Schizophrenia and bipolar affective disorder are chronic, disabling illnesses that together affect 2% of the population. Genetic factors are known to be important in their development, but there are, as yet, no confirmed susceptibility genes. Here we discuss important issues in terms of alternative genetic strategies (linkage, association and/or cytogenetics) in the identification of candidate genes for the major psychoses. We discuss the impact of the Human Genome Project, the role of comparative genetics in finding and testing positional candidates, and the prospects for rational drug design and personalized medicine.

Mental illnesses are among the most common causes of chronic morbidity worldwide. Two severe forms, schizophrenia and bipolar affective disorder (BPAD) each affect around 1 in 100 individuals, often with onset in early adult life. In spite of their high prevalence and decades of research in neuroscience, neuropsychology, neuropsychology, brain imaging and, indeed, genetics, the causes of these conditions remain unknown and treatments mainly empirical.

Both conditions tend to run in families, and the risk to a first-degree relative of an affected person is about ten times that to a member of the general population. Twin and adoption studies indicate that inherited factors are responsible for a major part of this increased risk. For a number of reasons, however, progress towards identifying genes has been difficult. Diagnosing psychiatric illness is imprecise because psychiatric phenotypes are mainly based on symptom profiles reported by patients. The use of standardized diagnostic criteria has ensured good reproducibility of diagnoses between researchers, but there is wide overlap of symptoms between schizophrenia, BPAD and unipolar depression (Box 1). In the absence of a biological or genetic markers specific for schizophrenia or BPAD, the validity of existing classification remains uncertain.

Approaches to finding genes underlying psychiatric disorders

Linkage and association approaches

In linkage studies, the problem of diffuse diagnostic boundaries is usually met by carrying out analyses...
Box 1. Overlapping symptoms of schizophrenia and related psychoses

**Schizophrenia**
Characterized by acute symptoms (delusions, hallucinations and disorganized thinking, speech and behaviour) which often respond to treatment. The more chronic symptoms (reduced emotional responsiveness, impaired fluency of speech and volition) are a major cause of long-term social and occupational impairment.

**BPAD I**
Characterized by bouts of major depression and episodes of mania (elevated or irritable mood).

**BPAD II and cyclothymia**
Similar symptoms to BPAD I, but less severe.

**Unipolar depression**
For example, major depressive disorder where depression lasts for over two weeks and changes in appetite, sleep, concentration, thoughts and energy lead to serious impairment of almost all daily activities.

Many genetic and environmental risk factors.

One of the strongest findings is a recent study of 22 extended families with a high rate of schizophrenia, produced mixed results, confirming that the illness cannot entirely be accounted for by one or a few major loci. Two genetic models have been considered to explain these findings (Fig. 1). The first hypothesis proposes genetic heterogeneity, such that mutations in one of a large number of different genes can result in a very similar phenotype (as is seen, for instance, in retinitis pigmentosa or nonsyndromal deafness). If this is the case, then linkage will be more readily detectable in single extended families where a single major gene is segregating (most probably in a background of genes of minor effect), while studies of affected sibling pairs or groups of small families will fail with the sample sizes generally available. A second possibility is that these psychiatric disorders show complex inheritance, in which the phenotype is the result of additive or interactive effects of several genes, each of small effect. Under this quantitative-trait model, disease alleles conferring risks to siblings of less than two fold are unlikely to be detectable by current linkage approaches as this would require unachievable numbers of families. Although linkage analysis has failed to detect loci in a number of study sets, it is striking that several genome-wide scans, in both single extended pedigrees and large sets of families and sibling pairs, have identified loci showing highly significant linkage. One of the strongest findings is a recent study of 22 extended families with a high rate of schizophrenia,

Fig. 1. Identification of disease genes in simple and complex disorders. This figure shows examples of different inheritance patterns and how their properties affect the detection of susceptibility loci. (a) Autosomal-dominant locus, fully penetrant. Linkage analysis can be used to localize the causative locus and haplotype analysis to narrow down the interval containing the gene; that is, the smallest region where a haplotype is shared by all of the affected individuals, but not the unaffected individuals. (b) Autosomal-dominant locus, incomplete penetrance. Additional genetic factors and/or environmental factors segregate in this family, altering the expression of the phenotype. Related diagnosis refers to a milder phenotype and/or a subset of the defining symptoms. Linkage and haplotype analysis can still be used as described above, but more careful definition of the affected phenotype is required. (c) Partially dominant autosomal locus of major effect or quantitative trait loci? This inheritance pattern is compatible with both models of complex inheritance described in the text. Linkage analysis will determine whether a single major locus can be detected, in which case haplotype analysis can be carried out. However, unaffected individuals with part of the disease haplotype cannot be used to narrow down the candidate region, as their status might be due to a lack of disease penetrance, rather than non-inheritance of the susceptibility gene.
Positional cloning strategy for a locus at chromosome 4p16 linked to bipolar affective disorder as a worked example. On average, 1 cM is equivalent to 1 Mb of the human genome, however, in this genomic region 11 cM is equivalent to 7 Mb. Stages A, B, and C are complete for this locus and stage D is in progress. The core haplotype is that part of the haplotype where the alleles are common to more than one linked family. This is reliant upon the physical map in the form of a set of partially overlapping bacterial artificial chromosome (BAC) clones (horizontal magenta lines). The magnified, solid magenta line indicates the distance over which linkage disequilibrium might stretch (stage E) and the vertical, blue lines, the individual genes (stage F) that would need to be screened for functional polymorphisms (stage G). The C→T polymorphism shown would have the effect of introducing a stop codon, but other, more subtle polymorphisms (e.g. base change, for example, an intragenic deletion or triplet repeat expansion). The candidate gene approach to psychiatric genetics is well tried, but has met with little real success, perhaps because it is remarkably easy to construct a hypothesis around any gene expressed in the brain, including those transiently expressed during development. By contrast, testing candidates within a tightly defined region can pay off. The major histocompatibility complex (MHC) region on human chromosome 6p has been highlighted by linkage studies in schizophrenia. NOTCH4, which is thought to have a neurodevelopmental function in Drosophila, maps to the MHC region. A strikingly significant association (P = 0.0000078) between susceptibility to schizophrenia and two putative functional polymorphisms at the NOTCH4 locus has recently been reported. However, replication tests of this association in other data sets are needed to validate this finding.

The MHC region has been studied extensively, is now sequenced completely and, to the best of current abilities, all the genes are defined. This best-case scenario contrasts with the bulk of the human genome sequence, which is in draft form, containing numerous small gaps, with only preliminary information on gene position and structure, and few data on predicted function. The annotation process will take several more years of concerted effort, both experimental and computational. In the meantime, two sequential genetic strategies can be adopted to narrow down the search region for positional candidates. The first step is based upon the reasonable assumption of a common founder in two or more families from the same stable population that show linkage to the same chromosomal region. It should be possible to define a smaller common haplotype between such affected families by typing polymorphic markers at high density across the regions identified within individual families. In the second step (which is dependent upon the actual population prevalence of the locus in question), it should be possible to define a sub-region by testing for linkage disequilibrium (LD) or association, seen as a distortion between the allele frequencies in case series and those of controls from the population where linkage was demonstrated. With the physical map in place, the
annotation of the draft human genome sequence building, and experimental approaches to transcript identification, it might then be possible to construct a (very) short-list of candidate genes for direct mutation scanning and functional testing (Fig. 2).

It could be argued that any susceptibility locus identified by this approach might be peculiar to a small sub-set of ‘at risk’ individuals and families, and therefore make only a very small contribution to the overall cause of the conditions. A counter justification for family based approaches is that this approach is well proven for other genetically complex neurological disorders including retinitis pigmentosa, nonsyndromal deafness, epilepsy and Alzheimer’s disease, where the identification of susceptibility loci has provided valuable windows into the underlying physiology.

**Cytogenetic approaches**

Linkage analysis tends to define broad genomic regions. By contrast, cytogenetic abnormalities tend to delineate smaller candidate regions. There are numerous examples of human disorders where chromosome translocations, deletions or insertions alter gene expression, either by direct disruption or by positional effects and lead to the same or a similar phenotype to that caused by point mutation. A single, sporadic observation can be interesting and worthy of speculative investigation, but demonstration of causality requires one of the following:

- finding multiple cases of a rare cytogenetic abnormality associated with the development of the condition;
- co-localization of the cytogenetic abnormality with a candidate gene or region of positive linkage or association;
- familial cosegregation (linkage) of the cytogenetic abnormality with the condition.

An example of the last category is a t(1:11) balanced reciprocal translocation found to cosegregate with major psychiatric illness including schizophrenia through multiple generations in a large Scottish pedigree (Fig. 3). Clinical follow-up found illness in additional translocation carriers within the family and the LOD score for linkage between major mental disorder and the chromosome abnormality is now 7.0 (Fig. 3). The t(1:11) breakpoint directly disrupts not one, but two, novel overlapping genes, DISC1 and DISC2, that are expressed in the brain and are located on chromosome 1 (Ref. 23).

**Impact of the Human Genome Project**

The two examples above illustrate what is now possible working from families towards gene identification in psychiatry. This has only become realistic with the recent and rapid advances arising from the HGP (http://www.sanger.ac.uk/HGP).

Four complementary features of the HGP - physical maps, annotated sequence, sequence variation and comparative genomics - each at a different stage of execution, join to give a sense of promise to psychiatric genetics and, indeed, all investigation of common complex disorders. Complete and accurate physical maps and whole-genome sequence data of each chromosome establish the order of linked genetic markers, define the size of the critical region and aid the identification of positional candidate genes. Contigs are a source for the isolation of new polymorphic markers, and they facilitate high-resolution mapping of genes relative to linked or associated markers and provide a physical template for transcript-identification procedures. The annotation of raw DNA sequence is non-trivial. Although annotation

http://tig.trends.com
will no doubt improve as the total learning set expands, it is unlikely to substitute completely for laboratory experimentation. We will derive great benefit from parallel genome sequencing efforts in experimental organisms more amenable to functional analysis.

The availability of the human genome sequence provides a starting point for the identification and characterization of individual sequence variation, including variation that confers susceptibility to psychiatric illness. DNA sequence variation in the form of single nucleotide polymorphisms (SNPs) are remarkably common, occurring on average once every 1000 bp between any two individuals picked at random from the population. The SNP Consortium (http://snp.cshl.org), a unique collaboration between academic institutions and industry, aims, by the end of 2001, to produce a freely available, generic map of ~300 000 SNPs distributed evenly across the genome. This will facilitate efficient genome-wide association studies, replication studies and meta-analyses and has the potential to become a standard tool for individual pharmacogenetic and population stratification studies.

There has been considerable debate about the number of SNPs that would be required for a successful whole genome scan, but current estimates range from 1 per 3 kb to 1 per 100 kb. This is because the extent of LD in the human population is relatively unknown, and varies at different chromosomal locations. There are less data available on variation between different populations, but initial studies suggest, contrary to some previous predictions, that LD in relatively isolated populations might not be substantially greater than those in more mixed populations. Given the unknown extent of LD in a particular genomic region, reliable high-resolution physical maps are vitally important to establish precise marker order, allowing efficient planning of association studies and reliable detection of LD gradients. Notwithstanding the global contribution of the SNP Consortium, focused searches for functional SNPs; that is, those that alter the product or expression of one or more genes in candidate regions will still be necessary.

Summary and future directions

Do genetics and genomics hold all the answers? Some commentators challenge the whole reductionist approach of genetics applied to psychiatry, dismissing it as a distraction from the importance of social and cultural influences. In truth, studies of genetic modulation and/or genetic variation are perfectly well suited to explore environmental influences, and to move beyond mere description towards a mechanistic and mathematical understanding of biological connectivity and emergent properties.

For the immediate future, a key task is to identify candidate loci and functional polymorphisms through family-based and case-control studies. Locus-by-locus assessment of relative risk, epistasis, pleiotropy and penetrance can then follow. Identification of each and every genetically defined locus also identifies a relevant biological pathway that, by inductive reasoning or experimental study, might lead directly to the nomination of other candidates. Interpreting the biological function of nominated genes and the effects of mutation will call upon a suite of tools encompassing bioinformatics, biochemistry, structural biology, physiology, functional imaging and behavioural studies. The example of Alzheimer's disease and the family based discoveries of APP, PS-1 and PS-2 best illustrate this principle; their discovery has unlocked doors to the pathophysiology of dementia, quite disproportionate to their population prevalence as genetic risk factors. There is every reason to expect similar breakthroughs in the major psychoses.

Intriguingly, animal studies could have a key role in understanding these most human of conditions. There is abundant evidence of ancient similarity in genome structure and function across the phyla. It is fascinating to witness the growing evidence for conservation of function and modules of behaviour in species as diverse as the nematode, the fruitfly and the laboratory mouse. These organisms give us the unrivalled opportunity to test empirically the effect of gene modulation through natural variant, random and site-directed mutagenesis, or transgenic studies.

To model quantitative traits in general, and neurological traits in particular, the onus will be to devise experimental regimes to tease out cause and effect, genetic interaction and environmental influence. This will require:

- refinement of classical transgenic methods to create both subtle, sequence-specific genomic modifications and conditional (tissue-specific and/or temporal) mutants;
- establishment of allelic series (both natural and induced mutations);
- intercrossing of independent mutations (at one or more loci) all with careful control of genetic background and of environmental exposures.

Focusing upon the laboratory mouse, a number of encouraging studies have already been reported. For example, Flint et al. identified three loci for the psychological trait of emotional retardation in a second generation (F2) intercross of ‘high’ and ‘low’ emotionality strains of mice, originally selected from a cross between the C57BL/6j and BALB/cj strains. Similarly, Caldarelli et al. reported quantitative trait loci for contextual fear-conditioning that complement findings from gene knockout studies. Mouse knockout studies have linked the N-methyl-D-aspartate (NMDA) receptor complex directly to synaptic plasticity, long-term potentiation, learning and memory. In keeping with the credo ‘don’t just ablate, modulate’, mice completely deficient in NMDA receptor expression die perinatally, but, when engineered to retain just 5% of normal expression, they survive to adulthood and then display a range of pronounced behavioural abnormalities (increased motor activity, stereotypy and defects in social and sexual interactions) that resemble some features of the schizophrenic phenotype. Intriguingly, these deficits are
ammed or treated by the dopamine-ergic and serotonergic antagonists haloperidol and dozapine that are used to treat the symptoms of schizophrenia. The tentative conclusion is that this mouse may serve as a useful model of schizophrenia. Most recently, Nolan et al.34 described an important, phenotype-driven approach that is complementary to the genotype-driven approaches illustrated above. They generated large numbers of novel mouse variants by chemical mutagenesis and screened these for a variety of phenotypes, including analogues of psychiatric dysfunction, generating 500 new behavioural mutants in the first round. The post-genomic promise of a merging of genotype- and phenotype-driven approaches is truly exciting.

To summarize, the clinical evidence in favour of a biological predisposition to psychiatric illness and to individual variation in genetic vulnerability is compelling. Many early studies were under-powered or undermined by diagnostic uncertainties. The field has moved on, however. Very convincing primary evidence and encouraging evidence for replication has recently emerged for a number of loci. The credibility of the field, however, will only be assured when definitive answers to their precise nature and number are forthcoming. In this regard, psychiatric genetics is in a very similar boat to most areas of complex genetics. The only real difference between solving the problem for psychiatry over cardiology, for example, is the paucity of robust, agreed endophenotypes. These will surely come, but perhaps only after the first gene discoveries are made.

Finally, genetics is providing some explanation for the perplexing and problematic issue of differential drug response in psychiatric patients. This is now recognized to be dependent in part upon intrinsic, genetically controlled levels of drug metabolism35. A polymorphism in cytochrome P450 CYTP2D enzyme accounts for the adverse response to tricylic antidepressants in a proportion of depressives and might arguably become a standard assessment in patient management.26,27 Similarly, a recent report suggests that dozapine response in schizophrenic patients can be predicted with about 75% success on the basis of typing multiple polymorphisms in the serotonin transporter, histamine H2 and the serotonin receptors, 5-HT2A and 5-HT2C (Ref. 36). But understanding the pharmacology of current antidepressants and antidepressants is just the beginning. These drugs have had a major impact on treatment, but their serendipitous discovery leaves the lack of specificity and unacceptable side-effects.

Beyond pharmacogenetics, the real challenge to the pharmaceutical industry is to convert gene discovery into biological understanding, which, through computational and experimental biology, can lead to target discovery followed by rational drug discovery and optimized patient prescription.

Acknowledgements

We thank K. Miller for constructive comments on this manuscript.

References

3 World Health Organization (1992) The CD-10 Classification of Mental Health and Behavioural Disorders. World Health Organization
16 Williams, N.M. et al. (1999) A two-stage genome scan for schizophrenia susceptibility genes in 196 affected sibling pairs. Hum. Mol. Genet. 9, 1729–1739
21 St Clair, D. et al. (1990) Association within a family of a balanced autosomal translocation with major mental illness. Lancet 336, 13–16