

Article

Innovative Family-Based Genetically Informed Series of Analyses of Whole-Exome Data Supports Likely Inheritance for Grammar in Children with Specific Language Impairment

Erin M. Andres ¹, Kathleen Kelsey Earnest ², Hao Xuan ³, Cuncong Zhong ³, Mabel L. Rice ^{2,4}
and Muhammad Hashim Raza ^{4,*}

- ¹ Thompson Center for Autism and Neurodevelopment, University of Missouri, Columbia, MO 65201, USA; eandres@health.missouri.edu
² Language Acquisition Studies Lab, University of Kansas, Lawrence, KS 66045, USA
³ Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045, USA
⁴ Child Language Doctoral Program, University of Kansas, Lawrence, KS 66045, USA
* Correspondence: razam@ku.edu

Abstract: Individuals with specific language impairment (SLI) struggle with language acquisition despite average non-verbal intelligence and otherwise typical development. One SLI account focuses on grammar acquisition delay. The current study aimed to detect novel rare genetic variants associated with performance on a grammar assessment, the Test of Early Grammatical Impairment (TEGI), in English-speaking children. The TEGI was selected due to its sensitivity and specificity, consistently high heritability estimates, and its absence from all but one molecular genetic study. We performed whole exome sequencing (WES) in eight families with SLI ($n = 74$ total) and follow-up Sanger sequencing in additional unrelated probands ($n = 146$). We prioritized rare exonic variants shared by individuals with low TEGI performance ($n = 34$) from at least two families under two filtering workflows: (1) novel and (2) previously reported candidate genes. Candidate variants were observed on six new genes (*PDHA2*, *PCDHB3*, *FURIN*, *NOL6*, *IQGAP3*, and *BAHCC1*), and two genes previously reported for overall language ability (*GLI3* and *FLNB*). We specifically suggest *PCDHB3*, a protocadherin gene, and *NOL6* are critical for ribosome synthesis, as they are important targets of SLI investigation. The proposed SLI candidate genes associated with TEGI performance emphasize the utility of precise phenotyping and family-based genetic study.

Keywords: whole-exome sequencing; language phenotypes; specific language impairment; family-based; pedigree; grammar impairment



Citation: Andres, E.M.; Earnest, K.K.; Xuan, H.; Zhong, C.; Rice, M.L.; Raza, M.H. Innovative Family-Based Genetically Informed Series of Analyses of Whole-Exome Data Supports Likely Inheritance for Grammar in Children with Specific Language Impairment. *Children* **2023**, *10*, 1119. <https://doi.org/10.3390/children10071119>

Academic Editors: Irene Cadime, Iolanda Ribeiro and Maria Luisa Lorusso

Received: 17 May 2023
Revised: 22 June 2023
Accepted: 26 June 2023
Published: 28 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Is human language inherited? Familial aggregation and behavioral genetic studies have consistently suggested that genes have a greater influence on language expression than the environment [1–6]. The current study adds supportive evidence to the claims of the innateness of language and its specificity to humans, specifically adding to our understanding of the genetic basis of specific language impairment (SLI). SLI is characterized by a delay in language acquisition and a persistent language deficit in the absence of hearing loss and other neurological or developmental disorders [7]. The estimated prevalence of SLI is 7–10% in English-speaking populations [8,9]. SLI remains a valid phenotype in the scientific literature, but we also note a recently updated term, developmental language disorder (DLD), which includes those who do not meet the specific criteria for SLI, as well as those who do meet the diagnostic criteria for SLI [10,11]. The advantage of the SLI criteria is a greater measurement precision for group status. Categorical labels for children with language impairments can be viewed as clinical labels for eligibility for services in contrast to group labels for scientific studies such as the one reported here. Further, the

pathways differ for clinical services across countries. In some countries, they are nested within school special education services (e.g., the United States of America (USA)), whereas, in other countries, they are nested within public health/medical services (e.g., the United Kingdom (UK)). In the USA, the term ‘SLI’ arose as a scientific label to differentiate children with multiple developmental disorders from children whose language disorders are their single developmental disorder [12]. ‘Developmental Language Disorders’ could be applied to children with co-existing clinical neurological disorders [13].

Discussions surrounding standardizing the diagnostic criteria for SLI/DLD have tended toward increasing the language and intelligence standard score cut-offs and many have discussed how variable diagnostic criteria have impacted our continued questions about the genetics of language acquisition [14]. In contrast, the current study defines the SLI phenotype according to performance on the Test of Early Grammatical Impairment (TEGI), a sensitive and specific assessment of tense and agreement marking abilities in individuals speaking mainstream American English [15]. A deficit in tense and agreement marking is a known clinical behavioral marker of SLI [16,17]. The emphasis on the grammar deficit also supports our interest in the genetic influences specifically on language acquisition, given that the participants represented in the current study show a deficit in language despite average or above average non-verbal intelligence [16,18,19]. There is continued interest in strengthening our interdisciplinary approaches to speech and language impairments, especially in thinking about how speech–language pathologists can expand and use their knowledge of behavioral genetics in practice [20]. Additionally, there has been a push for utilizing larger cohorts of unrelated individuals and existing medical records for the genetic investigation of language traits [21–23]. Significantly, our study relies foremost on the sensitivity and specificity of the TEGI and the genetic relatedness of the participants and the power it may provide to the genetic investigation of SLI.

The TEGI is regarded as one of the most psychometrically sound instruments in terms of sensitivity and specificity in the assessment of children with SLI (after age three) [15,24–26]. The TEGI was developed through a longitudinal study, and the evidence supporting the specificity of the test is rooted in the linguistic theory of grammar acquisition [15,17,27]. Specifically, Wexler’s (1994) theory of optional infinitive (OI), which is based on the assumption of Universal Grammar (UG)-constrained Maturation (UGCM) [27], contributed to the development of the TEGI and the theory of a clinical linguistic marker of SLI [17]. UGCM assumes children have an innate capacity for adult grammar, which matures, but during that maturation, it generally does not allow the child to produce utterances that go against the UG [27]. The OI theory focuses on the optionality that children appear to have in their grammar, causing them to produce infinitive verb forms (e.g., “She teach”), when morphological endings are obligatory [27]. The absence of inappropriate inflection supports the assumption of UGCM [27]. Children with SLI show growth trajectories of multiple language abilities that run parallel to their typically developing peers (same slope of development) about two years delayed [16,17]. The delay period was observed to extend the time that children with SLI optionally use tense, motivating the theory of Extended OI (EOI) and EOI as a clinical marker of SLI [16,17,28]. A recent report of individuals with SLI from the current study and additional individuals from the larger longitudinal study provides evidence that the difficulty in finiteness marking observed in childhood extends through age 18, as measured by performance on tasks with more grammatically challenging linguistic structures [29]. Crucially relevant to the current study, receptive and expressive grammar phenotypes revealed significant heritability estimates (up to 0.92) in two twin cohorts at time points ranging from 2 to 16 years [3–5,30]. Despite significant heritability estimates, the TEGI phenotype was used only once across previous molecular genetic investigations of SLI [31]. The current study aims to prioritize rare genetic variants from whole-exome sequencing (WES) data that segregate with the TEGI phenotype.

Two epidemiological twin cohorts tested for grammar impairment (i.e., neither ascertained based on SLI status) revealed that the lowest performing groups yielded the highest heritability estimates [3–5]. These results suggest that genetic factors may explain more of

the variance in grammar abilities among individuals with language impairment (LI) than in those without LI and that grammar impairment may be the most valid clinical marker in need of study at the molecular genetic level [4,5].

Family-based linkage studies provided several gene targets and multiple genes have been suggested for SLI through WES [18,19,32,33]. Family-based studies, in conjunction with next-generation sequencing (NGS), have the potential to provide promising gene targets to explain the biological basis of language acquisition. Three studies of SLI utilizing WES output are of note. First, WES of select individuals ($n = 5$) from a founder population of Robinson Crusoe Island ($N = 117$) resulted in a strong candidate, *NFXL1* [33], which expresses in the cerebellum, a region previously implicated in language development [34]. Second, another study used multiple variant filtering criteria to identify variants of interest from WES output from select SLI Consortium probands ($n = 43$) [32]. One variant prioritization approach targeted variants within candidate genes previously suggested for SLI, as recommended by the guidelines put forth for evaluating the causality of sequencing variants [32,35]. The other streams of variant filtering prioritization identified rare stop-gain variants, with bioinformatic in silico scores predicted to be deleterious, and sought compound heterozygotes and cases of what the authors called “multiple-hits” [32]. Ultimately, the WES findings led to the hypothesis that transmission of a complex disorder, like SLI, is likely to be explained by a combination of genetic variants (rare and common), including those on previously identified genes [32]. If related individuals were included at the whole-exome level, co-segregation analysis could have been used to prioritize variants. Third, we recently used a similar approach in the study of family 5886 (reported as family 489); we used three workflows to prioritize rare variants of interest and observed a co-segregating protein-coding rare variant in *BUD13*, which was not previously reported for SLI or related phenotypes. Targeted sequencing of this gene in unrelated SLI probands from the same population revealed more *BUD13* variants in additional probands with statistical significance [19]. Overall, these findings support the value of WES investigation, especially utilizing related individuals who are both affected and unaffected, to identify novel SLI candidate genes.

In the current study, we utilized WES output from select individuals who have completed the TEGI, from eight informative families from the University of Kansas (KU) SLI cohort (Figure 1, $n = 74$). Note genetic findings from many of these families have been reported previously [18,19,31,36]. We focused our genetic analysis on (Figure 2): (1) novel candidate genes and (2) variants in 113 previously reported candidate genes implicated in SLI and related phenotypes (Table S1) in the family members who have completed the TEGI (Table S2, $n = 34$). We hypothesized that the sensitivity and specificity of the TEGI and the consistent reports of high heritability would support the precise detection of rare genetic variants associated with SLI. We prioritized variants shared by at least two families and predicted the identified variants may also be observed in the larger KU SLI cohort, who have completed the TEGI (Table S3, $n = 146$).

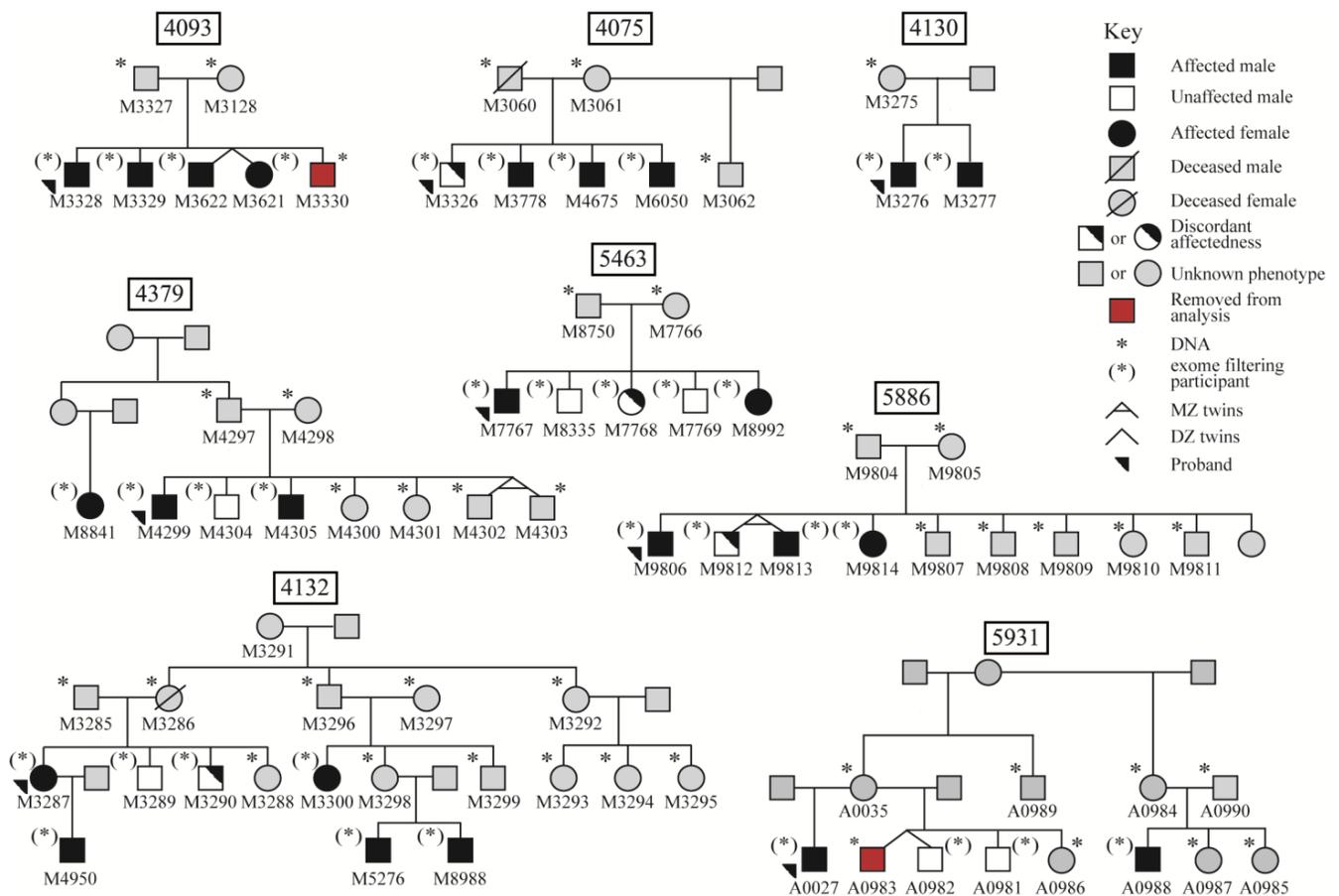


Figure 1. Eight Families included in WES ($n = 74$) with Categorical Affectedness Status for the TEGI in a Subset of the Family Members ($n = 34$). Note. Discordant affectedness refers to performance in the unaffected range on the screener or composite probes, but not both. Family 4132 Branch 1 (proband branch) includes descendants of M3286 and M3285, Family 4132 Branch 2 includes descendants of M3296 and M3297, and Family 4132 Branch 3 includes descendants of M3292. Family 5931 Branch 1 (proband branch) includes descendants of A0035, and Family 5931 Branch 2 includes descendants of A0984 and A0990. One proband (M3287), in family 4132, had only a TEGI screener score available. Another proband (M3326) performed in the unaffected range on the TEGI screener probes, but their composite performance was in the affected range.

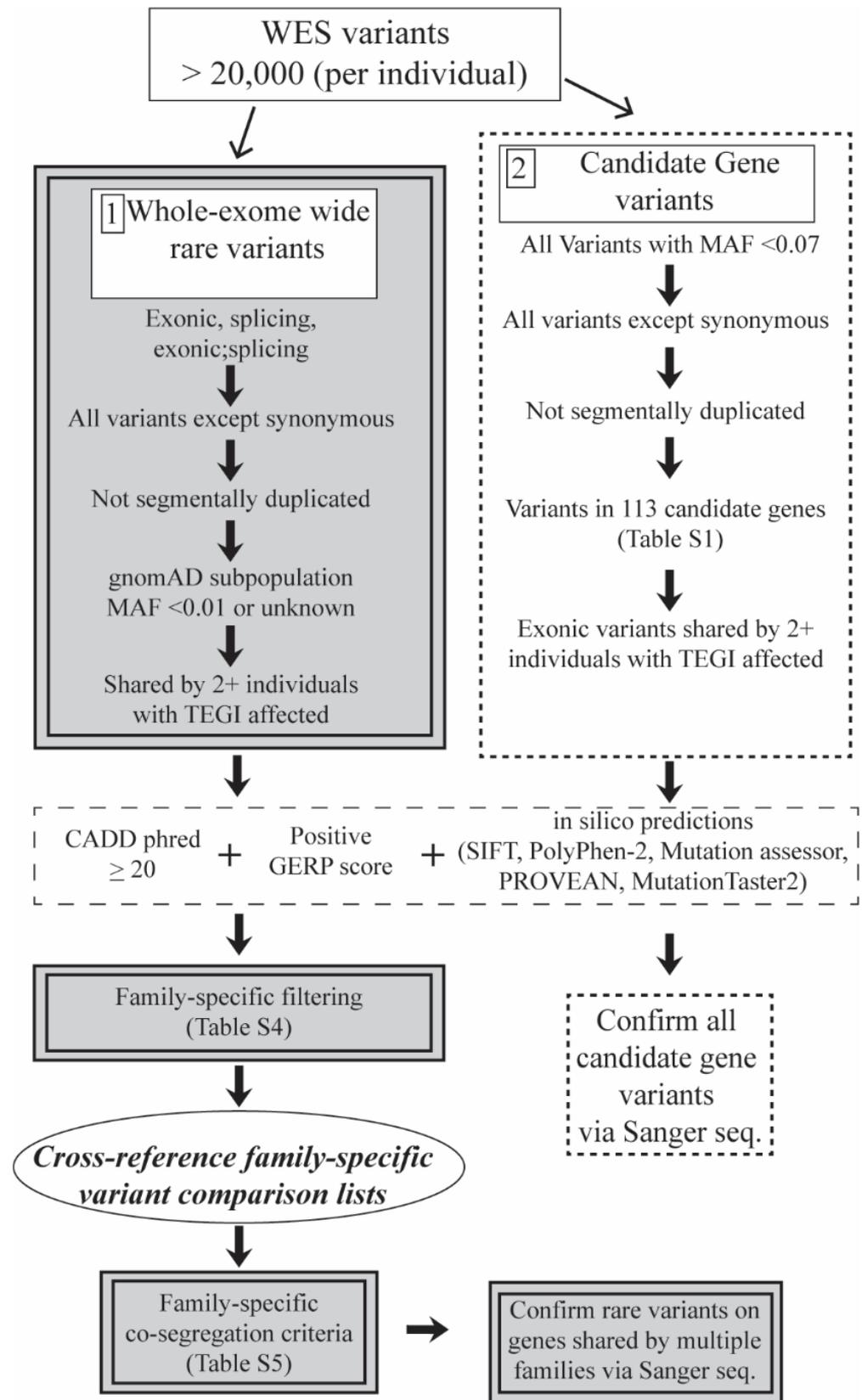


Figure 2. Variant Prioritization Workflow. MAF = minor allele frequency; familywise ‘co-segregating variant lists’ presented in Table S11a–h.

2. Materials and Methods

The institutional review board (IRB #8223) at the University of Kansas approved this study for behavioral data collection on 25 January 1993, and it has been annually reviewed and approved since that time. Relevant genetics amendments included: (1) Collection of DNA via blood draw and cheek cell samples, approved on 19 January 1999; (2) Collection of DNA via saliva samples, approved on 31 January 2006; (3) consent form update to include the National Institutes of Health Certificate of Confidentiality, approved on 3 January 2018. Participants provide their signatures for informed consent to all amended genetics protocols. All methods were performed in accordance with the relevant guidelines and regulations of the Declaration of Helsinki and University of Kansas Human Research Protection Program. All participants provided appropriate informed consent.

Participants in the current study were part of a larger ongoing longitudinal study of probands with SLI and their family members. Details concerning all assessments administered as part of the larger study are described in an earlier publication by Rice, Smith, and colleagues [31]. The term ‘proband’ refers to the individual originally targeted for the study. The proband entrance criteria for the study include (i) average or above average performance on a standardized non-verbal intelligence (NV-IQ) measure (standard score > 85), (ii) typical hearing, (iii) no history of neurological disorders or autism diagnosis, and (iv) intelligible speech/articulation [18,19,31]. All participants are monolingual speakers of General American English [18,19].

Individuals from eight families in the larger study ($n = 74$) were included in the current study. A subset of the individuals ($n = 36$) have completed the TEGI (age-referenced for children ages 3 to 8; 11 years), the phenotype of interest, at least once (Figure 1) [15]. Two siblings (both males) were excluded from the subset due to potentially confounding patterns of lowered NV-IQ (Figure 1). The remaining individuals (Table S2; $n = 34$; 27 males and 7 females; referred to as TEGI-WES group) were the focus of the WES variant filtering (Figure 2). All eight probands (7 males and 1 female) are affected on the TEGI based on their elicited grammar composite or screener score, six of whom are affected according to both scores (Figure 1).

The current study assigned affection status categorically based on the participants’ lowest score across time points, consistent with how previous research established affectedness [18,19]. The TEGI probes tense marking and finiteness relative to mastery in adult grammar [15]. The screener and composite probes require elicitation. A phonological probe prior to the other probes ensures that the child can produce the required sounds; crucial because the marking is required in the final position of the word [15]. The screener score is the average of the third-person singular and past tense probes, while the elicited grammar composite score also includes the ‘be’ and ‘do’ probe scores [15].

The total number of additional probands available for participation in this study was 157 (provided DNA and completed the TEGI). Eleven children who may have shown potentially confounding patterns of lowered NV-IQ were excluded from our analysis, resulting in a total of 146 probands. The proband entrance criteria do not require the proband to score in the affected range on the TEGI. Therefore, 21 probands included did not show low performance on the TEGI (Table S3). The TEGI-WES group ($n = 34$) and the additional probands ($n = 146$) all completed an age-appropriate standardized omnibus language measure and a receptive vocabulary measure (Tables S2 and S3).

Participants provided saliva samples/buccal swabs using the Oragene-Discover OGR-500 or OGR-575 Kits (DNA Genotek, Oragene). DNA was purified according to the manufacturer’s instructions. WES and bioinformatic analyses were performed in eight families ($n = 74$) over two-time points. The first round of WES was performed in select individuals from six of the eight families ($n = 29$), using the Illumina Nextera Rapid Capture Enrichment kit (expanded; includes untranslated genomic regions [UTR]). The remaining individuals from the six families and all individuals from two additional families ($n = 45$) were included in the second round of WES using the Illumina NovaSeq6000 (UTRs were

not included). The sequencing data were mapped to the human reference genome (hg38), and variants were called as described in an earlier publication [19].

The exonic variant filtering relied on categorical affectedness status based on TEGI performance (Table S2). Figure 2 shows the two complementary variant filtering prioritization approaches: (1) whole-exome wide (novel candidate genes) and (2) targeted prioritization (candidate genes previously reported for SLI and related phenotypes). Both approaches employ criteria to prioritize variants shared by multiple affected individuals within a single family. We applied the following common *a priori* filtering criteria in workflow 1 and 2: (i) classified as ‘exonic’, ‘splicing’, or ‘exonic;splicing’; (ii) not classified as synonymous; (iii) not located within a segmentally duplicated region; (iv) a Combined Annotation Dependent Depletion (CADD) Phred score ≥ 20 ; (v) a positive Genomic Evolutionary Rate Profiling (GERP) score; (vi) shared by at least two family members affected on the TEGI; (vii) multiple damaging scores according to five *in silico* programs, including SIFT (Sorting Intolerant from Tolerant), PolyPhen-2 (Polymorphism Phenotyping v2), Mutation Assessor, PROVEAN (Protein Variation Effect Analyzer), and MutationTaster2 [37–44]. Articles commonly assess and present multiple *in silico* prediction scores to provide context for the significance of the identified variants [19,36,44–47]. Finally, we applied family-specific criteria (detailed in Tables S4 and S5). All cross-referencing steps were completed in R using the ‘dplyr’ and ‘tidyr’ packages [48–50].

The first variant filtering workflow prioritized rare and novel variants whole-exome wide. Rare variants were defined as those with a minor allele frequency (MAF) ≤ 0.01 in the subpopulation appropriate for the family within the Genome Aggregation Database (gnomAD). Variants with unknown MAF and a predicted deleterious effect were defined as novel variants. Family 5931 is of African American descent, while the other families are of European descent.

We cross-referenced ‘family-specific variant comparison lists’ to identify genes shared across the eight families (Figure 2; Table S4). Family-specific co-segregation criteria (Figure 2; Table S5) were applied (‘co-segregating variant lists’) to further reduce the prioritized list. In total, four individuals (across two families) with low performance on the TEGI were required to carry variant(s) on the genes on the final list of prioritized variants.

The second variant filtering workflow independently prioritized variants in 113 candidate genes compiled from reviews and candidate gene investigations, as recommended by MacArthur and colleagues (Table S1; used in our previous WES investigation of family 5886) [32,35,51,52]. If the candidate gene was also listed in a more recent review (Mountford et al., 2022), the reference is noted in Table S1 [14]. Within the targeted filtering workflow, we prioritized variants using a less stringent MAF of <0.07 (Figure 2). Using a less stringent MAF applied to the updated public databases allows the filtering workflow to pick up variants in the previously suggested genes, such that confirming and disconfirming evidence for the previous candidates can be added to the literature.

After the variants with a MAF > 0.07 , variants causing a synonymous change or variants located in a segmentally duplicated region were removed, the remaining variants were cross-referenced with the list of 113 candidate genes (Table S1) [48,49]. All variants on previously reported candidate genes that were shared by two individuals who showed low performance on the TEGI and met all other *a priori* filtering criteria were prioritized for confirmation via Sanger sequencing. Prioritized variants in the candidate genes were not required to be shared by two families, given that they share the gene with a previous report.

Oligos were designed using Primer 3 to amplify and confirm the prioritized variants (Table S6). Then, we analyzed the Sanger sequencing data in SeqMan Pro within the DNASTar suite.

Select confirmed variants were Sanger sequenced in the additional probands ($n = 146$). Finally, variants were classified as benign, likely pathogenic based on if they met a combination of *a priori* criteria. The TEGI was completed by a subset of each family, which limited co-segregation analysis, so we added the likely pathogenic category based solely on the predicted pathogenicity.

3. Results

We prioritized variants in 36 genes by applying filtering criteria to the WES data in eight families (workflow 1 = 23 (Tables S7–S10 and S11a–h); workflow 2 = 13 (Tables S12 and S13)). Following confirmation, 12 variants in nine genes co-segregated in their respective families, prompting follow-up sequencing in the additional probands ($n = 146$; Table 1). We observed multiple unrelated probands carrying variants in six genes not previously reported for language impairment (*PDHA2*, *PCDHB3*, *FURIN*, *NOL6*, *IQGAP3*, and *BAHCC1*). We also observed variants in two genes previously suggested for SLI and related phenotypes (*GLI3* and *FLNB*).

3.1. Variant Prioritization Workflow 1: Whole-Exome Wide Rare Variants

The eight families started with a range of 12,000 to just over 47,000 exonic variants (Table S7). When only the variants shared by two individuals with low performance on the TEGI were kept, the variants were reduced to under 700 for all families (Table S7). Then, family-specific filtering criteria (Table S4) reduced the ‘family-specific variant comparison lists’ to a range of 18 to 208 variants, and the unique genes were cross-referenced to reveal 55 shared genes (Table S7). Of the 55 shared genes, 8 genes were excluded for various reasons (described in Table S7). Family-specific co-segregation criteria (Table S5) further reduced the ‘co-segregation variant lists’ to a range of 6 to 37 variants (Table S7). The ‘co-segregating variant lists’ (Table S11a–h) were cross-referenced familywise with the 47 shared genes of interest (Table S8) and only 23 genes containing a variant in at least one family’s ‘co-segregating variant lists’ were further investigated (Tables S9 and S10). Seven genes were excluded according to the reported protein expression (Table S9). Variants on nine genes were either observed in all family members, were not confirmed, or the primers could not be optimized (Table S9). Variants on the remaining nine genes were confirmed through Sanger sequencing (Tables S9 and S10). Co-segregation of variants on six of these genes (*PDHA2*, *PCDHB3*, *FURIN*, *NOL6*, *IQGAP3*, and *BAHCC1*) with the TEGI was confirmed (Tables S9, S10 and 1).

3.2. Variant Prioritization Workflow 2: Candidate Gene Variants

A filtered list of variants (5000 to 29,000) was cross-referenced familywise with the established list of 113 candidate genes suggested for SLI and related phenotypes (Tables S1 and S12). No additional family-specific criteria were applied to filter the variants and in total, 14 variants in 13 previously reported genes were prioritized under filtering workflow 2 (Table S12). Variants on 12 of the candidate genes were prioritized for confirmation in family members via Sanger sequencing (variant shared by family 4132 and 5886 on *PTEN* was excluded; Table S13).

In sum, all 13 variants on 12 candidate genes were confirmed in their respective families via Sanger sequencing (Tables S12 and S13). Co-segregation analysis with the TEGI was confirmed for variants in three genes (*GLI3*, *FLNB*, and *KMT2D*; Tables 1, S12 and S13).

Table 1. Additional Information for Variants Tested in the Proband (n = 146).

Gene	Discovery Pedigree (s)	Additional Proband Carrying Variant (s)		Fisher's Test p-Value ²	rsID	AA Change	MAF	# of Damaging In Silico Scores ³	AA Change (HOPE)		Causality Prediction
		Affected	Unaffected						Size	Charge	
<i>PDHA2</i>	5463, 5886	2 ^	0	0.3590	rs147966234	p.Arg286Pro	0.0089 ^{1a}	5/5	∨	POS to neu	P
<i>PCDHB3</i>	4093, 4130	1	0	0.5112	rs61739886	p.Thr81Ile	0.0064 ^{1a}	3/5	∧	NCC	P
<i>FURIN</i>	4075, 5886	0	0	0.06445	rs150925934	p.Arg462Trp	0.0017 ^{1a}	4/5	∧	POS to neu	P
<i>NOL6</i>	4130	1	0	<0.0001 ^{*,1,4}	rs114465306	p.Pro134Leu	0.00008 ¹	4/5	∧	NCR	P
	5931	0	0		NA	rs114110943	p.His366Tyr	0.006 ^{1b}	2/5	∧	NCR
<i>IQGAP3</i>	4093	0	0	NA	rs147754283	p.Arg630Trp	0.00005 ^{1a}	3/5	∧	POS to neu	Likely P
	5886	2	0	0.1094	rs112144116	p.Ala562Thr	0.0034 ^{1a}	4/5	∧	NCC	P
<i>BAHCC1</i>	4093	0	0	NA	rs369588790	p.Arg2199Gln	0.00006 ^{1a}	2/5	NCR	NEG to neu	Likely P
	5931	0	0	NA	rs200719992	p.Gln2463Glu	0.0066 ^{1b}	4/5	∨	POS to neu	Likely P
<i>GLI3</i>	4093, 4130, 4132	6 ^	4	0.7946	rs35364414	p.Arg1537Cys	0.0536 ^{1a}	4/5	∨	POS to neu	B
<i>FLNB</i>	4132	2 + 1 !	0	0.3398	rs116826041	p.Ile2319Thr	0.0093 ^{1a}	3/5	∨	NCR	B
<i>KMT2D</i>	5463	0	0	NA	rs146044282	p.Asp3419Gly	0.0015 ^{1a}	3/5	∨	NCR	Likely P

Note. * significant, ¹ global MAF, ^{1a} Subpop = non-Finnish European MAF, ^{1b} Subpop = African MAF (all from gnomAD v2.1.1.); ² Fisher's Exact Test for Count Data compared allele count in unrelated probands (including discovery pedigree probands) to appropriate subpopulation gnomAD MAF; ³ # of damaging in silico scores include: SIFT, Poly-Phen2, MutationTaster, PROVEAN, and Mutation Assessor; ⁴ Unreadable sequences for 35 probands in nucleotides surrounding the variant; ^ same proband represented, ! Additional discovery proband (M4299 in family 4379; WES output showed that the variant was called in the proband, M4299, and their sibling who was unaffected on the TEGI (M4304) and therefore not prioritized under filtering workflow 2); AA = amino acid; HOPE (Have yOur Protein Explained) output: ∧ = size of AA increased, ∨ = size of AA decreased, + = more hydrophobic, NCR = no change reported, Charge change: POS = positive, neu = neutral, NEG = negative, NCC = no change in charge, NCR = no change reported; Causality classifications: pathogenic (P) = (1) MAF < 0.05 (in gnomAD v2.1.1 exomes) AND (2) EITHER co-segregating OR carried by >1 proband OR some significant change to amino acid structure AND (3) positive GERP score (conserved) AND (4) CADD Phred score ≥ 20 AND (5) ≥2 damaging in silico prediction scores (of those analyzed ³) AND (6) a change in size or charge in the amino acid according to HOPE output; benign (B) = missing one of the 6 classification criteria All in silico scores were acquired using the hg19 locations prior to 15 April 2022. I converted the hg38 locations using the 'Lift Genome Annotations' tool within UCSC Genome Browser [53].

3.3. Significance of Identified Variants in Additional Unrelated Probands

We Sanger sequenced 12 variants in nine genes (workflow 1 = nine variants in six genes and workflow 2 = three variants in three genes) in the additional probands ($n = 146$; Table 1). We observed variants in four genes from workflow 1 (*PDHA2*, *PCDHB3*, *NOL6*, and *IQGAP3*) in six additional probands (Table 1). A variant (rs35364414) in the candidate gene, *GLI3* from filtering workflow 2 was confirmed in 10 additional probands; four of these probands were not affected according to their TEGI composite or screener performance (Table 1). Another variant in *FLNB* was observed in two additional probands (Table 1). We performed a Fisher's exact test comparing the variant counts in the total number of discovery probands ($n = 8$) and additional probands sequenced ($n = 146$) to the variant counts reported in gnomAD for the non-Finnish European subpopulation. The observed variants on *GLI3*, *FLNB*, *PDHA2*, *PCDHB3*, *FURIN*, and *IQGAP3* in unrelated probands were not significantly different from the gnomAD reports ($p > 0.05$; Table 1). The variant observed on *NOL6* in family 4130 and an additional proband (rs114465306) has not been observed in the non-Finnish European subpopulation (Table 1).

In total, 17 additional probands carried a prioritized variant on a previously reported gene or a newly prioritized gene (Table 1). One additional proband carried two variants (on *GLI3* and *PDHA2*); the proband was affected according to their TEGI composite and receptive vocabulary performance but performed well on an omnibus measure and the TEGI screener. Four of the 17 probands were unaffected on both the TEGI composite and screener and all carried the common (gnomAD non-Finnish European subpopulation MAF = 0.0502) variant on *GLI3* (rs35364414; Table 1). Though the sample size of additional probands is small ($n = 146$) and the majority are affected on all four phenotypes of interest ($n = 77$; Table S7), the probands unaffected on the TEGI composite and screener were more likely to carry the common variant than a rare variant (Table 1).

4. Discussion

This family-based molecular genetic study of SLI defined the phenotype based on low performance on the Test of Early Grammatical Impairment (TEGI). The TEGI measures a particular part of the English grammar and has high specificity and sensitivity for distinguishing between children with and without SLI between the ages of 3 to 8; 11 years [15,24,25]. The TEGI phenotype allowed for the prioritization of rare variants on multiple genes not previously suggested for SLI and precise detection of variants involved in SLI as defined by a grammar phenotype. The identification of multiple rare and common variants within single families supports the hypothesis that a single variant (or even only rare variants) may not be able to explain the genetic basis of SLI on its own. Further, the current study underlines a continued role for family-based genetic study in the pursuit of genes involved in disordered language acquisition.

Select individuals in the current study were included in four previous genetic investigations [18,19,31,36]. The first utilized the TEGI phenotype; Rice, Smith, and Gayán (2009) performed targeted linkage and association analyses of regions previously associated with a reading disorder (RD) in a large portion of the KU SLI cohort ($N = 322$). The relatedness was not explicitly accounted for in the analyses. There was significant linkage at chr6p22 and marginally significant linkage at a portion of the targeted chr3p12-q13 region to the TEGI composite phenotype [31]. We also filtered and sequenced four variants on three genes in these regions, but the variants did not co-segregate with the TEGI phenotype in the respective families (Table S14a–e). We note that other phenotypes were linked to these regions, e.g., both the receptive vocabulary and omnibus phenotypes were linked to a region of markers on chromosome 6. Early LIs are predictive of later RDs [54,55], such that low language performance at a young age (when the TEGI is administered) was likely correlated with the other phenotypes. This may mean that the combination of low performance on the TEGI and low performance on other phenotypes was likely driving the linkage. The focus on only TEGI performance at the WES variant filtering level and the reduced number of WES families may have limited the power of the regions to target

variants of interest. However, the lack of variants of interest in the RD regions previously linked to the TEGI does not suggest that targeted investigation of related phenotypes is not of value. Findings from related phenotypes should always be considered, especially given that common causal pathways may be identified through this consideration and that targeted investigation adds confirming and disconfirming evidence for existing reports.

Across the nine genes of interest, *FURIN*, *BAHCC1*, *NOL6*, *FLNB*, and *KMT2D* show higher brain expression than *PCDHB3*, *IQGAP3*, *PDHA2*, and *GLI3* [56]. We present supporting evidence for *PCDHB3* and *NOL6* according to their functions and previous reports.

The protein product of *PCDHB3* is protocadherin beta 3. Protocadherins are a cadherin subfamily composed of three gene clusters, *PCDH- α* , *PCDH- β* , and *PCDH- γ* [57]. Cadherins and protocadherins are a large family of proteins involved in diverse functions, like hearing, balance, and neurodevelopmental and neurological processes among mammals [57]. The highest expression of these genes was observed in the nervous system [57]. Beta protocadherins localize to the synapse junctions during early development in mice, demonstrating their significance of neuronal connections in the mammalian nervous system [58]. Interestingly, the genetic investigation of a male child with severe non-syndromic language delay showed an intergenic deletion (220 Kb) at the homologous region of chromosomes X and Y spanning *PCDH11X/Y* [59]. A genetic study of another child with a sexual developmental disorder with severe language impairment and autistic behavior reported a concurrent deletion of *PCDH11Y* and *NLGN4Y*, indicating the role of protocadherin in developing even syndromic language impairment [60]. Another genetic study of a multiplex family with dyslexia observed ancestral genetic variations in *PCDHG*, showing the role of developing reading skills in humans [61]. More interestingly, the broad specificity of the antibodies to the isoforms was localized to the cortical areas related to language, indicating the significance of the splicing mechanisms in brain tissues for regulating the posttranscriptional regulation of protocadherins and other necessary transcripts for their diversity in the brain tissues [19,62]. We speculate that such family-based studies provide an excellent opportunity to replicate and discover new gene targets involved in language acquisition.

NOL6 (nucleolar protein 6) is essential in the biogenesis of ribosomes. Ribosomes are an integral component of protein synthesis in all cells. The ribosomal RNAs and proteins complete the biogenesis of ribosomes in the nucleolus. The translational efficiency of ribosomes is determined by several factors, including ribosome assembly and how the mRNAs load to the ribosomes. This translational efficiency is variable in the complex neuronal structures creating variability in the protein expression [63]. Deficits in ribosome biogenesis can result in multiple neurodevelopmental disorders. Neuronal cell types and the developmental period in which the deficiency was experienced determine the pathological consequences of these deficits [64]. We identified likely pathogenic variants in *NOL6* in multiple families, leading to our prediction that it is involved in gene pathways suggested in language impairment.

4.1. Limitations

A few key limitations of the current study should be noted: the lack of a grammar phenotype in parents and the possible missingness due to using WES versus whole-genome sequencing (WGS). The family-based approach was limited by the lack of a grammar phenotype in parents and any children who entered the study after the age 8; 11 years. However, given that rare variants of interest were identified without parent grammar phenotypes, we predict a study including grammar phenotypes in parents would be even more powerful. It is always important to consider possible missingness due to the genetic method utilized. Variants called by WES and WGS have been compared, showing about 3% of coding variants present in the WGS output were not present in the WES output [65]. This means additional coding variants segregating with the TEGI could have been missed due to low coverage and higher false positive call rate in the WES vs WGS.

4.2. Future Directions

In the future, grammar phenotypes capturing grammar at the same precise level as the TEGI should be utilized for older ages. Behavioral evidence suggests that such precise measurement is possible. In adolescence, at 15 and 16 years old, measurement of correct and incorrect grammatical judgments of questions where 'be' and 'do' were omitted is specific for the extended optional infinitive phenotype and shows high heritability in twins [4,66].

Additionally, future study should sequence all coding regions of genes prioritized for sequencing in the additional probands. Additional criteria may need to be considered before determining how many of the genes should be sequenced in full. Other criteria could include the expression data compiled from databases and provided in the current report, or additional expression data from BrainSpan about expression in the fetal brain [67]. Testing the larger proband group for additional rare variants in the novel candidate genes would allow for gene level testing of the rate of rare missense or loss of function variants. While variant level comparison can be informative, it can also be dependent on multiple factors of the downstream analysis, as shown by the recent WES investigations of family 5886 and 4075. Gene level significance testing provides stronger evidence for suggesting a candidate gene for a disorder, which should be considered carefully. However, the importance of variant level significance and a variant's possible deleterious role within their given causal pathway should not be disregarded in favor of gene-level significance.

In the long term, the newly suggested genes in this study can be helpful in determining where to look in neuroimaging for differences in groups with and without SLI. For example, the suggested genes can be further queried in gene pathway databases. One such database is STRING db, which assesses functional protein association networks [68]. Using the example of *NOL6*, STRING output shows strong connections to 10 other genes, all interconnected. The genes in the associated pathways could be checked in the existing WES output for variants and brain expression of these genes could be evaluated to further narrow the search for causal pathways of language acquisition and disordered language acquisition.

4.3. Implications of Family-Based Genetic Study for Understanding Factors Involved in Language Development

Long-standing questions surrounding the rapid acquisition of language by humans (i.e., adult-like language by age five in most individuals) have prompted investigations from multiple perspectives and an ongoing debate concerning the extent to which language ability is inherited [2,69,70]. Behavioral genetic studies of twins and families first showed the significant role of genetics in language acquisition and molecular genetic studies followed [1–6,18,19,30,32,33]. The current study added to the growing literature of molecular genetic studies by specifically targeting genetic influences on abstract shared grammar scaffolding under the assumption that humans have a specific universal aptitude toward language, such that our findings could contribute to the larger discussion of the role of genetic and environmental factors at play specifically in children's rapid early acquisition of complex structures in English. The investigation targeted individuals with SLI showing low performance on the TEGI, which measures a deficit in tense and agreement production or finiteness marking in English grammar [15].

Using performance on the TEGI ensured a precise phenotype. Precise phenotyping is required for precise genetic investigation and this study adds to the literature with a test of a precise phenotype that consistently shows significant heritability estimates [3–5,30]. The significant heritability estimates reported in 16-year-old twin pairs who completed a grammaticality judgment task indicates that the influence of genetics in grammar remains past the age measured by the TEGI [4]. Additionally, a recent report of individuals with SLI, which includes those from the current study and showed that the same pattern of difficulty in finiteness marking extends through age 18 when participants complete age-appropriate more challenging tasks concerning their understanding of complex linguistics structures [29]. Both results in older age groups support the importance of precisely defining this grammatical impairment in behavioral and genetic research.

The current study cross-referenced previously reported candidate genes identified from studies of individuals with SLI and related phenotypes. Similarly, the genes identified from our family-based investigation provide additional targets for future studies of larger cohorts of unrelated individuals, which are the focus of many other groups studying the genetics of language [21–23], just as the foundational findings of linkage to chromosome 16q and 19q to SLI have been continually cross-referenced in results from both families and larger and larger cohorts of individuals with language and reading phenotypes [21,36,51,71,72].

Our focus on possible genetic contributions to SLI recognizes the full complexities of possible causal pathways. Children’s language acquisition unfolds in the context of social and cognitive dimensions of development, along with other health-related factors. Identification of possible genetic pathways can identify sources of individual variance that further our understanding of individual differences that influence language development, which could aid in individualizing implementation of effective therapeutic approaches, which may include parent counseling, specialized peer social settings, focus on the cognitive underpinnings of vocabulary development, and other elements of a comprehensive intervention approach. Socioeconomic status is one such individual difference that is commonly considered in studies of language acquisition [73–76].

Children with SLI can be confused with children from low-income families, where “low income” is an indirect index of familial social resources for young children. Numerous studies found that children in families with lower-than-average incomes could be delayed in language acquisition [74,77]. For example, vocabulary development is predicted by maternal education in our longitudinal studies of vocabulary in children with and without SLI [78]. The design of this family-based study of SLI allowed us to explore this possible association, using maternal education as a proxy for family social resources in a total of 191 child participants, $n = 175$ SLI affected and $n = 16$ unaffected child family members in the eight discovery pedigrees (i.e., families selected for predominance of affectedness). In this supplemental analysis, children were grouped according to affectedness status on their omnibus language performance at entry to the study (consistent with the proband entrance criteria reported in the Methods and previous publications [18,19,31]). Maternal education was scored on a scale of 1 = some high school, no diploma, 2 = high school graduate diploma or GED, 3 = some college, no degree, 4 = bachelor’s degree, 5 = some graduate studies, and 6 = graduate degree. Thus, if low social resources, as indexed by maternal education, were driving the high proportion of affected children, we would expect low levels of maternal education in the SLI group. The means on the education scale were the unaffected group $M = 2.94$ ($SD = 1.00$) and SLI affected group, $M = 3.03$ ($SD = 1.28$). An independent samples t -test that accounted for variance at the family level showed no statistically significant group differences in maternal education: $t(42.7) = -1.33$, $p = 0.190$, indicating that the findings from our sample do not support a close association of maternal education and SLI.

5. Conclusions

Our family-based investigation prioritized multiple genes not previously suggested for SLI based on the sharing of rare variants between unrelated individuals affected by the TEGI. These findings indicate the TEGI phenotype has the potential to play a vital role in the future genetic inquiry of SLI. More broadly, focusing on the TEGI phenotype at the genetic level can inform causal pathways involved in language acquisition and the genetic underpinnings of brain structures uniquely provided to humans.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/children10071119/s1>, Table S1: Candidate Gene List ($n = 113$, based on a combination of reviews); Table S2: Distribution of Affectedness Across the Four Possible Phenotypes in the Whole exome Sequenced Individuals ($n = 34$); Table S3: Distribution of Affectedness Across the Four Possible Phenotypes in the Additional Proband ($n = 146$); Table S4: Family-specific Criteria for ‘family-specific variant comparison lists’; Table S5: Family-specific Criteria applied for ‘co-segregating variant lists’; Table S6: Primers used for confirmation via sequencing; Table S7: Number of Rare Variants: Familywise Filtering Workflow 1; Table S8: 47 Genes Resulting from Cross-Referencing of ‘Family-Specific Variant Comparison Lists’; Table S9: Summary of 44 Variants on 23 Genes Prioritized from Filtering Workflow 1 Cross-referenced List; Table S10: Summary of Prioritized Variants from Filtering Workflow 1 Sequenced and Confirmed in Family Members; Table S11a: 18 Rare Variants Prioritized in Family 4093 Shared by All Four Family Members Affected on the TEGI; Table S11b: 31 Rare Variants Prioritized in Family 4130 Shared by Both Family Members Affected on the TEGI; Table S11c: 31 Rare Variants Prioritized in Family 4132 Shared by Two or More Family Members from Different Branches Affected on the TEGI; Table S11d: 37 Rare Variants Prioritized in Family 4075 Shared by Three or More of the Four Family Members Affected on the TEGI; Table S11e: 17 Rare Variants Prioritized in Family 4379 Shared by Two or More of the Three Family Members Affected on the TEGI; Table S11f: 6 Rare Variants of Large Effect Prioritized in Family 5463 Shared by All Three of the Four Family Members Affected on the TEGI; Table S11g: 23 Rare Variants Prioritized in Family 5886 Shared by All Four of the Four Family Members Affected on the TEGI; Table S11h: 18 Rare Variants Prioritized in Family 5931 Shared by Both Family Members Affected on the TEGI; Table S12: Number of Candidate Gene Variants: Familywise Filtering Workflow 2b; Table S13: Summary of Confirmation Notes for the 13 Variants Prioritized from Filtering Workflow 2 Output Sequenced in Family Members; Table S14a: Number of Variants on Chromosome 3q13.12-q13.31: Familywise Filtering Workflow 2a; Table S14b: Variants prioritized familywise within RD region: chr3q13.12-q13.31, previously significantly associated with TEGI phenotype; Table S14c: Number of Variants on Chromosome 6p21.1-p22.3: Family Filtering Workflow 2a; Table S14d: Variants prioritized familywise within RD region: chr6p21.1-p22.33q13.12-q13.31, previously significantly associated with TEGI phenotype; Table S14e: Primers used for confirmation of variants in RD regions via sequencing.

Author Contributions: M.L.R. is PI of the Kansas Cohort. Conceptualization, M.H.R. and M.L.R.; methodology, E.M.A., H.X., C.Z., M.H.R. and M.L.R.; software, H.X. and C.Z.; formal analysis, E.M.A., K.K.E., H.X., C.Z., M.H.R. and M.L.R.; investigation, E.M.A., K.K.E., H.X., C.Z., M.H.R. and M.L.R.; resources, M.H.R. and M.L.R.; data curation, E.M.A. and K.K.E.; writing—original draft preparation, E.M.A. and M.H.R.; writing—review and editing, E.M.A., K.K.E., C.Z., M.H.R. and M.L.R.; visualization, E.M.A. and M.H.R.; supervision, M.H.R. and M.L.R.; project administration, E.M.A., K.K.E., M.L.R. and M.H.R.; funding acquisition, M.L.R. Most of the effort by E.M.A. on the project occurred while at the University of Kansas, Lawrence, KS as a doctoral candidate in the Child Language Doctoral Program. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Institute on Deafness and Other Communication Disorders (T32DC000052 and R01DC001803; MLR) and (R21DC017830; MHR). During the bench work completed for this project, EMA was supported by a grant from the National Institute on Deafness and Other Communication Disorders (R21DC017830, PI: Raza).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Kansas (IRB #8223), for behavioral data collection on 25 January 1993, and has been annually reviewed and approved since that time. Relevant genetics amendments included: (1) Collection of DNA via blood draw and cheek cell samples, approved on 19 January 1999; (2) Collection of DNA via saliva samples, approved on 31 January 2006; (3) consent form update to include the National Institutes of Health Certificate of Confidentiality, approved on 3 January 2018. Participants provide their signatures for informed consent to all amended genetics protocols.

Informed Consent Statement: Appropriate informed consent was obtained from all participants.

Data Availability Statement: We have provided parts of the dataset generated during and/or analyzed for our study in the Supplemental Material. Additional parts are available from the corresponding author on reasonable request.

Acknowledgments: We acknowledge the children and families of the KU Cohort who contributed to this research. We also thank the examiners in the Language Acquisition Studies Lab and students in the Child Language Doctoral Program for collecting behavioral data and saliva samples. The KU Cohort is funded by R01DC001803 (awarded to MLR, principal investigator). We would like to thank Shelley Smith at the Medical Center, University of Nebraska for initial bioinformatic processing of the WES data. The first round of whole-exome sequencing (WES) was performed at the University of Nebraska Medical Center (UNMC) Genomics Core Facility. The UNMC Genomics Core Facility receives partial support from the National Institute for General Medical Science (NIGMS) INBRE—P20GM103427-19, as well as the National Cancer Institute The Fred & Pamela Buffett Cancer Center Support Grant- P30CA036727, The Center for Root and Rhizobiome Innovation (CRRRI) 36-5150-2085-20, and the Nebraska Research Initiative. This publications' contents are the sole responsibility of the authors and do not necessarily represent the official views of the NIH or NIGMS. The second round of WES was performed at the University of Kansas Medical Center Genomics Core. The University of Kansas Medical Center Genomics Core is supported by the following NIH grants—Kansas Intellectual and Developmental Disabilities Research Center (NIH U54 HD 090216), the Molecular Regulation of Cell Development and Differentiation—COBRE (5P20GM104936-10) and the NIH S10 High-End Instrumentation Grant (NIH S10OD021743) at the University of Kansas Medical Center, Kansas City, KS 66160. In addition, the initial consultation and library prep for the second round of WES was completed at the KU Genome Sequencing Core. Therefore, research reported in this dissertation was made possible in part by the services of the KU Genome Sequencing Core. This lab is supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health under award number P20GM103638. Finally, the article processing charges related to the publication of this article were supported by The University of Kansas (KU) One University Open Access Author Fund sponsored jointly by the KU Provost, KU Vice Chancellor for Research, and KUMC Vice Chancellor for Research and managed jointly by the Libraries at the Medical Center and KU—Lawrence.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript or in the decision to publish the results.

References

1. Tallal, P.; Hirsch, L.S.; Realpe-Bonilla, T.; Miller, S.; Brzustowicz, L.M.; Bartlett, C.W.; Flax, J.F. Familial Aggregation in Specific Language Impairment. *J. Speech Lang. Hear. Res.* **2001**, *44*, 1172–1182. [[CrossRef](#)]
2. Stromswold, K. Genetics of Spoken Language Disorders. *Hum. Biol.* **1998**, *70*, 297–324.
3. Bishop, D.V.M.; Adams, C.V.; Norbury, C.F. Distinct Genetic Influences on Grammar and Phonological Short-term Memory Deficits: Evidence from 6-year-old Twins. *Genes Brain Behav.* **2006**, *5*, 158–169. [[CrossRef](#)]
4. Dale, P.S.; Rice, M.L.; Rimfeld, K.; Hayiou-Thomas, M.E. Grammar Clinical Marker Yields Substantial Heritability for Language Impairments in 16-Year-Old Twins. *J. Speech Lang. Hear. Res.* **2018**, *61*, 66–78. [[CrossRef](#)]
5. Rice, M.L.; Taylor, C.L.; Zubrick, S.R.; Hoffman, L.; Earnest, K.K. Heritability of Specific Language Impairment and Nonspecific Language Impairment at Ages 4 and 6 Years across Phenotypes of Speech, Language, and Nonverbal Cognition. *J. Speech Lang. Hear. Res.* **2020**, *63*, 793–813. [[CrossRef](#)]
6. Rice, M.L.; Haney, K.R.; Wexler, K. Family Histories of Children with SLI Who Show Extended Optional Infinitives. *J. Speech Lang. Hear. Res.* **1998**, *41*, 419–432. [[CrossRef](#)] [[PubMed](#)]
7. National Institute of Deafness and Other Communication Disorders Developmental Language Disorder. Available online: <https://www.nidcd.nih.gov/health/developmental-language-disorder> (accessed on 8 May 2023).
8. Tomblin, J.B.; Records, N.L.; Buckwalter, P.; Zhang, X.; Smith, E.; O'Brien, M. Prevalence of Specific Language Impairment in Kindergarten Children. *J. Speech Lang. Hear. Res.* **1997**, *40*, 1245–1260. [[CrossRef](#)] [[PubMed](#)]
9. Norbury, C.F.; Gooch, D.; Wray, C.; Baird, G.; Charman, T.; Simonoff, E.; Vamvakas, G.; Pickles, A. The Impact of Nonverbal Ability on Prevalence and Clinical Presentation of Language Disorder: Evidence from a Population Study. *J. Child. Psychol. Psychiatry* **2016**, *57*, 1247–1257. [[CrossRef](#)]
10. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T.; Consortium, C. CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLoS ONE* **2016**, *11*, e0158753. [[CrossRef](#)] [[PubMed](#)]
11. Bishop, D.V.M.; Snowling, M.J.; Thompson, P.A.; Greenhalgh, T.; CATALISE-2 consortium. Phase 2 of CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study of Problems with Language Development: Terminology. *J. Child Psychol. Psychiatry* **2017**, *58*, 1068–1080. [[CrossRef](#)]
12. Leonard, L.B. *Children with Specific Language Impairment*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2014. ISBN 9780262324021.
13. *Developmental Language Disorders: From Phenotypes to Etiologies*; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2004; pp. xi–411, ISBN 978-0-8058-4662-1.

14. Mountford, H.S.; Braden, R.; Newbury, D.F.; Morgan, A.T. The Genetic and Molecular Basis of Developmental Language Disorder: A Review. *Children* **2022**, *9*, 586. [[CrossRef](#)]
15. Rice, M.L.; Wexler, K. *Rice/Wexler Test of Early Grammatical Impairment Examiner's Manual*; The Psychological Corporation: San Antonio, TX, USA, 2001.
16. Rice, M.L.; Wexler, K.; Hershberger, S. Tense Over Time: The Longitudinal Course of Tense Acquisition in Children with Specific Language Impairment. *J. Speech Lang. Hear. Res.* **1998**, *41*, 1412–1431. [[CrossRef](#)]
17. Rice, M.L.; Wexler, K. Toward Tense as a Clinical Marker of Specific Language Impairment in English-Speaking Children. *J. Speech Lang. Hear. Res.* **1996**, *39*, 1239–1257. [[CrossRef](#)]
18. Andres, E.M.; Earnest, K.K.; Smith, S.D.; Rice, M.L.; Raza, M.H. Pedigree-Based Gene Mapping Supports Previous Loci and Reveals Novel Suggestive Loci in Specific Language Impairment (SLI). *J. Speech Lang. Hear. Res.* **2020**, *63*, 4046–4061. [[CrossRef](#)]
19. Andres, E.M.; Earnest, K.K.; Zhong, C.; Rice, M.L.; Raza, M.H. Family-Based Whole-Exome Analysis of Specific Language Impairment (SLI) Identifies Rare Variants in BUD13, a Component of the Retention and Splicing (RES) Complex. *Brain Sci.* **2022**, *12*, 47. [[CrossRef](#)]
20. Peter, B.; Bruce, L.; Finestack, L.; Dinu, V.; Wilson, M.; Klein-Seetharaman, J.; Lewis, C.R.; Braden, B.B.; Tang, Y.-Y.; Scherer, N.; et al. Precision Medicine as a New Frontier in Speech-Language Pathology: How Applying Insights from Behavior Genomics Can Improve Outcomes in Communication Disorders. *Am. J. Speech-Lang. Pathol.* **2023**, 1–16. [[CrossRef](#)] [[PubMed](#)]
21. Eising, E.; Mirza-Schreiber, N.; de Zeeuw, E.L.; Wang, C.A.; Truong, D.T.; Allegrini, A.G.; Shapland, C.Y.; Zhu, G.; Wigg, K.G.; Gerritse, M.L.; et al. Genome-Wide Analyses of Individual Differences in Quantitatively Assessed Reading- and Language-Related Skills in up to 34,000 People. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2202764119. [[CrossRef](#)] [[PubMed](#)]
22. Nudel, R.; Christensen, R.V.; Kalnak, N.; Schwinn, M.; Banasik, K.; Dinh, K.M.; Erikstrup, C.; Pedersen, O.B.; Burgdorf, K.S.; Ullum, H.; et al. Developmental Language Disorder—a Comprehensive Study of More than 46,000 Individuals. *Psychiatry Res.* **2023**, *323*, 115171. [[CrossRef](#)] [[PubMed](#)]
23. Toseeb, U.; Vincent, J.; Oginni, O.A.; Asbury, K.; Newbury, D.F. The Development of Mental Health Difficulties in Young People with and without Developmental Language Disorder: A Gene–Environment Interplay Study Using Polygenic Scores. *J. Speech Lang. Hear. Res.* **2023**, *66*, 1639–1657. [[CrossRef](#)]
24. Ash, A.C.; Redmond, S.M. Using Finiteness as a Clinical Marker to Identify Language Impairment. *Perspect. Lang. Learn. Educ.* **2014**, *21*, 148–158. [[CrossRef](#)]
25. Weiler, B.; Schuele, C.M. Tense Marking in the Kindergarten Population: Testing the Bimodal Distribution Hypothesis. *J. Speech Lang. Hear. Res.* **2021**, *64*, 593–612. [[CrossRef](#)] [[PubMed](#)]
26. Weiler, B.; Schuele, C.M.; Feldman, J.I.; Krimm, H. A Multiyear Population-Based Study of Kindergarten Language Screening Failure Rates Using the Rice Wexler Test of Early Grammatical Impairment. *Lang. Speech Hear. Serv. Sch.* **2018**, *49*, 248–259. [[CrossRef](#)]
27. Wexler, K. Optional Infinitives, Head Movement and the Economy of Derivations. In *Verb Movement*; Lightfoot, D., Hornstein, N., Eds.; Cambridge University Press: Cambridge, UK, 1994.
28. Rice, M.L.; Wexler, K.; Cleave, P.L. Specific Language Impairment as a Period of Extended Optional Infinitive. *J. Speech Lang. Hear. Res.* **1995**, *38*, 850–863. [[CrossRef](#)] [[PubMed](#)]
29. Rice, M.L.; Earnest, K.K.; Hoffman, L. Longitudinal Grammaticality Judgments of Tense Marking in Complex Questions in Children with and without Specific Language Impairment, Ages 5–18 Years. *J. Speech Lang. Hear. Res.* *in press*.
30. Rice, M.L.; Zubrick, S.R.; Taylor, C.L.; Hoffman, L.; Gayán, J. Longitudinal Study of Language and Speech of Twins at 4 and 6 Years: Twinning Effects Decrease, Zygosity Effects Disappear, and Heritability Increases. *J. Speech Lang. Hear. Res.* **2018**, *61*, 79–93. [[CrossRef](#)]
31. Rice, M.L.; Smith, S.D.; Gayán, J. Convergent Genetic Linkage and Associations to Language, Speech and Reading Measures in Families of Proband with Specific Language Impairment. *J. Neurodev. Disord.* **2009**, *1*, 264–282. [[CrossRef](#)]
32. Chen, X.S.; Reader, R.H.; Hoischen, A.; Veltman, J.A.; Simpson, N.H.; Francks, C.; Newbury, D.F.; Fisher, S.E. Next-Generation DNA Sequencing Identifies Novel Gene Variants and Pathways Involved in Specific Language Impairment. *Sci. Rep.* **2017**, *7*, 1–17. [[CrossRef](#)]
33. Villanueva, P.; Nudel, R.; Hoischen, A.; Fernández, M.A.; Simpson, N.H.; Gilissen, C.; Reader, R.H.; Jara, L.; Echeverry, M.M.; Francks, C.; et al. Exome Sequencing in an Admixed Isolated Population Indicates NFXL1 Variants Confer a Risk for Specific Language Impairment. *PLoS Genet.* **2015**, *11*, 1–24. [[CrossRef](#)]
34. Nudel, R. An Investigation of NFXL1, a Gene Implicated in a Study of Specific Language Impairment. *J. Neurodev. Disord.* **2016**, *8*, 13. [[CrossRef](#)]
35. MacArthur, D.G.; Manolio, T.A.; Dimmock, D.P.; Rehm, H.L.; Shendure, J.; Abecasis, G.R.; Adams, D.R.; Altman, R.B.; Antonarakis, S.E.; Ashley, E.A.; et al. Guidelines for Investigating Causality of Sequence Variants in Human Disease. *Nature* **2014**, *508*, 469–476. [[CrossRef](#)]
36. Martinelli, A.; Rice, M.L.; Talcott, J.B.; Diaz, R.; Smith, S.D.; Raza, M.H.; Snowling, M.J.; Hulme, C.; Stein, J.; Hayiou-Thomas, M.E.; et al. A Rare Missense Variant in the ATP2C2 Gene Is Associated with Language Impairment and Related Measures. *Hum. Mol. Genet.* **2021**, *30*, 1160–1171. [[CrossRef](#)] [[PubMed](#)]
37. Adzhubei, I.A.; Schmidt, S.; Peshkin, L.; Ramensky, V.E.; Gerasimova, A.; Bork, P.; Kondrashov, A.S.; Sunyaev, S.R. A Method and Server for Predicting Damaging Missense Mutations. *Nat. Methods* **2010**, *7*, 248–249. [[CrossRef](#)]

38. Choi, Y.; Chan, A.P. PROVEAN Web Server: A Tool to Predict the Functional Effect of Amino Acid Substitutions and Indels. *Bioinformatics* **2015**, *31*, 2745–2747. [[CrossRef](#)]
39. Schwarz, J.M.; Cooper, D.N.; Schuelke, M.; Seelow, D. MutationTaster2: Mutation Prediction for the Deep-Sequencing Age. *Nat. Methods* **2014**, *11*, 361–362. [[CrossRef](#)]
40. Sim, N.L.; Kumar, P.; Hu, J.; Henikoff, S.; Schneider, G.; Ng, P.C. SIFT Web Server: Predicting Effects of Amino Acid Substitutions on Proteins. *Nucleic Acids Res.* **2012**, *40*, W452–W457. [[CrossRef](#)] [[PubMed](#)]
41. Reva, B.; Antipin, Y.; Sander, C. Predicting the Functional Impact of Protein Mutations: Application to Cancer Genomics. *Nucleic Acids Res.* **2011**, *39*, e118. [[CrossRef](#)] [[PubMed](#)]
42. Rentzsch, P.; Witten, D.; Cooper, G.M.; Shendure, J.; Kircher, M. CADD: Predicting the Deleteriousness of Variants throughout the Human Genome. *Nucleic Acids Res.* **2018**, *47*, D886–D894. [[CrossRef](#)]
43. Cooper, G.M.; Stone, E.A.; Asimenos, G.; Program, N.C.S.; Green, E.D.; Batzoglou, S.; Sidow, A. Distribution and Intensity of Constraint in Mammalian Genomic Sequence. *Genome Res.* **2005**, *15*, 901–913. [[CrossRef](#)]
44. Dong, C.; Wei, P.; Jian, X.; Gibbs, R.; Boerwinkle, E.; Wang, K.; Liu, X. Comparison and Integration of Deleteriousness Prediction Methods for Nonsynonymous SNVs in Whole Exome Sequencing Studies. *Hum. Mol. Genet.* **2015**, *24*, 2125–2137. [[CrossRef](#)]
45. Andres, E.M.; Neely, H.L.; Hafeez, H.; Yasmin, T.; Kausar, F.; Basra, M.A.R.; Raza, M.H. Study of Rare Genetic Variants in TM4SF20, NFXL1, CNTNAP2, and ATP2C2 in Pakistani Proband and Families with Language Impairment. *Meta Gene* **2021**, *30*. [[CrossRef](#)]
46. Hendam, A.; Al-Sadek, A.F.; Hefny, H.A. In Silico Deleterious Prediction of Nonsynonymous Single Nucleotide Polymorphisms in Neurexin1 Gene for Mental Disorders. *Int. J. Bioinform. Res. Appl.* **2020**, *16*, 1–24. [[CrossRef](#)]
47. Ding, J.; Miao, Q.-F.; Zhang, J.-W.; Guo, Y.-X.; Zhang, Y.-X.; Zhai, Q.-X.; Chen, Z.-H. H258R Mutation in KCNAB3 Gene in a Family with Genetic Epilepsy and Febrile Seizures Plus. *Brain Behav.* **2020**, *10*, e01859. [[CrossRef](#)] [[PubMed](#)]
48. Wickham, H.; François, R.; Henry, L.; Müller, K. Dplyr: A Grammar of Data Manipulation. R Package Version 1.0.7. 2022. Available online: <https://dplyr.tidyverse.org> (accessed on 8 May 2023).
49. Wickham, H.; Girlich, M. Tidy: Tidy Messy Data. R Package Version 1.2.0. 2022. Available online: <https://tidyr.tidyverse.org/> (accessed on 8 May 2023).
50. R Core Team R: A Language and Environment for Statistical Computing. 2016. Available online: <https://www.R-project.org/> (accessed on 8 May 2023).
51. Mountford, H.S.; Villanueva, P.; Fernández, M.A.; De Barbieri, Z.; Cazier, J.B.; Newbury, D.F. Candidate Gene Variant Effects on Language Disorders in Robinson Crusoe Island. *Ann. Hum. Biol.* **2019**, *46*, 109–119. [[CrossRef](#)] [[PubMed](#)]
52. Guerra, J.; Cacabelos, R. Genomics of Speech and Language Disorders. *J. Transl. Genet. Genom.* **2019**. [[CrossRef](#)]
53. Lee, B.T.; Barber, G.P.; Benet-Pages, A.; Casper, J.; Clawson, H.; Diekhans, M.; Fischer, C.; Gonzalez, J.N.; Hinrichs, A.S.; Lee, C.M.; et al. The UCSC Genome Browser Database: 2022 Update. *Nucleic Acids Res.* **2022**, *50*, D1115–D1122. [[CrossRef](#)]
54. Adlof, S.M.; Hogan, T.P. If We Don't Look, We Won't See: Measuring Language Development to Inform Literacy Instruction. *Policy Insights Behav. Brain Sci.* **2019**, *6*, 210–217. [[CrossRef](#)]
55. Catts, H.W.; Adlof, S.M.; Ellis Weismer, S. Language Deficits in Poor Comprehenders: A Case for the Simple View of Reading. *J. Speech Lang. Hear. Res.* **2006**, *49*, 278–293. [[CrossRef](#)] [[PubMed](#)]
56. GTEx Consortium The Genotype-Tissue Expression (GTEx) Project. *Nat. Genet.* **2013**, *45*, 580–585. [[CrossRef](#)]
57. Jaiganesh, A.; Narui, Y.; Araya-Secchi, R.; Sotomayor, M. Beyond Cell–Cell Adhesion: Sensational Cadherins for Hearing and Balance. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a029280. [[CrossRef](#)]
58. Junghans, D.; Heidenreich, M.; Hack, I.; Taylor, V.; Frotscher, M.; Kemler, R. Postsynaptic and Differential Localization to Neuronal Subtypes of Protocadherin Beta16 in the Mammalian Central Nervous System. *Eur. J. Neurosci.* **2008**, *27*, 559–571. [[CrossRef](#)]
59. Speevak, M.D.; Farrell, S.A. Non-Syndromic Language Delay in a Child with Disruption in the Protocadherin11X/Y Gene Pair. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **2011**, *156*, 484–489. [[CrossRef](#)]
60. Nardello, R.; Antona, V.; Mangano, G.D.; Salpietro, V.; Mangano, S.; Fontana, A. A Paradigmatic Autistic Phenotype Associated with Loss of PCDH11Y and NLGN4Y Genes. *BMC Med. Genom.* **2021**, *14*, 98. [[CrossRef](#)]
61. Naskar, T.; Faruq, M.; Banerjee, P.; Khan, M.; Midha, R.; Kumari, R.; Devasenapathy, S.; Prajapati, B.; Sengupta, S.; Jain, D.; et al. Ancestral Variations of the PCDHG Gene Cluster Predispose to Dyslexia in a Multiplex Family. *EBioMedicine* **2018**, *28*, 168–179. [[CrossRef](#)] [[PubMed](#)]
62. Priddle, T.H.; Crow, T.J. Protocadherin 11X/Y a Human-Specific Gene Pair: An Immunohistochemical Survey of Fetal and Adult Brains. *Cereb. Cortex* **2013**, *23*, 1933–1941. [[CrossRef](#)] [[PubMed](#)]
63. Dastidar, S.G.; Nair, D. A Ribosomal Perspective on Neuronal Local Protein Synthesis. *Front. Mol. Neurosci.* **2022**, *15*, 823135. [[CrossRef](#)] [[PubMed](#)]
64. Hetman, M.; Slomnicki, L.P. Ribosomal Biogenesis as an Emerging Target of Neurodevelopmental Pathologies. *J. Neurochem.* **2019**, *148*, 325–347. [[CrossRef](#)]
65. Belkadi, A.; Bolze, A.; Itan, Y.; Cobat, A.; Vincent, Q.B.; Antipenko, A.; Shang, L.; Boisson, B.; Casanova, J.L.; Abel, L. Whole-Genome Sequencing Is More Powerful than Whole-Exome Sequencing for Detecting Exome Variants. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5473–5478. [[CrossRef](#)]

66. Rice, M.L.; Hoffman, L.; Wexler, K. Judgments of Omitted BE and DO in Questions as Extended Finiteness Clinical Markers of Specific Language Impairment (SLI) to 15 Years: A Study of Growth and Asymptote. *J. Speech Lang. Hear. Res.* **2009**, *52*, 1417–1433. [[CrossRef](#)]
67. Miller, J.A.; Ding, S.L.; Sunkin, S.M.; Smith, K.A.; Ng, L.; Szafer, A.; Ebbert, A.; Riley, Z.L.; Royall, J.J.; Aiona, K.; et al. Transcriptional Landscape of the Prenatal Human Brain. *Nature* **2014**, *508*, 199–206. [[CrossRef](#)]
68. Szklarczyk, D.; Gable, A.L.; Nastou, K.C.; Lyon, D.; Kirsch, R.; Pyysalo, S.; Doncheva, N.T.; Legeay, M.; Fang, T.; Bork, P.; et al. The STRING Database in 2021: Customizable Protein-Protein Networks, and Functional Characterization of User-Uploaded Gene/Measurement Sets. *Nucleic Acids Res.* **2021**, *49*, D605–D612. [[CrossRef](#)]
69. Crain, S. Language Acquisition in the Absence of Experience. *Behav. Brain Sci.* **1991**, *14*, 597–612. [[CrossRef](#)]
70. Ambridge, B.; Lieven, E.V.M. *Child Language Acquisition: Contrasting Theoretical Approaches*; Cambridge University Press: Cambridge, UK, 2011; ISBN 978-1-139-50051-7.
71. SLI Consortium A Genomewide Scan Identifies Two Novel Loci Involved in Specific Language Impairment**Members of the Consortium Are Listed in the Appendix. *Am. J. Hum. Genet.* **2002**, *70*, 384–398. [[CrossRef](#)]
72. SLI Consortium Highly Significant Linkage to the SLI1 Locus in an Expanded Sample of Individuals Affected by Specific Language Impairment. *Am. J. Hum. Genet.* **2004**, *74*, 1225–1238. [[CrossRef](#)]
73. Hoff, E.; Tian, C. Socioeconomic Status and Cultural Influences on Language. *J. Commun. Disord.* **2005**, *38*, 271–278. [[CrossRef](#)] [[PubMed](#)]
74. Hoff, E. The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Dev.* **2003**, *74*, 1368–1378. [[CrossRef](#)]
75. Schuele, C.M. Socioeconomic Influences on Children’s Language Acquisition. *J. Speech-Lang. Pathol. Audiol.* **2001**, *25*, 77–88.
76. Stromswold, K. Why Aren’t Identical Twins Linguistically Identical? Genetic, Prenatal and Postnatal Factors. *Cognition* **2006**, *101*, 333–384. [[CrossRef](#)] [[PubMed](#)]
77. Piot, L.; Havron, N.; Cristia, A. Socioeconomic Status Correlates with Measures of Language Environment Analysis (LENA) System: A Meta-Analysis. *J. Child Lang.* **2022**, *49*, 1037–1051. [[CrossRef](#)]
78. Rice, M.L.; Hoffman, L. Predicting Vocabulary Growth in Children with and without Specific Language Impairment: A Longitudinal Study from 2;6 to 21 Years of Age. *J. Speech Lang. Hear. Res.* **2015**, *58*, 345–359. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.