

Chapter

5

Identifying genetic factors contributing to individual differences in behaviour



Identifying genetic factors contributing to individual differences in behaviour

Introduction

- 5.1 Estimates of heritability (see Chapter 4) provide evidence that genetic factors contribute, to a greater or lesser extent, to variation in behavioural traits. A great deal of effort is currently being made to identify these factors using molecular genetic approaches. This chapter reviews these methodologies. Research in molecular genetics is different from research in quantitative genetics in that it attempts to identify the function of particular genes, whereas research in quantitative genetics examines unspecified genetic influences.
- 5.2 Research focused on examining the genetic contribution to normal variation in behavioural traits is a branch of neuroscience. It thus represents another method used to seek understanding of how the brain works. Research in molecular genetics can also play a role in psychosocial and epidemiological research, enabling confounding genetic effects to be identified and controlled for, and to allow for further study of the non-genetic factors that influence a characteristic. A further reason for undertaking research on normal variation is that it will provide information relevant to disorders and diseases. As already observed, there may be common genetic influences on behaviour in the normal range, such as anxiety, and disorders, such as clinical depression. Further, a particular genetic influence may be linked both to behaviour in the normal range, as well as to extremes of that behaviour.

Approaches to identifying susceptibility alleles

- 5.3 Attempts to identify susceptibility alleles that influence traits represent various blends of 'bottom-up' and 'top-down' approaches. The 'bottom-up' approach starts with knowledge of the biochemistry of the system in question, and investigates, in a logical fashion, how the system may be varied. With behaviour, the problem is that in most cases, the biochemistry is understood imperfectly, if at all.¹
- 5.4 If there is some background biochemistry to direct researchers, a 'candidate gene' approach can be taken, by studying genetic variation that is known to affect the function of proteins suspected of having a role in behaviour, for example, those that act in the brain. An example of this approach is the dopamine D4 receptor (DRD4), which is discussed in paragraphs 5.9 - 5.10.
- 5.5 The next level of approach is to identify a variation in the relevant gene that is speculated to affect the function of the protein. The best candidates for polymorphisms to study are those that involve amino acid substitutions that are chemically significant, or are surmised

¹ One exception to this statement relates to alcoholism. Alcohol is rapidly broken down in the body, or metabolised, to form acetaldehyde, a highly toxic chemical that causes nausea, facial flushing, dizziness, headaches and other unpleasant symptoms. In most gene pools, individuals possess variant alleles of the aldehyde dehydrogenase gene that enables acetaldehyde to be easily broken down. However, there is a particular variant allele of the ALDH gene, common in people from South East Asia, that leads to a slow-metabolising form of the protein, ALDH2*2. When these individuals (particularly ALDH 2*2 homozygotes) drink alcohol, they experience toxic levels of acetaldehyde and the accompanying symptoms. Possession of one or two slow-ALDH alleles protects against alcoholism. In fact, from hundreds of individuals screened, only a single Asian alcoholic was reported to be an ALDH2*2 homozygote (Chen, C.-C. *et al.* (1999). Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. *Am. J. Hum. Genet.* **65**, 795–807). Thus, this single gene confers substantial protection against alcoholism. This is a rare example of a single gene that is conclusively linked to a behavioural trait.

to affect the levels of expression of proteins that have a role in brain function. An indirect way to determine an effect on function is to try to find an association between the protein variation and behaviour; then, if a significant association is found, to try to establish what the functional effect might be.

5.6 Finally, the most extreme 'top-down' approach (the hypothesis-free method) examines the relationship between genetic variation of unknown function (usually in the form of multiallelic variants, termed microsatellites,² or SNPs) and the behaviour under study. A significant association might indicate one of three things:

(i) the effects of chance;

(ii) that the variation serves as a proxy signal for a nearby variant that influences the behaviour that occurs through a process termed linkage disequilibrium, which describes the tendency for closely spaced markers to be inherited together, only becoming separated by rare recombinations;

(iii) that the variation might itself be influencing the behaviour (for example, the variation might turn out to lie in a region outside a gene, influencing the gene's expression).

Linkage studies

5.7 Two general methodologies are used to find or test for susceptibility alleles. These are linkage and association studies. Linkage studies follow the inheritance of traits through families in comparison with polymorphic genetic markers. The consistent co-inheritance of variation at a polymorphic locus with a trait would support the hypothesis that the trait was influenced, at least in part, by genetic variation close to the polymorphism being studied. Conversely, the random inheritance of the polymorphism and the trait would be evidence against linkage. Linkage analysis in large families has been very successful in the identification of single gene disorders, but is less applicable to the study of behaviour, which does not segregate in a simple dominant or recessive fashion.

5.8 Simplified types of linkage analysis, for example, using pairs of affected siblings, have the advantage that they do not assume a particular mode of inheritance. Also they can better accommodate variation in the trait along a continuous scale (rather than as present/absent categories) in a so-called quantitative trait locus (QTL) design. This approach has been used to map susceptibility alleles for traits (including behaviour) in animals. In humans, relatively large samples (several hundred pairs of siblings) are required, and even these only have the power to detect quite major effect sizes.³ Perhaps for this reason, linkage studies have not been applied very widely to the study of behaviour. A notable exception, though, is the early reported linkage between male homosexuality and genetic variation in the Xq28 region of the X chromosome,⁴ which remains controversial (see paragraphs 10.14 – 10.17).

² Microsatellites contain tandem repeats of a simple sequence such as the dinucleotide CA, that vary in number and are usually of no functional significance. At a particular locus, one person might have (for example) 10 CA repeats on one chromosome and 14 on the other. Another person might have 11 and 13 repeats on their two chromosomes. Because these repeat lengths are usually stably inherited from generation to generation, differences in the distribution of repeat lengths between two populations may indicate differences in their genetic origins.

³ Risch, N. & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* **273**, 1516–17.

⁴ Hamer, D. H., Hu, S., Magnuson, V. L., Hu, N. & Pattatucci, A. M. (1993). A linkage between DNA markers on the X chromosome and male sexual orientation. *Science* **261**, 321–7.

However, several further linkage studies are in progress, examining, for example, anxiety, depressive symptoms and neuroticism.

Association studies

- 5.9 Association studies are more commonly used for genetic studies of behaviour. In its simplest form, an association study compares the frequency of a particular genetic variant in a cohort of cases (a group of people with a particular behavioural characteristic) with a matched set of controls (a similar group of people not displaying the characteristic). In the study of behaviour, 'cases' as such may not exist, since behavioural traits are unlikely to be easily categorised as present or absent; QTL designs can accommodate this. Two major advantages of association over linkage studies are, first, that they are more powerful for detecting susceptibility alleles of small effect size, such as those anticipated in genetic influences on behaviour, and second, that the samples are easier to collect because only single affected individuals are needed in each family. An example of an association study is the occurrence of different genetic variants of the DRD4 gene in novelty-seeking behaviour. As dopamine is a key neurotransmitter, this DRD4 polymorphism is a plausible candidate as a contributor to genetic variation in behaviour.
- 5.10 Two highly influential papers published in 1996 suggested that a particular allele of DRD4 was associated with novelty-seeking behaviour.⁵ However, the effect of the allele was modest: the papers concluded that this polymorphism accounted for only 3–4% of overall variation in novelty-seeking. Nevertheless, this work sparked a deluge of studies of possible associations of DRD4 with many aspects of behaviour, including alcoholism, drug abuse and Attention Deficit Hyperactivity Disorder (ADHD). A subsequent critique concluded that the associations with novelty-seeking that were originally reported, as well as those with alcoholism and drug abuse, were not statistically robust. However, a weak association might exist between the 7-repeat allele and ADHD,⁶ a conclusion also supported by a recent meta-analysis.⁷
- 5.11 A frequent criticism of association studies is that if there are subtle, but undetected differences in the populations from which the cases and matched controls were sampled, then differences in allele frequency might simply reflect the background evolutionary differences between the two samples, rather than reflecting true trait-specific differences. This problem is termed stratification. One way to avoid this is to incorporate parents or siblings into the design and examine differences in the frequency with which the two parental alleles are passed down to the offspring (transmission disequilibrium tests). This approach provides a 'halfway house' between linkage and association.

⁵ The allele is called the 7-repeat allele, because it contains seven repeats of a particular series of 48 base pairs found in the gene. Other alleles have been identified that contain between two and eleven repeats of this section. Ebstein, R. P. *et al.* (1996). Dopamine D4 receptor (D4DR) exon III polymorphism associated with the human personality trait of Novelty-seeking. *Nat. Genet.* **12**, 78–80; Benjamin, J. *et al.* (1996). Population and familial association between the D4 dopamine receptor gene and measures of Novelty-seeking. *Nat. Genet.* **12**, 81–4.

⁶ Paterson, A. D., Sunohara, G. A. & Kennedy, J. L. (1999). Dopamine D4 receptor gene: novelty or nonsense? *Neuropsychopharmacol.* **21**, 3–16. The authors state that 'evidence for the role of DRD4 in novelty-seeking is inconclusive, with a number of methodological concerns'.

⁷ Faraone, S. V., Doyle, A. E., Mick, E. & Biederman, J. (2001). Meta-analysis of the association between the 7-repeat allele of the dopamine D(4) receptor gene and attention deficit hyperactivity disorder. *Am. J. Psychiatry* **158**, 1052–7.

- 5.12 Another major problem with association studies is that the testing of very large numbers of loci will lead to numerous spurious associations for purely statistical reasons, using normal criteria for significance. A common approach to combat this effect is to carry out a replication study, in which potential associations identified in an initial screen are re-examined in an independent group of individuals.
- 5.13 In practice, two approaches have been dominant in the study of behaviour. The first are association studies employing variants in candidate genes of either known or potential functional significance. This has been the most popular approach to date.⁸ The second is to use a whole genome, hypothesis-free approach to study association, in which all gene variants are of interest, not just selected candidates. Research by Plomin *et al* on the attempted identification of QTLs for general cognitive ability (*g*) provides one of the first examples of a fully hypothesis-free, genomic approach to the study of a behavioural trait (see paragraphs 7.15 – 7.24).⁹ It also illustrates many of the difficulties with this approach.

Identification of alleles that influence behaviour

- 5.14 The robust replication of a linkage or association identifies a small segment of the genome that contains a susceptibility allele, but does not necessarily identify the allele that actually influences the behavioural trait. This is inevitably the case for linkage, which examines gene loci, not specific alleles. In the case of association, the occurrence of linkage disequilibrium complicates the interpretation as the identified allele might be a neutral hitch-hiker with another (unidentified) allele close by. The next step in a linkage approach is to reduce the extent of the chromosomal segment which needs to be examined, by recruiting additional families and testing additional markers. The DNA sequence of the defined interval is then scrutinised for regions likely to encode genes; a list of 'candidate genes' is drawn up and the DNA sequence of each is obtained, looking for sequence changes from normal. In the case of association, attempts are made to identify all the SNPs near the site showing the initial association, then all the SNPs are tested to determine whether any show a stronger association with the trait than the allele originally identified.
- 5.15 As may be deduced from this abbreviated account, the identification of alleles that have an influence on complex traits is by no means straightforward, even when a robust linkage or association can be identified. Whereas in research involving animals this process is facilitated by the use of specific breeding strategies, this is clearly impossible in humans. Moreover, in contrast to Mendelian traits, it is unlikely that any specific allele will be both necessary and sufficient to cause the trait, so the genetic evidence for causation will be of a statistical nature. Further evidence of a causal link must be sought through functional studies. Most often these will involve experiments on animals, discussed in Chapter 6.

Scaling up the analysis: new methods in genetics

- 5.16 The past few years have witnessed the move of genetics from small-scale science conducted in individual laboratories to a larger-scale approach similar to that employed for many years in substantial physics projects as well as in industry. This has been prompted by various factors, including the availability of the sequence of the human genome, the development of partially

⁸ An example is provided by the work of Comings, D. E. *et al.* (2000). A multivariate analysis of 59 candidate genes in personality traits: the temperament and character inventory. *Clin. Genet.* **58**, 375–85.

⁹ Plomin, R. *et al.* (2001). A genome-wide scan of 1842 DNA markers for allelic associations with general cognitive ability: a five-stage design using DNA pooling and extreme selected groups. *Behav. Genet.* **31**, 497–509.

automated high-throughput technologies for analysing DNA sequence and its variation, and the strategic move of the pharmaceutical industry into genetics, as well as the sheer complexity of human genetic variation. It can be anticipated that this trend will continue apace.

- 5.17 Apart from the use of rapid methods of analysing genotypes involving large numbers of samples (essential for the approaches outlined above), another technology that promises to yield significant insights is the use of gene chips or microarrays. The best validated application of this technology is for simultaneous examination of the expression of thousands of different genes from a particular source of tissue. DNA sequences from these genes are arrayed onto a glass slide or synthesised chip, then RNA (the intermediate between DNA and protein) from the tissue of interest is prepared and matched by hybridisation to the array. This enables widespread changes in gene expression to be examined. Such methods are likely to come into widespread use in behavioural genetics. It is anticipated that they will lead to a more sophisticated view both of the biology of behavioural processes and to new ways of classifying these processes. Large-scale proteomics approaches, which examine changes in protein, rather than RNA, expression, are also being developed.

Conclusion

- 5.18 It is likely that there will be a significant increase in the application of molecular genetics to the study of behaviour. It can be anticipated that very large amounts of data about the function of particular genes will be generated over the coming years, and many claims will be made about the significance of these data. Box 5.1 contains a number of points that should be borne in mind when evaluating such claims.

Box 5.1: Central points about research in molecular genetics

- Research in molecular genetics tries to identify variation in particular genes that influences behaviour, by examining the DNA of individuals.
- This is difficult because there are usually many genes involved, each of which may only have a small effect. Many associations between a genetic variant and a behavioural trait have been reported but have not been successfully repeated by other researchers.
- In most cases, the research does not explain how the gene influences the behaviour. However, some researchers predict that they will overcome these difficulties and that genes that influence behaviour will be reliably identified.
- When associations are reported by researchers, it is important to consider the following questions:
 - How convincing is the evidence, in terms of both statistical analysis and the supposed pathway of causation, that the claim is correct? Much more credibility can be attached to findings that have been independently replicated by a different research group, and first reports of gene–behaviour associations should be treated with caution until they are replicated.
 - Over what range of populations and environmental conditions has the effect been tested?
 - If claims are made about the practical application of the findings to influence human behaviour, what is the size of the effect of the genetic variant? Is it large enough to have any relevance for the testing of individuals?
 - What are the implications for the pathway of causation of the behaviour?