Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population

Sergey A. Kornilov, PhD, a, b, c, d, e Natalia Rakhlin, PhD, a, f Roman Koposov, MD, PhD, f Maria Lee, BSc, a Carolyn Yrigollen, PhD, h Ahmet Okay Caglayan, MD, f, i James S. Magnuson, PhD, b, c Shrikant Mane, PhD, a Joseph T. Chang, PhD, a Elena L. Grigorenko, PhD a, c, e, j

BACKGROUND AND OBJECTIVE: Developmental language disorder (DLD) is a highly prevalent neurodevelopmental disorder associated with negative outcomes in different domains; the etiology of DLD is unknown. To investigate the genetic underpinnings of DLD, we performed genome-wide association and whole exome sequencing studies in a geographically isolated population with a substantially elevated prevalence of the disorder (ie, the AZ sample).

METHODS: DNA samples were collected from 359 individuals for the genome-wide association study and from 12 severely affected individuals for whole exome sequencing. Multifaceted phenotypes, representing major domains of expressive language functioning, were derived from collected speech samples.

RESULTS: Gene-based analyses revealed a significant association between SETBP1 and complexity of linguistic output ($P = 5.47 \times 10^{-7}$). The analysis of exome variants revealed coding sequence variants in 14 genes, most of which play a role in neural development. Targeted enrichment analysis implicated myocyte enhancer factor–2 (MEF2)-regulated genes in DLD in the AZ population. The main findings were successfully replicated in an independent cohort of children at risk for related disorders ($n = 372$).

CONCLUSIONS: MEF2-regulated pathways were identified as potential candidate pathways in the etiology of DLD. Several genes (including the candidate SETBP1 and other MEF2-related genes) seem to jointly influence certain, but not all, facets of the DLD phenotype. Even when genetic and environmental diversity is reduced, DLD is best conceptualized as etiologically complex. Future research should establish whether the signals detected in the AZ population can be replicated in other samples and languages and provide further characterization of the identified pathway.

WHAT'S KNOWN ON THIS SUBJECT: Genetic underpinnings of common forms of pediatric disorders of language are heavily understudied. Recent association studies identified several tentative candidate genes. However, thus far, none of these candidates has received strong support in replication or confirmation analyses.

WHAT THIS STUDY ADDS: We established a statistically significant association between SETBP1 and language disorders in a geographically isolated population. Whole exome sequencing convergently implicated the myocyte enhancer factor–2–regulated pathways (of which SETBP1 is part) in language disorders in this special population.
Developmental language disorder (DLD) is a prevalent neurodevelopmental disorder, with 7–10% of children exhibiting atypical patterns of language development despite not having apparent sensorimotor/cognitive impairments or other medical conditions. DLD is lifelong, comorbid with other neurodevelopmental and psychiatric disorders, and associated with adverse academic and socio-emotional outcomes. It is phenotypically complex and genetically heterogeneous; although highly heritable, the etiology and pathogenesis of DLD are poorly understood.

A rare Mendelian type of DLD has been attributed to deleterious variants in the FOX2 gene (7q31); however, it is not associated with the disorder's common forms. For the latter, linkage studies have identified 3 susceptibility regions: 16q24, 19q13, and 13q2. Targeted association studies implicated CNTNAP2 (7q35; downregulated by FOX2) and CMIP and ATP2C2 (16q) genes in phonological memory deficits. Four genome-wide association studies (GWAS) divulged no genome-wide significant signals, with the exception of gene-based associations for CD2CL1, CD2CL2, LOC728661, and RCAN1. A whole exome sequencing (WES) study of DLD in an admixed Chilean founder population suggested the involvement of a nonsynonymous single nucleotide variant (SNV) in NFXL1; however, its location is not in the previously identified linkage regions in this population.

This pattern of findings highlights the complexity of DLD’s etiology, driven by the exclusionary nature of the diagnosis, the multicomponential nature of the phenotype, and the heterogeneity of the samples studied. The main objective of the present study was to identify genetic bases of DLD in a unique population (small, geographically secluded, and with an elevated prevalence of DLD [hereafter, the AZ population]) in which genetic and environmental variability is constrained. Genetic profiles of isolated populations are characterized by restricted genetic and allelic heterogeneity, thus rendering them ideal for studying the genetic bases of complex disorders.

The study population resides in a remote cluster of villages in Russia’s rural north; it was founded in the 15th century by several nuclear families. Currently, the AZ population comprises ~860 individuals (~120 children aged 3–18 years). Of these, 74.6% are represented by a set of multigenerational family structures (6391 individuals), of whom 82% are interconnected through a single 11-generational pedigree. The environmental conditions in the population are relatively uniform: all children go to the same kindergarten and school, and the socioeconomic indicators such as parental education and income show little variation. The AZ population is relatively geographically isolated and is characterized by an atypically high prevalence of DLD compared with 9% in the control rural population. This finding suggests the presence of a shared genetic component, potentially attributable to the founder effect(s).

METHODS
Population and Sample
Altogether, 474 AZ individuals donated DNA. When considered in combination with first-degree relatives, 405 of these donors represented 79 nuclear and extended pedigrees (N = 1152; range, 3–474; median, 6). Of these, 359 underwent phenotyping and constituted the GWAS sample: 124 children (62 male subjects; age, 5.33–17.92 years) and 235 adults (102 male subjects; age, 18.83–83.42 years). A total of 149 were classified as affected (DLD) and 210 as typically developing individuals (Supplemental Information).

Phenotyping
Phenotyping was performed by clinical linguists using elicited semi-structured speech samples. These samples were scored by using previously described phenotyping procedures to produce 5 quantitative phenotypes representing the major facets of DLD: phonetic/prosodic characteristics (eg, phonological omissions, misarticulations); well-formedness (rate of grammatical/lexical errors); complex structures (frequency of complex syntactic structures); mean length of utterance in words; and semantic/pragmatic errors (rate of errors in sentence meaning). Age-adjusted z scores were computed by using data from healthy control subjects from the comparison population to determine impairment status (ie, a z score below −1). Individuals were classified as overall DLD if they met the impairment criterion for ≥2 facets. Principal component analysis revealed that the 5 phenotypes formed 2 independent components: linguistic errors (phonetic/prosodic characteristics, well-formedness, and semantic/pragmatic errors) and syntactic complexity (complex structures and mean length of utterance in words).

Single Nucleotide Polymorphism Genotyping
The DNA extracted from peripheral blood (n = 384) or saliva/buccal swabs (n = 21) underwent quality control (QC) assessment for purity and degradation after standard collection, storage, and extraction procedures recommended by the manufacturers (Qiagen N.V. [Hilden, Germany] and DNA Genotek, Inc [Ottawa, ON, Canada]), and prepared at a concentration of 50 ng/μL.
Samples were genotyped at the Yale Center for Genome Analysis using HumanCNV 370k-Duo \((n = 315)\) or 610k-Quad \((n = 90)\) BeadChips (Illumina, Inc, San Diego, CA). Language status and gender distributions across the plates were not statistically different from random. Allele calling was performed by using the GenCall algorithm in GenomeStudio version 2011.1.

Samples and markers underwent QC review with GenomeStudio and SNP & Variation Suite (SVS) version 7.7.8 (GoldenHelix, Inc, Bozeman, MT). Samples with call rates >95% and verified gender were retained. A total of 223,580 autosomal single nucleotide polymorphisms (SNPs) common to 2 genotyping platforms were retained after QC so that the GenCall score was >0.30, the call rate was >95%, and minor allele frequency was >1%.

**Whole Exome DNA Sequencing**

Four subpedigrees were chosen for WES based on the results of complex segregation analysis\(^{26}\) that suggested possible Mendelian transmission. From these subpedigrees, 12 severely affected individuals were selected. Nine control non-AZ individuals without DLD from the same geographical region also underwent sequencing.

Exome capture was completed by using NimbleGen EZ Exome SeqCap v2 (Roche NimbleGen, Madison, WI). One microgram of fragmented genomic DNA was used to prepare the library using the manufacturer’s protocol (Supplemental Information). The bar-coded libraries were sequenced by using Illumina’s HiSeq 2500 platform, producing 75-bp paired-end reads that were aligned to the hg19 human genome build using NovoAlign (http://www.novocraft.com; Novocraft Technologies Sdn Bhd, Selangor, Malaysia). Variant calling was performed jointly for all samples by using the HaplotypeCaller algorithm in GATK.

**Genetic Association Analysis**

All of the quantitative trait loci association analyses were performed within the AZ sample for a set of 5 quantitative phenotypes. SNP-based association analysis of age- and gender-adjusted quantile-normalized phenotypes was performed by using mixed linear modeling (MLM) as implemented in GEMMA\(^{27}\) version 0.94. MLM tests for genetic association of SNPs with quantitative traits were performed under the additive model while controlling for sample structure estimated directly from data as a genetic relatedness matrix. MLM can be considered an example of the de-correlation approach to family-based data, and we chose it as our analytical framework for several reasons. First, although a number of transmission-based approaches (eg, family-based association testing, FBAT) have been developed, their use in a large complex multigenerational pedigree is problematic and computationally intensive in the presence of missing genotypic or phenotypic data, requiring splitting the larger pedigree into smaller units; this approach, coupled with conditioning on the founders’ genotypes, can lead to a loss of power. Second and most importantly, comparative studies suggest that decorrelation approaches (and MLM among them) tend to have higher (or at least comparable) statistical power than transmission approaches even in large and complex pedigrees.\(^{28}\)

All 5 phenotypes were first used in a multivariate MLM analysis. Two multivariate MLMs were then fitted: 1 that modeled the genetic effects on the indicators of linguistic errors and the second that used syntactic complexity. We performed gene-based association analyses as implemented in KGG3 software\(^{29}\) version 3.0 by using the hybrid set-based test. Copy number variant (CNV) association analysis was performed in the FBAT\(^{30}\) framework. Samples underwent additional CNV-specific QC (Supplemental Information). CNVs were identified by using a univariate Copy Number Analysis Method algorithm as implemented in SVS with a minimum of 5 markers per segment. Permutation testing was used to identify cut-points, and average segment intensity was used in the analyses.

Homozygosity mapping was completed by using the runs of homozygosity (ROH) detection algorithm in SVS. The minimum size was set to 250 kb and 25 SNPs, allowing for up to 1 heterozygote and 5 missing genotypes. The maximum gap between SNPs was 100 kb, and the minimum density was 18 kb. The total length of ROHs in 5 different length brackets was log-transformed to ensure normality before analysis. ROH association and burden analyses were performed by using univariate linear and logistic regression in \(R\) (R Foundation for Statistical Computing, Vienna, Austria).

WES data were analyzed by using a set of annotation and filtering tools. This analysis assumed that the most severely affected individuals in the AZ population from familial substructures with suggestive evidence for Mendelian transmission could provide additional information about the genetic architecture of DLD in the sample by focusing on: (1) the coding variants in candidate genes highlighted in the larger GWAS sample; or (2) the disruptive coding variants in other genes that could be conferring additional DLD risk in a subsample of the AZ population. Thus, we focused on coding sequence variants that were frequent among severely affected AZ probands (present in at least 4 of the 12 affected AZ individuals) but were not present in the control sample of 9 exomes. We then excluded variants observed in >5% of the National Heart, Lung, and Blood Institute Go Exome Sequencing Project (https://
results (Supplemental Information).

**Gene-based Associations**

We found no genome-wide significant gene-based associations for the 5-phenotype multivariate analysis or linguistic errors. Importantly, such an association was established between SETBP1 (SET binding protein 1; 18q21) and the multivariate syntactic complexity phenotype \( P = 5.47 \times 10^{-7} \) (Fig 2). The nuclear protein encoded by SETBP1 binds the SET nuclear oncogene protein involved in DNA replication, apoptosis, transcription, and nucleosome assembly. Rare variants in SETBP1 are associated with Schinzel-Giedion syndrome (MIM#269150) characterized by severe developmental delays.

Table 2 presents the top 10 genes for the gene-based analyses. After SETBP1, the 2 next strongest associations with syntactic complexity were found for 2 genes on chromosome 11q23: PPP2R1B (\( P = 4.77 \times 10^{-5} \)), encoding a constant regulatory subunit of protein phosphatase 2A, and SIK2 (\( P = 5.00 \times 10^{-5} \)), a gene hypothesized to play a role in neuronal protection. These findings are likely driven by the top hit SNP rs585149 (\( P = 1.70 \times 10^{-5} \)), assigned to both genes and located in the 3′-UTR region of SIK2. We also found a nominally significant association of syntactic complexity with TNC (\( P = .0068 \)).

Nominally significant associations were also established between linguistic errors and several genes (ABCG4, HYOU1, and HINFP) 7 Mb away from PPP2R1B (11q23); these genes and DPAGT1 and H2AFX were associated with the combined multivariate DLD phenotype, with the top hits being rs639373 (\( P = 1.21 \times 10^{-5} \)) and rs643788 (\( P = 1.22 \times 10^{-5} \)). The functional significance of the ABCG4 product is unknown; the DPAGT1 product is crucial for glycoprotein biosynthesis; and H2AFX encodes a histone involved in the maintenance of chromatin structure. The transcription factor encoded by HINFP plays an important role in DNA methylation. We found an association between the combined multivariate phenotype and estrogen-receptor 1 (ESR1; \( P = 4.76 \times 10^{-5} \)), with rs722208 (\( P = 3.09 \times 10^{-6} \)) as a top hit. There was a nominally significant association between linguistic errors and the HCLS gene (\( P = 4.40 \times 10^{-5} \)). Neither linguistic errors nor syntactic complexity was associated with previously identified candidate DLD genes.

**CNV Analysis and Homozygosity Mapping**

The multivariate FBAT CNV analysis revealed several nominally statistically significant and 1 highly statistically significant CNV. However, follow-up confirmation using real-time polymerase chain reaction (PCR) failed to substantiate the presence of these CNVs. An alternative pipeline that integrated 3 CNV detection algorithms yielded no genome-wide significant associations (Supplemental Information).

Overall, AZ-affected individuals, compared with unaffected individuals, had longer cumulative lengths of ROHs that were 250 to 500 kb long (\( P = .006 \)) and 1000 to 1750 kb long (\( P = .004 \), corresponding to \( ∼ \)10% and \( ∼ \)1% increases in estimated autosomal homozygosity, respectively (Supplemental Fig 3). The association analysis did not reveal any ROHs that were genome-wide significantly enriched in affected individuals. None of the top 20 regions overlapped with the regions identified in the SNP analyses. Several potentially relevant identified regions are discussed in Supplemental Information.

**Whole Exome DNA Sequencing**

We identified 14 coding sequence variants, frequent in affected AZ
individuals: 4 frameshift indels, 1 inframe insertion, 2 stop gain/loss, and 7 missense variants (Table 3). SNVs were predicted by polymorphism phenotyping (PolyPhen) to be possibly or probably damaging. Although any or all of these 14 variants could be implicated in the etiology of DLD in AZ, 2 sets of findings deserve special attention.

First, multiple individuals in the AZ population carried coding sequence variants in genes that regulate neural development or are highly expressed in the brain; that is, a frameshift insertion in \textit{NT5DC2} (3p21.1) and missense SNVs in \textit{NECAB1} (8q21.3) and \textit{ILK} (11p15.4). \textit{NT5DC2} has been implicated in schizophrenia\textsuperscript{34} and borderline personality disorder.\textsuperscript{35} \textit{NECAB1} is a member of the neuronal calcium-binding family of proteins essential to Ca\textsuperscript{2+}-mediated signaling and is highly expressed in the temporal lobe.\textsuperscript{36} The protein encoded by \textit{ILK} is 1 of the key regulators of neural stem cell astrocytic differentiation\textsuperscript{37} and neurite outgrowth.\textsuperscript{38} We also found that 7 (58\%) of 12 individuals in the AZ population carried a known missense variant in \textit{CDH2} (18q12) that was found only at a 2\% frequency in the 1000 Genomes data set. \textit{CDH2} codes for a major cadherin that is widely expressed prenatally in neural stem cells and supports their differentiation and migration,\textsuperscript{39} regulating the laminar organization of the cortex.\textsuperscript{40} Moreover, 7 of 12 AZ individuals carried a stop-gain variant in \textit{TCP10L2} (6q27). It is unknown whether \textit{TCP10L2} codes for a functional protein; it is highly similar to \textit{TCP10L}, a primate-specific transcription factor thought to evolve via segmental duplication\textsuperscript{41} from \textit{TCP10L2} or \textit{TCP10}.

Second, a missense SNV in \textit{TRIP6} (7q22.1) and a frameshift deletion in \textit{ENTHD1} (22q13) indicate commonalities between the genetic pathways identified through GWAS.
<table>
<thead>
<tr>
<th>SNP</th>
<th>Phenotype</th>
<th>Chr</th>
<th>Min/Maj</th>
<th>MAF</th>
<th>PPC</th>
<th>WF</th>
<th>CS</th>
<th>MLU</th>
<th>SPE</th>
<th>P</th>
<th>Gene</th>
<th>Variant Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs722208</td>
<td>All</td>
<td>6q25</td>
<td>G/A</td>
<td>0.35</td>
<td>0.38</td>
<td>0.18</td>
<td>-0.17</td>
<td>0.08</td>
<td>0.22</td>
<td>3.09 × 10^-6</td>
<td>ESR1</td>
<td>Intron</td>
</tr>
<tr>
<td>rs1559831</td>
<td>All</td>
<td>15q14</td>
<td>T/C</td>
<td>0.06</td>
<td>-0.17</td>
<td>-0.04</td>
<td>0.65</td>
<td>0.17</td>
<td>-0.34</td>
<td>3.81 × 10^-6</td>
<td>Intergenic</td>
<td></td>
</tr>
<tr>
<td>rs498012</td>
<td>All</td>
<td>15q14</td>
<td>G/A</td>
<td>0.06</td>
<td>-0.17</td>
<td>-0.04</td>
<td>0.65</td>
<td>0.17</td>
<td>-0.34</td>
<td>3.81 × 10^-6</td>
<td>Intergenic</td>
<td></td>
</tr>
<tr>
<td>rs3787751</td>
<td>All</td>
<td>21q22</td>
<td>C/T</td>
<td>0.05</td>
<td>-0.16</td>
<td>-0.39</td>
<td>0.04</td>
<td>0.04</td>
<td>-0.07</td>
<td>4.88 × 10^-6</td>
<td>HLC5</td>
<td>Intergenic</td>
</tr>
<tr>
<td>rs839373</td>
<td>All</td>
<td>11q23</td>
<td>C/T</td>
<td>0.35</td>
<td>0.30</td>
<td>0.08</td>
<td>0.19</td>
<td>0.05</td>
<td>-0.10</td>
<td>1.37 × 10^-5</td>
<td>C2CD2L</td>
<td>Upstream</td>
</tr>
<tr>
<td>rs2180386</td>
<td>All</td>
<td>14q32</td>
<td>A/C</td>
<td>0.38</td>
<td>-0.35</td>
<td>-0.27</td>
<td>0.05</td>
<td>-0.11</td>
<td>-0.22</td>
<td>1.60 × 10^-5</td>
<td>MEG8</td>
<td>Intron</td>
</tr>
<tr>
<td>rs780382</td>
<td>All</td>
<td>11p15</td>
<td>A/G</td>
<td>0.42</td>
<td>-0.26</td>
<td>-0.04</td>
<td>0.22</td>
<td>0.09</td>
<td>-0.04</td>
<td>1.89 × 10^-5</td>
<td>SBF2</td>
<td>Intron</td>
</tr>
<tr>
<td>rs3789868</td>
<td>Syn. comp.</td>
<td>9q33</td>
<td>G/A</td>
<td>0.06</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>3.81 × 10^-6</td>
<td>TNC</td>
<td>Intron</td>
</tr>
<tr>
<td>rs3789867</td>
<td>Syn. comp.</td>
<td>9q33</td>
<td>T/G</td>
<td>0.47</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>4.02 × 10^-6</td>
<td>TNC</td>
<td>Intron</td>
</tr>
<tr>
<td>rs2480933</td>
<td>Syn. comp.</td>
<td>9q33</td>
<td>A/G</td>
<td>0.46</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>5.51 × 10^-6</td>
<td>TNC</td>
<td>Intron</td>
</tr>
<tr>
<td>rs2482071</td>
<td>Syn. comp.</td>
<td>9q33</td>
<td>A/G</td>
<td>0.42</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>6.57 × 10^-6</td>
<td>TNC</td>
<td>Intron</td>
</tr>
<tr>
<td>rs2011604</td>
<td>Syn. comp.</td>
<td>14q32</td>
<td>T/C</td>
<td>0.43</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.25 × 10^-5</td>
<td>Intergenic</td>
<td></td>
</tr>
<tr>
<td>rs581419</td>
<td>Syn. comp.</td>
<td>11q23</td>
<td>T/C</td>
<td>0.16</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.70 × 10^-5</td>
<td>SIK2, PPP2R1B</td>
<td></td>
</tr>
<tr>
<td>rs293587</td>
<td>Syn. comp.</td>
<td>7p14-p13</td>
<td>G/A</td>
<td>0.27</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>2.04 × 10^-5</td>
<td>VPS41</td>
<td>Intron</td>
</tr>
<tr>
<td>rs1537722</td>
<td>Syn. comp.</td>
<td>9q33</td>
<td>T/C</td>
<td>0.36</td>
<td>—</td>
<td>—</td>
<td>0.03</td>
<td>—</td>
<td>0.33</td>
<td>2.62 × 10^-5</td>
<td>HAP1</td>
<td>Missense</td>
</tr>
<tr>
<td>rs4798064</td>
<td>Syn. comp.</td>
<td>17q21</td>
<td>T/C</td>
<td>0.44</td>
<td>—</td>
<td>—</td>
<td>0.36</td>
<td>—</td>
<td>0.04</td>
<td>2.96 × 10^-5</td>
<td>HAP1</td>
<td>Missense</td>
</tr>
<tr>
<td>rs2993756</td>
<td>Syn. comp.</td>
<td>9q33</td>
<td>T/C</td>
<td>0.19</td>
<td>—</td>
<td>—</td>
<td>0.07</td>
<td>—</td>
<td>0.39</td>
<td>3.15 × 10^-5</td>
<td>CNNA5P5</td>
<td>Intron</td>
</tr>
<tr>
<td>rs3787751</td>
<td>Ling. err.</td>
<td>21q22</td>
<td>C/T</td>
<td>0.45</td>
<td>-0.16</td>
<td>-0.39</td>
<td>—</td>
<td>—</td>
<td>-0.07</td>
<td>5.53 × 10^-7</td>
<td>HLC5</td>
<td>Intergenic</td>
</tr>
<tr>
<td>rs10504229</td>
<td>Ling. err.</td>
<td>14q32</td>
<td>A/G</td>
<td>0.16</td>
<td>0.25</td>
<td>0.25</td>
<td>—</td>
<td>—</td>
<td>0.50</td>
<td>5.59 × 10^-6</td>
<td>LIN00588</td>
<td>Intron</td>
</tr>
<tr>
<td>rs2180386</td>
<td>Ling. err.</td>
<td>14q32</td>
<td>A/G</td>
<td>0.38</td>
<td>-0.33</td>
<td>-0.27</td>
<td>—</td>
<td>—</td>
<td>-0.21</td>
<td>7.37 × 10^-6</td>
<td>MEG8</td>
<td>Intron</td>
</tr>
<tr>
<td>rs12121894</td>
<td>Ling. err.</td>
<td>1p36</td>
<td>C/A</td>
<td>0.09</td>
<td>0.64</td>
<td>0.14</td>
<td>—</td>
<td>—</td>
<td>0.28</td>
<td>8.00 × 10^-6</td>
<td>Intergenic</td>
<td></td>
</tr>
<tr>
<td>rs10480031</td>
<td>Ling. err.</td>
<td>7p21</td>
<td>G/A</td>
<td>0.16</td>
<td>0.33</td>
<td>-0.05</td>
<td>—</td>
<td>—</td>
<td>-0.32</td>
<td>1.06 × 10^-5</td>
<td>DGK8</td>
<td>Intergenic</td>
</tr>
<tr>
<td>rs639373</td>
<td>Ling. err.</td>
<td>11q23</td>
<td>C/T</td>
<td>0.38</td>
<td>0.31</td>
<td>0.08</td>
<td>—</td>
<td>—</td>
<td>-0.09</td>
<td>1.21 × 10^-5</td>
<td>C2CD2L</td>
<td>Upstream</td>
</tr>
<tr>
<td>rs643788</td>
<td>Ling. err.</td>
<td>11q23</td>
<td>T/C</td>
<td>0.41</td>
<td>0.32</td>
<td>0.11</td>
<td>—</td>
<td>—</td>
<td>0.06</td>
<td>1.22 × 10^-5</td>
<td>DPAGT1, H2AFX</td>
<td>Missense</td>
</tr>
<tr>
<td>rs17445065</td>
<td>Ling. err.</td>
<td>20q13</td>
<td>G/T</td>
<td>0.11</td>
<td>0.29</td>
<td>0.53</td>
<td>—</td>
<td>—</td>
<td>0.09</td>
<td>2.18 × 10^-5</td>
<td>Intergenic</td>
<td></td>
</tr>
<tr>
<td>rs722208</td>
<td>Ling. err.</td>
<td>6q25</td>
<td>G/A</td>
<td>0.35</td>
<td>0.37</td>
<td>0.18</td>
<td>—</td>
<td>—</td>
<td>0.22</td>
<td>2.48 × 10^-5</td>
<td>ESR1</td>
<td>Intron</td>
</tr>
<tr>
<td>rs4989070</td>
<td>Ling. err.</td>
<td>6q25</td>
<td>T/C</td>
<td>0.50</td>
<td>0.36</td>
<td>0.15</td>
<td>—</td>
<td>—</td>
<td>0.14</td>
<td>2.75 × 10^-5</td>
<td>Intergenic</td>
<td></td>
</tr>
</tbody>
</table>

Coefficients for phonetic/prosodic characteristics (PPC), well-formedness (WF), complex structures (CS), mean length of utterance (in words) (MLU), and semantic/pragmatic errors (SPE) represent effect size estimates (regression coefficients in SD units) for the minor allele. Chr, chromosomal location (cytoband); Ling. err, linguistic errors; MAF, minor allele frequency; Min/Maj, minor/major allele; Syn. comp., syntactic complexity; — coefficient was not estimated in the analysis and is not available.
**TRIP6** is a transcription factor that has been identified as a regulator of postnatal neural stem cell maintenance in the subventricular zone.42 **ENTHD1** codes for ENTH domain-containing protein 1. ENTH domain-containing proteins are involved in synaptic vesicle endocytosis at nerve terminals at the crucial stages that precede synapse formation.43 Importantly, TRIP6 interacts with and ENTHD1 is upregulated by the same family of genes, myocyte enhancer factor–2 (MEF2), labeled MEF2A-D. MEF2 are transcription factors implicated in muscle and central nervous system differentiation. In addition to ENTHD1, MEF2 targets in human neural stem cells include SETBP1, TNC, and DKGB (3 genes highlighted by our GWAS), as well as individual genes (BDNF, DMD, and NCAM2) and gene families (cadherins, contactins, semaphorins, and serpins) implicated in (a)typical central nervous system development. A targeted formal analysis of gene list enrichment using the Enrichr tool44 suggested that, combined, GWAS and WES hits in this population are indeed enriched for

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Gene</th>
<th>( P )</th>
<th>Chr</th>
<th>Length, bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>ESR1</td>
<td>( 4.76 \times 10^{-5} )</td>
<td>6q25</td>
<td>297 602</td>
</tr>
<tr>
<td>All</td>
<td>ABCG4</td>
<td>( 6.53 \times 10^{-5} )</td>
<td>11q23</td>
<td>13626</td>
</tr>
<tr>
<td>All</td>
<td>SFB2</td>
<td>( 6.58 \times 10^{-5} )</td>
<td>11p15</td>
<td>515 542</td>
</tr>
<tr>
<td>All</td>
<td>PPP2R1B</td>
<td>( 7.90 \times 10^{-5} )</td>
<td>11q23</td>
<td>28566</td>
</tr>
<tr>
<td>All</td>
<td>SK2</td>
<td>( 8.30 \times 10^{-5} )</td>
<td>11q23</td>
<td>124 464</td>
</tr>
<tr>
<td>All</td>
<td>HYOU1</td>
<td>( 8.44 \times 10^{-5} )</td>
<td>11q23</td>
<td>130 022</td>
</tr>
<tr>
<td>All</td>
<td>HINFP</td>
<td>( 8.77 \times 10^{-5} )</td>
<td>11q23</td>
<td>135 354</td>
</tr>
<tr>
<td>All</td>
<td>SNORD113-9,7,8</td>
<td>( 9.05 \times 10^{-5} )</td>
<td>14q32</td>
<td>72 74</td>
</tr>
<tr>
<td>All</td>
<td>SNORD113-4,5</td>
<td>( 9.06 \times 10^{-5} )</td>
<td>14q32</td>
<td>75 78</td>
</tr>
<tr>
<td>All</td>
<td>H2AFX</td>
<td>( 9.10 \times 10^{-5} )</td>
<td>11q23</td>
<td>1594</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>SETBP1</td>
<td>( 5.47 \times 10^{-7} )</td>
<td>18q21</td>
<td>388 337</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>PPP2R1B</td>
<td>( 4.77 \times 10^{-5} )</td>
<td>11q23</td>
<td>28566</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>SK2</td>
<td>( 5.00 \times 10^{-5} )</td>
<td>11q23</td>
<td>124 464</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>SERPINA1</td>
<td>( 7.83 \times 10^{-5} )</td>
<td>14q32</td>
<td>139 47</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>SPATA2</td>
<td>( 2.05 \times 10^{-4} )</td>
<td>20q13</td>
<td>12 153</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>EIF1</td>
<td>( 2.14 \times 10^{-4} )</td>
<td>17q21</td>
<td>2773</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>ST7-073</td>
<td>( 2.33 \times 10^{-4} )</td>
<td>7q31</td>
<td>27 258</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>RNF114</td>
<td>( 2.40 \times 10^{-4} )</td>
<td>20q13</td>
<td>17 515</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>HAP1</td>
<td>( 2.52 \times 10^{-4} )</td>
<td>17q21</td>
<td>12 009</td>
</tr>
<tr>
<td>Syn. comp.</td>
<td>GAST</td>
<td>( 2.54 \times 10^{-4} )</td>
<td>17q21</td>
<td>36 45</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>LINC00588</td>
<td>( 3.44 \times 10^{-5} )</td>
<td>9q12</td>
<td>5190</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>RP11:515017.3</td>
<td>( 3.60 \times 10^{-5} )</td>
<td>9q12</td>
<td>5387</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>SNORD113-9,7,8,4,5</td>
<td>( 4.17 \times 10^{-5} )</td>
<td>14q32</td>
<td>72 78</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>HLC5</td>
<td>( 4.40 \times 10^{-5} )</td>
<td>21q22</td>
<td>238 358</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>SNORD113-3</td>
<td>( 4.76 \times 10^{-5} )</td>
<td>14q32</td>
<td>72</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>SNORD113-6</td>
<td>( 4.90 \times 10^{-5} )</td>
<td>14q32</td>
<td>75</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>SNORD113-2,1</td>
<td>( 5.46 \times 10^{-5} )</td>
<td>14q32</td>
<td>71 72</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>ABCG4</td>
<td>( 5.75 \times 10^{-5} )</td>
<td>11q23</td>
<td>136 26</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>HYOU1</td>
<td>( 6.34 \times 10^{-5} )</td>
<td>11q23</td>
<td>130 022</td>
</tr>
<tr>
<td>Ling. err.</td>
<td>HINFP</td>
<td>( 6.59 \times 10^{-5} )</td>
<td>11q23</td>
<td>135 354</td>
</tr>
</tbody>
</table>

Chr, chromosomal location (cytoband); Ling. err., linguistic errors; Syn. comp., syntactic complexity.

*Statistically significant after Bonferroni corrections for multiple testing.*
<table>
<thead>
<tr>
<th>ID</th>
<th>Type</th>
<th>Variant</th>
<th>N</th>
<th>Gene (GWAS P)</th>
<th>Brain Exp.</th>
<th>NHLBI Freq</th>
<th>rsid</th>
<th>Protein Change</th>
<th>PolyPhen</th>
<th>GERP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZW1</td>
<td>Del (FS)</td>
<td>NC_000001.10:g.6727803_6727804delTC</td>
<td>4/12</td>
<td>DNAJC1 (P = .0080)</td>
<td>Low</td>
<td>0.0000</td>
<td>rs374290353</td>
<td>NP_060968.2:p.Glu119fs</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>AZW2</td>
<td>Ins (FS)</td>
<td>NC_000003.11:g.52559298_52559299insCAAA</td>
<td>4/12</td>
<td>NT5DC2 (P = .3050)</td>
<td>High (pn)</td>
<td>0.0002</td>
<td>rs745485897</td>
<td>NP_001127703.1:p.Asp378fs</td>
<td>NA</td>
<td>5.1</td>
</tr>
<tr>
<td>AZW3</td>
<td>Stop gain (SNV)</td>
<td>NC_000006.11:g.167591956C&gt;T</td>
<td>7/12</td>
<td>TOP10L2 (P = .0198)</td>
<td>Medium</td>
<td>0.0000</td>
<td>rs2297482</td>
<td>NP_001138593.1:p.Gln195Ter</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>AZW4</td>
<td>Stop loss (SNV)</td>
<td>NC_000009.11:g.71152177A&gt;G</td>
<td>4/12</td>
<td>TMEM252 (P = .0921)</td>
<td>Low/medium</td>
<td>0.0030</td>
<td>rs147022534</td>
<td>NP_084969.1:p.Ter171fs</td>
<td>NA</td>
<td>0.4</td>
</tr>
<tr>
<td>AZW5</td>
<td>Ins</td>
<td>NC_000011.9:g.16510985insGGCGCTGGCTCCGGCTGTGGG</td>
<td>4/12</td>
<td>KRTAP5-5 (P = .0183)</td>
<td>NA</td>
<td>0.0000</td>
<td>NA</td>
<td>NP_001001480.2:p.Gly9_Cys10LeuTrpLeuTrpLeuTrpGly</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>AZW6</td>
<td>Del (FS)</td>
<td>NC_000019.9:g.13255238delG</td>
<td>4/12</td>
<td>STX10 (P = .3610)</td>
<td>Medium (us)</td>
<td>0.0018</td>
<td>NA</td>
<td>NP_003758.1:p.Leu248fs</td>
<td>NA</td>
<td>1.1</td>
</tr>
<tr>
<td>AZW7</td>
<td>Del (FS)</td>
<td>NC_000022.10:g.40161342_40161343delGA</td>
<td>4/12</td>
<td>ENTHD1 (P = .4250)</td>
<td>High (us)</td>
<td>0.0075</td>
<td>rs200712517</td>
<td>NP_089725.2:p.Pro369fs</td>
<td>NA</td>
<td>5.9</td>
</tr>
<tr>
<td>AZW8</td>
<td>Missense (SNV)</td>
<td>NC_000007.13:g.100465824G&gt;A</td>
<td>4/12</td>
<td>TRIP6 (P = .4260)</td>
<td>Medium (us)</td>
<td>0.0205</td>
<td>rs2437100</td>
<td>NP_003293.2:p.Asp111Gln</td>
<td>PrD</td>
<td>4.5</td>
</tr>
<tr>
<td>AZW9</td>
<td>Missense (SNV)</td>
<td>NC_000008.10:g.9153077G&gt;T</td>
<td>4/12</td>
<td>NECAB1 (P = .9248)</td>
<td>High (esp. temporal)</td>
<td>0.0001</td>
<td>rs11555424</td>
<td>NP_071746.1:p.Ala271Thr</td>
<td>PrD</td>
<td>4.4</td>
</tr>
<tr>
<td>AZW10</td>
<td>Missense (SNV)</td>
<td>NC_000009.11:g.124072932G&gt;A</td>
<td>4/12</td>
<td>GSN (P = .8910)</td>
<td>Medium (us)</td>
<td>0.0047</td>
<td>rs14305623</td>
<td>NP_000168.1:p.Val178Met</td>
<td>PrD</td>
<td>5.3</td>
</tr>
<tr>
<td>AZW11</td>
<td>Missense (SNV)</td>
<td>NC_000011.9:g.3700876C&gt;T</td>
<td>4/12</td>
<td>NOP98 (P = .9480)</td>
<td>Low/medium</td>
<td>0.0164</td>
<td>rs34446537</td>
<td>NP_057404.2:p.Glu1611Lys</td>
<td>PrD</td>
<td>5.6</td>
</tr>
<tr>
<td>AZW12</td>
<td>Missense (SNV)</td>
<td>NC_000011.9:g.6629345T&gt;A</td>
<td>4/12</td>
<td>ILK (P = .1540)</td>
<td>Low</td>
<td>0.0001</td>
<td>rs200536808</td>
<td>NP_001014794.1:p.Leu53Met</td>
<td>PrD</td>
<td>3.9</td>
</tr>
<tr>
<td>AZW13</td>
<td>Missense (SNV)</td>
<td>NC_000017.10:g.50555309G&gt;A</td>
<td>4/12</td>
<td>MPV1 (P = .1420)</td>
<td>No</td>
<td>0.0135</td>
<td>rs28730837</td>
<td>NP_000241.1:p.Ala332Val</td>
<td>PoD</td>
<td>4.3</td>
</tr>
<tr>
<td>AZW14</td>
<td>Missense (SNV)</td>
<td>NC_000018.9:g.25552304T&gt;C</td>
<td>6/12</td>
<td>CDH2 (P = .0041)</td>
<td>High (pn)</td>
<td>0.0268</td>
<td>rs2289864</td>
<td>NP_001783.2:p.Asn845Se</td>
<td>PoD</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Brain expression (Brain Exp.) information is based on the combination of the data from the BrainSpan Atlas of the Developing Human Brain (www.brainspan.org), the Human Protein Atlas (www.proteinatlas.org), and the Expression Atlas (www.ebi.ac.uk/gxa/). Del, deletion; FS, frameshift; GERP, Genomic Evolutionary Rate Profiling sequence conservation score; Ins, insertion; N, number of carriers (of 12 individuals); NA, not available; NHLBI Freq., variant frequency in National Heart, Lung, and Blood Institute European American exomes; pn, prenatally; PoD, possibly damaging; PrD, probably damaging; us, ubiquitous; rsid, Single Nucleotide Polymorphism Database submission number.

* P value for the HYST gene-based analysis (multivariate, all phenotypes).
MEF2 targets (for MEF2A, $P = 1.28 \times 10^{-6}$) (Supplemental Information), providing support to this hypothesis.

Our WES analysis also revealed the presence of 2 heterozygous missense mutations in SETBP1, carried by 2 (rs3744825) and 1 (rs1064204) sequenced AZ individual, respectively. Both were common (for European ancestry, minor allele frequency >10% in National Heart, Lung, and Blood Institute exome database) known SNPs, projected to be tolerated according to 5 different functional prediction algorithms.

**Replication**

We interrogated the main loci highlighted in the GWAS or WES analyses of DLD in the AZ population in an independent sample ($n = 372$) of children at risk for developmental disorders of language (spoken and written) by using teachers’ ratings of student’s spoken and written language skills as the main phenotype (details are given in the Supplemental Information). Association analysis controlled for age and gender and was performed by using EMMAX, a MLM algorithm implemented in SVS.

Both main findings were replicated. First, a significant gene-based association was found between language scores and SETBP1 ($P = 0.009360$). The top signal originated at exon 1383999/rs11082414 ($P = .000359$), a missense SNP located within exon 4 of SETBP1 that explained 3.41% of the variance in children’s language skills. Predicted to be tolerated according to sorting intolerant from tolerant/PolyPhen, this SNP may play a role in the regulation of expression of SETBP1.

The analysis of the Brainiac46 brain expression quantitative trait loci database suggested that it differentiates levels of SETBP1 expression in the brain, including the cerebellar cortex, hippocampus, and temporal cortex.

Second, genes nominally associated (at $P < .05$) with teacher ratings of students’ spoken and written language skills were enriched for MEF2A targets ($P = .0007024$), replicating the finding from the discovery cohort.

**DISCUSSION**

We established a genome-wide association between syntactic complexity and the SETBP1 gene in the AZ sample and then replicated it in an independent sample. SETBP1 is relatively large (388 337 bp), has 2 isoforms, and is expressed widely. Although little is known about its function, it is implicated in several neurodevelopmental conditions: SETBP1 haploinsufficiency is documented in expressive DLD47–49 and intellectual disability.50 Moreover, several tentative SNP associations were found between syntactic complexity and TNC that encodes tenasin, an extracellular matrix glycoprotein involved in neural development; TNC-deficient mice exhibit structural and functional cortical abnormalities, including atypical neuronal density and abnormal dendrite morphology.51 However, the combined multivariate phenotype was also nominally associated with ESR1, a nuclear hormone receptor involved in regulation of gene expression, cell proliferation, and differentiation. Estrogen is involved in synaptogenesis, regulates neurotransmission, and modulates the activity of all types of neural cells.52 This finding is intriguing given the male bias in incidence of DLD and the recent report of associations between early postnatal gender hormone concentrations and later language development.53

Our WES highlighted 14 coding variants in a set of genes implicated in neural development and/or differentiation. Intriguingly, 2 of the WES-identified genes (ENTHD1 and TRIP6) and 3 of the GWAS-identified genes (SETBP1, TNC, and DRG5) interact with or are regulated by the MEF2 transcription factors. MEF2 isoforms are widely expressed in neural cells,54 and their activity is regulated by extracellular factors (eg, in neurons via neurotrophin stimulation or Ca2+ influx after the release of neurotransmitters). MEF2 targets show enriched expression in the central nervous system and implicate multiple signaling pathways, rendering MEF2 as a key regulator of activity-dependent synapse development.55 The complex transcriptional program of MEF2 results in the restriction of excitatory synaptic transmission via the reduction of the number of excitatory neurons, elimination of glutamatergic synapses,56 and postsynaptic differentiation of neurons (dendrite morphogenesis).57

The cascade of events regulated by the transcriptional activity of MEF2 is critical for learning and memory.58, 59 A recent electrophysiological study partially attributed the DLD phenotype in the AZ population to atypicalities in the functioning of neural circuits that support attention and memory60 that were linked to syntactic complexity. It is plausible they at least partially stem from the dysregulation of common genetic pathways that orchestrate neural development.

This dysregulation can take multiple forms. Given the partial convergence of the results from the GWAS and WES, we hypothesized that the DLD phenotype in the AZ population emerged as the result of the interaction between common genetic variants that conferred background DLD susceptibility and rare variants that altered the development of language and memory circuits against that background. This extension of the threshold-dependent response model suggests that common variants in several genes (eg, SETBP1, TNC)
formed the probabilistic landscape(s) of DLD vulnerability, and that coding variants in multiple different genes (eg, regulated by MEF2 such as such \textit{ENTHD1} and \textit{TRIP6} or other genes important for neural development such as \textit{CDH2} or \textit{NECAB1}) conferred the critical amount of vulnerability and pushed this landscape into a critical state.

Finally, we established a higher rate of autosomal ROH burden among the affected AZ individuals compared with unaffected AZ individuals; this finding is not surprising given the isolated nature of the population and the role of ROHs in several developmental disorders.\textsuperscript{61} However, no single specific ROH was strongly associated with DLD. In addition, there was little overlap between the genetic loci identified in the GWAS analyses of the 2 multivariate phenotypes; this outcome raises an interesting hypothesis that the 2 global facets of DLD may be relatively independent at the level of their molecular neurobiology.

Our study has several limitations. First, it has a small sample size. Although it was modest for a GWAS study, however, the sample size was almost one-half of the total AZ population. Second, the unique nature of the population poses a complex issue for future research seeking to replicate these signals in other samples. Although we replicated the association finding for \textit{SETBP1} and the enrichment findings for GWAS-highlighted DLD genes for MEF2 targets in an independent sample of children at risk for a related disorder, further molecular and analytical studies in larger samples are necessary to better characterize the joint contribution of common and rare variants in the identified genes to DLD susceptibility and decipher the molecular pathways they affect.

\textbf{CONCLUSIONS}

This study presented a set of novel candidate genes and coding DNA sequence variants contributing to DLD phenotypes in the AZ population; the chief findings from this population have been replicated in an independent sample. Overall, the findings suggest that multiple genes (including a novel genome-wide significant candidate \textit{SETBP1}) and genetic pathways (including the suggested MEF2-regulated pathway) are involved in DLD. This study underlines the complexity of the genetic architecture of DLDs and illustrates that even in populations with reduced genetic and environmental diversity, DLD is best conceptualized as a polygenic and etiologically complex disorder.

\textbf{ACKNOWLEDGMENTS}

The authors thank the families who participated in the study for their cooperation and patience, and the local medical, kindergarten, and school officials of the AZ community for their help with data collection. They also thank Igor Pushkin, Anastasia Strelina, Liudmila Kniazeva, and various students, trainees, and employees from Northern State Medical University for their help with the logistics of the study; Dr Lesley Hart for her contributions to the early stages of the project; Drs Seongmin Han and Dean Palejev for their involvement at various stages of the project; Ms Mei Tan for her editorial assistance; and the late Dr Maria Babyonyshev for her contribution to the linguistic component of the study.

\textbf{ABBREVIATIONS}

\begin{itemize}
\item ACES: Academic Competence Evaluation Scales
\item CNV: copy number variant
\item DLD: developmental language disorder
\item GWAS: genome-wide association study
\item HWE: Hardy-Weinberg equilibrium
\item MEF2: myocyte enhancer factor–2
\item MLM: mixed linear modeling
\item PCR: polymerase chain reaction
\item PolyPhen: polymorphism phenotyping
\item QC: quality control
\item ROH: runs of homozygosity
\item SNP: single nucleotide polymorphism
\item SVN: single nucleotide variant
\item SVS: SNP & Variation Suite
\item WES: whole exome sequencing
\end{itemize}
REFERENCES


25. Rakshit N, Kornilov SA, Paleyev D, Koposov RA, Chang JT, Grigorenko EL. The language phenotype of a small geographically isolated Russian-speaking population: implications for genetic and clinical studies of
Suzuki Y, Yang X, Aoki Y, Kure S, 32. Sifrim A, Popovic D, Tranchevent LC,
Ionita-Laza I, Perry GH, Raby BA, et 31.  Li MX, Kwan JS, Sham PC. HYST: a
Li D, Zhou J, Thomas DC, Fardo DW. 27. Zhou X, Stephens M. Genome-wide
complex pedigrees in the sequencing era: to track transmission or decorrelate? Genet Epidemiol. 2014;38(suppl 1):S29–S36
methylation analysis.


Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population
Sergey A. Kornilov, Natalia Rakhlin, Roman Koposov, Maria Lee, Carolyn Yrigollen, Ahmet Okay Caglayan, James S. Magnuson, Shrikant Mane, Joseph T. Chang and Elena L. Grigorenko
Pediatrics 2016;137; DOI: 10.1542/peds.2015-2469 originally published online March 25, 2016;

Updated Information & Services
including high resolution figures, can be found at:
http://pediatrics.aappublications.org/content/137/4/e20152469

Supplementary Material
Supplementary material can be found at:
http://pediatrics.aappublications.org/content/suppl/2016/03/25/peds.2015-2469.DCSupplemental

References
This article cites 39 articles, 2 of which you can access for free at:
http://pediatrics.aappublications.org/content/137/4/e20152469.full#ref-list-1

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
Developmental/Behavioral Pediatrics
http://classic.pediatrics.aappublications.org/cgi/collection/developmental_issues_sub
Cognition/Language/Learning Disorders
http://classic.pediatrics.aappublications.org/cgi/collection/cognition:language:learning_disorders_sub
Genetics
http://classic.pediatrics.aappublications.org/cgi/collection/genetics_sub

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
https://shop.aap.org/licensing-permissions/

Reprints
Information about ordering reprints can be found online:
http://classic.pediatrics.aappublications.org/content/reprints

Pediatrics is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. Pediatrics is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2016 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005.
Genome-Wide Association and Exome Sequencing Study of Language Disorder in an Isolated Population

Sergey A. Kornilov, Natalia Rakhlin, Roman Koposov, Maria Lee, Carolyn Yrigollen, Ahmet Okay Caglayan, James S. Magnuson, Shrikant Mane, Joseph T. Chang and Elena L. Grigorenko

Pediatrics 2016;137;
DOI: 10.1542/peds.2015-2469 originally published online March 25, 2016;

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://pediatrics.aappublications.org/content/137/4/e20152469