Association versus linkage studies in psychosis genetics

One after another monogenic diseases are being successfully mapped within the human genome. In a number of cases, knowledge of the localisation has led to positional cloning of the genes. However, in diseases with a complex mode of transmission, such as the major psychoses, linkage studies with DNA markers have been far less conclusive than initially hoped. Nevertheless, genetic factors certainly exist in the major psychoses, as twin, family, and adoption studies consistently show. These genetic factors must be represented in variation of expressed or regulatory DNA sequences. In complex diseases the application of the linkage strategy may be hampered by both theoretical and practical problems. In some instances it might be more straightforward to use population samples and look for an association between DNA variants and disease. It is a major issue of current discussion to what extent association studies might complement the linkage approach in order to open the black box of psychosis genetics.

Linkage analysis in complex traits
In mendelian traits linkage is inferred from cosegregation of a genetic marker and the gene in question where the observed likelihoods are compared with the expected likelihoods assuming no linkage. The lod score is a measure of the likelihood of linkage. Linkage studies in monogenic diseases using maximum likelihood methods are based on precise specification of the genetic model. Genetic analysis of complex traits may be complicated by various factors such as incomplete penetrance, multiple disease susceptibility loci, epistasis, variable age of onset, assortative mating, gene-environment interaction, mode of inheritance, proportion of phenocopies, and diagnostic uncertainties. Classical segregation analysis to sort out these factors appears hopeless, because there are seldom enough family data to estimate the large number of unknown parameters required to accurately model the complex trait. Chromosomal exclusion maps that were successfully used to pin down gene loci in a variety of mendelian traits are therefore of questionable value for complex phenotypes. The results are only valid if the applied genetic assumptions are correct.

In the last few years a series of simulation and modelling studies have been carried out aiming to examine the power to detect linkage in families. In this context a major issue was the potential influence of genetic heterogeneity. These studies focused on the problem that the combination of families with different genetic bases renders it more difficult to find linkage (interfamilial heterogeneity). Given the existence of more than one disease locus, however, sampling of families with higher numbers of affected subjects ('high density' pedigrees) also increases the proportion of families with intrafamilial heterogeneity, especially when the disease forms are relatively similar in frequency.

Because of the problems with model specification encountered in maximum likelihood methods, non-parametric linkage approaches have been developed for the analysis of complex disorders. Application of the affected-sib-pair (ASP) method, based on a concept originally proposed by Penrose, does not require specification of the genetic model underlying the disease. In this approach one compares the distribution of the number of haplotypes shared by descent between two or more affected sibs with the theoretical distribution of shared haplotypes under the assumption of no linkage. The power of the ASP method to find linkage for a complex trait depends on a number of factors: number of sibships, degree of genetic heterogeneity (number of loci involved), risk ratio of the risk to sibs of a proband versus population prevalence, and informativeness of the marker, the recombination fraction, and the degree of heterogeneity. Since not only genetic heterogeneity but also assortative mating affect the power, study designs based on data from sib trios with one or two affected members have been recommended. Weeks and Lange extended the affected-sib-pair method of linkage analysis to larger pedigrees. This affected-pedigree-member (APM) method has been successfully applied to Alzheimer’s disease, bilateral breast cancer, and early onset breast cancer. However, it is worth pointing out that the APM method depends on shared alleles which are identical by state. The method is therefore sensitive to assumptions about gene frequency and can produce evidence suggestive of ‘linkage’ even in uninformative matings. Furthermore, as in maximum likelihood methods, the power of the APM method is limited by underlying genetic heterogeneity.

The simplicity of the association approach
Association studies proved to be particularly useful in diseases in which certain HLA alleles
are directly or indirectly involved in the aetiology. When molecular genetic polymorphisms are applied, a candidate gene for the disease in question is usually selected and a nearby polymorphic marker is used as a tool to compare patients and controls. The contribution of a marker allele is expressed as the relative risk (RR) that carriers of a certain allele or genotype have of developing a disease as compared to non-carriers. In association studies the RR is practically determined as the odds ratio, that is, the cross product ratio.

To understand the results of association studies one has to understand the rationale behind these studies. True association between a marker and a disease may occur for two reasons. (1) There is strong linkage disequilibrium between a disease causing locus and the locus under study. (2) The marker locus itself contributes to the pathogenesis of the illness. Restriction fragment length polymorphisms investigated to date in the major psychoses are almost exclusively located in non-coding regions in close proximity to suspected candidate genes. It is unlikely that the polymorphic DNA sequence itself causes the disease susceptibility, since the relationship between marker and disease susceptibility can only be interpreted in terms of linkage disequilibrium. Linkage disequilibrium is said to be present when certain alleles at one locus occur with certain alleles at another locus on the same chromosome at frequencies greater than can be attributed to chance alone. In practice, linkage disequilibrium can only be expected to occur under certain conditions, such as low mutation rate, close proximity between marker and disease locus to prevent frequent recombination events, and selective advantage of the haplotype containing the disease gene.

Sobell et al. pointed out the advantages of investigating prevalences of sequence variations affecting protein structure or expression (VAPSE) in large samples of unrelated patients and controls. Obviously, this approach does not rely on the existence of linkage disequilibrium between marker and disease locus, because the polymorphic DNA sequence itself is the object of interest. In this case the finding of an association is not interpreted as indirect evidence for linkage between marker and a nearby disease mutation. The sequence variation itself is regarded as potentially influencing susceptibility to a disease and an association study provides a powerful tool for investigation.

A recent example is the examination of the debrisoquine hydroxylase gene polymorphism (CYP2D6) in Parkinson’s disease. Subjects carrying two mutant alleles leading to the poor metaboliser phenotype have a relative risk of 2.5 to 2.7 of developing Parkinson’s disease. It is questionable whether such an association would ever be detected by family studies.

The availability of sequence information of cloned genes coding for CNS proteins offers the possibility of systematically looking for base pair variation in the population. Recently, allelic variations in the coding sequence of the human dopamine D4 receptor were found. The cloned receptor variants showed different properties with respect to clozapine and serpine binding. Variation of this locus may determine individual differences in susceptibility to neuropsychiatric diseases or responsiveness to antipsychotic medication or both. Another example is the sequence variation leading to an amino acid exchange in the N-terminal end of the dopamine D3 receptor protein.

Power to detect linkage as a function of sample sizes for six values of relative risk (RR). From right to left: RR = 2.5, 5, 7.5, 10, 12.5, 15. Dominant and recessive models with different allele frequencies are assumed.
Genetic variation is likely to exist in other DNA sequences coding for receptors, enzymes, or structural proteins. Association studies offer an easy way of studying their role in the pathogenesis of manic depression, schizophrenia, or any other puzzle of schizophrenia and affective disorder having a substantial genetic contribution.

In this context we felt it would be useful to examine the power of linkage studies to detect susceptibility genes when association has been found (figure A–D). To test for linkage we used the affected-sib-pair method, which is commonly applied in the analysis of non-mendelian traits. As a rule of thumb, sample sizes below 100 sib pairs have very limited power to detect a disease gene, which leads to a relative risk of less than 5. For example, the debrisoquine hydroxylase association in Parkinson’s disease (RR = 2.6) would escape detection even in large scale affected-sib-pair studies. What especially weakens the power of the linkage approach in this context is the nature of a susceptibility gene. The allelic form, which predisposes to the disease, is neither necessary nor sufficient to develop the disease, but increases the chance for a subject to be found in the patient group. However, a certain proportion of patients have no associated allele. In this case the segregation of the alleles will be random.

One major problem that has been encountered in association studies besides the reliance on linkage disequilibrium is the effect of population stratification. It is well known that minor biases in sampling can produce errors in estimates of the allele frequencies because allele frequencies can vary considerably between populations. To avoid stratification effects patients and controls have to be carefully matched regarding their ethnic, geographical, and social background. Falk and Rubinstein19 20 have proposed an ingenious solution to the problem of matching in association studies by examining not only the patients but also their parents. The control sample is derived from the parental alleles not inherited by the proband. Thus, the method avoids the problem of population stratification. It can be used to test a qualitative trait for association with any codominant marker system regardless of its degree of polymorphism.

About 30,000 genes are expressed in the brain12 of which 55 to 90% are brain specific. Since all of them are more or less potential candidates in the psychoses, the sheer number may be frightening. However, some of them may be favoured because of pharmacological, biochemical, or other data. Furthermore, the improvement of screening methods for uncovering sequence variation and the pace of large scale sequencing efforts at present continuing should make it feasible to look for expressed variation in many genes. The implications of such DNA sequence variations should be examined in panels of independent psychotic patients and controls as well as in Falk-Rubinstein trios.

The controversial data generated by applying molecular genetic methods in psychosis research should not be regarded as casting doubt on the genetic hypothesis in the aetiology of psychoses rather than on the methods applied so far. Compared to the powerful methods of molecular genetics there is no alternative approach that can be convincingly applied to the puzzle of schizophrenia and affective disorder. If there are single major genes implicated in the development of the major psychoses, large scale linkage analyses will probably be able to pinpoint those genes, as has been successfully done for diseases such as Alzheimer’s disease and familial breast cancer. On the other hand, if major genes are only responsible for a minority of cases, association studies offer a promising tool to identify susceptibility genes and by this means shed light on the complex nature of gene involvement in the aetiology of the major psychoses.

We thank Drs J. Körner, M. Rietschel, and J. Hebebrand for helpful discussions. This work was supported by the Deutsche Forschungsgemeinschaft and a Centre grant from the University of Bonn Medical Centre and the State of Nordrhein-Westfalen.

MARKUS M NÖTHEN
PETER PROPPING
Institute of Human Genetics,
University of Bonn,
Wilhelmsstrasse 31,
5300 Bonn 1,
Germany.

ROLF FIMMERS
Institute for Medical Statistics,
University of Bonn,
5300 Bonn-Venuela,
Germany.

Association versus linkage studies in psychosis genetics