minor children to participate in the study, and additionally, school-age children gave written assent and preschool-age children gave oral assent. Extensive family history interviews were conducted with the participating adults in each family to obtain background information regarding presence of an SSD diagnosis and history of speech therapy services for the interviewed persons themselves as well as other family members. In addition, each adult filled out a questionnaire regarding her/his educational, developmental, and health history. Parents provided details regarding the developmental history of each of their children. Copies of any available written assessment reports were obtained. Affection status was assigned based on this information and, for young children who had not yet been professionally assessed for the presence of SSD, additionally on performance on standardized and nonstandardized speech measures. In a few cases where sufficient evidence was not available, unknown affection status was assigned.

Family A consists of 24 members in three generations with a familial SSD consistent with CAS (Fig 1; note that the text refers to individual ID numbers with the family identifier as a prefix for clarity). All participants are of European descent, with a small admixture of Japanese descent in six of the participants. Phenotypes and DNA were available for two founders, four adult offspring and their spouses, and 13 grandchildren, 11 of whom could be classified with respect to CAS affection. The oldest grandchild, A-301, was unable to contribute DNA or participate in the testing; only his developmental history was available. The proband, ID A-304, age 10 years at the time of testing, had a history of severe CAS requiring intense and prolonged speech therapy. The grandfather, ID A-101, reported receiving speech services as an elementary school student whereas the grandmother, A-102, did not report receiving such services. Both grandparents reported individuals biologically related to them with difficulties in the area of speech and language acquisition. No written records were available regarding the grandparents' speech development. Two of their four participating adult offspring (A-206, A207) had received speech services for five or more years during their early elementary and middle school years. Of the 14 grandchildren, four (A-304, A-305, A-310, A-311) had previously been given a diagnosis of CAS and were currently receiving speech therapy or had completed their course of speech therapy, two (A-312, A-314) were diagnosed based on the speech testing conducted as part of this study, two (A-301, A302) had been diagnosed with a mild speech delay not consistent with CAS as preschoolers, one was too young (15 months) to be diagnosed unambiguously, and five had never received an SSD diagnosis of any type. Details regarding the behavioral findings have been reported previously [42].

Family B also has a history of familial CAS. The family consists of 39 members in five generations, all of European descent except for six individuals with an admixture of African American descent (Fig 2). DNA was available for 14 participants (B-202, B-204, B-205, B-206, B-301,

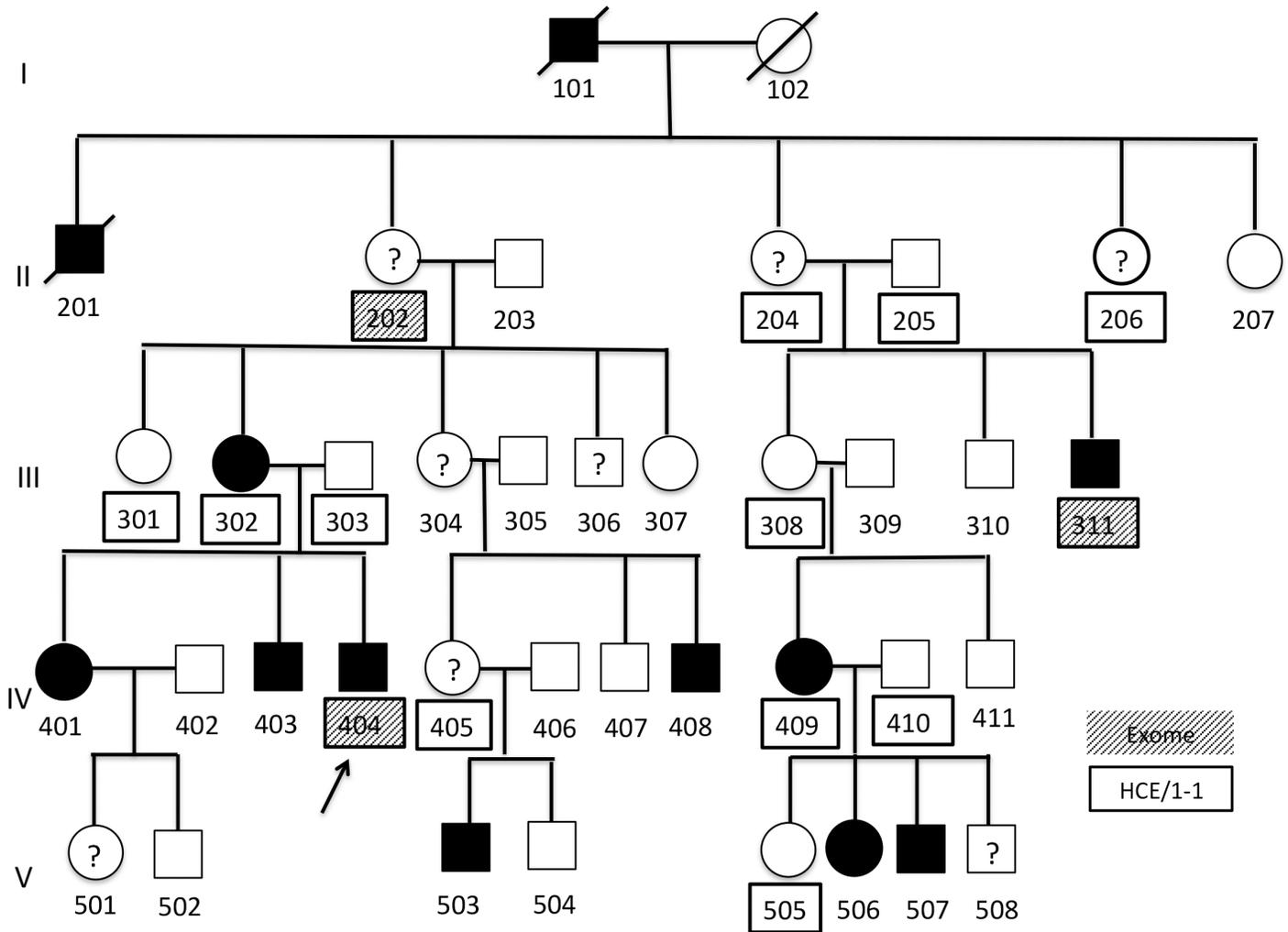


**Fig 1. Family diagram for Family A.** Square shape = male, circle shape = female, black fill = affected, white fill = unaffected, ? = affection status unknown, arrow = proband, HCS = Illumina HumanCytoSNP-12v2, HCE/1-0 = Illumina HumanCoreExome-12v1-0. B. Numbers underneath each symbol are individual IDs. Boxes around an ID identify individuals with SNP array data. Filled boxes indicate IDs that also have whole exome sequence data.

doi:10.1371/journal.pone.0153864.g001

B-302, B-303, B-308, B-311, B-404, B-405, B409, B-410, B-505). Questionnaire and interview information was available for these participants and also for B-506 and B-507. All of these participants except B-206, B-405, and B-410 participated in behavioral testing.

The proband, B-403, was 14 years old at the time of testing. He was born at term after an uncomplicated pregnancy and delivery and passed regularly scheduled health, vision, and hearing checks throughout the preschool years. He began receiving speech and expressive language services at age 2;5 due to severe delays in these areas. His diagnosis of CAS at this time was based on severely impaired articulation skills in the presence of severe oral apraxia, not further described in the assessment report. At age 3;8, speech testing with the Structured Photographic Articulation Test II (SPAT-II) [53] resulted in a standard score of 66 (population mean = 100, SD = 30; 1<sup>st</sup> percentile, far below normal limits), consistent with a severe SSD. His consonant inventory was extremely restricted, consisting of only /d, b, m, n/. Oral motor testing showed deficits in imitating tongue movements. Language testing using Clinical Evaluation of Language Fundamentals-Preschool (CELF-P) [54] showed an Auditory Comprehension standard score of 95 (37<sup>th</sup> percentile, within normal limits) and an Expressive Comprehension standard score of 50 (1<sup>st</sup> percentile, far below normal limits). At age 5;8, the proband underwent an occupational therapy evaluation that revealed severe fine motor deficits, especially in grasping and eye-hand coordination skills, qualifying him for services in this area, whereas his gross



**Fig 2. Family diagram for Family B.** Square shape = male, circle shape = female, black fill = affected, white fill = unaffected, ? = affection status unknown, arrow = proband, HCE/1-1 = Illumina HumanCoreExome-12v1-1\_B. Numbers underneath each symbol are individual IDs. Boxes around an ID identify individuals with SNP array data. Filled boxes indicate IDs that also have whole exome sequence data.

doi:10.1371/journal.pone.0153864.g002

motor development was found to be within normal limits. Upon entering school, the proband showed difficulty with reading and spelling. For instance, at age 8;10, when tested with the Woodcock-Johnson Tests of Achievement III [55], he obtained a standard score of 63 in Broad Written Language and 64 in Broad Reading (both 1<sup>st</sup> percentile). The ASHA technical report on CAS lists a small consonant inventory, poorer expressive than receptive language skills, oral apraxia, fine motor deficits, and difficulty with written language as frequently co-occurring conditions (<http://www.asha.org/docs/html/PS2007-00277.html>). Testing at age 14 showed severe difficulty with nonword imitation [45–47], especially in the form of rearranged phoneme sequences. During diadochokinetic testing, his syllable durations for the monosyllables /pa/, /ta/, and /ka/ were longer than expected for his age, indicative of slow syllable production speed ( $z = -1.11, -1.84, \text{ and } -1.13$ , respectively), but excessively long for the trisyllable /pataka/ ( $z = -4.32$ ), indicative of severe difficulties with motor planning of complex sequences. Increased difficulty with multisyllabic diadochokinetic tasks, compared to monosyllabic tasks, was reported in our previous studies of children and adults with CAS histories [39, 42, 52, 56].

A similar history of SSD and delays in written language was also reported by his mother, B-302, and another relative, B-311, whereas B-409 reported a history of SSD only in the absence of difficulties with written language, and B-405, a history of difficulties with written language in the absence of SSD. Two family members, B-101 and B-201, both deceased, were reported to have had severely disordered speech during childhood but written records were not available for them or any other members in generations I and II.

Because of her concerns that the severe speech disorder could be of genetic origin, the proband's mother had sought genetic testing for the proband and herself five years prior to participating in this study. According to the clinical report, a microarray analysis of 622 loci using 1,887 BAC clones was performed on DNA derived from peripheral blood. Two interstitial duplications, separated by a normal intervening sequence, were detected on 15q26.3 ([CTD-3210F22, RP11-947PI-631H11]x3, [RP11-262p8, RP11-654A16]x2, [RP11-20G13, CTD-3221M10]x3), summing to 2 Mb in size. The centromeric duplication contains the entire *FAM169B* gene and the telomeric one, part of the *MEF2A* gene. Fluorescence in situ hybridization (FISH) analysis using two BAC clones from the two regions (CTD-3210F22, RP11-20G13) showed a pattern consistent with duplication. The same duplication was also found in the mother's DNA using microarray analysis. The clinical significance of this abnormality could not be determined at the time of the clinical report.

Family A provided more direct phenotypic observations and fewer missing samples, compared to Family B. Therefore, the main focus of this study was placed on Family A and data from Family B were used for purposes of comparison.

## Genetic and Statistical Methods

**Overview.** Complementary genomic approaches were selected because the genetic etiologies of CAS cases in the literature to date include not only a point mutation [13, 14, 37] but also deletions and duplications [26, 35–37]. To investigate the presence of single, relatively rare alleles in the families, we conducted linkage analysis. To detect duplications and/or deletions, we performed copy-number variation (CNV) analysis. Identity-by-descent (IBD) analysis was used to investigate more common segregating variants in Family A, where the grandfather had received speech therapy during childhood but the possibility of childhood speech difficulties in the grandmother could not be ruled out completely. Whole exome sequencing (WES) followed by variant filtering was performed in both families to identify candidate variants. Because of greater statistical power to detect linkage in Family A, compared to Family B, selected candidate variants were genotyped and checked for segregation in Family A only.

**Genotyping and Sequencing.** DNA was extracted from peripheral blood using standard laboratory procedures. The samples passed quality control checks for sample swaps and incorrectly specified parentage. Because of the phenotypic overlaps with the previously described KE family where a point mutation in the *FOXP2* gene caused a severe speech and language disorder [13, 14], this gene was ruled out by exclusion mapping [57] prior to genome-wide analysis procedures.

The University of Washington (UW) Center for Mendelian Genomics (CMG) provided single nucleotide polymorphism (SNP) genotypes based on three arrays, as well as WES. In Family A, genotypes for eight participants (Fig 1) were obtained using the Illumina HumanCytoSNP-12v2 array (henceforth HCS) with 298,563 markers. Genotypes for these and eight additional participants (Fig 1) were obtained using the Illumina HumanCoreExome-12v1-0\_B array (henceforth HCE/1-0) with 538,448 markers. In Family B, all 14 available DNA samples (Fig 2) were genotyped using the Illumina HumanCoreExome-12v1-1\_B (henceforth HCE/1-1) with 542,585 markers.

In Family A, DNA samples from two cousins, ID A-304 and A-312, both with a diagnosis of CAS and highly informative based on position in the family pedigree, and samples from the two grandparents were selected for WES (Fig 1). Similarly, B-202, B-311, and B-404 were selected for WES in Family B (Fig 2). Following methods previously described in detail [58], the NimbleGen in-solution SeqCap EZ Exome Library v2.0 (Roche, Basel, Switzerland) was used to capture the exome and adjoining regions, following the manufacturer's instructions. Short-read sequencing was done on an Illumina HiSeq 2000 platform.

For Family B, to evaluate whether the previously reported duplications on 15q26.3 segregated with the disorder, one probe within each of the duplicated regions (Hs02820990\_cn, located within *FAM169B* at bp 98,981,473, and Hs01667266\_cn, located within *MEF2A* at bp 100,250,891), and two control probes (Hs03312008\_cn at bp 97,806,447 and Hs05387770\_cn at bp 101,256,220) were typed in seven strategically selected samples.

**Statistical Analyses.** Prior to the SNP-based linkage analyses in the two families, power analysis with 1,000 simulations was conducted using the SLINK package [59, 60]. Under the assumed model of autosomal dominant inheritance, there was one case of nonpenetrance in Family A (A-202) and one in Family B (B-308). As in other genome-wide family-based studies with similar mode of inheritance and evidence for reduced penetrance [61], we assumed parameters of penetrance = 0.50 in the two high-risk genotypes and 0.01 in the low-risk genotype. A simple reduced penetrance model similar to this that allows for sporadic cases works well in situations where the penetrance is unknown but incomplete, outside information to inform the parameters further is not available, and the genotype-phenotype relationship is likely to have at least some complexity [62]. In Family A, the resulting maximum log odds (LOD) score in the power analysis at  $\theta = 0$  was 2.75 with the grandfather coded as affected and the grandmother, as unaffected, and 2.45 with both grandparents coded with unknown affection status. The maximum LOD score in Family B was 2.21. Although both these maxima are below the traditional LOD score requirement of 3 for declaring strong evidence of autosomal linkage [63], this threshold was designed to be conservative, and is actually overly conservative [64, 65]. In addition, with current easy access to sequence data, the original concern about cost of follow-up no longer carries the same concern as it did when the original threshold was proposed.

The SNP markers were checked for genotyping errors using the PLINK [66] and PEDSTATS [67] packages and SNPs with genotyping errors were removed from the analysis. Files were formatted for MCMC linkage analysis and an ideal set of SNPs was chosen for a marker panel with the Pedigree-Based Analysis Pipeline (PBAP) [68], targeting marker spacing of 0.5 centimorgan (cM), minor allele frequency (MAF) > 0.2, and LD between markers < 0.04. Minor allele frequencies (MAFs) for the SNP arrays were based on the 1000 Genomes Project Europeans (<http://www.1000genomes.org>). Genetic locations (cM) were obtained from the Rutgers Maps, Build 134 [69] to establish marker order. These positions were then converted to positions based on the Haldane map function to comply with the requirements of the analysis methods. Affection status for the grandparents in Family A was conservatively set to unknown; two additional models, each with one grandparent coded as affected, were run. MCMC-based linkage analysis was conducted with the *gl\_auto* and *gl\_lods* programs of the MORGAN 3.2 package [70–72]. The *gl\_lods* program calculates LOD scores based on the phenotype information, penetrance model, and the inheritance vectors that are estimated by *gl\_auto* for each marker given the available pedigree constellation, the marker data, and the genetic map. For *gl\_auto*, the run conditions were 100,000 total run iterations, 15% burn-in iterations, and 2,000 saved iterations. Chromosomal regions retained for further analysis were required to have  $LOD_{max}$  scores > 1. The approximate 95% confidence interval (CI) about the peak was defined as the region between the boundaries about the peak where  $LOD = LOD_{max} - 1$  [73].

For CNV analyses in the two families, two sources of input were used. First, genotypes from the exomes were entered into the Copy Number Inference from Exome Reads (CoNIFER) package [74]. For CoNIFER-based CNV discovery, reads from each exome were split into up to two consecutive 36mers and mapped using the single-end mode of mrsFAST [75], then aligned to the hg19 reference genome. Reads per kilobase per million (RPKM) values were calculated and targets with a median RPKM of 1 were excluded. Standardized RPKM values were calculated and a single value decomposition (SVD) algorithm was applied. The output from this analysis, SVD-ZRPKM, was used as the normalized relative copy number of a given exon in a sample. To exclude naturally occurring regions that are duplicated or repeated in the genome, CNVs were filtered using a 50% reciprocal overlap mask. The second source of input for CNV analysis was the set of 16 Illumina HCE/1-0 (Family A) and 14 Illumina HCE/1-1 (Family B) SNP genotypes. Here, we calculated CNVs with two software packages, PennCNV [76] and cnvHap [77]. PennCNV uses a hidden Markov model (HMM) approach, incorporating several types of information including total signal intensity, allelic intensity ratio at each marker, distance between SNPs, and allele frequencies. To avoid biased results, we did not use pedigree information [78]. Like PennCNV, cnvHap uses an HMM approach but additionally incorporates chromosome-wide haplotypic information and cluster-based models of allele frequencies at each marker position. Specifically regarding the previously reported deletion regions on chrs 2 and 16 [26, 35–37], Illumina HCE/1-0 and HCE/1-1 genotypes from two affected members per family were examined for presence of heterozygous genotypes.

In Family A only, IBD analysis was performed using the HCS genotypes and the BEAGLE software package, Version 3.3.2 [79]. The SNP base calls were normalized to the forward genomic reference strand and converted to PLINK [66] format with the participants coded as unrelated. The unphased genotypes of 165 unrelated HapMap3 Caucasians (CEU) were merged by PLINK with the genotypes of the eight participants. Duplicated SNPs and SNPs with inconsistent locations were deleted. The genotypes of the participants and the HapMap3 Caucasians were phased as unrelated subjects in BEAGLE. The fastIBD routine of BEAGLE was then used to estimate the shared haplotype frequencies among all pairs, inputting default parameters. Ten haplotype pairs were sampled for each participant during each iteration of phasing. Very rare shared haplotypes between pairs (a threshold of a fastIBD score of  $1.0e-10$ ) are likely to be identical by descent. The results of ten independent FastIBD analyses were combined. Exclusive regions of haplotype sharing unique to affected participants were compared to the results from linkage analyses. Specifically, a region shared exclusively by the six affected grandchildren selected for SNP typing was required to be shared in all 15 pairwise comparisons. To determine IBD sharing with one of the grandparents, the region in question was required to be shared by the grandparent and all six selected grandchildren.

Selected variants in Family A were tested for segregation using single-marker parametric linkage analysis based on the same parameters as the genome-wide multipoint linkage analysis, here using MERLIN [80] with customized bit size to accommodate the pedigree size. This step was repeated for two additional models, one with the grandfather but not the grandmother coded as affected, and one with the reverse affection assignment.

**Exome Variant Annotation, Filtering, and Single-Variant Genotyping.** Exome variants were annotated using ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar>) and Seattle Seq 137 (<http://snp.gs.washington.edu/SeattleSeqAnnotation137/HelpHowToUse.jsp>), Variant Effect Predictor, Release 76 [81, 82], and searched with GEMINI [83]. All DNA physical map locations reported in this study refer to the hg19 reference genome. In the exome sequences, an important filtering criterion was position within regions implicated in linkage analysis. Because of the assumption of autosomal dominant inheritance, heterozygous variant genotypes were prioritized. In Family A, the possibility that the children inherited causal variants from either

of the two grandparents was considered. Based on the assumption that the causative change is relatively rare in the population, allele frequencies in control exomes obtained to date by the National Heart, Lung and Blood Institute's (NHLBI) Exome Sequencing Project (ESP) (<http://evs.gs.washington.edu/EVS/>) and the 1000 Genomes project for European as well as all populations were consulted to prioritize MAFs of 15% or lower. To maximize reliability, variants with read depths < 10 and variants that failed quality control by GATK [84] were excluded. The average read depth of the retained variants was 73.5. Variants were further evaluated with respect to their functions (e.g., missense, coding-synonymous), using the in-house Genome Variation Server, and predicted functional effects (e.g., benign, possibly damaging), using PolyPhen [85] and the Combined Annotation Dependent Depletion (CADD) scores [86].

Genotyping of selected candidate variants was done using polymerase chain reactions in a thermal cycler (DNA Engine Tetrad 2; MJ Research) followed by Sanger sequencing using an ABI 3130xl DNA Analyzer for capillary electrophoresis and ABI BigDye fluorescent dye terminator cycle sequencing kits (Applied Biosystems, Grand Island, NY). In one case, (*NIPBL* variant), ExoSAP-IT purified PCR products were submitted to Genewiz (Seattle, WA) for Sanger sequencing on ABI 3730xl DNA Analyzers. To obtain genotypes of two variants from *C4orf21*, ExoSAP-IT purified PCR products from all available Family B members were submitted to GenScript (Piscataway, NJ) for sequencing on 3730xl DNA Analyzers.

For exome variant filtering and single-marker linkage analysis, we considered not only variants in the regions with positive evidence for linkage but also variants in previously reported candidate regions for CAS [13–15, 18–24, 26, 35–38] including regions implied in reading/spelling disorders due to reported comorbidities with CAS [9–11]. For exome variant filtering, variants shared by all affected individuals in one or both families were considered most plausible.

## Web Resources

BEAGLE

<https://faculty.washington.edu/browning/beagle/b3.html#beagle4>

cnvHap

<http://www.imperial.ac.uk/people/l.coim>

CoNIFER

<http://conifer.sourceforge.net>

GEMINI

<http://faculty.washington.edu/wijsman/software.shtml>

MORGAN

<https://www.stat.washington.edu/thompson/Genepi/MORGAN/Morgan.shtml>

PBAP (Pedigree Based Analysis Pipeline)

<http://faculty.washington.edu/wijsman/software.shtml>

PEDSTATS

<http://csg.sph.umich.edu/abecasis/Pedstats/download/>

PennCNV

<http://penncnv.openbioinformatics.org/en/latest/>

PLINK

<http://pngu.mgh.harvard.edu/~purcell/plink/download.shtml>

## Results

### Linkage Analysis

In Family A, MCMC linkage analysis using PBAP and MORGAN resulted in two regions of interest with peak LOD scores > 1 (Table 1 and Figs 3 and 4). With both grandparents coded

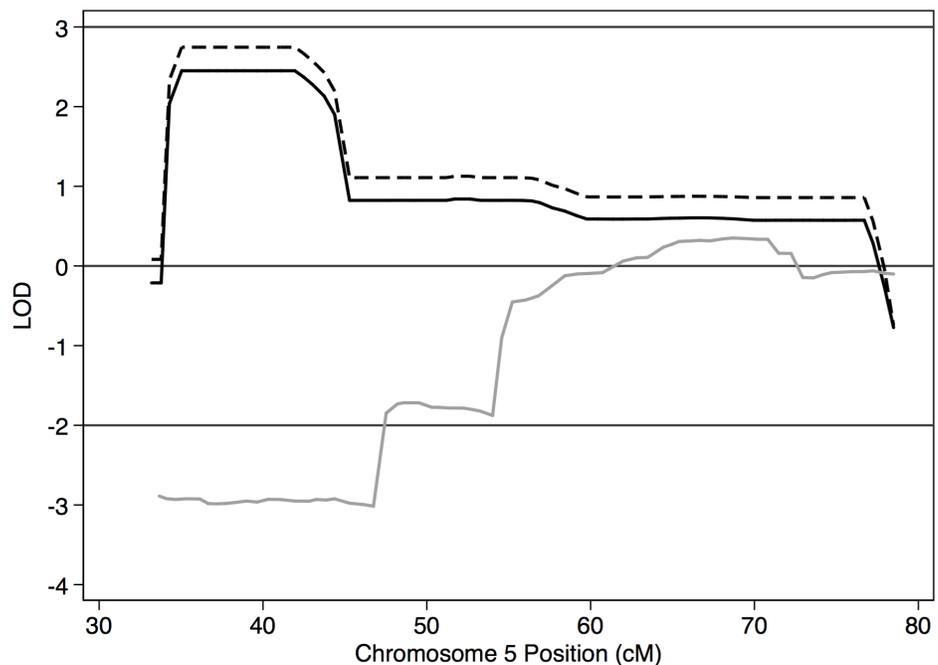
**Table 1. Regions of interest with LOD > 1 in both families.**

Fam.	Chr.	Phys. Pos.	Band	cM	LOD <sub>max</sub>	# Genes in Region	95% CI (bp)	95% CI (Band)	cM	# Genes in CI Region
A	5	15,117,438–63,394,447	p15.1-q12.3	34.3–77.2	2.45 (2.75*)	340	15,117,438–24,806,593	p15.1-p14.1	34.3–44.4	60
A	17	12,582,787–28,550,814	p13.1-q11.1	35.5–56.1	1.79 (2.09**)	341	12,582,787–28,550,814	p13.1-q11.1	35.5–56.1	341
B	1	95,781,217–154,626,705	p21.3-q21.3	118.8–149.9	2.02	779	99,733,964–109,352,915	p21.1-p13.2	121.7–129.3	79
B	4	109,931,951–131,629,960	q25-q28.3	112.4–128.1	1.97	186	111,267,671–131,013,140	q25-q28.2	113.8–128.1	163
B	4	155,182,784–175,836,048	q31.3-q34.1	149.7–169.9	1.46	156	162,595,918–175,443,156	q32.1-q34.1	155.6–169.1	100
B	6	138,299,093–147,799,122	q23.3-q24.3	140.8–148.3	1.15	76	138,599,285–147,799,122	q23.3-q24.3	142.5–148.3	73
B	10	32,584,506–56,662,603	p11.22-q21.1	60.2–74.3	1.42	276	32,584,506–54,037,188	p11.23-q11.22	60.2–71.8	266
B	12	101,877,346–115,526,747	q23.2-q24.21	115.1–135.5	1.98	193	104,311,731–105,281,654	q23.3-q24.11	117.7–120.3	15
B	13	19,263,735–23,953,924	q11-q12.12	0–11.91	1.87	101	19,263,735–23,953,924	q11-q12.12	0–10.5	101
B	17	75,839,026–80,986,540	q25.3	130.7–143.8	1.44	143	75,839,026–80,986,540	q25.3	130.7–143.8	143
B	21	14,827,698–16,611,077	q11.2	5.4–10.2	2.03	40	14,827,698–16,611,077	q11.2	5.4–10.2	40

\* Family A grandfather coded as affected

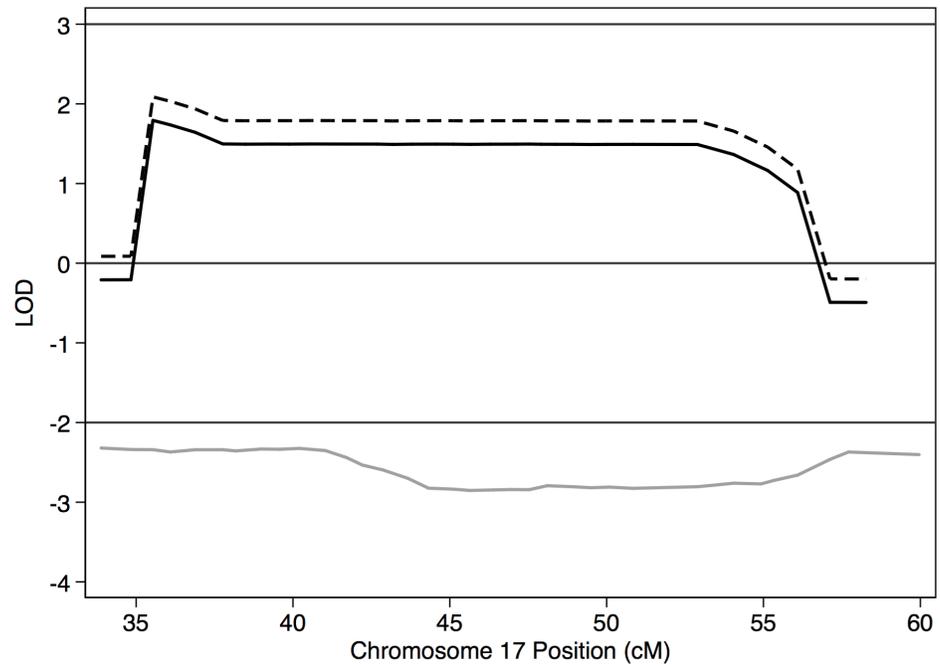
\*\* Family A grandmother coded as affected

doi:10.1371/journal.pone.0153864.t001



**Fig 3. Linkage results for chr 5.** Solid black line = Family A with both grandparents coded as affectionation unknown; dashed black line = Family A with grandfather coded as affected; solid gray line = Family B.

doi:10.1371/journal.pone.0153864.g003



**Fig 4. Linkage results for chr 17.** Solid black line = Family A with both grandparents coded as affection unknown; dashed black line = Family A with grandmother coded as affected; solid gray line = Family B.

doi:10.1371/journal.pone.0153864.g004

as having unknown affection status, a region of interest with  $LOD_{max} = 2.45$ , the maximum score possible under this model, was seen at the 95% CI region of 5p15.1-p14.1 (bp 15,117,438–24,806,593), containing 60 genes. When the linkage analysis for this chromosome was run under the assumption that the grandfather but not the grandmother was affected, consistent with the IBD results described below,  $LOD_{max}$  increased to 2.75, again the estimated maximum score given this model. When both grandparents were coded with unknown affection status, evidence for linkage with  $LOD_{max} = 1.79$  was found at the 95% CI region of 17p13.1-q11.1 (bp 12,582,787–28,550,814), where 341 genes are located. When the linkage analysis was run under the assumption that the grandmother was affected but not the grandfather, as supported by the IBD results for this region,  $LOD_{max}$  increased to 2.09. [S1 Fig](#) shows linkage results for all autosomes in both families.

In Family B, linkage analysis resulted in 9 linkage regions with  $LOD > 1$  ([Table 1](#), [S1 Fig](#)). The regions on chrs 1, 4 (q25-q28.3), 12, 13, and 21 are more plausible than the other regions based on the size of the region and LOD scores approaching the estimated maximum scores in this family.

Regions with LOD scores  $\leq -2$  are considered exclusionary, under the assumption of the correct penetrance mode. Several regions and genes previously identified as loci of interest for CAS overlapped with such regions, as follows: *DCDC2* (Family B), *KIAA0319* (Family B), part of the previously reported 6p21.2-p12.3 region of interest [40] (Family B), *FOXP2* (Family B), *CNTNAP2* (both families), part of the previously reported 7q36.1-q316.3 region of interest [40] (both families), *GALT* [20] (both families), part of the 16p11.2 deletion region [35–37] (Family A), and *CNTNAP1* [38] (both families). Several regions of interest in one family coincided with LOD scores  $\leq -2$  in the other family ([S1 Fig](#)).

## Copy-Number Variation (CNV)

In Family A, CNV analysis based on the four exomes using CoNIFER [74] and the HCE/1-0 SNP array genotypes obtained from the same four samples and 12 additional ones, using PennCNV [76] and cnvHap [77] did not result in any deletions or duplications that segregated with the speech disorder. In the HCE/1-0 SNP genotypes from two affected family members, A-304 and A-311, the regions previously reported deleted in individuals with CAS on 2p16.1 [26] and 16p11.2 [35–37] contained 33.3% and 46.7% heterozygous genotypes, respectively, in the 2p16.1 region and 5.7% and 10.9%, respectively, in the 16p11.2 region.

In Family B, the 15q26.3 duplication encompassing the *FAM169B* gene that had been previously identified in the proband and his mother was confirmed with our probes as well as with the PennCNV, cnvHap, and CoNIFER procedures. The probes and results from PennCNV and cnvHap also identified this duplication in B-202 (affected), B-206 (affectation unknown), and B-301 (unaffected) but not in B-204 (affectation unknown) and B-311 (affected). The second microdeletion involving part of the *MEF2A* gene was not confirmed with any of our CNV methods.

No CNVs shared by the affected family members were identified with PennCNV, cnvHap, or CoNIFER. HCE/1-0 SNP genotypes from two affected participants, B-403 and B-311, contained 41.2% heterozygous genotypes in the *BCL11A* region. In the 16p11.2 region, they contained 4.5% and 11.9% heterozygous genotypes, respectively.

## Identity-by-Descent (IBD) Analysis

In Family A, regions shared IBD as estimated with BEAGLE were observed only in two regions, 5p15.2-p14.1 and 17p12-q11.2. In these regions, there was IBD-sharing by all 15 pairwise comparisons of the six grandchildren selected for HCS SNP genotyping and simultaneously one of the grandparents. The region on chr 5 mutually shared IBD by all 6 grandchildren ranged from bp 14,776,897 to 25,340,617, a region nearly identical to the 95% CI region of interest obtained with linkage analysis. Pairwise comparisons between each of the grandparents and the six grandchildren resulted in matches between the grandfather and each of the six grandchildren between bp 3,265,689 and 25,390,490, and no matches between the grandmother and any of the six grandchildren. On chr 17, pairwise IBD sharing among the six selected grandchildren showed a region common to all 15 pairs, ranging from bp 12,348,755 to 29,783,244, again nearly identical to the results from linkage analysis. The grandfather shared this region with none of the grandchildren but the grandmother shared a region ranging from bp 9,614,556 to 29,943,842 with all six grandchildren.

## Exome Variants in Regions of Interest Based on Linkage Analysis

In Family A, the 95% CI region of interest on chr 5 contained 15 variants, of which the grandfather carried three in the heterozygous state (S1 Table). Of these, the only variant shared exclusively by the grandfather and both grandchildren was rs17285716 in *CDH18* (Cadherin 18 Type 2, OMIM #603019) (Table 2). This variant has a MAF of 0.18 in 1000 Genomes Europeans (0.10 in all populations combined) and a CADD score of 9.71. Two common variants carried by the grandfather, rs11750538 in *MYO10* (Myosin X, OMIM #601481) and rs162848 in *FAM134B* (Family with Sequence Similarity 134, Member B, OMIM #613114), were shared not only by both grandchildren but also the grandmother; the *MYO10* variant was additionally shared by B-311. The extended region with non-negative LOD scores (Table 1) contained nine additional variants with  $MAF \leq 0.15$  that were exclusively shared by the grandfather and the two grandchildren. In addition, one rare variant located in the *IL7R* gene (Interleukin 7 receptor, OMIM \*146661) was shared by the grandfather, the two grandchildren, and all three

exomes in Family B. Three variants shared by the grandfather and the two grandchildren were also seen in one or two of the Family B exomes. In the entire chr 5 region of interest, no variants were seen exclusively in the grandmother and the two grandchildren.

Within the 95% CI region of interest on chr 17, the grandmother carried 111 variants in the heterozygous state. Of these, 7 variants had  $MAF \leq 0.15$  and were shared with both grandchildren but not the grandfather. None of these 7 variants were found in Family B. In this region of interest, no variants were shared by the grandfather and the grandchildren but not the grandmother.

In Family B, within the 95% CI regions of interest, 40 variants had  $MAF \leq 0.15$ . Of these, 16 were shared only by the three Family B exomes; none segregated in the exomes of both families. Also of these 40 rare variants, seven were located in one of the more compelling regions of interest, which was 4q25-q28.2 for five of them and 1p21.1-p13.2 for the other two (Table 2). Two of the chr 4 variants were rs76187047 and rs61745597, both located in the *C4orf21* gene (Chromosome 4 Open Reading Frame 21, no OMIM number) and not found in Family A. An alias for *C4orf21* is *ZGRF1* (Zinc finger, GRF-type containing 1). The rs76187047 variant is a missense variant with a high CADD score of 25.3. It occurs in <1% of Europeans and all populations. The other *C4orf21* variant is equally rare, also has a missense function, and has a CADD score of 7.6. The rs61732241 variant in the *SYNPO2* gene (Synaptopodin-2, no OMIM number) on chr 4 is not extremely rare and was also seen in two Family A exomes. The rs12499000 variant in the *PRDM5* gene (PR Domain-containing protein 5, OMIM #614161) was only seen in the Family B exomes; it is not rare in Europeans. The rs6848868 variant in the *KIAA1109* gene (OMIM \*611565) was also seen in one of the Family A grandchildren. The two chr 1 variants were only rare in all populations combined and also shared by a subset of the Family A exomes. Table 2 lists segregating variants in the 95% CI in the most compelling regions based on linkage analysis. S1 Table lists all segregating variants in the regions of interest.

**Table 2. Exome variants in the regions of interest based on linkage analysis results in 95%CI regions and regions with strongest evidence.**

Fam.	Chr.	Gene	rs ID	hg19 Position	MAF (1KG EUR)	MAF (1KG All)	CADD	Carrier in Other Family
A*	5	<i>CDH18</i>	rs17285716	19,591,174	0.18	0.1	9.71	
A**	17	<i>NCOR1</i>	rs61754982	16,004,888	<0.01	<0.01	4.79	
A**	17	<i>FLCN</i>	rs3744124	17,124,815	0.03	0.1	0.08	
A**	17	<i>TOM1L2</i>	rs143069395	17,786,070	<0.01	<0.01	8.65	
A**	17	<i>EPN2</i>	rs55883526	19,216,576	0.01	0.01	1.11	
A**	17	<i>KSR1</i>	rs2293180	25,909,816	0.12	0.17	9.37	
A**	17	<i>TNFAIP1</i>	rs145418568	26,671,614	0.01	<0.01	0.15	
A**	17	<i>RPL23A</i>	rs2288595	27,052,358	0.07	0.11	1.59	
A**	17	<i>NEK8</i>	rs200972000	27064924	<0.01	<0.01	15.47	
B	1	<i>VCAM1</i>	rs3176878	101,203,698	0.15	0.13	0.01	102, 312
B	1	<i>COL11A1</i>	rs17127270	103,405,892	0.18	0.12	0.38	101, 304
B	4	<i>C4orf21</i>	rs76187047	113,506,711	0.01	0.02	25.30	
B	4	<i>C4orf21</i>	rs61745597	113,544,993	0.01	0.02	7.57	
B	4	<i>SYNPO2</i>	rs61732241	119,952,955	0.11	0.04	8.81	101, 304
B	4	<i>PRDM5</i>	rs12499000	121,706,201	0.27	0.14	10.32	
B	4	<i>KIAA1109</i>	rs6848868	123,150,286	0.08	0.03	10.64	312

\* Consistent with inheritance from the grandfather

\*\* Consistent with inheritance from the grandmother

doi:10.1371/journal.pone.0153864.t002

### Exome Variants in Other Regions of Interest

In 17 previously reported regions or genes of interest for CAS [13–15, 18–24, 26, 35–38], variants with  $MAF \leq 0.15$  shared by one grandparent and both grandchildren in Family A or all three exomes in Family B were only seen in the 6p21.2-p12.3 region of interest, the 7q11.23 duplication region, the 7q36.1-q38.3 region of interest, and the Senotaxin (*SETX*; OMIM # 608465) gene on chr 9. Most of the variants in 6p21.2-p12.3 were seen in the Family B exomes, whereas the variants in the 7q11.23 duplication region were seen in Family A. One variant, rs386701097 in the Polycystic and Hepatic Disease 1 (*PKHD1*; OMIM # 606702) gene, was seen in all exomes except the grandmother in Family A. S2 Table lists variants segregating in the exomes of either of the two families and carriers in the other family.

### Selected Candidate Variants

In the single-marker linkage analysis in Family A, the highest LOD score, 2.45, was obtained for rs17285716 in the *CDH18* gene in the 95% CI region of interest on chr5p15.1-p14.1 when both grandparents were coded as unknown affection status and 2.75 when the grandfather was coded as affected (Table 3). A *MYO10* variant in the 5p region of interest had a LOD score of 1.24 regardless of the affection coding of the grandparents. A variant in the *NIPBL* gene (Nipped-B-Like; OMIM #608667) in the extended 5p region of interest had a LOD score of 1.14 when the grandfather was coded as affected. The second highest LOD score, 1.49, was obtained for two variants on chr 17, rs61754982 (Nuclear Co-receptor Repressor 1, *NCOR1*; OMIM #600849) and rs3744124 (Folliculin, *FLCN*; OMIM #607273), with an increase to  $LOD = 1.79$  when the grandmother, who carries the variants, was coded as affected. Also in the linkage region on chr 17 and inherited from the grandmother, an unannotated variant at position 27,064,924 in the *NEK8* gene (Never in Mitosis A-Related Kines 8, OMIM #609799) had a LOD score of 1.20, increasing to 1.50 when the grandmother was coded as affected. Variants in the *SMCR8* gene (Smith-Magenis Syndrome Chromosome Region, candidate 8; no OMIM #) and the *GLP2R* (Glucagon-Like Peptide 2 Receptor, OMIM #603659) gene also showed increased LOD scores when the grandmother was coded as affected. Two variants from genes outside the regions of interest, Ankyrin Repeat Domain 12 (*ANKRD12*, OMIM # 610616) and

**Table 3. Family A LOD scores from single-marker analysis for variants with LOD > 1.** See S3 Table for all tested markers.

Chr. Band	Gene	rs ID	CADD	hg19 Position	Rationale	LOD (Grandp. Aff. Unknown)	LOD (Grandf. Aff.)	LOD (Grandm. Aff.)
5p15.1	<i>MYO10</i>	rs396514	8.27	16,794,916	Linkage ROI	1.24	1.24	1.24
5p14.3	<i>CDH18</i>	rs17285716	9.71	19,591,174	Linkage ROI; exomes	2.45	2.75	0.88
5p13.2	<i>NIPBL</i>	NA	10.98	37,064,663	Linkage ROI	0.87	1.14	-0.09
12p13.33	<i>CACNA1C</i>	rs216008	10.35	2,721,137	Cand. gene for dev. dis.	0.90	-0.23	1.19
17p13.1	<i>GLP2R</i>	rs1113915	(intron)	9,770,685	Linkage ROI	1.06	-0.42	1.35
17p11.2	<i>NCOR1</i>	rs61754982	4.79	16,004,888	Linkage ROI; exomes	1.49	-0.03	1.79
17p11.2	<i>FLCN</i>	rs3744124	0.08	17,124,815	Linkage ROI; exomes	1.49	-0.03	1.79
17p11.2	<i>SMCR8</i>	rs8080966	9.13	18,220,674	Linkage ROI	0.96	-0.31	1.25
17q11.2	<i>NEK8</i>	NA	15.47	27,064,924	Linkage ROI	1.20	-0.28	1.50
18p11.22	<i>ANKRD12</i>	rs116726679	0.60	9,255,539	Dyslexia cand. gene	1.09	1.38	-0.34

doi:10.1371/journal.pone.0153864.t003

**Table 4. Number of risk alleles in Family A members.**

ID	Aff.	<i>MYO10</i> rs396514	<i>CDH18</i> rs17285716	<i>NIPBL</i> NA	<i>CACNA1C</i> rs216008	<i>GLP2R</i> rs1113915	<i>NCOR1</i> rs61754982	<i>FLCN</i> rs3744124	<i>SMCR8</i> rs8080966	<i>NEK8</i> NA	<i>ANKRD12</i> rs116726679
101	(GF)	1	1	1	0	0	0	0	1	0	1
102	(GM)	1	0	0	1	2	1	1	2	1	0
206	Aff.	1	1	1	1	1	1	1	1	1	1
207	Aff.	1	1	1	1	1	1	N/A	1	1	1
304	Aff.	1	1	1	1	1	1	1	1	1	1
305	Aff.	1	1	1	1	1	1	1	1	1	1
310	Aff.	1	1	1	1	2	1	1	1	N/A	1
311	Aff.	1	1	0	1	1	1	1	1	1	1
312	Aff.	1	1	1	1	2	1	1	1	1	1
314	Aff.	1	1	1	2	1	1	1	2	1	1
202	Unaff.; OC	0	1	1	1	1	1	N/A	2	1	1
201	Unaff.	1	0	0	1	0	0	0	0	0	1
204	Unaff.	1	0	0	0	1	1	1	1	1	1
205	Unaff.	2	0	0	1	1	0	0	0	0	0
208	Unaff.	2	0	0	1	0	0	0	1	0	0
303	Unaff.	1	0	0	0	0	1	1	1	1	1
309	Unaff.	2	0	0	1	2	0	0	0	0	0
302	Unk.	0	0	0	1	0	0	N/A	1	0	2
313	Unk.	2	0	0	2	2	1	1	1	1	1

doi:10.1371/journal.pone.0153864.t004

Calcium Channel, Voltage-Dependent, L Type, Alpha-1C subunit (*CACNA1C*, OMIM #114205), also showed increased LOD scores > 1 when the grandmother or the grandfather were coded as affected, respectively. The *ANKRD12* variant has predicted downstream effects and the *CACNA1C* variant is synonymous. The average number of risk alleles across these 10 variants in the affected and unaffected members, respectively, was 11.25 (range: 9,12) and 4.8 (range: 3, 7), not counting the grandparents and the obligate carrier.

In Family B, the *C4ord21* variants, both of which are rare and deleterious, were linked and found in all affected family members, one unaffected member, and three members of unknown affection, one of whom, B-405, had a history of difficulties with written language. The average number of risk alleles among the affected and unaffected members, respectively, was 2 (no variation) and 0.4 (range: 0, 2).

Table 3 summarizes the results from Sanger genotyping and single-marker linkage analysis for all markers with single-marker LOD scores > 1 in Family A. S3 Table shows results for all tested markers. Tables 4 and 5 summarize the number of risk alleles in the Family A and B members, respectively, by variant and affection status.

## Discussion

The purpose of this study was to investigate the genetic etiology of a severe form of speech sound disorder, childhood apraxia of speech (CAS), in two multigenerational families. We used complementary approaches and compared the results from the two families to each other and to previously described findings in individuals with CAS. Results are consistent with different genetic etiologies in the two families as well as a heterogeneous etiology more broadly, because previously reported candidate genes were not confirmed in either of the two families.

In Family A, linkage analysis resulted in two regions of interest located at 5p15.1-p14.1 and 17p13.1-q11.1, both of which overlapped partially with regions that provide exclusionary

**Table 5. Number of risk alleles in Family B members.**

ID	Affectation	<i>C4orf21</i> rs61745597	<i>C4orf21</i> rs76187047
404	Aff.	1	1
302	Aff.	1	1
409	Aff.	1	1
311	Aff.	1	1
303	Unaff.	0	0
205	Unaff.	0	0
505	Unaff.	1	1
410	Unaff.	0	0
301	Unaff.	0	0
308	Unaff.; OC	1	1
204	Unkn.	1	1
206	Unkn.	0	0
405	Unk. OC	1	1
202	Unk. OC	1	1

doi:10.1371/journal.pone.0153864.t005

evidence in Family B. The results from linkage models with one or the other of the grandparents coded as affected, IBD testing, exome variant analysis, and genotyping the candidate variants in the whole family suggest that affected individuals in the family inherited the 5p15.1-p14.1 region from the grandfather and the 17p13.1-q11.1 region, from the grandmother. According to the family interviews, the grandfather but not the grandmother had undergone speech therapy as a child, although both had biological relatives with difficulty in the area of speech and language. It is therefore plausible to suspect that the region of interest on chr 5 harbors one or more variants influencing speech development, whereas the region of interest on chr 17 may harbor variants influencing other inherited traits in the family or may represent a false positive result. Alternatively, it is possible that variants in the 17p13-q11.1 region or both of these regions influenced speech development in the affected family members. The comparison of risk alleles in both regions and additional candidate loci in the affected and unaffected members of the family shows that the affected group had more than twice as many risk alleles as the unaffected group. This finding is consistent with multiple factors influencing the phenotype, possibly with additive effects, similar to findings in other neurological disorders. The unaffected obligate carrier had the risk allele of most of these loci; the reasons for her lack of speech difficulties are not clear.

In the single-marker linkage analysis, the highest obtainable LOD score of 2.75, as estimated by power analysis with one grandparent coded as affected, was seen for only one tested variant, rs17285716 in the *CDH18* gene located in the 5p15.1-p14.1 region of interest. This variant is synonymous but has a high conservation score based on 100 vertebrates basewise conservation score by PhyloP and a high GERP score based on 35 mammalian alignments, and its scaled CADD score of 9.71 places it close to the top ten percent of variants in pathogenicity. Synonymous variants have been implicated in other disorders, for instance in a parkinsonian disorder where a synonymous variant was associated with exon skipping [87]. *CDH18* is specifically expressed in the central nervous system and is thought to influence synaptic adhesion, axon outgrowth, and axon guidance, thus regulating the development of the central nervous system [88–90]. The Allen Human Brain Atlas (AHBA; <http://human.brain-map.org/>) [91] shows maximum *CDH18* expression levels throughout the cerebellar

cortex. A DECIPHER (<https://decipher.sanger.ac.uk>) [92] search yielded 44 syndromic cases with CNVs involving *CDH18* including a case with apraxia, speech and language development, hyperactivity, intellectual disability and a 2.95 Mb duplication partially involving *MYO10* and *CDH18*, and a case with attention deficit hyperactivity disorder, speech apraxia, and a 6.96 Mb deletion ranging from *DNAH5* to part of *CDH18*. Speech and language phenotypes were noted in two additional cases with *CDH18* CNVs.

It is possible that *CDH18* influences speech development by acting in concert with other, functionally related genes via regulatory mechanisms. The functional network of *CDH18* includes 15 cadherin genes (*CDH1* through *CDH13*, *CDH17*, and *CDH24*), all influencing cell adhesions, as well as three cadherin-associated protein genes, *CTNNA1*, *CTNNB1*, and *CTNND1*; four of these genes are expressed in the cerebellum. The 5p region of interest harbors two of the cadherin genes, *CDH12* (OMIM #600562) and *CDH10* (OMIM #604555), as well as the *MYO10* gene, where one variant had a single-marker LOD score of 1.24 regardless of affectation status of the grandparents, both of whom are carriers. *MYO10* encodes a protein that belongs to the myosin superfamily and is expressed in epithelia-rich tissues [93]. Its protein product is expressed in many tissue types; in the central nervous system, the AHBA shows high gene expression levels in the basal ganglia and the thalamus. *MYO10* plays a role in axon development, neurite outgrowth, and radial neuron migration in the developing cortex and cell-matrix adhesion [94]. One variant in the *NIPBL* gene on 5p13.2 has no dbSNP rs number and is extremely rare in the population, found neither in the deep population sequencing ESP (n = 6,503) nor in 1000 Genomes. *NIPBL* codes for a protein necessary for the cohesion of sister chromatids during mitosis [95] and is disrupted by some translocations in Cornelia-de-Lange syndrome [96]. This gene is expressed in many tissue types including brain. The AHBA shows maximum expression levels in the basal ganglia, cerebellum, and corpus callosum.

Additional support for the 5p15.1-p14.1 region of interest is found in the autism literature, where common variants in this region have been implicated [97, 98]. Autism spectrum disorder and CAS co-occur in proportions of cases greater than expected by chance [99].

The second highest single-marker LOD score, 1.79, was seen for two variants in the 17p13.1-q11.1 region of interest with the grandmother coded as affected, rs61754982 in *NCOR1* and rs3744124 in *FLCN*. *NCOR1* is involved in thyroid hormone and retinoid acid repression [100]. The AHBA shows maximum expression levels in the basal ganglia and the cerebellar cortex. DECIPHER lists 20 syndromic cases with CNVs involving *NCOR1* including one with truncal ataxia, one with apraxia, and one with delayed speech and language development. *FLCN* mutations are implicated in Birt-Hogg-Dube syndrome [101], a disease involving fibrofolliculomas, renal tumors, lung cysts, and pneumothorax. The AHBA shows maximum expression levels in the dentate gyrus and the cerebellar cortex. DECIPHER lists 92 syndromic cases with CNVs involving *FLCN* including 10 cases with delayed speech and language development. Also in the linkage region on chr 17 and inherited from the grandmother, an unannotated variant in the *NEK8* gene had a LOD score of 1.50 when the grandmother was coded as affected. *NEK8* is expressed in various brain regions including cerebellar nuclei and is thought to play a role in fetal organ development and polycystic kidney disease [102]. A variant within the *SMCR8* gene in the Smith-Magenis Syndrome region was also inherited from the grandmother. This syndrome is characterized by a diverse set of traits including mild intellectual disability, delayed speech and language abilities, distinctive facial features, abnormal sleep patterns, and challenging behaviors. Only the speech trait fits the phenotypic profile observed in Family A. Also inherited from the grandmother was a variant in the *GLP2R* (glucagon-like peptide 2 receptor) gene, which plays a role in intestinal growth and nutrient absorption. According to the AHBA, *SMCR8* expression is widely dispersed throughout regions of the brain whereas *GLP2R* is mainly expressed in cortical regions. Of five cases with CNVs

involving *GLP2R* listed in DECIPHER, a 2.45 Mb deletion was associated with autism and delayed speech and language development.

Single-marker linkage analysis resulted in LOD scores  $> 1$  for variants in two genes outside the regions of interest from linkage analysis. *ANKRD12* is located within a dyslexia candidate region, DYX6, on 18p11.22. It is expressed in many tissues including brain [103], most strongly in the cerebellar cortex according to the AHBA. A single-point LOD score of 1.38 when the grandmother was coded as affected was obtained for rs116726679 within this gene. This variant is synonymous with predicted downstream effects. *CACNA1C* is involved in cellular processes including contraction and electric signaling [104]. Channelopathies associated with *CACNA1C* have been observed in psychiatric disorders [105], although these were not reported for Family A. According to the AHBA, strongest expression levels are in the thalamus. This variant is synonymous. A LOD score  $> 1$  was only seen when the grandmother was coded as affected.

In Family B, the two most plausible exome variants, rs61745597 and rs76187047, are located in the *C4orf21* (*ZGRF1*) gene. These highly deleterious variants are linked and carried by all affected members, the three obligate carriers (two with unknown affection and one unaffected), and only one other unaffected member; they were not found in four unaffected members and one member with unknown affection. One carrier with unknown speech affection status had a history of written language difficulties, raising the possibility of variable expressivity. *C4orf21* is not yet well characterized in terms of function and functional networks. One paralogous protein is encoded by *SETX*, one of the genes of interest for CAS [38]. Although *SETX* is not located within a region with positive evidence for linkage, the grandfather and the two grandchildren in Family A share a rare variant in this gene. *SETX* is associated with autosomal recessive spinocerebellar ataxia-1 and ataxia-oculomotor apraxia-2 [106]. *C4orf21* may encode similar functions related to motor praxis. Similar to the most plausible genes in Family A, *C4orf21* is highly expressed in the cerebellum. Of 10 cases with CNVs involving this gene listed in the DECIPHER database, one, a 23 Mb deletion, is characterized by craniosynostosis and delayed speech and language.

Because *CDH18*, several functionally related genes, *C4orf21*, and other genes of interest in both families are strongly expressed in the cerebellum, it is possible that genetic influences converge on the cerebellum, producing downstream effects on speech and other behaviors. The cerebellum plays a crucial role not only in complex motor processes but also in linguistic and cognitive activities, observable in infants during speech perception [107] and in children and adults during tasks requiring sequential processing such as applying syntactic rules [108, 109]. Cerebellar anomalies were implicated in functional and structural imaging studies of the KE family with severe CAS caused by a *FOXP2* mutation [16, 17]. The deficits in motor sequencing as well as those in linguistic and cognitive tasks observed in our CAS studies [39, 42, 52] are consistent with cerebellar involvement, although brain imaging data were not collected.

Only two variants, rs2228141 in the *IL7R* gene, located in the 5p region of interest but outside the 95% CI, and rs386701097 in the *PKHD1* gene in a region previously reported as a region of interest, were shared by all relevant samples in both families (one grandparent and both grandchildren in Family A and all Family B exomes). *IL7R* plays a role in the immune system and is implicated in multiple sclerosis, whereas *PKHD1* influences kidney and liver functions. It is uncertain whether these variants contributed to the phenotype.

We found no evidence of causal CNVs in either of the families. In Family B, one region of a previously identified duplication on 15q was confirmed in the proband and his mother, but was not found in an affected family member, making it unlikely to be causal in this family. CNVs occur frequently without any pathogenic effects [110]. In the two families, we found no evidence for CNVs that had previously been reported to associate with CAS [23, 26, 35, 36], and we also effectively ruled out deletions in the 2p16.1 and 16p11.2 regions by showing

heterozygous genotypes in affected family members of both participating families in these regions. Similarly, we found no evidence for most of the previously reported validated or candidate single nucleotide variations [13, 14, 38]. Exceptions are the *SETX* variant in Family A, variants in regions of interest in 6p21.2-p12.3 (Family B), 7q11.23 and 7q36.1-q36.3 (Family A). Whether or not these variants contributed to the phenotype is uncertain.

Taken together, our results are consistent with the hypothesis that the CAS phenotype, like many other neurological phenotypes including Alzheimer's disease and autism spectrum disorder, is a complex and genetically heterogeneous disorder with several discoverable variants, each of which segregates and confers risks of varying levels of impact. The authors of a recent study of an extended family with Alzheimer's disease report multiple segregating risk factors of high impact including *ApoE4* and *TREM2*, where the effects of the variants were interpreted as additive [111]. Multiple-hit risks for autism spectrum disorder have been observed in a sporadic case [112]. In simplex cases and multigenerational families with autism spectrum disorder, imbalances in multiple genes were found to contribute to the disease state [113]. We posit that we discovered at least one variant of high impact in each of the two families but that there may be other factors influencing the trait. Whether or not distinctly different genetic etiologies have clinical implications for diagnosis and therapy remains to be investigated.

Ongoing efforts to characterize the genetic etiology of CAS and other forms of speech sound disorders will lead to early identification of infants at genetic risk and motivate the development of effective preventative measures. Detailed knowledge of genotype-phenotype associations will also provide the basis for subtype-specific, customized therapy approaches.

## Limitations and Future Directions

Learning to speak is a complex task thought to be influenced by many variables related to genes and environment. Additional sources of variability in the phenotype of the affected family members are their relative ages and type and duration of intervention. It is possible that individuals who married into the family contributed additional genetic risk factors that modified the speech development in the children, potentially adding genetic heterogeneity.

The assumptions underlying the present study included relatively low allele frequencies, few causal variants, and heterozygous genotypes exclusively shared by the affected family members. If CAS were the result of many variants, each common and of very small effect, then our methodology would be unsuitable to identify them. Similarly, it is possible that one or more causal variants reside outside the exome in regulatory DNA regions, in which case we would have been unable to detect them.

We obtained CNV data from two sources (exomes, SNP array) and three algorithms (CoNIFER, cnvHap, PennCNV). It is possible that some CNVs were missed or reported as false positives.

Future plans include investigating these and other multigenerational families with severe forms of speech sound disorder consistent with CAS using whole genome sequences of all informative family members to look for segregating variants. Such studies have the potential to identify causal variants, even in cases of polygenic etiologies where multiple genes, each of moderate effect, shape the ability to learn to speak. In addition, we will investigate samples from many smaller families for shared variations, as these families would not provide sufficient statistical power for linkage analysis in individual families but pooled together, they would provide useful data toward discovery of contributing variants. Efforts to characterize the genetic of CAS and other forms of speech sound disorders will lead to early identification of infants at genetic risk and motivate the development of effective preventative measures.

## Supporting Information

**S1 Fig. Linkage analysis results for all autosomes in both families.** Previously reported candidate regions are marked along the cM axis.  
(TIF)

**S1 Table. Exome variants in the regions of interest based on linkage analysis results.**  
(DOCX)

**S2 Table. Segregating exome variants in previously identified CAS genes or regions of interest in Family A.**  
(DOCX)

**S3 Table. Family A LOD scores from single-marker analysis for all tested variants.**  
(DOCX)

## Acknowledgments

The authors are deeply grateful to the participants whose time and effort made this study possible. The following funding sources are gratefully acknowledged: American Speech-Language-Hearing Foundation New Century Scholars Research Grant (B. Peter), NIDCD T32DC00033 (B. Peter), NIDCD R03DC010886 (B. Peter), R01HD054562 (W. H. Raskind), and R01MH094293 (E. Wijsman). SNP genotyping and exome sequencing were provided by the University of Washington Center for Mendelian Genomics (UW CMG) and was funded by the National Human Genome Research Institute and NIH grant 1U54HG006493 to the three UW CMG members, Drs. M. Bamshad, D. Nickerson, and J. Shendure. Ian Stanaway was supported under grants from the Environmental Health Sciences of the National Institutes of Health (T32ES015459 and 5P01 ES009601) and the EPA (RD-83451401). Many thanks to Drs. E. Thompson and E. Blue for help with MORGAN program implementation and to Drs. Bamshad, Nickerson, and Shendure for helpful consultation.

## Author Contributions

Conceived and designed the experiments: BP EW WR. Performed the experiments: BP MM IS JW KO VG. Analyzed the data: BP AN UWCMG MM KC. Contributed reagents/materials/analysis tools: EW AN UWCMG KC. Wrote the paper: BP EW AN WR.

## References

1. Shriberg LD, Lewis BA, Tomblin JB, McSweeney JL, Karlsson HB, Scheer AR. Toward diagnostic and phenotype markers for genetically transmitted speech delay. *Journal of speech, language, and hearing research: JSLHR*. 2005; 48(4):834–52. doi: [10.1044/1092-4388\(2005/058\)](https://doi.org/10.1044/1092-4388(2005/058)) PMID: [16378477](https://pubmed.ncbi.nlm.nih.gov/16378477/).
2. Lewis BA, Freebairn LA, Hansen AJ, Stein CM, Shriberg LD, Iyengar SK, et al. Dimensions of early speech sound disorders: A factor analytic study. *J Commun Disord*. 2006; 39(2):139–57. doi: [10.1016/j.jcomdis.2005.11.003](https://doi.org/10.1016/j.jcomdis.2005.11.003) PMID: [16386753](https://pubmed.ncbi.nlm.nih.gov/16386753/).
3. Vick JC, Campbell TF, Shriberg LD, Green JR, Truemper K, Rusiewicz HL, et al. Data-driven subclassification of speech sound disorders in preschool children. *Journal of speech, language, and hearing research: JSLHR*. 2014; 57(6):2033–50. doi: [10.1044/2014\\_JSLHR-S-12-0193](https://doi.org/10.1044/2014_JSLHR-S-12-0193) PMID: [25076005](https://pubmed.ncbi.nlm.nih.gov/25076005/); PubMed Central PMCID: [PMC4300961](https://pubmed.ncbi.nlm.nih.gov/PMC4300961/).
4. Lousada M, Jesus LM, Hall A, Joffe V. Intelligibility as a clinical outcome measure following intervention with children with phonologically based speech-sound disorders. *International journal of language & communication disorders / Royal College of Speech & Language Therapists*. 2014; 49(5):584–601. doi: [10.1111/1460-6984.12095](https://doi.org/10.1111/1460-6984.12095) PMID: [24861159](https://pubmed.ncbi.nlm.nih.gov/24861159/).
5. Hall BJC. Attitudes of Fourth and Sixth Graders Toward Peers With Mild Articulation Disorders. *Language, Speech, and Hearing Services in Schools*. 1991; 22(1):334–40.

6. Shriberg LD, Potter NL, Strand EA. Prevalence and phenotype of childhood apraxia of speech in youth with galactosemia. *Journal of speech, language, and hearing research: JSLHR*. 2011; 54(2):487–519. Epub 2010/10/23. doi: [10.1044/1092-4388\(2010/10-0068\)](https://doi.org/10.1044/1092-4388(2010/10-0068)) PMID: [20966389](https://pubmed.ncbi.nlm.nih.gov/20966389/); PubMed Central PMCID: PMC3070858.
7. Namasivayam AK, Pukonen M, Goshulak D, Hard J, Rudzicz F, Rietveld T, et al. Treatment intensity and childhood apraxia of speech. *International journal of language & communication disorders / Royal College of Speech & Language Therapists*. 2015. doi: [10.1111/1460-6984.12154](https://doi.org/10.1111/1460-6984.12154) PMID: [25581372](https://pubmed.ncbi.nlm.nih.gov/25581372/).
8. Murray E, McCabe P, Ballard KJ. A systematic review of treatment outcomes for children with childhood apraxia of speech. *American journal of speech-language pathology / American Speech-Language-Hearing Association*. 2014; 23(3):486–504. doi: [10.1044/2014\\_AJSLP-13-0035](https://doi.org/10.1044/2014_AJSLP-13-0035) PMID: [24686844](https://pubmed.ncbi.nlm.nih.gov/24686844/).
9. Lewis BA, Freebairn LA, Hansen AJ, Iyengar SK, Taylor HG. School-age follow-up of children with childhood apraxia of speech. *Lang Speech Hear Serv Sch*. 2004; 35(2):122–40. PMID: [15191325](https://pubmed.ncbi.nlm.nih.gov/15191325/).
10. Sices L, Taylor HG, Freebairn L, Hansen A, Lewis B. Relationship between speech-sound disorders and early literacy skills in preschool-age children: impact of comorbid language impairment. *J Dev Behav Pediatr*. 2007; 28(6):438–47. Epub 2007/12/20. doi: [10.1097/DBP.0b013e31811ff8ca](https://doi.org/10.1097/DBP.0b013e31811ff8ca) 00004703-200712000-00003 [pii]. PMID: [18091088](https://pubmed.ncbi.nlm.nih.gov/18091088/); PubMed Central PMCID: PMC2755217.
11. Nathan L, Stackhouse J, Goulandris N, Snowling MJ. Educational consequences of developmental speech disorder: Key Stage 1 National Curriculum assessment results in English and mathematics. *Br J Educ Psychol*. 2004; 74(Pt 2):173–86. Epub 2004/05/08. doi: [10.1348/000709904773839824](https://doi.org/10.1348/000709904773839824) PMID: [15130186](https://pubmed.ncbi.nlm.nih.gov/15130186/).
12. Shriberg LD, Aram DM, Kwiatkowski J. Developmental apraxia of speech: I. Descriptive and theoretical perspectives. *Journal of speech, language, and hearing research: JSLHR*. 1997; 40(2):273–85. PMID: [9130199](https://pubmed.ncbi.nlm.nih.gov/9130199/).
13. Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME. Localisation of a gene implicated in a severe speech and language disorder. *Nature genetics*. 1998; 18(2):168–70. doi: [10.1038/ng0298-168](https://doi.org/10.1038/ng0298-168) PMID: [9462748](https://pubmed.ncbi.nlm.nih.gov/9462748/).
14. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*. 2001; 413(6855):519–23. doi: [10.1038/35097076](https://doi.org/10.1038/35097076) PMID: [11586359](https://pubmed.ncbi.nlm.nih.gov/11586359/).
15. Fisher SE, Lai CS, Monaco AP. Deciphering the genetic basis of speech and language disorders. *Annual review of neuroscience*. 2003; 26:57–80. doi: [10.1146/annurev.neuro.26.041002.131144](https://doi.org/10.1146/annurev.neuro.26.041002.131144) PMID: [12524432](https://pubmed.ncbi.nlm.nih.gov/12524432/).
16. Belton E, Salmond CH, Watkins KE, Vargha-Khadem F, Gadian DG. Bilateral brain abnormalities associated with dominantly inherited verbal and orofacial dyspraxia. *Hum Brain Mapp*. 2003; 18(3):194–200. doi: [10.1002/hbm.10093](https://doi.org/10.1002/hbm.10093) PMID: [12599277](https://pubmed.ncbi.nlm.nih.gov/12599277/).
17. Liegeois F, Morgan AT, Connelly A, Vargha-Khadem F. Endophenotypes of FOXP2: dysfunction within the human articulatory network. *Eur J Paediatr Neurol*. 2011; 15(4):283–8. doi: [10.1016/j.ejpn.2011.04.006](https://doi.org/10.1016/j.ejpn.2011.04.006) PMID: [21576028](https://pubmed.ncbi.nlm.nih.gov/21576028/).
18. Vernes SC, Newbury DF, Abrahams BS, Winchester L, Nicod J, Groszer M, et al. A functional genetic link between distinct developmental language disorders. *The New England journal of medicine*. 2008; 359(22):2337–45. doi: [10.1056/NEJMoa0802828](https://doi.org/10.1056/NEJMoa0802828) PMID: [18987363](https://pubmed.ncbi.nlm.nih.gov/18987363/); PubMed Central PMCID: PMC2756409.
19. Peter B, Raskind WH, Matsushita M, Lisowski M, Vu T, Berninger VW, et al. Replication of CNTNAP2 association with nonword repetition and support for FOXP2 association with timed reading and motor activities in a dyslexia family sample. *Journal of neurodevelopmental disorders*. 2011; 3(1):39–49. PMID: [21484596](https://pubmed.ncbi.nlm.nih.gov/21484596/); PubMed Central PMCID: PMC3163991.
20. Elsas LJ 2nd, Langley S, Paulk EM, Hjelm LN, Dembure PP. A molecular approach to galactosemia. *Eur J Pediatr*. 1995; 154(7 Suppl 2):S21–7. Epub 1995/01/01. PMID: [7671959](https://pubmed.ncbi.nlm.nih.gov/7671959/).
21. Shriberg LD, Potter NL, Strand EA. Prevalence and Phenotype of Childhood Apraxia of Speech In Youth with Galactosemia. *J Speech Lang Hear Res*. 2010. Epub 2010/10/23. 1092–4388\_2010\_10-0068 [pii] doi: [10.1044/1092-4388\(2010/10-0068\)](https://doi.org/10.1044/1092-4388(2010/10-0068)) PMID: [20966389](https://pubmed.ncbi.nlm.nih.gov/20966389/).
22. Pal DK, Li W, Clarke T, Lieberman P, Strug LJ. Pleiotropic effects of the 11p13 locus on developmental verbal dyspraxia and EEG centrotemporal sharp waves. *Genes Brain Behav*. 2010; 9(8):1004–12. Epub 2010/09/10. doi: [10.1111/j.1601-183X.2010.00648.x](https://doi.org/10.1111/j.1601-183X.2010.00648.x) PMID: [20825490](https://pubmed.ncbi.nlm.nih.gov/20825490/).
23. Velleman SL, Mervis CB. Children with 7q11.23 Duplication Syndrome: Speech, Language, Cognitive, and Behavioral Characteristics and their Implications for Intervention. *Perspect Lang Learn Educ*. 2011; 18(3):108–16. doi: [10.1044/1le18.3.108](https://doi.org/10.1044/1le18.3.108) PMID: [22754604](https://pubmed.ncbi.nlm.nih.gov/22754604/); PubMed Central PMCID: PMC3383616.

24. Morris CA, Mervis CB, Paciorkowski AP, Abdul-Rahman O, Dugan SL, Rope AF, et al. 7q11.23 Duplication syndrome: Physical characteristics and natural history. *American journal of medical genetics Part A*. 2015. doi: [10.1002/ajmg.a.37340](https://doi.org/10.1002/ajmg.a.37340) PMID: [26333794](https://pubmed.ncbi.nlm.nih.gov/26333794/).
25. Thevenon J, Callier P, Andrieux J, Delobel B, David A, Sukno S, et al. 12p13.33 microdeletion including ELKS/ERC1, a new locus associated with childhood apraxia of speech. *European journal of human genetics: EJHG*. 2013; 21(1):82–8. doi: [10.1038/ejhg.2012.116](https://doi.org/10.1038/ejhg.2012.116) PMID: [22713806](https://pubmed.ncbi.nlm.nih.gov/22713806/); PubMed Central PMCID: PMC3522191.
26. Peter B, Matsushita M, Oda K, Raskind W. De novo microdeletion of BCL11A is associated with severe speech sound disorder. *American journal of medical genetics Part A*. 2014; 164A(8):2091–6. doi: [10.1002/ajmg.a.36599](https://doi.org/10.1002/ajmg.a.36599) PMID: [24810580](https://pubmed.ncbi.nlm.nih.gov/24810580/).
27. Rajcan-Separovic E, Harvard C, Liu X, McGillivray B, Hall JG, Qiao Y, et al. Clinical and molecular cytogenetic characterisation of a newly recognised microdeletion syndrome involving 2p15-16.1. *J Med Genet*. 2007; 44(4):269–76. Epub 2006/09/12. doi: [10.1136/jmg.2006.045013](https://doi.org/10.1136/jmg.2006.045013) PMID: [16963482](https://pubmed.ncbi.nlm.nih.gov/16963482/); PubMed Central PMCID: PMC2598046.
28. de Leeuw N, Pfundt R, Koolen DA, Neefs I, Scheltinga I, Mieloo H, et al. A newly recognised microdeletion syndrome involving 2p15p16.1: narrowing down the critical region by adding another patient detected by genome wide tiling path array comparative genomic hybridisation analysis. *J Med Genet*. 2008; 45(2):122–4. Epub 2008/02/05. doi: [10.1136/jmg.2007.054049](https://doi.org/10.1136/jmg.2007.054049) PMID: [18245392](https://pubmed.ncbi.nlm.nih.gov/18245392/).
29. Chabchoub E, Vermeesch JR, de Ravel T, de Cock P, Fryns JP. The facial dysmorphism in the newly recognised microdeletion 2p15-p16.1 refined to a 570 kb region in 2p15. *J Med Genet*. 2008; 45(3):189–92. Epub 2008/03/04. doi: [10.1136/jmg.2007.056176](https://doi.org/10.1136/jmg.2007.056176) PMID: [18310269](https://pubmed.ncbi.nlm.nih.gov/18310269/).
30. Liang JS, Shimojima K, Ohno K, Sugiura C, Une Y, Ohno K, et al. A newly recognised microdeletion syndrome of 2p15-16.1 manifesting moderate developmental delay, autistic behaviour, short stature, microcephaly, and dysmorphic features: a new patient with 3.2 Mb deletion. *J Med Genet*. 2009; 46(9):645–7. Epub 2009/09/03. doi: [10.1136/jmg.2008.059220](https://doi.org/10.1136/jmg.2008.059220) PMID: [19724011](https://pubmed.ncbi.nlm.nih.gov/19724011/).
31. Felix TM, Petrin AL, Sanseverino MT, Murray JC. Further characterization of microdeletion syndrome involving 2p15-p16.1. *Am J Med Genet A*. 2010; 152A(10):2604–8. Epub 2010/08/28. doi: [10.1002/ajmg.a.33612](https://doi.org/10.1002/ajmg.a.33612) PMID: [20799320](https://pubmed.ncbi.nlm.nih.gov/20799320/); PubMed Central PMCID: PMC2946431.
32. Prontera P, Bernardini L, Stangoni G, Capalbo A, Rogaia D, Romani R, et al. Deletion 2p15-16.1 syndrome: case report and review. *Am J Med Genet A*. 2011; 155A(10):2473–8. Epub 2011/09/13. doi: [10.1002/ajmg.a.33875](https://doi.org/10.1002/ajmg.a.33875) PMID: [21910216](https://pubmed.ncbi.nlm.nih.gov/21910216/).
33. Piccione M, Piro E, Serraino F, Cavani S, Ciccone R, Malacarne M, et al. Interstitial deletion of chromosome 2p15-16.1: report of two patients and critical review of current genotype-phenotype correlation. *Eur J Med Genet*. 2012; 55(4):238–44. Epub 2012/03/13. doi: [10.1016/j.ejmg.2012.01.014](https://doi.org/10.1016/j.ejmg.2012.01.014) PMID: [22406401](https://pubmed.ncbi.nlm.nih.gov/22406401/).
34. Hancarova M, Simandlova M, Drabova J, Mannik K, Kurg A, Sedlacek Z. A patient with de novo 0.45 Mb deletion of 2p16.1: the role of BCL11A, PAPOLG, REL, and FLJ16341 in the 2p15-p16.1 microdeletion syndrome. *Am J Med Genet A*. 2013; 161(4):865–70. Epub 2013/03/16. doi: [10.1002/ajmg.a.35783](https://doi.org/10.1002/ajmg.a.35783) PMID: [23495096](https://pubmed.ncbi.nlm.nih.gov/23495096/).
35. Raca G, Baas BS, Kirmani S, Laffin JJ, Jackson CA, Strand EA, et al. Childhood Apraxia of Speech (CAS) in two patients with 16p11.2 microdeletion syndrome. *European journal of human genetics: EJHG*. 2013; 21(4):455–9. doi: [10.1038/ejhg.2012.165](https://doi.org/10.1038/ejhg.2012.165) PMID: [22909774](https://pubmed.ncbi.nlm.nih.gov/22909774/); PubMed Central PMCID: PMC3598318.
36. Newbury DF, Mari F, Sadighi Akha E, Macdermot KD, Canitano R, Monaco AP, et al. Dual copy number variants involving 16p11 and 6q22 in a case of childhood apraxia of speech and pervasive developmental disorder. *European journal of human genetics: EJHG*. 2013; 21(4):361–5. doi: [10.1038/ejhg.2012.166](https://doi.org/10.1038/ejhg.2012.166) PMID: [22909776](https://pubmed.ncbi.nlm.nih.gov/22909776/); PubMed Central PMCID: PMC3598310.
37. Laffin JJ, Raca G, Jackson CA, Strand EA, Jakielski KJ, Shriberg LD. Novel candidate genes and regions for childhood apraxia of speech identified by array comparative genomic hybridization. *Genetics in medicine: official journal of the American College of Medical Genetics*. 2012; 14(11):928–36. doi: [10.1038/gim.2012.72](https://doi.org/10.1038/gim.2012.72) PMID: [22766611](https://pubmed.ncbi.nlm.nih.gov/22766611/); PubMed Central PMCID: PMC3563158.
38. Worthey EA, Raca G, Laffin JJ, Wilk BM, Harris JM, Jakielski KJ, et al. Whole-exome sequencing supports genetic heterogeneity in childhood apraxia of speech. *Journal of neurodevelopmental disorders*. 2013; 5(1):29. doi: [10.1186/1866-1955-5-29](https://doi.org/10.1186/1866-1955-5-29) PMID: [24083349](https://pubmed.ncbi.nlm.nih.gov/24083349/); PubMed Central PMCID: PMC3851280.
39. Peter B, Raskind WH. A multigenerational family study of oral and hand motor sequencing ability provides evidence for a familial speech sound disorder subtype. *Topics in language disorders*. 2011; 31(2):145–67. PMID: [21909176](https://pubmed.ncbi.nlm.nih.gov/21909176/); PubMed Central PMCID: PMC3169648.
40. Peter B, Matsushita M, Raskind WH. Motor sequencing deficit as an endophenotype of speech sound disorder: a genome-wide linkage analysis in a multigenerational family. *Psychiatr Genet*. 2012; 22

- (5):226–34. Epub 2012/04/21. doi: [10.1097/YPG.0b013e328353ae92](https://doi.org/10.1097/YPG.0b013e328353ae92) PMID: [22517379](https://pubmed.ncbi.nlm.nih.gov/22517379/); PubMed Central PMCID: PMC3407825.
41. Konig IR, Schumacher J, Hoffmann P, Kleensang A, Ludwig KU, Grimm T, et al. Mapping for dyslexia and related cognitive trait loci provides strong evidence for further risk genes on chromosome 6p21. *Am J Med Genet B Neuropsychiatr Genet.* 2011; 156B(1):36–43. Epub 2010/12/25. doi: [10.1002/ajmg.b.31135](https://doi.org/10.1002/ajmg.b.31135) PMID: [21184582](https://pubmed.ncbi.nlm.nih.gov/21184582/).
  42. Peter B, Button L, Stoel-Gammon C, Chapman K, Raskind WH. Deficits in sequential processing manifest in motor and linguistic tasks in a multigenerational family with childhood apraxia of speech. *Clinical linguistics & phonetics.* 2013; 27(3):163–91. doi: [10.3109/02699206.2012.736011](https://doi.org/10.3109/02699206.2012.736011) PMID: [23339324](https://pubmed.ncbi.nlm.nih.gov/23339324/); PubMed Central PMCID: PMC3875160.
  43. Robbins J, Klee T. Clinical assessment of oropharyngeal motor development in young children. *The Journal of speech and hearing disorders.* 1987; 52(3):271–7. PMID: [3455449](https://pubmed.ncbi.nlm.nih.gov/3455449/).
  44. Fletcher SG. Time-by-count measurement of diadochokinetic syllable rate. *Journal of speech and hearing research.* 1972; 15(4):763–70. PMID: [4652397](https://pubmed.ncbi.nlm.nih.gov/4652397/).
  45. Wagner RK, Torgesen JK, Rashotte CA. CTOPP, Comprehensive Test of Phonological Processing. Austin, Tex.: PRO-ED; 1999. 1 case. p.
  46. Dollaghan C, Campbell TF. Nonword repetition and child language impairment. *J Speech Lang Hear Res.* 1998; 41(5):1136–46. Epub 1998/10/15. PMID: [9771635](https://pubmed.ncbi.nlm.nih.gov/9771635/).
  47. Shriberg LD, Lohmeier HL, Campbell TF, Dollaghan CA, Green JR, Moore CA. A nonword repetition task for speakers with misarticulations: the Syllable Repetition Task (SRT). *J Speech Lang Hear Res.* 2009; 52(5):1189–212. Epub 2009/07/29. doi: [10.1044/1092-4388\(2009/08-0047\)](https://doi.org/10.1044/1092-4388(2009/08-0047)) PMID: [19635944](https://pubmed.ncbi.nlm.nih.gov/19635944/); PubMed Central PMCID: PMC2930205.
  48. Wolf M, Denckla M. Test of rapid automatic naming. Austin: Pro-Ed; 2004.
  49. Woodcock RW. Woodcock Reading Mastery Tests—Revised. Minneapolis: Pearson; 1998.
  50. Torgesen JK, Wagner RK, Rashotte CA. Test of word reading efficiency. Austin: Pro-Ed; 1999.
  51. Wechsler D. Wechsler Individual Achievement Test—Second Edition. New York: Psychological Corporation; 2002.
  52. Button L, Peter B, Stoel-Gammon C, Raskind WH. Associations among measures of sequential processing in motor and linguistics tasks in adults with and without a family history of childhood apraxia of speech: a replication study. *Clinical linguistics & phonetics.* 2013; 27(3):192–212. doi: [10.3109/02699206.2012.744097](https://doi.org/10.3109/02699206.2012.744097) PMID: [23339292](https://pubmed.ncbi.nlm.nih.gov/23339292/); PubMed Central PMCID: PMC3875157.
  53. Dawson J, Tattersall P. Structured Photographic Articulation Test II: Janelle Publications; 2001.
  54. Wiig E, Secord W, Semel E. Clinical Evaluation of Language Fundamentals—Preschool. San Antonio: Psychological Corporation; 1992.
  55. Woodcock R, McGrew K, Mather N. Woodcock-Johnson Tests of Achievement. Itasca: Riverside Publishing; 2001.
  56. Peter B, Matsushita M, Raskind WH. Motor sequencing deficit as an endophenotype of speech sound disorder: a genome-wide linkage analysis in a multigenerational family. *Psychiatric genetics.* 2012; 22(5):226–34. doi: [10.1097/YPG.0b013e328353ae92](https://doi.org/10.1097/YPG.0b013e328353ae92) PMID: [22517379](https://pubmed.ncbi.nlm.nih.gov/22517379/); PubMed Central PMCID: PMC3407825.
  57. Edwards JH. Exclusion mapping. *J Med Genet.* 1987; 24(9):539–43. PMID: [3669048](https://pubmed.ncbi.nlm.nih.gov/3669048/); PubMed Central PMCID: PMC3407825.
  58. Chapman NH, Nato AQ Jr., Bernier R, Ankenman K, Sohi H, Munson J, et al. Whole exome sequencing in extended families with autism spectrum disorder implicates four candidate genes. *Human genetics.* 2015; 134(10):1055–68. doi: [10.1007/s00439-015-1585-y](https://doi.org/10.1007/s00439-015-1585-y) PMID: [26204995](https://pubmed.ncbi.nlm.nih.gov/26204995/); PubMed Central PMCID: PMC3722055.
  59. Ott J. Computer-simulation methods in human linkage analysis. *Proceedings of the National Academy of Sciences of the United States of America.* 1989; 86(11):4175–8. PMID: [2726769](https://pubmed.ncbi.nlm.nih.gov/2726769/); PubMed Central PMCID: PMC287412.
  60. Weeks DE, Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. *Am J Hum Genet.* 1990; 47(3):A204 (Supplement).
  61. Marchani EE, Chapman NH, Cheung CY, Ankenman K, Stanaway IB, Coon HH, et al. Identification of rare variants from exome sequence in a large pedigree with autism. *Human heredity.* 2012; 74(3–4):153–64. doi: [10.1159/000346560](https://doi.org/10.1159/000346560) PMID: [23594493](https://pubmed.ncbi.nlm.nih.gov/23594493/); PubMed Central PMCID: PMC3722055.
  62. Greenberg DA, Abreu P, Hodge SE. The power to detect linkage in complex disease by means of simple LOD-score analyses. *American journal of human genetics.* 1998; 63(3):870–9. doi: [10.1086/301997](https://doi.org/10.1086/301997) PMID: [9718328](https://pubmed.ncbi.nlm.nih.gov/9718328/); PubMed Central PMCID: PMC3722055.

63. Morton NE. Sequential tests for the detection of linkage. *Am J Hum Genet.* 1955; 7(3):277–318. Epub 1955/09/01. PMID: [13258560](#); PubMed Central PMCID: PMC1716611.
64. Ziegler A, König IR. A statistical approach to genetic epidemiology. 2nd ed. Weinheim: Wiley-VCH; 2010. xxiv, 497 p. p.
65. Kleensang A, Franke D, Alcais A, Abel L, Muller-Myhsok B, Ziegler A. An extensive comparison of quantitative trait Loci mapping methods. *Hum Hered.* 2010; 69(3):202–11. Epub 2010/03/06. doi: [10.1159/000289596](#) PMID: [20203525](#).
66. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet.* 2007; 81(3):559–75. PMID: [17701901](#)
67. Wigginton JE, Abecasis GR. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics.* 2005; 21(16):3445–7. Epub 2005/06/11. doi: [10.1093/bioinformatics/bti529](#) PMID: [15947021](#).
68. Nato AQ Jr., Chapman NH, Sohi HK, Nguyen HD, Brkanac Z, Wijsman EM. PBAP: a pipeline for file processing and quality control of pedigree data with dense genetic markers. *Bioinformatics.* 2015; 31(23):3790–8. doi: [10.1093/bioinformatics/btv444](#) PMID: [26231429](#); PubMed Central PMCID: PMCPMC4668752.
69. Matisse TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, et al. A second-generation combined linkage physical map of the human genome. *Genome research.* 2007; 17(12):1783–6. doi: [10.1101/gr.7156307](#) PMID: [17989245](#); PubMed Central PMCID: PMC2099587.
70. Tong L, Thompson E. Multilocus lod scores in large pedigrees: combination of exact and approximate calculations. *Human heredity.* 2008; 65(3):142–53. doi: [10.1159/000109731](#) PMID: [17934317](#); PubMed Central PMCID: PMC2701716.
71. Thompson E. The structure of genetic linkage data: from LIPED to 1M SNPs. *Human heredity.* 2011; 71(2):86–96. doi: [10.1159/000313555](#) PMID: [21734399](#); PubMed Central PMCID: PMC3136382.
72. Thompson EA, Lin S, Olshen AB, Wijsman EM. Monte Carlo analysis on a large pedigree. *Genetic epidemiology.* 1993; 10(6):677–82. doi: [10.1002/gepi.1370100658](#) PMID: [8314080](#).
73. Conneally PM, Edwards JH, Kidd KK, Lalouel JM, Morton NE, Ott J, et al. Report of the Committee on Methods of Linkage Analysis and Reporting. *Cytogenet Cell Genet.* 1985; 40(1–4):356–9. PMID: [3864600](#).
74. Krumm N, Sudmant PH, Ko A, O’Roak BJ, Malig M, Coe BP, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res.* 2012; 22(8):1525–32. Epub 2012/05/16. doi: [10.1101/gr.138115.112](#) PMID: [22585873](#); PubMed Central PMCID: PMC3409265.
75. Hach F, Hormozdiari F, Alkan C, Hormozdiari F, Birol I, Eichler EE, et al. mrsFAST: a cache-oblivious algorithm for short-read mapping. *Nat Meth.* 2010; 7(8):576–7. Epub 2010/08/03. doi: [10.1038/nmeth0810-576](#) PMID: [20676076](#); PubMed Central PMCID: PMC3115707.
76. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research.* 2007; 17(11):1665–74. doi: [10.1101/gr.6861907](#) PMID: [17921354](#); PubMed Central PMCID: PMC2045149.
77. Coin LJ, Asher JE, Walters RG, Moustafa JS, de Smith AJ, Sladek R, et al. cnvHap: an integrative population and haplotype-based multiplatform model of SNPs and CNVs. *Nature methods.* 2010; 7(7):541–6. doi: [10.1038/nmeth.1466](#) PMID: [20512141](#).
78. Kim S, Millard SP, Yu CE, Leong L, Radant A, Dobie D, et al. Inheritance model introduces differential bias in CNV calls between parents and offspring. *Genetic epidemiology.* 2012; 36(5):488–98. doi: [10.1002/gepi.21643](#) PMID: [22628073](#); PubMed Central PMCID: PMCPMC3678551.
79. Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. *Am J Hum Genet.* 2011; 88(2):173–82. Epub 2011/02/12. doi: [10.1016/j.ajhg.2011.01.010](#) PMID: [21310274](#); PubMed Central PMCID: PMC3035716.
80. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics.* 2002; 30(1):97–101. doi: [10.1038/ng786](#) PMID: [11731797](#).
81. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics.* 2010; 26(16):2069–70. doi: [10.1093/bioinformatics/btq330](#) PMID: [20562413](#); PubMed Central PMCID: PMC2916720.
82. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. *Nucleic acids research.* 2014; 42(Database issue):D749–55. doi: [10.1093/nar/gkt1196](#) PMID: [24316576](#); PubMed Central PMCID: PMC3964975.

83. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS computational biology*. 2013; 9(7):e1003153. doi: [10.1371/journal.pcbi.1003153](https://doi.org/10.1371/journal.pcbi.1003153) PMID: [23874191](https://pubmed.ncbi.nlm.nih.gov/23874191/); PubMed Central PMCID: PMC3715403.
84. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. 2010; 20(9):1297–303. doi: [10.1101/gr.107524.110](https://doi.org/10.1101/gr.107524.110) PMID: [20644199](https://pubmed.ncbi.nlm.nih.gov/20644199/); PubMed Central PMCID: PMC2928508.
85. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002; 30(17):3894–900. Epub 2002/08/31. PMID: [12202775](https://pubmed.ncbi.nlm.nih.gov/12202775/); PubMed Central PMCID: PMC137415.
86. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*. 2014; 46(3):310–5. doi: [10.1038/ng.2892](https://doi.org/10.1038/ng.2892) PMID: [24487276](https://pubmed.ncbi.nlm.nih.gov/24487276/); PubMed Central PMCID: PMC3992975.
87. Korvatska O, Strand NS, Berndt JD, Strovast T, Chen DH, Leverenz JB, et al. Altered splicing of ATP6AP2 causes X-linked parkinsonism with spasticity (XPDS). *Human molecular genetics*. 2013; 22(16):3259–68. doi: [10.1093/hmg/ddt180](https://doi.org/10.1093/hmg/ddt180) PMID: [23595882](https://pubmed.ncbi.nlm.nih.gov/23595882/); PubMed Central PMCID: PMC3723311.
88. Redies C, Hertel N, Hubner CA. Cadherins and neuropsychiatric disorders. *Brain research*. 2012; 1470:130–44. doi: [10.1016/j.brainres.2012.06.020](https://doi.org/10.1016/j.brainres.2012.06.020) PMID: [22765916](https://pubmed.ncbi.nlm.nih.gov/22765916/).
89. Balsamo J, Ernst H, Zanin MK, Hoffman S, Lilien J. The interaction of the retina cell surface N-acetylgalactosaminylphosphotransferase with an endogenous proteoglycan ligand results in inhibition of cadherin-mediated adhesion. *The Journal of cell biology*. 1995; 129(5):1391–401. PMID: [7775582](https://pubmed.ncbi.nlm.nih.gov/7775582/); PubMed Central PMCID: PMC2120464.
90. Shibata T, Shimoyama Y, Gotoh M, Hirohashi S. Identification of human cadherin-14, a novel neurally specific type II cadherin, by protein interaction cloning. *The Journal of biological chemistry*. 1997; 272(8):5236–40. PMID: [9030594](https://pubmed.ncbi.nlm.nih.gov/9030594/).
91. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. 2012; 489(7416):391–9. doi: [10.1038/nature11405](https://doi.org/10.1038/nature11405) PMID: [22996553](https://pubmed.ncbi.nlm.nih.gov/22996553/); PubMed Central PMCID: PMC4243026.
92. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American journal of human genetics*. 2009; 84(4):524–33. doi: [10.1016/j.ajhg.2009.03.010](https://doi.org/10.1016/j.ajhg.2009.03.010) PMID: [19344873](https://pubmed.ncbi.nlm.nih.gov/19344873/); PubMed Central PMCID: PMC2667985.
93. Liu KC, Jacobs DT, Dunn BD, Fanning AS, Cheney RE. Myosin-X functions in polarized epithelial cells. *Molecular biology of the cell*. 2012; 23(9):1675–87. doi: [10.1091/mbc.E11-04-0358](https://doi.org/10.1091/mbc.E11-04-0358) PMID: [22419816](https://pubmed.ncbi.nlm.nih.gov/22419816/); PubMed Central PMCID: PMC3338435.
94. Yu H, Wang N, Ju X, Yang Y, Sun D, Lai M, et al. PtdIns (3,4,5) P3 recruitment of Myo10 is essential for axon development. *PloS one*. 2012; 7(5):e36988. doi: [10.1371/journal.pone.0036988](https://doi.org/10.1371/journal.pone.0036988) PMID: [22590642](https://pubmed.ncbi.nlm.nih.gov/22590642/); PubMed Central PMCID: PMC3349655.
95. Watrin E, Schleiffer A, Tanaka K, Eisenhaber F, Nasmyth K, Peters JM. Human Scc4 is required for cohesin binding to chromatin, sister-chromatid cohesion, and mitotic progression. *Current biology*: CB. 2006; 16(9):863–74. doi: [10.1016/j.cub.2006.03.049](https://doi.org/10.1016/j.cub.2006.03.049) PMID: [16682347](https://pubmed.ncbi.nlm.nih.gov/16682347/).
96. Krantz ID, McCallum J, DeScipio C, Kaur M, Gillis LA, Yaeger D, et al. Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of *Drosophila melanogaster* Nipped-B. *Nature genetics*. 2004; 36(6):631–5. doi: [10.1038/ng1364](https://doi.org/10.1038/ng1364) PMID: [15146186](https://pubmed.ncbi.nlm.nih.gov/15146186/).
97. St Pourcain B, Wang K, Glessner JT, Golding J, Steer C, Ring SM, et al. Association between a high-risk autism locus on 5p14 and social communication spectrum phenotypes in the general population. *Am J Psychiatry*. 2010; 167(11):1364–72. doi: [10.1176/appi.ajp.2010.09121789](https://doi.org/10.1176/appi.ajp.2010.09121789) PMID: [20634369](https://pubmed.ncbi.nlm.nih.gov/20634369/); PubMed Central PMCID: PMC2667985.
98. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009; 459(7246):528–33. doi: [10.1038/nature07999](https://doi.org/10.1038/nature07999) PMID: [19404256](https://pubmed.ncbi.nlm.nih.gov/19404256/); PubMed Central PMCID: PMC2943511.
99. Tierney C, Mayes S, Lohs SR, Black A, Gisin E, Veglia M. How Valid Is the Checklist for Autism Spectrum Disorder When a Child Has Apraxia of Speech? *J Dev Behav Pediatr*. 2015; 36(8):569–74. doi: [10.1097/DBP.000000000000189](https://doi.org/10.1097/DBP.000000000000189) PMID: [26114615](https://pubmed.ncbi.nlm.nih.gov/26114615/).
100. Mottis A, Mouchiroud L, Auwerx J. Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes & development*. 2013; 27(8):819–35. doi: [10.1101/gad.214023.113](https://doi.org/10.1101/gad.214023.113) PMID: [23630073](https://pubmed.ncbi.nlm.nih.gov/23630073/); PubMed Central PMCID: PMC3650221.

101. Nickerson ML, Warren MB, Toro JR, Matrosova V, Glenn G, Turner ML, et al. Mutations in a novel gene lead to kidney tumors, lung wall defects, and benign tumors of the hair follicle in patients with the Birt-Hogg-Dube syndrome. *Cancer cell*. 2002; 2(2):157–64. PMID: [12204536](#).
102. Frank V, Habbig S, Bartram MP, Eisenberger T, Veenstra-Knol HE, Decker C, et al. Mutations in NEK8 link multiple organ dysplasia with altered Hippo signalling and increased c-MYC expression. *Human molecular genetics*. 2013; 22(11):2177–85. doi: [10.1093/hmg/ddt070](#) PMID: [23418306](#).
103. Nagase T, Ishikawa K, Suyama M, Kikuno R, Hirose M, Miyajima N, et al. Prediction of the coding sequences of unidentified human genes. XII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA research: an international journal for rapid publication of reports on genes and genomes*. 1998; 5(6):355–64. PMID: [10048485](#).
104. Tsien RW, Ellinor PT, Horne WA. Molecular diversity of voltage-dependent Ca<sup>2+</sup> channels. *Trends in pharmacological sciences*. 1991; 12(9):349–54. PMID: [1659003](#).
105. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nature genetics*. 2008; 40(9):1056–8. doi: [10.1038/ng.209](#) PMID: [18711365](#); PubMed Central PMCID: PMC2703780.
106. Moreira MC, Klur S, Watanabe M, Nemeth AH, Le Ber I, Moniz JC, et al. Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nature genetics*. 2004; 36(3):225–7. doi: [10.1038/ng1303](#) PMID: [14770181](#).
107. Deniz Can D, Richards T, Kuhl PK. Early gray-matter and white-matter concentration in infancy predict later language skills: a whole brain voxel-based morphometry study. *Brain Lang*. 2013; 124(1):34–44. Epub 2013/01/01. doi: [10.1016/j.bandl.2012.10.007](#) PMID: [23274797](#); PubMed Central PMCID: PMC3551987.
108. Koziol LF, Budding D, Andreasen N, D'Arrigo S, Bulgheroni S, Imamizu H, et al. Consensus Paper: The Cerebellum's Role in Movement and Cognition. *Cerebellum*. 2013. Epub 2013/09/03. doi: [10.1007/s12311-013-0511-x](#) PMID: [23996631](#).
109. Pliatsikas C, Johnstone T, Marinis T. Grey Matter Volume in the Cerebellum is Related to the Processing of Grammatical Rules in a Second Language: A Structural Voxel-based Morphometry Study. *Cerebellum*. 2013. Epub 2013/08/31. doi: [10.1007/s12311-013-0515-6](#) PMID: [23990323](#).
110. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nature reviews Genetics*. 2015. doi: [10.1038/nrg3871](#) PMID: [25645873](#).
111. Korvatska O, Leverenz JB, Jayadev S, McMillan P, Kurtz I, Guo XD, et al. R47H Variant of TREM2 Associated With Alzheimer Disease in a Large Late-Onset Family Clinical, Genetic, and Neuropathological Study. *Jama Neurol*. 2015; 72(8):920–7. doi: [10.1001/jamaneurol.2015.0979](#) WOS:000359404000013. PMID: [26076170](#)
112. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature genetics*. 2011; 43(6):585–9. doi: [10.1038/ng.835](#) PMID: [21572417](#); PubMed Central PMCID: PMC3115696.
113. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, et al. Refinement and discovery of new hotspots of copy-number variation associated with autism spectrum disorder. *American journal of human genetics*. 2013; 92(2):221–37. doi: [10.1016/j.ajhg.2012.12.016](#) PMID: [23375656](#); PubMed Central PMCID: PMC3567267.