A Functional Genetic Link between Distinct Developmental Language Disorders


BACKGROUND
Rare mutations affecting the FOXP2 transcription factor cause a monogenic speech and language disorder. We hypothesized that neural pathways downstream of FOXP2 influence more common phenotypes, such as specific language impairment.

METHODS
We performed genomic screening for regions bound by FOXP2 using chromatin immunoprecipitation, which led us to focus on one particular gene that was a strong candidate for involvement in language impairments. We then tested for associations between single-nucleotide polymorphisms (SNPs) in this gene and language deficits in a well-characterized set of 184 families affected with specific language impairment.

RESULTS
We found that FOXP2 binds to and dramatically down-regulates CNTNAP2, a gene that encodes a neurexin and is expressed in the developing human cortex. On analyzing CNTNAP2 polymorphisms in children with typical specific language impairment, we detected significant quantitative associations with nonsense-word repetition, a heritable behavioral marker of this disorder (peak association, $P=5.0 \times 10^{-5}$ at SNP rs17236239). Intriguingly, this region coincides with one associated with language delays in children with autism.

CONCLUSIONS
The FOXP2–CNTNAP2 pathway provides a mechanistic link between clinically distinct syndromes involving disrupted language.
Developmental Disorders of Speech, Language, and Communication Account for 40% of Referrals to Pediatric Services

Although many children grow out of early language delay, others have persistent difficulties with language expression and comprehension, despite normal nonverbal ability and lack of an obvious reason. In some children, developmental speech or language impairments are part of a broader syndrome such as autism, in which these deficits are accompanied by unusual repetitive behaviors and disturbances in social interaction. More commonly, such impairments occur in the absence of autistic features. Longitudinal studies have indicated that when language impairments persist to school age, they are likely to be associated with enduring academic and psychiatric problems.

Developmental speech and language disorders are highly heritable, with most cases showing complex multifactorial inheritance. The isolation of relevant genetic effects will yield new insights into the causes of such impairments, along with improved classification, diagnosis, and treatment. One notable success in this area was the discovery that heterozygous disruptions of the FOXP2 gene cause a rare mendelian speech and language disorder. Point mutations and chromosomal abnormalities that affect FOXP2 are associated with difficulties in the learning and production of sequences of oral movements, which impair speech (also called developmental verbal dyspraxia or childhood apraxia of speech). The affected persons also have variable levels of impairment in expressive and receptive language, extending to problems with production and comprehension of grammar. However, FOXP2 disruptions are rare. It has been estimated that approximately 2% of people with verbal dyspraxia carry etiologic point mutations in this gene.

Specific language impairment is the most frequently diagnosed form of developmental language disorder, affecting up to 7% of children who are 5 or 6 years of age. Although there is considerable variation in the profile of linguistic deficits observed and in the functions affected (expressive, receptive, or both), specific language impairment often occurs without accompanying difficulties in speech articulation. For example, an epidemiologic study showed that only about 5 to 8% of children with persistent specific language impairment had a significant speech delay. Moreover, analyses of FOXP2 in persons with typical forms of specific language impairment have not detected etiologic mutations or evidence of association. Mutation of FOXP2 itself is therefore unlikely to be a major risk factor for common language impairments. Indeed, to date we know of no report of a gene associated with typical specific language impairment.

Because FOXP2 encodes a neurally expressed transcription factor, we reasoned that one or more of the genes that it regulates in the brain might be implicated in common language-related phenotypes. Here we describe the isolation of a novel FOXP2-regulated target with neural functions and provide evidence of its association with language-related deficits in a large set of well-characterized families with specific language impairment.

**Methods**

**Screening for Targets of FOXP2**

We engineered the human neuroblastoma SH-SY5Y cell line to stably express FOXP2 and then, using this transfected cell line, carried out unbiased screening for genomic sites bound by FOXP2 protein. This involved the use of chromatin immunoprecipitation with anti-FOXP2 antibodies, followed by shotgun sequencing of purified DNA, a process of randomly cloning fragments of DNA and then determining their sequence (for details, see the Supplementary Appendix, available with the full text of this article at www.nejm.org). We determined the positions of DNA sequences that were isolated with chromatin immunoprecipitation, using BLAT on the University of California, Santa Cruz, Genome Server (http://genome.ucsc.edu/), which enabled identification of putative target genes.

**Validation of Binding and Regulation by FOXP2**

Binding of FOXP2 to target sites was independently verified and further localized with the use of semiquantitative polymerase-chain-reaction (PCR) assay of chromatin isolated from additional chromatin-immunoprecipitation experiments and electrophoretic mobility shift assays (EMSAs), according to protocols reported previously. Regulation of putative target genes was assessed with...
the use of quantitative reverse-transcriptase PCR (RT-PCR) of RNA extracted from SH-SY5Y cells expressing different FOXP2 levels, as described previously,\textsuperscript{18} (see Table S1 in the Supplementary Appendix for primer sequences). In situ hybridization was performed on human fetal brains,\textsuperscript{19} as described in the Methods section in the Supplementary Appendix.

**STUDY SUBJECTS**

The study subjects were members of epidemiologically and clinically ascertained families identified by the Specific Language Impairment Consortium.\textsuperscript{20,21} These families were recruited from four sites in the United Kingdom: the Comen Centre at Guy’s Hospital,\textsuperscript{20,21} the Cambridge Language and Speech Project,\textsuperscript{22} the Child Life and Health Department at Edinburgh University,\textsuperscript{23} and the Manchester Language Study.\textsuperscript{24} Families were selected through a proband with specific language impairment whose past or current language skills were 1.5 SD or more below the normative mean for the child’s age on the Clinical Evaluation of Language Fundamentals–Revised (CELF-R) scale,\textsuperscript{25} a tool that is routinely used for diagnosis and follow-up evaluation of language disorders in school-age children. (Scores on the scale range from 50 to 150, with a mean of 100 and a standard deviation of 15 in the general population. Lower scores indicate poorer performance.) We excluded any children with a nonverbal IQ of less than 80, a clinical diagnosis of an autistic-spectrum disorder, or another known medical or developmental condition that can impair language, such as hearing loss, cleft lip, or cleft palate. Moreover, for clinically ascertained samples, children were comprehensively assessed on scales evaluating language, IQ, and behavior, and those with overt pragmatic difficulties, behavioral characteristics associated with autism, or a family history indicative of autism were also excluded.

We collected quantitative phenotypic data from probands and all available siblings. We then determined composite CELF-R scores for expressive and receptive language abilities. We also used a measure of the ability to repeat nonsense words, the Children’s Test of Nonword Repetition,\textsuperscript{26} which has been established as a robust endophenotype of specific language impairment.\textsuperscript{12} (Scores on the scale range from 46 to 141, with a mean of 100 and a standard deviation of 15 in the general population. Lower scores indicate poorer performance.) This measure is thought to provide an index of phonologic short-term memory.\textsuperscript{12} Children with specific language impairment perform particularly poorly on nonsense-word repetition, and impaired phonological short-term-memory has been proposed as a core deficit in the disorder. An impairment in the ability to repeat nonsense words is highly heritable, persists in persons with historical language problems that have otherwise resolved,\textsuperscript{27} and appears to be relatively unaffected by environmental factors.\textsuperscript{28} Additional information on the consortium families has been reported previously,\textsuperscript{20,21} and is available in Table S2 in the Supplementary Appendix, which shows means, standard deviations, and intertrait correlations for language measures used in this study. Written informed consent was obtained from all subjects or their parents; assent was obtained from children of appropriate age.

**SINGLE-NUCLEOTIDE-POLYMORPHISM GENOTYPING**

To directly test the hypothesis that variants of the identified FOXP2 target (the CNTNAP2 gene) may increase susceptibility to common language impairments, we genotyped single-nucleotide polymorphisms (SNPs) in consortium families, followed by quantitative association analyses of measures of expressive and receptive language abilities and nonsense-word repetition. We genotyped and validated 38 SNPs from the CNTNAP2 locus on chromosome 7q35 in samples from 847 members of 184 consortium families, using Golden Gate assays on the Illumina platform (for details, see the Methods section and Table S3 in the Supplementary Appendix).

**STATISTICAL ANALYSIS**

For analyses of differences in gene expression in SH-SY5Y cells, we assessed statistical significance using unpaired t-tests (two-tailed). For family-based association analyses of SNP data from the consortium series, we used a quantitative transmission disequilibrium test (QTDT), adopting an orthogonal association model that considers only the within-family variance and is robust to population stratification.\textsuperscript{29} After identifying significant single SNP associations, we used the Merlin package\textsuperscript{30} to generate haplotypes for the cluster of nine associated SNPs, which were similarly analyzed.
with QTDT. Finally, we investigated the possibility of an effect of sex or imprinting within QTDT, using these nine SNP-tag haplotypes.

RESULTS

CNTNAP2 AS A TARGET OF FOXP2

To identify candidate genes that might be involved in typical specific language impairment, we used an unbiased screening method to isolate genomic fragments bound by FOXP2 protein in chromatin of human neuronlike cells. We thereby discovered a FOXP2-bound fragment that was of particular interest, because it was located within intron 1 of CNTNAP2 (Fig. 1A). This gene encodes Caspr2, a member of the neurexin superfamily of transmembrane proteins, found at the nodes of Ranvier in myelinated nerve fibers. In mice, Caspr2 is important for the regulation of the localization and maintenance of Shaker-type voltage-activated potassium channels and is implicated in neuronal recognition and cell adhesion. In humans, it has been suggested that CASPR2 is involved in cortical development, possibly mediating intercellular interactions during neuroblast migration and laminar organization.

We used PCR to amplify sequences spanning the FOXP2-bound fragment of CNTNAP2 in independent FOXP2 chromatin-immunoprecipitation samples and in control samples in which no antibodies were used and observed evidence of enrichment only when FOXP2-specific antibodies had been used to isolate the chromatin (Fig. 1A). Primers amplifying regions of 1000 bp or more away from the bound fragment did not display FOXP2–chromatin immunoprecipitation enrichment. FOXP2 is thought to bind chromatin as a dimer, and our in silico analyses of the chromatin immunoprecipitation–enriched fragment identified two adjacent sites, separated by 48 bases, matching a known consensus sequence for FOXP2 binding (CAAAATT). EMSA analyses indicated that FOXP2 could bind both sites (data not shown). At each site, binding could be disrupted by the mutation of three core nucleotides of the recognition sequences (CAAAATT→CGGGTT), with more dramatic effects observed for the 5′ site. Full competition assays for this site showed highly efficient and specific binding by FOXP2 (Fig. 1B).

We then used quantitative RT-PCR to directly test whether modulation of FOXP2 protein levels would yield altered CNTNAP2 expression. Indeed, CNTNAP2 messenger RNA levels were consistently and significantly reduced in neuronlike cells that were stably transfected with FOXP2, as compared with sham-transfected control samples (Fig. 2A).

A recent genomewide analysis of differential gene expression in the developing human cerebral cortex independently highlighted CNTNAP2 as a gene with substantial enrichment in frontal gray matter, which is primarily restricted to the region between the orbital gyrus and superior frontal anlage, spanning the inferior and middle frontal gyri. Because FOXP2 is also expressed in the developing human cortex, we carried out expression analyses of this structure in fetal tissue (18 to 22 weeks’ gestation) through in situ hybridization. We observed complementary patterns with respect to cortical lamination: CNTNAP2 expression was lowest in layers that showed the highest levels of FOXP2 (Fig. 2B). These in vivo findings are consistent with our data from neuronal models, supporting negative regulation of human CNTNAP2 expression by FOXP2.

ASSOCIATION ANALYSES OF CNTNAP2

Several studies underscored CNTNAP2 as a particularly compelling candidate gene to test for association with specific language impairment. In addition to our identification of it as a direct neural target of FOXP2, it has known neuronal functions, and its expression is enriched in hu-
A Genetic Link in Distinct Developmental Language Disorders

man language-related circuitry. Furthermore, the gene is disrupted in a family with Tourette's syndrome, and a rare point mutation causes a severe recessive disorder involving cortical dysplasia and focal epilepsy, associated with language regression and autistic characteristics. Recent independent studies have implicated variants at the CNTNAP2 locus in autistic-spectrum disorders, with one study showing association with a measure of language delay (the age at the first spoken word) in multiplex autism families.

We therefore went on to assess CNTNAP2 involvement in specific language impairment by genotyping polymorphisms across the locus in the large series of consortium families and testing for marker-trait association, using a family-
based association method. Using an approach that was consistent with previous studies of this series, we analyzed quantitative phenotypes from probands and all available siblings, regardless of the diagnosis of specific language impairment, and focused on composite diagnostic measures of expressive and receptive language abilities, as well as a test of nonsense-word repetition, which was previously established as a robust endophenotype.

We observed significant associations (with \( P \) values from 0.01 to \( 5.0 \times 10^{-5} \)) between nonsense-word repetition and nine intronic SNPs (rs851715, rs10246256, rs2710102, rs759178, rs1922892, rs2538991, rs17236239, rs2538976, and rs2710117), all mapping between exons 13 and 15. The most strongly associated SNP was rs17236239 (\( P = 5.0 \times 10^{-5} \)) (Fig. 3A, and Table S3 in the Supplementary Appendix). Even after an overly conservative Bonferroni correction for testing of multiple SNPs was made, this \( P \) value remained significant (\( P = 0.002 \)). The rs17236239 SNP was also the marker showing strongest evidence of association with expressive language abilities (\( P = 0.008 \)). The exon 13–15 region was similarly implicated in analyses of receptive language abilities, but in this case the strongest association was observed for a different SNP, rs4431523 (\( P = 0.003 \)).

We constructed multimarker haplotypes with the 9 SNPs implicated in the ability to repeat nonsense words and observed 11 different combinations. Four haplotypes represented 94% of subjects (Table S4 in the Supplementary Appendix). The most common haplotype, \( h t l \), negatively influenced the ability to repeat nonsense words; it was more often present in family members with poor scores (a frequency of 40% among those with scores of \( >2 \) SD below the population mean) than in those with good scores (a frequency of 29% among those with scores of \( >0.5 \) SD above the population mean). We classified probands and siblings according to the number of copies (none, one, or two) of this putative risk haplotype they had and calculated the mean score for nonsense-word repetition for each group (Fig. 3B). There was a decrease of approximately 6 points (0.4 SD) between the mean scores for nonsense-word repetition of children carrying no copies of \( h t l \) (mean, 95.2) and those carrying two copies (mean, 89.4). We observed a difference of similar magnitude in scores between children carrying no copies of \( h t l \) and those carrying one
copy of $ht1$ (mean, 89.7), a finding that suggests a dominant effect. Although only 49 children carried two copies of $ht1$, as compared with 176 with no copy and 174 with one copy, we observed very similar results in an augmented data set incorporating all available parental scores (Fig. 3B). Moreover, family-based quantitative association analyses of the nine-marker haplotypes and scores on nonsense-word repetition yielded a P value of $6.0 \times 10^{-4}$ for $ht1$, again indicating that this haplotype is significantly associated with impaired language performance (Table S4 in the Supplementary Appendix).

Investigations of CNTNAP2 in patients with autism showed an increased association in families with affected males\(^{37}\) and also suggested the possibility of imprinting.\(^{38}\) We therefore repeated our QTDT analysis of the haplotype associated with impaired nonsense-word repetition in the families with specific language impairment, using sex as a covariate and testing for differences in transmission of paternal and maternal alleles. We found no evidence of a role of sex ($P=6.0 \times 10^{-4}$ with no adjustment for sex, $P=8.0 \times 10^{-4}$ with adjustment for sex) or of imprinting ($P=0.27$) at this locus.

**DISCUSSION**

We have shown that FOXP2, which is mutant in people with a rare speech and language disorder, directly regulates expression of the CNTNAP2 gene. We went on to demonstrate that variants of CNTNAP2 are associated with deficits in common forms of language impairment. In so doing, we provide an example of how knowledge of the genetic cause of a rare single-gene disorder provides an entry point into the causes of a more complex phenotype. Further analyses of the relevant regulatory networks — including the FOXP2–CNTNAP2 pathway identified here — may lead to a better understanding of neurogenetic mechanisms involved in typical language disorders.

The FOXP2-bound fragment of CNTNAP2 lies outside the classically defined regulatory regions of the genome represented on available promoter-based microarrays. It therefore escaped detection in recently published efforts in which chromatin immunoprecipitation with FOXP2 antibodies was coupled to screening of such microarrays.\(^{18,40}\) Indeed, large-scale surveys of transcription-factor binding have indicated that functional regulatory sequences often lie far from known promoters, with many of such sequences mapping within introns.\(^{41}\)

Thus far, CNTNAP2 is the only FOXP2 target that we have tested for association in specific language impairment. Of all FOXP2 targets identi-
fied to date, \textsuperscript{18,19,40} we prioritized \textit{CNTNAP2} for association testing in common language disorders because it represented the most compelling of candidates, with converging support from multiple lines of independent investigation. That these first association analyses were positive illustrates the promise of our function-based approach. In future work, we will go on to assess other neural targets of \textit{FOXP2} in a similar manner.

There is considerable debate over the existence of shared causes of neurodevelopmental syndromes that are treated as diagnostically distinct, such as autism and specific language impairment.\textsuperscript{5,42,43} In this study, we observed an association between the endophenotype of nonsense-word repetition and polymorphisms in the exon 13–15 region of \textit{CNTNAP2} in children with specific language impairment. A study of the gene in children with autism\textsuperscript{47} showed an association between polymorphisms in the exon 13–15 region (similarly centered on tagging-SNP rs2710102) and the age at the first spoken word. The same SNP alleles were associated with susceptibility in both studies (Table S3 in the Supplementary Appendix). Therefore, similar \textit{CNTNAP2} variants may represent susceptibility factors for language-related deficits in both specific language impairment and autism.

The \textit{CNTNAP2} associations we describe here for specific language impairment are not simply a replication of those reported previously for patients with autistic-spectrum disorders. We made a rigorous effort to exclude persons with autistic-spectrum disorders from our analysis. Although diagnostic boundaries are not always clear, it is unlikely that persons who received a misdiagnosis of autistic-spectrum disorder remained in such numbers that they accounted for the strength of the association that we observed. Moreover, we used a quantitatively defined endophenotype previously proposed to underlie typical specific language impairment, rather than a categorical designation of affected status.

Instead, our findings are compatible with the idea that different components of autistic-spectrum disorders (communication deficits, impaired social interaction, and rigid or repetitive behaviors) may be under different genetic influences.\textsuperscript{44} In this view, language impairments are observed in relatively pure form in specific language impairment, but when they occur in association with other social and behavioral deficits, such impairments can result in a diagnosis of autism.\textsuperscript{45} Thus, altered \textit{CNTNAP2} function or regulation could represent a shared mechanism contributing to language-related endophenotypes in both specific language impairment and autism. These findings illustrate the value of using endophenotypes for the genetic dissection of such disorders.

In conclusion, by integrating functional genomics and quantitative trait analyses, we have identified a shared neurogenetic pathway that is disturbed in distinct forms of language impairment. This work represents a move away from isolated studies of individual genes and toward an understanding of molecular networks that may go awry in neurodevelopmental disorders affecting language.

Supported by project and program grants from the Wellcome Trust (to Drs. Fisher and Monaco, respectively), a pilot grant from Autism Speaks (to Dr. Fisher), awards (MH75028-R21, MH081754-R01, MH60233-R37, and HD055784-P50) from the National Institute of Mental Health (to Dr. Geschwind), a Christopher Welch Biological Sciences Scholarship from the University of Oxford (to Dr. Verres), a postdoctoral fellowship from the Tourette Syndrome Association (to Dr. Abrahams), and a Marie Curie Intra-European Fellowship (to Dr. Nicod). Dr. Monaco is a Wellcome Trust Principal Research Fellow, and Dr. Fisher is a Royal Society Research Fellow. Drs. Newbury, Monaco, and Fisher and Ms. Winchester are members of the Specific Language Impairment Consortium.

No potential conflict of interest relevant to this article was reported.

We thank all the families who participated in this study through the Specific Language Impairment Consortium and other consortium members for their contributions to this work: L. Addis, Wellcome Trust Centre for Human Genetics, University of Oxford; I.M. Goodyer, Cambridge Language and Speech Project; E. Simonoff and P.F. Bolton, Child and Adolescent Psychiatry Department and Medical Research Council Centre for Social, Developmental and Genetic Psychiatry, Institute of Psychiatry; V. Sionimens and G. Baird, Newcomen Centre, Guy's Hospital; A. Everitt, E. Hennessy, M. Main, D. Shaw, and P.J. Helms, Department of Child Health, University of Aberdeen; A.D. Kindley, Raeden Centre and Grampian University Hospitals Trust; A. Clark and J. Watson, Speech and Hearing Sciences, Queen Margaret University; A. O’Hare, Department of Reproductive and Developmental Sciences, University of Edinburgh; J. Seckl, Molecular Medicine Centre, University of Edinburgh; H. Cowie, Department of Speech and Language Therapy, Royal Hospital for Sick Children, Edinburgh; W. Cohen, Department of Educational and Professional Studies, University of Strathclyde; J. Nasir, Clinical Developmental Sciences, St. George’s University of London; D.V.M. Bishop, Department of Experimental Psychology, University of Oxford; Z. Simkin and G. Conti-Ramsden, Human Communication and Deafness, School of Psychological Sciences, University of Manchester; and M. Falcaro and A. Pickles, Biostatistics Group, School of Epidemiology and Health Science, University of Manchester; as well as D.V.M. Bishop, G. Baird, G. Conti-Ramsden, and A. O’Hare for their assistance with the preparation of the manuscript.
A GENETIC LINK IN DISTINCT DEVELOPMENTAL LANGUAGE DISORDERS

REFERENCES


