Psychiatric genetics in the post-genome age

AIDEN CORVIN and MICHAEL GILL

Less than 50 years after publication of the structure of DNA, we possess a working draft of the human genome and expect a completed draft listing all genes and their location by 2003 (International Human Genome Sequencing Consortium, 2001). For 21st-century medicine, the genome will offer a gateway to new biological mechanisms that integrate our genetic and environmental understanding of behaviour and disease. In psychiatry a genetic contribution to common disorders, cognition and personality traits is well established, and susceptibility loci have been identified for schizophrenia, bipolar affective disorder, autism and Alzheimer's disease. Genetic mutations causing hundreds of Mendelian disorders are known; many present with rare neuropsychiatric phenotypes, such as Huntington's disease and familial forms of Alzheimer's disease and Parkinson's disease (http://www.ncbi.nlm.nih.gov/OMIM). Nevertheless, as in other branches of medicine, locating susceptibility genes of minor effect has proved difficult. The common neuropsychiatric disorders have a complex genetic aetiology, probably involving interaction between genes and interaction between genome and environment. Until now this complexity has (mostly) frustrated investigators, but the prospect of a complete genome draft offers hope that genotype–phenotype interactions can be unlocked to expose underlying biology. This may be the first challenge of the post-genome era, but it echoes older motivations that are not exclusive to psychiatric geneticists:

> 'The intention is to furnish a psychology that shall be a natural science: that is, to represent psychical processes as quantitatively determinate states of specifiable material particles, thus making these psychical processes perspicuous and free of contradiction.' (Freud, 1914)

FROM GENOME TO SUSCEPTIBILITY GENE

Genetic principles can be applied to directly identify relationships between genomic variation and disease. Hypothesis-free approaches of this type are attractive because genes can be identified where underlying disease biology is unknown (as is often the case in psychiatry). Draft human genomic information will improve the power of traditional methods such as linkage analysis to uncover relationships between DNA variation and disease. Linkage studies examine for co-segregation of disorder and genetic markers in families. Early studies had a limited choice of markers; now there are extensive marker maps for all chromosomes. Unfortunately, in such analyses marker density reaches a saturation point where statistical evidence may implicate regions (loci) spanning millions of base pairs. Linkage studies also have limited power to identify small gene effects, but this can be partially resolved by collaboration or meta-analysis to increase sample size. By increasing study power in this way, susceptibility loci have now been identified for a number of complex disorders, including non-insulin-dependent diabetes mellitus (Horikawa et al, 2000), Alzheimer's disease and schizophrenia. Where loci are identified the completed genome draft will improve our ability to explore these large regions, which until now have required extensive mutation screening or DNA sequencing to uncover causative genetic mutations (Risch, 2000).

For complex disorders, individual gene effects may be too small to identify by linkage, so new positional cloning methods and expansion in genotyping and analytical capacity will be necessary to identify all loci. Ultimately all DNA variation, or at least all variation that influences function, may need to be tested (Risch, 2000). This is a considerable challenge, as total genomic variation is immense. A major collaborative effort by the SNP Consortium (http://www.ncbi.nlm.nih.gov/SNP/get_html.cgi?whichHtml=overview) has identified 4 million single-nucleotide polymorphisms (SNPs), the commonest form of DNA variation, across the human genome. Comprehensive data on functionally relevant SNPs and their frequencies in different populations are only in development. Exploiting this information, when it becomes available, to directly test all functional SNPs may require a 100-fold increase in genotyping capacity. Modern genotyping methods using hybridisation (e.g. DNA chip microarrays), primer extension (e.g. mass spectrometry, pyrosequencing), oligonucleotide ligation (e.g. microarray ligation) or allele-specific polymerase chain reaction (PCR) are all limited by the necessity of a PCR amplification step. Novel genotyping methods, perhaps avoiding this step, will be required before tests of all functional SNPs will be feasible (Syvänen, 2001).

Recent studies of genome structure may make it possible to capture total genomic variation using information from only a fraction of all DNA variants. This economy depends on linkage disequilibrium, the non-random association between variants at adjacent loci. Linkage disequilibrium occurs in populations because approximately two recombination events occur per chromosome at meiosis. Consequently, markers that are closely physically located tend to remain together over many generations – analogous to playing cards sticking together until disrupted by repeated shuffling. New evidence suggests that linkage disequilibrium follows highly structured block patterns and each block might be characterised by a small number of variants (Daly et al, 2001). Understanding patterns of linkage disequilibrium across the genome in different populations is a major goal of current research, but even optimistic predictions of the number of markers ($\sim$50 000) required for genome-wide association are far beyond current capacity. However, linkage disequilibrium structural information will be useful and will allow much more efficient mapping of susceptibility loci, but at a cost. Association studies identify haplotypes (sequences of DNA uninterrupted by recombination) containing susceptibility genes, but may not allow localisation of a specific mutation because of complete linkage disequilibrium between variants within the haplotype.

GENE MUTATION IDENTIFICATION INFORMED BY NEUROBIOLOGY

The positional cloning strategies outlined will not develop in a vacuum. Direct

examination for association (or linkage disequilibrium) between DNA variants and phenotypes is a more powerful method of detecting small gene effects than linkage. Until all variation can be examined, we are dependent on scientific hypotheses to select candidate genes to be tested for association with disease. Because the aetiology of many neuropsychiatric disorders is obscure, the number of plausible candidate genes is large and prior probability that a given variant is associated with disease is low. Where success has been achieved it has depended on strong test hypotheses (e.g. DAT1 in attention-deficit hyperactivity disorder).

Research in neuroscience is improving our molecular understanding of neuro-development and brain function. The structure of brain regions and neural systems at a molecular level are being delineated; for example, at least 16 distinct gene clusters (~1000 genes) are involved in neuronal proliferation, differentiation, synapse formation and maturation in normal hippocampus development (Mody et al, 2001). As our aptitude for directly testing DNA variation improves, so too will our molecular understanding of neural systems, allowing more comprehensive and refined candidate gene studies. Gene expression studies will also directly inform candidate gene selection: first, by allowing the expression of all genes in susceptibility loci to be tested directly (Shoemaker et al, 2001), and second, by demonstrating differential gene expression between disease tissue from case and control samples. Studies of this kind in schizophrenia may allow more focused candidate gene studies by demonstrating abnormal expression patterns for a tiny proportion of the thousands of genes tested.

BIOINFORMATICS

Requirements for storage, analysis and interpretation of the vast amount of biological data generated by the Human Genome Project have spawned a new science, bioinformatics. Bioinformatics is the Rosetta stone that has rendered the genome intelligible. Genes initially located experimentally from expressed sequences can now be identified computationally directly from the sequence. Algorithms can predict genes and gene structure *inter alia* by homology with either known human genomic/protein sequences or by comparison with sequences from other organisms.

As the draft genome is completed, integrated databases of gene structure, transcripts and proteomics are being developed which will contribute to our understanding of gene function and relationships (http://www.ensembl.org/genome/central/). Many secrets of the genome are still out of reach. The Human Genome Project found fewer genes than expected, indicating that much of our genetic complexity may be because of gene regulation, alternative splice variation and post-translational change. Characterising subtle genomic effects and establishing the molecular mechanisms involved will be a major challenge for future experimental and computational research.

FUTURE POST-GENOMIC RESEARCH

As understanding of neurobiology improves, the methods described will merge and susceptibility genes will be identified. Confirming their role will require interpretation of expression (the transcriptome) and protein products. Animal models will be needed to study the behavioural impact of genes *in vivo* in specific systems and through neurodevelopment. Because many genetic effects will be complex, multifaceted transgenic models will be necessary to examine subtle interactions between genes. In psychiatry particularly, the sophistication of behavioural phenotypes will be a major obstacle: current models of schizophrenia, for example, depend on basic phenotypes such as gating deficits. New epidemiological approaches will also be required to examine for and interpret interactions between genetic and environmental risk factors in people. Gene identification may be only the first step, but it begins a process that will expose the genetic architecture and biology of neuropsychiatric disorders. Future researchers will follow stronger neurobiological leads than were available to earlier generations:

'. . . those who being in search of any thing, that lies conceal'd from them, and not finding it in the place they expected, beat about all neighbouring

fields, without any certain view or design, in hopes their good fortune will at last guide them to what they search for' (Hume, 1739).

Post-genomic research may forever change how we classify and treat mental disorders. Genetic insights into behaviour may fundamentally alter how we view the individual and his or her responsibilities. But for common disorders and traits, susceptibility genes will only be risk factors and we must never forget the importance of environmental influences.

REFERENCES

Daly, M. J., Rioux, J. D., Schaffner, S. F., et al (2001) High-resolution haplotype structure in the human genome. *Nature Genetics*, **29**, 229–232.

Freud, S. (1914) Project for a Scientific Psychology. Reprinted (1953–1974) in the *Standard Edition of the Complete Psychological Works of Sigmund Freud* (trans. and ed. J. Strachey), vol. I, p. 281. London: Hogarth Press.

Hume, D. (1739) *A Treatise of Human Nature: Being an Attempt to Introduce the Experimental Method of Reasoning into Moral Subjects, etc.* London: John Noon.

Horikawa, Y., Oda, N., Cox, N. J., et al (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nature Genetics*, **26**, 163–175.

International Human Genome Sequencing Consortium (2001) A physical map of the Human Genome. *Nature*, **409**, 860–921.

Mody, M., Cao, Y., Cui, Z., et al (2001) Genome-wide gene expression profiles of the developing mouse hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 8862–8867.

Risch, N. J. (2000) Searching for genetic determinants in the new millennium. *Nature*, **405**, 847–856.

Shoemaker, D. D., Schadt, E. E., Armour, C. D., et al (2001) Experimental annotation of the human genome using microarray technology. *Nature*, **409**, 922–927.

Syvänen, A. C. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature Reviews, Genetics*, **2**, 930–942.

AIDEN CORVIN, MRCPsych, MICHAEL GILL, MRCPsych, Department of Psychiatry, St James's Hospital, Dublin 8, Ireland

Correspondence: Dr Aiden Corvin, Department of Psychiatry, Trinity Centre for Health Sciences, St James's Hospital, St James's Street, Dublin 8, Ireland