Genome-wide analysis of genetic susceptibility to language impairment in an isolated Chilean population

Pia Villanueva¹,²,³,⁵, Dianne F Newbury*⁴,⁵, Lilian Jara¹, Zulema De Barbieri², Ghazala Mirza⁴, Hernán M Palomino², Maria Angelica Fernández², Jean-Baptiste Cazier⁴, Anthony P Monaco⁴ and Hernán Palomino¹

Specific language impairment (SLI) is an unexpected deficit in the acquisition of language skills and affects between 5 and 8% of pre-school children. Despite its prevalence and high heritability, our understanding of the aetiology of this disorder is only emerging. In this paper, we apply genome-wide techniques to investigate an isolated Chilean population who exhibit an increased frequency of SLI. Loss of heterozygosity (LOH) mapping and parametric and non-parametric linkage analyses indicate that complex genetic factors are likely to underlie susceptibility to SLI in this population. Across all analyses performed, the most consistently implicated locus was on chromosome 7q. This locus achieved highly significant linkage under all three non-parametric models (max NPL = 6.73, P = 4.0 × 10⁻¹¹). In addition, it yielded a HLOD of 1.24 in the recessive parametric linkage analyses and contained a segment that was homozygous in two affected individuals. Further, investigation of this region identified a two-SNP haplotype that occurs at an increased frequency in language-impaired individuals (P = 0.008). We hypothesise that the linkage regions identified here, in particular that on chromosome 7, may contain variants that underlie the high prevalence of SLI observed in this isolated population and may be of relevance to other populations affected by language impairments.

European Journal of Human Genetics (2011) 19, 687–695; doi:10.1038/ejhg.2010.251; published online 19 January 2011

Keywords: Specific language impairment (SLI); Robinson Crusoe Island; linkage; language

INTRODUCTION

Specific language impairment (SLI) is a profound deficit in the acquisition of language despite adequate intelligence and opportunity, in the absence of any possible medical aetiology.¹ This disorder is a common developmental condition affecting between 5% and 8% of pre-school children, and thus places a heavy burden upon health-related and educational services.² It is well documented that SLI has a strong genetic basis (reviewed by Stromswold³). However, it is proposed that susceptibility to this disorder is complex in nature involving multiple genes, in combination with environmental factors.⁴ The genetic basis of complex disorders are notoriously difficult to characterise, as the contributing factors can vary greatly between affected individuals and may be masked by undetermined environmental effects. This is reflected in the fact that, to date, only four genetic loci⁵–⁷ and three associated candidate genes⁸,⁹ have been described for SLI (OMIM no. 606711 (SLI1), OMIM no. 606712 (SLI2), OMIM no. 607134 (SLI3), OMIM no. 612514 (SLI4), OMIM no. 612514 (CNTNAP2, SLI4)), OMIM no. 613082 (ATP2C2, in SLI1) and OMIM no. 610112 (CMIP in SLI1)).

Isolated founder populations can provide an important resource in the identification of causal genes underlying complex disorders.¹⁰ Such populations are derived from a small number of relatively recent ancestors and thus are relatively homogeneous, a point which can greatly assist gene mapping processes.¹¹ Furthermore, one may postulate that loci identified in founder populations may hold more relevance to the general population than those yielded by the study of rare monogenic forms of impairment. In 2008, Villanueva et al.¹² described a Chilean founder population with an increased incidence of SLI (known as TEL in Spanish-speaking countries). This population inhabit the Robinson Crusoe Island, which forms part of the Juan Fernandez Archipelago, 677 km to the west of Chile, South America. Robinson Crusoe Island is the only inhabited island in the archipelago and has 633 residents. The most recent colonisation dates to the late nineteenth century when the island was repopulated by a group of eight families. A total of 77% of the current population has at least one of the colonising surnames supporting a high degree of consanguinity. Linguistic profiling of the colonising children indicated that 35% met current criteria for SLI (expressive or comprehensive language > 2SD below that expected for their age), 27.5% had language deficits associated to other pathologies (eg, delayed psychomotor development, intellectual deficit or auditory impairment) and 37.5% displayed normal language skills.¹² In contrast, the frequency of SLI

¹Human Genetics Division, Faculty of Medicine, University of Chile, Santiago, Chile; ²School of Speech and Hearing Services, Faculty of Medicine, University of Chile, Santiago, Chile; ³Department of Paediatric Dentistry and Dentomaxillary Orthopaedics, Faculty of Dentistry, University of Chile, Santiago, Chile; ⁴Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
²Correspondence: Professor P Villanueva, Faculty of medicine, University of Chile, Av. Independencia 1027, Independencia, Santiago, CP 8380453, Chile. Tel: +56 2 978 6606; Fax: +56 2 978 6608; E-mail: pavihltah@gmail.com
¹¹Dr DF Newbury, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford, OX3 7BN, UK. Tel: +44 1865 287510; Fax: +44 1865 287501; E-mail: dianne@well.ox.ac.uk
¹²These authors contributed equally to this work.

Received 27 April 2010; revised 17 November 2010; accepted 3 December 2010; published online 19 January 2011
in the non-colonising children (3.8%), coincided with that reported for mainland Chile (~4%).

Genealogical reconstruction indicated that 75% of known affected individuals were descended from a single pair of founder brothers. This population therefore represents a rare resource, which may be valuable in the identification of genetic loci contributing to susceptibility to SLI.

In this study, we perform genome-wide loss of heterozygosity mapping and parametric and non-parametric linkage analysis of the Robinson Crusoe population. We identify five regions (on chromosomes 6, 7, 12, 13 and 17) that meet genome-wide significance, and several loci, which are consistently implicated across alternative analyses. We hypothesise that these regions may contain variants underlying the high prevalence of SLI observed in this isolated population.

**SUBJECTS AND METHODS**

This work was approved by the ethics department of the University of Chile. Informed consent was given by all participants and/or, where applicable, their parents.

DNA was extracted from EDTA whole blood samples collected from all available SLI and language-normal probands and their immediate families (125 individuals from 34 families, Table 1) using a standard chloroform extraction protocol.

All Island inhabitants between 3 and 8 years, 11 months of age (*n* = 66) were subjected to a linguistic battery, which included tests of phonology (Test para Evaluar Procesos de Simplificación Fonológica (TEPROSIF)) and expressive and receptive morphosyntax (Toronto Spanish Grammar Exploratory test).

Table 1 Sample structure

| N   | SLI (%) | Language normal (%) |
|-----|---------|---------------------|
| Probands | 34 | 12 (35) | 22 (65) |
| Sibs | 22 | 5 (23) | 17 (77) |
| Half-sibs | 6 | 4 (67) | 2 (33) |
| Parents | 61 | 21 (34) | 40 (66) |
| **Total** | **123** | **42 (34)** | **81 (66)** |

Male: 55 17 (39) 38 (47)

Female: 68 25 (61) 43 (53)

Abbreviation: SLI, specific language impairment.

A total of 123 samples were analysed. These included 42 language impaired individuals and 81 language normal individuals.

All genotypes were called within Beadstudio (Version 3, Illumina Inc., San Diego, CA, USA). Any SNP with a genotype score below 0.9 was manually inspected and, if necessary, the clusters adjusted. A total of 18 samples were duplicated across arrays. Any SNP with a genotype score below 0.5 (n = 27), a call rate below 0.97 (n = 4) or a minor allele frequency below 2.5% (n = 2) was excluded from further analyses.

Allele-sharing between individuals was examined using the Graphical Representation of Relationships (GRR). This software calculates mean Identity by State (IBS) values for all possible pairs of samples and clusters individuals accordingly. Any individual found to cluster outside the expected IBS values were further examined. This error checking stage identified two DNA samples that had been mislabelled and were therefore excluded.

**Generation of linkage pedigrees**

Genealogical information was collated from birth and marriage certificates, family names and parent and relative interviews. Known relationships between
Genotype data from all affected individuals were analysed for loss of heterozygosity within PLINK.24 Sliding windows of 20-SNP genotypes were examined for runs of homozygosity. In all, 42 affected individuals from 23 nuclear families were examined including 2 affected sib-pairs, an affected trio of siblings and 3 affected half-sib-pairs. Previous studies have found that runs of homozygosity <4 Mb are common in outbred individuals.25 Segments were therefore defined as homozygous tracts if 10 homozygous SNPs were found to extend across a region greater than 4 Mb in size.

Homozygosity mapping
Genotype data of linkage within MERLIN (autosomes) and the MERLIN extension, MINX (X chromosome).21 As linkage packages were unable to analyse genome-wide data for the 242-bit pedigree as a whole, it was broken into sub-pedigrees. This segmentation was manually performed on the basis of closest shared ancestor. Seven extended families of 20–24 bits (where a bit is defined as 2\(^x\) the number of non-founders—the number of founders) were defined and included 41 affected individuals and 63 of the 123 genotyped individuals (\(n=123\)). These individuals are derived directly from the population under study and can therefore be expected to provide representative expected allele frequencies. Nonetheless, these data are derived from related individuals and can therefore lead to a bias. We therefore repeated the analyses using allele frequency data from genotyped founder individuals of the generated sub-pedigrees (ie, those who marry into the pedigree, \(n=9\)). Although this reduces the dependence between individuals, it relies upon a small number of data points. We therefore also performed linkage analysis using allele frequency data from 60 unrelated CEPH individuals. The Y chromosome SNP data of the Robinson Crusoe population indicated that the founder males were European in origin (data not shown). Non-parametric results are reported as NPL scores and threshold levels for genome-wide significance are in line with that suggested by Kruglyak and Lander.26 Namely, NPL scores of > 3.8 (\(P=7.4 \times 10^{-4}\)) are described as suggestive linkage, NPL scores > 4.08 (\(P=2.2 \times 10^{-5}\)) as significant and NPL scores > 4.99 (\(P=3.0 \times 10^{-7}\)) as highly significant. Using a Bonferroni multiple testing correction for the three non-parametric analyses run, these thresholds equate to \(P=2.46 \times 10^{-4}\), \(P=7.3 \times 10^{-6}\) and \(P=1.0 \times 10^{-7}\), respectively. In this instance, we expect the Bonferroni correction to be over-conservative because of the high-expected correlation between the three analyses.

Haplotype analyses
Haplotypes were reconstructed for the chromosome 7 region of linkage within nuclear 2-generation families using MERLIN.23 Two-SNP sliding windows were visually inspected for allele combinations that co-segregated with affection status. All haplotypes that were found to have odds ratios of > 2.0 or < 0.5 (\(n=5\)) were analysed for association within PLINK using all genotyped cases and controls under a linear model.24 In these analyses, no correction was made for the relationships between cases and controls. Association analyses of simulated data-sets yielded a distribution of empirical \(P\)-values that fit well with those expected under the theoretical model indicating that, in this particular case, the relationships between individuals do not inflate the significance of the results obtained (data not shown). Measures of linkage disequilibrium (LD) were calculated within haploview.22

RESULTS
Pedigree reconstructions confirmed that of the 44 affected individuals from whom we had DNA, 37 (84%) were descendants of a pair of
founder brothers and 4 (9%) had unknown ancestry. Following quality control, genotypes were available for 6009 SNPs (5666 autosomal) with an average spacing of one SNP every 490 kb. The average genotype call rate was 99.9%.

### Table 2 Homozygous segments shared between more than two affected individuals

| Chromosome | Start | End  | Size    | Number of SNPs | Number of Inds | Homozygous individuals |
|------------|-------|------|---------|----------------|----------------|------------------------|
| 2          | 169542195 | 173937368 | 4395173 | 9              | 2              | relationship unknown    |
| 4          | 73731890   | 78761621  | 5029731 | 11             | 2              | relationship unknown    |
| 6          | 71779542   | 77471874  | 5692332 | 15             | 3              | unrelated               |
| 6          | 77572235   | 77572235  | 0       | 1              | 2              | unrelated               |
| 6          | 87364428   | 87532681  | 168253  | 3              | 2              | unrelated               |
| 6          | 88115604   | 92044752  | 3929148 | 17             | 3              | unrelated               |
| 6          | 92098625   | 93639259  | 1540634 | 9              | 2              | unrelated               |
| 7          | 108674847  | 114462759 | 5787912 | 13             | 2              | relationship unknown    |
| 8          | 18755221   | 19196467  | 441246  | 3              | 2              | relationship unknown    |
| 8          | 19559214   | 23746576  | 4187362 | 16             | 3              | relationship unknown    |
| 9          | 83685047   | 93408941  | 9723894 | 20             | 2              | relationship unknown    |
| 10         | 3791436    | 9111974   | 5320561 | 22             | 2              | relationship unknown    |
| 11         | 548867814  | 548867814 | 0       | 1              | 3              | relationship unknown    |
| 11         | 55360988   | 59674738  | 4313750 | 12             | 4              | relationship unknown    |
| 11         | 59597022   | 60616462  | 659440  | 2              | 2              | relationship unknown    |
| 12         | 41575238   | 46437276  | 4856038 | 17             | 3              | relationship unknown    |
| 14         | 20992444   | 24950428  | 4051384 | 14             | 2              | relationship unknown    |
| 14         | 94718410   | 98165306  | 344626  | 8              | 2              | relationship unknown    |
| 14         | 98298832   | 99813174  | 1514342 | 8              | 3              | relationship unknown    |
| 14         | 100345436  | 101474494 | 1129058 | 4              | 2              | relationship unknown    |
| 15         | 36837208   | 36837208  | 0       | 1              | 2              | relationship unknown    |
| 15         | 37016395   | 37119086  | 102691  | 5              | 3              | relationship unknown    |
| 15         | 37318605   | 43474548  | 6155943 | 16             | 4              | relationship unknown    |
| 15         | 81012306   | 88150562  | 7138256 | 20             | 2              | relationship unknown    |
| 15         | 15723647   | 19953169  | 4229522 | 18             | 2              | relationship unknown    |
| 15         | 50678730   | 57149118  | 6470388 | 19             | 2              | relationship unknown    |
| 16         | 52260700   | 58384823  | 6124123 | 18             | 2              | relationship unknown    |
| 21         | 32754546   | 35223308  | 2468762 | 5              | 2              | unrelated               |
| 21         | 35233892   | 38156688  | 2922796 | 18             | 3              | unrelated               |
| 21         | 38599459   | 39524326  | 924867  | 3              | 2              | unrelated               |

Abbreviation: SNPs, single nucleotide polymorphisms.
Start and end positions give positions of the extremities of overlapping segments between all individuals (in bp, B36). Boxed segments are contiguous.
region identified five 2-SNP combinations that were present in at least 90% of affected individuals. Further investigations in all genotyped individuals, indicated that one of these haplotypes (rs727714/rs969356, AG) occurred at a significantly lower frequency in unaffected individuals than affected (Supplementary Table 1). The AG genotype of the rs727714/rs969356 haplotype was present in 98% of cases and 76% of controls and had an allele frequency of 67% in cases.

The AG haplotype occurred at a significantly lower frequency in unaffected individuals, indicated that one of these haplotypes (rs727714/rs969356, AG) occurred at a significantly lower frequency in unaffected individuals than affected (Supplementary Table 1). The AG genotype of the rs727714/rs969356 haplotype was present in 98% of cases and 76% of controls and had an allele frequency of 67% in cases.

Affected individuals do not present with a specific core phenotype as may be predicted under a monogenic model, but instead show extensive heterogeneity in the severity and nature of impairment between affected individuals, as is typical of complex genetic forms of SLI.

The most consistent region of linkage extended across 48 Mb of chromosome 7q (chromosome position 111 285 062–158 710 965). This region reached a maximum NPL score of 6.73 (P=4.0 × 10⁻¹¹) and achieved genome-wide significance in all three non-parametric analyses performed and overlapped with a peak of parametric linkage (recessive model max HLOD=1.24) and two segments of homozygosity. Although these are not independent observations and a number of alternative analyses were performed, the reliability of the linkage in this region is consistent with that expected from a true positive.

Segregation analyses identified a two-SNP haplotype that was found at a marginally increased frequency in cases than controls (P=0.008). This haplotype fell across the NOBOX (OMIM no. 610934) and TPK1 (OMIM no. 606370) genes. NOBOX is a homeobox gene, which is preferentially expressed in oocytes, but not reported to be expressed in brain. TPK1 encodes the thiamine pyrophosphokinase 1 enzyme, which catalyses the conversion of thiamine to thiamine pyrophosphate. Thiamine (or vitamin B1) is essential for the metabolism of carbohydrates into glucose and acts as a co-enzyme in the production of acetylcholine. Thiamine deficiency forms part of numerous disorders including ataxia, confusion and impaired memory. Interestingly, a recent study suggested a link between thiamine deficiency and syntactic and lexical disorder. The chromosome 7 peak also overlaps with the AUTS1 locus of linkage to autism and includes both the FOXP2 and CNTNAP2 genes, both of which have previously been associated with language disorders. The genotyping panels utilised in this study were optimised for linkage investigations and thus involve a relatively sparse map of SNPs (~1SNP every 500 kb). The fine mapping of these regions is therefore required to enable the identification of candidates in an unbiased manner. We found that the two-SNP haplotype on chromosome 7 showed moderate long-range

![Figure 3](image)

**Figure 3** Genome-wide linkage analyses. Traces are shown for parametric analyses using both dominant and recessive models with full penetrance and three non-parametric models utilising expected allele frequencies derived from CEPH population, from genotyped founders in the sub-pedigrees and from all genotyped individuals. Traces are also shown for identified stretches of homozygosity (where the X-axis represents the number of individuals found to be homozygous across the region).
### Table 3 Chromosome regions highlighted by linkage analysis

| Chr | Max rec (chr posn) | SNP (chr posn) of max HLOD | CEPH max HLOD (P-value) | SNP (chr posn) of max NPL score | Founder max NPL score (P-value) | All max found NPL score (P-value) | No. Homo Runs |
|-----|--------------------|-----------------------------|-------------------------|-------------------------------|-------------------------------|---------------------------------|--------------|
| 1   | 1.23               | rs2129975 (92054668)        |                         | 1.26 (0.10)                    | rs792321–rs481387             |                                 | 1            |
| 1   | 1.52               | rs1906255 (186438670)       |                         | 1.05 (0.15)                    | rs2093759                     |                                 | 1            |
| 2   |                    | rs1098966 (12956420)        |                         | 3.90 (5.0 × 10⁻⁵)             | rs3102960 (8369948)           |                                 | 1            |
| 4   |                    | rs809000 (167059469)        |                         | 7.66 (3.2 × 10⁻¹⁴)            | rs724750–rs675162             |                                 | 1            |
| 5   |                    | rs1476640–rs768055 (141058779–141095920) |                 | 6.73 (4.0 × 10⁻¹¹)           | rs1524341–rs1024676          |                                 | 2            |
| 7   |                    | rs390950 (11633238)         |                         | 2.40 rs749540 (41600253)      | 3.95 (4.0 × 10⁻⁵)            |                                 | 3            |
| 8   |                    | rs153277 (116148616)        |                         | 3.69 rs717081 (116179966)     | 3.04 rs1868280–rs1375062     |                                 | 1            |
| 9   |                    | rs3345 (131346779)          |                         | 3.72 (0.0001)                 | rs1868280–rs1375062          |                                 | 2            |
| 11  |                    | rs1495906 (81238725)        |                         | 1.27 rs2044727 (13164201)     | 1.44 rs3345 131346779        |                                 | 1            |
| 12  |                    | rs975388–rs7960480 (132120315–1322288239) |             | 1.13 rs975388–rs7960480     | 1.44 rs3345 131346779        |                                 | 2            |
| 13  |                    | rs1572372–rs3847993 (19738004–20193194) |            | 4.78 rs80285 (37937932)       | 2.66 rs80285 (37937932)       |                                 | 3            |

Genome analysis of SLI in a Chilean isolate

P Villanueva et al

European Journal of Human Genetics
We did not observe any linkage to chromosomes 16 or 19, which have previously been implicated in SLI.\(^5,6,33\) Again, this may be caused by the low density of markers investigated in the present study. Alternatively, as the loci on chromosome 16 and 19 were identified by a quantitative genome screen of language-related measures, this may reflect differences in study design. As the Chilean quantitative linguistic data was collected only for subjects within a restricted age range (3 and 9 years), the current study utilised a binary affection status. This is similar to the approach applied by Bartlett et al. (2002, 2004) in their genome screen for SLI in which they identified a region of linkage on chromosome 13 (SLI3), which overlaps with that found by the present study. This region has also been linked to autism,\(^34\) a result which was strengthened by the selection of families on the basis of linguistic data.\(^5,6\) Our chromosome 13 linkage consisted of two adjacent peaks. The distal peak (34–48 Mb) overlapped with a segment of homozygosity and achieved a maximum NPL score of 4.8 \((P = 8.0 \times 10^{-7})\) using CEPH allele frequencies. The proximal peak (83–94 Mb) reached an NPL of 3.5 \((P = 0.0002)\) under all non-parametric analyses performed and coincided with an area of marginal linkage under a recessive parametric model.

In addition to the linkages on chromosome 7 and 13, we also observed significant linkage (NPL > 4.08 \((P < 2.2 \times 10^{-5})\)) to chromosome 17 and highly significant linkages (NPL > 4.99 \((P < 3.0 \times 10^{-7})\)) to chromosomes 6q and 12 (Figure 3, Table 3). However, these peaks were only observed under a single non-parametric model and not in models using alternative expected allele frequencies. It is therefore likely that these divergent results may be driven by differences in the allele frequencies of the control populations used and illustrate the importance of correctly estimating allele frequencies, especially for markers that are in linkage disequilibrium.\(^36\) Indeed, we found that the correlation of expected allele frequencies between the three different control groups was moderate (0.41–0.70 across all SNPs) and was lower than average across the conflicting regions of linkage on chromosomes 6 and 12 (as low as 0.29 and 0.09, respectively), but remained moderate across the region of linkage on chromosome 7 (0.48–0.67). Importantly, simulation studies indicate that although allele frequency misspecification can lead to false positives, this artefact is not expected to affect the power to detect true linkages.\(^37\) Thus, although the loci on chromosome 6 and 12 reached a threshold of highly significant linkage, as these were observed with only one non-parametric analysis, we must recognise the possibility that they represent false positives, especially given the high number of tests performed. Instead, a more fruitful avenue of investigation may be provided by the examination of regions found to be consistently implicated across all three analyses performed, even in cases where this linkage did not reach genome-wide significance (eg, chromosome 2, 6p, 8, 9, 15 and 17, Table 3, Supplementary Figure 1).

In conclusion, this study has applied a genome-wide approach to identify loci which may contain genes underlying susceptibility to SLI in an isolated population. This study represents the first step in the detection of genetic variants that underlie the increased frequency of language impairments in this population. It is envisaged that the fine mapping of the identified loci will allow the detection of associated polymorphisms. It is likely that the variants identified by the further study of this population will have a significant role in furthering our understanding of the genetic basis of language impairments and language development.

### Table 3 (Continued)

| No. | SNP                  | Rec. NPL score | NPL score | P-value |
|-----|----------------------|----------------|-----------|---------|
| 14  | rs961700-rs1015023   | 1.05           | 0.009     |         |
| 15  | rs2596156            | 1.42           | 0.003     |         |
| 16  | rs1540297            | 1.31           | 0.003     |         |

Abbreviations: Chr, chromosome; posn, position; max, maximum; SNP, single nucleotide polymorphism.
Figure 4 Chromosome 7. Chromosome 7 represented the most consistently linked locus across analyses. Traces are shown for parametric analyses using both dominant and recessive models with full penetrance, three non-parametric models utilising expected allele frequencies derived from CEPH population, from genotyped founders in the sub-pedigrees and from all genotyped individuals. Traces are also shown for identified stretches of homozygosity (where the y axis represents the number of individuals found to be homozygous across the region) and association of P-values (relative to the secondary y axis).

Table 4 SNPs that are in LD with associated haplotype

| SNP          | Chr posn | D’ nuc | D’ Confidence Intervals | LOD nuc | Fams | R² nuc | Fams | D’ | D’ Confidence Intervals | LOD | R² | Fams | Fams5 | Genes | Peak of Linkage? |
|--------------|----------|--------|--------------------------|---------|------|--------|------|----|--------------------------|------|----|------|-------|-------|-----------------|
| rs2040587    | 114 462 759 | 0.23   | 0.04–0.44                | 0.66    | 0.04 | 0.50   | 0.27–0.66 | 2.99 | 0.14 | 300 kb downstream of FOXP2 |
| rs1962522    | 134 436 889 | 0.48   | 0.25–0.65                | 2.77    | 0.17 | 0.39   | 0.20–0.54 | 2.69 | 0.14 | AGBL3             |
| rs10488598   | 136 238 383 | 0.56   | 0.28–0.76                | 2.58    | 0.11 | 0.53   | 0.21–0.73 | 1.74 | 0.10 | CHRM2             |
| rs1371463    | 137 933 620 | 0.46   | 0.26–0.62                | 3.32    | 0.15 | 0.31   | 0.10–0.49 | 1.39 | 0.07 | SVOPL             |
| rs1467498    | 138 982 540 | 0.64   | 0.36–0.81                | 3.25    | 0.15 | 0.62   | 0.39–0.77 | 4.26 | 0.20 | HIPK2             |
| rs1476640    | 141 058 779 | 0.10   | 0.00–0.31                | 0.17    | 0.01 | 0.27   | 0.05–0.48 | 0.74 | 0.04 | E2EE3             |
| rs768055     | 141 059 520 | 0.80   | 0.47–0.99                | 3.61    | 0.17 | 0.76   | 0.34–0.91 | 2.04 | 0.12 | Peak of parametric linkage |
| rs727714     | 143 729 925 | 1.00   | 0.88–1.00                | 18.1    | 0.67 | 1.00   | 0.92–1.00 | 23.1 | 0.70 | NOBOX exon 3     |
| rs969356     | 143 804 256 | 1.00   | 0.89–1.00                | 19.8    | 0.71 | 1.00   | 0.91–1.00 | 21.5 | 0.65 | Associated haplotype SNP |
| rs802200     | 145 736 404 | 0.42   | 0.23–0.58                | 3.00    | 0.17 | 0.39   | 0.18–0.55 | 2.21 | 0.13 | Associated haplotype SNP |
| rs1524341    | 146 337 622 | 0.23   | 0.05–0.41                | 0.87    | 0.05 | 0.21   | 0.04–0.37 | 0.78 | 0.04 | Associated haplotype SNP |
| rs1024676    | 146 346 794 | 0.22   | 0.03–0.43                | 0.52    | 0.03 | 0.23   | 0.03–0.45 | 0.51 | 0.03 | Peak of non-parametric linkage |
| rs4431523    | 147 228 099 | 0.45   | 0.23–0.63                | 2.66    | 0.12 | 0.38   | 0.15–0.57 | 1.7  | 0.07 | Peak of non-parametric linkage |

Abbreviations: Chr, chromosome; posn, position; SNP, single nucleotide polymorphism.
Any SNP that has a D’>0.4 and a pairwise LOD>2.0 with the associated haplotype is shown. Measures of LD were evaluated both in nuclear families and in the extended pedigree as shown in Figure 1.
The associated haplotype was formed from SNPs rs727714 and rs969356. These two SNPs gave the maximum NPL score of the non-parametric linkage analyses using allele frequencies from all genotyped individuals. The peak of linkage in the non-parametric analyses using allele frequencies from founder and CEPH individuals fell across SNPs rs1524341 whereas the peak of parametric linkage fell at SNPs rs1476640 and rs768085.
All SNPs are intronic unless otherwise stated.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGEMENTS
The authors are extremely grateful to the inhabitants of Robinson Crusoe Island who have agreed to participate in this study. We would also like to thank Mr Leopoldo Gonzalez the mayor of the Ilustre Municipalidad de Juan Fernandez for his infinite assistance and patience in the development of this research. Also to the authorities of schools of medicine and dentistry for giving us the necessary permits to travel to the island of Juan Fernandez. We would also like to thank Simon Fisher for his comments on the original manuscript. This work was specifically funded by a John Fell Fund award from Oxford University (http://www.ox.ac.uk) and was supported by core genotyping and statistical support facilities funded by a Wellcome Trust (http://www.wellcome.ac.uk) core award grant (grant no. 075491). The collection of DNA samples and characterisation of the Robinson Crusoe population was funded by Vicerrectoria de Investigación, Universidad de Chile (www.uchile.cl), UCHILE DID TNAC 01-02/01, UCHILE DI MULT 05-05/02 grants.

1 Law J, Boyle J, Harris F, Harkness A, Nye C: Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. Int J Lang Commun Disord 2000; 35: 165–188.
2 Harel S, Greenstein Y, Kramer U et al: Clinical characteristics of children referred to a child development center for evaluation of speech, language, and communication disorders. Pediatr Neurol 1996; 15: 305–311.
3 Stromswold K: The heritability of language: a review and meta analysis of twin adoption and linkage studies. *Language 2001; 77: 647–723.

4 Bishop DV: Genetic and environmental risks for specific language impairment in children. *Philos Trans R Soc Lond B Biol Sci 2001; 356: 369–380.

5 SLIC: A genomewide scan identifies two novel loci involved in specific language impairment. *Am J Hum Genet 2002; 70: 384–398.

6 SLIC: 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.

7 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

8 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

9 Vernes SC, Newbury DF, Abrahams BS et al: A functional genetic link between distinct developmental language disorders. *N Engl J Med 2008; 359: 2337–2345.

10 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

11 Stromswold K: The heritability of language: a review and meta analysis of twin adoption and linkage studies. *Language 2001; 77: 647–723.

12 Bishop DV: Genetic and environmental risks for specific language impairment in children. *Philos Trans R Soc Lond B Biol Sci 2001; 356: 369–380.

13 SLIC: A genomewide scan identifies two novel loci involved in specific language impairment. *Am J Hum Genet 2002; 70: 384–398.

14 SLIC: 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.

15 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

16 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

17 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

18 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

19 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

20 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

21 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

22 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

23 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

24 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

25 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

26 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

27 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

28 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

29 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

30 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

31 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

32 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

33 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

34 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

35 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

36 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

37 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

38 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

39 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

40 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

41 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

42 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

43 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

44 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

45 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

46 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

47 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.

48 Newbury DF, Winchester L, Addis LS et al: CMIP and ATP2C2 modulate phonological short-term memory in language impairment. *Am J Hum Genet 2009; 85: 264–272.

49 Bartlett CW, Flax JF, Logue MW et al: A major susceptibility locus for specific language impairment is located on 13q21. *Am J Hum Genet 2002; 71: 45–55.