Report

Genomewide Significant Linkage to Stuttering on Chromosome 12

Naveeda Riaz,[1] Stacy Steinberg,[2,*] Jamil Ahmad,[3] Anna Pluzhnikov,[2] Sheikh Riazuddin,[3] Nancy J. Cox,[2] and Dennis Drayna[1]

[1]National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Rockville, MD; [2]Department of Human Genetics, University of Chicago, Chicago; and [3]Centre of Excellence in Molecular Biology, University of the Punjab, Lahore, Pakistan

Stuttering is a common and sometimes severe communication disorder, of unknown primary etiology, that exists in populations worldwide. Many types of evidence suggest a genetic contribution to stuttering; however, the complex inheritance of this disorder has hindered identification of these factors. We have employed highly inbred families to increase the power of linkage analysis of this disorder. Forty-four Pakistani families with documented or probable consanguinity, from the city of Lahore and surrounding areas, were included. Each family contained multiple cases of stuttering, which were diagnosed using the Stuttering Severity Instrument. Using the Marshfield Weber 9 marker panel, we performed a genomewide linkage scan focused on affected individuals and their parents. The analysis included 199 genotyped individuals, 144 affected and 55 unaffected. The Pedigree Relationship Statistical Test (PREST) was used to identify pedigrees that required additional specification of inbreeding. Initial nonparametric analysis gave evidence of linkage on chromosomes 1, 5, 7, and 12. Additional genotyping was performed on chromosome 12 to a 5-cM level of resolution, and 16 additional individuals were then included, bringing the number of families to 46. Analysis of the enlarged data set provided consistent evidence of linkage on chromosome 12: the S_{homoz} scoring function gave a nonparametric LOD score of 4.61, and a LOD score of 3.51 was obtained using the S_{all} scoring function. These results suggest that a locus on chromosome 12q may contain a gene with a large effect in this sample.

Stuttering is a speech disorder characterized by the presence of syllable repetitions, syllable prolongations, and interruptions in the smooth flow of speech known as blocks. The disorder typically arises in children, who often spontaneously recover. However, in a minority of cases, it persists into adulthood, with the result that up to 1% of the population worldwide displays significant symptoms (Bloodstein 1995). Approximately half of stutterers have a family history of the disorder, and twin and adoption studies support a genetic contribution (Howie 1981; Yairi et al. 1996; Felsenfeld and Plomin 1997; Drayna et al. 1999; Shugart et al. 2004). However, Mendelian inheritance is typically not observed, supporting the view of stuttering as a complex trait.

Because of this, we sought a specialized population to perform linkage studies. We have collected these families from in the city of Lahore, Pakistan, and surrounding areas. These families are highly inbred, with stuttering occurring in two or more generations, chosen with preference for a high number of affected individuals. We made video and audio recordings of speech samples by use of the Stuttering Severity Instrument (Riley 1980), and diagnosis was performed independently by two different clinicians specializing in stuttering, according to standard criteria, with 4% dysfluent words or syllables as the minimum for diagnosis as affected. We initially ascertained a group of 100 families, and we selected 56 for collection of blood and diagnostic speech samples and subsequent genotyping.

By use of the Marshfield Weber 9 marker panel (see the Center for Medical Genetics Web site), a total of 162 affected and 62 unaffected individuals were genotyped from these 56 Pakistani families affected with stuttering. For allele-frequency estimation, 27 population-matched unaffected controls were also genotyped. Multiplex PCR was done under standard conditions and analysis was

carried out on an ABI377 using GeneScan 3.1 software, with allele sizes assigned using Genotyper v. 2.0.

Different numbers of families were informative for the various linkage analyses. In table 1, the number of families and the number of genotyped individuals—affected and unaffected—used in each analysis are shown. At most, 44 families, including 199 genotyped individuals (155 affected and 44 unaffected), were informative for the genomewide scan. A maximum of 46 families, including 220 genotyped individuals (160 affected and 60 unaffected), were informative for linkage performed using additional markers on chromosome 12. The supplementary tables (tables A1 and A2 [online only]) present information on the structure of the families used for the analyses. Here, we note that, in the maximum set of families informative for linkage, 61% (28/46) had at least one known consanguineous marriage. Known consanguineous marriages were always between cousins, with ~60% being between first cousins. The power to detect linkage in this sample was not explicitly calculated, because of the unknown effects of the reduction in genetic heterogeneity caused by the use of pedigrees from this relatively isolated and inbred population.

Mendelian inconsistencies were identified with Ped-Check (O'Connell and Weeks 1998). Genotypes creating incompatibilities were re-examined and, if possible, corrected (but not removed if they could not be corrected). Next, both the Pedigree Relationship Statistical Test (PREST) (McPeek and Sun 2000) and the graphical method described by Sun et al. (2001) were employed to detect relationship misspecification. Four instances in which individuals appeared less closely related to other family members than specified ("underrelated") and four in which individuals appeared to be more closely related to other family members than specified ("overrelated") were identified. To remedy the cases of underrelatedness, the genotypes of three problematic individuals were deleted and a pair of full sibs was recoded as half sibs. (After this recoding, a rerun of PREST resulted in no detectable relationship incompatibilities in the pedigree.) The instances of overrelatedness appeared to be the result of unspecified inbreeding. Additional inbreeding loops were then added to these pedigrees. Because we could not be confident that these inbreeding loops were correctly specified—with existing tools, it is much more difficult to detect misspecification in inbred relationships than in outbred ones—all analyses were carried out two ways: (1) with these additional inbreeding loops included and (2) with all individuals who appeared to be overrelated removed. After relationships were respecified, all genotypes creating Mendelian inconsistencies were deleted.

Marker allele frequencies were estimated using the maximum likelihood method described by Boehnke (1991), as implemented in MENDEL (Lange et al. 2001). All genotyped individuals (224 persons from the families and 27 controls for the genomewide scan; 240 individuals from the families and 27 controls for the additional markers on chromosome 12) were used for allele-frequency estimation. With the exception of the markers TPO and F13A1, marker positions were taken from the sex-averaged Marshfield map (see the Center

Table 1

Results of Genomewide Scan and Additional Genotyping on Chromosome 12

	No. of Families Used		No. of Subjects Genotyped		LOD Score at Marker[a]			
Scoring Function and Data Set	As Originally Presented	After Split	Affected	Unaffected	D1S1677	D5S408	D7S559	PAH
S_{homoz}, with overrelated individuals	44	47	144	55	1.59	1.15	1.67	3.11
S_{homoz}, without overrelated individuals	41	43	130	48	1.40	.92	2.20	3.45
S_{all}, with overrelated individuals	40	43	140	51	2.93	2.80	.09	3.11
S_{all}, without overrelated individuals	37	39	126	44	2.77	2.41	.09	3.28
S_{homoz}, with overrelated individuals, with additional genotyping	46	49	160	60	...	...	...	4.61
S_{homoz}, without overrelated individuals, with additional genotyping	43	45	143	53	...	...	...	5.20
S_{all}, with overrelated individuals, with additional genotyping	43	46	157	56	...	...	...	3.51
S_{all}, without overrelated individuals, with additional genotyping	40	42	140	49	...	...	...	3.64

NOTE.—Regions achieving LOD scores ≥2 in at least one analysis. Shown are results with the S_{homoz} or S_{all} scoring function, including or excluding the individuals "corrected" for overrelatedness, and, for chromosome 12, including additional genotyping.

[a] The scores presented are exactly at the marker listed for *D1S1677, D5S408,* and *D7S559.* In the case of *PAH,* the score shown is exactly at the marker when S_{homoz} is used; when S_{all} is used, however, the score shown is the largest in the region, which was found between *D12S1041* (104.12 cM) and *PAH* (109.47 cM) in the genomewide scan and between *PAH* and *D12S78* (111.87 cM) when extra markers were added to this region.

for Medical Genetics Web site). TPO and F13A1 were located with PYGMALION (Pluzhnikov et al. 2003), making use of information on their physical positions. Only genotyped affected individuals and the persons necessary to specify relationships among these individuals were included in the pedigrees used for linkage analysis. Three pedigrees had to be split (into two families each), and a distant inbreeding loop had to be removed from one pedigree to make linkage analysis computationally possible.

Multipoint allele-sharing methods (Kong and Cox 1997), as implemented in Allegro (Gudbjartsson et al. 2000a), were used to assess evidence for linkage. An exponential model (Kong and Cox 1997) and two scoring functions, S_{homoz} (Gudbjartsson et al. 2000b) and S_{all} (Whittemore and Halpern 1994), were used. Allegro was modified so that, for S_{homoz} only, scores from families containing a single genotyped affected whose parents were related could be included. Nonzero scores were possible with these families (when S_{homoz} was used), because S_{homoz} contains a term for identical by descent (IBD) sharing within a single individual. (Except for this term, S_{homoz} is identical to S_{pairs}.) To take some account of the differing sizes of the pedigrees, families were weighted by the square roots of the standard deviations of their scores under the null hypothesis of no linkage.

Figure 1 and table 1 demonstrate that the inclusion of individuals with evidence of overrelatedness—"corrected" for overrelatedness—did not greatly change the allele-sharing logarithm of the odds (LOD) scores. The use of different scoring functions, on the other hand, often led to dramatically altered LOD scores.

LOD scores $\geqslant 2$ were achieved at four positions: on chromosome 1 at *D1S1677* (175.62 cM), on chromosome 5 at *D5S408* (195.49 cM), on chromosome 7 at *D7S559* (181.97 cM), and on chromosome 12 close to *PAH* (109.47 cM) (table 1). On chromosomes 5 and 7, the largest scores were found at the most telomeric marker on the chromosome arms. Since LOD scores at the terminal marker of a chromosome estimate evidence for linkage less precisely than scores at markers located between two other markers—because scores at the final marker on a chromosome are multipoint on one side only—the large scores at the ends of chromosomes 5 and 7 should be interpreted with caution. In only one position, on chromosome 12 near *PAH*, were scores $\geqslant 2$ achieved in all four analyses. This region of chromosome 12 was also the location of the largest score in each of the analyses (table 1).

Because the *PAH* region on chromosome 12 appeared to be most promising, eight supplementary markers in this area were genotyped. Additional individuals were also genotyped for these markers, resulting in a maximum of 46 families with 220 individuals genotyped (160 affected) that were informative in this region. The ad-

ditional genotyping led to higher LOD scores in each of the four analyses, with LOD scores increasing modestly with S_{all}, and by at least 1.5 LOD units with S_{homoz} (table 1).

To assess the significance of the genome-scan results, we compared the results from the framework scan with the "corrected" pedigree set to those from linkage analyses of 100 simulated data sets generated by gene dropping in Merlin (Abecasis et al. 2002), using identical pedigrees, allele frequencies, marker maps, and patterns of missing data. In the observed data, LOD scores >3 were achieved with both scoring functions in the same region of chromosome 12; in the simulated data sets, no region (here defined as 40 cM) was identified that gave scores >3 with both scoring functions. The maximum LOD score observed with either scoring function was 3.11 when the "corrected" pedigree set was used; scores of this magnitude or higher (with either scoring statistic) were achieved only twice in the simulated data set.

The contribution of individual families to the chromosome 12 peak was also examined. With either scoring statistic, regardless of the inclusion of the families that showed evidence of overrelatedness, the same three families (two of which were actually parts of a large family that had to be split) gave the largest contributions at *PAH*. By use of S_{pairs} and the pedigrees, including the overrelated "corrected," the LOD scores at *PAH*, before the fine-mapping markers were added, were 1.81 and 0.85 in the two families created by splitting a single family and 1.07 in the third high-scoring family. By use of S_{all} with the same pedigrees and markers, the LOD scores at the peak were 2.38 and 0.84 in the two parts of the split family and 0.25 in the third high-scoring family. Interestingly, in these families, the same allele of *PAH* exhibited excess sharing. Because this allele is relatively uncommon in the Pakistani population (estimated frequency .08) and both families are from the same geographical area, the observations are compatible with the hypothesis that there is a common mutational origin for stuttering in these families. Although the lack of a clearly identifiable mode of inheritance and the high degree of inbreeding in these families preclude confident assessment of disease-associated haplotypes, these families will provide a powerful substrate for follow-up studies. Given the complex genetics of stuttering, it is possible that alleles at numerous loci contribute to this phenotype, even in this relatively homogeneous sample. In the only previously published genomewide linkage survey for stuttering, a locus on chromosome 18 was identified (Shugart et al. 2004). That result was obtained in a collection of outbred North American families, and much of the score was due to a single family within the data set. Our results in the Pakistani population provide no significant evidence of linkage on chromosome 18, consistent with the view that stuttering, like numerous
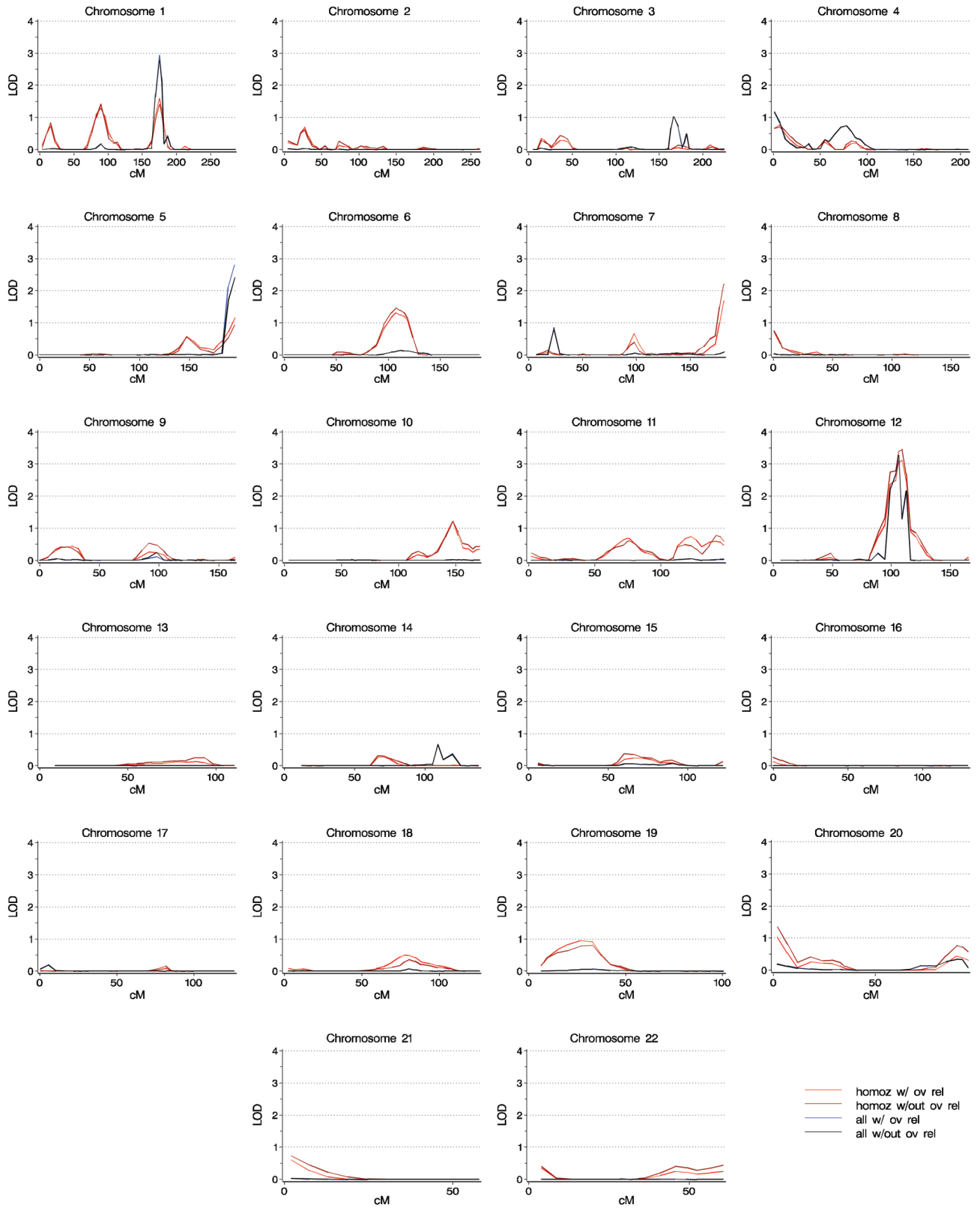
Figure 1   Genomewide linkage scan for stuttering. Shown are results with S_{homoz} ("homoz") and S_{all} ("all") scoring functions and with the "corrected" overrelated individuals included ("w/ ov rel") and excluded ("w/out ov rel"). The horizontal axis is distance, in cM, from the p-terminus.

Mendelian and non-Mendelian genetic disorders, can be caused by mutations at many different loci.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Center for Medical Genetics Web site, http://research .marshfieldclinic.org/genetics/Map_Markers/maps/Index-MapFrames.html (for sex-averaged Marshfield map)

References

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97–101

Bloodstein O (1995) A handbook of stuttering. 5th ed. Singular Press, San Diego

Boehnke M (1991) Allele frequency estimation from data on relatives. Am J Hum Genet 48:22–25

Drayna D, Kilshaw J, Kelly J (1999) The sex ratio in familial persistent stuttering. Am J Hum Genet 65:1473–1475

Felsenfeld S, Plomin R (1997) Epidemiological and offspring analyses of developmental speech disorders using data from the Colorado Adoption Project. J Speech Lang Hear Res 40: 778–791

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000a) Allegro, a new computer program for multipoint linkage analysis. Nat Genet 25:12–13

Gudbjartsson DF, Jonasson K, Frigge ML, Kong A (2000b) Fast multipoint linkage analysis and the program Allegro. Technical report, deCODE Genetics, Reykjavik

Howie PM (1981) Concordance for stuttering in monozygotic and dizygotic twin pairs. J Speech Hear Res 24:317–321

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Lange K, Cantor R, Horvath S, Perola M, Sabatti C, Sinsheimer J, Sobel E (2001) MENDEL version 4.0: a complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. Am J Hum Genetics Suppl 69:A1886

McPeek MS, Sun L (2000) Statistical tests for detection of misspecified relationships by use of genome-screen data. Am J Hum Genet 66:1076–1094

O'Connell JR, Weeks DE (1998) PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet 63:259–266

Pluzhnikov A, Wen W, Luban S, Chen Y, Cox NJ (2003) Efficient construction of genetic maps based on publicly available physical and genetic map data. Am J Hum Genet Suppl 73:A438

Riley GD (1980) SSI-3, stuttering severity instrument for children and adults, 3rd ed. Western Psychological Services, Los Angeles

Shugart YY, Mundorff J, Kilshaw J, Doheny K, Doan B, Wanyee J, Green ED, Drayna D (2004) Results of a genome-wide linkage scan for stuttering. Am J Med Genet 124:133–135

Sun L, Abney M, McPeek MS (2001) Detection of misspecified relationships in inbred and outbred pedigrees. Genet Epidemiol 21:S36–S41

Whittemore AS, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127

Yairi E, Ambrose N, Cox N (1996) Genetics of stuttering: a critical review. J Speech Hear Res 39:771–784