Genetics of Childhood Disorders: IV. Linkage Analysis

DAVID PAULS, PH.D.

Over the past decade, investigators have isolated many genes in which mutations have been shown to be responsible for specific disorders. Genetic linkage studies, often among the first steps in these efforts, provide a powerful approach to help elucidate the underlying genetic mechanisms for inherited disorders. However, the application of genetic linkage studies to more complex psychiatric disorders has been less than satisfactory. The purpose of this column is to provide a brief overview of linkage methodology and to discuss some of the possible reasons why most initial linkage studies of complex disorders have not been successful.

Genetic linkage is designed to estimate the distance between genes. Normally, immediately before the gametes (sperm or eggs) are produced, there is a lining up of parental chromosomes in preparation for the separation of genetic material into gametes. An exchange of genetic material occurs between parental chromosomal pairs, which is termed *recombination*, or crossing over between chromosomes. The chromosomes are then separated and packaged into the gametes.

Two genes that lie on separate chromosomes will be transmitted independently of each other from parent to child. The child has an equal chance of receiving the gene from his

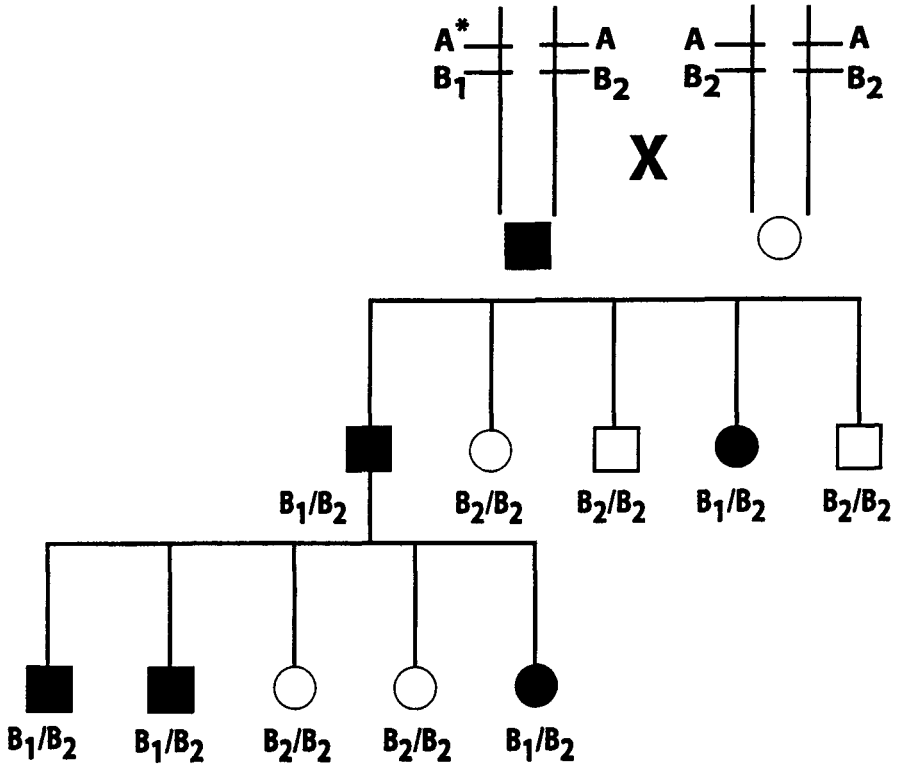

Fig. 1 Analyses of large pedigrees are useful for demonstrating linkage between a gene of interest and marker loci. In the example, the father has a gene (A*) that is mutated and causes a particular disorder. The abnormal gene lies on one chromosome, and a normal variant (A) of the gene is present on his second chromosome. The mother has the normal variant on both of her chromosomes. Close to this gene lies a sequence of DNA that serves as a marker. Two variants of the marker sequence have been discovered in the population (B_1 and B_2). The father has both variants, with the B_1 variant being very close to and tightly linked to the A gene. Therefore, all children who have the B_1 polymorphism will also have the A* gene. Figure adapted from Lombroso PJ, Pauls D, Leckman JF (1990), The importance of genetic factors in the etiology of child psychiatric disorders. In: *Brain and Behavior in Child Psychiatry,* Rothenberger A, ed. New York: Springer-Verlag, p. 116, Figure 3a.

mother or from his father. This phenomenon is encapsulated in Mendel's law of independent assortment.

What happens when the 2 genes are on the same chromosome? If they are located at opposite ends, then they will once again be transmitted independently of each other. This is because they are so far away from each other that a recombination event is very likely to occur between the 2 loci. However, the closer the 2 genes lie to each other, the less likely it is that a genetic crossover will occur between them. Finally, 2 genes may lie so close that it is much more likely that they will remain together and be transmitted together into the forming gamete. This is the violation of Mendel's law of independent assortment. Two examples should make this clearer.

If an individual has genotype A_1A_2 at locus A and genotype B_1B_2 at locus B and the loci are not linked to each other, the alleles at locus A and locus B will assort independently and 4 different types of gametes (A_1B_1, A_1B_2, A_2B_1, A_2B_2) will be produced in equal frequencies. This is termed *independent assortment.*

If locus A is very close to locus B on the same chromosome, an individual will again produce 4 types of gametes, but now the alleles found will not be in equal frequencies. The most common types of gametes will be those that represent the alleles that occurred in each parent. The less frequent types of gametes will contain a mixture of the parental alleles that has occurred as a result of infrequent recombination events between the 2 loci.

Historically, 2 approaches to linkage analyses were developed: (1) the allele sharing method and (2) the family pedigree method. Each of these methods has been extended over the past 2 decades and now includes methods that examine cosegregation of marker alleles and disease as well as allele sharing between 2 affected individuals that are either sibs or some other relative pair. The allele sharing method is a model-free procedure and does not require any assumptions about the nature of transmission involved in the disease. The family pedigree method examines cosegregation of marker alleles and disease, and it requires that the mode of inheritance of the trait under study be specified. The first approach is generally referred to as a *nonparametric* method and the second as a *parametric* one.

Because the mode of inheritance of any psychiatric disorder is not well understood, the allele sharing approach would seem to be the method of choice. However, allele sharing analyses are generally statistically less powerful than family pedigree analyses and until recently have not been used extensively in genetic linkage studies. It has become evident that the allele sharing approach can provide important preliminary evidence for linkage that can form the basis for more powerful techniques.

Early linkage studies in human genetics involved inherited disorders with known patterns of inheritance. The disease in question was clearly passed in an autosomal dominant or an autosomal recessive fashion, for example. The family pedigree approach became the more widely used analytic method to test for linkage between the disorder and a marker locus. That is, researchers asked whether the 2 genes traveled together within the pedigree or sorted independently. If they sorted independently, the research team moved on to the next DNA marker and repeated the analysis. Eventually they hoped to find a marker that was so close to the gene that caused the disorder that they no longer sorted independently, but were rather found together in affected members of the pedigree. All that was needed were markers evenly scattered throughout the chromosomes, and these are now available.

The family pedigree approach became even more widely used after the development of a computer program (LIPED) that facilitated the analysis of data from large families. The applicability of this approach to complex disorders was enhanced when LIPED was modified to allow for the incorporation of age- and sex-specific variables. The use of this approach has had limited success, however, in the localization of genes for complex psychiatric disorders.

The family pedigree parametric method involves the comparisons of likelihoods for specific genetic linkage hypotheses. First, the likelihood of observing a specific pattern of transmission in a pedigree is calculated assuming the null hypothesis of no genetic linkage to be true. That is, the likelihood of observing the distribution of the disease and a set of marker genotypes is calculated assuming independent assortment of the disease and marker alleles. Next, the likelihood of observing the pattern of disease and marker alleles for each of several alternative hypotheses of linkage is calculated and compared with the likelihood of the null hypothesis by means of an "odds ratio." The odds ratio consists of the likelihood of an alternative hypothesis divided by the likelihood of the null hypothesis. An odds ratio greater than 1,000 to 1 is taken as evidence for linkage, whereas an odds ratio of less than 1 in 100 is taken as evidence against linkage. For ease of comparison, the base 10 logarithm of the odds ratio is reported. In linkage analyses, these so-called "lod" scores ($\log_{10}$ [odds ratio]) are calculated for various hypotheses of linkage.

The alternative hypotheses of linkage are specified by different values of the recombination fraction. That is, the alternative hypotheses propose that (1) linkage is so close that no recombinations have occurred (i.e., $\theta = 0.0$); (2) linkage is tight, but a one-recombination event has occurred in 100 meioses ($\theta = 0.01$); (3) linkage is quite close, but relatively more recombinations have occurred ($\theta = 0.05$); and so on for as many increments of θ as the investigator specifies. The null hypothesis specifies that $\theta = 0.50$. A lod score greater than 3.0 ($\log_{10}$ [1,000/1]) is taken as evidence for linkage, and a lod score of less than –2.0 ($\log_{10}$ [1/100]) is taken as evidence against linkage.

The replicated demonstration of genetic linkage constitutes proof that a gene that confers susceptibility to some

illness exists. Thus, this approach has the potential to be extremely powerful in the search for genetic factors important for the expression of psychiatric illness.

However, there are limitations to the parametric approach. All of the calculations of likelihoods assume that the underlying genetic mechanisms for the illnesses being studied are known. The calculation of the lod scores for these families requires that the genetic model be specified. For psychiatric disorders, the underlying genetic mechanisms are not known. Furthermore, in those cases in which the patterns of transmission within families are consistent with a single gene, the estimate of penetrance is always less than 1.0. The ability to estimate correctly and thus detect linkage decreases dramatically when there is reduced penetrance. If penetrance is reduced from 1.0 to 0.8, 20% of the individuals with the susceptibility genotype will not manifest the disease. Because these individuals carry the genetic susceptibility (although they do not manifest the disorder), they will have the genetic marker associated with the illness.

Because the individuals are unaffected (and it is not possible to know what their disease genotype is), it is not possible to know whether they have the susceptibility genotype or whether there has been a crossover between the susceptibility locus and the marker locus. It follows, then, that when there is reduced penetrance, it becomes more difficult to estimate accurately the strength of linkage. Likewise, the misspecification of other parameters of the genetic model for the disease results in inaccurate estimates of the strength of the linkage relationship. Changes in gene frequency, penetrance estimates, and diagnostic criteria can significantly affect the results of linkage analyses. Alternative nonparametric analytic strategies have been and are being developed and will be discussed in the subsequent column.

WEB SITES OF INTEREST

http://www.ncbi.nlm.nih.gov/
http://linkage.rockefeller.edu/soft/list.html
http://www.dana.org/dana/manic2.html

ADDITIONAL READINGS

Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J (1986), Effects of misspecifying genetic parameters in lod score analysis. *Biometrics* 42:393–400

Lander E, Kruglyak L (1995), Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genet* 11:241–247

Morton NE (1956), Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277–318

Ott J (1991), *Analysis of Human Genetic Linkage.* Baltimore: Johns Hopkins University Press

Pauls DL (1994), Genetic linkage studies in psychiatry: strengths and weaknesses. In: *Genetic Studies in Affective Disorders: Overview of Basic Methods, Current Directions and Critical Research Issues* (Einstein/Montefiore Monograph Series in Clinical and Experimental Psychiatry), Papolos DF, Lachman HM, eds. New York: Brunner/Mazel, pp 91–104

Risch N, Botstein D (1996), A manic depressive history. *Nature Genet* 12:351–353

Accepted January 21, 1999.

Dr. Pauls is Professor of Psychiatry and Neurobehavioral Genetics, Child Study Center, Yale University School of Medicine, New Haven, CT.

Correspondence to Dr. Lombroso, Child Study Center, Yale University School of Medicine, 230 South Frontage Road, New Haven, CT 06520; e-mail: paul.lombroso@yale.edu.

To read all the columns in this series, visit the web site at: http://info.med.yale.edu/chldstdy/plomdevelop/