REVIEW

Genetic association studies of complex neurological diseases

P M Abou-Sleiman, M G Hanna, N W Wood

J Neurol Neurosurg Psychiatry 2006;**77**:1302–1304. doi: 10.1136/jnnp.2005.082024

Genetic association studies offer a powerful approach to identify the multiple variants of small effect that modulate susceptibility to common, complex disease. They, however, have a poor reputation, mainly because of the consistent lack of replication of all but a few. Thousands of genetic studies have been carried out on multifactorial diseases in the past 30 years, yielding only about 50 variants that can be considered to be true positives. Although the positive studies show proof of principle, the multitude of negative studies indicate fundamental problems in the design and execution of association studies. Here, we discuss some of the more pertinent study design and data analysis issues which can affect the outcome of genetic association studies.

Genetic association studies rely on shared common ancestry in populations to detect small genetic effects on disease susceptibility.¹ A basic case-control study, the most commonly used design for allelic association studies, compares the frequency of common allelic variants in a set of samples with those in controls. A statistically significant difference in allele, genotype or haplotype frequencies in the case-control partition should then indicate the presence of a disease-causing variant at or around the sampled polymorphism. Although association studies are simplistic in principle, they are considerably more challenging at a practical level. The first hurdle is identifying the right candidate genes. Unlike linkage studies, where whole genome scans have been used with great success to identify the genetic basis of monogenic disorders, association studies have, to date, been carried out on specific candidate genes ascertained through either biological plausibility or regions of linkage from sib-pair studies.^{2,3} Genomewide association studies would end the reliance on prior assumption, eliminating the guess work associated with candidate gene selection; however, they remain technically, statistically and economically challenging. Genomewide linkage studies have proved hugely successful in identifying the genetic basis of mendelian disorders, as the highly penetrant low-frequency variants, which are causal of monogenic diseases, can be easily detected by the low-resolution (5–20 cM) cover of linkage studies. By contrast, the variants associated with multifactorial diseases are by definition more likely to be higher in frequency, and incompletely penetrant. Several factors will probably be

involved, each delivering a modest contribution to the overall heritability. Association studies therefore require the genotyping of a far denser set of variants in large sample numbers. The preponderance of human genomic data available to researchers from resources such as the HapMap project, which currently contains 4.2 million genotypes representing one single-nucleotide polymorphism (SNP) every 5 kb in four different populations, and the development of low-cost genotyping technologies have converged to realise the possibility of whole-genome association studies. However, it remains to be seen whether the resolution of these studies will be sufficiently high to detect the disease-causing variants, and whether the data analysis protocols are capable of reliably dissecting out the true associations from the false positives.⁴ The candidate gene approach, despite its limitations, will therefore continue to be used in the immediate future.

Having selected candidate genes, the case-control study must be appropriately designed to maximise the possibility of identifying causal variants. Well-designed studies will also deliver reliable negative results, which are important to reduce duplication of effort and conflicting reports in the literature. A well-designed study will take account of the following:

1. population substructure and selection bias
2. sample selection (phenotype definition)
3. sample size
4. genetic architecture
5. genotyping efficiency
6. appropriate statistical analysis of the data
7. replication of positive results in an independent cohort.

We will expand on each of these points, showing their importance in the design of a reliable and informative genetic association study.

Population substructure and selection bias

Homogenous genetic ancestry, both within and between case-control cohorts, is an essential prerequisite of association studies as allele and haplotype frequencies can differ considerably between populations, resulting in both false positives and negatives. Population-based association studies are particularly susceptible to stratification owing to ethnic admixture as differences in disease prevalence between ethnic groups can result in the over-representation of a subgroup in the case cohort. As the most widely discussed source of systematic bias in association studies, the effect of stratification on

See end of article for authors' affiliations

Correspondence to:
N W Wood, Department
of Molecular
Neuroscience, Institute of
Neurology, Queen
Square, London WC1N
3BG, UK; n.wood@ion.ucl.
ac.uk

Received 23 October 2005
Revised 5 July 2006
Accepted 1 August 2006

well-matched studies remains controversial. Two powerful approaches have been proposed for the detection and correction of stratification. The first uses genetic markers which differ in frequency between founder populations to identify individuals with different ancestries. Ancestry can then be included as a covariate in the association analysis.⁵ This method can be powerful when applied to admixed populations such as African Americans, where the founders are genetically very different and well characterised, but its usefulness in less stratified populations such as northern Europeans remains to be determined. The second method, genomic control, provides an estimate of stratification by comparing the distribution of association statistics at unlinked variants distributed throughout the genome in cases and controls with that of the test allele. The false positive rate can then be reduced by increasing the statistical significance threshold. Two studies have examined the effects of population stratification using genomic control,^{5,6} concluding that whereas low-level stratification may be present, well-matched studies are unlikely to be affected by high-level stratification.^{7,8} What level of stratification is acceptable for a given study and how many unrelated markers should be tested? To begin with, given that genomic control deals with only the false positive rates, it needs to be considered only in studies reporting positive associations. The number of markers that should be typed therefore depends on the genetic effect being reported. Where highly significant *p* values are obtained, high-level stratification can probably be ruled out using a few dozen markers, whereas studies reporting marginal *p* values may require several hundred markers to ensure the absence of low-level stratification. At a practical level, it is advisable to avoid stratification, although it can be detected and corrected for, by ensuring that the sample cohorts are well matched. The most obvious approach would be to maintain homogeneity by sampling ethnically matched individuals from outbred populations in a defined geographical region. Selection of appropriate controls is often more problematic than case selection (see next section); ideally, controls would be nationally representative, with good longitudinal phenotype data such as the National Child Development Study, which takes as its subjects all the people born in the UK during one week in March 1958. Many studies are carried out on combined sample sets sourced from multiple centres to increase patient numbers. Although the large sample sizes should translate into higher power, the increased genetic heterogeneity can cancel out the advantages of the sample increment. It is therefore paramount that publications state clearly how and where the samples have been collected.

Family-based association studies, such as the analysis of pedigrees,⁹ case–parent triad¹⁰ and case–parent–grandparent¹¹ designs, have been proposed as alternatives to population-based studies, as they are unaffected by population stratification. In these designs, the differential transmission of alleles to affected individuals is indicative of linkage; controls are therefore exactly matched to the case samples. However, the benefits of family-based studies can be outweighed by the difficulty in obtaining samples, especially when the disease under investigation is late onset. Genotyping costs are also increased owing to the higher density of markers required compared with population samples.

Sample selection (phenotype definition)

Complex diseases often present with heterogeneous phenotypes that are likely to represent the clinical manifestations of diverse genetic or environmental insults. Although it is unrealistic to create a general rule of thumb for disease classification, consideration should be given to the appropriate phenotyping of samples during study design to avoid

heterogeneity which will reduce the power of the study. However, for most complex diseases, obtaining a phenotypically heterogeneous cohort that is large enough for a case–control study may be unrealistic. Ultimately study-specific decisions have to be made on the appropriate balance between phenotypic heterogeneity and sample size, based on the cost of the phenotyping tools and their ability to dissect out differential latent genetic architecture. These decisions can have major repercussions on the outcome of the study. For example, most of the complex diseases can also be caused by rare, highly penetrant, low-frequency alleles (<0.01–0.001). Case cohorts enriched for genetic susceptibility by the inclusion of samples with known family histories or early age of onset may have reduced power to detect higher frequency susceptibility alleles of small effect. On the other hand, the inclusion of intermediate phenotypes, where the risk of heterozygotes is intermediate to the risk of the two homozygotic genotypes, will not only enrich a cohort for genetic disease but also provide convincing proof of functionality.

Sample size

One of the most important and consistently overlooked considerations in association studies is power. The power of a study is the probability of detecting a hypothesised effect of a given size, and is a function of the effect of size on the causal variant, the allele frequencies of both causal and test variants, the linkage disequilibrium between the test SNP and the causal variant, and, of course, the sample size. The number of samples required to reach significance at a predetermined level of power (usually $\geq 80\%$) can be estimated using one of several genetic power calculators.¹² The power obtainable from a given sample set should be estimated to approximate the smallest detectable effect size. All association studies, and particularly negative studies, should then clearly report the minimal detectable effect size to facilitate the differentiation between underpowered studies and those that are genuinely negative. To illustrate this point, it is generally accepted that 1000 cases and controls are needed to detect a minimum risk ratio of 1.5, assuming an allele frequency of $\geq 10\%$ and given that there is high linkage disequilibrium between the marker and the causal variant, and that their allele frequencies are similar. If we maintained the above parameters, and reduced the number to 500 cases and controls, the minimum detectable effect size would increase to a risk ratio of above 2.

Genetic architecture

Indirect association studies, those that do not directly assay known functional variants, rely on linkage disequilibrium to detect disease-causing variants in the proximity of the test SNPs. The use of linkage disequilibrium in association studies has been extensively reviewed elsewhere.^{13,14} In summary, a given pair of intrachromosomal markers are said to be in linkage disequilibrium if they segregate together more often than would be expected under independent assortment. Markers that are in high or complete linkage disequilibrium will therefore cosegregate either always or most of the time. Hence, the measure of linkage disequilibrium between two markers will indicate the extent to which the genotype of one marker will predict the state of the other. Crucially, for association studies, the same principle applies to unseen polymorphisms, circumventing the need to identify all variants before study design. Currently, a large international effort, the HapMap project, is dedicated to the assembly of a genomewide haplotype map. The data from this project are freely available and should be used in all indirect association studies, in combination with techniques such as haplotype tagging,¹⁰ to ensure that future studies are comprehensively assaying as much genetic diversity as possible in an economic way.

Genotyping efficiency

Genotyping projects, regardless of the platform used, are susceptible to misclassification errors. Despite major advances in genotyping accuracy, a proportion of assays will remain inherently unreliable. Case-control studies are particularly vulnerable to genotyping errors, compared with family-based association tests, as misclassifications cannot be identified through segregation analysis. Small-scale or random misclassifications will not increase type I error, provided they occur independently of disease status; they will, however, result in a loss of power to detect association. Gordon *et al.*¹⁵ report that sample sizes must be increased by 2–8% to maintain constant type I and II error rates for each percentage point increase in genotyping error. Loss of power due to genotyping error could theoretically inflate conflicting reports of positive association, whereas systematic genotyping error could result in spurious association, particularly if only one SNP is being assayed per gene. On a practical level, commonly used countermeasures include the use of duplicate samples both within and between plates, the use of negative controls, and the checking of classification by resequencing a subset of samples of each genotype. Errors can also be detected during data analysis—for example, in family-based studies, genotypes can be checked for mendelian errors; genotype frequencies that depart significantly from the Hardy–Weinberg equilibrium can also be indicative of error, particularly if both cases and controls are affected; large numbers of very low-frequency haplotypes can also be indicative of genotyping error. Therefore, although genotyping error can be a definite problem, it is relatively easy to detect and control against. All association studies should report on the countermeasures used to detect genotyping errors and the frequencies at which they occur.

Statistical analysis

The trend in association studies is turning increasingly towards multiple hypothesis testing. Increasing numbers of candidate genes are being studied simultaneously, and approaches such as haplotype tagging require the analysis of several SNPs in a candidate gene to achieve coverage of the underlying genetic diversity. However, by increasing the number of tests, the probability of reaching a statistically significant false positive also increases. Therefore, to avoid inflation of type I error, the data should be corrected to reflect the increased probability of achieving spurious significance. The most widespread method currently used in molecular epidemiology is the Bonferroni correction, where the significance threshold is adjusted as a function of the number of tests carried out. Alternative Bayesian methods are increasingly being used in studies where the number of multiple tests is high, as the Bonferroni correction would be far too stringent.¹⁶ Bayesian methods use the prior probability of association to calculate the posterior probability of a false positive; an estimate of the prior probability is therefore required. The choice of which multiple correction test to use will ultimately depend on the individual study, but given that most association studies are attempting to uncover relatively small effect sizes and the numbers of tests are on the increase, all studies where multiple comparisons have been carried out should apply suitable corrections to their datasets.

Replication

Despite the inclusion of precautionary measures such as those discussed above in the design of an association study, the single most effective way of ensuring a true positive result is to replicate it in an independent cohort. Although the onus of proof has to date been transferred to independent groups after publication of a positive association, this practice has

resulted in the publication of a large number of false positive studies and further ambiguous and conflicting reports. This system has two inherent problems: the first is heterogeneity in study design and sample selection, which can result in the failure to replicate true positive associations, and the second is the publication bias in favour of positive reports, which results in an inflation of false positives in the literature. Therefore, to tackle the inconsistency of genetic association studies, we recommend that all positive associations be replicated in at least one appropriately sized independent cohort before publication.

CONCLUSIONS

The ultimate aim of all mapping strategies is to gain a better understanding of underlying disease pathways through the identification and functional characterisation of the genes involved. Positive genetic associations will therefore provide the groundwork in the process of delineating the functional variants that are associated with the predisposition or modulation of complex diseases. Although it is practically impossible to set definite rules for an association study, the implementation of a few safeguards in both study design and data analysis should decrease the number of false positive associations reported. There is also a need for greater transparency in the reporting of association studies to allow for more rigorous critical review; data sharing may also improve meta-analyses.

Authors' affiliations

P M Abou-Sleiman, N W Wood, Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London, UK
M G Hanna, Centre for Neuromuscular Disease, Institute of Neurology, Queen Square, London, UK

Competing interests: None declared.

REFERENCES

- 1 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996;**273**:1516–17.
- 2 Ioannidis JP, Trikalinos TA, Ntzani EE, *et al.* Genetic associations in large versus small studies: an empirical assessment. *Lancet* 2003;**361**:567–71.
- 3 Lohmueller KE, Pearce CL, Pike M, *et al.* Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;**33**:177–82.
- 4 Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet* 2005;**6**:95–108.
- 5 Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999;**65**:220–8.
- 6 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;**55**:997–1004.
- 7 Freedman ML, Reich D, Penney KL, *et al.* Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004;**36**:388–93.
- 8 Ardlie KG, Lunetta KL, Seielstad M. Testing for population subdivision and association in four case-control studies. *Am J Hum Genet* 2002;**71**:304–11.
- 9 Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;**361**:598–604.
- 10 Schaid DJ, Sommer SS. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am J Hum Genet* 1993;**53**:1114–26.
- 11 Weinberg CR. Studying parents and grandparents to assess genetic contributions to early-onset disease. *Am J Hum Genet* 2003;**72**:438–47.
- 12 Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;**19**:149–50.
- 13 Goldstein DB, Ahmadi KR, Weale ME, *et al.* Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet* 2003;**19**:615–22.
- 14 Wall JD, Pritchard JK. Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet* 2003;**4**:587–97.
- 15 Gordon D, Finch SJ, Nothnagel M, *et al.* Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum Hered* 2002;**54**:22–33.
- 16 Cordell HJ, Clayton DG. Genetic association studies. *Lancet* 2005;**366**:1121–31.