

DECIPHERING THE GENETIC BASIS OF SPEECH AND LANGUAGE DISORDERS

Simon E. Fisher, Cecilia S.L. Lai, and Anthony P. Monaco

Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford OX3 7BN, United Kingdom; email: simon.fisher@well.ox.ac.uk, cecilia.lai@well.ox.ac.uk, anthony.monaco@well.ox.ac.uk

Key Words linkage analysis, quantitative trait locus, specific language impairment, *SPCH1*, *FOXP2*

■ **Abstract** A significant number of individuals have unexplained difficulties with acquiring normal speech and language, despite adequate intelligence and environmental stimulation. Although developmental disorders of speech and language are heritable, the genetic basis is likely to involve several, possibly many, different risk factors. Investigations of a unique three-generation family showing monogenic inheritance of speech and language deficits led to the isolation of the first such gene on chromosome 7, which encodes a transcription factor known as *FOXP2*. Disruption of this gene causes a rare severe speech and language disorder but does not appear to be involved in more common forms of language impairment. Recent genome-wide scans have identified at least four chromosomal regions that may harbor genes influencing the latter, on chromosomes 2, 13, 16, and 19. The molecular genetic approach has potential for dissecting neurological pathways underlying speech and language disorders, but such investigations are only just beginning.

THE ENIGMA OF SPEECH AND LANGUAGE ACQUISITION

One of the most intriguing aspects of the human condition is our unique capacity for rapidly acquiring speech and language in the early years of life. Although clearly dependent on a certain amount of input from the environment, the vast majority of children develop intricate speech and language abilities effortlessly and without formal instruction. Usually by the time a child is four years of age he or she can employ a vocabulary of a few thousand words to construct a vast number of complex meaningful sentences, which can be conveyed to others via precise motor control of the articulatory apparatus. Aspects of this elusive ability to acquire speech and language may be encoded within the genetic makeup of our species (reviewed in Pinker 1994). However, not every child who grows up in a language-rich environment goes on to develop normal communication skills. In some cases this is an obvious consequence of a condition such as mental retardation, hearing loss, cleft palate, autism, or cerebral palsy. Nevertheless, having ruled out

these and other similar causes, there remain a significant number of children with unexplained deficits in speech and language acquisition (Bishop 2001a).

Since diagnostic criteria of developmental speech and language disorders are largely exclusionary (that is, they require the absence of specific symptoms), there can be considerable heterogeneity among those individuals who are classified as affected. Over the years, clinicians and researchers have attempted to resolve this by dividing speech and language disorders into subtypes. Current diagnostic schemes recognize the existence of multiple forms of language impairment. For example, the DSM-IV classification system (APA 1994) includes three subtypes—a failure to use speech sounds that are appropriate for age and dialect (phonological disorder), excessive deficits in expressive language despite normal receptive skills (expressive disorder), and significant impairment in both expressive and receptive language abilities (mixed expressive-receptive disorder). As yet there are little data to support the notion that these subtypes truly represent discrete entities with distinct aetiologies (Bishop et al. 1995). Another problem for diagnosis is that there can be substantial variability in cognitive profile for the same individual at different ages. Finally, in commonly used classification schemes, a positive diagnosis requires significant discrepancy between language ability and nonverbal intelligence. The validity of this practice has been disputed by studies suggesting that the nature of language problems is similar in children with “low-language” regardless of discrepancy criteria (Tomblin & Pandich 1999), and that early language difficulties influence subsequent nonverbal development (Rutter & Mawhood 1991, Tallal et al. 1991).

Thus, unravelling the phenotypic complexity of speech and language impairment represents a major challenge for those seeking to ascertain the underlying causes. Although there is still a formidable amount of work remaining to be done in exploration of phenotype, the past few years have seen the emergence of a new approach for dissection of speech and language impairments, which exploits recent developments in molecular genetic technology. The current review evaluates the state of the field and discusses the implications of molecular genetic strategies for future research.

DEVELOPMENTAL DISORDERS OF SPEECH AND LANGUAGE ARE HIGHLY HERITABLE

There is little evidence that environmental factors, such as inadequate stimulation from caregivers or exposure to perinatal hazards, are common causes of developmental language disorders (Bishop 2001a). During the course of the past two decades, it has been reliably established that speech and language problems tend to cluster in families (Neils & Aram 1986, Lewis et al. 1989, Tallal et al. 1989, Tomblin 1989, Lahey & Edwards 1995). Familial aggregation is compatible with a role for genetic risk factors but could be confounded by environmental influences shared between individuals of the same family. This issue can be resolved to a large extent by investigating samples of twin pairs, in which at least one member

is affected with a speech and language disorder (Lewis & Thompson 1992, Bishop et al. 1995, Tomblin & Buckwalter 1998). Such studies have consistently demonstrated elevated concordance for speech and language difficulties in monozygotic (MZ) twins, who have a virtually identical genetic makeup, versus dizygotic (DZ) twins, who are as genetically similar as ordinary siblings, sharing roughly half of their segregating alleles. One of the largest of these studies, involving 90 twin pairs, reported 70% concordance in MZ pairs, as compared to 46% in DZ pairs, when employing a strict definition of speech and language disorder (Bishop et al. 1995). When diagnostic criteria were broadened to include children with a history of problems, plus those with low language but no substantial discrepancy between verbal and nonverbal skills, Bishop and coworkers found that MZ concordance increased to almost 100%, versus a DZ concordance approaching 50%.

Support for genetic involvement in the aetiology of speech problems has also arisen from studies of phenotypic outcomes in adopted children. In a study of 156 adopted and nonadopted children, Felsenfeld & Plomin (1997) demonstrated that a positive history of speech difficulties in biological parents leads to a significant increase in a child's risk of developing similar problems, even if living with adoptive parents who have no impairment. In contrast, they found no risk increase for adopted children as a consequence of living with an affected parent. Note that these data suggest the importance of genetic factors or early prenatal influences (or a combination of the two) on speech development, but the adoption design is unable to distinguish between these possibilities.

QUANTITATIVE EVALUATION OF SPEECH AND LANGUAGE DEFICITS FOR GENETIC ANALYSIS

A central issue for genetic analysis of any complex trait is the way in which the phenotype is defined. For speech and language impairments, clinical or research-based diagnostic procedures usually involve assessment of a subject's performance on a series of psychometric tests. An all-or-none diagnosis of disorder is derived from applying thresholds to quantitative measures of language ability. These thresholds are set somewhat arbitrarily with reference to the range of variability in normal populations (Bishop 1994, Cole et al. 1995). For example, under ICD-10 guidelines, a positive diagnosis in a young subject requires that language skills must fall outside the 2-SD limit for the child's chronological age (WHO 1993). However, this is an artificial boundary; it is debatable whether the deficits of subjects who meet this criterion are qualitatively distinct from language-delayed children with less severe impairment. Accordingly, some researchers believe that language disorders simply represent the extreme lower tail of normal developmental variation, with similar aetiological factors operating across the entire range of ability (see Bishop 2001a).

An alternative approach to the problem of phenotype definition, which surmounts some problems associated with categorical diagnoses, is to employ quantitative scores directly in genetic analysis. Twin-based methods allow estimation of

heritability of a particular quantitative measure, that is, the proportion of observed phenotypic variability in that measure that can be attributed to genetic factors. (Note that heritability is a statistical property that is specific to the population under study and may change if the environment is altered.) DeFries & Fulker (1985) developed regression-based techniques for estimating the heritability of extreme deficits in ability (referred to as group heritability). Their method exploits selected samples of twin pairs in which at least one twin has a phenotypic score in the extreme lower tail of the distribution. If DZ co-twins regress further toward the unselected population mean than MZ co-twins, then this may indicate a role for genetic factors. The significance of the relationship between twin zygosity and phenotypic similarity in a selected sample is assessed via a simple statistical test, and direct estimates of group heritability can be obtained (DeFries & Fulker 1985). Using this method, Bishop and coworkers (1995) demonstrated significant heritability (close to 100%) for deficits in measures of expressive and receptive language ability in their twin sample. Similarly, a study of selected pairs taken from a large epidemiological sample of two-year-old twins yielded a group heritability of ~73% for extreme deficits in a measure of productive vocabulary (Dale et al. 1998).

Direct genetic analysis of quantitative data circumvents issues regarding arbitrary categorical definition and external validity of subtyping. As we discuss elsewhere in this article, quantitative approaches can be powerful for localizing genetic risk factors to particular genomic regions. However, such methods have associated problems of their own. In particular, a subject's speech and language abilities are usually evaluated with a variety of different tests. Given the uncertainty over the primary deficit(s) responsible for developmental speech and language disorders, it is seldom clear which measures will be the most appropriate indices of severity. As a consequence, researchers often have to choose between multiple testing of correlated measures and analysis of a composite score that does not fully capture all aspects of the phenotype. The results of heritability studies can help to focus attention on those measures most relevant to genetic mapping efforts. In addition, they may provide extra insight into the aetiological significance of different processes.

For example, although Bishop et al. (1995) estimated high heritabilities for absolute levels of language deficit, they found little evidence for genetic influence on measures of discrepancy between language ability and nonverbal IQ, which suggests that the latter may be of limited use in genetic mapping studies. In later work, Bishop and colleagues (1999) investigated two different tests, each of which can be used as a reliable indicator to distinguish language-impaired children from unaffected control subjects. One of these was a task that evaluates a subject's ability to repeat orally presented nonsense words of varying complexity, and this was found to be highly heritable. It has been suggested that deficits in this "non-word" repetition test reflect an underlying impairment in phonological working memory, a cognitive system involved in short-term storage of phonetic sequences (Gathercole et al. 1994), and the data from heritability studies provide some support

for this hypothesis. The second task evaluated by Bishop et al. required the auditory discrimination of tone sequences presented at variable rates (Tallal & Piercy 1973). Poor performance in this task appeared to be predominantly influenced by environmental factors shared between siblings rather than by genetic predisposition. Thus, although deficits in auditory temporal processing may be associated with language impairment (Tallal & Piercy 1973), there is little evidence that such deficits have a genetic basis. Bishop and colleagues used small sample sizes for their DeFries-Fulker analyses, so this inference of differential heritability for the two tasks requires further study. Nevertheless, this work provides an excellent illustration of how quantitative methods can be used for exploring the biological underpinnings of complex disorders.

STRATEGIES FOR GENETIC DISSECTION OF SPEECH AND LANGUAGE DISORDERS

Although there is a substantial heritable component contributing to speech and language impairment, the underlying genetic basis is complex. Family studies indicate that there are likely to be several, possibly many, different genetic risk factors, with distinct combinations of these involved in different affected individuals. As a consequence, the correspondence between genotype and phenotype is eroded, which can be highly problematic for genetic analyses (Fisher 2002). Families segregating speech and language impairment usually include cases of reduced penetrance—subjects with high-risk genotypes who do not have problems—and/or cases of phenocopy—individuals who manifest the disorder but do not have a high-risk genotype. Faced with the challenge of deciphering the complex genetic basis of speech and language impairment, there are several routes that can be taken.

Candidate Gene Approaches

In some cases one can exploit preexisting insights into the pathology of a disorder and established data on protein function to identify genes that are likely to be involved (Collins 1995). Targeted screening of these functional candidates in affected individuals might then yield evidence that they are indeed implicated in disease aetiology. Although such strategies can also take advantage of genetic mapping information from families inheriting the trait (as discussed below), this is not a necessary prerequisite for success. Key examples of the value of candidate gene studies for analysis of complex neurological traits come from research into Alzheimer's disease, a common cause of progressive cognitive decline in humans (Selkoe & Podlisny 2002). Biochemical investigations of brain lesions associated with Alzheimer's disease have helped to guide searches for genetic factors, which have led to identification of risk variants in the genes encoding beta-amyloid precursor protein and apolipoprotein-E. Similarly, molecular genetic studies of attention-deficit/hyperactivity disorder, a common behavioral disorder with childhood onset, have predominantly focused on genes of the dopaminergic system because

drugs targeting dopamine-related pathways are known to be effective in treating symptoms of the disorder (see Fisher et al. 2002b). By contrast, for developmental disorders of speech and language, almost nothing is known about the underlying molecular mechanisms, and there are no convincing functional hypotheses that could drive the selection of candidate genes (Fisher 2002). Thus, in this case a strategy based solely on choosing candidate genes is currently unfeasible.

The Positional Cloning Paradigm

One of the most significant advances in human genetics in the past few decades has been the development of positional cloning, which allows identification of gene variants contributing to a trait of interest without requiring any prior knowledge of relevant biological pathways or gene function. This research paradigm has become a standard tool for dissecting genetic aetiology of monogenic disorders (Collins 1995) and is now beginning to yield fruit in studies of more complex traits (Korstanje & Paigen 2002). Positional cloning essentially relies on tracking the inheritance of polymorphic genetic markers in families displaying a trait of interest and using such information to highlight chromosomal regions that are most likely to contain genes influencing that trait. The results of this process of linkage mapping (so-called because it evaluates whether transmission of a genetic marker is “linked” to inheritance of the trait) can restrict the search to a manageable subset of genes in the chromosomal region(s) of interest. These genes can be investigated further, eventually leading to isolation of specific gene variants that contribute to the trait. Exploiting the wealth of data now available from human genomic sequencing and other studies, it is often possible to prioritize which genes should be investigated from any given region of interest on the basis of their predicted functions or expression patterns. This combination of positional cloning and candidate gene strategies is sometimes referred to as a positional candidate approach.

A typical positional cloning effort begins with a genome-wide scan, a systematic search for linkage across all chromosomes, using several hundred markers. Developments in genotyping technology in recent years have greatly reduced the costs and increased the efficiency of undertaking such a scan, even when screening the large numbers of individuals required for complex trait analyses. A review of the literature estimated that by December 2000 more than one hundred genome-wide scans had been conducted for common traits such as diabetes, hypertension, obesity, asthma, and psychiatric disorders (Altmüller et al. 2001), and there have been many more since. While genome-wide searches for monogenic diseases have in general been hugely successful, scanning efforts for complex disorders have been less lucrative. The lower success of complex trait mapping is most likely due to reduced power resulting from factors like genetic heterogeneity and small effect sizes of relevant genes. Genome-wide scans for complex traits commonly yield only weak or suggestive results as to the locations of potential genetic risk factors, and there is often a lack of concordance between results obtained from different samples (Altmüller et al. 2001). Nevertheless, recent findings suggest that the

potential of positional cloning for genetic dissection of complex traits may eventually be met (Korstanje & Paigen 2002).

By adopting the positional cloning paradigm, geneticists are now starting to make progress in the search for loci that influence speech and language disorders. We describe below how these ongoing studies are employing a variety of study designs, ranging from traditional linkage analysis of multigenerational pedigrees, to contemporary quantitative methods for gene mapping in large numbers of nuclear families.

THE KE FAMILY: A UNIQUE CASE OF MONOGENIC INHERITANCE

In 1990 Hurst and coworkers reported an intriguing case of a large three-generation pedigree from the United Kingdom, in which about half of the members are affected with a severe speech and language disorder. What makes the finding so unique is the simple way in which the trait is passed down through the family, consistent with the action of a single autosomal gene with a dominant effect (Figure 1). To date, this represents the only documented case of unambiguous monogenic inheritance for a developmental speech and language disorder. The discovery led to the suggestion that disruption of a single gene could make a direct impact on

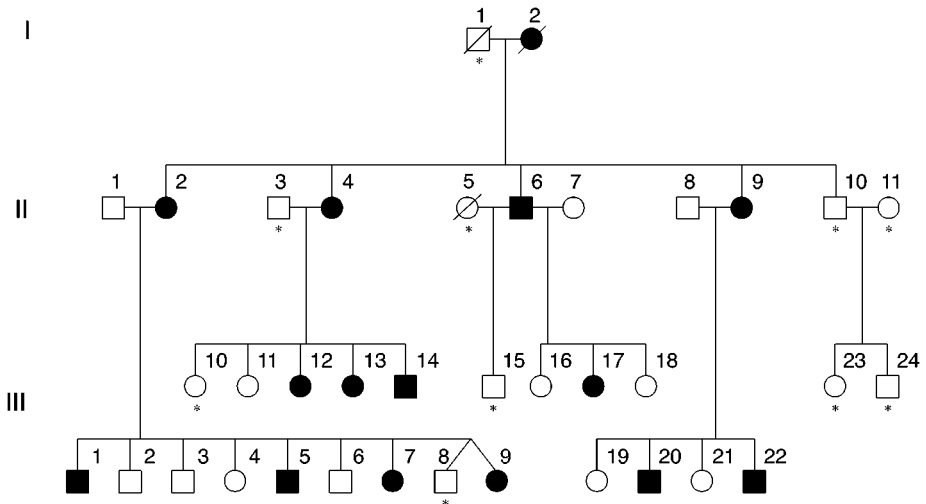


Figure 1 Pedigree diagram of family KE. Squares represent males; circles represent females. A diagonal line indicates that the individual is deceased. Family members with speech and language impairment are shaded. Asterisks indicate those individuals who were unavailable for linkage analysis. From Fisher et al. (1998) © Nature America Inc. Adapted with permission.

speech and language acquisition and further fueled the debate surrounding innate properties of language (Pinker 1994). Consequently, the KE family have become the focus of over a decade of intensive study, as researchers attempt to delimit the precise nature of the phenotype (Gopnik & Crago 1991, Vargha-Khadem et al. 1995, Gopnik & Goad 1997, Vargha-Khadem et al. 1998, Alcock et al. 2000, Watkins et al. 2002a).

Simple Inheritance: Complex Phenotype

A number of descriptions of the KE family have been published, and there have been inconsistencies in the various characterizations of the phenotype, which has led to some confusion regarding the nature of the disorder. A comprehensive review of the literature reveals that, although their pathology is clearly confined to the central nervous system (CNS), the affected members of the KE family are impaired in several related aspects of brain function. In other words, they do not have a tidy lesion of one single specific feature of cognition. This does not constitute an argument against involvement of a single genetic mutation. In most monogenic disorders simple inheritance is associated with complex phenotypes involving multiple facets. In the classic example of cystic fibrosis, severe inflammatory lung disease is accompanied by salty sweat, pancreatic insufficiency, intestinal obstruction, and male infertility, and all these problems result from mutation of a single gene encoding a chloride transporter. Thus, the mapping of simple inheritance to multifaceted phenotype observed in the KE family should not be a surprise, especially given the particular complexity of the human brain. The different aspects of the KE phenotype are as follows.

ARTICULATORY PROBLEMS Affected members have an underlying difficulty with controlling the complex coordinated mouth movements that are required for speech, referred to as developmental verbal dyspraxia (Hurst et al. 1990, Vargha-Khadem et al. 1995), which remains detectable in adulthood even after speech therapy. This is not a result of abnormalities in facial musculature, and affected individuals perform normally on all tests of limb praxis (Watkins et al. 2002a). In addition, affected individuals are not impaired when making single, simple oral movements (Alcock et al. 2000).

LANGUAGE IMPAIRMENT The disorder involves impairment in a wide range of language-related skills (Vargha-Khadem et al. 1995, Watkins et al. 2002a). It might be suggested that some of the deficits in expressive language, such as those involving word repetition, are secondary consequences of the articulation problems described above. However, there are three critical points to consider here. First, impairment is evident not just in spoken language but also in written language. Watkins et al. (2002a) demonstrated that the affected individuals perform significantly worse than unaffected members of the family on written tests of verbal fluency and nonword spelling. Second, the disorder is not confined to

expressive language but extends to the receptive domain; for example, the affected individuals show significant deficits in receptive vocabulary, assessed via a test of lexical decision, in which they have to indicate whether a presented word is real or a nonword (Watkins et al. 2002a). Third, the disorder affects both comprehension and production of grammar (Gopnik & Goad 1997, Watkins et al. 2002). Affected individuals are significantly impaired when understanding complex sentences (as assessed by picture-selection tasks) and generating word inflections (changes in tense or number) or derivations (changes in meaning).

COGNITIVE DEFICITS The mean nonverbal IQ of affected KE family members is reported to be significantly lower than that of the unaffecteds (Vargha-Khadem et al. 1995, Watkins et al. 2002a), which has led some to suggest that the disorder represents a form of general cognitive impairment. However, the moderate reduction in nonverbal ability seen in some family members does not tend to co-segregate with the disorder. The family includes cases of unaffected individuals with low nonverbal IQ (but no speech and language problems) as well as affected individuals who have normal nonverbal IQ (despite the presence of severe speech and language difficulties). Furthermore, observed deficits in verbal cognition appear to be more severe and wide-ranging than those found in the nonverbal domain. Examination of scores from the separate subtests used to measure IQ indicates that the affected members are impaired relative to the unaffecteds on only one nonverbal subtest, which involves learning of arbitrary associations between symbols and digits (Watkins et al. 2002a). In contrast, the affected individuals show significant deficits on each of the separate subtests that contribute to estimates of verbal IQ. Furthermore, Watkins and coworkers studied longitudinal test data and found that the measured nonverbal intelligence of some affected individuals had declined with increasing age. Investigations of children with other forms of language impairment have shown similar reductions in nonverbal abilities as they get older, as indexed by IQ testing (Tallal et al. 1991). The presence of a speech and language disorder puts a child at increased risk of developing educational, behavioral, and social problems in later life (Rutter & Mawhood 1991). Overall, the available data suggest that nonverbal deficits are not a primary characteristic of the KE disorder, but further study may help to clarify how these are related to other aspects of the phenotype.

Although there are clearly several aspects to the KE phenotype, debate continues regarding which of these represents the primary or core deficit (or if indeed there is one). For example, using discriminant function analysis, Watkins et al. (2002a) demonstrated that performance on a nonword repetition task involving complex articulation could, by itself, successfully discriminate affected from unaffected groups of the KE family and might thus be considered the best biological marker of the disorder. It has been suggested that poor articulation, resulting from low-level problems with controlling oral movements, leads to impaired phonological representations, which then impact development of language and learning of rules of syntax (see Alcock et al. 2000, Watkins et al. 2002a). However, Gopnik & Goad

(1997) argue that the pattern of linguistic errors made by affected individuals is not concordant with such a hypothesis. In addition, investigations of nonvocal individuals affected with cerebral palsy indicate that a complete lack of speech does not necessarily lead to significant impairment in language and grammatical skills (Bishop et al. 1990). It has also been proposed that the verbal and nonverbal deficits of the KE family may all arise from a basic deficit in sequencing or procedural learning (Watkins et al. 2002a). However, such an explanation should lead to a much broader profile of cognitive impairment than that which is actually found in the KE family (Bishop 2002).

Searching for *SPCH1*

While the question of core deficit remains unanswered, the consensus from the above studies is that the gene disrupted in the KE family is probably implicated in pathways that are important (even though they may not necessarily be specific) to the acquisition of speech and language ability. Furthermore, the apparent monogenic nature of transmission suggested that it might be possible to isolate the gene of interest with standard positional cloning techniques. Therefore, Fisher and coworkers (1998) initiated a genome-wide search in the pedigree and used traditional parametric linkage methods to localize the gene responsible to a small interval of chromosome 7. The locus, in cytogenetic band 7q31, was assigned the name *SPCH1*. Evidence for linkage to the region was highly convincing, yielding a lod score of 6.62, which dramatically exceeded the conventional threshold for significant linkage (lod of 3) when analyzing monogenic traits. This study provided the first formal demonstration that disruption of a single locus could lead to a disorder of speech and language development. Furthermore, it represented an initial step toward identifying the precise causative mutation, narrowing the search to a small percentage of the genome.

The *SPCH1* interval identified by Fisher et al. (1998) was still likely to contain around a hundred or so genes. At the time this work was carried out, large-scale sequencing of the human genome was underway, but the available data were highly fragmented, not integrated and annotated to the extent they are today. Therefore, Lai et al. (2000) used a bioinformatic approach, exploiting sequence-similarity search tools and information from mapping studies, to assemble fragments into a sequence-based map of the *SPCH1* region, containing almost 8 million bases of complete nucleotide sequence. Additional *in silico* analyses allowed Lai et al. to position 20 known genes and 50 unknown transcripts precisely, with respect to this map. Given the full penetrance and lack of phenocopy in the KE family, it was possible to localize accurately the sites of meiotic recombinations, yielding unambiguous boundaries for the *SPCH1* region. Lai et al. developed novel polymorphic markers from 7q31 and genotyped these in the KE family, thereby narrowing the critical interval for *SPCH1*, eliminating more than 2.65 million bases, and excluding 4 known genes. They then began to sift through the most promising of the candidate genes remaining in the *SPCH1* region, searching for mutations in the DNA of affected individuals from the KE family.

Lai et al. also investigated new subjects, unrelated to the KE family, who had developmental verbal dyspraxia and language impairment. They identified a patient, known as CS, who had severe articulation difficulties accompanied by expressive and receptive language delay. The disorder of CS was associated with a chromosomal rearrangement involving reciprocal exchange of genetic material between the long arms of chromosomes 7 and 5. This translocation was not present in the patient's parents, and there was no family history of speech and language difficulties. The chromosome 7 breakpoint appeared to involve the *SPCH1* region of 7q31. In many cases of balanced translocation, breakpoints are located in regions of noncoding DNA and do not affect expression of neighboring genes, so that the individuals carrying them are phenotypically normal. However, in some subjects, a translocation may disrupt or alter expression of a critical gene, thereby leading to disorder, and the study of such cases can prove valuable for positional cloning efforts.

Lai et al. used fluorescence in situ hybridization, employing genomic fragments from their sequence-based map of the *SPCH1* interval, to identify the position of the 7q31 breakpoint in patient CS. They discovered that the breakpoint mapped very close to a partially characterized gene, called *CAGH44*, which was predicted to encode a protein with long stretches of consecutive polyglutamine residues. Expansion of polyglutamine repeats causes several neurodegenerative diseases, including Huntington disease and a number of hereditary ataxias (Cummings & Zoghbi 2000). This gene was thus an attractive candidate for *SPCH1*. It was apparent from initial bioinformatic analyses that the *CAGH44* transcript, which contained a few hundred coding nucleotides, represented just a portion of the gene of interest. Lai et al. (2001) investigated genomic sequence data around the CS translocation breakpoint using *in silico* gene-prediction methods to assemble the entire coding region of the *CAGH44* gene, which they then confirmed with laboratory-based analyses of messenger RNA (mRNA) (Figure 2). In addition to the polyglutamine region, the complete gene also encodes a DNA-binding motif, known as a forkhead/winged-helix domain (Kaufmann & Knöchel 1996). This observation placed the gene in the large (and rapidly growing) forkhead box (FOX) gene family, and it was therefore renamed *FOXP2* in accordance with recent nomenclature guidelines (Kaestner et al. 2000).

Disruption of *FOXP2* Leads to Severe Speech and Language Difficulties

Lai et al. (2001) demonstrated that the translocation breakpoint of patient CS directly disrupts one copy of *FOXP2*, between exons 3 and 4 (Figure 2). Sequencing of the entire *FOXP2* coding region in the KE family revealed a G-to-A nucleotide transition in exon 14 that cosegregated perfectly with the speech and language disorder in that pedigree. The mutation, which was not present in 364 control chromosomes from normal individuals, is predicted to cause a substitution of histidine for arginine at a critical residue of the DNA-binding motif of the *FOXP2* protein

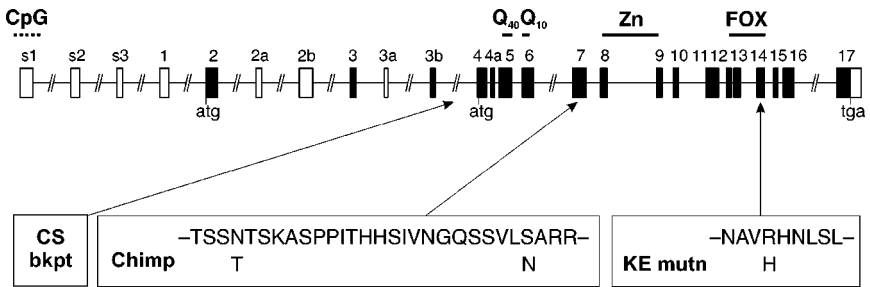


Figure 2 Schematic of the human *FOXP2* locus, spanning more than 600,000 bases of genomic DNA in 7q31. Boxes represent exons (present in processed mRNA). Lines represent introns (removed from mRNA by splicing). Black shading indicates those exons that code for protein. Positions of initiation (atg) and termination (tga) codons are shown, corresponding to start and end sites for protein translation, respectively. The main *FOXP2* protein isoform, encoded by exons 2–17, contains 715 amino acids, including two polyglutamine tracts of 40 (Q_{40}) and 10 residues (Q_{10}), a zinc-finger motif (Zn), and a forkhead domain (FOX). Exons 3b and 4a are alternatively spliced to give isoforms with small insertions of extra amino acids. Alternative splicing of noncoding exons 2a, 2b, and 3a is predicted to give a shorter protein isoform beginning in exon 4. Exon s1 overlaps with a CpG island and is likely to represent a promoter of *FOXP2* expression, but variation in splicing suggests that there is also an alternative promoter elsewhere. See Lai et al. (2001), Newbury et al. (2002), and Bruce & Margolis (2002) for full details of known splicing patterns. The CS translocation breakpoint lies between exons 3b and 4 (Lai et al. 2001). The point mutation in affected KE individuals maps within exon 14, altering the amino acid sequence of the FOX domain, as shown. Two amino acid differences, encoded by exon 7, separate the chimpanzee and human versions of *FOXP2* (Enard et al. 2002).

(Lai et al. 2001). Forkhead/winged-helix domains form a characteristic structure involving at least three alpha helices, followed by two large loops (or “wings”). The substitution occurs within the third alpha helix, adjacent to a residue that makes direct contact with the major groove of target DNA (Clark et al. 1993). Furthermore, an arginine is found at this point of the third alpha helix in every known FOX protein, in a wide range of organisms such as yeast, nematode, fruit fly, mouse, and man (Kaestner et al. 2000). A substitution mutation of the corresponding residue in one forkhead protein (FOXN1) causes an immunodeficient nude phenotype in mice (Schlake et al. 2000). In the case of FOXN1, studies of transfected mammalian cell lines have indicated that alteration of the critical arginine completely abolishes protein function (Schlake et al. 2000).

FOX genes encode transcription factors with diverse roles in cellular differentiation/proliferation, signal transduction, and pattern formation (Kaufmann & Knöchel 1996, Kaestner et al. 2000). Many members of this family are important regulators of embryogenesis, and several are implicated in developmental

disorders in mice and humans. These include *FOXC1* in glaucoma (Nishimura et al. 1998), *FOXE1* in thyroid agenesis (Clifton-Bligh et al. 1998), *FOXN1* and *FOXP3* in different forms of immune deficiency (Schlake et al. 2000, Wildin et al. 2001, Bennett et al. 2001), *FOXL2* in ovarian failure (Crisponi et al. 2001), and *FOXC2* in lymphedema syndromes (Finegold et al. 2001). In such cases, disease is often associated with amino acid substitutions in the forkhead domain (e.g., Nishimura et al. 1998, Clifton-Bligh et al. 1998, Schlake et al. 2000, Bennett et al. 2001) or alteration of FOX gene expression due to chromosomal rearrangement (Nishimura et al. 1998, Crisponi et al. 2001). Dosage of functional forkhead protein can be critical for certain developmental processes (Nishimura et al. 2001). In patient CS and affected members of the KE family, speech and language disorder is associated with disruption of only one copy of *FOXP2*; the other copy remains intact. Lai et al. (2001) suggested that in both cases insufficient functional *FOXP2* at a key stage of brain embryogenesis leads to abnormal development of neural structures that are important in speech and language acquisition. This hypothesis is supported by the observation that human *FOXP2* mRNA is strongly expressed in human fetal brain (Lai et al. 2001, Bruce & Margolis 2002). Initial studies of murine *FOXP2* indicate that the gene shows a restricted pattern of expression during development of the CNS (Shu et al. 2001). As is the case for other transcription factors, the *FOXP2* protein may play multiple roles during embryogenesis, since these mouse studies also showed expression in defined areas of developing lung, intestinal, and cardiovascular tissue (Shu et al. 2001). However, given that the KE and CS pathology is confined to the brain, it is likely that a single functional copy of *FOXP2* is sufficient for normal development of these other tissues.

MOLECULAR GENETICS OF COMMON FORMS OF LANGUAGE IMPAIRMENT

In the unique example of the KE pedigree, the simple inheritance pattern and large number of affected family members yielded substantial power for identifying the gene responsible. However, certain aspects of the KE and CS phenotypes distinguish their form of speech and language disorder from more common cases of developmental language impairment (often referred to as specific language impairment or SLI). As discussed previously in this article, there may be considerable heterogeneity in phenotypic profile among the latter, the validity of current schemes for subtyping remains a contentious issue, and there is a lack of consistency between studies in the operational definitions of disorder that they employ. Regardless, many researchers consider that the presence of developmental verbal dyspraxia in the KE family precludes a diagnosis of SLI. The observation of non-verbal deficits in some of the affected individuals has also been taken as evidence that the KE phenotype is not relevant for typical SLI. (Recall, however, the following points. First, the affected members of the KE family do have linguistic and grammatical difficulties, beyond their problems with speech motor control, which are rather similar to those displayed by SLI sufferers. Second, nonverbal deficits

are not central to the KE phenotype and the adoption of IQ-discrepancy-based diagnoses of SLI has often been called into question.)

Two recent studies have attempted to localize genes influencing language-related traits in families with typical SLI. Although both involved linkage mapping across the entire genome, the family structures and methods adopted for statistical analysis differed. In each case, novel loci were identified that might influence susceptibility to SLI.

Categorical Linkage Analysis of Extended Families Segregating SLI

Bartlett and coworkers (2002) studied five Canadian families of Celtic ancestry who had a history of language and reading impairment. The sample was a subset of families previously identified for a schizophrenia study, but only 7 of the 86 subjects displayed psychotic symptoms, and these individuals were coded as “unknown” phenotype for the Bartlett et al. investigation. Each family included at least two subjects who met strict diagnostic criteria for SLI: poor spoken language (based on scoring below a specific threshold on a standardized test of language development), a performance IQ exceeding 80, no evidence of hearing abnormalities, no deficits in oral motor function, and no comorbid diagnosis of autism, psychoses, or neurological disorder. Sixty-nine subjects were genotyped for 381 polymorphic markers, spanning the entire genome with an average spacing of ~ 9 cM. Linkage analyses were performed using three classifications of disorder—language impairment (spoken-language-quotient standard score of ≤ 85), reading impairment (1-SD discrepancy between performance IQ and nonword reading ability), and clinical impairment (poor performance on language or reading subtests, and/or history of at least 2 years of therapeutic intervention). These classifications were not mutually exclusive, yielding overlapping groups of affected subjects. Each phenotype was analyzed using traditional parametric linkage analysis under two models, one dominant and one recessive. Penetrance levels, phenocopy rates, and disease allele frequencies were set to correspond to a 7% prevalence rate of SLI in the general population. Bartlett et al. acknowledged that the parameters for these analyses were very unlikely to be correct. However, they cited previous studies indicating that the use of both dominant and recessive models with arbitrary penetrance levels can be a viable method for detecting linkage in complex diseases.

Regions yielding preliminary evidence of linkage were genotyped with additional markers in an expanded sample that included a further 17 individuals from the families. These analyses identified strong linkage of reading impairment to chromosome 13q21, under the recessive model, with a multipoint lod score of 3.92. To account for multiple testing of different phenotypes, genetic models, and markers, Bartlett et al. simulated 1500 genome-wide scans *in silico* under the assumption of no linkage. A lod score of 3.92 was observed in $<1\%$ of simulated scans, which suggests that this finding meets accepted guidelines for declaration of significant linkage. [Conventionally, a significant result is one that should occur

by chance in 1, or less, out of every 20 genome-wide scans conducted, that is to say in <5% of scans (Lander & Kruglyak 1995)]. Two other regions of potential interest were reported; 2p22 with a recessive language impairment model (multipoint lod = 2.79) and 17q23 with a dominant reading impairment model (multipoint lod = 2.19). Simulations showed that these results were expected to occur in 6% and 20% of genome-wide scans, respectively. No linkage was observed for the *SPCH1/FOXP2* interval on 7q31, although this may not be surprising, given that cases of verbal dyspraxia had been specifically excluded.

It is curious to note that while each of these families was selected for study on the basis of strictly defined SLI in at least two subjects, the only significant finding (that on 13q21) involved a reading-discrepancy phenotype. Linkage of 13q21 to the other categorical definitions of “language impairment” or “clinical impairment” was weak at best, regardless of inheritance model. This might suggest a counterintuitive conclusion: that genetic variants at a putative major SLI locus do not cosegregate with language impairment or with a broader phenotype involving problems with language and/or reading but do cosegregate with reading discrepancy in these five families. It is not standard practice to diagnose SLI on the basis of a subject’s reading performance. However, as discussed elsewhere in this article, there is evidence of comorbidity between language impairment and specific reading disability (dyslexia), which may partly reflect a common genetic aetiology (Bishop 2001b). Bartlett et al. argued that their reading-impairment phenotype assesses the reading outcome of an underlying language deficit, but they did not explain why they did not therefore see linkage of 13q21 to either language-based phenotypes or a broader definition of clinical impairment. This highlights the difficulty of interpreting results when conducting separate analyses of related phenotypic definitions in a single dataset. In this kind of situation, use of alternative diagnoses often leads to discrepancies in the observed evidence for linkage. It is not clear whether such discrepancies truly reflect the underlying genetic aetiology or are a consequence of differences in the sensitivity of psychometric tests, age distribution of affected subjects, and/or stochastic effects (Fisher et al. 1999).

Quantitative Trait Analyses in a Large Collection of Nuclear Families

A rather different approach for mapping loci influencing language impairment was taken by the SLI consortium (2002). The consortium identified 98 small nuclear families, each of which contained at least one child with SLI. Probands had receptive or expressive language skills that were ≥ 1.5 SD below the mean for their chronological age. Individuals were excluded from the study if they had a performance IQ below 80, deafness, chronic illness, autism, or a known neurological disorder. Families for this study came from two UK-based sources, a clinical sample collected at Guy’s Hospital and an epidemiological sample recruited by the Cambridge Language and Speech Project (CLASP). All available siblings in each family were assessed with a battery of language-related tests. Each family (473

individuals in total) was then genotyped for over 500 polymorphic markers across all chromosomes, yielding a final marker density of <8 cM.

Instead of employing categorical definitions of impairment to search for marker-trait linkage, the SLI consortium handled phenotypic complexity with a quantitative trait locus (QTL) mapping strategy. This involved direct genetic analyses of quantitative data from all siblings, regardless of whether they had a positive diagnosis of impairment. Three heritable measures were investigated: the receptive and expressive scales of a standardized clinical assessment battery and a test of nonword repetition, comparable to that assessed by the twin studies of Bishop et al. (1999). Linkage was evaluated at each point of the genome via two complementary methods for QTL mapping: Haseman-Elston regression (Haseman & Elston 1972) and variance-components analysis (Amos 1994). The basic Haseman-Elston method is a simple robust technique that assesses whether there is a significant correlation between the phenotypic similarity of siblings (indexed by comparison of their quantitative scores) and their genetic similarity at the chromosomal region under investigation (determined from genotype data). The variance-components method is more powerful but is computationally intense and relies on certain assumptions that are often violated in real datasets. For variance-components analysis, the full variability of a measure is partitioned into components owing to major-gene, unlinked polygenic, and residual environmental effects. At each chromosomal region, a statistical test compares likelihood of the data under the null hypothesis of no linkage (no major-gene effect) to that under the alternative hypothesis of linkage (where the major-gene component is unconstrained).

Two prominent areas of linkage stood out from the background of the rest of the genome—one on 16q24, the other on 19q13. The putative 16q24 QTL was linked to nonword repetition, with multipoint lod scores of 3.55 and 2.57 for Haseman-Elston and variance-components methods, respectively. To account for possible violations of the inherent assumptions made by these analytical methods, the SLI consortium derived empirical p-values from 100,000 single-marker simulations for the nonword repetition trait, run under the assumption of no linkage. These simulations indicated that the evidence supporting the 16q24 QTL was on the borderline of significant linkage (Lander & Kruglyak 1995), although no adjustment was made for analysis of three correlated traits. Evidence supporting a 19q13 locus was found for the expressive language measure. Although the multipoint lod scores on 19q13 were similar in strength to those on 16q24, with 3.55 for Haseman-Elston and 2.84 for variance-components, trait-specific simulations demonstrated that these results were not as significant and only qualified as suggestive linkage. When the SLI consortium dataset was separated into the two constituent samples of families (that is, the Guy's Hospital and CLASP collections) and reanalyzed for chromosomes 16 and 19, each sample was found to contribute equally to each putative QTL linkage. Given that the two groups had different origins—one a severe clinical sample, the other taken from an epidemiological study—the concordance of linkage results was particularly encouraging. No linkage was observed to the *SPCH1/FOXP2* region on 7q31.

As with the Bartlett et al. study, there was trait discordance at the peaks of linkage identified by the SLI consortium. The 16q24 locus was only detected with the nonword repetition task, whereas 19q13 appeared to be linked only to the measure of expressive language, despite the fact that the two traits were moderately correlated (0.538). Again, it is important to note that strength of linkage in such studies is not a reliable indicator of the trait-specific effect size of a locus, so one should be cautious about drawing conclusions from these discrepancies (Fisher et al. 1999). For QTL methods, it is now possible to perform simultaneous multivariate analysis of correlated traits, rather than conduct separate univariate analyses, to help clarify issues of trait specificity. Multivariate approaches have recently been applied to a genome-wide scan of reading-disability (Marlow et al. 2003) and should be applicable to the data from the SLI consortium study.

The *FOXP2* Gene and Common Forms of Language Impairment

As noted above, Lai et al. (2001) demonstrated that disruption of one copy of *FOXP2* causes the severe speech and language disorder observed in the KE family and in case CS. The exon-14 *FOXP2* mutation of family KE is fully penetrant (all who inherit it manifest the disorder), and there is no known history of language-related problems beyond the grandmother (individual I.2 on Figure 1). It is highly probable that this particular mutation arose in the germ cells of one of the grandmother's parents or at a very early stage of her fetal development. Similarly, the chromosomal rearrangement that disrupted *FOXP2* in subject CS was a de novo event. Thus, it is extremely unlikely that the specific mutation of the KE family or the translocation found in CS should themselves represent common risk alleles for language impairment, but it remains possible that other alterations in *FOXP2* may contribute to risk in a proportion of SLI cases. As such, for any study that seeks to evaluate whether *FOXP2* variants play a significant role in more typical SLI, it is important to systematically search the coding region for potential mutations.

Therefore, Newbury and coworkers (2002) conducted a mutation screen of all known *FOXP2* exons in 43 probands taken from the Guy's Hospital collection of the SLI consortium. Although no evidence had been found for 7q31 linkage in the QTL-based genome scan of this collection, the region had not been formally excluded (SLI consortium 2002). Newbury et al. (2002) identified a single case where there was a potential change in the coding region of *FOXP2*—an insertion of two CAG repeats within the second polyglutamine tract, preserving the reading frame—but this variant did not cosegregate with disorder in other members of the family. The polyglutamine regions of *FOXP2* are relatively stable, as compared to those implicated in neurodegenerative disorders, because they are encoded by a mixture of CAG and CAA triplets (Lai et al. 2001, Bruce & Margolis 2002), but they still have elevated levels of polymorphism (Enard et al. 2002). No other coding changes were detected, leading the authors to conclude that *FOXP2* coding variants are unlikely to make a major contribution to risk for common forms of language

impairment. Recently Bruce & Margolis (2002) reported identification of novel *FOXP2* isoforms, generated by alternative splicing of exons that were previously uncharacterized (Figure 2). The significance of these additional isoforms has not yet been addressed, but it will be necessary to screen all newly identified coding sequence to fully exclude *FOXP2* involvement in the SLI consortium collection.

Finally, it is possible that variation in *FOXP2* protein levels, rather than coding changes, might increase risk to SLI. Newbury et al. (2002) identified a CpG-rich region, likely to represent a promoter for *FOXP2* gene expression, and studied a polymorphic marker that mapped nearby (Figure 2). There was no evidence for any association between genetic variation at this marker and quantitative variability in language abilities in the SLI consortium families.

COMMON GENETIC AETIOLOGY FOR LANGUAGE-RELATED TRAITS?

Individuals who have speech and language difficulties at a young age are at increased risk of literacy failure when they begin to read, so that many go on to develop features of dyslexia, and it is possible that there are common genetic influences on deficits in language and literacy (Bishop 2001b). Molecular genetic investigations of developmental dyslexia have implicated loci on diverse chromosomes, including 2, 3, 6, 15, and 18, although no risk gene has been identified (reviewed by Fisher & DeFries 2002). Thus far, there is little evidence of overlap between the most significant regions implicated by genome-wide scans of speech and language disorders and the strongest findings from studies of dyslexia. The major 13q21 locus identified by analysis of the “reading impaired” phenotype in the Bartlett et al. (2002) SLI study is adjacent to a weak finding of linkage to 13q22 in a genome-wide scan of U.S. families with dyslexia (Fisher et al. 2002a). Bartlett et al. also reported suggestive linkage of SLI to 2p22, but this putative risk locus is quite distant from 2p15–16, a replicated region of linkage to dyslexia (Fagerheim et al. 1999, Fisher et al. 2002a). Overall, current linkage data remain inconclusive regarding the possibility that comorbidity between SLI and dyslexia might be genetically mediated.

Another heritable trait involving language deficits is autism—a neurodevelopmental disorder involving impairment of social interaction and communication, associated with repetitive and stereotyped behavior (APA 1994, WHO 1993). Many children with autism have severely limited speech output. Those who do develop normal linguistic abilities with respect to phonology and structure still retain deficits in pragmatics, the use of language in a social context (Tager-Flusberg et al. 2001). Autism and SLI are usually treated as distinct clinical entities; the language profiles of autistic children can vary substantially, and pragmatic impairments are seldom found in typical SLI. However, it has been proposed that these traits may lie within a spectrum of language-related disorders, involving overlapping sets of genetic risk factors (see Tager-Flusberg et al. 2001). Several genome-wide scans have been conducted for autism (reviewed by Folstein & Rosen-Sheidley 2001). The 13q21 locus identified by Bartlett et al. in their SLI sample is coincident with

a region that has been suggestively implicated in autism (Bradford et al. 2001). It has been reported that 13q21 linkage to autism is increased in families where the autistic proband has language delay and where there is a parental history of language-related difficulties (Bradford et al. 2001). There is also overlap between the 19q13 linkage found by the SLI consortium (2002) and a potential region of interest in autism, as identified by Liu et al. (2001). However, in the Liu et al. study, linkage to 19q13 appeared to be associated with a narrow diagnosis that would exclude children who had greatest overlap with SLI.

One of the most consistent findings of molecular studies of autism is linkage to a locus on 7q31, referred to as *AUTS1* (IMGSAC 2001), in a region that overlaps with the *SPCH1* interval identified in the KE family. Chromosomal abnormalities associated with autism have also been mapped to this area (see Folstein & Rosen-Sheidley 2001). As for the 13q21 locus discussed above, Bradford et al. (2001) found increased linkage to 7q31 when analyzing subgroups of autistic families with strong evidence of language deficits. Despite an absence of autistic features in the KE family and the presence of severe verbal dyspraxia, which is not usually associated with autism, it was suggested that *SPCH1* and *AUTS1* might be equivalent. Following the discovery that mutation of *FOXP2* underlies the *SPCH1* linkage, two independent groups evaluated the role of this gene in large numbers of families segregating autism, taken from collections that showed suggestive or strong linkage to *AUTS1* (Newbury et al. 2002, Wassink et al. 2002). Neither study found any evidence of association between polymorphic markers in introns of the *FOXP2* gene and risk of autism. Screening of the entire known *FOXP2* coding sequence in autistic probands identified two independent cases where a small number of glutamines had been deleted from the first polyglutamine tract (Wassink et al. 2002). No other nonconservative coding changes were identified in autistic probands from a total of 183 families (48 from Newbury et al., 135 from Wassink et al.), and both studies concluded that *FOXP2* variants are unlikely to make a significant contribution to genetic risk for autism.

DISSECTING NEUROLOGICAL MECHANISMS UNDERLYING SPEECH AND LANGUAGE DISORDERS

Following the genome-scans of Bartlett et al. (2002) and the SLI consortium (2002), considerable effort is now being invested into positional cloning of the putative risk genes that underlie each linkage. The lack of concordance between regions implicated by these two studies may be discouraging, although such disparity is often encountered for complex traits. Regardless, moving from initial observations of linkage to identification of aetiological gene variants is far from straightforward. Typically, linkage mapping implicates regions of several million base pairs at best. Furthermore, the breakdown of phenotype-genotype concordance usually observed in complex traits means that the boundaries of critical intervals will seldom be clearly established. A methodical and laborious mutation search through many candidate genes in a chromosomal region of interest may not be fruitful. Ultimately success will depend on converging information from a

variety of sources and approaches (Fisher 2002), including studies of chromosomal abnormalities and association-based methods (in which gene-trait association is evaluated at population, rather than family, level).

The eventual isolation of genetic factors that predispose to common forms of speech and language impairment may lead to improvements in diagnosis and treatment of such conditions. For example, it might be possible to develop genetic tests for identification of children who are at increased risk, facilitating earlier environmental intervention. In addition, functional investigations of risk genes will highlight key biological pathways that have gone awry to cause disorder and reveal novel insights into aetiology that may not be detected by other approaches. Although *FOXP2* is only involved in a rare severe form of disorder, it still provides an entry point into relevant molecular mechanisms and serves as an example of the future shape this field may take. Studies are underway to establish the expression patterns of *FOXP2* in early human embryogenesis and to identify the target genes that the transcription factor regulates during neuronal development, some of which may themselves be considered candidates for common forms of language impairment. The impact of disruption of *FOXP2* on mouse neurological development is also being investigated through the use of gene-targeting methods. The results from these investigations may complement data from brain imaging studies of individuals who have *FOXP2* mutations (Watkins et al. 2002b, Liegeois et al. 2002).

Finally, it is possible that molecular genetic studies of speech and language impairment might shed light on more fundamental questions such as the role of genes in the evolution of the human faculty for language. Recently, Enard et al. (2002) reported that the *FOXP2* protein is highly conserved, with only three amino acid substitutions between mouse and man, but two of those changes (both in exon 7) appear to have occurred on the human lineage after separation from the common ancestor with the chimpanzee. One of these human-specific changes may have functional consequences, creating a potential target site for phosphorylation. Enard et al. investigated the intraspecific sequence variation among humans for a large intronic segment adjacent to exon 7 and found a pattern of nucleotide polymorphism that strongly suggests *FOXP2* has been a target of selection in recent human history. The authors speculated that fixation of the human-specific changes occurred within the last 200,000 years, which raises the possibility that, at a point when spoken language was emerging, modifications in *FOXP2* led to improvements in vocal communication. It must be emphasized that this is likely to be just one of many genes involved in speech and language, and functional comparisons of the human and chimpanzee *FOXP2* proteins will be necessary to support the hypothesis. Nevertheless, the findings of the *FOXP2* studies demonstrate the future potential of this intriguing area of research.

ACKNOWLEDGMENTS

SEF is a Royal Society Research Fellow. APM is a Wellcome Trust Principal Research Fellow.

The Annual Review of Neuroscience is online at <http://neuro.annualreviews.org>

LITERATURE CITED

- Alcock KJ, Passingham RE, Watkins KE, Vargha-Khadem F. 2000. Oral dyspraxia in inherited speech and language impairment and acquired dysphasia. *Brain Lang.* 75:17–33
- Altmüller J, Palmer LJ, Fischer G, Scherb H, Wjst M. 2001. Genomewide scans of complex human diseases: True linkage is hard to find. *Am. J. Hum. Genet.* 69:936–50
- Amos CI. 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* 54:535–43
- APA. 1994. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. Washington, DC: Am. Psychiatr. Assoc.
- Bartlett CW, Flax JF, Logue MW, Vieland VJ, Bassett AS, et al. 2002. A major susceptibility locus for specific language impairment is located on 13q21. *Am. J. Hum. Genet.* 71:45–55
- Bennett CL, Christie J, Ramsdell F, Brunkow ME, Ferguson PJ, et al. 2001. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. *Nature Genet.* 27:20–21
- Bishop DVM. 1994. Is specific language impairment a valid diagnostic criteria? Genetic and psycholinguistic evidence. *Philos. Trans. R. Soc. Lond. B.* 346:105–11
- Bishop DVM. 2001a. Genetic and environmental risks for specific language impairment in children. *Philos. Trans. R. Soc. Lond. B.* 356:369–80
- Bishop DVM. 2001b. Genetic influences on language impairment and literacy problems in children. *J. Child Psychol. Psychiatry* 42:189–98
- Bishop DVM, Bishop SJ, Bright P, James C, Delaney T, Tallal P. 1999. Different origin of auditory and phonological processing problems in children with language impairment: evidence from a twin study. *J. Speech Lang. Hear. Res.* 42:155–68
- Bishop DVM, Brown BB, Robson J. 1990. The relationship between phoneme discrimination, speech production, and language comprehension in cerebral-palsied individuals. *J. Speech Hear. Res.* 33:210–19
- Bishop DVM, North T, Donlan C. 1995. Genetic basis for specific language impairment: evidence from a twin study. *Dev. Med. Child Neurol.* 37:56–71
- Bradford Y, Haines J, Hutcheson H, Gardiner M, Braun T, et al. 2001. Incorporating language phenotypes strengthens evidence of linkage to autism. *Am. J. Med. Genet.* 105:539–47
- Bruce HA, Margolis RL. 2002. FOXP2: novel exons, splice variants, and CAG repeat length stability. *Hum. Genet.* 111:136–44
- Clark KL, Halay ED, Lai E, Burley SK. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364:412–20
- Clifton-Bligh RJ, Wentworth JM, Heinz P, Crisp MS, John R, et al. 1998. Mutation of the gene encoding human TTF-2 associated with thyroid agenesis, cleft palate and choanal atresia. *Nature Genet.* 19:399–401
- Cole KN, Schwartz IS, Notari AR, Dale PS, Mills PE. 1995. Examination of the stability of two methods of defining specific language impairment. *Appl. Psycholinguist.* 16:103–23
- Collins FS. 1995. Positional cloning moves from perdictional to traditional. *Nature Genet.* 9:347–50
- Crisponi L, Deiana M, Loi A, Chiappe F, Uda M, et al. 2001. The putative fork-head transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. *Nature Genet.* 27:159–66
- Cummings CJ, Zoghbi HY. 2000. Fourteen and counting: unraveling trinucleotide repeat diseases. *Hum. Mol. Genet.* 9:909–16
- Dale PS, Simonoff E, Bishop DVM, Eley TC, Oliver B, et al. 1998. Genetic influence on

- language delay in two-year-old children. *Nature Neurosci.* 1:324–28
- DeFries JC, Fulker DW. 1985. Multiple regression analysis of twin data. *Behav. Genet.* 15:467–73
- Enard W, Przeworski M, Fisher SE, Lai CSL, Wiebe V, et al. 2002. Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* 418:869–72
- Fagerheim T, Raeymaekers P, Tonnessen FE, Pedersen M, Tranebjaerg L, Lubs HA. 1999. A new gene (*DYX3*) for dyslexia is located on chromosome 2. *J. Med. Genet.* 36:664–69
- Felsenfeld S, Plomin R. 1997. Epidemiological and offspring analyses of developmental speech disorders using data from the Colorado Adoption Project. *J. Speech. Lang. Hear. Res.* 40:778–91
- Finegold DN, Kimak MA, Lawrence EC, Levinson KL, Cherniske EM, et al. 2001. Truncating mutations in *FOXC2* cause multiple lymphedema syndromes. *Hum. Mol. Genet.* 10:1185–89
- Fisher SE. 2002. Isolation of the genetic factors underlying speech and language disorders. In *Behavioral Genetics in the Postgenomic Era*, ed. R Plomin, JC DeFries, IW Craig, P McGuffin, pp. 205–26. Washington, DC: APA Books
- Fisher SE, DeFries JC. 2002. Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nature Rev. Neurosci.* 3:767–80
- Fisher SE, Francks C, Marlow AJ, MacPhie IL, Newbury DF, et al. 2002a. Independent genome-wide scans identify a chromosome 18 quantitative-trait locus influencing dyslexia. *Nature Genet.* 30:86–91
- Fisher SE, Francks C, McCracken JT, McGough JT, Marlow AJ, et al. 2002b. A genome-wide scan for loci involved in attention-deficit/hyperactivity disorder. *Am. J. Hum. Genet.* 70:1183–96
- Fisher SE, Marlow AJ, Lamb J, Maestrini E, Williams DF, et al. 1999. A quantitative-trait locus on chromosome 6p influences different aspects of developmental dyslexia. *Am. J. Hum. Genet.* 64:146–56
- Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME. 1998. Localisation of a gene implicated in a severe speech and language disorder. *Nature Genet.* 18:168–70
- Folstein SE, Rosen-Sheidley B. 2001. Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Rev. Genet.* 2:943–55
- Gathercole SE, Willis CS, Baddeley AD, Emmslie H. 1994. The children's test of nonword repetition: a test of phonological working memory. *Memory* 2:103–27
- Gopnik M, Crago MB. 1991. Familial aggregation of a developmental language disorder. *Cognition* 39:1–50
- Gopnik M, Goad H. 1997. What underlies inflectional error patterns in genetic dysphasia? *J. Neurolinguist.* 10:109–37
- Haseman JK, Elston RC. 1972. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* 2:3–19
- Hurst JA, Baraitser M, Auger E, Graham F, Norell S. 1990. An extended family with a dominantly inherited speech disorder. *Dev. Med. Child Neurol.* 32:352–55
- Int. Mol. Genet. Stud. Autism Consort. (IMGSAC). 2001. Further characterization of the autism susceptibility locus *AUTS1* on chromosome 7q. *Hum. Mol. Genet.* 10:973–82
- Kaestner KH, Knöchel W, Martinez DE. 2000. Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev.* 14:142–46
- Kaufmann E, Knöchel W. 1996. Five years on the wings of fork head. *Mech. Dev.* 57:3–20
- Korstanje R, Paigen B. 2002. From QTL to gene: the harvest begins. *Nature Genet.* 31:235–36
- Lahey M, Edwards J. 1995. Specific language impairment: preliminary investigation of factors associated with family history and with patterns of language performance. *J. Speech Hear. Res.* 38:643–57
- Lai CSL, Fisher SE, Hurst JA, Levy ER, Hodgson S, et al. 2000. The *SPCH1* region on human 7q31: genomic characterization of the critical interval and localization of

- translocations associated with speech and language disorder. *Am. J. Hum. Genet.* 67: 357–68
- Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. 2001. A novel forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–23
- Lander E, Kruglyak L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet.* 11:241–47
- Lewis BA, Ekelman BL, Aram DM. 1989. A familial study of severe phonological disorders. *J. Speech Hear. Res.* 32:713–24
- Lewis BA, Thompson LA. 1992. A study of developmental speech and language disorders in twins. *J. Speech Hear. Res.* 35:1086–94
- Liegeois F, Connelly A, Baldeweg T, Gadian D, Vargha-Khadem F. 2002. *Functional abnormalities associated with the FOXP2 gene mutation in the KE family: a covert language study*. Presented at Int. Conf. Func. Map. Human Brain, 8th, Sendai, Japan
- Liu J, Nyholt DR, Magnussen P, Parano E, Pavone P, et al. 2001. A genomewide screen for autism susceptibility loci. *Am. J. Hum. Genet.* 69:327–40
- Marlow AJ, Fisher SE, Francks C, MacPhie IL, Cherny SS, et al. 2003. Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am. J. Hum. Genet.* 72:In press
- Neils J, Aram DM. 1986. Family history of children with developmental language disorders. *Percept. Motor. Skill.* 63:655–58
- Newbury DF, Bonora E, Lamb JA, Fisher SE, Lai CSL, et al. 2002. *FOXP2* is not a major susceptibility gene for autism or specific language impairment. *Am. J. Hum. Genet.* 70:1318–27
- Nishimura DY, Searby CC, Alward WL, Walton D, Craig JE, et al. 2001. A spectrum of FOXC1 mutations suggests gene dosage as a mechanism for developmental defects of the anterior chamber of the eye. *Am. J. Hum. Genet.* 68:364–72
- Nishimura DY, Swiderski RE, Alward WL, Searby CC, Patil SR, et al. 1998. The forkhead transcription factor gene FKHL7 is responsible for glaucoma phenotypes which map to 6p25. *Nature Genet.* 19:140–47
- Pinker S. 1994. *The Language Instinct*. London: Allen Lane
- Rutter M, Mawhood L. 1991. The long-term psychosocial sequelae of specific developmental disorders of speech and language. In *Biological Risk Factors For Psychosocial Disorders*, ed. M Rutter, P Casaer, pp. 233–59. Cambridge, UK: Univ. Cambridge Press
- Schlake T, Schorpp M, Maul-Pavicic A, Malashenko AM, Boehm T. 2000. Forkhead/winged-helix transcription factor Whn regulates hair keratin gene expression: molecular analysis of the nude skin phenotype. *Dev. Dyn.* 217:368–76
- Selkoe DJ, Podlisny MB. 2002. Deciphering the genetic basis of Alzheimer's disease. *Annu. Rev. Genom. Hum. Genet.* 3:67–99
- Shu W, Yang H, Zhang L, Lu MM, Morrissey EE. 2001. Characterization of a new subfamily of winged-helix/forkhead (Fox) genes that are expressed in the lung and act as transcriptional repressors. *J. Biol. Chem.* 276:27,488–97
- SLI Consort. 2002. A genomewide scan identifies two novel loci involved in specific language impairment. *Am. J. Hum. Genet.* 70:384–98
- Tager-Flusberg H, Joseph R, Folstein S. 2001. Current directions in research on autism. *Ment. Retard. Dev. Disabil. Res. Rev.* 7:21–29
- Tallal P, Piercy M. 1973. Defects of non-verbal auditory perception in children with developmental aphasia. *Nature* 241:468–69
- Tallal P, Ross R, Curtiss S. 1989. Familial aggregation in specific language impairment. *J. Speech Hear. Disord.* 54:167–73
- Tallal P, Townsend J, Curtiss S, Wulfeck B. 1991. Phenotypic profiles of language-impaired children based on genetic/family history. *Brain Lang.* 41:81–95
- Tomblin JB. 1989. Familial concentration of developmental language impairment. *J. Speech Hear. Disord.* 54:287–95
- Tomblin JB, Buckwalter PR. 1998. Heritability

- of poor language achievement among twins. *J. Speech Lang. Hear. Res.* 41:188–99
- Tomblin JB, Pandich J. 1999. Lessons from children with specific language impairment. *Trends Cogn. Sci.* 3:283–85
- Vargha-Khadem F, Watkins K, Alcock K, Fletcher P, Passingham R. 1995. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proc. Natl. Acad. Sci. USA* 92:930–33
- Vargha-Khadem F, Watkins KE, Price CJ, Ashburner J, Alcock KJ, et al. 1998. Neural basis of an inherited speech and language disorder. *Proc. Natl. Acad. Sci. USA* 95:12,695–700
- Wassink TH, Piven J, Vieland VJ, Pietila J, Goedken RJ, et al. 2002. Evaluation of FOXP2 as an autism susceptibility gene. *Am. J. Med. Genet.* 114:566–69
- Watkins KE, Dronkers NF, Vargha-Khadem F. 2002a. Behavioural analysis of an inherited speech and language disorder: comparison with acquired aphasia. *Brain* 125:452–64
- Watkins KE, Vargha-Khadem F, Ashburner J, Passingham RE, Connelly A, et al. 2002b. MRI analysis of an inherited speech and language disorder: structural brain abnormalities. *Brain* 125:465–78
- Wildin RS, Ramsdell F, Peake J, Faravelli F, Casanova JL, et al. 2001. X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. *Nature Genet.* 27:18–20
- WHO. 1993. The ICD-10 classification for mental and behavioural disorders: diagnostic criteria for research. Geneva, Switzerland: World Health Organ.