Eur Arch Psychiatry Clin Neurosci (1993) 243:143–149

European Archives of Psychiatry and Clinical Neuroscience

Linkage studies of psychiatric disorders

Neil Risch[1,2], **Kathleen Ries Merikangas**[1,3]

Department of [1]Epidemiology and Public Heath, [2]Genetics and [3]Psychiatry, Yale University School of Medicine, New Haven, CT 06510, USA

Summary. Linkage analysis has been successful in identifying the genetic basis of numerous Mendelian diseases. These successes were due in part to the rapid developments in molecular biology, which have yielded a plethora of informative genetic markers. Although there is strong evidence that the manifestation of schizophrenia and bipolar affective disorders is controlled by genes, no evidence for linkage has been established. For psychiatric disorders, the most important limiting factor is likely to be the lack of single loci with very large effects that occur with any relevant frequency. The difficulties of linkage studies in psychiatric disorders are discussed with reference to non-psychiatric genetic diseases for which linkage to genetic markers has been successful. Recommendations for collecting information to clarify the patterns of transmission of the psychiatric disorders are described.

Key words: Genetics – Linkage – Psychiatric disorders – Genetic epidemiology

Introduction

Rapid developments in molecular biology have introduced a new era in human genetics. Although only a decade ago linkage depended upon inferences about underlying alleles from phenotypic expression of known markers, advances in the identification of polymorphic DNA markers and the methods for processing and sequencing DNA have dramatically enhanced our ability to detect linkage between these markers and diseases for which no aberrant gene product has been identified. Family pedigrees may be examined to determine whether a particular disease or trait is associated with a specific DNA marker (i.e. linkage). Since the exciting discovery of a linked marker for Huntington's disease in 1983 (Gusella et al. 1983), numerous other diseases have followed, and the primary gene defect has now been discovered for many such disorders.

It is now more than two decades since the first linkage study for a psychiatric disorder was performed (Reich et al. 1969). Despite the initial excitement generated by the successful application of molecular genetics technology to linkage studies of severe diseases, the lack of positive findings in psychiatry has been a sobering experience. Nevertheless, this approach could have the greatest impact on diseases, such as those within the realm of psychiatry, for which the etiological basis in unknown or poorly understood. This paper describes the complex features of the psychiatric disorders that need to be examined in order to optimize the chance of success in the next generation of linkage studies.

Correspondence to: K. R. Merikangas

Principles of population genetics

Although most human diseases have an inherited component to susceptibility, there is a wide range of heritability. A disease which is due entirely to the effect of a single mutated gene is referred to as Mendelian. Mendelian diseases follow very specific patterns of inheritance in families (Table 1). These patterns follow from the fact that genes are inherited in pairs, one from each parent. Genes are located on chromosomes, 22 of which are referred to as autosomes (i.e. non-sex determining chromosomes) and one pair of sex determining chromosomes (labelled X and Y). Females possess two X chromosomes and always transmit an X chromosome to each child. Males possess an X and Y chromosome, and transmit either the X or the Y to each child. If the X is transmitted, the child is a female, while if the Y is transmitted it is a male; hence fathers determine the sex of the children.

When possession of only a single copy of a mutant gene is sufficient for disease expression, the disease is referred to as *dominant*. If compatible with reproduction, dominant diseases are transmitted directly from parent to child, with half the children carrying the same mutation (and therefore being affected), on average. If the genes is on an autosome, males and females are affected equally frequently. Often, pedigrees can be constructed, whereby the disease is transmitted through many generations of affected individuals. Normally, only affected individuals produce affected children; the children of unaffected individuals are at no risk.

If the mutant gene is situated on the X chromosome, the sex-specific pattern of inheritance is quite different. In this case, affected mothers will transmit the disease

Table 1. Observed patterns of transmission for the major genetic models[a]

Model	Observed patterns
Single major locus (SML)	
Autosomal dominant	Every generation, no skipping
	Unilineal transmisssion
	Affected persons transmit to 1/2 (on average) of the offspring
	Males = females
Autosomal recessive	Horizontal transmission
	1/4 of sibs of affected persons are affected
	Males = females
	Consanguinity may be increased in parents
X-linked recessive	Males >>> females
	Absence of male-to-male transmission
	1/2 sons of carrier female are affected
	Males are carriers
X-linked dominant	Females > males
	Affected females transmit trait to 1/2 sons and 1/2 daughters
	Affected males transmit to all daughters and not sons

[a] (From Merikangas, 1987)

half the time (on average) to both their sons and daughters, as in the autosomal case. By contrast, affected fathers will transmit the disease to all of their daughters and to none of their sons. It is this lack of father-son transmission that is the hallmark of X-linked inheritance. Furthermore, since females carry two X-chromosomes, as opposed to males who have just one, the disease (i.e. mutation carrier) frequency in females is twice as large as in males. Because females also carry a normal X chromosome, they may be less severely affected than males, who carry only the single abnormal X chromosome.

When two copies of a mutant gene are required for disease expression, it is referred to as *recessive*. In this case, disease transmission appears horizontal rather than vertical; that is, affected individuals appear together in sibships, while the parents and offspring of affected individuals are generally unaffected. Parents are usually unaffected carriers of the mutant gene (i.e. heterozygotes), while the affected children are homozygotes. By Mendel's law of segregation, an average of 25% of the male and female offspring are affected. It the disease is rare, one can expect to find an increased frequency of parental consanguinity (i.e. parental relatedness), because the chance of both parents carrying the same mutant gene may be significantly increased when they are related.

If the recessive disease gene lies on the X chromosome, usually only males are affected, because females also carry a normal X chromosome. However, the disease is transmitted only through carrier females, because there is no father-to-son transmission. Half of the sons of a carrier female are affected, but none of the daughters (although half will be carriers). Despite the recessive nature of the disease, multigeneration pedigrees can be obtained because males, who carry a single X chromosome, can be affected in consecutive generations. All of these males will be related to each other through carrier females.

Table 2. Risk ratio for relatives of schizophrenics

Type of relative	Recurrence risk[a]	Risk ratio[b]
Offspring	13%	10.0
Siblings	9%	8.6
Monozygotic twins	48%	52.1
Dizygotic twins	17%	14.2
Half sibs	6%	3.5
Nieces/nephews	4%	3.1
Grand children	5%	3.3
First cousins	2%	1.8

[a] (From McGue et al. 1983)
[b] (From Risch 1990a)

A number of features may lead to deviation from precise conformity to these rules. For example, not all individuals who possess a disease-predisposing genotype may be affected. This phenomenon is referred to as *reduced penetrance*. Indeed, in some cases of a dominant disease, the disease gene may pass through an individual who is unaffected, but has an affected parent (and/or sib) and an affected child. This may give the appearance of a "skipped" generation. Akin to reduced penetrance is *variable expressivity*, where the expression of disease among those carrying the genotype can vary greatly, ranging from extremely severe to extremely mild or even non-detectable. Also, the age at which the disease expresses itself may be highly variable, ranging from childhood to late age (as in Huntington's disease). A further complication can arise when non-genetic, or environmental causes yield phenotypes that resemble those produced by genes, referred to as *phenocopies*. This means that individuals without the disease predisposing genotype may present with a disease quite similar to the genetic form; often, however, phenocopies can be separated from a genetic form on clinical grounds.

Although reduced penetrance and phenocopies eliminate the precise one-to-one correspondence between genotype and phenotype, neither phenomenon should necessarily abrogate the Mendelian nature of a disease. The essential characteristic is that the familial nature be attributable primarily to the effect of a single genetic locus. Other genetic loci may contribute to the expression of the disease, but they are neither necessary nor sufficient, and play a "modifying" role.

Another important characteristic of Mendelian disease is that they are *rare*. The reason for this is that deleterious mutant alleles are eliminated from a population through selection; that is, individuals who are affected tend to have fewer children than unaffecteds, and therefore pass on fewer of the deleterious alleles to future generations. As a general but not absolute rule, the stronger the selection against a deleterious gene, the rarer will that gene be in the population. This is especially true for dominant diseases, where everyone pos-

sessing the disease allele can be affected, and therefore selection is particularly effective. The frequency of a dominant disease in a population is often due to a balance between new mutations being produced and then eliminated through selection. The most common dominant diseases tend to have a population frequency of 1 in 10,000 or less (such as achondroplasia, tuberous sclerosis, Huntington's disease, torsion dystonia). However, there are a few cases of dominant diseases with higher frequency; usually, such diseases are very mild and have only a small effect on reproductive capacity. The most common dominant disease yet defined is hypercholesterolemia, due to a defect in the receptor for low density lipoprotein (LDL) cholesterol, with a frequency of 1 in 500 in most populations (McKusick, 1990). Since this defect only becomes clinically manifest in terms of significant heart disease in the fourth and fifth decades of life, it is reasonable to assume that selection against this defect is mild.

Although recessive diseases also tend to be quite rare, especially those that are selectively disadvantageous, there are some notable exceptions. These exceptions are primarily due to heterozygote advantage – that is, that individuals who are carriers (heterozygotes) for the recessive gene actually have a selective advantage over the normal homozygotes. The classic example of this situation is the hemiglobinopathies such as sickle cell disease which is lethal, yet persists at very high frequency in some populations (as high as 1–2%). It has been shown that heterozygotes for these diseases have a significant selective advantage over normal homozygotes against malaria, explaining the high frequency of these diseases in malarial environments.

Stochastic (chance) effects can also influence disease frequencies. This is particularly the case for populations founded by a small number of individuals (where the gene frequency may differ markedly from the original population from which the founders derive), or populations experiencing a significant bottleneck, where the number of surviving and reproducing individuals may be dramatically reduced, leaving gene frequencies altered form the original frequencies by chance.

Another important characteristic of Mendelian diseases, attributable to their highly familial nature and population infrequency, is the very high ratio of risk to close (e.g. first-degree) relatives of an affected individual versus the general population frequency. For example, for a dominant disease with complete penetrance and a population frequency of 1 in 10,000, this ratio would be 1 in 2 to 1 in 10,000 or 5,000 to 1. For a recessive disease with a population frequency of 1 in 10,000, the risk ratio for sibs would be 1 in 4 to 1 in 10,000, or 2,500 to 1.

Finally, Mendelian diseases, or indeed any disease in which familial recurrence is attributable to a single locus, exhibit a decrement of 0.5 in the risk of relatives according to the decrease in the degree of genetic relatedness (Risch 1990a). For example, the risk to offspring of an affected individual should be double the risk to nieces/nephews, whose risk should be double that of first cousins. As described below, a pattern of risks that decreases more rapidly suggests a more complicated genetic basis for inheritance.

Genetic heterogeneity

What appears to be a single Mendelian disease entity may in fact be attributable to multiple different mutations. For example, for a dominant disease, different mutant alleles may run in different families. This phenomenon is referred to as *genetic heterogeneity*. Such heterogeneity can be caused by mutations occurring at the same genetic locus (referred to as allelic or intra-locus heterogeneity), or at different genetic loci (called non-allelic or inter-locus heterogeneity).

There are three ways such heterogeneity can be detected. First, different mutations may be associated with distinct clinical characteristics. For example, Duchenne (severe) and Becker (mild) *X*-linked muscular dystrophy are attributable to mutations at the same genetic locus (the dystrophin gene) (i.e. allelic heterogeneity); however, it is the severity of the mutation that determines the clinical picture (i.e. Duchenne versus Becker), (Kunkel et al. 1986). The essential feature of such heterogeneity is that the clinical picture tends to "run true" within families. Another classic example is neurofibromatosis, where the predominant peripheral von-Recklinghausen form (NF1) is due to a locus on chromosome 17, while the central, acoustic form (NF2) is due to a locus on chromosome 22 (an example of non-allelic heterogeneity). Again, these forms run true within families.

The second way genetic heterogeneity can be detected is by observing different Mendelian inheritance patterns in different families.

A classic example of this type of heterogeneity relates to the eye disease retinitis pigmentosa, which can manifest in an autosomal dominant, autosomal recessive, or X-linked recessive pattern. Presumably, distinct mutations underlie these differently inherited forms.

The third way to detect genetic heterogeneity is in linkage analysis. Although the pattern of inheritance may appear identical in different families and the clinical aspects similar, only a subset of families may demonstrate genetic linkage to a particular chromosomal region. Obviously, linkage analysis can detect only non-allelic heterogeneity. Often (but not always), once linkage and heterogeneity are detected, linked and unlinked families can be distinguished clinically. Although genetic heterogeneity is classically applied to dominant diseases, it also applies to recessive diseases. A typical example would be cystic fibrosis, in which some mutations in the CF gene are associated with pancreatic insufficiency, while other (allelic) mutations are associated with pancreatic sufficiency. The clinical picture is clouded, however, by the occurrence of compound homozygotes (i.e. those who possess one allele of each type).

The degree of genetic heterogeneity for a particular Mendelian disease may be difficult to predict accurately. As a general rule, one might expect the degree of mutational heterogeneity to increase with the degree of selection operating against the mutation and with the population frequency. In particular, a (relatively) frequent disease with a large selective disadvantage would be likely to have a large number of different underlying mutations (allelic and/or non-allelic). A rarer disease with lower

selection would be likely to be more genetically homogeneous. However, as mentioned previously, it is also important to consider the history of the population under study, and the possible role of genetic drift. A disease which has drifted to a high frequency in a given population by chance is most likely to be largely homogeneous.

The concept of genetic heterogeneity generally applies to Mendelian disease, and is based on the concept that the disease is due, in any particular family or families, to the effect of a single genetic locus. Caution needs to be taken in applying this term more generally to a disease that is non-Mendelian, lest the Mendelian hypothesis be implied inappropriately.

Complex inheritance

The above discussion relates to Mendelian disorders. In fact, many diseases of interest, especially psychiatric ones, tend to be familial but do not conform to Mendelian patterns. First, most psychiatric disorders, even severe ones, tend to be far more common than typical Mendelian diseases, and the ratios of risk to first-degree relatives compared to population frequency are far below Mendelian rates. For example, these ratios tend to be only in the range of 2–10-fold for most psychiatric conditions. The critical question then becomes, what is the genetic basis for these disorders? The answer to this question is important for designing and interpreting linkage studies to map genes for such disorders. As a general rule, however, one can suspect that the larger the effect of a gene that contributes to disease susceptibility the rarer it will be, particularly when the disease has a selective disadvantage. It is likely that under such circumstances, more than one locus contributes to disease susceptibility, and no single locus accounts for the majority of the familial aggregation.

If there are interactive effects between predisposing loci (i.e, the contributions of various loci to susceptibility are non-additive when considering penetrance, or disease risk), the recurrence patterns in families may reveal such interactions. Specifically, if the genetic contribution to a disease is determined by a single locus, or if multiple loci act in an additive fashion on risk, then the risk to relatives should decrease approximately by a factor as 0.5 with each degree of relationship; for example, from first-degree relatives (such as parents or children) to second-degree relatives (such as uncles or nieces) to third-degree relatives (such as first cousins) (Risch 1990a). Furthermore, if there are no recessive genes operating, as indicated by comparable risks to siblings and offspring, then the ratio of concordance between identical (MZ) twins and fraternal (DZ) twins should also be twofold. If this ratio is greater than twofold, and the recurrence risks decrease more rapidly than by a factor of 0.5 with each degree of relationship, interactive effects between loci are suggested. This automatically implies that more than a single locus must contribute to disease risk, and no single locus can largely predominate. A summary of familial recurrence data for schizophrenia according to the type of relative is shown in Table 2. The MZ-to-DZ twin concordance ratio is nearly two times greater than the expected ratio of 2. Moreover, the decrements in risk between first- and second-, and between second- and third-degree relatives are also greater than 2 (Risch 1990a). These data suggest that multiple loci (perhaps three or more) contribute to susceptibility to schizophrenia, with no one locus predominating. This type of interactive effect among loci contributing to risk for common, familial disorders appears to be a recurring theme among various diseases, such as type-1 diabetes, multiple sclerosis, and cleft lip and palate.

While the general familial pattern of a disease may appear consistent with multiple, common interacting genes (also known as epistasis), this does not automatically preclude the possible existence of a single locus with a very large, or even deterministic effect (i.e., a Mendelian-type gene). However, we can expect that if such loci do exist, they would be quite rare (as is typical for Mendelian genes), and account for only a small proportion of all cases.

How can one know whether such genes exist, a priori? Essentially, there are two lines of evidence. The first, and perhaps most convincing, is the case where a subset of the disease can be determined on clinical grounds that defines it as a Mendelian disease. One such subgrouping can sometimes be made based on age of onset, where early-onset disease appears to have a Mendelian basis. Alzheimer's disease is an excellent example of a disease that is very common in late adulthood (80s and 90s), but very rare in early adulthood (40s and 50s). Yet some families have been characterized that have many cases of Alzheimer disease, with onset in the 30s or 40s, with many cases (as many as 50 or more) appearing in consecutive generations of the same family (St. George-Hyslop et al. 1987; Schellenberg et al. 1988). The infrequency of early-onset disease in the general population coupled with the high fequency and autosomal dominant inheritance pattern of early-onset cases in some families suggests the presence of a gene mutation at a single locus with a strong effect (i.e. a Mendelian gene). It is critical to note that the relatives in these families all tend to be affected with early-onset disease, a hallmark characteristic of the mutation. In fact, if other family members are affected, but at a late age, it would be unclear, and perhaps unlikely, that they carry the same mutation.

Recent advances in identifying the genes for colon cancer also illustrate the successful application of molecular genetic approaches to subtypes of complex conditions based on familial aggregation of early age of onset. The gene for an autosomal dominant form of colon cancer, familial adenomatous polyposis (FAP), has been mapped to a marker on chromosome 2 (Kinzler et al. 1991). However, FAP accounts for only about 1% of the cases of colon cancer. More recently, another gene on chromosome 2 has been linked to a series of families with early-onset colon cancer, accounting for up to 15% of the additional cases of this disease (Pettomaki et al. 1983). It is particularly noteworthy that these families also exhibited a high frequency of other types of cancer.

Diabetes mellitus may also be subdivided into two forms based on the requirement of insulin, namely type-1

or insulin-dependent diabetes (IDDM), and type-2, or non-insulin-dependent diabetes (NIDDM). Type 1 generally has early onset, and is by far less common than IDDM, with a general population frequency of about 0.4% by age 25 years. By contrast, type-2 diabetes generally has later onset, and is quite common, with a late adult frequency of about 5–10%. The genetic basis of type 1 is likely to be complex, involving interactions between multiple interacting loci. This conclusion is supported by the high MZ–DZ twin concordance ratio, as well as a rapid drop in risk from first-to second-degree relatives (Rich 1990). One contributing locus, the HLA complex, has already been identified, but clearly constitutes only a portion of the genetic susceptibility (Risch 1987). Furthermore, the mouse model of type-1 diabetes, known as the NOD mouse, also shows a complex genetic picture with at least several contributing loci (Todd et al. 1991; Risch et al. 1993).

There is also a form of type-2 diabetes that occurs in youth; this form is called MODY, for maturity-onset-type diabetes of youth. Although MODY is very infrequent, constituting only about 5% of all early-onset diabetes (or a population prevalence of about 1/5,000), it is this rare form of diabetes that appears to have a Mendelian, autosomal dominant inheritance pattern (Fajans 1990). Despite the comparability in the ages of onset of MODY and type-1 diabetes, MODY is the far more clinically benign of the two; that is, insulin therapy is not necessary and it may not be diagnosed for a long period of time, even for decades. Hence, it is quite compatible with survival and reproduction.

By contrast, type-1 diabetes is far more severe clinically, requiring insulin for survival. Indeed, before the discovery of insulin, such individuals would likely have died from their disease, and even today there is a selective disadvantage. Yet it is the severe type-1 form that is by far more common than the benign MODY. Undoubtedly, the reason for this is that IDDM is genetically complex, having an important non-genetic component, while MODY is Mendelian. Thus, Mendelian genes are likely to be quite rare, even when the effects are relatively benign; when the clinical effects have a strong selective disadvantage, they will be particularly rare.

How does this conclusion relate to psychiatric disorders, such as schizophrenia? Since schizophrenia, in general, has a relatively strong selective disadvantage, a Mendelian gene for schizophrenia would likely be quite rare and account for only a very small proportion of all cases. Even if such a gene (or genes) does exist, it would be difficult to find it without some clinical or age-of-onset characteristic which defines it as a Mendelian syndrome, enabling informative pedigrees to be identified for linkage studies. As yet, no such subgroup has been defined for any psychiatric disorder.

The second line of evidence that can be eymployed to detect a Mendelian subgroup is segregation analysis. If a rare Mendelian subgroup of a common familial disease exists, and this subgroup is characterized by high penetrance, then a small subset of families should have particularly high risk compared to the large remainder. It is also helpful in this case if the Mendelian subgroups have a discriminatory characteristic, such as an early age of onset.

Breast cancer, which achieves a frequency of 11% at late age among women, may also be subdivided by age at onset. Two observations have suggested that an early-onset autosomal dominant Mendelian form exists: first, the risk to female relatives increases dramatically with decreasing age of onset of the proband; and second, the risk to sisters of probands increases substantially when either the mother and/or another sister of the proband is also affected (Schwartz et al. 1985; Claus et al. 1990). These impressions have been confirmed when family data have been subjected to formal segregation analysis, where the results have generally indicated the presence of a rare, early-onset autosomal dominant form with high penetrance, with the vast majority of cases being non-genetic (Williams and Anderson 1984; Bishop et al. 1988; Newman et al. 1988; Claus et al. 1991). These results have subsequently been confirmed with the detection of linkage on chromosome 17 in early-onset families (Hall et al. 1990; Easton et al. 1993).

Linkage strategies

In mapping genes for common, complex diseases, it is important to consider linkage strategies in terms of what types of families are best, which phenotypes to include as affected, what statistical methods to employ, and the statistical inference of results. The classic paradigm for Mendelian traits is to identify informative families segregating the trait of interest. For a recessive disease, this generally consists of identifying nuclear families with normal parents and as many affected offspring as possible. Inbred families, where the parents are related to each other, can be especially useful. For psychiatric disorders, it is unlikely that any recessive loci could be identified, if they exist, unless they could be identified in advance by other means. Because the recurrence risks among siblings and offspring of probands with psychiatric disorders are equal, recessive modes of transmission are not likely to characterize any of the major psychiatric disorders.

For a trait which is dominant, extended, multigenerational pedigrees with many affected members are the most useful. For a rare dominant trait, affected individuals can be assumed to be heterozygotes at the disease locus; therefore, they are informative for linkage analysis if they are also heterozygotes at the marker. Hence, all affected individuals with children in an extended pedigree are potentially informative. Furthermore, for a typical Mendelian dominant disease, all affected individuals in the pedigree will be heterozygous for the same disease locus, precluding the possible complication of genetic heterogeneity, although combining evidence across pedigrees can still be influenced by heterogeneity.

The statistical analysis for this case is the lod score method, which incorporates knowledge of the dominant mode of disease inheritance. The log of likelihood ratio of observing the co-segregation of the disease and marker under linkage versus no linkage is the lod score statistic;

a value of 3 or greater (or likelihood ratio of 1000 or greater) is taken as statistical evidence supporting linkage.

The same approach, namely identifying multigenerational pedigrees and applying the lod score method, is also appropriate when studying a Mendelian, autosomal dominant subgroup of a common, multifactorial disease. For example, this approach has been appropriately and successfully employed to detect linkage for early-onset Alzheimer's disease on chromosomes 14 and 21 (Goate et al. 1989; Schellenberg et al. 1992) for early onset breast cancer on chromosome 17 (Hall et al. 1990; Easton et al. 1993) and for MODY on chromosomes 20 and 7 (Bell et al. 1991; Froguel et al. 1992). All three cases may involve non-allelic genetic heterogeneity even among the dominantly inherited families; indeed, the latter pattern has been clearly demonstrated for MODY (Froguel et al. 1992).

Optimal strategies for diseases with no evidence for an autosomal dominant subgroup, either clinically or from segregation analysis, such as psychiatric disorders, are less certain. In the absence of a Mendelian inheritance pattern, some have advocated using small constellations of affected relatives, such as affected sib pairs or other affected relative pairs. The advantage of such an approach is that it does not require the specification of the mode of inheritance of the disease susceptibility locus being sought (which, indeed, is generally unknown); robust methods of analysis, such as calculating the amount of marker gene sharing identical by descent by pairs of relatives, can be performed.

The relative power of pedigrees vs. pairs of relatives is difficult to determine without evidence of a precise genetic model. There could even be a serious disadvantage in using dense families with many affecteds, if the true genetic model underlying the disease is oligogenic (i.e. several interacting genes of small or moderate effect). Detection of linkage requires that parents be heterozygous at the disease locus, and not homozygous for the normal allele and the disease-predisposing alleles. Parents with many affected children, and parents who are themselves affected, may have an increased probability of being homozygous at the disease locus, making them uninformative. Therefore, one needs to sample families to maximize parental heterozygosity.

As a rule of thumb, a general approach that is likely to achieve this across a broad range of circumstances is to collect a large sample of affected sibling pairs with both parents normal. The only situation where such a strategy is likely to be disadvantageous is when a Mendelian subform exists and extended, genetically homogeneous pedigrees can be obtained. However, as stated before, the ability to obtain such pedigrees, particularly for a disease such a schizophrenia, seems unlikely, a priori.

In collecting relative pairs, special emphasis should be placed on obtaining pairs where both members are affected with the core (severe) diagnosis; this is because core diagnoses usually show the greatest degree of familial aggregation (i.e. relative risk ratio), whereas milder diagnostic phenotypes show lower ratios. The power to detect linkage is directly related to the size of this ratio (Risch 1990b); hence, diagnoses which maximize the ratio should be employed. For example, when studying schizophrenia, pairs where both relatives are schizophrenic should be given high priority as opposed to those where one has schizotypal personality disorder. Similarly, pairs with bipolar disorder are preferable to those with a bipolar and a unipolar.

As an empirical example of the preferability of sampling sibling pairs (or multiplex sibships with normal parents) as opposed to multigenerational pedigrees, we can reconsider data of the two types published by Barbosa et al. (1977, 1980) on type-1 diabetes and the HLA system. In the first series, there were 23 nuclear families with at least two affected siblings and normal parents and one three-generational pedigree with affecteds in three generations. These data were analyzed using the lod score method assuming a recessive model; with 50% penetrance, the maximum lod score was about 4.0, providing significant evidence for linkage. The second set of data contained 28 multigenerational pedigrees (with the disease in at least two generations). The lod score method was also used, but assuming a dominant rather than recessive model. However, the maximum lod score was only around 0.4, again assuming 50% penetrance. Hence, there was no evidence for linkage in the latter analysis. The difference between the significant lod score for the first set of data and the non-significant score for the second set cannot be attributed to model misspecification in the lod score analysis for the second set of data, nor to HLA not being involved in susceptibility in the latter set of families; rather, it is because there was less linkage information in the second data set because of the manner in which the families were sampled.

One should also keep in mind that the lod score statistic of 3.0 for significance was defined for Mendelian traits, where there is strong a priori evidence of a single locus underlying the trait being mapped. For this case, the lod score of 3 corresponds to a 5% posterior false-positive rate, due to the low prior probability that two loci chosen at random are linked to each other; that is, about 5% of lod scores of 3 can be expected to actually be false (i.e. not linkage between both loci). When a disease being mapped is non-Mendelian, there may be no way of knowing whether a locus whose effect is of sufficient magnitude to be mapped actually exists. Hence, the posterior false-positive rate in this case may actually be greater than 5%.

Conclusions

Genetic linkage studies to identify the genes underlying psychiatric disorders are truly in their infancy. One major limitation to such studies, the lack of informative genetic markers, has now been eliminated through the development of a very large number (thousands) of markers based on modern, molecular genetic technology, such as restriction fragment length polymorphism (RFLPs), variable number of tandem repeat loci (VNTRs), dinucleotide, trinucleotide and tetranucleotide repeats, and microsatellite DNA sequences. Hence, the lack of infor-

mative markers is no longer the limiting factor. For psychiatric disorders, the most important limiting factor will probably be the number of potential genes involved, and the lack of single loci with very large effect that occur with any relevant frequency. As yet, no clear Mendelian forms of psychiatric illness have been defined. In general, the genetic basis may involve at least several genes of only mild influence with interactive effects. If this is true, then a large number of families will have to be studied before a reliable linkage finding can be obtained and replicated. There is a lack of well-controlled, large-scale family studies of the major psychiatric disorders with systematic interviews of first- and second-degree relatives. When taken together with twin studies, these studies may clarify the patterns of transmission of these complex conditions, yield information of clinical subtypes that may adhere to Mendelian patterns of transmission, and elucidate the mechanisms for comorbidity between disorders. The results of these studies would also provide the basis for selecting the disorders to which the application of linkage studies would be most likely to succeed in the identification of genes. Although these large-scale studies are extremely laborious, they are an important source of information in identifying the complexity of the major psychiatric disorders.

Acknowledgements. This work was supported in part by U.S. Public Health Service, National Institutes of Health Research Scientist Development Awards HG00438 (Dr. Risch) and MH00499 (Dr. Merikangas).

Adapted in part from: Risch N. Mapping genes for psychiatric disorders. Gershon ES, Cloninger CR (eds) Genetic approaches to mental disorders. Washington D.C.: American Psychiatric Association Press (in press).

References

Barbosa J, King R, Noreen H, Yunis EJ (1977) The histocompatibility (HLA) system in juvenile insulin-dependent diabetic multiplex kindreds. J Clin Invest 60:989–998

Barbosa J, Chern MM, Anderson VE, et al (1980) Linkage analysis between the major histocompatibility system (HLA) and insulin dependent diabetes in families with patients in two consecutive generations. J Clin Invest 65:592–601

Beaudet A, Bowcock A, Buchwald M, et al (1986) Linkage of cystic fibrosis to two tightly linked DNA markers: Joint report from a collaborative study. Am J Hum Genet 39:681–693

Bell GI, Xiang K, Newman MV, et al (1991) Gene for non-insulin dependent diabetes mellitus (maturity-onset diabetes of the young subtype) is linked to DNA polymorphism on human chromosome 2q. Proc Natl Acad Sci USA 88:1484–1488

Bishop DT, Cannon-Albright L, McLellan T, et al (1988) Segregation and linkage analyis of nine Utah breast cancer pedigrees. Genet Epidemiol 5:151–169

Claus EB, Risch N, Thompson WD (1990) Age of onset as an indicator of familial risk of breast cancer. Am J Epidemiol 131: 961–972

Claus EB, Risch N, Thompson WD (1991) Genetic analysis of breast cancer in the cancer and steroid hormone study. Am J Hum Genet 48:232–242

Easton DF, Bishop DT, Ford D, Crockford GP, and the Breast Cancer Linkage Consortium (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. Am J Hum Genet 52:678–701

Fajans SS (1990) Scope and heterogeneous nature of MODY. Diabetes Care 13:49–64

Froguel PH, Vaxillaire M, Sun F, et al (1992) Close linkage of glucokinase locus on chromosome 7p to early onset non-insulin-dependent diabetes mellitus. Nature 356:162–164

Goate AM, Hayns AR, Owen MJ, et al (1989) Predisposing locus for Alzheimer's disease on chromosome 21. Lancet 1:352–355

Gusella JF, Wexler NS, Conneally PM, et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306:234–238

Hall JM, Lee MK, Newman B, et al (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250: 1684–1689

Kinzler KW, Nilbert MC, Li-Kuo S, et al (1991) Identifikation of FAP Locus Genes from Chromosome 5q21. Science 253:661–665

Kunkel LM, et al (1986) Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. Nature 322:73–77

McKusick VC (1990) Mendelian Inheritance in Man, Ninth Edition. The Johns Hopkins University Press, Baltimore

Merikangas KR (1987) Genetic epidemiology of psychiatric disorders. In: Frances AJ, Hales RE (eds) American psychiatric association annual review, vol 6. Psychiatry update. American Psychiatric Press, Washington, DC

Newman B, Austin MA, Lee M, King MC (1988) Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. Proc Nacl Acad Sci USA 85:1–5

Nishisho I, Nakamura Y, Miyoshi Y, et al (1991) Mutations of Chromosome 5q21 Genes in FAP and Colorectal Cancer Patients. Science 253:665–669

Peltomäki P, Aaltonen LA, Sistronen P, Pylkkänen L, Mecklin JP, Järvinen H, Green JS, Jass JR, Weber JL, Leach FS, Petersen GM, Hamilton SR, Chapelle A de la, Vogelstein B (1993) Genetic Mapping of a Locus Predisposing to Human Colorectal Cancer. Science, 260:810

Risch SS (1990) Mapping genes in diabetes: genetic epidemiological perspective. Diabetes 39:1315–1319

Risch N (1987) Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 40:1–14

Risch N (1990a) Linkage strategies for genetically complex traits. I. Multilocus models. Am J Hum Genet 46:222–228

Risch N (1990b) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am J Hum Genet 46:229–241

Risch N, Ghosh S, Todd J (1993) Statistical evaluation of multiple locus linkage data in experimental species and relevance to human studies: application to NOD in mouse and human IDDM. Am J Hum Genet (in press)

Schellenberg GD, Bird TD, Wijsman EM, et al (1988) Absence of linkage of chromosome 21q21 markers of familial Alzheimer's disease. Science 241:1507–1510

Schellenberg G, et al (1992) Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. Science 258: 668–671

Schwartz AG, King MC, Belle SH, et al (1985) Risk of breast cancer to relatives of young breast cancer patients. J Natl Cancer Inst 75:665–668

St. George-Hyslop PH, Tanzi RH, Polinsky RJ, et al (1987) The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science 235:885–890

Todd JA, Aitman TJ, Cornall RJ, et al 61991) Genetic analysis of autoimmune type 1 diabetes mellitus in mice. Nature 351:542–547

Willams WR, Anderson DE (1984) Genetic epidemiology of breast cancer: segregation analysis of 200 Danish pedigrees. Genet Epidemiol 1:7–20