Strategies for Linkage Studies of Schizophrenia: Pedigrees, DNA Markers, and Statistical Analyses

by Scott R. Diehl and Kenneth S. Kendler

Abstract

A current overview of the genetics of schizophrenia is presented. Recently reported linkage studies of schizophrenia using polymorphic DNA markers are discussed and critically evaluated. Linkage studies of schizophrenia that we are currently undertaking are described, and the rationale underlying the specific strategies that we are using is explained. Our major focus is on extended pedigrees ascertained in Ireland containing a high density of schizophrenic cases, but we are also planning to use the technique of homozygosity mapping in an attempt to localize a recessively inherited gene that may be segregating in consanguineous families ascertained in Saudi Arabia. Issues relating to diagnostic criteria, the choice of which DNA markers to type with the highest priority, and appropriate statistical methodology are also considered.

With a few exceptions that are mostly attributable to poor methodology or small sample size, studies have consistently shown schizophrenia to aggregate strongly within families (Kendler 1988a). The results of twin and adoption studies have repeatedly indicated that genetic factors are responsible for most or all of this familial aggregation (Kendler 1988a). This finding has recently been confirmed again by results from the Provincial Danish Adoption study of schizophrenia (Kety 1987). In a recent thorough review of the epidemiology of schizophrenia, genetic factors were shown to be far and away the most important known risk factor (Eaton 1985). This is consistent with results based on a multifactorial threshold model in which the heritability of liability to schizophrenia is estimated to be around 0.70 (Rao et al. 1981; Kendler 1983). Family and adoption studies have also conclusively demonstrated that biological relatives of schizophrenic probands are at increased risk not only for classical schizophrenia, but also for a variety of schizophrenia-like psychotic conditions (e.g., schizoaffective disorder) and schizophrenia-like personality disorders (e.g., schizotypal personality disorder) (Kendler 1988a).

While evidence in favor of a genetic contribution to schizophrenia has been available for nearly 50 years, we are still ignorant about the mode of transmission of this genetic liability. Nearly all early investigators favored simple Mendelian models (e.g., Rüdin [1916]—two recessive loci, Kallmann [1938]—a simple recessive). In 1965, Falconer introduced the multifactorial threshold model, which was then applied to schizophrenia by Gottesman and Shields (1967) and gained wide acceptance. Innovations in analytic methods have recently sparked an increasing interest in more sophisticated single gene models for schizophrenia which, for example, incorporate incomplete penetrance and a proportion of phenocopies. Until recently, the major problem has been the lack of tools sufficiently powerful to discriminate these varying modes of transmission. Three major methods have

For definitions of technical terms, see glossary on p. 366.

Reprint requests should be sent to Dr. S.R. Diehl, Dept. of Psychiatry, Box 710, Medical College of Virginia, Richmond, VA 23298.
been applied in attempts to resolve the mode of transmission of schizophrenia: model fitting to morbid risk estimates, complex segregation analysis, and linkage analysis.

In general, morbid risk estimates alone have relatively little power to resolve mode of transmission (Smith 1971). For example, Kidd and Cavalli-Sforza (1973) found that risks of schizophrenia reported in the literature for various classes of relatives of schizophrenic patients could be explained as well by single gene and polygene models. Using risk estimates, O'Rourke et al. (1982) and McGue et al. (1985) have claimed they can rule out a single gene model for schizophrenia. These conclusions, however, can be questioned for several reasons including their (1) reliance on early family studies performed without operationalized criteria or blind diagnoses, (2) pooling across many potentially heterogeneous samples, (3) lack of correction for fitness effects, and (4) requirement of genetic homogeneity for all cases of schizophrenia.

Complex segregation analysis uses the pattern of phenotypes in families to estimate the likelihood of a given mode of transmission. Four major segregation analyses of schizophrenia have been published to date. Two of them (Elston et al. 1978; Tsuang et al. 1982) found evidence for vertical transmission of schizophrenia but strongly rejected a single gene model. Two others (Carter and Chung 1980; Risch and Baron 1984) also found strong evidence for vertical transmission of schizophrenia but found the pattern to be consistent with a single major gene. The significance of these findings is unclear. Two of the studies (Elston et al. 1978; Carter and Chung 1980) were based on family data sets that would not meet current methodological criteria. Only one of the four studies systematically assessed schizophrenia spectrum disorders, used a “mixed model” in the segregation analysis, or attempted any correction for reduced fertility in schizophrenia (Risch and Baron 1984). This study, however, was based on a small sample size of interviewed relatives (276). Furthermore, the power of segregation analysis to determine mode of transmission of “complex” human genetic disorders, especially in the face of possible genetic heterogeneity, remains unclear (e.g., Reich et al. 1981).

**Linkage Analyses**

Shortly after the rediscovery of Mendelian inheritance at the turn of the century, violations were observed to Mendel’s law of independent assortment. The term “crossover” was first used in the context of linkage effects on the segregation of genes for different characters by Morgan and Cattell (1912), and extensive linkage maps were constructed for many experimentally manipulable animals and plants. Until quite recently, however, human linkage analysis has been of very limited utility due to the severe paucity of polymorphic genetic markers. Developments in molecular biology over the past 10 years have now completely removed this limitation. At present, over 1,500 polymorphic DNA markers are available for human gene mapping studies (Pearson et al. 1987; K. Kidd, personal communication). In addition, the cloning and characterization of very highly polymorphic VNTR (variable number of tandem repeat) loci (Nakamura et al. 1987) are especially valuable, since the power to detect linkage in a particular set of available human families increases in direct proportion to the informativeness (i.e., the degree of polymorphism) of the markers used.

These DNA markers have already been applied successfully to map several major human genetic diseases where the physiological mechanism underlying the disease is unknown, an approach that has come to be known as “reverse genetics” (Orkin 1986). Examples of these successes include several neuropsychiatric disorders (Huntington’s disease, Wilson’s disease, and at least a subset of familial Alzheimer’s disease), as well as other important human genetic diseases such as chronic granulomatous disease, cystic fibrosis, muscular dystrophy, neurofibromatosis, and retinoblastoma (Watkins 1988). That this approach can be successful with psychiatric disorders not classically considered to be Mendelian is demonstrated by the evidence of linkage DNA markers on chromosome 11 to bipolar affective disorder in the Amish by Egeland et al. (1987). The acquisition of linked markers for some of these disease genes permits prenatal and presymptomatic testing. In the case of von Recklinghausen neurofibromatosis, linked markers have also provided evidence against previously suspected genetic heterogeneity of the disease, since virtually all families tested thus far have provided consistent evidence of linkage to markers located in the same region of chromosome 17 (e.g., Diehl et al. 1989). For several of the examples listed including muscular dystrophy (Hoffman et al. 1988), the linked markers have already facilitated the
cloning and characterization of the disease gene itself, thereby providing significant insights into the basic physiological processes underlying the disease.

To date, linkage studies of schizophrenia using conventional markers (i.e., polymorphisms detected at the protein level) have been few and relatively disappointing. In a sibpair analysis, Elston et al. (1973) reported possible linkage between schizophrenia and the group specific protein (Gc) and the immunoglobulin Gm locus. Turner (1979) found possible evidence for linkage between schizophrenia and HLA (human leukocyte antigen). However, neither McGuffin et al. (1983), in 12 multiplex families, nor Andrew et al. (1987), in 17 multiplex families, could replicate positive linkage to Gc or HLA or any other of the approximately 20 conventional markers used. Furthermore, two other studies (Chadda et al. 1986; Goldin et al. 1987) were unable to replicate Turner’s finding of linkage between schizophrenia and HLA. Given the relatively small sample sizes of all these studies and the small proportion of the human genome that was examined, these negative results should not be too surprising.

Two very recent findings now suggest that a major gene influencing liability to schizophrenia in some families may be located in the proximal region of the long arm of chromosome 5. First, Bassett et al. (1988) noted an uncle and nephew who both have schizophrenia and also trisomy of 5q11.2-13.3, with the trisomic portion of chromosome 5 translocated to chromosome 1. This observation alone could easily be attributed to coincidence, but Sherrington et al. (1988) have found strong evidence of linkage to DNA markers from this same region in families primarily from Iceland (but also including families from England that alone contribute a lod score of 2.0). However, Kennedy et al. (1988) have conducted linkage analyses on a set of seven related Swedish families using the same DNA markers used by Sherrington et al. (1988), and have obtained evidence excluding a major gene causing schizophrenia from this region of chromosome 5. A dominant mode of transmission and relatively high penetrance levels were used in both of these linkage analyses. One obvious conclusion that could be drawn from the discordant results of these two studies is that of genetic locus heterogeneity, but caution seems advisable in their interpretation until more extensive replications of these findings are reported. The glucocorticoid receptor, a potential schizophrenia candidate gene, was previously thought to reside in the region of proximal 5q implicated by these studies. However, Giuffra et al. (in press) have demonstrated by linkage analysis that this gene actually maps to distal 5q, probably more than 1 Morgan away from the region implicated by the partial trisomy/schizophrenia concordance and linkage data.

Two aspects of the results of these linkage studies are noteworthy. First, the penetrance of the dominantly inherited schizophrenia that is assumed and/or estimated is very high, compared to what might be expected based on other sources of information, such as concordance rates of monozygotic twins (Kendler 1988a). While it is certainly possible that these families may not be representative of all schizophrenic cases in the general population, it is then incorrect to estimate an allele frequency for the putative schizophrenia major gene at such a high level as to account for the overall population incidence of the disorder. The main problem is that no attempt was made in either of these studies to account for ascertainment bias that is highly likely due to preferential sampling of families with a high density of schizophrenic cases. Ignoring this bias could result in invalid assumptions of both dominant transmission and very high penetrance of the schizophrenia gene. This problem is clearly relevant to evaluating the true likelihood of exclusion of this region of chromosome 5 by the data of Kennedy et al. (1988). The data presented only exclude this region completely if relatively high penetrance is assumed. On the other hand, it is more difficult to imagine how the positive evidence of linkage presented by Sherrington et al. (1988) could result from incorrect assumptions about penetrance or mode of transmission, since incorrect assumptions of this sort are expected, on average, to reduce lod scores if a disease gene is actually linked to a marker being tested, and to have no systematic effect if the disease gene and marker are truly unlinked.

The second surprising aspect of the results of Sherrington et al. (1988) is the fact that substantially higher lod scores were obtained when psychiatric conditions (“fringe phenotypes”) that most previous studies indicate do not coaggregate in families with schizophrenia (Kendler 1988a) were categorized as “affected.” It was curious to note that the maximum likelihood estimates of penetrance obtained from a limited segregation analysis decrease from 86 percent to 73 percent when this heterogeneous collection of disorders is grouped together with schizophrenia (figure 2, Sherrington et al. 1988). We have
recently learned that the estimated penetrance values were erroneously reported by Sherrington et al. (1988) and Gurling et al. (1989). In fact, the highest penetrance estimated is obtained when the “fringe phenotypes” are grouped together with schizophrenia. But in any case, these small changes in estimated penetrance are minor compared to the associated increase of about 3 lod score units in support of linkage to the chromosome 5 markers.

Inspection of the segregation of marker and disease phenotypes in the pedigrees presented by Sherrington et al. (1988) indicates that about half of the potentially informative meioses are currently uninformative for assessing evidence of linkage, due to insufficient marker polymorphism. Therefore, typing of the half dozen additional polymorphic DNA markers that map to this region of chromosome 5q11-13 should be undertaken with the highest priority. The linkage information potentially available from the study of these presently uninformative meioses in the Icelandic and English pedigrees could either refute or strongly confirm the inference that a major gene influencing liability to schizophrenia has been genetically mapped in these families.

**Irish High-Density Extended Pedigrees**

Given the tentative evidence for genetic heterogeneity in schizophrenia with regard to chromosome 5 discussed above, it is important to note that the positive findings for linkage come from two countries: Iceland and England. Although culturally Nordic, the Icelandic population has a major Celtic genetic component. For example, Saugstad (1975) writes:

> The remarkable resemblance between Iceland and Ireland, in respect of several genetic markers (including Rh, PGM and Kell systems) is considered to be an expression of a similar proportion of people of Celtic origin in each of the two countries. Their identical, high incidence rates of PKU are regarded as further evidence of this.

Thus, the original Celtic population of Ireland has close genetic links with Iceland. However, historical and population genetic studies also demonstrate substantial evidence (more marked in the eastern and northern than western regions) that English (Anglo-Saxon) genes have been introduced into the original Celtic population of Ireland (Relethford 1983). Because of these genetic relationships among Irish, English, and Icelandic populations, the Irish families with a high density of schizophrenia to be studied in this proposal represent an ideal population in which to attempt to replicate the findings of Sherrington et al. (1988) based on Icelandic and English pedigrees.

Our studies in Ireland consist of two major components: a “family” study and a “high-density” study. The family study selects as probands all cases of schizophrenia (ICD 295) born after December 31, 1929, from the Roscommon County Case Register, in County Roscommon, Ireland, which has been in operation since 1972. When completed by November 1989, this sample will include approximately 290 schizophrenic probands diagnosed using ICD-8 (World Health Organization 1967) criteria and over 1,000 personally interviewed first-degree relatives. In addition, the family study includes both a matched normal control group (selected from the electoral register, n = 150) and their first-degree relatives and a psychiatric control group of affective disorder probands (n = 100) and their first-degree relatives. All traceable relatives living in Ireland or England will be contacted and interviewed face to face. As of September 9, 1988, 163 schizophrenic and 337 total probands have been interviewed, with a refusal rate of 8 percent. Of the total 2,248 identified relatives of probands, 1,472 have been interviewed to date and 94 (6 percent of those approached) have refused. To date, over 100 interrater reliability interviews have been conducted in the field. For the first year of the project, data are now available, and interrater reliability is good to excellent: (1) interview-based diagnoses (n = 33, Kappa = 1.00), (2) family history diagnoses (n = 115, Kappa = 0.85 ± 0.06), and (3) assessment of schizotypal symptoms and signs (mean of 16 scales, n = 33, mean intraclass correlation = 0.80).

The goal of the high-density study is to ascertain and study families with a high density of schizophrenia or schizophrenia-like disorders. Support has been obtained for conducting psychiatric interviews and establishing immortalized cell lines on 300 individuals from these families, and a proposal to continue and expand this ascertainment project is pending. While the majority of interviews in the family study have been conducted by trained interviewers with backgrounds in social work or psychiatric nursing, the high-density study has been largely conducted by Irish psychiatrists. It is easiest to describe the high-den-
sity study in three phases: (1) ascertainment, (2) screening, and (3) assessment and blood drawing. Ascertainment is based on the county psychiatric hospitals that serve catchment areas throughout Ireland and provide over 97 percent of the psychiatric care in the country. Research has shown that over 95 percent of individuals with schizophrenia in Ireland are known to treatment services. Ascertainment has two major steps. First, we approach the hospital medical director and administrator to explain the study and request their cooperation. Second, we meet with the medical and nursing staffs (including the community nurses who nearly all live in the areas they serve and hence have detailed “local knowledge” of these families) to explain the study and ask them to work together to develop lists of potential multiplex families.

Screening begins with this list of potential informative families and itself consists of two phases, diagnostic verification and investigation of the extended pedigree. Diagnostic verification is done, in almost all cases, by review and dictation of hospital case notes. Inclusion criteria for the high-density study require the family to contain at least two members (who are first-, second-, or third-degree relatives) with definite schizophrenia or schizoaffective disorder by DSM-III-R (American Psychiatric Association 1987). Schizoaffective disorder is included because of the strong evidence that, as defined by DSM-III-R, this disorder is, from a familial-genetic perspective, very closely related to classic schizophrenia (for review, see Kendler 1988a). If after the diagnostic review of each family by K.S. Kendler there is uncertainty about the diagnosis of a relative, then the relative in question is personally interviewed. This is done in about 15 percent of families. It should also be noted that in western Ireland, the use of stimulant and psychedelic drugs of abuse is virtually unknown, which considerably simplifies problems of diagnosis. Furthermore, contrary to cultural stereotype, rates of alcohol intake, cirrhosis, and alcoholism in Ireland are lower than those found in Western Europe and North America (2.3 percent in our pilot sample of relatives of controls [Brown 1978; see also Davies and Walsh 1983]). We have found only one case in the family study and none in the high-density study where the differential diagnosis between schizophrenia and alcoholic psychosis was problematic.

During the past year, we have been piloting the use of smooth pursuit eye movements, an attentional battery (including both the normal and attenuated stimulus continuous performance task [CPT]), and an extensive schizotypy self-report instrument in the family study based in County Roscommon. We plan to follow up a large sample of relatives of schizophrenics and controls previously interviewed in the family study and reassess them both by interview and by this “battery” of putative markers of the vulnerability to schizophrenia. As we now have been trained and have acquired experience with these assessment methods, we are also actively considering applying them to high-density families as well. Certainly, if we find evidence for linkage in any of our high-density families, we will move quickly to apply these methods to all cooperative relatives in an attempt to clarify further the phenotypic effect of the putative “schizophrenia” allele: More sensitive diagnostic evaluation, including these new measurements, might increase the penetrance of the schizophrenia allele thereby increasing the power of linkage analyses.

In the large majority of instances, the affected relatives identified through the hospital and nursing staff are first- or second-degree relatives. Usually, the affected members of the family are poor informants about other members of their family. Therefore, we routinely take an extended family history, which results in a pedigree being drawn out to at least third-degree relatives (i.e., great-grandparents, grand-uncles, and first cousins), from one or more of the healthy family members.

The hospital dictations of case notes and pedigree sketches are then reviewed by K.S. Kendler, and the decision is made whether to study the family and, if so, which members should be interviewed and sampled. The interview protocol is identical to that used in the family study and includes the major sections of the Structured Clinical Interview for DSM-II (SCID; Spitzer and Williams 1985) as well as a section specifically focused on schizotypal signs and symptoms. We have shown, in pilot data, that this section successfully detects relatives of schizophrenic cases versus controls. In addition, for individuals with psychotic illness, a special appendix is used which contains information on social function, negative symptoms, course of illness, precipitants, and premorbid functioning.

Our current sampling strategy, in brief, involves conducting diagnostic interviews and establishing immortalized cell lines for all individuals in a family where
preliminary information indicates a high likelihood of either definite schizophrenia or schizoaffective disorder. Cell lines are established, and diagnostic evaluations are conducted (including face-to-face interviews) for all available first-degree relatives of the affected family members. If unaffected parents of affected individuals are deceased or otherwise unavailable for sampling, an attempt is made to establish cell lines and obtain face-to-face interviews from the deceased parent’s siblings to improve the chances of inferring the genotype of the deceased parent. The ascertainment stage has been completed in eight hospitals and initially suggested 197 possible families. Further screening has identified 70 of these families (36 percent) as being potentially appropriate for linkage studies.

**Homozygosity Mapping in Consanguineous Families**

The second sources of families segregating for schizophrenia that we are planning to use for linkage studies involve first- or second-cousin marriages. We are planning to ascertain these pedigrees in Saudi Arabia in collaboration with K. Chaleby (Head, Section of Psychiatry, King Faisal Specialist Hospital & Research Center, Riyadh). As noted above (and in contrast to the finding of Sherrington et al. [1988]), recessive models of the genetics of schizophrenia have long attracted attention, from the early speculations of Rüdin (1916) and Kallmann (1938) to the complex segregation analysis of Risch and Baron (1984). Recent studies of Saudi Arabian consanguineous families similarly suggest possible recessive transmission of a major gene for schizophrenia (Chaleby and Tuma 1987). The availability of such families represents a valuable resource suitable for applying the linkage analysis technique of homozygosity mapping, as recently elaborated by Lander and Botstein (1987). This method is limited to recessively inherited disorders and requires consanguineous families, but it is generally much more powerful than “traditional” linkage analysis (in terms of numbers of individuals that need to be genotyped) for detecting both linkage to a major gene and heterogeneity among different families that might be segregating for different major genes. While the power calculations of Lander and Botstein (1987) assume complete recessivity (which is highly unlikely for schizophrenia), the homozygosity mapping method may still retain a substantial power advantage even in the case of only partially recessive transmission.

### Which DNA Markers to Use?

For reasons discussed above, DNA markers from chromosome 5 are being typed with the highest priority. Markers from chromosome 11p will also be typed early in this study, because of the report of linkage to another psychiatric disorder, bipolar affective disorder (Egeland et al. 1987). In addition to the partial trisomy of 5q noted above, other cytogenetic studies weakly implicate sites in chromosomes 2, 3, and 19, and for a subgroup of cases, the X chromosome (often involving “fragile sites”) as possibly offering a better than average prospect for finding a major gene influencing schizophrenia (DeLisi et al. 1988). DNA probes will, therefore, be typed from these regions before typing markers in other areas without any a priori interest. Although present evidence for “candidate genes” for schizophrenia appears weak (Propping and Friedl 1988), it is still logical to type either (1) markers that code for a protein known to be involved in some way with the nervous system or (2) anonymous DNA markers (especially those that are highly polymorphic) that are known to map very closely to genes involved in some way with the nervous system. Examples of such “candidate genes” for which RFLPs (restriction fragment length polymorphisms) have been identified include neuron peptide Y (Detera-Wadleigh et al. 1987), the proopiomelanocortin gene (Feder et al. 1985), the nerve growth factor receptor (Breakfield et al. 1986), the myelin basic protein (Kamholz et al. 1987), phenylalanine hydroxylase (Lichter-Konecki et al. 1988), and tyrosine hydroxylase (Moss et al. 1986). The recent cloning of a rat D₂ dopamine receptor (Bunzow et al. 1988) is also of special interest. A small portion of our total effort will be invested in developing new RFLPs or improving the informativeness of probes that are already polymorphic, if the necessary DNA clones can be obtained. However, if none of these “preferred” markers provide evidence of linkage to a major schizophrenia, a broader search will be undertaken toward the goal of eventually excluding the entire genome.

### High Resolution Mapping of a “Schizophrenia Gene”

If our project succeeds in achieving its major goal of finding strong evidence of linkage to a major schizophrenia gene in a significant
portion of the Irish high-density families or the Saudi-Arabian consanguineous families, its focus would immediately change. Instead of continuing to screen additional markers located elsewhere in the genome, in hope of perhaps finding a second major schizophrenia gene, research efforts would be largely redirected toward fine resolution mapping of the region of that part of a chromosome where the major schizophrenia gene appeared to be located. First, all DNA markers (and perhaps also some polymorphic protein-level markers too) from the candidate region would be screened in all families showing evidence of the major gene. Multipoint mapping analyses would be applied in an attempt to order the loci in the area, both with respect to the schizophrenia gene and also with respect to each other. If the order of the markers and their positions flanking the schizophrenia gene could not be determined with a high degree of certainty, the markers would be typed in reference families such as the CEPH (Centre d'Etude du Polymorphisme Humain) pedigrees (if not already typed) or other available pedigrees. Additional information obtained from typing such reference families could significantly increase the knowledge of the genetic map of the region of the chromosome of interest and thus indirectly increase the power of the multipoint analysis as applied to mapping the schizophrenia gene with respect to these marker loci.

Efforts would simultaneously be undertaken to clone additional DNA markers from the region of interest. A variety of methods are currently available to enrich clones obtained for specific regions of the genome, but significant further advances are occurring so regularly that the following suggested approaches should be viewed as tentative. The most readily available sources of clones enriched with DNA from a particular chromosome are the “chromosome specific” DNA libraries. These libraries have been constructed by flow-sorting human chromosomes derived from rodent-human somatic cell hybrids containing a limited number of human chromosomes. However, even if the library contained only DNA from a single human chromosome (and these libraries are usually not that pure), the size of most chromosomes (tens of millions of bases of DNA) still limits the chance that a randomly chosen clone would map into the region of interest (a target size of perhaps less than 5 million base pairs).

One method offering higher precision is that of somatic cell genetics. Currently available techniques permit the construction of somatic hybrid cells that contain only about 5–10 million base pairs of human DNA in a rodent background. Genetic libraries can be made from these hybrid cell lines, and clones containing human DNA can be detected by probing the library with human-specific repetitive DNA. This method has recently been used to generate a large number of new clones in the vicinity of the neurofibromatosis type 1 (NF1) gene in order to map its location more precisely and to clone this disease gene (O’Connell et al. 1989).

Another possible relevant technique is the cloning of “jumping” and “linking” libraries (Collins et al. 1987; Smith et al. 1987). Briefly, each “jumping” clone contains two pieces of DNA that are located from 100 to 500 kilobase (kb) pairs apart in the genome. With the exception of newly developed “yeast artificial chromosome” vectors (Burke et al. 1987) that are still not technically practical for routine use, available vectors for cloning human DNA are currently limited to insert sizes of 40 kb at most. This is one reason why “jumping” clones can be of special value. For example, if some DNA marker is mapped by linkage analysis to within a million bases from a disease gene (roughly 1 centimorgan [cM]), by finding the “jumping” clone that contains the linked marker, one obtains a second piece of DNA that is located from 100 to 500 kb from the original marker. Thus, it is possible to obtain additional markers that may be closer to the actual disease gene much faster than if one attempted to clone all the DNA between the marker and the disease gene by more standard “chromosome walking” techniques, where progress is limited to an average of about only 20 kb per step. “Linking” clones are used to join together adjacent “jumping” clones, allowing the process to be continued again.

As additional markers are cloned, these would be used to enhance the resolution of the genetic map surrounding the putative schizophrenia disease gene. Mapping of new clones would best be accomplished by a combination of genetic mapping in both high-density schizophrenia families and reference families, and physical mapping carried out at several levels. A rough estimate of the physical location of new clones could be obtained by using a panel of somatic cell hybrids, each of which contains a different portion of the human chromosome of interest. These can be obtained from naturally occurring translocations and by the somatic cell methods described above. A finer resolution physical map can be obtained by pulse field electrophoresis tech-
niques (Smith et al. 1987) that are capable of resolving DNA molecules differing in the size range of up to millions of base pairs, taking advantage of restriction enzymes that recognize sites that occur very rarely in human DNA and cut it precisely into very large fragments. DNA markers can be physically mapped by this approach by inferring that if two DNA probes consistently hybridize to the same set of DNA fragments, their location in the genome can be no greater than the size of the smallest fragment they share. Examples of both of these physical mapping methods can be seen in their current application aimed at more precisely localizing the NF1 gene in 17q11.2 (Fountain et al. 1989; O’Connell et al. 1989).

Possession of a set of DNA markers known to reside quite close to a major schizophrenia gene would permit screening for chromosomal rearrangements in the area of interest with very high precision. Very small deletions or other rearrangements, far smaller than those detectable with the highest resolution cytogenetic techniques, could be searched for by Southern blotting of standard and pulsed field DNA gels. It is possible that some of these rearrangements have actually caused mutations in the schizophrenia gene and thereby caused its malfunction leading to the development of this disease. Thus, the identification of such rearrangements could greatly facilitate the ultimate goal of the reverse genetic approach, which is the cloning and characterization of the disease gene. Rearrangements have been an essential tool used for the successful cloning of several disease genes such as muscular dystrophy and retinoblastoma (Orkin 1986).

**Statistical Analyses**

There remains considerable uncertainty about the best possible conceptual and statistical approach to the linkage analysis of complex phenotypes such as schizophrenia. We will restrict ourselves to a general description of our approach to data analysis.

**Segregation Analyses.** One of the major problems in pedigree-based linkage analysis is that such analyses require knowledge of the allele frequencies and penetrance vector at the putative disease locus. Many investigators deal with this problem by “searching” in their linkage analysis over a wide array of values of penetrance and, less thoroughly, allele frequency. Due to the post hoc aspect of choosing penetrances that produce the best evidence for linkage, it is clear that if such a “search” strategy is used, the traditional lod score criterion of 3 is no longer valid. A related problem is that the optimal phenotype for defining “schizophrenia” in linkage analysis is unknown. Again, investigators traditionally try multiple definitions, but this approach is not without its risks.

A major advantage of the data collection strategy of the Irish family study is that, as outlined above, it includes both the ascertainment of multiplex families and a case-controlled, large sample, epidemiologically based family study of schizophrenia. Morbidity risk based analyses can inform us what specific phenotypes aggregate in families of schizophrenic patients in an Irish population. Our sample size will be sufficiently large to develop and test criteria for schizophrenia-related personality disorder. The family study can, therefore, provide for linkage analysis, an a priori definition of the affected phenotype obtained from the same population.

We also intend to apply a variety of complex segregation analysis methods to the family study data. Previous power analyses (e.g., Reich et al. 1981) indicate that our sample size should be large enough to detect single gene variation in a discontinuous trait, even given some “polygene background.” These analyses will include the standard “unified mixed model” approach with or without multiple thresholds as well as more experimental methods that will (1) include putative markers of environmental etiology (e.g., season of birth, birth complications, and precipitating factors) (Eaves 1984), (2) correct for fitness effects, or (3) deal with correlated ages of onset in affected relatives. In addition, we will attempt to detect individual families that have a high probability that a single major gene may be segregating (e.g., Moll et al. 1984).

Parameter estimates that will be informative for linkage analysis can be derived from the segregation analysis of the large family study in three ways. First, we can constrain the model over all pedigrees to a single Mendelian locus and require it only to estimate allele frequency and penetrances. Second, we can repeat this analysis only in the pedigrees shown to have a high probability of having a major gene segregating. Third, we can allow the model more freedom (e.g., polygene background) and test for (a) the presence of single gene variation and (b) the allele frequency and penetrance vector for the best fitting model (if it contains a major gene) in all pedigrees. Clearly, we will be in the best position if the
estimates from these different approaches converge. However, per K. Kidd (personal communication), even when segregation analysis cannot detect unambiguous evidence for a major locus, the first or second approach can still be useful for generating a penetrance vector for linkage analysis.

Although we consider the availability of our large family data set to be an important advantage in our analytic approach to linkage analysis, it is not without problems alluded to above in our discussion of the study reported by Sherrington et al. (1988). Particularly, the high-density families that will be genotyped for the linkage analyses proposed here may not be representative of the general population of schizophrenic patients. It is possible that the penetrance of a putative schizophrenia gene may differ (probably be higher) in high-density families than in an epidemiologically representative sample. Nonetheless, the large family data set will provide a rational starting point for estimates of allele frequency and penetrance. As noted below, segregation analyses of the high-density families themselves may also provide information about allele frequency and penetrances.

**Linkage Analyses.** Given our uncertainty about the most appropriate and powerful method with which to approach the linkage analysis of schizophrenia and related phenotypes, we intend to apply the three current major methods to detect linkage in human populations: (1) affected sib-pair methods of testing for association and their extensions, (2) single marker “classic” pedigree-based methods, and (3) multipoint markers in pedigrees.

The affected sib-pair method has long been available to test for departures from independent segregation of disease and marker phenotypes without any required assumption about the mode of inheritance (Penrose 1935). This method has recently been expanded in several potentially useful ways. First, Suarez et al. (1982) and Suarez and Van Eerdewegh (1984) have shown how the method can be expanded to consider higher numbers of affected siblings. As might be expected, sibships with more than two affected (which are found not infrequently in our Irish high-density pedigrees) can be quite powerful at detecting linkage. Second, Weeks and Lange (1988) have shown how this method may be extended to other nonsibling affected relatives by an “affected-pedigree-member” method of linkage analysis based on observations of “identity by state” at marker loci. Likewise, Bishop and Williamson (1988) have developed affected-pedigree-member statistical methods that rely on inference of “identity by descent” and a likelihood ratio test on the recombination fraction to replace the $\chi^2$ test that may be invalid for some sample sizes. Lastly, Risch (1988a) has developed methods for the use of multiple markers, and this multipoint approach allows the precise localization of the disease locus.

As noted above, classic pedigree-based methods of linkage analysis require knowledge of allele frequency and mode of transmission of both the disease locus and markers. It is worth noting, however, that simulation studies have suggested that detection of a major gene influencing a phenotype like schizophrenia will be feasible even without highly accurate estimates of genetic parameters (Goldin et al. 1984; Cox et al. 1988). Goldin et al. (1984) demonstrated that in some cases where a major locus could not be detected by segregation analysis, linkage to a close marker locus could still be detected. In the study by Cox et al. (1988), it was found that linkage was likely to be detected, if several different models were examined, even without knowledge of the mode of genetic transmission. Misspecification of the mode of transmission does lead to inaccurate estimates of the recombination fraction between the marker and disease loci. However, this problem is probably of minor concern (and could be addressed in subsequent studies) compared to the immense value of simply demonstrating the existence of a major gene for schizophrenia. We will conduct classic pairwise linkage analyses using the program LIPE (Ott 1974) which calculates the well-known lod score (Morton 1955). Pairwise lod scores assessing linkage between each marker and a putative schizophrenia disease locus will be combined to produce an exclusion map (Edwards 1987). This method of generating an exclusion map provides an assessment of the relative probabilities of different locations in the genome for the major disease locus (subject to the assumptions of the original lod score analysis), and may under some circumstances complement multipoint analyses.

We will also use the more powerful techniques of multipoint linkage analysis (e.g., Leppert et al. 1987) that take advantage of the recent availability of a large number of highly polymorphic DNA markers dispersed throughout most of the human genome (Pearson et al. 1987). These methods assess
whether a group of markers that are linked to each other cosegregate with a disease locus. Some of the markers in the linkage group may not be informative in particular pedigrees or in parts of a single pedigree. For this reason, considerably more information is obtained by the simultaneous assessment of the cosegregation of many linked markers, because the odds are much higher that at least one marker on both sides of the disease locus will be informative for all potentially informative meioses. Multipoint linkage analyses will be performed by calculating location scores using the programs LINKAGE (Lathrop et al. 1984) and MENDEL (Lange et al. 1988). As with pairwise analyses, if enough markers that are evenly spaced throughout the genome are tested and consistently negative results are obtained, the hypothesis that a single major locus influences susceptibility to schizophrenia could be excluded (subject to the assumptions underlying the multipoint analyses). Additional multipoint methods are under development that even more fully exploit the power of an existing linkage map of DNA markers, and these will be used in our data analysis as they become available. Principal examples of these new methods are the strategies of “interval mapping” and “simultaneous search” currently under development by Lander and Botstein (1986).

Thus, our plan is to use the affected-pedigree-member method of testing for association, and both pairwise and multipoint linkage analysis methods for estimating recombination fractions and map location relating DNA markers to any detectable major locus for schizophrenia. The affected-pedigree-member method would be favored because it does not require assumptions about the mode of transmission of the disease, but it is almost certainly less powerful than pairwise linkage analysis methods that do require such assumptions (Weeks and Lange 1988). Until recently (Risch 1988a), the affected-pedigree-member method was limited to testing for association between a single marker and a putative disease locus on a pairwise basis only. It is clear that the abundance of polymorphic markers now mapped throughout the human genome now favor the application of multipoint linkage analyses as a more powerful method to detect the presence of a major locus influencing liability to schizophrenia.

The statistical methods to be used to assess evidence of linkage in the Saudi-Arabian consanguineous pedigrees do not actually differ from the standard calculation of pairwise or multipoint lod scores. The uniqueness of this approach and its potential power lie in the structure of the inbred families themselves. For example, genotyping a single affected offspring of a first-cousin marriage permits up to six meioses to be studied. The lod scores can be calculated using standard programs that compute multilocus likelihoods or with a special program (Hommap) that is specially designed for consanguineous families and therefore performs its calculations much faster than standard programs (Lander and Botstein 1987). In addition, since we are sampling as many individuals from these consanguineous pedigrees as are available, we can consequently also use these families to test for linkage to a dominantly inherited schizophrenia gene such as that reported to reside on chromosome 5q11–13.

Among the many unresolved issues in the linkage analysis of complex phenotypes, two are of particular relevance to schizophrenia: optimal diagnostic criteria and fitness effects. In simulation studies based on the affected sib-pair method, Kendler (1988b) has shown that (1) altering diagnostic thresholds can have dramatic effects on the power to detect linkage and (2) power to detect linkage is a near Gaussian function of diagnostic stringency. Simulation studies for schizophrenia indicate that optimal power for the detection of linkage will occur with a diagnostic threshold somewhat broader than classic schizophrenia. These results, in addition to the large family data set described above, should put us in a good position to choose, a priori, the optimal diagnostic criteria to be used for linkage analysis.

Individuals with schizophrenia reproduce at a rate approximately half that of the general population (Haverkamp et al. 1982). This produces a large alteration in the pattern of risk of illness in relatives from that predicted by any standard genetic model (parents < sibs < offspring; Kendler 1986). Studies of discordant monozygotic twins suggest that for schizophrenia fitness is phenotype dependent (e.g., Fischer 1971). For incompletely penetrant single gene models, phenotype dependent reduced fitness selects for incomplete penetrance in all progenitors because they have been ascertained conditional upon their having reproduced. Kendler and MacLean (in press) have derived a “correction term” for the impact of reduced fitness on penetrance in progenitors and has validated this by extensive stochastic simulations. Although a fully satisfactory treatment of fitness effects in linkage analysis is not currently available, the application of the derived correction term should prevent a
serious distortion of results by the large fitness reduction associated with the disorder.

Given the a priori probability of genetic heterogeneity in schizophrenia, even in a racially and culturally relatively homogeneous population such as that of Ireland, it will be essential to test whether a linked single major locus is segregating in only a portion of our families. Risch (1988a) has developed a new method for testing for such linkage heterogeneity, which shows greater power under most circumstances than an admixture test developed by Smith (1963) and the original test developed by Morton (1956). The multipoint "interval mapping" and "simultaneous search" methods of Lander and Botstein (1986), although not yet fully developed, offer the promise of greatly increasing the power to detect heterogeneity, and these methods will also be applied when available. As mentioned above, the homozygosity mapping method is especially powerful for detecting locus heterogeneity, assuming that all loci are inherited in a completely recessive manner.

Combined Segregation and Linkage Analyses. As discussed above, analysis of linkage depends upon the mode of inheritance estimated by segregation analysis. Many investigators deal with this problem by "searching" in their linkage analysis over a wide array of values of penetrance and allele frequency. Conversely, linkage with a marker, together with association data, often yields the most important information concerning the segregation model. If segregation is misspecified, however, evidence obtained for linkage is inaccurate. Especially in schizophrenia, where the penetrance of any major genes may be low, the two models interact. Simultaneous estimation of linkage and segregation parameters (e.g., MacLean et al. 1984; Bonney et al. 1988) constrains genetic inference and yields more unbiased estimates of recombination, as well as gene frequency and penetrance. Simultaneous estimation invalidates the traditional lod score criterion of 3, however, so that statistical inference under complex interactive models must be used. This combined analysis approach may be especially useful after linkage to a major locus is found, to help refine estimates of genetic parameters, and confirm the initial findings.

Power Analyses. Unfortunately, relatively little is known about the sample size required to detect linkage for a complex phenotype like schizophrenia. We know that for two fully informative codominant traits, with a lod score criterion of 3.0, 80 percent power for a recombination rate of 0.10 and 0.30 requires approximately 30 and 130 fully informative (i.e., double back-cross) gametes, respectively (Kidd and Ott, unpublished manuscript). However, both incomplete penetrance (McGuffin and Sturt, unpublished manuscript) and genetic heterogeneity (Cavalli-Sforza and King 1986; Goldin and Gershon 1988; Risch 1988b) can increase significantly the sample size required to detect linkage. For example, for a recombination frequency of 10 percent, McGuffin and Sturt (unpublished manuscript) show that for a dominant trait, penetrances of 50 percent and 80 percent require sample sizes to detect linkage that are 9 and 2.6 times larger, respectively, than for a fully penetrant trait. Cavalli-Sforza and King (1986) show that for a fully penetrant dominant disorder, at a map distance of 10 cM from a marker locus, for 50 percent power only 14 families with 2 parents (phase unknown) and 3 children are needed given genetic homogeneity. If the locus under consideration were responsible for 80, 60, or 40 percent of cases, however, the numbers of families required would increase to 21, 37, and 80, respectively.

Recently, Martinez and Goldin (1989) have estimated the number of phase-unknown nuclear families needed to give 50 percent power to detect linkage and to detect heterogeneity as a function of mode of transmission, size of sibship, percentage of affected cases of the linked form, and minimum number of affected individuals per family. They optimistically assume a phenocopy rate of 5 percent and a closely linked single marker that is informative in every family, or a pair of flanking markers, both of which are similarly always informative. We estimate that it would require an approximate tripling of the size of the samples that are reported to have the ability to detect linkage 80 percent of the time, considering the levels of informativeness of most marker loci available today.

On the basis of this estimate, and of the data presented by Martinez and Goldin (1989), we arrive at the following conclusions: If schizophrenia is transmitted as a dominant trait with 50 percent penetrance and we obtain an average nuclear family size of four siblings (with at least two of the siblings affected with schizophrenia and with both parents available for sampling), and if 50 percent of the families have the linked form of the illness, then the sample size that we are proposing to collect in Ireland (300 families) will have
approximately 80 percent power to detect linkage. Only if the penetrance is around 90 percent, however, would we have high power to detect linkage heterogeneity. This last point calls into question whether much weight should be attached to the fact that significant evidence of heterogeneity was not detected among the few extended families studied by Sherrington et al. (1988). We further estimate that if only 25 percent of the Irish families that we plan to study are linked to a common locus, with a penetrance of 50 percent, our study population would only have about 50 percent power to detect linkage. With 90 percent penetrance, however, our sample would have about 75 percent power to detect linkage in this situation.

The families we are proposing to study in this linkage project (approximately 300 pedigrees, which will produce the rough equivalent of 320-400 affected sib pairs) would represent, to our knowledge, the largest consistently ascertained sample to be collected. It is fair to conclude that we would have considerable power to exclude linkage anywhere in the genome, be it a place that we previously suspect (e.g., chromosome 5, candidate genes) or close to randomly chosen informative markers if (1) the nearby disease locus accounted for a substantial proportion of the cases for schizophrenia, (2) penetrance was not too low, and (3) the marker was relatively close (e.g., < 15 cM) from the disease gene. At present, a more precise assessment of the power of this linkage study cannot be made; none of the studies cited above examined conditions close to those found in the families and markers studied here. The reality of field studies such as ours in Ireland is that among different families, individuals related to different degrees are usually collected, resulting in a very wide assortment of family structures. Similarly, the markers we will use also vary widely in their degree of informativeness.

A computer simulation approach recently developed by Boehnke (1986) offers considerable promise to assess the power of linkage studies despite these complexities. This method uses the actual pedigree structures that will be used in the linkage study. Power is assessed by running many computer simulations using the pedigrees, under a variety of assumptions about (1) the mode of transmission of the disease locus, (2) the degree of polymorphism of the markers, and (3) how close the markers are to the disease gene. The results permit an assessment of both the chances of detecting linkage and the power to exclude linkage using the families and markers available. This method is currently being expanded to allow for incomplete penetrance (L.M. Ploughman and M. Boehnke, personal communication), an essential feature for use in linkage studies of schizophrenia. When this software becomes available, we will apply this simulation method to the Irish schizophrenia pedigrees and should then obtain a more precise estimate of their power for detecting and excluding linkage.

As mentioned above, power calculations presented by Lander and Botstein (1987) indicate that very few (15-20) affected offspring from consanguineous families are needed to map major genes that are inherited with complete or almost complete recessivity. We expect to obtain a sample at least this large in our studies of Saudi-Arabian families. However, we are unaware of any estimates of the loss of power if the disorder under study is transmitted by some intermediate mode of inheritance between complete dominance or recessivity, or if nongenetic phenocopies are sometimes sampled, both situations being highly likely in studies of schizophrenia.

Significance of Linkage Studies

The demonstration and/or replication of linkage of schizophrenia to a genetic marker will have major implications for the entire field of schizophrenia research. First, it will provide conclusive evidence, at least for a subgroup of individuals with schizophrenia, that variation at a major identifiable genetic locus profoundly influences liability to illness. Second, once the major gene is precisely mapped with close flanking markers (so that genotype at the disease locus can be reliably inferred from marker phenotypes), it will be possible to clarify to a far greater extent than currently feasible the range of phenotypic manifestations of the "schizophrenia gene." This will not be limited to clinical phenomena (e.g., schizotypal symptoms) but can also include such "endophenotypes" as smooth pursuit eye movements (permitting a definitive test of the latent trait model [Holzman et al. 1988]), attentional measures such as the CPT and soft neurological signs. Identification of a gene for schizophrenia will add an entirely new dimension to longitudinal high-risk studies where high-risk status can be directly inferred from genotype rather than indirectly from parental phenotype. Finally, and perhaps most importantly, mapping of a gene for schizophrenia will facilitate the cloning and characterization of such
a gene. While no simple technical matter, it would be in the range of current techniques to be able to determine, at the level of the DNA sequence, the variation responsible for altering susceptibility to schizophrenia. This could in turn lead to a “reverse genetic” or “bottom-up” understanding of the pathophysiology of this disorder, which would in turn open new vistas for treatment and prevention.

While the scientific value of positive findings is obvious, completely negative results would also provide considerable information about the genetics of schizophrenia. First, vis-à-vis the report of linkage by Sherrington et al. (1988), the inability to replicate this finding in a large Irish sample would call into question either the validity or the generalizability of this finding. Second, ruling out a number of “candidate genes” that will be tested in this study as causally related to disease transmission within pedigrees is important in narrowing the focus to more plausible theories of the genetic substrate of liability to schizophrenia. Third, ruling out large sections of the genome by typing many highly informative markers in a sufficient sample size provides other investigators at least some leads as to which sections of the genome future efforts should address. Finally, if a search of the entire genome future efforts should address. Finally, if a search of the entire genome were unsuccessful, this would suggest either that (1) no gene “of large effect” exists for schizophrenia in the Irish population or that (2) because of a high degree of heterogeneity, or interaction with another genetic locus (epistasis) or an unknown environmental risk factor (G x E interaction), current methods are insufficiently powerful to detect such a gene if present.

References

American Psychiatric Association. 


DeLisi, L.E.; Reiss, A.L.; White, B.J.; and Gershon, E.S. Cytogenetic


Kamholz, J.; Spielman, R.; Gogolin, K.; Modi, W.; O'Brien, S.; and


**Acknowledgments**

Field studies in Ireland are supported by National Institute of Mental Health grant MH–41953 and by grants from the Schizophrenia Research Program of the Scottish Rite, N.M.J., U.S.A., and from the McDonough Foundation. The field study project is a collaborative endeavor with the Health Research Board (Dublin) and the Western Health Board (Republic of Ireland). Dr. Dermot Walsh of the Health Research Board is a co-principal investigator on this project. Preliminary support for the typing of DNA markers on the Irish families is supported by an A.D. Williams Faculty Research grant from the Medical College of Virginia.

**The Authors**

Scott R. Diehl, Ph.D., is Assistant Professor, and Kenneth S. Kendler, M.D., is Professor, Departments of Psychiatry and Human Genetics, Medical College of Virginia, Richmond, VA.

**Announcement**

NARSAD ARTWORKS, a non-profit organization affiliated with The National Alliance for Research on Schizophrenia and Depression, has produced notepaper illustrated by mentally ill artists. The informal art cards are available for sale from the National Alliance for the Mentally Ill, the National Depressive and Manic Depressive Association, and the National Mental Health Association.

For further information, please write to:

**NARSAD ARTWORKS**
P.O. Box 941
La Habra, CA 90633-0941