Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Vernes SC, Newbury DF, Abrahams BS, et al. A functional genetic link between distinct developmental language disorders. N Engl J Med 2008;359:2337-45. DOI: 10.1056/NEJMoa0802828.

SUPPLEMENTARY METHODS

Cell culture and reagents

The human SH-SY5Y cell-line, a commonly used cellular model for neuronal function, was employed for chromatin immunoprecipitation and expression analyses. This cell-line was successfully used in earlier functional analyses of FOXP2.[1] SH-SY5Y cells were stably transfected with the pcDNA3.1/FOXP2 vector (which carries the major isoform of FOXP2; isoform I) or the empty vector, as described.[2] HEK293T cells were employed to generate protein for electrophoretic mobility shift assays (EMSAs).[1] FOXP2 protein was detected using a commercially available goat polyclonal antibody (Santa Cruz Biotechnology).

Chromatin Immunoprecipitation (ChIP)

ChIP was carried out using SH-SY5Y cells stably expressing FOXP2 isoform I, as described.[2] Purified chromatin was amplified via Ligation Mediated PCR (LMPCR) according to published protocols.[3]

Shotgun cloning of ChIP DNA

ChIP DNA, amplified via LMPCR, was cloned into pGEM-T Vectors (Promega) and a library of clones was constructed. Clones were directly sequenced using T7 and SP6 primers via Big Dye chemistry on the ABI3700 automated capillary sequencer. The position of ChIP-isolated DNA sequences in the human genome was determined using the BLAT program (http://genome.ucsc.edu/) on the UCSC Genome Bioinformatics Server (May 2004 human genome assembly).[4]

Semi-quantitative PCR

DNA isolated during ChIP was amplified using a semi-quantitative PCR technique, as described previously.[2] Primer sequences are given in Supplementary Table S1.

EMSA

DNA binding was assessed using EMSAs according to published protocols.[2] Probes were designed as oligomers, 24-29 nucleotides in length (Consensus: 5'-AGCTTTATTTATGTTGTTTTGTAT-3', CNTNAP: 5'-AGCTGCTTTCAAATTTAAGCAATCAAGTG-3', CNTNAP-b: 5'-AGCTTTCCCAAATTGTCTTCATTTACATT-3'). When an unlabelled competitor probe (CNTNAP-M: 5'-AGCTGCTTTCGGGTTTAAGCAATCAAGTG-3', NFK: 5'-AGCTCCGGGGGTGATTTCACTCCCCG-3') was used to confirm specificity of DNA binding, it was added in 1-fold, 5-fold or 10-fold excess and incubated at room temperature for 15 minutes before addition of labelled probe.

Quantitative RT-PCR

RNA was extracted from SH-SY5Y cells transfected with either FOXP2 or the empty control vector as described previously.[2] For stable transfectants multiple independent passages of a single clone were used. Reverse transcription and PCR amplification was performed as described.[2] Primers were designed using PrimerBank[5] and sequences are given in Table S1. Data analysis was performed with iCycler software (BioRad), followed by quantification using the comparative CT method.[6] Fold changes are reported in response to FOXP2 expression compared to control cells transfected with empty vector, following normalisation to the *GAPDH* internal control. Statistical significance was assessed using unpaired t-tests (two-tailed).

In situ hybridisations

Isotope-based in situ hybridisation was performed as described[7] on human foetal brains obtained from the University of Maryland Brain and Tissue Bank under an approved UCLA IRB protocol. Emulsions were imaged using a Nikon Eclipse E600 Microscope with a digital capture system built around a spot cooled CCD camera. Probes employed for *FOXP2* (2210–2462 bp of NM_148900.1) and *CNTNAP2* (550–4300 bp of NM_014141.1) were previously reported.[8, 9] All hybridisations were performed on adjacent sections from at least three separate brains. Sense controls showed no signal (data not shown).

Participants

The collection and phenotypic characterisation of the SLIC cohort has been described in detail.[10, 11] As noted in the main text, the cohort includes epidemiologically and clinically selected families from four UK sites,[10-14] ascertained through a proband whose past or current language skills (expressive and/or receptive) was ≥1.5SD below the normative mean for his/her age. Diagnosis was made using the Clinical Evaluation of Language Fundamentals (CELF-R)[15] diagnostic tool. The battery is split into receptive and expressive scales, consisting of several subtests; the combination of individual tests that are used is dependent on the subject's age at testing. Additive raw scores are transformed to derive a standardized receptive (RLS) and expressive language score (ELS), each with a mean of 100 and a SD of 15 in the general population calibration sample. The majority of the SLIC probands (65%) were ascertained from a clinical setting and thus can be considered to represent a self-referred sample of children with persistent language difficulties. Many require continued learning support or specialist schooling and thus represent the more severe end of the SLI continuum. Nonetheless,

as detailed in the main text, all children in the study had a non-verbal IQ exceeding 80 and no evidence of autism, hearing loss, dyspraxia or cleft lip/palate.

Consistent with previous work,[10, 11] we employed a quantitative approach using phenotypic data collected from all siblings regardless of SLI status. This included the standardised ELS and RLS scores, but also performance on a nonsense-word repetition task (NWR), well established in prior studies as a core endophenotype of SLI.[16-18] In this task, subjects are asked to repeat tape-recorded pronounceable nonsense words of increasing length and complexity (e.g. “brufid” and “contramponist”). NWR performance is highly heritable, and children with current language impairments perform poorly on this test, as do individuals who have had language difficulties in early childhood that later resolved.[16-18] Means, standard deviations and inter-trait correlations for the language measures in the cohort used in this study are shown in Table S2.

Single nucleotide polymorphism (SNP) genotyping

To specifically test the hypothesis that our identified FOXP2 target (the *CNTNAP2* gene) might be implicated in language deficits in the SLIC cohort, we genotyped 38 SNPs across the *CNTNAP2* locus on chromosome 7q35. For practical reasons these SNPs were genotyped on a Golden Gate 1536-SNP array (Illumina),[19] which primarily consisted of unrelated SNPs from chromosome 16. These unrelated SNPs were part of an ongoing positional-cloning effort to identify the gene underlying a previously identified chromosome-16 linkage[10, 11] and are independent of the present study. To facilitate verification of genotype quality, 2 CEPH samples were included in the plates and 10 family samples were duplicated across plates. In addition, 10 SNPs that had previously been genotyped in these families were included on the array. The call mismatch rate

across all SNPs was 0.76% for the CEPH and duplicated samples, and 0.73% for duplicated SNPs. Genotypes were uploaded into the Integrated Genotyping System[20] and checked for pedigree inconsistencies and Hardy Weinberg Equilibrium (HWE) prior to analysis. SNPs or pedigrees with >1% inconsistencies or a HWE probability of below 0.001 were removed. In addition, haplotypes were created within Merlin[21] and those with a HWE P-value of <0.01 were re-examined at the call level and corrected if necessary. Any genotypes with a P<0.001 that could not be corrected were deleted. The final dataset included validated data from 38 SNPs in 847 individuals from 184 SLIC families (minimum genotyping rate of 83%; average success rate of 91%).

Quantitative association analyses

Given that the SLIC cohort comprised families that were ascertained and assessed by different teams spread across the UK, we carried out quantitative family-based association analyses using the orthogonal association model within the QTDT package.[22] This option considers only the within-family variance and is therefore robust to biases caused by population stratification. Haplotype generation was performed within Merlin[21] using the --best option and was successful for 81% of genotyped individuals. Following our identification of significant association, the possibility of a sex or imprinting effect was investigated within QTDT using nine-SNP-tag haplotypes with sex as a covariate (-cs option) and by testing for differences in the transmission of paternal and maternal alleles (-ot option).

SUPPLEMENTARY TABLES

Table S1. Oligonucleotide primers used for semi-quantitative ChIP-PCR and expression analyses (qRT-PCR)

Method	Gene	Forward primer (5'-to-3')	Reverse primer (5'-to-3')
qPCR	CNTNAP2-P1	GGAAGCAGAGACCACTCCAG	ACTCAGGCCAGTTCTCTCCA
	CNTNAP2-P2	ATGGAGAGAACTGGCCTGAG	AATAAGTCATGGCGCATTCA
	CNTNAP2-P3	TTCCTGCTTCCCAAATTGTC	AAAAGCAAGGGTGAATGGAA
qRT-PCR	CNTNAP2_A	TCCCTCCACGTCCCAAAAATG	TCTTGGCATAGCCGGGAGAA
	CNTNAP2_B	TCCCGGCTATGCCAAGATAAA	TTCCGATTGCCAAAGTCAACC
	CNTNAP2_C	TGACTTTGGCAATCGGAAGCA	CCTGTGTCGCTGTAGAGCAT
	FOXP2	CCTTCAGCGTCAGGGACTCA	CACTTCTTTCCATAACTGCTGAATCTC
	GAPDH	CAGTCCATGCCATCACTGC	TTCGTTGTCATACCAGGAAATG

Table S2A. Descriptive statistics for quantitative phenotypes in SLIC probands, and available siblings, genotyped in this study.

Group	Statistic	ELS	RLS	NWR
All	**Mean**	78.60	88.39	91.75
	Std Error	0.78	0.90	0.93
	Std Deviation	15.72	18.17	19.97
	Count	405	405	464
Probands	**Mean**	71.96	81.92	85.83
	Std Error	1.08	1.26	1.59
	Std Deviation	14.38	16.73	21.53
	Count	176	177	183
Siblings	**Mean**	83.69	93.41	95.60
	Std Error	0.98	1.17	1.07
	Std Deviation	14.79	17.68	17.90
	Count	229	228	281

Note: All traits have a mean of 100 and a standard deviation of 15 in the normal population. 81 of the 229 siblings with available data (35.4%) were also affected under the SLIC criteria of ≥1.5SD below the mean on ELS or RLS. In a small number of cases of severe SLI, probands were unable to complete all subtests of the CELF, and so their full ELS and RLS scores could not be calculated.

Table S2B. Correlations between phenotypes in the SLIC sample genotyped in this study.

	ELS	RLS	NWR
ELS	1.000		
RLS	0.788	1.000	
NWR	0.566	0.499	1.000

Table S3. Quantitative association analyses of SNPs in the CNTNAP2 gene in children with SLI.

					NWR		ELS		RLS	
Marker	**Start (bp)**	**End (bp)**	**SNP**	**MAF**	**P-value**	**Effect**	**P-value**	**Effect**	**P-value**	**Effect**
Exon 1	145251477	145251713								
rs7806058	145445440		A/G	0.655	0.998	0.98	0.886	0.42	0.257	1.73
rs6946112	145496865		C/T	0.713	0.375	1.69	0.870	1.10	0.905	1.35
rs12703803	145500557		T/G	0.708	0.434	2.01	0.813	1.08	0.779	1.69
rs2058377	145527718		A/G	0.684	0.433	2.08	0.820	0.76	0.971	1.44
rs12667234	145548824		A/G	0.697	0.296	2.46	0.680	1.55	0.870	1.61
rs2888335	145562222		T/C	0.674	0.402	2.21	0.883	1.17	0.977	1.39
rs7805539	145597926		G/A	0.714	0.500	2.22	0.734	1.27	0.818	1.46
rs4726793	145714538		A/G	0.796	0.497	2.31	0.999	1.62	0.352	2.55
rs2191295	145715642		A/G	0.795	0.343	2.31	0.751	1.62	0.241	2.54
rs10277654	145790224		T/C	0.571	0.546	1.09	0.884	1.06	0.338	2.16
Exon 2	145909011	145909121								
rs7794745	145927254		A/T	0.669	0.934	0.89	0.540	1.66	0.790	1.84
Exon 3	145974451	145974644								
rs17170287	146015977		T/G	0.755	0.475	0.55	0.294	1.40	0.686	1.66
rs6945085	146128868		T/C	0.916	0.519	2.73	0.197	3.99	0.080	7.98
rs4725699	146136016		C/T	0.952	0.164	9.08	0.665	3.56	0.569	1.05
rs1024676	146153509		C/T	0.580	0.524	0.96	0.649	0.55	0.833	1.33
rs10282158	146175715		T/A	0.953	0.095	10.24	0.578	3.52	0.502	0.99
rs7812091	146178225		T/C	0.569	0.495	0.86	0.503	0.51	0.719	1.34
Exon 4	146178647	146178794								
rs6975159	146204462		T/C	0.686	0.842	0.47	0.796	0.66	0.529	1.16
Exon 5	146242887	146243090								
Exon 6	146255719	146255903								
Exon 7	146263433	146263576								
Exon 8	146266985	146267249								
rs10500170	146285899		A/G	0.833	0.829	2.65	0.715	1.21	0.683	1.78
rs1603453	146346567		T/A	0.882	0.177	3.86	0.588	3.14	0.309	4.40
rs1603450	146351188		G/A	0.855	0.192	3.92	0.423	1.21	0.654	1.41
Exon 9	146434881	146435030								
Exon 10	146530349	146530520								
rs10251377	146555102		A/G	0.770	0.308	2.13	0.635	0.69	0.665	0.02
Exon 11	146620675	146620781								
Exon 12	146696878	146696997								
Exon 13	146773846	146774046								
rs851715	146964554		**A**/G	0.687	**0.002**	4.54	0.085	2.96	0.046	3.89
rs1177007	146984019		A/G	0.675	0.087	5.25	0.727	3.73	0.520	4.40
rs1186173	146984251		C/T	0.669	0.016	3.93	0.069	3.29	0.038	4.25
rs10246256	146992455		**T**/C	0.683	**0.001**	4.89	0.032	3.29	0.027	4.13
rs2710102	147012038		**C**/T	0.517	**0.002**	4.26	0.143	2.40	0.074	2.54
rs759178	147012760		**G**/T	0.518	**0.002**	4.36	0.169	2.50	0.065	2.62
rs1922892	147014059		**T**/C	0.518	**0.002**	4.38	0.114	2.50	0.053	2.63
rs2538991	147017267		**C**/A	0.515	**0.002**	4.60	0.116	2.47	0.062	2.79
rs17236239	147019953		A/**G**	0.654	**0.00005**	5.53	**0.008**	3.21	0.015	2.95
rs2538976	147023467		**G**/A	0.511	**0.002**	4.38	0.100	2.37	0.061	2.60
rs4431523	147034814		A/**G**	0.692	0.014	2.88	0.036	2.42	**0.003**	3.37
rs2538963	147037094		G/T	0.625	0.517	0.99	0.714	0.34	0.602	0.57
Exon 14	147038305	147038461								
rs2710117	147039420		**A**/T	0.653	**0.0004**	4.53	0.033	2.77	0.019	3.79

Table S3 (continued).

					NWR		ELS		RLS	
Marker	**Start (bp)**	**End (bp)**	**SNP**	**MAF**	**P-value**	**Effect**	**P-value**	**Effect**	**P-value**	**Effect**
Exon 15	147112602	147112729								
rs10240503	147112626		A/G	0.889	0.662	0.92	0.492	1.80	0.231	2.40
Exon 16	147252858	147253028								
Exon 17	147282231	147282449								
rs12155129	147294513		A/G	0.897	0.710	0.81	0.377	0.27	0.182	0.41
Exon 18	147306982	147307218								
Exon 19	147352028	147352264								
Exon 20	147364386	147364519								
rs11980146	147394381		A/G	0.680	0.993	0.93	0.337	1.11	0.260	0.85
Exon 21	147401773	147401866								
Exon 22	147518389	147518628								
Exon 23	147544131	147544211								
Exon 24	147550157	147555734								

Note: Positions of *CNTNAP2* exons and SNP markers with respect to genomic sequence are given in the lefthand columns. The **SNP** column gives the two possible alleles for the typed SNP, with the major allele stated first. Where appropriate, the putative risk allele is highlighted in red. For all SNPs, the identified risk allele was consistent across all traits. **MAF** indicates the major allele frequency in the SLIC families tested. The **P-value** column gives the p-value for association between the given SNP and phenotype within an orthogonal QTDT model. The most significant P-values (those that are <0.01) are indicated in red. There are clusters of SNPs with P-values <0.05 in the exon 13-15 region for all three phenotypes, with NWR showing the highest significance. **Effect** denotes the effect size of the major allele upon the trait (e.g. it is estimated that each risk allele of rs17236239 carried confers a 5.53 drop in the NWR score, a 3.21 drop in the ELS score, and a 2.95 drop in the RLS score).

Table S4. QTDT analyses of 9-snp haplotypes

Allele	**Haplotype**	**Frequency**	**NWR**	**ELS**	**RLS**
1	ATCGTCGGA	0.353	0.0006	0.0686	0.0457
2	GCTTCAAAT	0.285	0.0155	0.2279	0.0467
3	ATCGTCAGA	0.169	0.2499	0.411	0.4149
4	ATTTCAAAA	0.129	0.3627	0.2395	0.2477
5	ATTTCAAAT	0.045	0.0219	0.1466	0.5449
rare alleles		0.021	NT	NT	NT

Note: Frequency of MERLIN-constructed haplotypes for the 9 SNPs that showed single-SNP NWR associations at P<0.01 (rs851715, rs10246256, rs2710102, rs759178, rs1922892, rs2538991, rs17236239, rs2538976, & rs2710117). See Table S3 for information on individual SNPs. Note that the alleles which make up the significantly associated haplotype *ht1* are identical to the putative NWR risk alleles identified in the single-SNP analyses of these markers (Table S3). P-values are given for association between the trait shown and the marker haplotype, obtained from QTDT using an orthogonal model.

SUPPLEMENTARY REFERENCES

1. Vernes SC, Nicod J, Elahi FM, et al. Functional genetic analysis of mutations implicated in a human speech and language disorder. Hum Mol Genet 2006;15(21):3154-67.

2. Vernes SC, Spiteri E, Nicod J, et al. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. Am J Hum Genet 2007;81(6):1232-50.

3. Oberley MJ, Farnham PJ. Probing chromatin immunoprecipitates with CpG-island microarrays to identify genomic sites occupied by DNA-binding proteins. Methods Enzymol 2003;371:577-96.

4. Karolchik D, Baertsch R, Diekhans M, et al. The UCSC Genome Browser Database. Nucleic Acids Res 2003;31(1):51-4.

5. Wang X, Seed B. A PCR primer bank for quantitative gene expression analysis. Nucleic Acids Res 2003;31(24):e154.

6. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 2001;25(4):402-8.

7. Abu-Khalil A, Fu L, Grove EA, Zecevic N, Geschwind DH. Wnt genes define distinct boundaries in the developing human brain: implications for human forebrain patterning. J Comp Neurol 2004;474(2):276-88.

8. Teramitsu I, Kudo LC, London SE, Geschwind DH, White SA. Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. J Neurosci 2004;24(13):3152-63.

9. Abrahams BS, Tentler D, Perederiy JV, Oldham MC, Coppola G, Geschwind DH. Genome-wide analyses of human perisylvian cerebral cortical patterning. Proc Natl Acad Sci U S A 2007;104(45):17849-54.

10. The SLI Consortium. A genomewide scan identifies two novel loci involved in specific language impairment. Am J Hum Genet 2002;70(2):384-98.

11. The 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(6):1225-38.

12. Burden V, Stott CM, Forge J, Goodyer I. The Cambridge Language and Speech Project (CLASP). I .Detection of language difficulties at 36 to 39 months. Dev Med Child Neurol 1996;38(7):613-31.

13. Clark A, O'Hare A, Watson J, et al. Severe receptive language disorder in childhood--familial aspects and long-term outcomes: results from a Scottish study. Arch Dis Child 2007;92(7):614-9.

14. Conti-Ramsden G, Botting N. Characteristics of children attending language units in England: a national study of 7-year-olds. Int J Lang Commun Disord 1999;34(4):359-66.

15. Semel EM, Wiig EH, Secord W. Clinical Evaluation of Language Fundamentals - Revised. San Antonio, Texas: Psychological Corporation; 1992.

16. Gathercole SE, Willis CS, Baddeley AD, Emslie H. The Children's Test of Nonword Repetition: a test of phonological working memory. Memory 1994;2(2):103-27.

17. Bishop DV, North T, Donlan C. Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. J Child Psychol Psychiatry 1996;37(4):391-403.

18. Newbury DF, Bishop DV, Monaco AP. Genetic influences on language impairment and phonological short-term memory. Trends Cogn Sci 2005;9(11):528-34.

19. Fan JB, Oliphant A, Shen R, et al. Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 2003;68:69-78.

20. Fiddy S, Cattermole D, Xie D, Duan XY, Mott R. An integrated system for genetic analysis. BMC Bioinformatics 2006;7:210.

21. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 2002;30(1):97-101.

22. Abecasis GR, Cardon LR, Cookson WO. A general test of association for quantitative traits in nuclear families. Am J Hum Genet 2000;66(1):279-92.