Molecular genetics and the major psychoses

M GILL and C WALSH

The hypothesis that the major psychoses, schizophrenia and manic depression, are largely genetic disorders is supported by an impressive body of data. What remains in question is the extent and mode of transmission of that genetic contribution, which genes are involved, and how DNA mutations might interact with the environment to produce the constellation of symptoms known as schizophrenia and manic depression.

Defining the phenotype

Defining the phenotype for genetic study presents a number of problems for the major psychoses. First, diagnoses which depend on the identification and assessment of diverse groups of clinical symptoms may not define discrete phenotypes in the genetic sense, and what is considered a clinically homogenous entity may in fact represent several disorders of distinct etiology. Indeed, in the case of schizophrenia, a number of subtypes have been proposed which, if valid, are likely to differ in the extent to which genetic and environmental factors contribute to their etiology[1,2].

Secondly, it is not clear which related diagnoses should be included in the phenotypic definition. In the case of schizophrenia, family studies have shown that non psychotic diagnoses such as schizotypal or paranoid personality disorder cluster within the families of schizophrenic probands (for review see Leviston and Mowry[3]). It is likely, therefore, that these 'spectrum disorders' are genetically related to schizophrenia and may represent a modified expression of the same disease gene or genes. A similar situation arises in the case of major depression and cyclothymic personality in relation to manic depression. However, the specificity with which these disorders represent the phenotype is low. Further doubt arises with regard to the status of schizoaffective disorder, found in families of both schizophrenia and bipolar probands. On the basis of this observation, it has been suggested that the two disorders are not distinct from each other, but rather that they occupy polar

All correspondence to: Dr M Gill, Genetics Section, Institute of Psychiatry, De Crespigny Park, Denmark Hill, London SE5 8AF, UK.

Cambridge Medical Reviews: Neurobiology and Psychiatry Volume 2

positions on a disease continuum. The decision as to where the phenotypic boundary exists may therefore be an arbitrary exercise, but have serious implications for the interpretation of genetic studies.

Finally, there is a large literature describing how the major psychoses can be mimicked by organic conditions[4–6]. The exact prevalence of such 'phenocopies' is unknown; however, one study[7] identified underlying organic pathology in as many as 6% of patients with 'functional' psychoses.

Genetic investigations of the psychoses would be considerably enhanced by the identification of stable biological markers or 'endophenotypes' which distinguish etiologically homogenous forms of disorder. Various neuropsychological and neurophysiological measures, eg smooth eye pursuit movements[8], have been proposed but none is, as yet, generally applicable.

Clinical genetics of the major psychoses

Since the early 1900s, family studies have consistently shown that schizophrenia clusters within families[9,10]. In addition, morbid risk is found to correlate with genetic proximity to an affected individual. A 10% lifetime risk of developing schizophrenia in first degree relatives, and a 3% risk in second degree relatives, contrasts with a risk of 1% in the general population[10]. Adoption studies indicate that this increased liability is the result of shared genes and not shared environment[11,12]. Furthermore, pooled data from twin studies demonstrate illness concordance rates of 53% in genetically identical monozygotic twins in contrast to 15% in dizygotic pairs[13].

Recent work suggests the possibility that environmental hazards are of greater relevance to the development of schizophrenia in young males whilst genetic factors may be of greater importance in females[14–16]. Variables other than gender may also influence risk. For example, some studies show that early onset in schizophrenic probands is associated with greater morbidity in relatives[16], although this is not a consistent finding[17]. This may suggest that age of onset and liability to develop schizophrenia are associated. Alternatively, it may indicate etiological heterogeneity, with families of early onset probands expressing a more highly genetic form of the disease.

Similar studies conducted for manic depression demonstrate a strong genetic influence in that disorder. An 8% lifetime risk of illness in relatives of manic depressive probands represents a marked increase above a population risk of 0.5%. Adoption studies support the genetic hypothesis as do twin studies, with MZ twins up to three times more likely to show illness concordance than DZ twin pairs[18].

Proposed modes of inheritance

In the case of Mendelian disorders such as Huntington's disease and cystic fibrosis, information regarding the mode of transmission is easily obtained from family data. The transmission of other disorders may resemble

Mendelian inheritance but with minor transgressions; a number of individuals with a disease genotype may fail to manifest the disorder (reduced penetrance) or some individuals may express the disorder in an attenuated form (variable expression).

In yet another group, genetic factors are known to be of etiological importance but there is no adherence or resemblance to Mendelian transmission. These are 'complex genetic disorders' and include such conditions as coronary heart disease, diabetes, many autoimmune disorders and the major psychoses. For these, the mode of inheritance is likely to be obscured by variable age of onset, reduced penetrance, nongenetic cases (phenocopies), environmental influences, unclear phenotypic boundaries and genetic heterogeneity.

Mathematical models of disease transmission

Mathematical models of disease transmission predict patterns of familial aggregation which can be tested against existing family data. This data may take the form of morbid risk figures for various classes of relatives (prevalence analysis), or consist of information on the segregation of illness within entire pedigrees (segregation analysis)[19]. In general, segregation analysis is favoured for its greater statistical power.

If values predicted by a model are significantly different from observed findings, the model does not fit, and is rejected. A more problematic and not uncommon situation arises when several models are equally acceptable, each providing a moderate fit. Models considered in this manner, some of which incorporate a liability threshold construct[20], include the generalized single major locus (GSL) and the multifactorial threshold (MFT) models. For the GSL model, liability to develop a disorder is conferred by a single gene whose penetrance can take any value between 0 and 1. In contrast, liability in the MFT model is determined by both polygenes and environmental factors. The GSL and MFT models are not mutually exclusive and a 'mixed' model has been proposed[21] in which both major genes and polygenes play a role.

Genetic transmission of schizophrenia

Some studies have found compatibility between the GSL model and family data but nearly all noted an underprediction of MZ twin concordance[22–24]. Several other studies were able to reject the model outright[25–27]. Overall, existing evidence is strongly against a single gene effect when schizophrenia is treated as a unitary disease.

Many findings in schizophrenia are, however, more compatible with the MFT model. Morbid risk in relatives increases as a function of the number of affected individuals in the family. It also increases as a function of severity of illness in the proband; this is supported by twin studies which show higher concordance rates when the proband twin has a severe and chronic illness. In

addition, most multifactorial/polygenic disorders occur with a lifetime risk of greater than 1 in 500[28].

Genetic transmission of manic depression

Family data do not consistently support any model of inheritance in manic depression[19]. Early family studies suggested X-linked dominant inheritance[29], demonstrating an excess of affected females and a deficit of father to son transmission. Risch and Baron[30] estimated that up to 30% of the manic depressive population might carry the putative disease gene on the X chromosome (based on linkage results). However, the concept of an X-linked subgroup remains controversial[31] and consistent evidence for autosomal dominant inheritance is also lacking.

Models of transmission: implications for molecular genetics

For the psychoses, as with other complex genetic disorders, there is an unclear relationship between phenotype and genotype, and Mendelian patterns of segregation are not normally observed within families. Undoubtedly, for some of these disorders, Mendelian traits remain to be recognized, having been obscured so far by genetic heterogeneity, late onset or variable penetrance. Whether this applies to the major psychoses remains to be seen.

On the other hand, the major psychoses may more closely resemble traits such as height and intelligence, both with large genetic components. Thus they would be due to the additive effects of many genes, each contributing a minor amount of the total genetic variance. Indeed, many animal behaviours, such as alcohol sensitivity, nest building and aggressiveness appear to be truly polygenic, in that repeated selected breeding of animals for such traits fail to divide them into two behavioural extremes[32]. If only one or two major genes were responsible for the genetic effects of these behaviours, the relevant alleles would be sorted into the high and low lines in a few generations.

Some traits and disorders, previously considered polygenic, are yielding single gene subtypes. Hypertension in rats, for example, can be caused by major genes[33], and a subtype of diabetes in humans is caused by a dominant mutation at the glucokinase gene[34]. Alzheimer's disease, previously considered polygenic, can be caused by single gene mutations, detected with the aid of linkage analysis.

Demonstrating single gene subtypes in many disorders has, however, depended on obvious candidate genes, animal models or clear clinical subdivisions. In diabetes, the mutation on chromosome 7 causes an early onset, noninsulin dependent form of disease[34]. Moreover, the definition of the phenotype is assisted by biochemical tests. In the case of Alzheimer's disease[35], early onset delineated a clinical subtype, and the association with Down's syndrome indicated chromosome 21. The absence of animal models or clinical subdivision for the major psychoses makes the researchers task considerably more difficult. However, recent discoveries in other complex

genetic disorders suggest that for the major psychoses, etiological and genetic heterogeneity are to be expected. Under these circumstances clinical genetic studies and mathematical modelling of disease transmission are unlikely to indicate which molecular biological techniques are appropriate. A broad approach is therefore justified in preliminary investigations. The remainder of this chapter will discuss current molecular strategies and results to date.

The molecular genetics of the psychoses: candidate genes and gene finding techniques

A candidate gene is one for which there is some a priori reason to suspect its involvement in the etiology of the disorder in question. In the case of the major psychoses, there are many genes that fall into the category of 'possible candidates', but few, if any, that fall into the category of 'probable candidates'. Although all brain specific genes could be considered candidates for the major psychoses, choice of genes for study should be theory led. Thus the hypothesis that neurodevelopmental abnormality may lead to schizophrenia suggests a possible role for genes which determine or modify neurodevelopment processes[36].

In manic depression, as in schizophrenia, there is a range of etiological theories; one of these concerns cation transport. It has been suggested that, in normal controls, the enzyme Na+K+ATPase is increased in lymphocytes following incubation in lithium or ethacrynate[37]. This 'upregulation' is absent or attenuated in both euthymic drug-free manic depressive patients, and in those taking lithium. These results confirmed earlier work by Naylor and Smith[38]. The authors suggest that this altered response of the enzyme is an enduring trait marker in manic depression. The enzyme is composed of two subunits, a large catalytic subunit, and a smaller glycoprotein subunit. The catalytic subunit is encoded by at least three separate genes. All of these are potential candidate genes.

Once a particular gene is identified as a candidate, it remains to be assessed if, and to what degree, it contributes to the disease liability. Clarification of a particular gene's involvement can be approached in a number of ways: gene structure and function may be examined directly or, alternatively, the techniques of genetic linkage or association can be employed. One potential advantage of genetic linkage methods is that they can identify the location of abnormal genes without requiring prior knowledge of the disease process.

Candidate genes: functional analysis

Candidate genes may be studied directly, their structure and function compared between patients and controls. Functional abnormalities may occur anywhere from transcription of the gene into RNA, through processing of that RNA, to translation into a protein. The function of a gene is studied by examination of messenger RNA (mRNA). The relative site of action of the

gene in the brain and the abundance of its expression is determined using in situ hybridization with cloned segments of transcribed gene. It is important to note, however, that the function of a set of genes may be disturbed as a secondary consequence of abnormality elsewhere.

For schizophrenia, one set of genes studied in this manner are those receptors stimulated by the excitatory neurotransmitter L-glutamate. This neurotransmitter provides the major afferent input to the hippocampus. Its receptors are divided pharmacologically into 5 classes; *N*-methyl-D-aspartate (NMDA), alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate (AMPA), kainate and 2-amino-4-phsophonobutyrate (AP4), known as ionotrophic receptors and one metabotrophic receptor (mGluR)[39]. In addition, there is evidence that alternative splicing of the mRNA occurs, thereby adding to receptor diversity[40]. Abnormalities in the binding of glutamate receptor ligands in the brains of schizophrenics have been reported and molecular studies are a natural progression from such findings[41]. Recently, Harrison et al.[42] have reported a decrease in mRNA encoding a non-NMDA glutamate receptor in hippocampal tissue from schizophrenic patients compared to normals. Whether this is specific, or part of a more widespread abnormality in gene expression they are unable to say. Nor has a secondary effect of medication been ruled out as an explanation.

Candidate genes: structural analysis

To date, there has been little work examining the structure of any particular candidate gene in patients. One study reported fully sequencing the functionally significant parts of the D2 receptor in 14 schizophrenic patients, but found no structural changes[43].

Gene finding techniques: genetic linkage and its application to the major psychoses

Genetic linkage refers to the observation that two genetic traits may be coinherited rather than independently inherited as Mendel prediction in his second law. If 2 genetic traits are caused by genes which exist close together on the same chromosome then during meiosis, recombination between them will occur only rarely, and they will be passed on to offspring together.

A genetic trait or a disease is the result of a genetic polymorphism; a base pair sequence difference between two homologous strands of DNA that can be detected. In the case of a disease or a trait, the genotypic difference is inferred from the phenotype. In the case of a DNA polymorphism, it is detected by direct examination of the DNA. DNA polymorphisms, or markers, are known throughout the genome, and their segregation in families can be determined and compared with the segregation of a genetic disease. If cosegregation is apparent, then the conclusion is that the disease gene is located close to the marker polymorphism.

The LOD (*Log of the odDs* ratio) score method of Morton is widely used in the assessment of linkage. It is a powerful method, based on the theory of likelihood, using all the information available within large pedigrees. For single gene disorders, a LOD score > 3 is taken as evidence of linkage, and a score of < -2 as evidence of exclusion of linkage; numbers in between are inconclusive.

The LOD score method requires the estimation of genetic parameters describing disease and marker transmission. The appropriate model and parameters are not known for the major psychoses, and the effects of misspecification on the reliability of the LOD score is not certain although such misspecification has been shown to decrease the power to detect, rather than to falsely infer, linkage[45].

The sib-pair method of Penrose[46], and its modern modifications[47,48] may be more robust to misspecification of genetic parameters, and are thought to offer an alternative to the LOD score method in complex genetic diseases. In the affected sib-pair method, if a disease and a genetic marker are truly unlinked then, in a large sample of affected sib-pairs, marker alleles will be shared and unshared in equal proportions. If alleles are shared significantly more than expected, then one conclusion is that linkage exists between marker and disease. With many highly polymorphic markers throughout the genome, 'significant' results will clearly occur by chance. Replication, as with LOD score results, will be essential.

In the major psychoses, the absence of a priori evidence for single gene subtypes and a reliable method of clinical subdivision causes difficulties for the linkage method of analysis. It remains at present no more than an article of faith whether or not single gene transmission exists in a significant proportion of families, multiply affected with one or other of the major psychoses.

Nevertheless, there are two reasons for using linkage analysis to search for genes of major effect. The first, as outlined above, is the evidence from other complex genetic traits that single gene subtypes may be much more common than has been considered to date. The second reason is the enormous potential for a serious breakthrough in our knowledge that may result from identification of such genes, even if mutations in them cause disease in only a minority of large multiply affected families.

It is the proven ability of molecular genetic techniques to facilitate significant advances in knowledge of pathophysiological mechanisms that caused such excitement when initial results in the case of the major psychoses were promising[49], and such disappointment when these results turned out to be false leads[50].

Linkage studies in schizophrenia: results to date

The casual reader in this subject would be forgiven if he thought that linkage studies in schizophrenia began with chromosome 5. One of the earliest

studies showing promising results was a study of HLA genotypes in families multiply affected by 'schizotaxia': a broadly defined clinical phenotype including schizophrenia and schizotypal personality[51]. Four further studies failed to confirm these findings[52–55]. One study employed the affected sib-pair method along with the LOD score method of analysis[52]. Overall, the conclusion was that a 'dominant' gene for schizophrenia was unlikely to be located close to the HLA locus.

The advent of DNA markers provides a catalogue of polymorphic markers scattered throughout the genome. On the basis of a trisomy of part of chromosome 5q apparently cosegregating with schizophrenia in a small family[56], Sherrington et al.[57] examined the segregation of three DNA markers mapping to the region in five Icelandic and two British families. Evidence for linkage was found; it was strongest when a variety of disorders, some probably related to schizophrenia, and some not, were included in the analysis. The quoted maximum LOD score of 6.49 was generated using an extension of the method which allows data from a number of closely linked markers to be used simultaneously. This procedure, although safe for known single gene disorders, can falsely inflate the evidence for linkage and is not recommended for a preliminary analysis. The maximum LOD score between a marker and the disease in the study by Sherrington et al. was 3.96; strong but not conclusive evidence for linkage. Furthermore, if, as has been recommended by some observers[58,59], the maximum 2-point LOD score is corrected to take into account multiple genetic and diagnostic models, then the result would have been reduced by 1.38, ie 2.58.

Simultaneously, Kennedy et al.[60] failed to confirm the findings of Sherrington et al. using a large Swedish family. That this result could have been due to linkage heterogeneity was suggested by Kennedy et al.[60] and Lander[61]. Subsequently, however, five further studies have been reported[62–66] all of which found no evidence of linkage. Finally, the group that produced the initial promising findings have now extended their studies on chromosome 5q by generating new, highly polymorphic markers within the region. They have determined the genotypes at these markers of both the original families and a new collection. None of their recent results confirms previous findings, and, indeed, effectively excludes the region from containing a single gene of major effect in the etiology of schizophrenia[67].

There is no evidence of X-linkage in schizophrenia. However, a role for the sex chromosomes is suggested by findings of an excess of schizophrenia in individuals with sex chromosome aneuplodies and by the finding that pairs of affected relatives are more likely to be of the same, than opposite, sex[68,69].

Crow[70] suggested that these results could be explained if the gene(s) for schizophrenia was located in the pseudoautosomal region of the sex chromosomes. This is a small area located at the distal end of both the X and Y chromosome. It contains sequence homology between the X and Y chromo-

somes and exchange of genetic material takes place during meiosis. Depending on its position within the region, a gene may be transmitted in either a sex-linked or 'pseudoautosomal' manner. In their study of 83 sib-pairs affected with schizophrenia Collinge and colleagues[71] found tentative evidence of linkage to pseudoautosomal markers. However, because of the limited power of the affected sib-pair method, the authors concluded that examination of the hypothesis was required in a larger sample. A second group have also reported evidence of nonrandom segregation of marker alleles to affected sib-pairs[72]. Others[73], while detecting an excess of same sex sib-pairs, found no evidence for linkage in this region.

More recently, there has been some interest in the long arm of chromosome 11. Several families have been reported in which balanced translocations involving chromosome 11q apparently cosegregate with psychotic illness[74–76]. In addition, genes encoding the D2 receptor, porphobilinogen deaminase, and tyrosinase map to this region, and could be considered candidate genes for the schizophrenia. In a systematic search using a large sample of multiply affected pedigrees, Gill et al.[77] found little evidence suggesting the presence of a gene of major effect in this region.

Linkage studies in manic depression

Early linkage studies in manic depression have concentrated on two areas of the genome; the X-chromosome and the HLA region of chromosome 6. The interest in the X-chromosome began when Rosanoff et al.[78] proposed this form of transmission on the basis of their family and twin studies. A number of early studies showed apparent cosegregation (linkage) between the X-linked genetic markers, colour blindness and G6PD[29,79]. Evidence of linkage to the blood group marker Xg was also found[80]. As these two markers are on opposite ends of the X-chromosome linkage to both in the same pedigrees is unlikely. Mendlewicz and Fleiss[81] studied both markers and concluded that a putative X-linked bipolar gene was situated between the markers. Their evidence for colour blindness linkage was stronger than that for Xg. Many further studies ensued, most demonstrating evidence of varying degrees in favour of linkage[82–85], but some showing evidence against X-linkage[86,87]. Detailed analysis of the pedigrees studied by Baron et al.[85] suggest that the LOD scores are robust to a variety of phenotype definitions and to variations in genetic parameters[88]. These authors suggest that the X-linked form of manic depression is characterized by severity, early age of onset, high familial prevalence of the bipolar phenotype and high recurrence rate of major depression.

There have also been reports of linkage to factor IX[89]. These genes are some distance apart, and linkage to both may not be compatible, thus raising the possibility of two X-linked genes or false positive results at one or other locus.

In affective disorders, there appears to have been no good reason to study HLA antigens other than their presence and extreme variability, and thus suitability as linkage markers. In 1981, Weitkamp et al.[94] reported that affected sibling pairs shared HLA haplotypes more often than would be expected by chance. A previous study had reported similar findings[90]. Further studies[91–93] failed to confirm these findings and cast doubt on the validity of the statistical methods of Weitkamp et al.[94].

Evidence for a susceptibility locus close to the tyrosine hydroxylase gene on chromosome 11 in a large Amish pedigree[95] was reported, but was not supported by studies of other pedigrees[96–98]. This was initially interpreted as evidence of genetic heterogeneity, but reanalysis and extension of the Amish pedigree led to a drop in the LOD score, thus casting doubt on the original report[99].

Genetic association and its use in the major psychoses

At present, there is no obvious method of subdividing the major psychoses such that one or more subdivisions are more likely (a) to be etiologically homogeneous or (b) to demonstrate Mendelian transmission. Applying linkage methods to the whole may fail if there is significant etiological or genetic heterogeneity. An alternative method of genetic analysis, less dependent on the underlying disease model, is to test for association between particular alleles of a polymorphism and the disease phenotype. This method requires a large sample of unrelated affected individuals and has been successful for HLA-associated diseases[100], for diseases associated with red blood cell surface antigens[101] and for heart disease[102]. DNA polymorphisms have greatly expanded the scope of these studies, and are available within, or close to, any gene of interest.

Genetic association is thought to arise as follows. If a mutation contributes to a disease then, clearly, it will occur more frequently in patients than controls. However, closely surrounding DNA will also tend to occur more frequently in patients. If, within this surrounding region, there is another polymorphism then certain alleles at that polymorphism will also occur more frequently in patients than controls. This phenomenon, also known as linkage disequilibrium, is often detected between DNA markers from the same location[103], and is very common between HLA classes. Whether two polymorphisms are indeed in linkage disequilibrium depends on the period of time since both polymorphisms arose, the distance between them and the frequency of recombination in the region.

Ferns et al.[104] describe DNA polymorphisms within or near genes as simply markers which may or may not be in linkage disequilibrium with gene variations (mutations) that may predispose to disease. Sobell et al.[105] point out that the strategy of testing markers within or near genes for evidence of allelic association with the disease phenotype depends on the presence of linkage