Exome sequencing of an isolated Chilean population affected by Specific Language Impairment (SLI).

D. F. Newbury [1], P. Villanueva [2,3,4], A. Hoischen [5], R. Nudel [1], C. Gilissen [5], L. Carvajal-Carmona [1,6], M. M. Echeverry [6], L. Jara [2], Z. De Barbieri [3], M. A. Fernández [3], J. Veltman [5], A. P. Monaco [1], H. Palomino [4], S. E. Fisher [1,7,8]

1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, United Kingdom.
2) Human Genetics Program, Institute of Biomedical Sciences (ICBM), Faculty of Medicine, University of Chile, Santiago, Chile.
3) School of Speech and Hearing Therapy, Faculty of Medicine, University of Chile, Santiago, Chile.
4) Department of Child and Dental Maxillary Orthopedics, Faculty of Dentistry, University of Chile, Santiago, Chile.
5) Dept Human Genetics, Nijmegen Centre for Molecular Life Sciences and Institute for Genetic and Metabolic Disorders, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.
6) Department of Biology, University of Tolima, Tolima, Colombia
7) Department of Language and Genetics, Max Planck Institute for Psycholinguistics, Nijmegen,. The Netherlands.
8) Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands.

SPECIFIC LANGUAGE IMPAIRMENT (SLI)

Speech and language impairments that are a primary deficit and have no obvious cause (e.g. a comorbid neurological disorder such as autism) are diagnosed as Specific Language Impairment (SLI). SLI affects 5-8% of preschool children and represents a lifelong disability, associated with an increased risk of behavioural disorders, social problems and literacy deficits (Law *et al.* 2000). SLI is highly heritable but the underlying genetic mechanisms are expected to be multifactorial (Stromswold 2001).

FOUNDER POPULATION

In collaboration with the University of Chile, we have been investigating an isolated population affected by an increased incidence of SLI. These individuals live on the Robinson Crusoe Island, 677km west of mainland Chile. The current population consists of 633 individuals derived from eight founder families over 5-6 generations. Linguistic profiling indicated that 35% of the colonising children aged 3-9 met current criteria for SLI (expressive or comprehensive language >2SD below expected, n=14), 27.5% had language deficits associated to other pathologies (e.g., delayed psychomotor development, intellectual deficit or auditory impairment, n=11) and 37.5% displayed normal language skills (n=15) (Villanueva *et al.* 2008). Genealogical reconstruction revealed that over 80% of the language impaired individuals were related to a single pair of founder brothers (Figure 1).

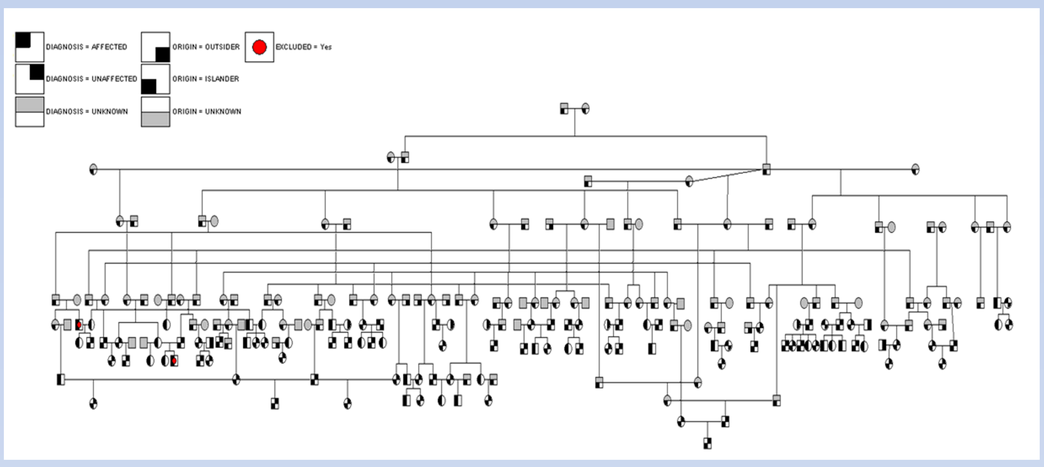

FIGURE 1 - Descendants of founder brothers.

METHODS

Given the presentation of this population, we hypothesised that the Founder brothers may have carried a mutation that directly contributes to the high incidence of SLI. We therefore performed exome sequencing in five individuals selected from separate branches of the pedigree (Figure 2).

Exome capture was performed using Agilent SureSelect v1 human exome kit (Agilent, Santa Clara, CA, USA). Sequencing was performed on the SOLiD 3 sequencer (Life Technologies, Carlsbad, CA, USA). Sequence reads were aligned to the human reference sequence (hg18). Variants were called using the DiBayes algorithm with a conservative call stringency and annotated by a custom pipeline, developed in Nijmegen.

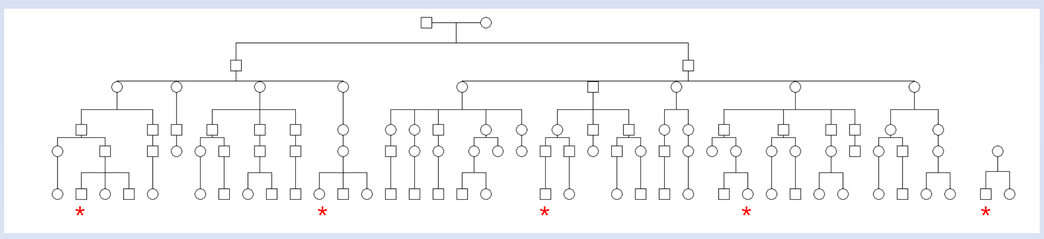

FIGURE 2 – Five individuals (marked by red stars) were sequenced.

FOR MORE INFORMATION ON OUR RESEARCH PLEASE VISIT HTTP://WWW.WELL.OX.AC.UK/NEWBURY.

RESULTS

We identified an average of 46,566 genic variants, 16,797 exonic variants, 8,096 missense variants and 120 nonsense variants per individual.

We did not identify any novel coding mutation (i.e. a change that has not previously been documented) that was present in all five affected individuals.

Nine potentially deleterious novel changes (non-synonymous coding change, splice site mutation or frame shift mutation) were found to occur in 3 of the 5 individuals sequenced.

These nine changes were sequenced in all available 123 Island-derived samples. One change was significantly over-represented in affected individuals ($P=4x10^{-5}$, Table 1). This variant introduced a non-synonymous change (Asparagine to Lysine) in the Nuclear transcription factor, X-box binding-like 1 (*NFXL1*) gene on chromosome 4.

	Allele Freq	Carrier Freq	
Unaffected	0.06	0.12	$P=4x10^{-5}$
Affected	0.23	0.46	
Islander	0.125	0.25	
Outsider	0.04	0.08	
UK controls	0.00	0.00	
Colombian controls	0.045	0.09	

TABLE 1 - Variant frequency in Islanders and control populations

The identified variant was not observed in 127 European controls, nor in 223 UK SLI probands (SLI Consortium). However, sequencing of 320 unselected Colombian controls, and later data from the 1000 Genomes project, indicated that this is a South American variant (designated rs144169475).

In the total Island population, we found the variant at a frequency of 12.5%, which is higher than that expected for a South American population (average reported allele freq across 3 South American populations – 6.3%), and that observed in Robinson Crusoe inhabitants who were not born on the Island (4%). This increase is in-line with that predicted from simulations, given Island population structure. Furthermore, the increased allele frequency was found to be driven by the language impaired Islanders (23%), and was not evident in unaffected individuals (6%).

CONCLUSION

We conclude that, although the identified *NFXL1* variant is not enough to fully explain the increased prevalence of SLI in the Robinson Crusoe population, it may contribute to the risk of language impairment within a multifactorial genetic model.

FUNDING AND ATTRIBUTIONS

The Chilean groups collected DNA, phenotypic and genealogical data from the Robinson Crusoe population.
The Netherlands groups performed the exome sequencing, variant calling and annotation pipeline.
The Oxford groups completed confirmation sequencing and downstream data analyses.
The Colombian groups provided control DNA samples for sequencing.
Exome sequencing was funded by the Max Planck Society through SE Fisher. The ongoing Robinson Crusoe project is funded by the John Fell Fund and the MRC through P Villanueva & DF Newbury.
Our thanks go to the Islanders, without whom our research would not be possible.

REFERENCES

- Law, J., J. Boyle, et al. (2000). "Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature." Int J Lang Commun Disord **35**(2): 165-88.
- Stromswold, K. (2001). "The heritability of language: A review and meta analysis of twin adoption and linkage studies." Language **77**: 647-723.
- Villanueva, P., Z. de Barbieri, et al. (2008). "[High prevalence of specific language impairment in Robinson Crusoe Island. A possible founder effect]." Rev Med Chil **136**(2): 186-92.