Modern molecular genetic approaches to psychiatric disease

Nick Craddock and Michael J Owen
Departments of Psychological Medicine and Medical Genetics, University of Wales College of Medicine, Cardiff, UK

Major advances in molecular genetics over the last 15 years have made it possible to identify the genes responsible for human diseases using purely genetic approaches that do not require knowledge about disease pathophysiology. Many successes have been achieved for single gene disorders and methods are being adapted and refined for complex diseases. The main strategies include linkage and association studies to map the position of disease genes followed by investigation of potential candidate genes within these genomic regions. Successes have already been achieved in complex disorders such as diabetes and Alzheimer's disease, and it is almost certain that genes predisposing to the major psychiatric disorders will be identified over the next few years. This will lead to major advances in treatment, prevention and classification of mental illness and is likely to have a dramatic impact on clinical practice.

When finally interpreted, the genetic messages encoded within our DNA molecules will provide the ultimate answers to the chemical underpinnings of human existence. They will not only help us understand how we function as healthy human beings, but will also explain, at the chemical level, the role of genetic factors in a multitude of diseases, such as cancer, Alzheimer's disease, and schizophrenia, that diminish the individual lives of so many millions of people. James D. Watson (1990)¹

Family, twin and adoption studies provide evidence of genetic predisposition to many psychiatric disorders including schizophrenia, bipolar disorder, unipolar depression, panic disorder, obsessive-compulsive disorder, personality disorder and alcoholism². The purpose of this chapter is to outline some of the ways in which the power of modern molecular genetics can be harnessed to identify the genetic mechanisms involved.

The advances in molecular genetics over the last 15 years have made it possible to investigate the pathophysiology of diseases having a genetic
basis using purely genetic approaches that do not require prior knowledge about disease pathophysiology (although such knowledge usually facilitates the task of identifying disease genes)\textsuperscript{3,4}. If sufficient family material is available, it is now only a question of time until the gene responsible for a specific single gene disorder is identified.

These successes are leading towards important advances in both treatment and prevention of those diseases for which genes have been identified, as well as adding to our understanding of normal functioning of biological systems. However, single gene diseases are rare and account for a relatively small proportion of total morbidity in human populations. The main challenge to molecular geneticists over the next decade is identification of genes predisposing to more common but ‘complex’ genetic disorders, such as diabetes, coronary artery disease and the major forms of mental illness, which do not have simple modes of inheritance. Success in identifying these genes offers the potential of enormous benefits. Improved understanding of disease pathophysiology, which in many cases is obscure, will lead to the development of more effective methods of treatment and prevention and lead to a major reduction in morbidity in human populations.

In this paper we shall first outline some of the methods and resources currently available to molecular geneticists and then show how they have been used with often dramatic success in single gene disorders. We shall then consider the ways in which apparently complex and non-Mendelian patterns of inheritance can arise. Next, we shall discuss how molecular genetic approaches are being refined and adapted to study complex diseases using Alzheimer’s disease, bipolar disorder and schizophrenia as examples relevant to biological psychiatry. Finally we shall speculate about future advances.

**Some tools and terminology of the new genetics**

A concise account of techniques available to modern molecular geneticists is given by Nicholl\textsuperscript{5} and an authoritative but readable introduction is given by Watson \textit{et al.}\textsuperscript{6}. In this section, we outline some of the more important terminology and tools of the trade. We assume some basic knowledge of the structure and properties of DNA, RNA and proteins.

Restriction endonucleases are enzymes that cut DNA at specific base sequences or recognition sites. Several hundred such enzymes, cutting at different sequences, are known. Human DNA can, thus, be cleaved into fragments of manageable size (typically 1000–10,000 base pairs, (bps) in length) having known sequences at their ends. Ligases are enzymes that
join together pieces of DNA with complementary sequences at their ends. Specific segments of DNA can, therefore, be removed from and repositioned into other DNA segments. DNA segments can be inserted into a vector, for example a bacterial plasmid, which can be introduced into bacteria which, in turn, will replicate to produce colonies of millions of bacteria containing the additional DNA segment. Resultant DNA molecules are known as recombinant molecules since they contain DNA from different sources. Multiple copies of the original DNA segment can be recovered by using a restriction enzyme to cut out the fragment. Such procedures are known as cloning and when used for gene sequences give origin to the emotive-sounding term gene cloning. The representation of a particular source of DNA as a large number of cloned inserts is called a DNA library. Libraries provide genetic material in packages of manageable size and in a reproducible and manipulable form. Techniques are available to determine the base sequence of any purified piece of DNA by either chemical or enzymatic means. Sequences of DNA can be used as probes to identify the occurrence of complementary sequence in fragments of DNA.

The polymerase chain reaction (PCR) is a technique having many important applications within molecular genetics. It allows a sequence of interest to be amplified selectively against a background of a large excess of irrelevant DNA. A sequence of up to 5000 bps or more can be amplified $10^5$–$10^6$-fold. PCR requires that unique sequences flanking the sequence of interest are known, so that specific oligonucleotide primers can be constructed which define the region to be amplified.

There is enormous variation of DNA sequences within human populations. Within genes, variation includes disease mutations and normal polymorphisms (common variants that usually do not have pathological significance). However, the greatest variability occurs within DNA that does not code for a protein and which comprises the vast majority of the genome. In general, sequence variations in these regions have no relationship to disease but they are immensely useful as genetic markers that can be used to map the position of disease genes in the genome using linkage and association studies. The aim of both types of study is detection of non-random sharing of alleles at a genetic marker in individuals affected by disease which may indicate close proximity of marker and disease gene.

The most useful class of DNA markers currently available is simple sequence repeats (also called microsatellites) which comprise repeated units of a simple DNA sequence (e.g. the dinucleotide, CA) for which the number of repeats varies within the population. Maps of dinucleotide repeat markers are available that span the genome at approximately 1 cM intervals. Genotypes at such markers are readily assayed by PCR using oligonucleotide primers that are labelled either by radioactivity or a fluorescent dye. Another common class of markers, restriction fragment...
length polymorphisms, RFLPs (which are often due to mutations that cause a change in a restriction enzyme recognition site) are usually assayed using a more laborious procedure known as Southern blotting.

Transcript maps are an intermediate stage on the way to a complete map of all human genes. They are maps of unique DNA sequences known as expressed sequence tags (ESTs). ESTs are made by converting cellular mRNA molecules into complementary DNA (cDNA) (using reverse transcriptase) and sequencing part of the cDNA. Each EST, therefore, consists of a small stretch of known DNA sequence that corresponds to part of a gene and which serves as a tag signifying the presence of that gene. ESTs are generally free from polymorphism and are, therefore, not of use in initial localization of disease genes. The benefit of a map of ESTs is that known genes recognizable by ESTs can be identified in the neighbourhood of any marker linked to a disease gene.

As more genes are identified, complete sequence information on genes is becoming available. Such data are stored on computer databases and software allows prediction of the amino acid sequence of the encoded protein as well as its three dimensional structure and likely function. Such databases, which are readily accessible via Internet, allow comparison of new sequences with those of known genes to determine sequence similarity (homology). This can help determine whether the new sequence is part of a gene and allow guesses to be made about its possible function.

Methods are available to locate the physical position of DNA sequences on the genome, e.g. fluorescent in situ hybridization (FISH), in which a DNA sequence of at least 1 kb is labelled with fluorescent dye and allowed to bind to its complementary sequence on a chromosome spread. The site of binding can then be visualized directly by microscopy.

All these developments are occurring within the context of the Human Genome Project, a large scale International scientific endeavour set up at the end of the 1980s with the objective of sequencing the human genome, as well as the genomes of several 'model' organisms (the bacterium, Escherichia coli, the roundworm, Caenorhabditis elegans, the fruit fly, Drosophila melanogaster, a yeast, Saccharomyces cerevisiae, and the mouse). The pace of technological development has been such that progress at each stage has exceeded targets and it is likely that the vast majority of human genes will have been identified and sequenced by early next century.

**Measures of distance**

Three common but distinct measures of the separation of two loci are used in genetics: Recombination fraction, $\theta$, is the probability that two loci will be separated by recombination at meiosis. Map distance,
measured in centiMorgans, cM, is the expected number of recombinations occurring between two loci at meiosis. Physical distance is measured in base pairs, bps, of DNA. The relationship between these measures is non-linear and variable in different parts of the genome and between the sexes but for small separations there is an approximate equivalence between $\theta = 0.01$, map distance $= 1$ cM and physical distance $= 10^6$ bp (one Megabase). The haploid human genome comprises approximately $10^9$ bps and has a sex-averaged map length of approximately 3300 cM.

Mapping genes for disorders with simple Mendelian inheritance

Simple genetic disorders are typically rare and follow simple Mendelian patterns of inheritance. Examples include cystic fibrosis which is autosomal recessive and affects 1 in 2000 live births$^9$ and Huntington’s disease which is autosomal dominant with a prevalence of 1 in 10,000$^{10}$. It is to disorders such as these that molecular genetic techniques have been so successfully applied and the approaches taken will form the foundation for current work in complex disorders.

Functional cloning

Single gene diseases are caused by mutations in the coding or control sequences of a gene which alter the structure or expression of the gene’s protein product. In the case of a disease for which the biochemical basis (i.e. the protein abnormality) is known it is possible to determine the responsible gene by searching the genome for the gene that encodes the relevant protein. This approach, called ‘forward genetics’ or ‘functional cloning’, has been used to identify genes involved in diseases such as phenylketonuria and haemophilia A$^4$.

Positional cloning

For many diseases, the biochemical basis is unknown and the forward approach not possible. For such diseases, the positional cloning approach (previously known as reverse genetics) has produced many successes$^{11}$. This depends upon determining the genomic location of the relevant gene using purely genetic methods and without prior knowledge of disease pathophysiology. The first stage is usually linkage analysis using DNA markers in extended pedigrees containing multiple cases of disease.
Typically, a set of evenly spaced markers is examined and statistical (lod score) analyses performed to detect non-random sharing of marker alleles between affected members of each family\textsuperscript{12}. Cosegregation of marker and disease alleles (usually different marker alleles in different families) in affected individuals is termed genetic linkage and indicates that the causative gene lies near to the marker. After preliminary localization, additional markers in that region are typed and the location of the disease gene is determined more precisely by further linkage analysis. If sufficient family material is available, the disease gene can, typically, be localized to within 1–2 cM using such an approach. The next stage is usually to search for allelic association between the disease and one or more alleles of markers in the region using samples of unrelated patients and unrelated controls from a genetically homogenous population. Such a finding may indicate the presence of linkage disequilibrium due to the extreme proximity of the disease gene to the marker. Such linkage disequilibrium occurs when marker and disease gene are so close together that the disease mutation is passed through many generations of a population without being split from the marker allele with which it was originally paired. Linkage disequilibrium is not always present but under favourable circumstances, this method can localize a disease gene to within less than 0.1 cM. Thus, purely positional approaches can localize a disease gene to a region of between 2 million and 100,000 bases (2 Megabases–100 kilobases) – much less than the total genome (approximately 3000 Megabases) but still a large region which may contain several hundred genes.

The next stage is to examine directly the DNA in this region in order to detect genes that might be involved in the disease. This stage is called physical mapping. Libraries of DNA are searched and overlapping clones ordered so that a contiguous representation of the region is obtained. A variety of methods can then be used to identify genes in the region. These include searching for sequences that are: (a) highly conserved across species; or (b) bind mRNA; or (c) contain an open reading frame (i.e. sequences that follow a start codon and/or do not contain stop codons); or (d) likely to be coding portions of genes or exons (so-called exon trapping); or (e) homologous to genes that have already been sequenced. Genes that are identified must then be examined in samples of patients and normal controls to determine whether mutations in that gene are pathogenic. It is necessary to demonstrate that mutations occur in the gene in patients but not controls and that the mutation segregates with disease within families. Even once this has been achieved, causation is not proven until the consequences of the mutation have been studied in a functional system such as a transgenic animal. Examples of successful applications of these laborious methods are provided by cystic fibrosis\textsuperscript{6} and the identification of mutations in the Presenilin-1 gene on
chromosome 14 in the majority of families with autosomal dominant familial Alzheimer’s disease

Use of cytogenetic abnormalities to provide positional information

Identification of a cytogenetic abnormality, such as a deletion or translocation, associated with disease can provide a rapid route to isolating the disease gene. The site of the abnormality provides positional information to guide the choice of genetic marker. Further, the abnormality itself can often be exploited in an elegant manner using special techniques that circumvent much of the laborious work of positional cloning in the absence of cytogenetic abnormality. An example is Duchenne muscular dystrophy, the gene for which was isolated following investigation of a boy who had a deletion of his X chromosome that resulted in him suffering simultaneously from Duchenne muscular dystrophy, chronic granulomatous disease and retinitis pigmentosa (three X-linked single gene disorders caused by mutations or absence of genes in the deleted region).

Candidate gene approach

Candidate genes are those which are known, or suspected, to have a role in the pathophysiology of the disease. For example, in the case of hereditary retinal degeneration, numerous genes encoding proteins involved in phototransduction had already been cloned. Investigation of one of these, a rhodopsin gene, led to identification of mutations in some patients with autosomal dominant retinitis pigmentosa.

Positional candidate approach

As more genes are being mapped, identified and sequenced, a combination of positional cloning and candidate gene approaches is becoming the method of choice for identifying disease genes. This positional candidate approach relies on a combination of mapping to the correct chromosomal subregion (generally by linkage analysis) followed by a survey of the interval to see if it contains attractive candidates. An example of this approach is provided by Marfan syndrome which was initially mapped to chromosome 15q by linkage analysis. The connective tissue protein fibrillin was almost simultaneously but independently mapped to 15q21.1 by in situ hybridization, making it an excellent candidate.
Modern molecular genetic approaches to psychiatric disease

candidate gene. Investigation of the structure of the fibrillin gene in normal individuals and Marfan patients led rapidly to identification of a mis-sense mutation, suggesting that fibrillin is the disease gene. Another example of neuropsychiatric interest is identification of mutations predisposing to familial autosomal dominant Alzheimer’s disease within the amyloid precursor protein gene on chromosome 21.

Mapping complex disorders

So-called complex disorders are characteristically common and aggregate in families. However, no simple pattern of inheritance is followed and these conditions have traditionally been attributed to a combination of multiple genes and environmental factors (multifactorial inheritance). Examples include type I diabetes mellitus which has a lifetime prevalence of approximately 0.4%, late onset Alzheimer’s disease which has a prevalence of approximately 15% by age 90 years and bipolar disorder and schizophrenia both of which have lifetime prevalences of the order of 1%. The hallmark of complex diseases is that, in spite of evidence for the involvement of genes, inheritance does not follow simple Mendelian patterns. Many different genetic mechanisms lead to complex inheritance and available data are usually insufficient to allow unambiguous determination of the precise mechanism operating in any particular disease. Genetic mechanisms that may produce complex inheritance include:

1. Epistasis – Multiple disease genes interact to determine disease susceptibility. Individual susceptibility alleles are usually not sufficient, but rather certain combinations of alleles at the various disease loci are required to increase disease susceptibility. Such a mechanism occurs in one form of retinitis pigmentosa and is a plausible model for bipolar disorder and schizophrenia.

2. Locus heterogeneity – Multiple disease genes, any one of which on its own may cause disease. Locus heterogeneity occurs in many apparently simple genetic disorders. For example, retinitis pigmentosa can result from mutation in any of at least 14 different loci and autosomal dominant familial Alzheimer’s disease can be caused by mutations in one of three genes. Heterogeneity models have been considered for both schizophrenia and bipolar disorder.

3. Allelic heterogeneity – Multiple alleles at a single disease locus with different pairs of alleles resulting in expression of different phenotypes. Such a model has been proposed for bipolar disorder.
4. **Dynamic mutation**—A single disease locus in which a disease allele mutates as it is passed from one generation to the next (e.g., trinucleotide repeat expansion). Such a mechanism underlies several neuropsychiatric disorders including Huntington’s disease and fragile X syndrome and is commonly associated with genetic anticipation, the clinical phenomenon of increasing severity and/or decreasing age at onset as the disorder is transmitted through successive generations. Such a mechanism has been proposed in both bipolar disorder and schizophrenia.

5. **Parent of origin effects** (e.g., imprinting)—Expression of an allele depends upon its parental origin. An example of neuropsychiatric relevance is the region of chromosome 15q11–q13 involved in Prader-Willi syndrome and in Angelman syndrome. Both maternal and paternal contributions are required for normal functioning. Absence of maternal chromosomal material leads to Angelman syndrome whereas absence of paternal material leads to Prader-Willi syndrome. It has been suggested that parent of origin effects may operate in bipolar disorder.

6. **Mitochondrial gene mutation**—The disease mutation lies in the mitochondrial genome rather than the nuclear genome, leading to a maternal pattern of inheritance (mitochondria are inherited only from the mother). Such inheritance occurs in several disorders including Leber optic atrophy and some forms of deafness and has been proposed in bipolar disorder.

This list illustrates the range of potential genetic mechanisms that may be operating and which molecular genetic studies should be capable of resolving. It is, of course, possible that more than one of these mechanisms may operate in a single disease.

After the impressive successes in single gene disorders, attention is turning to the more challenging problems presented by complex disorders. The fundamental genetic complication is that the straightforward correspondence between genotype and phenotype that is characteristic of simple disorders breaks down, either because the same genotype can result in different phenotypes (due to the effects of chance, environment, or interactions with other genes) or because different genotypes can result in the same phenotypes. For psychiatric disorders, there are additional problems associated with such issues as variable age at onset, diagnostic reliability and, crucially, validity of the diagnostic categories for genetic research.

Although major chromosomal abnormalities in psychiatric disorders are uncommon, cytogenetic methods are now capable of detecting much more subtle aberrations, such as small deletions, and it is possible that findings from high resolution chromosome studies may make useful
contributions to gene mapping\textsuperscript{30}. However, as in simple genetic disorders, the positional candidate approach seems likely to become the main method used to identify genes underlying complex disorders, and systematic mapping by linkage or association methods will usually be required for initial localization of genes.

\textit{Linkage studies}

\textbf{Method of linkage analysis} Traditional large family linkage studies have used lod score analyses which require specification of genetic parameters, such as gene frequencies and penetrances. For disorders in which mode of inheritance is unknown, analyses must be undertaken for several genetic models. Apart from the potential for inflating type I errors due to multiple testing, type II errors may occur if the true mode of inheritance is not included within the parameter space examined\textsuperscript{17}. An alternative is to use methods of linkage analysis that do not require specification of the genetic model and which, although less powerful, are robust to uncertainties concerning mode of transmission. A number of such methods exist for analysis of extended pedigrees\textsuperscript{31,32} as well as affected sib pairs\textsuperscript{33-35} (see Fig. 1).

\textbf{Type of sample} Although ideal for mapping single gene disorders, there are a number of difficulties in using large, densely affected families for genetically complex traits such as bipolar disorder and schizophrenia:

1. Linkage results in large families can be very sensitive to genotyping errors or changes of diagnosis in single individuals. This is well illustrated by the dramatic drop in lod score in the Amish study of bipolar disorder that resulted from previously well individuals becoming ill for the first time\textsuperscript{36}.

2. Large pedigrees typically contain a broad spectrum of phenotype and to retain power in the analysis a broad definition of illness is necessary. However, broader definitions are usually less stable than narrow ones\textsuperscript{37} and include an unknown proportion of cases that are not genetically related to narrowly defined disorder.

3. It is often claimed that in the face of genetic heterogeneity, large pedigrees should ensure sufficient homogeneity within each pedigree for lod score analysis. However, it has been shown that in heterogeneous disorders large densely affected pedigrees are actually more likely to have multiple disease genes segregating than are smaller less densely affected ones\textsuperscript{38}. 
Fig. 1  Affected sib-pair linkage analysis. The diagram shows three nuclear families, A, B and C, comprising parents and a pair of affected offspring. In each family the father has genotype 1/2 and mother genotype 3/4 at the marker locus. In family A both affected sibs have genotype 1/3 and are said to share 2 alleles identically by descent, ibd. In family B the sibs share one allele ibd and in family C they share no alleles ibd. Under the null hypothesis of no linkage between disease gene and marker locus, the expected proportions of affected sib-pairs sharing 2, 1 and 0 alleles ibd is 1/4, 1/2, 1/4. Sib pair linkage analysis uses statistical tests to detect significantly increased allele sharing above this null expectation.

4. Large pedigrees are rare and, therefore, difficult to collect. Further, there is an unknown relationship between disease in such families and more common cases in which few or no relatives are affected.

Investigation of nuclear families containing pairs of affected sibs has several advantages over the large family approach. Cases are more representative of typical disease, it is possible to focus effort on collecting affected subjects who satisfy narrow and stable definitions of illness, large samples are readily obtained and linkage results are less sensitive to laboratory or clinical errors.

When mode of inheritance is unknown or uncertain, the type and size of sample must be chosen to provide optimal power to detect susceptibility genes over a range of plausible genetic models. As we have pointed out, interacting oligogenes represent plausible models for bipolar disorder and schizophrenia. Under such models, in addition to the problems already mentioned, there is a serious disadvantage in using densely affected families for linkage studies: parents of several affected children have an increased probability of being homozygous at the disease locus, thus, making them uninformative for linkage. In contrast, collection of nuclear families with two affected siblings, preferably with both parents unaffected maximizes the probability that the parents will be heterozygous at the disease locus and, thus, informative for linkage. 

The main disadvantage of a sib-pair linkage approach to complex disorders is that, although robust, the power is relatively low, and so
large samples are required. Nevertheless, the feasibility of accomplishing a complete genome scan in sibling pairs has been demonstrated in a study of type I diabetes where it is probable that at least 5 different susceptibility genes are involved.\textsuperscript{16}

An important difference between linkage mapping of simple and complex disorders is that whereas for single gene diseases recombination events can define an exact interval in which the disease gene must lie, in complex diseases recombination events can only alter the probability that the susceptibility locus lies in a particular interval. Fine linkage mapping of genes for complex disorders, therefore, requires large samples. For example, localizing a susceptibility gene to a 1 cM interval requires a median of 200 sib pairs for a locus causing a fivefold increased risk to a first degree relative and 700 sib pairs for a locus causing a 2-fold increased risk.\textsuperscript{40}

Association studies

Association studies have led to localization of several important susceptibility loci for complex disorders including the HLA locus in type I diabetes mellitus and the ApoE locus in Alzheimer’s disease.\textsuperscript{41,42}

Association between a marker allele and disease status in samples of unrelated patients and unrelated controls may indicate close proximity of the marker to a disease gene (linkage disequilibrium). As for single gene disorders, such case-control studies can, therefore, be useful in narrowing down the region of interest after preliminary localization of a gene by linkage studies. An important attraction of the association method is that it is capable of detecting genes of relatively small effect that could only be detected by linkage studies in prohibitively large samples. A disadvantage is that detection is only possible if the marker is extremely close to the gene and this approach has generally been considered suitable only for the analysis of markers close to candidate genes. Systematic mapping by association requires a marker map of density 10–100-fold greater than that for linkage mapping. However, with very dense maps of markers and efficient, high-throughput genotyping becoming available, such an approach is becoming feasible for initial localization of susceptibility genes.

In addition to linkage disequilibrium, association between disease and marker allele can occur if the marker locus is itself involved in predisposition to disease. In such a situation the association study detects the direct effect of the disease gene. As increasing numbers of genes are being cloned and functional polymorphisms identified, a candidate association approach which seeks to detect polymorphisms leading to variations in protein structure or expression (VAPSEs) is becoming an increasingly attractive method for identifying susceptibility genes.\textsuperscript{43}
Fig. 2 The family-based association method. The diagram shows a nuclear family comprising an affected proband and both parents with genotypes at a marker locus. The proband has genotype 1/3. A well-matched, notional control for the proband is provided by the non-transmitted parental alleles, i.e. 2/4.

However, like all candidate gene approaches, this is only as good as understanding of disease pathophysiology permits.

Conventional association studies are susceptible to spurious associations resulting from inadequate matching of cases and controls, especially when there is ethnic heterogeneity within populations sampled. However, the recent development of family-based association methods helps overcome this problem by allowing an artificial well-matched notional control sample to be constructed from marker data from the family of each proband (see Fig. 2). This advance increases the attractiveness of association studies for genetic dissection of complex diseases.

Physical mapping

After initial localization of susceptibility genes by linkage and association methods, it is possible to use the range of physical methods described above. However, in complex disorders it will often be necessary to investigate a much larger genomic region than has typically been the case for single gene disorders, though this is likely to be less of a problem if linkage disequilibrium is detected. The potential problems should not be underestimated, and it is likely that progress will be critically dependent upon access to data in the form of EST maps and maps of cloned genes assembled as part of the Human Genome Project. This will require use of information technology allowing integration and synthesis of large quantities of data from diverse sources. It is probably not an exaggeration to say that many of the genes for mental illness will be identified in computers rather than in laboratories.
Quantitative traits

A general principle of statistics is that analysing continuous variables provides greater power than analysing categorical variables. Genes influencing scores on continuous measures are known as quantitative trait loci (QTL). Both linkage and association methods have been developed to map QTLs in humans\textsuperscript{45} and successes have been achieved for reading disability\textsuperscript{46}. Risch and Zhang\textsuperscript{47} have pointed out that the optimal design for a QTL sib-pair linkage study is to use a pair of sibs that are extremely discordant for the variable. This is in contrast to the situation of a dichotomous variable (e.g. affected/unaffected) where pairs of affected sibs are by far the most powerful design. Psychiatric genetics has traditionally focused on categorical phenotypes. However, if a valid continuous measure exists that is biologically related to the disorder of interest, a QTL approach is feasible and may provide important information about the disorder itself. A variable of potential relevance in schizophrenia which is currently being used to locate QTLs is schizotypy\textsuperscript{48}.

Animal models

Animals such as rodents offer many advantages over humans for genetic studies because they have large families, short intervals between generations, and experimental crosses can be set up. Mouse models of disease have been used to help in identification of human genes by mapping the relevant mouse gene and then determining its human counterpart (homologue). This is possible because homologues of closely linked human genes are often syntenic (on the same chromosome) in mouse. Such methods have been used to map genes involved in diabetes\textsuperscript{49}. Although there are obvious difficulties in obtaining suitable animal models for studying mental illness, some advances have been made. For example, rodents have been bred that display strain differences in susceptibility to developing 'learned helplessness'\textsuperscript{50} which may be of use in identifying genes involved in susceptibility to depression. Three loci that influence emotionality in mice, a trait that may be related to predisposition to anxiety in humans, have recently been mapped\textsuperscript{51}. Rodents have also been used successfully to map genes of potential importance in substance-abuse\textsuperscript{52}. It seems likely that animal models will play an increasing role in future genetic studies\textsuperscript{14}.

Another important role for animal models is investigation of the function of disease genes that have been cloned. Thus, the neuropathological effects of mutations in the amyloid precursor protein gene on...
Biological psychiatry

Chromosome 21 are being studied by introducing the relevant human gene into the mouse genome, creating a transgenic mouse. An obvious limitation on conclusions based on such an approach is that gene expression is within the context of murine variables which may not be applicable in the human (an example being that mice rarely live for more than 2 years, which obviously may affect conclusions about a process that may take 50 years or more in the human).

Current state of psychiatric genetics

Despite the implausibility of simple, single-gene explanations for bipolar disorder and schizophrenia, much effort has been directed at locating disease genes by linkage analysis in large families with multiple affected members. This is based on the assumption that a proportion of such multiplex families are segregating genes of major effect. Focusing research on highly familial sub-forms has proved useful in other complex disorders such as Alzheimer’s disease, non-insulin dependent diabetes and breast cancer. However, several groups have already conducted searches for susceptibility genes for the functional psychoses throughout a substantial proportion of the genome but no replicated linkages demonstrating genes of major effect have yet emerged. Currently, there is a move towards using non-parametric methods of analysis, collection of smaller, less densely affected families as well as analyses of very large datasets (usually obtained by collaboration) and some interesting linkage findings are beginning to emerge that are consistent with the existence of oligogenic susceptibility loci in both schizophrenia and bipolar disorder. It is likely that at least some of these findings will be convincingly replicated in the large independent samples that are required for complex diseases and undoubtedly additional loci await discovery.

The future

Large family linkage studies using parametric (lod score) methods of analysis have been the work-horse of human geneticists investigating single gene disorders. Linkage studies using non-parametric methods in nuclear families, together with association studies using functional polymorphisms are taking over as the standard approaches to complex disorders. Replication of positive findings will require large independent datasets. Further, as susceptibility genes of relatively large effect are discovered there will be a move toward investigation of larger and larger samples capable of detecting loci of smaller and smaller effect. Such
samples will inevitably require collaboration between groups. Multi-stage designs for linkage mapping are being developed that offer increased efficiency in that they provide the same power to detect linkage for fewer genotypings\textsuperscript{60,61}. Moreover, in the last 5 years genotyping methods have become increasingly automated with dramatic increases in efficiency, and this trend will continue. Extremely rapid and inexpensive automated methods of genotyping and sequencing using DNA attached to computer microchips are being developed and may lead to further increases in efficiency.

In genetic isolates the use of an approach based on detection of shared segments amongst affected subjects may be a particularly efficient strategy for detecting disease genes. This approach, which blurs the usual distinction between linkage and association methods, has been used to map the gene for benign recurrent intrahepatic cholestasis, a rare autosomal recessive disease\textsuperscript{62}. Although the efficacy of this approach for complex disorders is as yet unproven it can be expected to provide an important contribution to the genetic dissection of complex traits\textsuperscript{40}.

Novel molecular methods will undoubtedly become available and may eventually supersede current linkage and association studies. For example, whole genome subtraction methods such as representational difference analysis or genome mis-match scanning are being developed that are capable of investigating all loci in the genome simultaneously in a single reaction\textsuperscript{63}.

**Conclusion**

Genetic analysis of complex traits is difficult because the precise mode of inheritance cannot be deduced. Psychiatric disorders are more challenging than most because measuring behavioural phenotypes is difficult, no laboratory or pathological diagnostic tests exist and virtually nothing is known of the pathophysiology of the diseases. However, bipolar disorder and schizophrenia can be diagnosed reliably using modern operational diagnostic systems and the diagnostic categories have validity as determined by genetic and outcome studies. Powerful and efficient experimental methods have been, and continue to be, developed for investigation of complex disorders and are already yielding successes in disorders such as diabetes and Alzheimer's disease. We can, therefore, be extremely optimistic that over the next few years susceptibility genes for the functional psychoses will be identified and sequenced. Such findings will undoubtedly lead to improvements in treatment and prevention and will provide the first aetiologically-based diagnostic system for the functional psychoses. Identification of susceptibility genes will also
Biological psychiatry greatly facilitate investigation of environmental factors that contribute to disease.

We have focused on the functional psychoses because the genetic contributions to these disorders are relatively large and because much recent work has concentrated upon them. Successes in psychiatric genetics are, therefore, likely to come first for these disorders. However, the methods discussed are equally applicable to the many other psychiatric disorders for which genes confer susceptibility.

Acknowledgements

Nick Craddock is a Wellcome Trust Senior Research Fellow in Clinical Sciences. The work was supported in part by grants from the Wellcome Trust and the Medical Research Council.

References

5 Nicholl DST. An Introduction to Genetic Engineering. Cambridge: Cambridge University Press, 1994
Modern molecular genetic approaches to psychiatric disease