ESHG

ARTICLE

A genome-wide analysis in consanguineous families reveals new chromosomal loci in specific language impairment (SLI)

Erin M. Andres [ORCID][1] · Huma Hafeez[2] · Adnan Yousaf[1,3] · Sheikh Riazuddin[4] · Mabel L. Rice[1] · Muhammad Asim Raza Basra[2] · Muhammad Hashim Raza[1]

Abstract

Language is a uniquely human ability, and failure to attain this ability can have a life-long impact on the affected individuals. This is particularly true for individuals with specific language impairment (SLI), which is defined as an impairment in normal language development in the absence of any other developmental disability. Although SLI displays high heritability, family-based linkage studies have been hampered by an unclear mode of Mendelian segregation, variable disease penetrance, and heterogeneity of diagnostic criteria. We performed genome-wide parametric linkage analysis and homozygosity mapping in 14 consanguineous families from Pakistan segregating SLI. Linkage analysis revealed a multipoint LOD score of 4.18 at chromosome 2q in family PKSLI05 under a recessive mode of inheritance. A second linkage score of 3.85 was observed in family PKSLI12 at a non-overlapping locus on chromosome 2q. Two other suggestive linkage loci were found in family PKSLI05 on 14q and 22q with LOD scores of 2.37 and 2.23, respectively, that were also identified in homozygosity mapping. Reduction to homozygosity was observed on chromosomes 2q, 5p, 8q, 14q, 17q, and 22q. Each homozygosity region occurred in multiple PKSLI families. We report new SLI loci on chromosomes 2 and 8 and confirm suggestive SLI linkage loci on chromosomes 5, 14, 17, and 22 reported previously in the population of Robinson Crusoe Island. These findings indicate that linkage and homozygosity mapping in consanguineous families can improve genetic analyses in SLI and suggest the involvement of additional genes in the causation of this disorder.

Introduction

Specific language impairment (SLI) is defined as a delay in the mastery of language despite normal intelligence and no other developmental or hearing problems (NIDCD, 2018). SLI is common in children, with a prevalence of 7% in the United States [1]. Affected individuals often have life-long

✉ Muhammad Hashim Raza
    razam@ku.edu

[1] Child Language Doctoral Program (CLDP), University of Kansas, Lawrence, KS, USA

[2] Institute of Chemistry, University of the Punjab, Lahore, Pakistan

[3] Department of Biotechnology, International Islamic University, Islamabad, Pakistan

[4] Allama Iqbal Medical College, Lahore, Pakistan

issues with academics and in relationships with peers, and are reportedly prone to other behavioral and social difficulties [2]. The language abilities of individuals with SLI are assessed through multiple measures of both expressive and receptive language, grammar, vocabulary, and reading [3]. Psychometrically robust tests are not available for all languages. Of interest here is a widely used measure of vocabulary comprehension that requires pointing to pictures, with a large research literature on English-speaking children. The Peabody Picture Vocabulary Test (PPVT) is available in the fourth edition (PPVT-4) [4] with scores benchmarked to age expectations. Children with SLI are reported to score lower than unaffected children on the PPVT in longitudinal data from 2 to 18 years of age [5].

Twin and family aggregation studies show that genetic factors contribute to SLI [6–8]. Two independent family aggregation studies reported higher recurrence rate of SLI in proband's families compared with control families [1, 9, 10]. A longitudinal study of language and speech in twins generated heritability estimates ranging from 0.44 to 0.92 at 6 years of age across multiple assessments [11]. This

study reported heritability of 0.54 for the PPVT-3 for 6-year-old twins. Segregation studies are not well established in SLI but pedigree analysis has shown that SLI tends to follow complex, rather than simple Mendelian inheritance patterns [8, 12].

Linkage refers to the co-inheritance of alleles at loci that are close together and transmitted together in a family as a unit. Linkage studies are used to identify the trait loci that co-segregate with the specific genomic regions within families using polymorphic markers with known chromosomal positions. LOD scores (Logarithm of odds) are used to quantify linkage that allows an estimate of recombination frequency between a marker locus and the disease locus. High values of LOD scores favor the linkage hypothesis. A LOD score of 3 or more is taken as statistically significant evidence of linkage for Mendelian traits, while a LOD score of 1.5 or higher considered to be suggestive [13]. A LOD score of −2 or less is considered a significant evidence against linkage [13]. Pedigree-based parametric linkage analysis has historically been a powerful method for mapping genetic variants of large effect even with an incomplete disease penetrance and an undefined mode of inheritance [14]. Such an approach has proven successful in mapping genetic loci in complex genetic disorders [15, 16]. Family-based linkage studies have mapped SLI susceptibility loci on chromosome 16q24, chromosome 19q13 [17], chromosome 13q21 [18], and on chromosome 7q35–36 [19]. The linkage on chromosome 13q21 was replicated in US and Canadian SLI samples. However, this chromosomal region is shared with a locus associated with the autism spectrum disorders (ASD) phenotype [20]. Similarly, a founder population inhabiting Robinson Crusoe Island (RCI), off the coast of Chile, was utilized for genome-wide linkage analysis and homozygosity mapping [21]. Linkage analysis of the RCI families resulted in five regions with significant LOD scores, including chromosomes 6q, 7q, 12, 13, and 17 [21]. These studies have used model-based and family-based linkage analyses and obtained strong linkage loci. Another family-based genome-wide association study revealed a significant maternal parent of origin effect on chromosome 14q12 and suggestive paternal parent of origin effect on chromosome 5p13 [22]. This study emphasizes the parent of origin effects in SLI families.

Homozygosity mapping is a method that has the potential to identify regions of homozygosity present in the genome of individuals with recessive disorders. Regions of homozygosity in the genome of such individuals denote inheritance of DNA segments from an ancestor common to both parents to their descendants [23]. While this method has been highly successful in mapping disease genes for simple Mendelian recessive disorders in consanguineous families [24], it has also been effectively used for complex Mendelian disorders like SLI [21]. Family-based

homozygosity mapping has been less widely used in complex Mendelian disorders. However, population-based genome-wide homozygosity association mapping has been successfully used in several complex genetic disorders including Schizophrenia, bipolar disorder, and late-onset Alzheimer's disease [25–27]. In the study reported here, we performed family-based homozygosity mapping and parametric linkage analysis in consanguineous SLI families from Pakistan. The high rate of consanguinity in Pakistan favors the appearance of genetic disorders that have a recessive mode of inheritance, as observed in several genetic disorders, including hereditary deafness, intellectual disability (ID), and stuttering [24, 28, 29]. The primary goal of this project was to identify loci linked with SLI using extended consanguineous families. We enrolled families with multiple cases of SLI from Pakistan and used parametric linkage analysis and family-based homozygosity mapping to both confirm previous SLI loci and identify multiple novel SLI loci.

Methods

Identification of SLI families and phenotype assignment

Families were identified through public schools in Punjab, Pakistan. The study was approved by the University of Kansas Institutional Review Board (IRB #8223). To identify children with language impairment, we sent a brief introductory letter to school teachers, which included information to identify individuals with SLI from schools. The family history questionnaire developed by Rice and colleagues was used to prepare this Performa (information to identify individuals with SLI) for school teachers in Pakistan [30]. The statements were intended to help teachers identify children who might have differences in language abilities compared with their peers. The intent is teachers serve as a first-step in screening a clinically defined sample of children to be identified as probands for the subsequent family recruitment. Teachers were instructed not to include children with known developmental disabilities or overt signs of hearing impairment or neurological disorders. Direct behavioral assessments were conducted by trained researchers based in Pakistan. Follow-up assessments included interviews with the parents, a family history questionnaire (PhenX protocol #200401) and a version of the PPVT-4 developed for this study [30]. The PPVT-4 is a receptive vocabulary assessment in which the examiner provides a word orally and the participant chooses the picture that best describes the word from a grid of four pictures [4]. Individuals with speech deficits can complete the PPVT-4 because no spoken language is required and the

format is also suitable for persons who might be shy in the assessment setting. No version of the PPVT is available locally in Urdu, a native primary language of Pakistan. For the purpose of phenotyping in this study, the PPVT-4 test items were translated into Urdu for a U-PPVT-4 (Urdu), with the same number of items, sequence of presentation, and ways of obtaining a beginning and ending point for the test. The translated version of words for the test was vetted with the local research team and confirmed by local speakers. In order to obtain external standards for interpretation of the test scores we used the standard scores for the English-speaking population, allowing for estimates of performance relative to age, where a standard score of 100 is the mean, 15 is a standard deviation, and a standard score of 85 has been defined as affected in previous studies of children with SLI [5]. To date, we have collected data on a total of 116 local Urdu-speaking individuals ages 2.5–18 years on the U-PPVT-4 and parental report of a history of late talking which is a known antecedent of SLI. We evaluated the psychometric properties of U-PPVT-4 by examining the distribution within this clinically ascertained sample, comparing caregiver reports of being slow to talk to U-PPVT-4 outcomes, and examining the correlation of U-PPVT-4 and slow to talk. The scores of the clinically referred sample were normatively distributed, with a mean of 70.12 and a standard deviation of 22.33. As demonstrated in other family-based studies of clinically ascertained probands [5], this clinically ascertained sample distribution is shifted downward. For the purpose of our study we selected a standard score of 80 as a definition of "affected" because that value is a break in the distribution at the upper end. We adjusted further for females given their reduced literacy rate compared with males [31]. Correlation was calculated for a positive parental report of a child as slow to talk and the child's classification as affected on the U-PPVT-4, $r = 0.374$, $p = 0.000$. The "slow to talk" caregiver screener item correctly identified 37% of the children defined as affected on U-PPVT-4. Although the psychometric properties of U-PPVT-4 are not as robust as the English version, the preliminary indications of reliability and validity are encouraging and offer an opportunity for measuring individual differences in a phenotype within family pedigrees for a study of possible genetic differences. We note that any errors in identification of truly affected individuals would work against detection of genetic associations.

A total of 157 individuals from 14 families were included in the study, and the DNA was available from 156 of these. The U-PPVT-4 standard score was obtained from 96 individuals, of which 62 were assigned as affected and 34 were unaffected (data not shown). The family history questionnaire was used as described previously [30]. The questionnaire asked parents to answer yes/no questions about difficulties in learning to read, enjoyment of reading,

spelling, and general storytelling behaviors of all family members. Extended family members were assessed with the U-PPVT-4 and their saliva samples collected using Oragene DNA 500 kits (Fig. 1).

DNA purification and whole-genome SNP genotyping

Saliva samples were collected from participants at the time of behavioral assessment, using the Oragene-Discover OGR-500 Kit from DNA Genotek (Oragene). DNA was extracted following the protocol provided by the manufacturer (https://www.dnagenotek.com/us/products/collection-human/oragene-discover/500-series/OGR-500.html). DNA samples from 156 individuals were genotyped using the Illumina Infinum QC Array-24, containing 15,949 SNPs by the Johns Hopkins University School of Medicine, Genetic Resources Core Facility. The SNPs are distributed evenly throughout the genome. On the autosomal chromosomes, there are 11,994 SNPs and the rest are distributed to the sex chromosomes and the mitochondrial chromosome. The SNP genotype data of two DNA samples was excluded from the analysis due to a high rate of Mendelian inconsistency (parent–child and parent–parent–child error). The genotype call rate was 99.96% when the data of two individuals of inconsistent relationship was removed. The quality of genotype calls was consistent with the call rate of 6 CEPH samples used as positive controls. A total of 11,925 SNPs were used for linkage analysis as well as for homozygosity mapping. There were 49 additional SNPs excluded from the analysis due to low intensity of heterozygote SNPs.

Linkage analysis

Genome-wide parametric linkage analysis (single-point and multipoint) was done using SUPERLINK ONLINE SNP 1.1 (http://cbl-hapw.cs.technion.ac.il/superlink-snp/) [32]. Additional quality control on the data was performed during the linkage analysis. For example, inheritance inconsistencies at the pedigree level and homozygosity SNPs with low informativity were checked during linkage analysis. We did not observe a consistent Mendelian inheritance pattern in our PKSLI families. However, the transmission of SLI more frequently followed an autosomal recessive inheritance model, as expected in consanguineous families. Linkage analysis was performed on all families under the recessive mode of inheritance with variable disease penetrance as described below. Large extended families, PKSLI05 and PKSLI30 were divided into branches, a–d and a–e, respectively. Branches of extended large SLI families were analyzed individually as well as in combination, assuming genetic heterogeneity within families as previously described [15]. Traditional LOD score
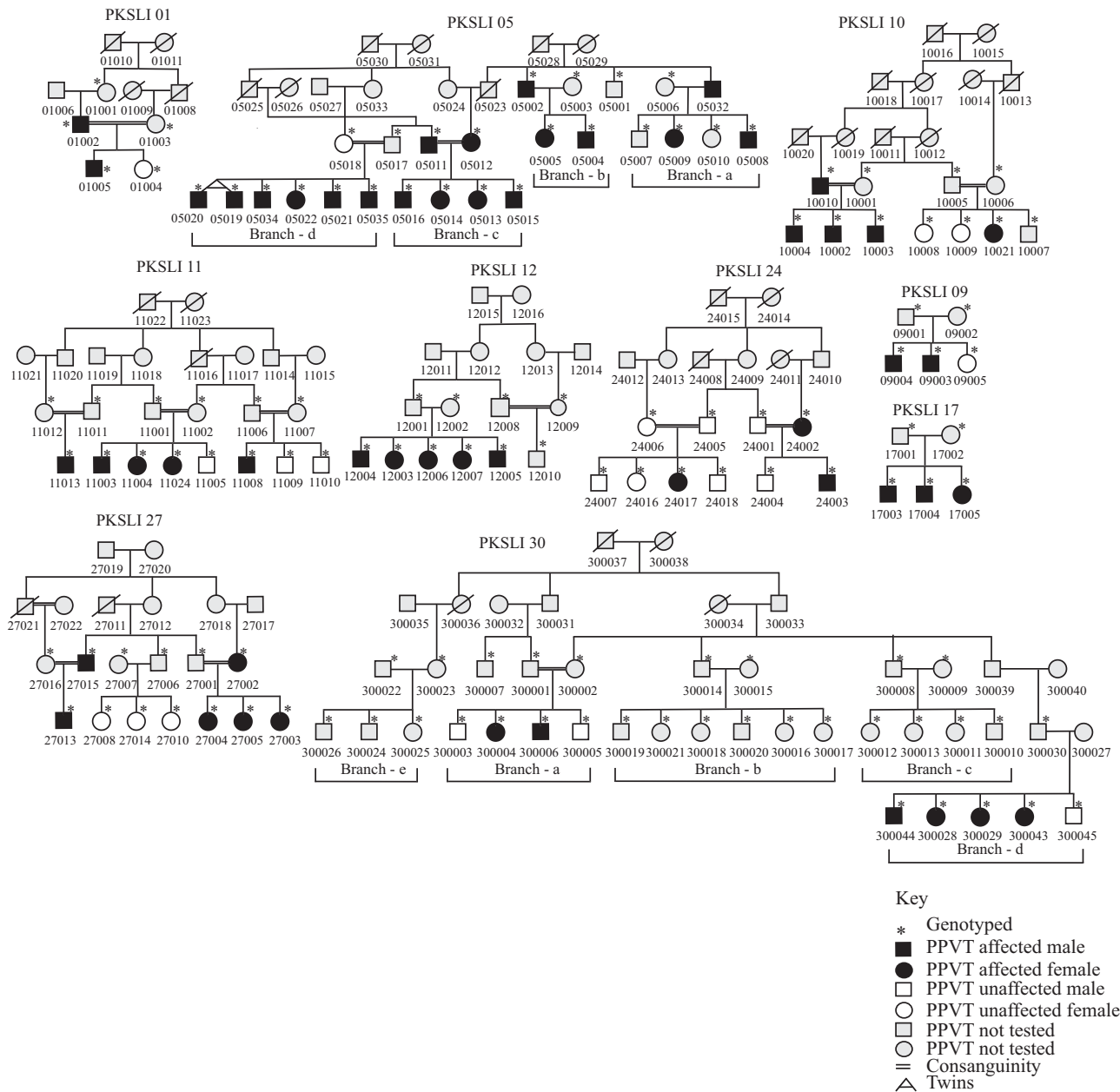
Fig. 1 Pedigrees of Pakistani specific language impairment (PKSLI) families. Large families are divided into branches as labeled in brackets. PKSLI 05: Branch-a includes individuals 05028, 05029, 05001, 05032, 05006, 05007, 05008, 05009, 05010. Branch-b includes individuals 05002, 05003, 05004, 05005. Branch-c includes individuals 05030, 05031, 05023, 05024, 05025, 05026, 05011, 05012, 05013, 05014, 05015, 05016. Branch-d includes individuals 05030, 05031, 05025, 05026, 05027, 05033, 05017, 05018, 05019, 05020, 05021, 05022, 05034, 05035. PKSLI 30: Branch-a includes individuals 300031, 300032, 300033, 300034, 300037, 300038, 300001, 300002, 300007, 300003, 300004, 300005, 300006. Branch-b includes individuals 300014, 300015, 300016, 300017, 300018, 300019, 300020, 300021. Branch-c includes individuals 300008, 300009, 300010, 300011, 300012, 300013. Branch-d includes individuals 300027, 300030, 300028, 300029, 300043, 300044, 300045. Branch-e includes individuals 300022, 300023, 300024, 300025, 300026

calculations under defined modes of inheritance provide more power to test linkage in complex genetic disorders. Statistical linkage significance testing was done using multiple penetrance functions previously reported in complex disorders [15, 33, 34]. These studies proposed that the LOD score threshold of genome-wide significance should be increased by 0.3 ($=\log_{10}2$) for each additional disease penetrance function tested. Affection status was categorically assigned in all the PKSLI families based on the U-PPVT-4 adjusted standard scores (Fig. 1). Under recessive inheritance, variable disease penetrance (0.99, 0.80, 0.70) was set for individuals with two copies of the

susceptibility alleles. The LOD scores are reported assuming a disease allele frequency of 0.001 and affection status on the U-PPVT-4. The LOD scores were calculated at the recombination fraction $\theta = 0$. The adjusted threshold of significance and suggestive linkage may be increased to 3.9 and 2.4, respectively, if we add 0.3 to each additional penetrance test.

Homozygosity mapping

Homozygosity mapping was performed using HomozygosityMapper (http://www.homozygositymapper.org/), which calculates the length of homozygous blocks across SNP markers [35]. We performed this analysis in each family independently. Allele frequency was set to zero and the option of genetic homogeneity was selected, with the aim of identifying homozygous regions shared in affected individuals. The homozygous stretches were excluded if the block length was longer than 20 markers in controls [35]. Runs of homozygosity <4 Mb are reportedly common in outbred populations [23]. Therefore, we initially selected regions of homozygosity >4 Mb in all PKSLI families, which were later tested in other PKSLI families for shared homozygosity regions (HRs). The HRs over 4 Mb and those identified in 2 or more affected individuals were subsequently investigated in other PKSLI families. There were two goals of doing this: (1) the regions of homozygosity were not considered significant if found in 2 or more unaffected individuals among all 14 families analyzed, (2) identification of other PKSLI families in which 2 or more affected individuals share the same HR.

The linkage loci and homozygosity regions were submitted to the LOVD database https://databases.lovd.nl/shared/individuals (Individual IDs: 206608, 222897, 223937, 224970–224974, 226087, 226089–226091).

Results

Most PKSLI families failed to generate pairwise LOD scores >1 (data not shown) during linkage analysis. Two families, PKSLI05 and PKSLI12 (Fig. 1) generated pairwise LOD scores over 2 on chromosomes 2q33 and 2q37, respectively (Tables 1 and 2). The highest LOD score of 3.2 was obtained on chromosome 2q33.3 in PKSLI05 branch-d. There were two other chromosomal regions (14q22 and 22q13) identified in PKSLI05 branch-d that showed a maximum LOD score of 1.77 under an autosomal recessive inheritance model (Table 1). Linkage was not detected in other branches of PKSLI05 and PKSLI30. Additive linkage analysis was also performed, including all PKSLI families and across all branches of PKSLI05 (Supplementary Figs. 2 & 3). Under the additive analysis of PKSLI families,

Table 1 Summary of linkage loci identified in PKSLI families under recessive inheritance model

Linkage loci (hg19)	Max. pairwise LOD score—complete penetrance	Max. multipoint LOD score, complete penetrance	Max. pairwise LOD score—reduced penetrance (0.70)	Mode of inheritance	Family/ branch	Homozygosity
2q33.3-q35	3.28	4.18	3.27	Recessive	PKSLI05 -branch-d	No
2q37.1-q37.2	3.05	3.85	2.28	Recessive	PKSLI12	No
14q22.3-q24.3	1.77	2.37	1.77	Recessive	PKSLI05 -branch-d	Yes
22q13.31-q13.32	1.77	2.23	1.77	Recessive	PKSLI05 -branch-d	Yes

hg19 = Human Genome Assembly 19 (Feb. 2009)

Table 2 Pairwise LOD scores at multiple markers in 2q33.3-2q35 in full PKSLI05 and branch-d under autosomal recessive inheritance model

Chr.	SNP ID	hg19 position (Mb)	LOD score PKSLI05—branch-d—complete penetrance	LOD score PKSLI05—branch-d—reduced penetrance (0.7)	LOD score PKSLI05 full—complete penetrance	LOD score PKSLI05 full reduced penetrance (0.7)
2	rs717973	207.97	−1.8147	−1.8186	−4.5245	−4.5245
2	rs1446596	210.29	1.803	1.7952	1.8599	1.8521
2	rs10208979	210.34	1.803	1.7952	1.3911	1.3834
2	rs731953	210.68	3.2821	3.2743	−3.0983	−3.106
2	rs1396828	211.05	1.803	1.7952	0.6506	0.6429
2	rs4356603	211.51	−0.3228	−0.3268	0.1054	0.1054
2	rs714393	212.70	1.7768	1.7768	−1.8314	−1.8314
2	rs2054615	213.48	1.803	1.7952	2.4883	–
2	rs13420690	215.24	−0.3689	−0.3585	−0.8236	−0.8306
2	rs2229571	215.65	−1.8147	−1.8186	−0.0809	−0.0809
2	rs1880586	216.20	1.803	1.7952	2.0291	2.0212
2	rs921970	219.21	1.803	1.7952	1.633	1.6252
2	rs9076	219.21	1.803	1.7952	1.5919	1.5842
2	rs585185	219.46	1.7768	1.7768	−2.2111	−2.2111
2	rs1554622	219.61	3.2821	3.2743	−0.1944	−0.2022
2	rs1109867	220.08	−1.8147	−1.8186	−2.192	−2.192

hg19 = Human Genome Assembly 19 (Feb. 2009), Mb = Megabases

there were multiple loci including chromosomes 2 and 22 where markers showed the highest LOD scores above 3. However, multiple markers on several chromosomes showed higher LOD scores (> 2) that indicates background noise and confounding results. The additive linkage analysis across the branches of PKSLI05 identified the highest LOD scores at chromosome 2q33 under the recessive inheritance model. The highest additive LOD score at chromosome 2q33 shows PKSLI 05 branch-d contributes most significantly when the analysis was done on each branch independently.

We identified six HRs in 10 PKSLI families during homozygosity mapping. Each HR was shared among multiple PKSLI families except chromosome 17q (Table 1, Figs. 2 and 3). Two HRs were found in the mapped linkage regions on chromosomes 14q22 and 22q13.

Chromosome 2q

There were two linkage loci identified on chromosome 2q, one found in family PKSLI05 branch-d and the other in family PKSLI12 (Table 1). Two loci mapped on chromosome 2q were ~14.2 Mb apart. Both loci are separated by a region of negative LOD scores, showing they are two independent loci. Markers that showed the highest LOD score at 2q33.3-q35 in PKSLI branch-d were rs731953 and rs1554622. The LOD score decreased to 3.27 when analyzed at a reduced disease penetrance of 0.70 (Table 2). The
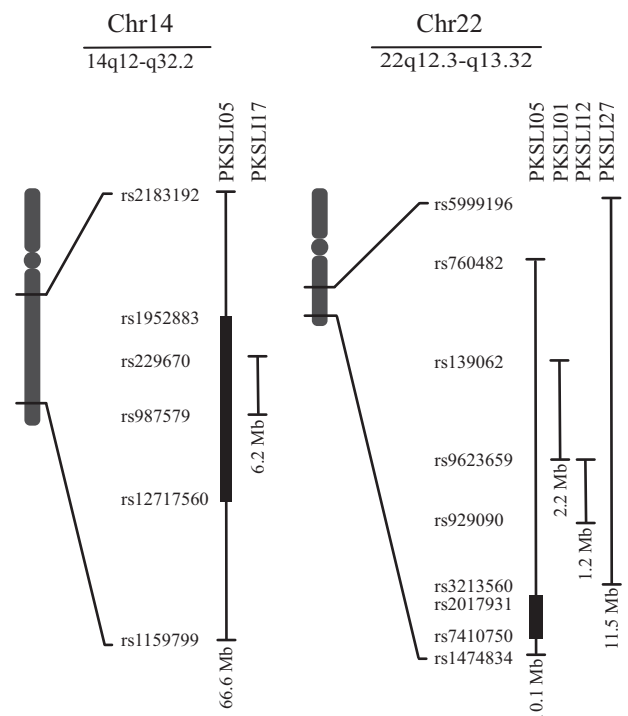

Fig. 2 Homozygosity and linkage regions identified in PKSLI05 on chromosomes 14q and 22q, which also showed evidence of convergent homozygosity in additional families. Vertical bars, designate homozygosity regions with their size in million base pairs (Mb). The solid rectangles indicate the regions showing linkage in PKSLI05. SNP markers show the boundaries of homozygosity and linkage regions
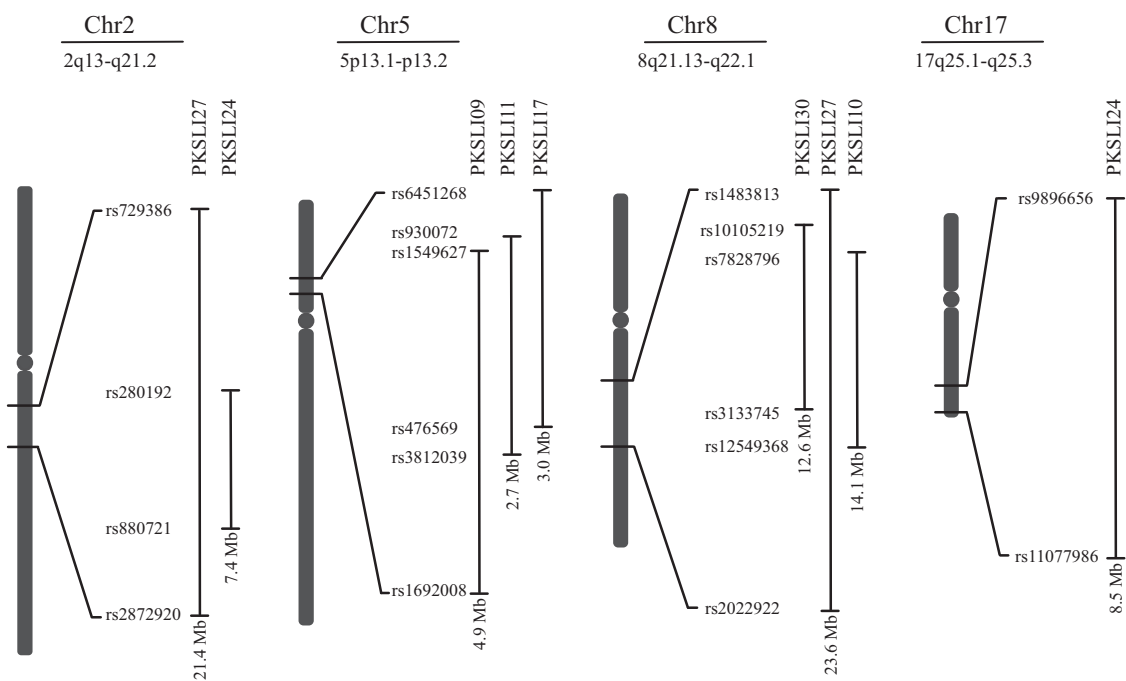
Fig. 3 Homozygous regions in PKSLI families identified on chromosomes 2, 5, 8, and 17. SNP IDs represent boundaries of the homozygous region in each family. Vertical lines show the corresponsding homozygosity regions (HR), underneath each verticle line the size of each region is mentioned in Mb. Mb (Megabases), Chr (chromosome), PKSLI (Paksitani SLI family)

LOD scores decreased significantly at most of the markers when the linkage analysis was performed in the full pedigree. The LOD scores went down to negative values on most of the markers in this locus when linkage analysis was performed in the other branches of family PKSLI05. Multipoint analysis of branch-d under a recessive mode of inheritance with complete disease penetrance resulted in a substantial increase in the LOD score at this locus, to 4.18.

The second linkage region on chromosome 2q, at 2q37.1-q37.2 (2.2 Mb), was identified in family PKSLI12. A maximum pairwise LOD score of 3.05 was obtained at rs1469375 under recessive inheritance with complete disease penetrance, and most of the markers in the vicinity showed a LOD score over 1. The LOD scores decreased when the analysis was performed under reduced disease penetrance. In multipoint analysis, a LOD score of 3.85 was identified at this locus in family PKSLI12 under recessive mode of inheritance with complete disease penetrance.

Homozygosity data mapping identified a 21.4 Mb homozygous region (HR) on chromosomes 2q13-q21.2 (rs729386 to rs2872920) in family PKSLI27 (Fig. 3). Four affected individuals (27002, 27003, 27004, and 27005) and one unaffected individual (27014) of PKSLI27 shared the HR on chromosome 2q (Table 3). A section (7.4 Mb) of the same HR was identified in two affected individuals (24002 and 24003) of family PKSLI24 (Fig. 3 and Table 3). This region was not found to be homozygous in individuals of other PKSLI families.

Chromosome 14q

We identified suggestive linkage with a maximum LOD score of 1.77 on chromosomes 14q22.3-q24.3 (rs1952883-rs12717560) in family PKSLI05 branch-d, under a recessive mode of inheritance with complete penetrance (Table 1). Multipoint linkage analysis at this locus in branch-d resulted higher LOD score of 2.37 under the same parameters (Table 1).

Homozygosity mapping produced additional support for this region where we identified a 66.6 Mb HR on chromosome 14 (14q12-q32.2) in family PKSLI05 branch-d. All U-PPVT-4 affected individuals in branch-d shared most of this HR except for individuals 05034 and 05035. The linkage region on chromosome 14q identified in branch-d overlaps with this HR as shown in Fig. 2. Upon investigating this HR of 66.6 Mb in other PKSLI families (as described in Methods), we found a 6.2 Mb shared homozygous region (rs229670 to rs987579) in two U-PPVT-4 affected (17003 and 17004) and one U-PPVT-4 unknown individual (17001) in family PKSLI17 (Fig. 2 and Table 3). The HR in PKSLI17 is also shared with the linkage region on chromosome 14q; however, this family was too small to generate significant linkage evidence.

Table 3 Summary of homozygosity regions in PKSLI families

Chromosome	Human Genome Assembly 19 (Feb. 2009) - Chromosomal coordinates and size	Family	Total number of individuals		Number of individuals homozygous	
			PPVT Affected	PPVT Unaffected/ unknown	PPVT Affected	PPVT Unaffected/ unknown
2	111,705,738-133,160,621 (21.4 Mb)	PKSLI27	6	7	4	1
2	121,737,198-129,153,304 (7.4 Mb)	PKSLI24	3	7	2	0
14	32,636,261-99,254,712 (66.6 Mb)	PKSLI05 branch-d	6	2	4	0
14	65,283,501-71,513,480 (6.2 Mb)	PKSLI17	3	2	2	0
22	39,178,701-49,324,829 (10.1 Mb)	PKSLI05 branch (a + c + d)	15	5	8	0
22	40,881,793-43,088,971 (2.2 Mb)	PKSLI01	2	3	2	0
22	43,088,971-44,313,849 (1.2 Mb)	PKSLI12	5	5	3	0
22	34,469,292-46,022,339 (11.5 Mb)	PKSLI27	6	7	3	0
5	36,676,161-41,585,557 (4.9 Mb)	PKSLI09	2	3	2	1
5	36,666,071-39,426,327 (2.7 Mb)	PKSLI11	5	6	4	1
5	36,255,364-39,342,308 (3.0 Mb)	PKSLI17	3	2	2	0
8	83,857,175-96,534,806 (12.6 Mb)	PKSLI30 branch (d + e)	4	7	4	1
8	77,844,519-101,472,165 (23.6 Mb)	PKSLI27	6	7	2	0
8	84,280,532-98,451,430 (14.1 Mb)	PKSLI10	5	6	3	0
17	71,774,648-80,280,865 (8.5 Mb)	PKSLI24	3	7	2	1

Chromosome 22q

Pairwise linkage analysis in family PKSLI05 branch-d identified a suggestive linkage on chromosomes 22q13.31-q13.32 (rs2017931 to rs7410750) with a LOD score of 1.77 under the recessive inheritance model with complete disease penetrance. Using multipoint analysis, the LOD score at this locus increased to 2.23 under the recessive mode of inheritance with complete disease penetrance (Table 1).

Homozygosity mapping identified an HR of 10.1 Mb on chromosomes 22q12.3-q13.32 (rs760482 to rs1474834) in family PKSLI05 (affected individuals 05008, 05009, 05016, 05019, 05020, 05021, 05022, and 05034) (Fig. 2 and Table 3). The linkage region of 1.66 Mb (22q13.31-q13.32) overlaps with the HR in PKSLI05. Upon investigating the HR in other families, two affected individuals (01002 and 01005) in family PKSLI01 shared the HR of 2.2 Mb (rs139062 to rs9623659) to the 10.1 Mb region on chromosome 22q. Three affected individuals (12003, 12004, and 12005) in family PKSLI12 shared an HR of 1.2 Mb (rs9623659 to rs929090) and three affected individuals (27003, 27004, and 27005) in PKSLI27 shared an HR of 11.5 Mb (rs5999196 to rs3213560) to the 10.1 Mb region on chromosome 22q (Fig. 2 and Table 3). The HR shared by these individuals was not found in individuals of other PKSLI families.

Chromosome 5p

Linkage was not identified on chromosome 5p. Homozygosity data mapping identified an HR of 4.9 Mb (rs1549627 to rs1692008) mapped on chromosomes 5p13.1-p13.2 in family PKSLI09. This HR was shared among affected individuals, 9003, 9004 and an unaffected individual 9005 (female) (Table 3). Upon investigating this HR in other PKSLI families, we identified a shared HR of 2.7 Mb (rs930072 to rs3812039) in four affected individuals and an unaffected individual of family PKSLI011 (Table 3). Two affected individuals from family PKSLI17 shared the HR of 5.3 Mb (rs6451268-rs 476569) on chromosome 5p (Fig. 3 and Table 3). This HR was not present in individuals of other PKSLI families.

Chromosome 8q

Linkage analysis did not reveal linkage on chromosome 8q in PKSLI families. Homozygosity mapping identified a region of 12.6 Mb (rs10105219 to rs3133745) on chromosomes 8q21.13-q22.1 in four U-PPVT-4 affected individuals of family PKSLI30 (Fig. 3 and Table 3). Upon further investigation in other PKSLI families, we identified a shared HR of 14.1 Mb (rs12549368 to rs7828796) on chromosome 8q in three affected individuals of family PKSLI10. The extended but shared HR of 23.6 Mb

(rs2022922 to rs1483813) on chromosome 8q was identified in two affected individuals of family PKSLI27 (Fig. 3 and Table 3). This HR was not observed in individuals of other PKSLI families.

Chromosome 17q

Linkage analysis did not reveal linkage to chromosome 17q. Homozygosity mapping identified an HR of 8.5 Mb (rs9896656 to rs11077986) on chromosomes 17q25.1-q25.3 in two U-PPVT-4 affected individuals of PKSLI24. This region was also identified in one U-PPVT-4 unaffected individual in this family (Fig. 3 and Table 3). This HR was not observed in other PKSLI families.

Discussion

Although twin studies have shown a high heritability for SLI, genetic studies are difficult due to a lack of Mendelian transmission in families, presumptive variable disease penetrance, availability of large extended families, and characterization of SLI phenotype. Studies in consanguineous families identified stuttering and SLI loci previously [21, 36, 37]. A stuttering locus on chromosome 12 led to identification of a causative gene for stuttering and subsequent sequencing of two functionally related genes in the lysosomal targeting pathway identified mutations in stuttering population that were found among 16% of stuttering cases [38, 39]. Similarly, variants in *NFXL1* were identified using a large consanguineous SLI family from RCI [12]. Our study differs from previous genetic studies of SLI in that we targeted consanguineous families from Pakistan, which has a very high rate of consanguinity. For phenotyping, as a screener we used teacher judgments of children with low language (relative to their peers), instructing them to identify children without other developmental neurological disorders. For the U-PPVT-4, the distribution and variation in the standard scores in our Urdu-speaking sample population is similar to that of the English-speaking population. The shifted distribution due to lower performance compared to age-matched peers is also similar to family-based SLI-proband-ascertained samples in English-speaking children. The family history questionnaire also asked parents to indicate if their child was a late talker, which can be indicative of later language ability though it is not perfectly correlated with SLI diagnosis. [40]. In population-ascertained samples, 20% of those who showed late language emergence at 24 months show general language ability deficits on standardized language measures at 7-years-old [40]. In our sample, the correlation between late talker status and affected status on the U-PPVT-4 was statistically significant. Further, 37% of the probands were

identified as late talkers, adding to the likelihood that the probands are perceived as less language proficient than unaffected children. Therefore, despite the lack of a standardized language measure in Urdu, the phenotype criteria distinguished between affected and unaffected individuals in the Pakistani population. A study of controls, or unaffected children, is in progress with the purpose of developing age estimates for adjusted U-PPVT-4 scores in the Pakistani population for research purposes.

The results of our study in consanguineous SLI families are consistent with previous genetic loci associated with SLI. There are four SLI loci (5p, 14q, 17q, and 22q) identified previously in the RCI population that we replicated in our PKSLI families. A maximum linkage score of 1–1.15 was achieved previously at 5p, 14q, and 22q under parametric linkage analysis. We achieved these regions with homozygosity mapping except in chromosome 14q where multipoint linkage of 2.37 was obtained as well. The locus on chromosome 17q was reported previously with a significant *p*-value under non-parametric linkage analysis. We did not identify linkage at 17q but found a HR that overlaps with the previously identified linkage region in the Chilean family. Our analysis identified SLI regions that were previously reported for the same phenotype, which may suggest the involvement of shared genes or gene pathways responsible of language impairment (Tables 1 and 3) [21]. Chromosomal loci, 14q and 22q are our potential candidates, as they are identified through linkage mapping as well as through homozygosity mapping. Other chromosomal loci, 5p and 17q are our suggestive loci as they are obtained through homozygosity mapping, though no linkage was observed in these loci. In the UK SLI families, chromosome 5p was reported as maternal parent-of-origin effects, while it was identified in the Chilean SLI family as a suggestive locus [21, 22]. These results suggest intrafamilial heterogeneity in our large PKSLI families, even with high rates of consanguinity. Our data suggest that multiple genes are implicated in familial SLI.

Linkage analysis was done in Canadian SLI families under autosomal dominant and recessive modes of inheritance with variable disease penetrance. A significant evidence of linkage was observed at 13q21 with reading impairment (RI) in Canadian SLI families. Two other loci, 2p22 and 17q23, produced suggestive linkage with language impairment (LI) and RI, respectively [18]. Our linkage loci reported on 2q33 and 2q37 do not overlap with the previously identified locus on chromosome 2p. The lack of overlap could be due to the use of different phenotype measures or these regions are simply two independent SLI loci, 2p22 with a stronger link to RI [18] and the 2q33 and 2q37 segregating with LI. Another study identified a 4 kb deletion in 2q36.3 that deleted a coding exon of *TM4SF20* in several South Asian families segregating early language

delay [41]. We did not find linkage or homozygosity of this region in our SLI families. However, the *TM4SF20* locus is located adjacent to the linkage locus we report on 2q (Table 1). This suggests additional studies may be helpful to refine these regions to chromosome 2q.

We obtained an evidence of linkage to chromosomes 2q33, 14q22, and 22q13 in PKSLI05 branch-d. The strongest linkage with a multipoint LOD score of 4.18 was identified at chromosome 2q33 but this LOD score was not obtained with the additive analysis of the whole family. In order to understand why the LOD scores increased in branch-d compared with the full PKSLI05 family, haplotypes were built on chromosome 2q33 (Supplementary Fig. 1). We used MERLIN to build haplotypes in all branches of this family and drew them in the pedigree using Haplopainter. The haplotypes were later analyzed and drawn manually in the full pedigree of PKSLI05 using Adobe Illustrator (Supplementary Fig. 1). Two haplotypes, red and dark blue are consistently present in all affected individuals of branch-d. It is obvious that both haplotypes are ancestrally inherited without any cross over events. It is likely that these haplotypes are inherited from ancestors 05026 and 05033, in which these haplotypes are inferred. The relationship of these individuals is unknown as drawn in PKSLI05 and the haplotypes ruled out the possibility of common ancestors of 05026 and 05033. Both haplotypes (red and dark blue) are not seen in other branches of PKSLI05. Distribution of the red haplotype in branch-d might be a fair explanation of the higher LOD score in this branch of the family.

There are several genes identified to each linkage region (Supplementary Tables 1–4). There were over 100 genes located on chromosome 2q linkage region (Supplementary Table 1). Two genes, *MAP2* and *CREB1* in this region-might be of interest to SLI. *MAP2* (microtubule-associated protein 2) plays a role in dendrite development in the developing brain [42]. A deletion of a region at 2q34 was reported in a 14-year old with ASD and symptoms of Rett syndrome [42]. Though these disorders differ significantly from SLI, *MAP2* could be an interesting candidate for SLI based on its involvement in dendritic development. The cAMP (cyclic adenosine monophosphate) pathway is involved in long-term memory formation [43]. There have been many reports of memory abilities in children with SLI. The SLI Consortium reported the first linkage analysis of SLI using the non-word repetition task, which measures a working memory, as one of the phenotype measures [17, 44]. *CREB1* regulates the transcription of multiple genes, including *BDNF* [45]. *BDNF*, located on chromosome 11, encodes brain-derived neurotrophic factor.

The linkage score on chromosome 14q caught our attention because the same region was found to be homozygous in the majority of affected individuals of branch-d. We identified several candidate genes based on their role in the process of intracellular trafficking. Some of them, among others, include *ACOT4, ACOT2*, and *AP5M1*. Acyl-CoA thioesterases (*ACOT4* and *ACOT2*) play an important role in maintaining the intracellular ratio between the activated form of free fatty acids and acyl-CoAs. Imbalance of the ratio of these molecules has serious consequence in the onset of chronic diseases, like diabetes and obesity, but this may also influence the signaling pathways in neurons [46]. The encoded protein of *AP5M1* is the *mu-1* subunit of the adaptor protein complex 5, which is of interest because stuttering mutations were reported in the *AP4E1*, another member of the AP family [47]. Adaptor proteins are involved in the sorting of selective transport of transmembrane proteins in the intracellular compartments. AP5 is one of the recently discovered adaptor protein complexes. Mutations in the *AP5* subunits cause a form of complicated hereditary paraplegia with intellectual disability and a thin corpus collosum [48]. The homozygosity region on chromosome 22q was mapped from 1.2 to 11.5 Mb (rs5999196 to rs1474834) in PKSLI05, PKSLI01, PKSLI12, and PKSLI27. There are six genes located in this linkage region, five of which are associated with noncoding RNA or microRNA (*LINC01644*, *LINC00898*, *LOC84930*, *LOC284933*, *MIR3201*) (LINC refers to long intergenic noncoding RNA). Analysis of 3′UTRs in NGS data from a study of SLI identified a variant within a microRNA binding site associated with SLI [49]. This small region clustered with genes related to noncoding RNA could be a candidate for future investigation. *FAM19A5* is another gene in this region, which is part of the TAFA family of proteins. TAFA proteins express exclusively in all parts of the brain [50]. TAFA genes are hypothesized to play a role in more biological response mechanisms, such as immune response, regulation of immune and nervous cells and including brain injury response, such as axonal growth and guidance [50]. However, with such high expression, exclusive to the brain, within this small region surrounded by genes related to noncoding RNA, this could also be a candidate for future investigation.

Our study focused on the parametric linkage analysis and homozygosity mapping in large, consanguineous SLI families. We found candidate SLI loci that replicated across multiple PKSLI families from this sample. Some SLI families showed more than one linkage and multiple regions of homozygosity, suggesting a polygenic effect [51]. Although most of the SLI loci reported in our study differ from those previously reported, our data provided confirmatory evidence for several previous linkage loci. We provide additional support for the role of genetic factors in familial SLI and point to new loci at which causative SLI genes may reside.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. J Speech Lang Hear Res. 1997;40:1245–60.
2. Rice ML. Overlooked by public health: specific language impairment. 1st edn. Lawrence, KS: Open Access Government; 2017.
3. Rice ML, Smith SD, Gayán J. Convergent genetic linkage and associations to language, speech and reading measures in families of probands with specific language impairment. J Neurodev Disord. 2009;1:264.
4. Dunn LM, Dunn DM. PPVT-4: Peabody picture vocabulary test. Bloomington, MN: NCS Pearson, Inc.; 2007. Pearson Assessments 2007.
5. Rice ML, 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;58:345–59.
6. Bishop DV, Adams CV, Norbury CF. Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. Genes Brain Behav. 2006;5:158–69.
7. Choudhury N, Benasich AA. A family aggregation study: the influence of family history and other risk factors on language development. J Speech Lang Hear Res. 2003;46:261–72.
8. De Barbieri Z, Fernández MA, Newbury DF, Villanueva P. Family aggregation of language impairment in an isolated Chilean population from Robinson Crusoe Island. Int J Lang Commun Disord. 2018;53:643–55.
9. Rice ML, Haney KR, Wexler K. Family histories of children with SLI who show extended optional infinitives. J Speech Lang Hear Res. 1998;41:419–32.
10. Tallal P, Hirsch LS, Realpe-Bonilla T, Miller S, Brzustowicz LM, Bartlett C, et al. Familial aggregation in specific language impairment. J Speech Lang Hear Res. 2001;44:1172–82.
11. Rice ML, Zubrick SR, Taylor CL, 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.
12. Villanueva P, Nudel R, Hoischen A, Fernandez MA, Simpson NH, Gilissen C, et al. Correction: exome sequencing in an admixed isolated population indicates NFXL1 variants confer a risk for specific language impairment. PLoS Genet. 2015;11: e1005336.
13. Morton NE. Sequential tests for the detection of linkage. Am J Hum Genet. 1955;7:277–318.
14. Ott J, Wang J, Leal SM. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet. 2015;16:275.
15. Raza MH, Gertz EM, Mundorff J, Lukong J, Kuster J, Schaffer AA, et al. Linkage analysis of a large African family segregating stuttering suggests polygenic inheritance. Hum Genet. 2013;132:385–96.
16. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet. 2007;39:319–28.
17. SLI Consortium. A genomewide scan identifies two novel loci involved in specific language impairment. Am J Hum Genet. 2002;70:384–98.
18. Bartlett CW, Flax JF, Logue MW, Vieland VJ, Bassett AS, Tallal P, et al. A major susceptibility locus for specific language impairment is located on 13q21. Am J Hum Genet. 2002;71:45–55.
19. Newbury DF, Monaco AP. Genetic advances in the study of speech and language disorders. Neuron. 2010;68:309–20.
20. Bartlett CW, Flax JF, Logue MW, Smith BJ, Vieland VJ, Tallal P, et al. Examination of potential overlap in autism and language loci on chromosomes 2, 7, and 13 in two independent samples ascertained for specific language impairment. Hum Hered. 2004;57:10–20.
21. Villanueva P, Newbury DF, Jara L, De Barbieri Z, Mirza G, Palomino HM, et al. Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population. Eur J Hum Genet. 2011;19:687.
22. Nudel R, Simpson NH, Baird G, O'Hare A, Conti-Ramsden G, Bolton PF, et al. Genome-wide association analyses of child genotype effects and parent-of-origin effects in specific language impairment. Genes Brain Behav. 2014;13:418–29.
23. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. Am J Hum Genet. 2008;83:359–72.
24. Imtiaz F, Taibah K, Ramzan K, Bin-Khamis G, Kennedy S, Al-Mubarak B, et al. A comprehensive introduction to the genetic basis of non-syndromic hearing loss in the Saudi Arabian population. BMC Med Genet. 2011;12:91.
25. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. Proc Natl Acad Sci USA. 2007;104:19942–7.
26. Nalls MA, Guerreiro RJ, Simon-Sanchez J, Bras JT, Traynor BJ, Gibbs JR, et al. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. Neurogenetics. 2009;10:183–90.
27. Vine AE, McQuillin A, Bass NJ, Pereira A, Kandaswamy R, Robinson M, et al. No evidence for excess runs of homozygosity in bipolar disorder. Psychiatr Genet. 2009;19:165–70.
28. Mir A, Kaufman L, Noor A, Motazacker MM, Jamil T, Azam M, et al. Identification of mutations in TRAPPC9, which encodes the NIK-and IKK-β-binding protein, in nonsyndromic autosomal-recessive mental retardation. Am J Hum Genet. 2009;85:909–15.
29. Raza MH, Riazuddin S, Drayna D. Identification of an autosomal recessive stuttering locus on chromosome 3q13.2-3q13.33. Hum Genet. 2010;128:461–3.
30. Hamilton CM, Strader LC, Pratt JG, Maiese D, Hendershot T, Kwok RK, et al. The PhenX Toolkit: Get the Most From Your Measures. American Journal of Epidemiology. 2011;174:253–260.
31. Chaudhry IS, Rahman S. The impact of gender inequality in education on rural poverty in Pakistan: an empirical analysis. Eur J Econ Fin Admin Sci. 2009;15:174–88.
32. Silberstein M, Weissbrod O, Otten L, Tzemach A, Anisenia A, Shtark O, et al. A system for exact and approximate genetic linkage analysis of SNP data in large pedigrees. Bioinformatics. 2012;29:197–205.

33. Greenberg DA, Abreu P, Hodge SE. The power to detect linkage in complex disease by means of simple LOD-score analyses. Am J Hum Genet. 1998;63:870–9.

34. Hodge SE, Abreu PC, Greenberg DA. Magnitude of type I error when single-locus linkage analysis is maximized over models: A simulation study. Am J Hum Genet. 1997;60:217–27.

35. Seelow D, Schuelke M, Hildebrandt F, Nürnberg P. Homo-zygosityMapper—an interactive approach to homozygosity mapping. Nucleic Acids Res. 2009;37(suppl_2):W593–9.

36. Raza MH, Amjad R, Riazuddin S, Drayna D. Studies in a consanguineous family reveal a novel locus for stuttering on chromosome 16q. Hum Genet. 2012;131:311–3.

37. Riaz N, Steinberg S, Ahmad J, Pluzhnikov A, Riazuddin S, Cox NJ, et al. Genomewide significant linkage to stuttering on chromosome 12. Am J Hum Genet. 2005;76:647–51.

38. Kang C, Riazuddin S, Mundorff J, Krasnewich D, Friedman P, Mullikin JC, et al. Mutations in the lysosomal enzyme-targeting pathway and persistent stuttering. N Engl J Med. 2010; 362:677–85.

39. Raza MH, Domingues CE, Webster R, Sainz E, Paris E, Rahn R, et al. Mucolipidosis types II and III and non-syndromic stuttering are associated with different variants in the same genes. Eur J Hum Genet. 2016;24:529–34.

40. Rice ML, Taylor CL, Zubrick SR. Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. J Speech Lang Hear Res. 2008;51:394–407.

41. Wiszniewski W, Hunter JV, Hanchard NA, Willer JR, Shaw C, Tian Q, et al. TM4SF20 ancestral deletion and susceptibility to a pediatric disorder of early language delay and cerebral white matter hyperintensities. Am J Hum Genet. 2013;93:197–210.

42. Pescucci C, Meloni I, Bruttini M, Ariani F, Longo I, Mari F, et al. Chromosome 2 deletion encompassing the MAP2 gene in a patient with autism and Rett-like features. Clin Genet. 2003;64:497–501.

43. Bartsch D, Casadio A, Karl KA, Serodio P, Kandel ER. CREB1 Encodes a Nuclear Activator, a Repressor, and a Cytoplasmic Modulator that Form a Regulatory Unit Critical for Long-Term Facilitation. Cell. 1998;95:211–23.

44. 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;74:1225–38.

45. Patterson SL, Pittenger C, Morozov A, Martin KC, Scanlin H, Drake C, et al. Some forms of cAMP-mediated long-lasting potentiation are associated with release of BDNF and nuclear translocation of phospho-MAP kinase. Neuron. 2001;32:123–40.

46. Hunt MC, Rautanen A, Westin MA, Svensson LT, Alexson SE. Analysis of the mouse and human acyl-CoA thioesterase (ACOT) gene clusters shows that convergent, functional evolution results in a reduced number of human peroxisomal ACOTs. FASEB J. 2006;20:1855–64.

47. Raza MH, Mattera R, Morell R, Sainz E, Rahn R, Gutierrez J, et al. Association between rare variants in AP4E1, a component of intracellular trafficking, and persistent stuttering. Am J Hum Genet. 2015;97:715–25.

48. Guardia CM, De Pace R, Mattera R, Bonifacino JS. Neuronal functions of adaptor complexes involved in protein sorting. Curr Opin Neurobiol. 2018;51:103–10.

49. Devanna P, Chen XS, Ho J, Gajewski D, Smith SD, Gialluisi A, et al. Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. Mol Psychiatry. 2018;23:1375–84.

50. Tom Tang Y, Emtage P, Funk WD, Hu T, Arterburn M, Park EE, et al. TAFA: a novel secreted family with conserved cysteine residues and restricted expression in the brain. Genomics. 2004;83:727–34.

51. Reader RH, Covill LE, Nudel R, Newbury DF. Genome-wide studies of specific language impairment. Curr Behav Neurosci Rep. 2014;1:242–50.