BRAIN
ORIGINAL ARTICLE

The clinical and genetic spectrum of paediatric speech and language disorders

Jan H. Magielski,[1,2,3] Sarah M. Ruggiero,[1,2] Julie Xian,[1,2,3] Shridhar Parthasarathy,[1,2,3] Peter D. Galer,[1,2,3,4] Shiva Ganesan,[1,2,3] Amanda Back,[1,2] Jillian L. McKee,[1,2,3,5] Ian McSalley,[1,2,3] Alexander K. Gonzalez,[3] Angela Morgan,[6,7] Joseph Donaher[8,9] and Ingo Helbig[1,2,3,5]

Speech and language disorders are known to have a substantial genetic contribution. Although frequently examined as components of other conditions, research on the genetic basis of linguistic differences as separate phenotypic subgroups has been limited so far.

Here, we performed an in-depth characterization of speech and language disorders in 52 143 individuals, reconstructing clinical histories using a large-scale data-mining approach of the electronic medical records from an entire large paediatric healthcare network.

The reported frequency of these disorders was the highest between 2 and 5 years old and spanned a spectrum of 26 broad speech and language diagnoses. We used natural language processing to assess the degree to which clinical diagnoses in full-text notes were reflected in ICD-10 diagnosis codes. We found that aphasia and speech apraxia could be retrieved easily through ICD-10 diagnosis codes, whereas stuttering as a speech phenotype was coded in only 12% of individuals through appropriate ICD-10 codes. We found significant comorbidity of speech and language disorders in neurodevelopmental conditions (30.31%) and, to a lesser degree, with epilepsies (6.07%) and movement disorders (2.05%). The most common genetic disorders retrievable in our analysis of electronic medical records were *STXBP1* ($n = 21$), *PTEN* ($n = 20$) and *CACNA1A* ($n = 18$). When assessing associations of genetic diagnoses with specific linguistic phenotypes, we observed associations of *STXBP1* and aphasia ($P = 8.57 \times 10^{-7}$, 95% confidence interval = 18.62–130.39) and *MYO7A* with speech and language development delay attributable to hearing loss ($P = 1.24 \times 10^{-5}$, 95% confidence interval = 17.46–infinity). Finally, in a sub-cohort of 726 individuals with whole-exome sequencing data, we identified an enrichment of rare variants in neuronal receptor pathways, in addition to associations of *UQCRC1* and *KIF17* with expressive aphasia, *MROH8* and *BCHE* with poor speech, and *USP37*, *SLC22A9* and *UMODL1* with aphasia.

In summary, our study outlines the landscape of paediatric speech and language disorders, confirming the phenotypic complexity of linguistic traits and novel genotype–phenotype associations. Subgroups of paediatric speech and language disorders differ significantly with respect to the composition of monogenic aetiologies.

1 Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
2 The Epilepsy NeuroGenetics Initiative (ENGIN), Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
3 Department of Biomedical and Health Informatics (DBHi), Children's Hospital of Philadelphia, Philadelphia, PA 19146, USA
4 Center for Neuroengineering and Therapeutics, University of Pennsylvania, Philadelphia, PA 19104, USA
5 Department of Neurology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA
6 Murdoch Children's Research Institute, Parkville, VIC 3052, Australia
7 Department of Audiology and Speech Pathology, University of Melbourne, Parkville, VIC 3052, Australia
8 Center for Childhood Communication, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
9 Department of Otorhinolaryngology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA

Received March 29, 2024. Revised July 11, 2024. Accepted July 18, 2024. Advance access publication October 16, 2024

Correspondence to: Ingo Helbig, MD
Division of Neurology, Children's Hospital of Philadelphia
Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
E-mail: helbigi@email.chop.edu

Keywords: speech disorder; language disorder; genetics; electronic medical records; Human Phenotype Ontology

Introduction

Speech and language differences are common clinical features associated with neurodevelopmental disorders. Differences in the neurological basis of communication have been characterized in individuals with specific neurodevelopmental conditions, including rare genetic disorders such as *GRIN2A*-related disorders, *FOXP2*-related disorders and *STXBP1*-related disorders.[1-4] There has been promising recent work that has identified novel monogenic and polygenic aetiologies of speech disorders.[5-7] However, there is still much of the genetic landscape to be elucidated. Accordingly, this represents a major gap in our understanding of speech and language disorders, given their presumed genetic component.[8,9]

With the widespread use of electronic medical records (EMRs), it becomes possible to study systematically conditions that have not yet received significant attention previously. In addition to making it possible to analyse data on these conditions at scale, EMRs allow for the analysis of clinical data over time. For speech disorders in children, this longitudinal component is particularly important, given the dynamic nature of neurodevelopment in childhood and adolescence. Hence, as there remains a need to characterize the full clinical spectrum of individuals with communication disorders and the underlying genetic aetiology that impacts differences in speech and language development, EMR-based approaches offer unprecedented opportunities to conduct targeted deep phenotypic analyses at scale.[10,11]

Paediatric speech disorders that have been investigated in the context of their genetic aetiologies include childhood apraxia of speech, childhood dysarthria and stuttering.[5,8] *FOXP2* was the first gene discovered to be associated with specific speech impairments, namely speech apraxia and dysarthria.[12-14] Since this characterization, a variety of genetic aetiologies have been suggested to be associated with neurobiological disruptions of speech and language, but these studies often lack the statistical support that is now available through our increased understanding of population genetics and the development of human genome databases.

Here, we used the wealth of information captured in the EMRs at a large paediatric specialty care network, including robust primary care, speech–language pathology, developmental and neurology departments and clinics, to retrieve and reconstruct the longitudinal clinical histories of 52 143 individuals with documented speech and language disorders. A subset of analysis was done on targeted epilepsy and neurogenetics cohorts. We tracked clinical features over time across cohorts and developed a framework for the prediction and identification of clinical subgroups with shared trajectories, allowing us to identify previously unrecognized clinical patterns and to build a more comprehensive understanding of the prevalence and landscape of communication disorders.

Materials and methods

Study inclusion and setting

The study was performed at the Children's Hospital of Philadelphia through the analysis of EMRs. We selected a group of the relevant International Classification of Diseases, Tenth Revision (ICD-10) codes [F01–F99 (mental, behavioural and neurodevelopmental disorders); G00–G99 (diseases of the nervous system); R25–R29 (symptoms and signs involving the nervous and musculoskeletal systems); R47–R49 (symptoms and signs involving speech and voice); R62 (lack of expected normal physiological development in childhood and adults); Z13 (encounter for screening for other diseases and disorders); Z14–Z15 (genetic carrier and genetic susceptibility to disease); Z81 (family history of mental and behavioural disorders); Z84 (family history of other conditions); and I69 (sequelae of cerebrovascular disease)] to define a broad neurological cohort.[15] Subsequently, we compiled a list of ICD-10 codes describing speech phenotype-related diagnoses (F80 and R47–R49) to delineate our speech cohort (Supplementary Table 1). Within this group, we then analysed ICD-10 codes that co-occurred with speech ICD-10 codes to assess their comorbidity with other neurological diagnoses: neurodevelopmental disorders (F84, F88 and F89), epilepsy (G40) and movement disorders (G20–G26). We were able to extract the genetic diagnoses of individuals from the broad neurological cohort from the dedicated ICD-10 code (Z15.89). We were able to access and retrieve these ICD-10 code data for individuals who were seen at the Children's Hospital of Philadelphia since 2015.

With regard to speech motor disorders, we focused particularly on three, namely speech apraxia, speech dysarthria and stuttering. These conditions all fall under the subdivision of motor/neurological speech disorders, according to the classification of the American Speech–Language–Hearing Association.[15] Speech apraxia is characterized by a difficulty with producing sounds needed for correct pronunciation and an inability to use prosody appropriately in the absence of muscle weakness.[6] This disorder, however, can co-occur with dysarthria, a condition associated with neuromuscular issues, such as abnormal tone, spasticity or ataxia, which makes the production of comprehensive speech more difficult.[6,16] Lastly, stuttering is a block in speech fluency that includes features such as repetitions, prolongations and blocks during fluent speech.[17]

Patient cohorts and data extraction

In the sub-cohort composed of individuals from the Pediatric Epilepsy Learning Health System (PELHS) and Epilepsy Genetics Research Project (EGRP), we analysed charts from all encounters; PELHS containing de-identified EMR data of individuals who were seen in our healthcare network and received an epilepsy diagnosis, and EGRP with paediatric patients who are known or believed to have a genetic epilepsy or neurodevelopmental disorder.

The study was conducted in accordance with the Declaration of Helsinki. Data extraction was performed under the following

Children Hospital of Philadelphia Institutional Review Board protocols: #15-12226 (EGRP) and #20-017641 [The Children's Hospital of Philadelphia Neuroscience Electronic Health Record Data Science Study (NeuroEHR)]. Individuals in this study were recruited from the Epilepsy Neurogenetics Initiative clinic, a programme that evaluates individuals with neurodevelopmental disorders and epilepsy. Study enrolment facilitated biospecimen collection and clinical genetic testing raw data release, allowing for information from the clinical charts and whole-exome sequencing data to be available for the EGRP sub-cohort in our analysis. In the PELHS sub-cohort, deidentified data from the EMRs of individuals who had an epilepsy diagnosis within the Children's Hospital of Philadelphia were made available. The data from the EGRP sub-cohort were collected since 2014, and the data from the PELHS sub-cohort were collected since 2010.

We extracted phenotypic data using Clinical Text and Knowledge Extraction System, a natural language processing tool, that were then mapped onto the Human Phenotype Ontology (HPO) terms.[18] This was performed independently from the ICD-10 extraction. By using a well-established controlled dictionary of HPO, we were able not only to record phenotypic information in a standardized computable manner, but also to harmonize our dataset, as used by our group in the past.[19,20] For example, if a chart of a given individual contained information about stuttering (HP:0025268), this framework enabled us to reason that the individual also had 'abnormality of speech or vocalization' (HP:0002167). Such methods allowed us simultaneously to capture broad and granular phenotypic information, ensuring a thorough phenotypic picture for each individual.

EMR visibility index

Furthermore, we developed a new EMR visibility index by comparing the frequency of clinical speech diagnoses based on the ICD-10 codes against the frequency of speech disorders mentioned in the full-text clinical notes that were mapped onto HPO terms. We developed this new measure in response to the need to capture as much EMR data as possible, while accounting for the 'blind spots' of this method by identifying clinical groups that tend to be undercharacterized because of low visibility in the medical charts. This disparity is particularly important in rare disease communities, who frequently advocate for the creation of new ICD-10 codes for rare conditions in order that providers and researchers might reliably track individuals with a given disorders across institutions.[21] The EMR visibility index allowed us to identify the extent of visibility for neurological disorders and speech impairment diagnostic codes, and how their visibility changes depending on the depth of phenotypic analysis. To that end, in the PELHS sub-cohort, we counted the individuals with seizures, speech apraxia, aphasia, autism, intellectual disability, attention deficit hyperactivity disorder and stuttering who had their diagnosis recorded in ICD-10 codes and divided that number by the number of individuals that had the diagnosis coded in their medical charts in HPO. This proportion gave us the EMR visibility index.

Data abstraction and genomic analysis

The documentation and analysis of neurological features associated with speech and language disorders was facilitated through clinical data captured in EMRs. Data collected included clinical diagnoses (ICD-10 codes), phenotypic features, neurodevelopmental histories and genetic findings and diagnoses. Clinical features were mapped across the age span for all individuals. Furthermore, in 726 individuals from the EGRP sub-cohort, among whom 541 individuals had a speech phenotype, we analysed raw exome data from whole-exome sequencing. Within the EGRP sub-cohort, 52% of individuals were male. The observed racial distributions were as follows: White, 62%; other, 16%; Black, 14%; multi-racial, 4%; Asian, 3%; Indian, 0.5%; and unknown or refused, 0.5%. Additionally, Hispanic or Latino ethnicity was seen in 13% of the EGRP sub-cohort. Whole-exome sequencing data were obtained from the internal Children Hospital of Philadelphia's real-world dataset, which contained both clinical and research-basis sequencing data. The raw data were aligned to the GRCh38 human reference genome using the Burrows Wheeler Alignment (v.0.7.12) MEM algorithm. After alignment for each sample, mate tags (MC and MQ) were added to the paired-end lines with Samblaster (v.0.1.20). Joint genotyping and variant calling were performed according to the Genome Analysis Toolkit (GATK) best practices (v.4.2.6.1). Variants with low genotyping quality were filtered based on GATK variant quality score recalibration. The filtered variants were annotated using Ensembl Variant Effect Predictor (VEP v.107.0). Variants that had <0.005 Genome Aggregation Database (gnomAD) frequency were classified into three groups: Class 1, protein-truncating variants (PTVs) with a probability of loss-of-function (pLI) score >0.95; Class 2, combined annotation dependent depletion (CADD) score, >20 and residual variation intolerance score (RVIS), <65; and Class 3, protein-truncating variants and missense combined. For individuals who had an established genetic diagnosis, only the gene from the genetic diagnosis was used in the analyses to avoid obtaining spurious relationships between other variants that these individuals had with specific speech phenotypes for which their genetic diagnosis would account. Using the Online Mendelian Inheritance in Man (OMIM) database, we identified genes with and without known phenotypic associations.[22] Variants without known phenotypic associations were then analysed further by evaluating the frequency in the gnomAD population database, and through Integrative Genomics Viewer (IGV) to assess reliable alignment of the exome sequencing reads.[23,24] Furthermore, variants were filtered based on the RVIS to help in prioritizing functional relevance.[25] Lastly, we leveraged the Database for Annotation, Visualization and Integrated Discovery (DAVID) 2021 bioinformatics resources to gain a better understanding of possible functional and physiological correlates within our findings.[26] Reactome Pathway annotation results were used and, following the DAVID guidelines, looked at the fold enrichment level of >1.5.[27,28] If this condition was met, we also analysed whether a nominal *P*-value was significant (<0.05) for a given association within the DAVID analysis and whether at least five genes were present in a given pathway. Subsequently, we explored the associations of such genes to speech phenotypes in our genotype–phenotype analysis.

Statistical analysis

All statistical analyses were conducted using the R statistical framework.[29] Statistical testing of associations of Fisher's exact test is reported with correction for multiple comparisons using a false discovery rate (FDR) of 5%. If statistical significance was not achieved following correction for multiple comparisons, results were described using their respective odds ratios (OR), with 95% confidence intervals (CI) provided. To assess the similarity of clinical sub-groups within the speech cohort in addition to those with and without a genetic diagnosis in the speech and language cohort, Welch's two-sample *t*-test was performed. For the genomic portion

analysis of the associations between the variants and speech phenotypes, we used logistic regression, with the covariates of age, biological sex, race and ethnicity; the testing was done with an FDR correction of 5%. Apart from the FDR of 5%, in the analysis of associations between specific variants and speech and language phenotypes, only variants that were seen in at least two individuals were designated as significant.

Results

Speech and language disorders span a wide range of clinical diagnoses

In a broad paediatric cohort of 5 519 989 encounters from 265 926 individuals with a neurological diagnosis, based on 26 ICD-10 codes, we identified 1 671 257 encounters across 52 143 individuals with speech and language disorders, spanning a total of 203 150 patient-years (Fig. 1). Among these individuals, we found that the most common speech-related ICD-10 diagnoses were mixed receptive–expressive language disorder (F80.2; $n = 27\,057$ individuals), developmental disorder of speech and language, unspecified (F80.9; $n = 17\,579$ individuals) and expressive language disorder (F80.1; $n = 9865$ individuals). These diagnoses were followed by functional speech sound disorders: phonological disorder (F80.0; $n = 6060$ individuals) and dysphonia (R49.0; $n = 3184$ individuals). The five most common speech disorders accounted for more than four-fifths (81.53%) of all speech diagnoses in the cohort. For motor speech disorders with a presumed genetic basis, speech apraxia (R48.2) was seen in 1099 individuals, stuttering (F80.81) in 1684 individuals, and dysarthria in 1056 individuals (R47.1); ICD-10 codes for these disorders represented 4.91% of all speech and language-specific phenotypes. We also observed that among speech and language phenotypes, speech apraxia and aphasia had the highest EMR visibility indices (0.74 and 0.52, respectively), while stuttering had the lowest EMR visibility index,0.12 (Fig. 2B).

The landscape of speech and language disorders is characterized by age-related phenotypes

We observed that speech phenotype-related diagnoses were most prevalent in the second year of life, with the majority of speech and language diagnoses made between 2 and 5 years of age and with the highest frequencies seen at 2 years of age (0.173, $n = 10\,938$ individuals), 1 year of age (0.134, $n = 7924$ individuals) and 3 years of age (0.109, $n = 6767$ individuals). After 3 years of age, the frequency of speech phenotype-related diagnoses dropped dramatically and was found in <10% of all individuals. Within the sub-cohorts of individuals who experience stuttering, speech apraxia and dysarthria, we observed that the highest frequency still occurred within the 2–5 years old window but slightly later than in the case of paediatric speech and language phenotypes at large (Fig. 2A). The frequency of individuals diagnosed with stuttering (frequency in the broad neurological cohort = 0.0141) or dysarthria (frequency in the broad neurological cohort = 3.95×10^{-4}) reached its peak at 3–4 years of age. Apraxia diagnoses reached its peak at 2–3 years of age (frequency in the broad neurological cohort = 2.09×10^{-3}). We found that 90% of individuals with a speech abnormality received their first speech and/or language diagnosis by 10.77 years of age.

Speech disorders overlap with neurodevelopmental disorders, epilepsies and movement disorders

Examining ICD-10 code diagnoses co-occurring with speech and language phenotypes, we assessed the landscape of speech and language disorders relative to other neurological and psychiatric diseases. We observed the strongest overlap with neurodevelopmental diagnoses: among 52 143 individuals with a speech diagnosis, 15 806 (30.31%; $P < 2.2 \times 10^{-16}$, OR 6.57, CI 6.40–6.74) also had a neurodevelopmental diagnosis. In our speech cohort, the most frequent co-existing developmental disorders were autism (F84.0, $n = 11\,940$) and other disorders of psychological development (F88, $n = 7239$). Epilepsy was found to be the second-most substantial comorbidity ($n = 3080$, 6.07%; $P = 0.0132$, OR 1.05, CI 1.01–1.10) among the broad neurological disorders, with the following most prevalent phenotypes: G40.909: epilepsy, unspecified ($n = 1587$); G40.209: focal epilepsy with complex focal seizures ($n = 896$); and G40.109: focal epilepsy with simple partial seizures ($n = 847$). Lastly, we investigated the overlap between speech and movement disorders, which represented 2.05% of comorbidities ($n = 1070$; $P = 0.443$, OR 0.97, CI 0.91–1.04). Among these, G24.9: dystonia, unspecified ($n = 290$) was the most frequent, followed by G25.3: myoclonus ($n = 214$) and G25.81: restless legs syndrome ($n = 172$; Fig. 3).

Next, we analysed how these broader co-existing phenotypes related to the age at which the first speech and language diagnoses were made. In the subgroup with comorbid speech and epilepsy diagnoses, 90% of individuals had a speech diagnosis documented by 14.6 years (mean age of diagnosis = 6.80 years), while for those in the speech and language cohort without an epilepsy diagnosis, that age was 10 years (mean age of diagnosis = 4.44 years); this difference in the diagnosis age distributions was also captured in Welch's two-sample *t*-test ($P < 2.2 \times 10^{-16}$). Conversely, in the speech–neurodevelopmental sub-cohort including individuals with co-occurring speech and neurodevelopmental disorders, 90% of individuals received their speech diagnosis at 10.2 years (mean age of diagnosis = 4.62 years), in comparison to 10.7 years for the individuals presenting with a speech phenotype, but without a neurodevelopmental disorder (mean age of diagnosis = 4.56 years). The difference between the mean age of diagnosis was not significant between the two groups ($P = 0.096$). The data might be limited by the under-documentation of speech phenotypes or the lack of availability of the entirety of the EMR data through one healthcare network system.

Specific speech and language phenotypes are associated with various genetic aetiologies

We next investigated the landscape of genetic diagnoses in our speech cohort. We found 273 unique genetic diagnoses found in at least one individual, and a total of 607 individuals (1.16%) with a genetic diagnosis. Analysis of the cumulative onset of age at which speech diagnoses were first reflected in the EMR demonstrated that 90% of individuals with both a speech/language and genetic diagnosis had documentation of both diagnoses by 12.0 years (mean age = 5.23 years). The accrual of speech diagnosis occurred slightly later in comparison to individuals without a genetic diagnosis (90% at 10.5 years, mean age = 4.57 years; Fig. 4); the distribution of speech diagnosis age was significantly different between the two groups, as evidenced by Welch's two-sample *t*-test ($P = 0.0002$). The most common genetic diagnoses included *STXBP1* ($n = 21$), *PTEN* ($n = 20$), *CACNA1A* ($n = 18$), *SCN2A* ($n = 14$) and *SYNGAP1* ($n = 11$). We next explored more granular

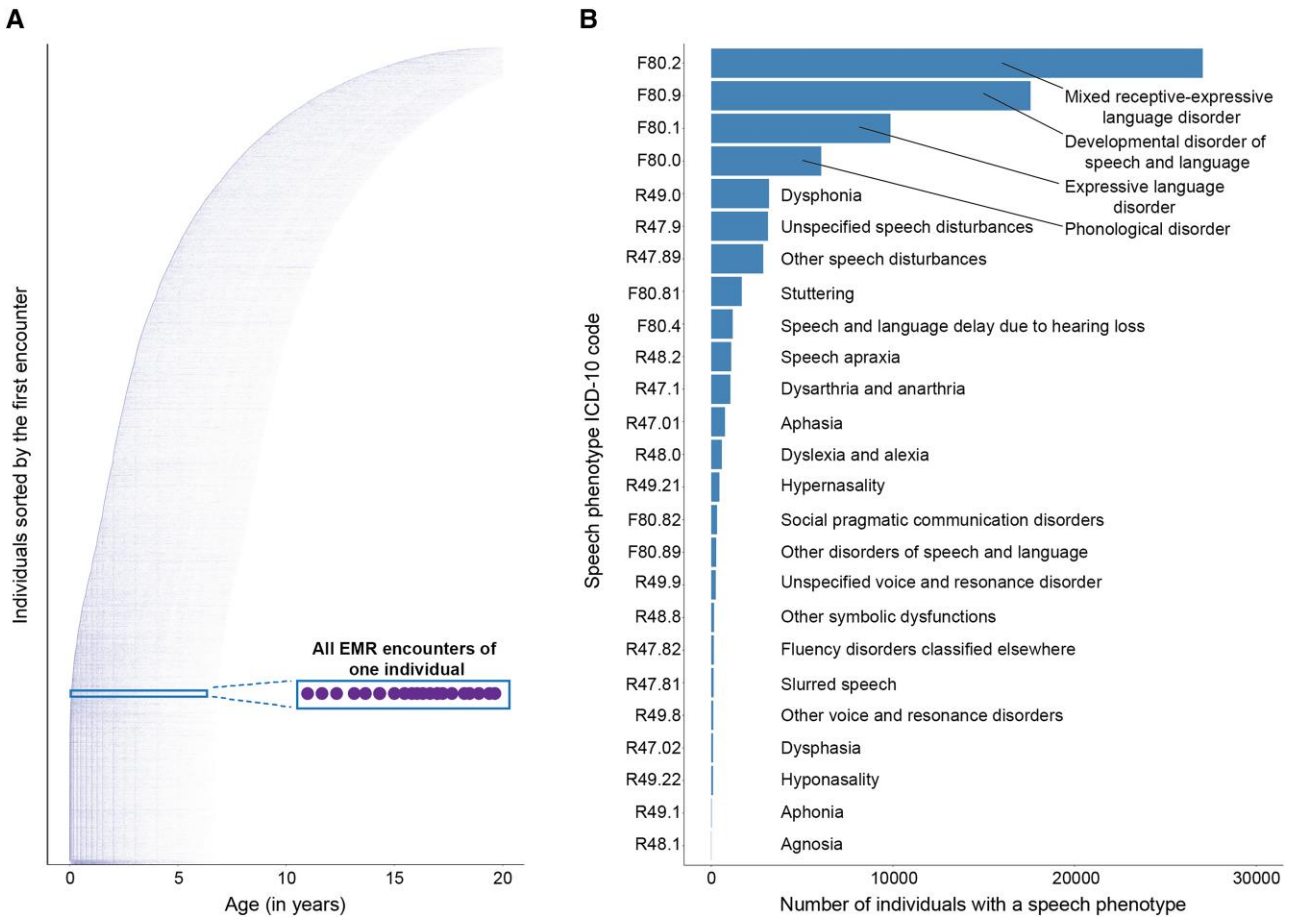

Figure 1 Overview of the speech cohort. (**A**) There were 1 671 257 encounters across 52 143 individuals with speech disorders, including data from a total of 203 150 patient years. (**B**) Distribution of speech diagnoses. EMR = electronic medical records.

relationships between specific speech and language disorder types and genetic diagnoses. After correcting for multiple testing, the following relationships were significant: *STXBP1* with aphasia ($P = 8.57 \times 10^{-12}$, OR 50.23, CI 18.62–130.39) and *MYO7A* with other developmental disorders of speech and language ($P = 1.24 \times 10^{-5}$, OR infinity, CI 17.46–infinity). The nominally significant relationships with the highest level of significance included *GRIN2A* with speech apraxia ($P = 3.3 \times 10^{-4}$, OR 34.06, CI 4.98–201.11), *MECP2* with other developmental disorders of speech and language ($P = 9.81 \times 10^{-4}$, OR 54.02, IC 5.45–284.24) and *POLG* ($P = 0.0013$, OR 65.87, CI 4.77–898.38) with aphasia (Fig. 5 and Table 1).

Exome sequencing analysis also shows there is an underlying genetic component to speech and language disorders

As expected, analysis of exome sequencing data in 726 individuals revealed a variety of rare variants. In total, we found 212 PTVs (Class 1), 6355 variants with CADD score > 20 and RVIS < 65 (Class 2), and 15 181 variants in the combined PTV–missense group (Class 3); 95 (13.09%) individuals had a clinically verified genetic diagnosis. We observed that variants in the following genes were significantly associated with speech and language phenotypes after correction for multiple testing: *UQCRC1* and expressive aphasia, *MROH8* and poor speech, *KIF17* and expressive aphasia, *BCHE* and poor speech, *USP37* and aphasia, *SLC22A9* and aphasia, and *UMODL1* and aphasia (Fig. 6, Table 2 and Supplementary Table 2). No PTVs showed nominally significant relationships with speech and/or language clinical features.

Genes contributory to speech disorders cluster in neurologically relevant pathways

We broadly entered all nominally significant Class 2 variants ($n = 925$) into DAVID in a manner consistent with prior studies, allowing for a meaningful integrative analysis and a reliable assessment of enrichment relative to background.[27] From this analysis, we found that genes involved in the pathways 'MECP2 regulates neuronal receptors and channels' (*GRIN2A*, *GRIN2B*, *TRPC3*, *FKBP5* and *SIN3A*) were most enriched (fold enrichment = 5.6, $P = 0.01$). Other neurologically relevant pathways that showed enrichment were 'transcriptional regulation by MECP2' (fold enrichment = 3.7, $P = 3.2 \times 10^{-4}$), 'NCAM1 interactions' (fold enrichment = 3.6, $P = 5.6 \times 10^{-3}$), 'netrin-1 signalling' (fold enrichment = 3.4, $P = 0.0041$) and 'NCAM signalling for neurite out-growth' (fold enrichment = 2.7, $P = 0.016$). The other enriched pathways spanned the three main physiological processes: cellular structure, cellular interactions and metabolism (Supplementary Table 3).

Discussion

In this study, we conducted a comprehensive analysis of the landscape of paediatric speech and language disorders, leveraging

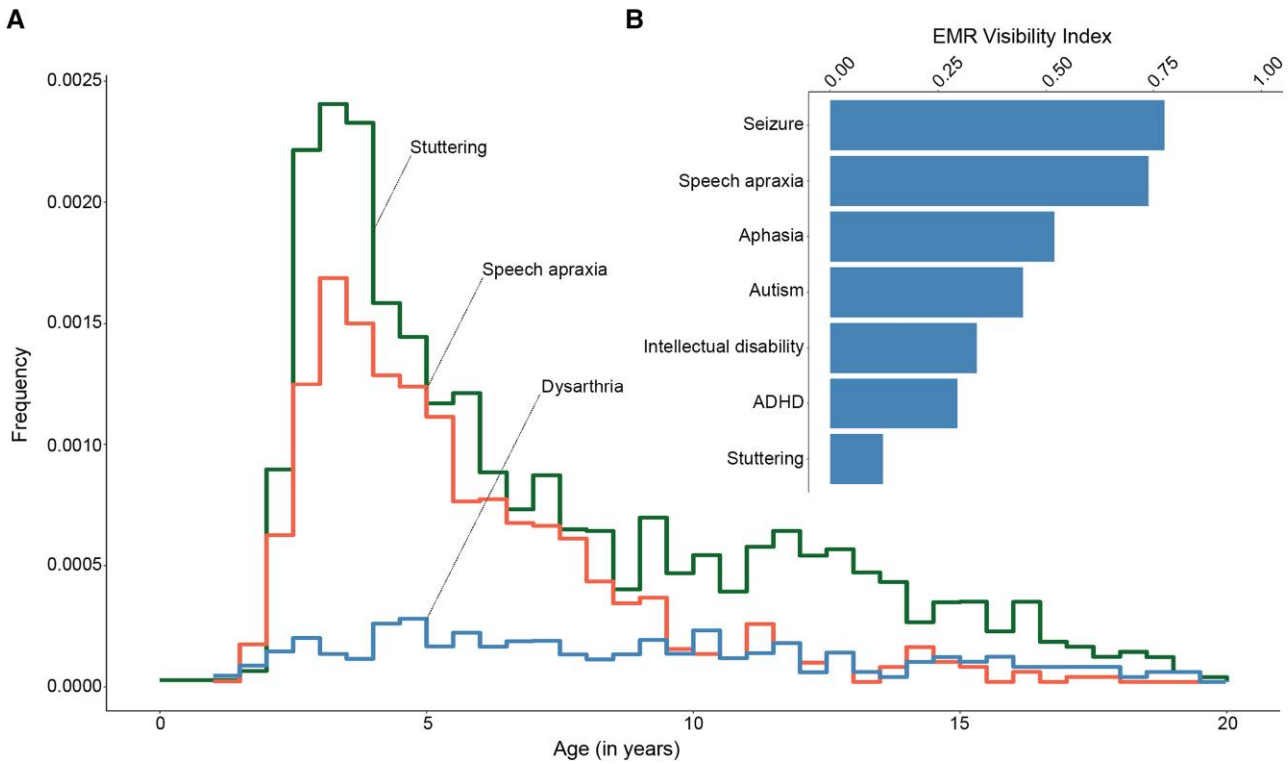

Figure 2 Frequency of specific speech phenotype diagnoses. (**A**) The frequency of stuttering, speech apraxia and dysarthria diagnoses in 265 926 patients with a neurological diagnosis. (**B**) Electronic medical records (EMR) visibility index plot.

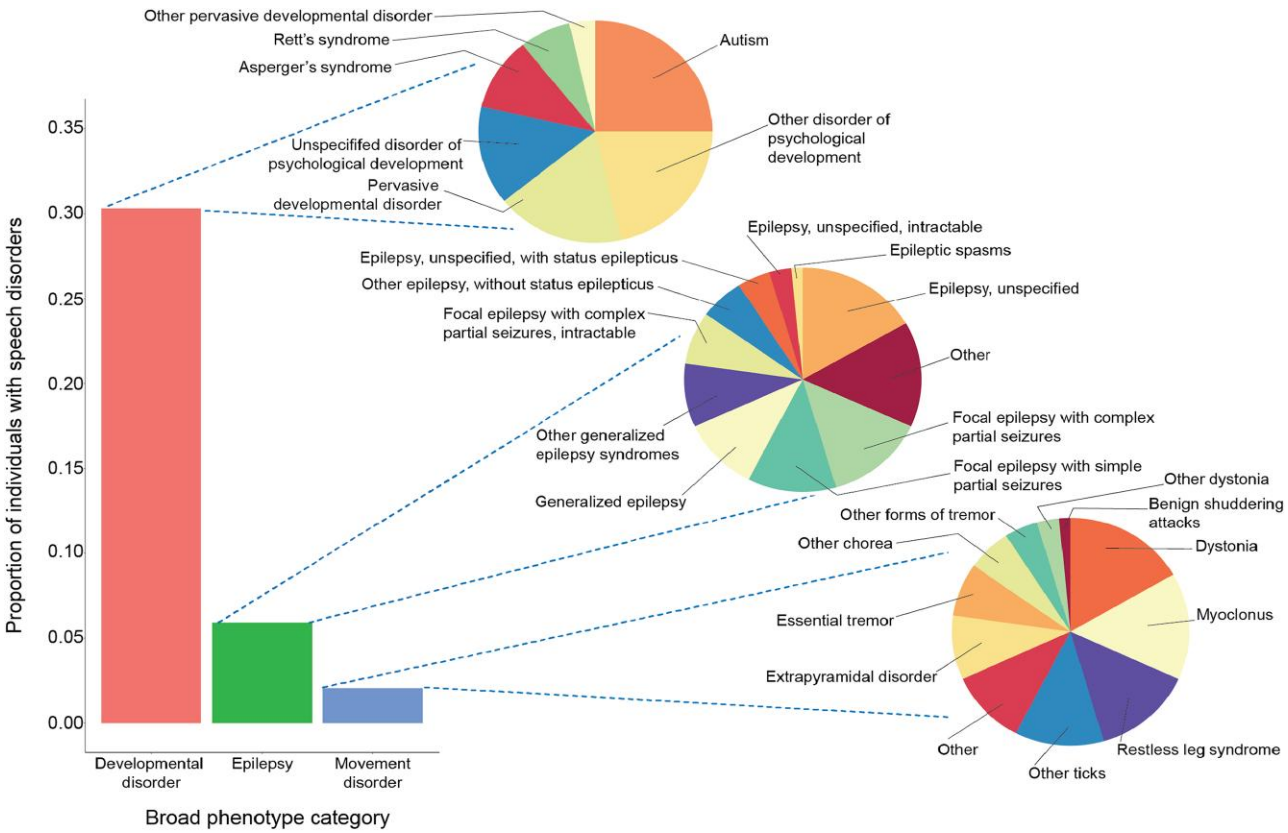

Figure 3 Diagnoses comorbid with speech phenotypes. Broad phenotypic categories co-occurring with speech diagnoses and most common developmental disorder, epilepsy and movement disorder diagnoses in the speech cohort.

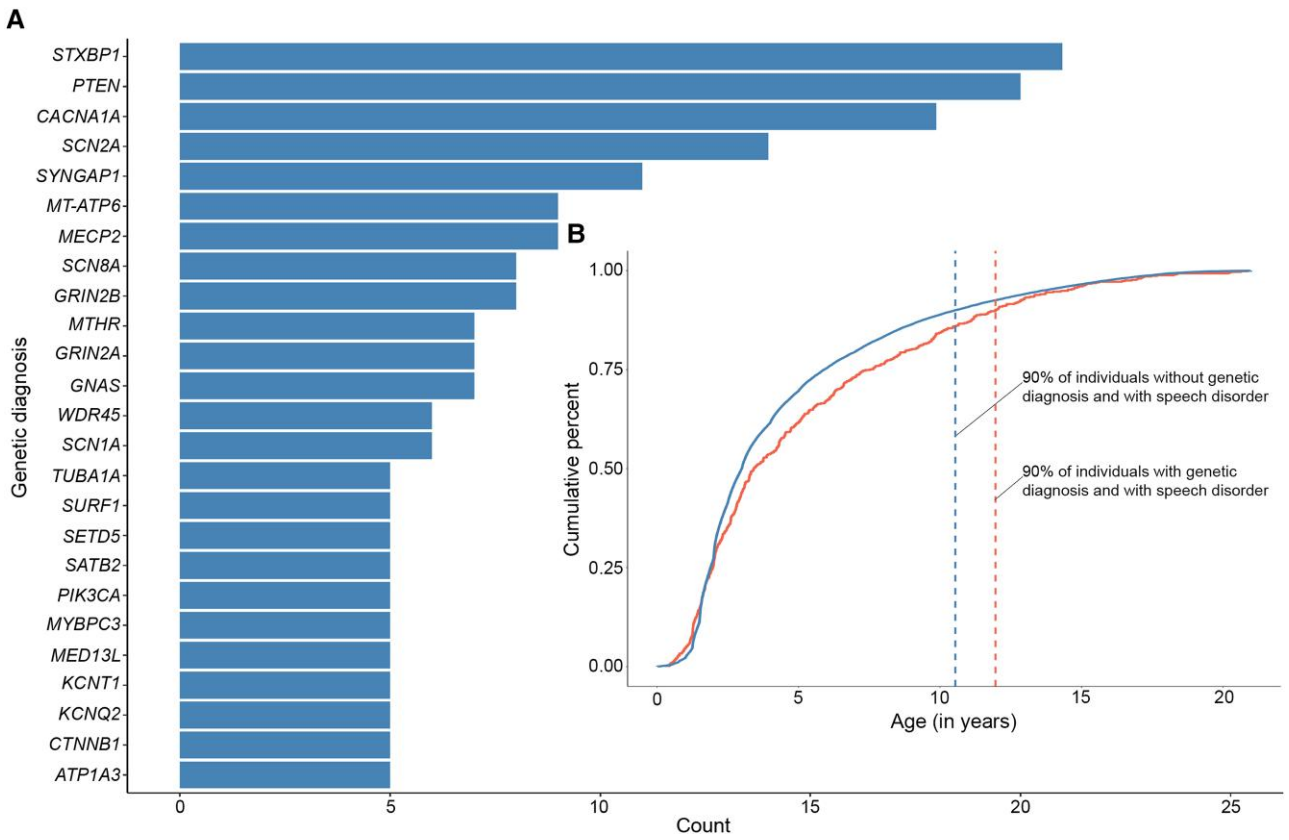

Figure 4 Genetic diagnoses in the speech cohort. (**A**) Distribution of the genetic diagnoses with $n \geq 5$ in the speech cohort. (**B**) Cumulative onset of speech diagnosis in individuals with and without a genetic diagnosis.

clinical information captured from routine care within EMR of 52 143 individuals across 203 150 years of patient data at a major US paediatric academic hospital. Overall, through this high-throughput EMR genomics approach, we confirmed the knowledge established previously by traditional phenotyping studies of smaller sample size, while expanding their findings. This approach allowed us to make three crucial observations. First, we found substantial heterogeneity of speech diagnoses, with mixed receptive–expressive language disorder and developmental disorder of speech and language being the most frequent diagnoses. Second, speech and language disorders have considerable overlap with neurodevelopmental disorders, movement disorders and epilepsy.[13,30] Third, distinct speech phenotypes can be associated with specific genotypic findings and demonstrate genetic overlap with known neurodevelopmental genetic conditions.[2,4,6]

Our analysis of speech diagnoses showed that, although there was a total of 26 ICD-10 codes corresponding to this broad clinical presentation, the broader phenotypic diagnoses were the most frequent. Terms describing mixed receptive–expressive language disorder, developmental disorder of speech and language or expressive language disorder were >11 times more prevalent than more specific speech disorder ICD-10 codes, such as speech apraxia, stuttering or dysarthria. Although the general speech diagnoses are undoubtedly useful in assessing high-level phenotypic associations, parsing out more granular features of speech impairment has proved to be difficult at the level of typical description in the EMRs. This observation reflects the need for deep speech phenotyping in order to describe this phenotypic landscape accurately, characterize clinical trajectories and allow for high-yield phenotype–gene association discoveries.[10] Speech and language impairment is often considered a feature of neurodevelopmental disorders, rather than an entity of its own, which might hinder precise characterization of these conditions. Our analysis supports this observation via the EMR visibility index; stuttering, a speech disorder with an elusive genetic underpinning, was least visible when assessing ICD-10 codes in our cohort. Here, only slightly more than 1 in 10 individuals had their stuttering diagnosis reflected in ICD-10 codes. This might account for prior observations that stuttering is a virtually absent diagnosis within large biobanks.[11] Additionally, our data might be affected by the fact that many individuals who stutter receive their care through community centres and school-based therapies. In short, genomic approaches using EMR data might not provide clear insight into a particular phenotype, requiring new approaches, such as phenotype classifiers[31] or, as in our study, analysis of full-text clinical notes through natural language processing.

Comorbidity with other conditions is a crucial aspect of the phenotypic spectrum of speech and language disorders. We appreciated substantial overlap with neurodevelopmental disorders, which was more than five times as high as that seen with epilepsies or movement disorders. This result is consistent with the general clinical presentation of neurodevelopmental disorders; speech and language impairment is a common domain affected in such conditions.[32] It is possible that, for this reason, speech and language differences are noticed more frequently in medical records of individuals with neurodevelopmental diagnoses[33] and are given attention in clinical care in these cases.

The clear relationship between speech and language and neurodevelopmental disorders was also reflected by the spectrum of

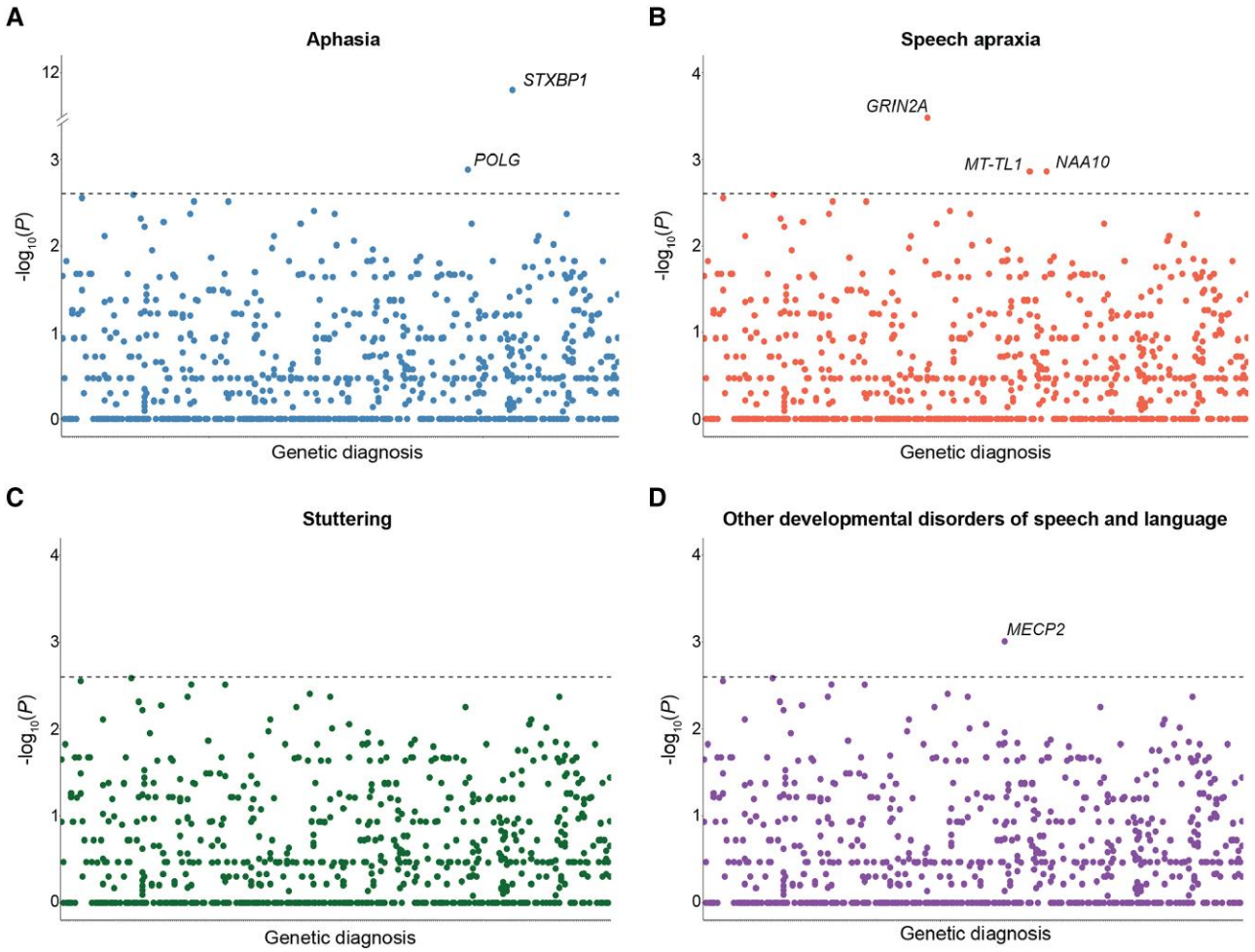

Figure 5 Associations between clinical genetic diagnoses and speech phenotypes. (**A**) Aphasia. (**B**) Speech apraxia. (**C**) Stuttering. (**D**) Other developmental disorders of speech and language. Labels are assigned to the diagnoses that maintained significance after false discovery rate of 10% after exploratory analysis shown by the dashed line. The genetic diagnoses on the x-axis are sorted alphabetically.

genetic diagnoses that we observed in our cohort. The genetic diagnoses that we identified here were related to genes known to be contributory in various neurodevelopmental disorders and epileptic encephalopathies (*STXBP1*, *GRIN2A*, *POLG* and *MECP2*), which is consistent with what was reported in the literature previously.[2,34-36] Furthermore, genes for which there was a nominally significant association with speech disorders were those contributing to movement disorders: *NKX2-1* is associated with chorea and *NUBPL* with ataxia and dystonia.[37,38] The last group of genes that showed nominally significant relationship with speech and or language phenotypes were known to be contributory to hearing loss: *GJB2* and *KCNQ1*.[39,40] The breadth of the genetic diagnoses spectrum illustrates the various dimensions of potential aetiologies of speech impairment, ranging from epileptic encephalopathies to movement disorders and hearing loss, mirroring the findings of our phenotype-based analysis. Disentangling speech and language phenotype–genotype association warrants further examination; we identified several relationships, but no genes that would be explanatory for speech and language impairments alone were identified in our cohort. It is worth noting that we identified genetic diagnoses with a frequency of occurrence equal to one in our cohort (Supplementary Table 4). Some of these included genes that are known to be contributing to conditions leading to speech or language impairment, such as *MYO7A* and hearing loss,[41] in addition to other genes that were identified in singular cases in our cohort but were not reported to be contributory elsewhere. This provides insight into the potential breadth of genes contributing to speech and language phenotypes.

With an increased search radius for both phenotypes—using more granular clinical data extracted from natural language processing of patient notes than clinical diagnoses—and genotypes—analysing exome sequencing in lieu of genetic diagnoses—we found more evidence for a genetic basis for speech and language phenotypes. We showed that variants in genes that have, and do not have, an established phenotype were found to contribute to speech and language disorders. Variants in *UQCRC1* have been established to be causative of parkinsonism with polyneuropathy.[42] Our work extends the spectrum of the disorders related to deleterious missense variants in this gene, revealing a prominent association with expressive aphasia. Likewise, we identified an association of poor speech in individuals with *BCHE* variants, which had been established previously as a cause of autosomal recessive butyrylcholinesterase deficiency (MIM: 617936) and associated with Alzheimer's disease, in addition to sudden infant death syndrome.[43,44] The remainder of the post-FDR significant genes (*MROH8*, *KIF17*, *USP37*, *SLC22A9* and *UMODL1*) do not have any established phenotypes, potentially representing novel genetic

Table 1 Associations between genetic diagnoses and speech phenotypes

Genetic diagnosis	Individuals	P-value	OR	95% CI	Frequency
Aphasia (R47.01)					
STXBP1	9	8.57×10^{-12}[a]	50.23	18.62–130.39	0.43
POLG	2	0.0013	65.87	4.77–898.38	0.5
CACNA1C	1	0.0297	65.79	0.84–4911.20	0.5
APC	1	0.0443	32.86	0.56–630.79	0.33
TUBA1A	1	0.0727	16.46	0.33–166.81	0.2
Speech apraxia (R48.2)					
GRIN2A	3	3.30×10^{-4}	34.06	4.98–201.11	0.43
NAA10	2	0.0014	90.60	4.71–5110.56	0.67
MT-TL1	2	0.0014	90.52	4.71–5106.15	0.67
CACNA1C	1	0.0428	45.21	0.576–3430.25	0.5
GABRB3	1	0.0428	45.21	0.576–3430.25	0.5
Dysarthria and anarthria (R47.1)					
NKX2-1	2	0.0013	92.88	4.83–5232.37	0.67
NUBPL	2	0.0013	92.88	4.83–5232.37	0.67
KCNQ2	2	0.0043	30.93	2.58–270.02	0.4
CTNNB1	2	0.0043	30.93	2.58–270.02	0.4
SURF1	2	0.0043	30.93	2.58–270.02	0.4
Speech and language development delay attributable to hearing loss (F80.4)					
MYO7A	3	1.24×10^{-5}[a]	Inf	17.46–Inf	1
GJB2	2	0.0016	84.97	4.42–4807.29	0.67
KCNQ1	2	0.0016	84.89	4.41–4803.38	0.67
Other developmental disorders of speech and language (F80.89)					
MECP2	2	9.81×10^{-4}	54.02	5.45–284.24	0.22
GLI3	1	0.0106	187.91	2.38–12 642.68	0.5
PACS1	1	0.0159	93.49	1.58–1817.74	0.33
DYRK1A	1	0.0211	62.60	1.19–795.86	0.25

CI = confidence interval; Inf = infinity; OR = odds ratio.
[a]If significant after the false discovery rate correction for multiple testing.

aetiologies that contribute to speech disorders. Importantly, *KIF17* was shown to be interacting with the glutamate *N*-methyl-D-aspartate receptor (NMDAR), presenting an interesting link between this gene and expressive aphasia, given that variants in genes encoding subunits of the NMDAR, e.g. *GRIN2A*, are an established cause of speech disorders.[2,45-48]

To gain a better understanding of the biological meaning and functional clustering of variants in genes nominally associated with speech phenotypes, we performed DAVID analyses, which showed that the most enriched pathways constitute central elements of neurologically crucial processes. Initially, these results confirmed what we established on the level of the ICD10-genetic diagnosis analysis; we observed nominally significant results for *GRIN2A*, *CACNA1C* and *MYO7A* in both analyses. This exhibits the high quality and sensitivity of the EMR genomics approaches, while highlighting the importance of comprehensive integrative bioinformatic analysis when dealing with rare variants. Glutamatergic neurotransmission appears to play a particularly prominent role in the genetics of speech impairment.[49] Although it was known before that *GRIN2A* had a characteristic speech and epilepsy phenotype, we determined that *GRIN2B* and *GRM1* are also associated with speech impairment.[2] This demonstrates a meaningful expansion of the existing knowledge of *GRIN2B*- and *GRM1*-related conditions, which have previously been associated with developmental epileptic encephalopathy and spinocerebellar ataxia, respectively.[50,51] Alhough these were absent in the DAVID analysis output, other glutamate receptor genes with both known (*GRIA3*) and unestablished phenotypes (*GRID1*, *GRIP1*, *GRIP2* and *GRIN3B*) showed nominally significant associations with speech and language differences in our exome analyses.[52,53] This analysis is consistent with what we observed on the phenotypic level through EMR analysis: the nature of speech disorders intersects with that of neurodevelopmental disorders.

To date, this is the first attempt to characterize speech disorders as their own entity and map them using longitudinal EMR data. We demonstrated that they tend to overlap both phenotypically and genetically with developmental, epilepsy and movement disorders. Novel variants we observed to be associated with speech phenotypes show a possible phenotypic plurality as conditions might have differing clinical characteristics depending on the genetic variation.

Further investigation into the landscape of the genetic architecture of speech disorders is necessary. Although we provide a comprehensive perspective on speech phenotypes here, the depth of phenotypic analysis is limited by the EMR-driven methods. Additionally, EMR genomics approaches can be influenced by specific centres of expertise contained within a particular healthcare network. It is possible that some genes causative of epilepsy and neurodevelopmental conditions emerged from our analysis owing to a large epilepsy genetics centre at Children's Hospital of Philadelphia, where children with these diagnoses are seen frequently. Future explorations might pursue phenotyping approaches in a similar computational manner, but in cohorts composed of individuals with a predefined speech disorder (e.g. stuttering, speech apraxia, dysarthria) which would allow for more finite analysis of associations between genetic changes and speech features. Targeted studies as described above are crucial for the discovery of novel genotype–phenotype associations, in addition to gene discovery, in the realm of speech disorder genetics.

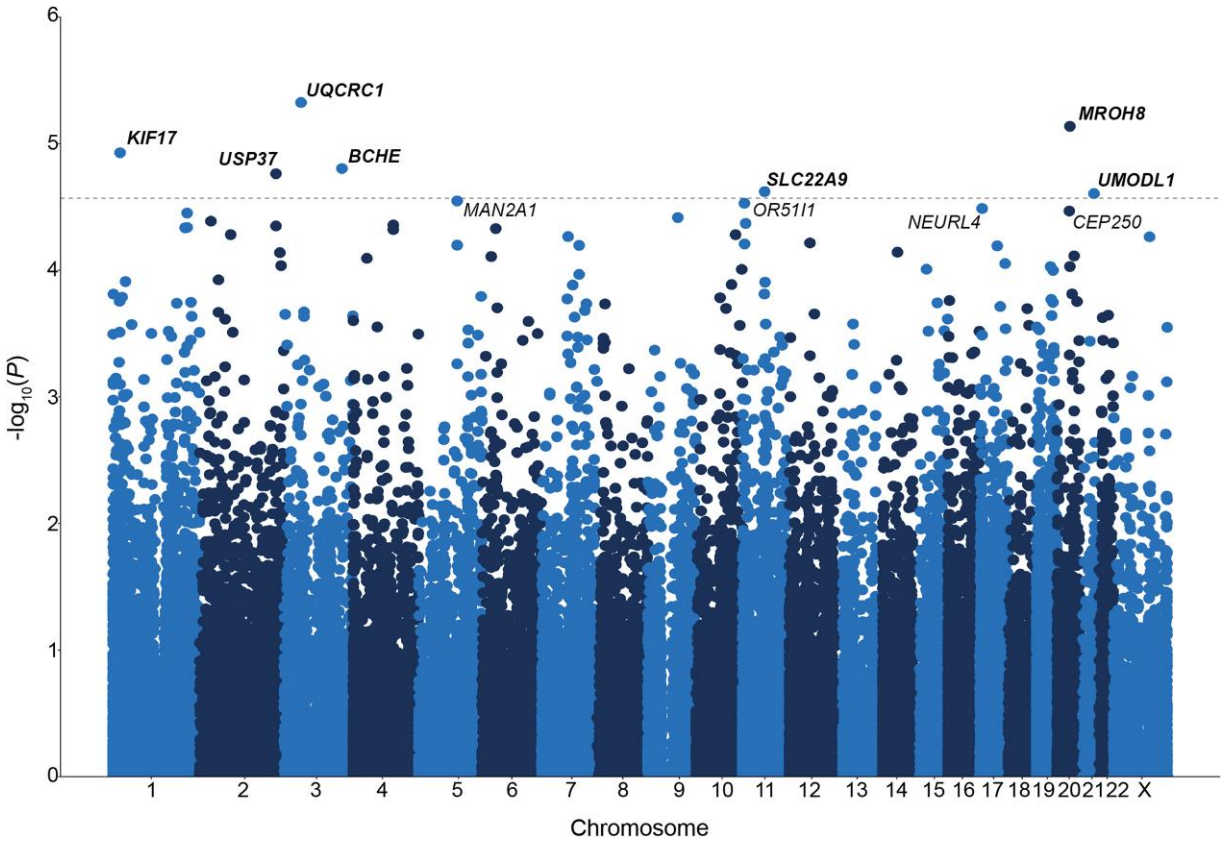

Figure 6 Variants present in the speech cohort. Distribution of all variants in the speech cohort. The dotted line represents the post-false discovery rate threshold of significance, and the names of the genes with significant associations are shown in bold text.

Table 2 Associations between speech phenotypes and rare variants

Gene	Speech phenotype	Total no. of individuals with a variant (proportion with the phenotype)	Control frequency	P-value	OR	95% CI
Class 2 variants (missense with CADD > 20)						
UQCRC1[a]	Expressive aphasia	8 (0.5)	0.02	4.73×10^{-6}[b]	50.27	9.39–269.16
NDST4	Incomprehensible speech	5 (0.4)	0.004	4.39×10^{-5}	511.50	25.67–10 193.46
POR	Alexia	9 (0.22)	0.006	5.42×10^{-5}	85.38	9.85–739.85
MAN2A1	Abnormal non-verbal communicative behaviour	11 (0.64)	0.11	6.34×10^{-5}	17.47	4.30–70.96
COG5	Mutism	9 (0.33)	0.02	6.36×10^{-5}	24.84	5.15–119.96
CHAC2	Receptive language delay	12 (0.33)	0.05	1.19×10^{-4}	15.00	3.78–59.57
TEKT2	Loss of speech	8 (0.38)	0.010	1.23×10^{-4}	42.17	6.25–284.68
GRID1	Poor speech	3 (0.67)	0.010	2.00×10^{-4}	720.03	22.46–23 083.16
Class 3 variants (PTVs and missense combined)						
MROH8	Poor speech	12 (0.25)	0.008	7.30×10^{-6}[b]	49.22	8.97–270.17
KIF17	Expressive aphasia	24 (0.21)	0.02	1.18×10^{-5}[b]	17.11	4.80–60.95
BCHE	Poor speech	16 (0.19)	0.008	1.57×10^{-5}[b]	43.49	7.85–240.95
USP37	Aphasia	10 (0.60)	0.10	1.73×10^{-5}[b]	21.31	5.28–85.98
SLC22A9	Aphasia	12 (0.58)	0.10	2.39×10^{-5}[b]	18.29	4.75–70.42
UMODL1	Aphasia	33 (0.33)	0.09	2.48×10^{-5}[b]	6.07	2.63–14.05
MAN2A1	Abnormal non-verbal communicative behaviour	17 (0.53)	0.11	2.83×10^{-5}	9.75	3.36–28.30
OR51I1	Mutism	14 (0.29)	0.02	2.96×10^{-5}	33.53	6.45–174.32
NEURL4	Aphasia	27 (0.37)	0.10	3.25×10^{-5}	6.33	2.65–15.12
CEP250	Aphasia	16 (0.50)	0.09	3.40×10^{-5}	10.34	3.43–31.22

CADD = Combined Annotation Dependent Depletion; CI = confidence interval; OR = odds ratio; PTVs = protein-truncating variants.
[a]*UQCRC1* Class 2 variants represented the entirety of *UQCRC1* Class 3 variants, hence all the statistical values for the association of Class 3 *UQCRC1* variants and expressive aphasia were the same as for *UQCRC1* Class 2 variants.
[b]If significant after the false discovery rate correction for multiple testing.

Data availability

Primary data used in this study are available upon request from the corresponding author. The computer code is available at https://github.com/jmagielski/EMR_Speech.

Funding

I.H. was supported by The Hartwell Foundation through an Individual Biomedical Research Award. This study received support from the National Institute of Neurological Disorders and Stroke (R01 NS127830-01A1, R01 NS131512-01, and K02 NS112600), and intramural funds of the Children's Hospital of Philadelphia from the Epilepsy NeuroGenetics Initiative (ENGIN).

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Thompson-Lake DGY, Scerri TS, Block S, et al. Atypical development of Broca's area in a large family with inherited stuttering. *Brain*. 2022;145:1177-1188.
2. Strehlow V, Heyne HO, Vlaskamp DRM, et al. *GRIN2A*-related disorders: Genotype and functional consequence predict phenotype. *Brain*. 2019;142:80-92.
3. Xian J, Parthasarathy S, Ruggiero SM, et al. Assessing the landscape of *STXBP1*-related disorders in 534 individuals. *Brain*. 2022;145:1668-1683.
4. Liégeois F, Morgan AT, Connelly A, Vargha-Khadem F. Endophenotypes of *FOXP2*: Dysfunction within the human articulatory network. *Eur J Paediatr Neurol*. 2011;15:283-288.
5. Morgan AT, Scerri TS, Vogel AP, et al. Stuttering associated with a pathogenic variant in the chaperone protein cyclophilin 40. *Brain*. 2023;146:5086-5097.
6. Morgan AT, Amor DJ, St John MD, Scheffer IE, Hildebrand MS. Genetic architecture of childhood speech disorder: A review. *Mol Psychiatry*. 2024;29:1281-1292.
7. Verhoef E, Allegrini AG, Jansen PR, et al. Genome-wide analyses of vocabulary size in infancy and toddlerhood: Associations with attention-deficit/hyperactivity disorder, literacy, and cognition-related traits. *Biol Psychiatry*. 2023;95:859-869.
8. Kang C, Drayna D. Genetics of speech and language disorders. *Annu Rev Genomics Hum Genet*. 2011;12:145-164.
9. den Hoed J, Fisher SE. Genetic pathways involved in human speech disorders. *Curr Opin Genet Dev*. 2020; 65:103-111.
10. Mei C, Fedorenko E, Amor DJ, et al. Deep phenotyping of speech and language skills in individuals with 16p11.2 deletion. *Eur J Hum Genet*. 2018;26:676-686.
11. Shaw DM, Polikowsky HP, Pruett DG, et al. Phenome risk classification enables phenotypic imputation and gene discovery in developmental stuttering. *Am J Hum Genet*. 2021;108:2271-2283.
12. 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;413:519-523.
13. Hildebrand MS, Jackson VE, Scerri TS, et al. Severe childhood speech disorder: Gene discovery highlights transcriptional dysregulation. *Neurology*. 2020;94:e2148-e2167.
14. Morgan A, Fisher SE, Scheffer I, Hildebrand M. FOXP2-related speech and language disorder. In: Adam MP, Feldman J, Mirzaa GM, et al., eds. *GeneReviews®*: University of Washington; 1993.
15. Maas E, Robin DA, Austermann Hula SN, et al. Principles of motor learning in treatment of motor speech disorders. *Am J Speech Lang Pathol*. 2008;17:277-298.
16. Stein CM, Benchek P, Miller G, et al. Feature-driven classification reveals potential comorbid subtypes within childhood apraxia of speech. *BMC Pediatr*. 2020;20:519.
17. Howell P. A model of serial order problems in fluent, stuttered and agrammatic speech. *Hum Mov Sci*. 2007;26:728-741.
18. Savova GK, Masanz JJ, Ogren PV, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. *J Am Med Inform Assoc*. 2010; 17:507-513.
19. Daniali M, Galer PD, Lewis-Smith D, et al. Enriching representation learning using 53 million patient notes through human phenotype ontology embedding. *Artif Intell Med*. 2023;139:102523.
20. Lewis-Smith D, Parthasarathy S, Xian J, et al. Computational analysis of neurodevelopmental phenotypes: Harmonization empowers clinical discovery. *Hum Mutat*. 2022;43:1642-1658.
21. Anderson C. Rare parents tackling rare diseases. *Inside Precis Med*. 2022;9(S1):6-13.
22. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res*. 2005; 33(Suppl 1):D514-D517.
23. Gudmundsson S, Singer-Berk M, Watts NA, et al. Variant interpretation using population databases: Lessons from gnomAD. *Hum Mutat*. 2022;43:1012-1030.
24. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24-26.
25. Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*. 2013;9:e1003709.
26. Dennis G Jr, Sherman BT, Hosack DA. DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biol*. 2003;4:P3.
27. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57.
28. Fabregat A, Jupe S, Matthews L, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2018;46(D1):D649-D655.
29. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Accessed 2 March 2023. http://www R-project org/
30. Kaspi A, Hildebrand MS, Jackson VE, et al. Genetic aetiologies for childhood speech disorder: Novel pathways co-expressed during brain development. *Mol Psychiatry*. 2023; 28:1647-1663.
31. Pruett DG, Shaw DM, Chen HH, et al. Identifying developmental stuttering and associated comorbidities in electronic health records and creating a phenome risk classifier. *J Fluency Disord*. 2021;68:105847.
32. Shriberg LD, Strand EA, Jakielski KJ, Mabie HL. Estimates of the prevalence of speech and motor speech disorders in persons with complex neurodevelopmental disorders. *Clin Linguist Phon*. 2019;33:707-736.
33. Brothers KB, Glascoe FP, Robertshaw NS. PEDS: Developmental milestones—An accurate brief tool for surveillance and screening. *Clin Pediatr (Phila)*. 2008;47:271-279.
34. Stamberger H, Nikanorova M, Willemsen MH, et al. *STXBP1* encephalopathy: A neurodevelopmental disorder including epilepsy. *Neurology*. 2016;86:954-962.

35. Rahman S, Copeland WC. *POLG*-related disorders and their neurological manifestations. *Nat Rev Neurol*. 2019;15:40-52.
36. Shahbazian MD, Zoghbi HY. Molecular genetics of Rett syndrome and clinical spectrum of *MECP2* mutations. *Curr Opin Neurol*. 2001;14:171-176.
37. Mencacci NE, Erro R, Wiethoff S, et al. *ADCY5* mutations are another cause of benign hereditary chorea. *Neurology*. 2015;85:80-88.
38. Balint B, Charlesworth G, Stamelou M, et al. Mitochondrial complex I *NUBPL* mutations cause combined dystonia with bilateral striatal necrosis and cerebellar atrophy. *Eur J Neurol*. 2019;26:1240-1243.
39. Snoeckx RL, Huygen PL, Feldmann D, et al. *GJB2* mutations and degree of hearing loss: A multicenter study. *Am J Hum Genet*. 2005;77:945-957.
40. Homma K. The pathological mechanisms of hearing loss caused by *KCNQ1* and *KCNQ4* variants. *Biomedicines*. 2022;10:2254.
41. Ma Y, Xiao Y, Zhang F, et al. Novel compound heterozygous mutations in *MYO7A* gene associated with autosomal recessive sensorineural hearing loss in a Chinese family. *Int J Pediatr Otorhinolaryngol*. 2016;83:179-185.
42. Jia F, Fellner A, Kumar KR. Monogenic Parkinson's disease: Genotype, phenotype, pathophysiology, and genetic testing. *Genes (Basel)*. 2022;13:471.
43. Ji H, Dai D, Wang Y, et al. Association of *BDNF* and *BCHE* with Alzheimer's disease: Meta-analysis based on 56 genetic case-control studies of 12,563 cases and 12,622 controls. *Exp Ther Med*. 2015;9:1831-1840.
44. Harrington CT, Hafid NA, Waters KA. Butyrylcholinesterase is a potential biomarker for sudden infant death syndrome. *EBioMedicine*. 2022;80:104041.
45. Liu Y, Tian X, Ke P, et al. KIF17 modulates epileptic seizures and membrane expression of the NMDA receptor subunit NR2B. *Neurosci Bull*. 2022;38:841-856.
46. Guillaud L, Setou M, Hirokawa N. KIF17 dynamics and regulation of NR2B trafficking in hippocampal neurons. *J Neurosci*. 2003;23:131-140.
47. Bai X, Karasmanis EP, Spiliotis ET. Septin 9 interacts with kinesin KIF17 and interferes with the mechanism of NMDA receptor cargo binding and transport. *Mol Biol Cell*. 2016;27:897-906.
48. Yin X, Feng X, Takei Y, Hirokawa N. Regulation of NMDA receptor transport: A KIF17–cargo binding/releasing underlies synaptic plasticity and memory *in vivo*. *J Neurosci*. 2012;32:5486-5499.
49. Li W, Kutas M, Gray JA, Hagerman RH, Olichney JM. The role of glutamate in language and language disorders – evidence from ERP and pharmacologic studies. *Neurosci Biobehav Rev*. 2020;119:217-241.
50. Endele S, Rosenberger G, Geider K, et al. Mutations in *GRIN2A* and *GRIN2B* encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nat Genet*. 2010;42:1021-1026.
51. Watson LM, Bamber E, Schnekenberg RP, et al. Dominant mutations in *GRM1* cause spinocerebellar ataxia type 44. *Am J Hum Genet*. 2017;101:451-458.
52. Salpietro V, Dixon CL, Guo H, et al. AMPA receptor GluA2 subunit defects are a cause of neurodevelopmental disorders. *Nat Commun*. 2019;10:3094.
53. Marafi D, Mitani T, Isikay S, et al. Biallelic *GRM7* variants cause epilepsy, microcephaly, and cerebral atrophy. *Ann Clin Transl Neurol*. 2020;7:610-627.