

Genetics of Childhood Disorders: XXXVII. News Flash: A Gene for Expressive Language

PAUL J. LOMBROSO, M.D.

The ability to speak is uniquely human. Although other species are capable of communicating with each other, the use of speech is a distinguishing feature of *Homo sapiens*. Is it possible that this trait is determined by specific genes? This question has been debated by linguists for many decades, and both sides of the argument have vocal supporters. The recent isolation of a gene that appears to be critical for the normal acquisition of language is a remarkable achievement and important to clinicians and researchers alike. The researchers' overall strategy is instructive because a similar strategy is likely to prove successful with other childhood psychiatric disorders including autism and Tourette syndrome. Each of the component steps in this

process have been reviewed in prior columns: identifying an extended family for study, using linkage or karyotyping techniques to pinpoint the chromosomal location of the gene, pulling out the mutated gene, and determining the normal function of the encoded protein.

Approximately 5% of children are significantly impaired in acquiring expressive and/or receptive language despite the absence of any detectable deficits in hearing, intelligence, or socioeconomic status. Twin studies have established that genetic factors play a significant role in certain developmental language disorders such as expressive language disorder without articulation difficulties, which is the most frequent expressive lan-

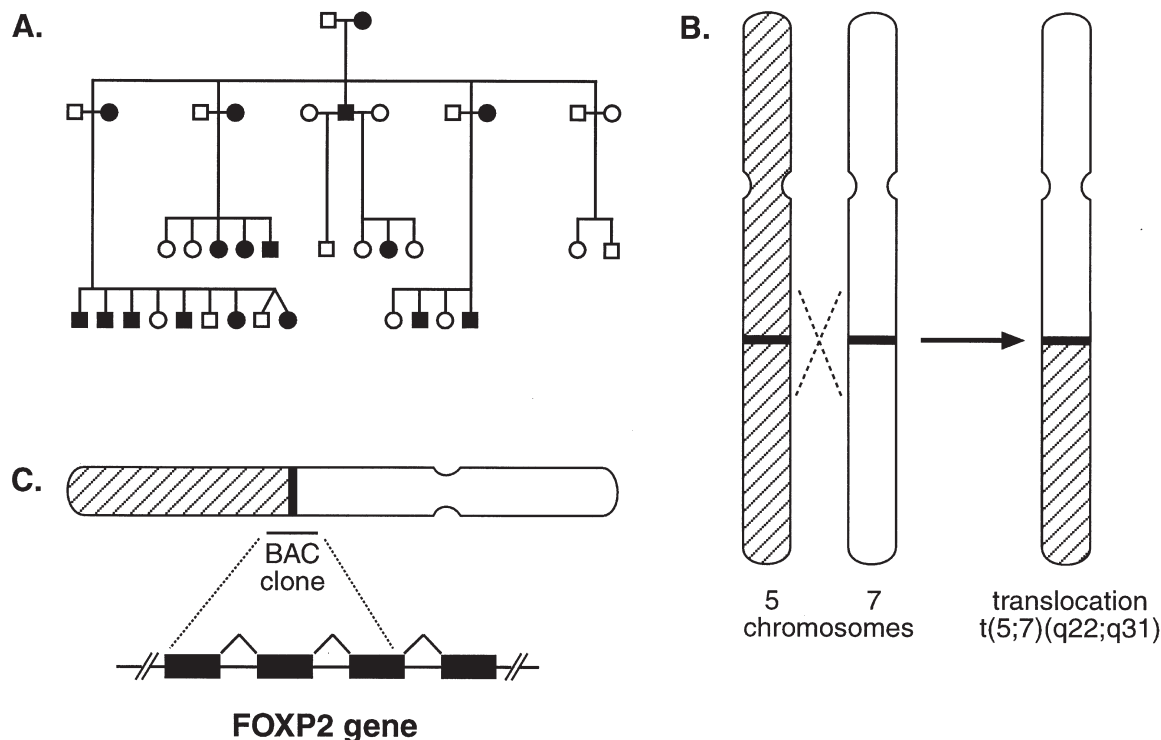


Fig. 1 Cloning a gene for language development. (A) Pedigree of the original extended family that have a dominantly inherited speech disorder. Linkage analyses allowed investigators to localize the gene to chromosome 7q31. (B) A second individual with a very similar language disorder was evaluated and found to have a visible chromosomal abnormality after karyotyping. A translocation had occurred between chromosomes 5 and 7, and researchers then attempted to fine-map the breakpoint. (C) The translocation on chromosome 7 was further characterized by isolating the bacterial artificial chromosome (BAC) that spanned the breakpoint. A portion of a gene (*FOXP2*) was found on the clone and sequenced. The full-length clone was isolated, and the coding sequence encoded a protein highly homologous to the forkhead family of transcription factors. It is thought that this gene is required during CNS development to establish normal neural pathways involved in the acquisition of language.

guage disability, and expressive language disorder with articulation difficulties. These disorders have significantly higher concordance rates in monozygotic twins than in dizygotic twins.

In 1990, researchers described a three-generation family containing members with a severe language disorder (Fig. 1). Initial characterization of the family showed that approximately half the family members were affected. There was an equal distribution among males and females, an inheritance pattern consistent with an autosomal dominant form of transmission.

The family was initially described as suffering from a very specific defect in the use of grammatical rules. However, continued characterization led to a broader phenotype not limited to difficulties in grammatical skills, but pervading virtually all facets of grammar and expressive language acquisition. Affected individuals suffered from deficits in several areas of language processing, including the ability to break up words into their constituent phonemes. There were also indications of cognitive impairments, with verbal components more profoundly depressed than performance scores. Moreover, all affected members had severe impairments in fine orofacial movements. The speech dyspraxia was so severe and disabling that the speech of a number of the subjects was unintelligible; many affected children were taught sign language to improve their ability to communicate. This profile indicates that the disorder disrupted intellectual, linguistic, and orofacial functions.

Functional and structural brain-imaging studies were then performed on some of the affected probands. The results suggest a key site of pathology in the basal ganglia. Bilateral reduction in the volume of the caudate nucleus was present in all affected family members studied. The authors speculate that these neural abnormalities arise early in development and persist into maturity, leading to the observed intellectual impairments and the orofacial dyspraxia. Functional abnormalities were also found in several motor-related areas of the frontal cortex that project to the affected neostriatal regions. It was suggested, however, that the abnormalities in the frontal regions were secondary to the primary bilateral lesions within the basal ganglia.

The availability of approximately 30 affected and unaffected family members allowed investigators to localize the gene responsible for the observed phenotype. Linkage analysis was performed with fluorescence-based typing techniques. Such techniques rely on the availability of microsatellite markers, which are DNA sequences evenly spaced throughout the human genome. The closer a marker is to a gene of interest, the less likely a recombination event will occur between that marker and the mutated gene. That is, there is a very high probability that DNA sequences on either side of the affected gene will be inherited together in affected family members. The markers need to be "informative," that is, one needs to be able to distinguish different variants between the various subjects. Only then can one determine whether one of these markers is tightly linked with the gene that causes the disorder.

A region of chromosome 7 (7q31) was implicated as the area of interest. Two closely linked markers gave maximal pairwise lod scores of approximately 6. A lod score of 3 means that the odds are 10 to the third power for the likelihood of linkage compared with odds of no linkage. Because of the logarithmic nature of lod scores, a lod score of 6 indicates that the odds for the likelihood of a gene being found at that site were now a million times greater than the odds for no linkage. The linkage analysis localized the gene to a region of chromosome 7 containing several million nucleotides.

At this point, a fortuitous event occurred that made the task of identifying and cloning the gene considerably easier. An unrelated patient was discovered with a very similar expressive language disorder. Moreover, this individual had a specific chromosomal rearrangement at 7q31, the same region implicated in the previous linkage analysis.

The chromosomal abnormality in this new individual was a translocation. Translocations are exchanges of genetic material between two chromosomes and, along with other chromosomal abnormalities such as microdeletions or inversions, are of considerable use to researchers as they often act as beacons to chromosomal regions in which mutated genes will be found. A well-known example of this is the Philadelphia chromosome, a translocation that occurs between chromosomes 9 and 22. The translocation is found in more than 90% of people with chronic myelogenous leukemia (CML). The breakpoints that disrupt each chromosome were finely mapped, and the genes that lie at each breakpoint were identified. As a consequence of the translocation, the gene for *ABL*, a tyrosine kinase, is translocated from its normal position on chromosome 9 to chromosome 22, where it is fused into the *BCR* gene. The functional consequence of this breakage and reconnection is the creation of a new gene containing a portion of both the *BCR* and *ABL* genes. The protein translated by this new gene is a novel, chimeric protein that contains the enzymatic portion of the protein tyrosine kinase but lacks its regulatory domain. Thus the protein is constitutively active and unable to be regulated. As the normal function of the *abl* protein is to regulate cell differentiation, cell division, and cell adhesion properties, the newly formed oncogene leads to the uncontrolled proliferation of white cells observed in CML.

A similar strategy was used to isolate and characterize the gene responsible for the language disorder described above. A first step was to identify any genes located at the translocation breakpoint. A team led by Anthony Monaco at the Wellcome Trust Center for Human Genetics at Oxford used fluorescence in situ hybridization (FISH) to map the 7q31 region.

To understand how this is done, one must first understand certain elements of the molecular technique used. Human genomic libraries contain all the sequences of DNA found on our chromosomes. The chromosomes are first cut into stretches of DNA much smaller than the original full complement of 3

billion nucleotides. The much smaller sections are easier to work with and manipulate. To be useful to the investigators, the DNA segments are inserted into other organisms that are capable of replicating not only their own DNA but also the inserted human fragments. When needed, it is relatively easy to separate the human DNA inserts from the other DNA present.

Considerable effort has gone into generating these genomic libraries. Human cells are often grown in the laboratory in cultures. Experimental tricks have been devised to isolate specific cell lines that contain only one or another human chromosome. Over time, each of the human chromosomes was isolated and cut into smaller portions. These smaller portions were placed into vectors to allow their replication by the host organism. In fact, these initial inserts were enormous (up to a million bases long) and could be replicated only after placing them in yeast. The resulting libraries were thus called yeast artificial chromosomes (YACs) and, although very useful, they still proved too large to easily isolate individual genes.

Libraries were then constructed in which the average genomic insert size was on the order of tens of thousands of bases. These smaller segments could now be inserted into bacteria instead of yeast. These libraries are called bacterial artificial chromosomes (BACs). Once again, the entire human genome was digested and the resulting component parts were placed into a suitable vector. Individual BAC clones are currently available, with each containing a relatively small piece of human genomic sequence. The position of the insert on the human genome has also been determined. Thus several BAC clones that were adjacent to the 7q31 region were labeled with fluorescent markers and allowed to hybridize with chromosomes isolated from the affected individual. The fluorescent label on the BAC clones allows microscopic visualization of the chromosomes to which the BAC clone binds. The BAC clone that spanned the breakpoint on chromosome 7 was identified in this way.

The researchers had thus reduced the search from billions of nucleotides on the full complement of human DNA to a more manageable tens of thousands of nucleotides present on the BAC clone. Moreover, because the sequence of each BAC clone is known, the researchers were now able to look at that sequence to determine whether any genes were present. This is now done through computer-assisted searches that identify particular patterns in the nucleotide sequence. Specifically, they look for long stretches of nucleotides that encode potential open-reading frames, that is, the sequence of nucleotides that can theoretically be translated into a protein. The researchers found one such open-reading frame on the isolated BAC clone and also discovered that the full sequence of the gene was not present. Additional sequences from adjacent BAC clones were obtained, and the longest open-reading frame that was present before encountering a STOP codon was generated.

This candidate gene was then translated into its predicted amino acid sequence. Comparisons were made between this

sequence and all known proteins, and highly homologous regions were sought. Such similarities between a known and an unknown protein provide hints about the potential function of the unknown protein. The carboxyl terminal of the protein sequence contained an 84-amino acid sequence that resembled a domain present in a group of transcription factors known as the forkhead family of proteins. The novel gene was designated *FOXP2* in accordance with the nomenclature standards for this rapidly growing family of transcription factors. The researchers found that the gene is expressed in fetal tissues, particularly the brain. Moreover, preliminary characterization of its expression pattern in mouse brain suggested that it could be found within the neopallial cortex and the developing cerebral hemispheres.

The forkhead gene family comprises transcription factors with a conserved 100-amino acid DNA-binding motif. It was originally identified in *Drosophila*, where mutations cause ectopic expression of head structures within the gut. Members of this family have been found in many species and are known to be key regulators of embryogenesis. Similar to all transcription factors, they function as master control genes that regulate the transcription of other genes required for normal development of the tissues in which they are found. A number of human disorders have already been found to be caused by mutations in specific *FOX* genes including congenital glaucoma and thyroid agenesis. Many of the mutations that have been found in these disorders are missense changes that result in the substitution of critical amino acids in functional domains of the transcription factor.

With this knowledge, the investigators went back to the original extended family to search for any mutations in the *FOXP2* gene. A single missense mutation that changed a guanine to an adenine was detected in exon 14 of the gene in all affected family members, and this mutation cosegregated perfectly with the language disorder. As it was possible that this change in nucleotide might be a normal variant within the population, 364 unrelated individuals were also screened. None of them had the observed change in sequence, indicating that it does not represent a naturally occurring polymorphism. Moreover, the functional consequence of the mutation was to change an arginine to a histidine within the critical DNA-binding motif. The amino acid arginine is found at that position in all forkhead proteins isolated to date and is believed to be required for either the binding of the transcription factor to DNA sequences at specific target genes or regulating transcription of those genes.

In conclusion, the isolation of a gene that affects the acquisition of expressive language is a remarkable achievement. It suggests that the transcription factor *FOXP2* is necessary for normal embryonic development of brain areas related to expressive language and that the genes that it regulates are required for the establishment of neural networks involved in this complex developmental milestone. It is likely that future work will

reveal many other genes, such as those regulated by *FOXP2*, that affect specific components of expressive language development. It should also be clear that none of this work was possible without the initial work and interest of astute clinicians.

WEB SITES OF INTEREST

<http://info.med.yale.edu/chldstudy/plomdevelop/development/april.html>
<http://info.med.yale.edu/chldstudy/plomdevelop/genetics/99julgen.htm>
<http://info.med.yale.edu/chldstudy/plomdevelop/genetics/99octgen.htm>
<http://info.med.yale.edu/chldstudy/plomdevelop/genetics/99sepgen.htm>

ADDITIONAL READINGS

Bishop DV, North T, Donlan C (1995), Genetic basis of specific language impairment: evidence from a twin study. *Dev Med Child Neurol* 37:56–71
 Fisher SE, Vargha-Khadem F, Watkins KE, Monaco AP, Pembrey ME (1998), Localization of a gene implicated in a severe speech and language disorder. *Nat Genet* 18:168–170
 Gopnik M, Crago MB (1991), Familial aggregation of a developmental language disorder. *Cognition* 39:1–50
 Hurst JA, Baraitser M, Auger E, Graham F, Norell S (1990), An extended family with a dominantly inherited speech disorder. *Dev Med Child Neurol* 32:352–355

Kaufman E, Knochel W (1996), Five years on the wings of fork head. *Mech Dev* 57:3–20
 Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP (2001), A fork-head-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–523
 Vargha-Khadem F, Watkins K, Alcock K, Fletcher P, Passingham R (1998), Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proc Natl Acad Sci U S A* 92:930–933
 Vargha-Khadem F, Watkins KE, Price CJ et al. (1998), Neural basis of an inherited speech and language disorder. *Proc Natl Acad Sci U S A* 95:12695–12700

Accepted October 18, 2001.

Dr. Lombroso is Associate Professor, Child Study Center, Yale University School of Medicine, New Haven, CT.

Correspondence to Dr. Lombroso, Child Study Center, 230 South Frontage Road, New Haven, CT 06520; e-mail: Paul.Lombroso@Yale.edu.

To read all the articles in this series, visit the Web site at <http://info.med.yale.edu/chldstudy/plomdevelop/>

0890-8567/02/4104-0482©2002 by the American Academy of Child and Adolescent Psychiatry.