Neurology 2001;57:1153–1154

Genetic association studies

Genes in search of diseases

Thomas D. Bird, MD; Gail P. Jarvik, MD, PhD; and Nicholas W. Wood, PhD

The urge to associate genetic markers with human traits or diseases has been around for centuries. The association of specific hair colors with certain personality types has a long tradition in folklore. The rate-limiting factor has been the availability of measurable genetic markers. When the ABO blood groups were identified as early biological markers under genetic control, they were promptly used for association studies. For example, the association of type O with peptic ulcers was reported nearly 50 years ago.[1] (Interestingly, these blood group markers still emerge as associated with personality traits, presumably seeming more "scientific" than hair color.) Then, HLA genotyping stimulated another round of association studies. Although many putative HLA associations could not be reproduced, several were remarkably strong, such as those with ankylosing spondylitis and two of neurologic interest, MS and narcolepsy.[2,3] Now, with DNA-based genotyping, the identification of single-nucleotide polymorphisms (SNP), and the completion of the Human Genome Project, the volume of genetic association studies has increased from a trickle to a veritable cascade.

The basic idea underlying genetic association studies is both simple and important. Normal phenotype characteristics as well as diseases represent an interplay of environmental factors operating on a genetic background. Many common diseases are said to be complex, meaning they are oligogenic and multifactorial. That is, they are the end result of the complex effects of several or many genes interacting with the environment. To discover the predisposing genes is just as important as finding the environmental factors, and should provide clues to pathogenesis, treatment, and prevention.

For all their importance, genetic association studies have many potential problems that have been reviewed in detail.[4] AD is illustrative. One of the most remarkable and successful genetic association studies of the past decade has been the association of APOE alleles with AD.[5] This association stimulated new ideas about the causes and biology of AD and related disorders. Conversely, more than 50 other associations of AD with various genetic markers have been reported.[6,7] Many have not been replicated, most are not generally accepted, and the rest remain controversial. This prompted the editors of *Nature Genetics* to discuss the strengths and weaknesses of genetic association studies and to set stringent standards for acceptance of such reports for publication.[8]

The potential problems with genetic association studies include:

1. Accuracy of diagnostic criteria for the disorder to be studied. Investigators should provide evidence that all the subjects have the same disease.
2. Selection of appropriate control subjects, especially regarding age, sex, and ethnic background. The recent *Neurology* report by Payami et al.[9] on the lack of association of PD with CYP2D6 alleles demonstrated that even with adequate and appropriate control subjects, association may not be confirmed.
3. Choice of study strategy, such as using a population-based, case-control study vs a family approach (various affected sibling or affected family member protocols using transmission by descent tests [TDT]).
4. The problem of multiple comparisons (i.e., the high likelihood of a false-positive result occurring by chance because of large numbers of comparisons in the study).
5. Choice of statistical analysis and threshold for significance.
6. The tendency of both investigators and journals to report only studies with positive rather than negative results. As a result, the literature becomes heavily weighted toward unconfirmed associations.

Neurologic diseases will continue to represent frequently studied populations searching for genetic predisposing factors, particularly common disorders

See also pages 1304 and 1341

From the VA Medical Center (Dr. Bird), University of Washington (Drs. Bird and Jarvik), Seattle; and National Hospital for Neurology and Neurosurgery (Dr. Wood), London, United Kingdom.

Address correspondence and reprint requests to Dr. Thomas D. Bird, Geriatrics Research Service (S-182-GRECC), University of Washington, Veterans Affairs Medical Center, 1660 South Columbian Way, Seattle, WA 98108.

such as AD, PD, MS, stroke, and migraine. *Neurology* receives many genetic association manuscripts and this number is likely to increase. Two such well-done studies concerning MS and restless legs can be found in the current issue of *Neurology*.[10,11] In a carefully controlled study, Weinshenker et al.[10] were not able to confirm a previously reported association of MS with a polymorphism in the TNF-α gene. Desautels et al.[11] studied polymorphisms in eight genes involved in dopaminergic transmission in 92 subjects with restless legs syndrome and 182 control subjects. Even though these represented reasonable candidate genes for this syndrome, no association could be demonstrated. Although these are negative results, both studies help define the limits of genetic factors in these two common disorders.

We propose the following guidelines to authors of genetic association studies in order to judge for themselves whether their results are likely to be considered of sufficient power and interest to warrant publication. These guidelines reflect the excellent discussion of the strengths and weaknesses of genetic association studies reported by Cardon and Bell,[4] a reference we recommend to all investigators in this field. These guidelines are not hard and fast rules, but important points to consider in preparing and analyzing genetic studies. The goal is to publish rigorous, high-quality studies that are likely to provide new insights into genetic predispositions to neurologic diseases and results that are likely to stand the test of time.

- Numbers of subjects: By definition, common diseases are common, so investigators should study large numbers of cases. A study with more than 200 cases will be more robust than one with 50 subjects.
- Control subjects: Control subjects should be carefully matched for ethnic background, age, and sex. Control groups should be fully described and it should be clear if more than one group is used or if control data is being reused from a previous study.
- Analysis: Adjustments should be made for multiple comparisons. Authors should acknowledge all of the polymorphisms that they evaluated, rather than submitting a single significant result without addressing or correcting for having made multiple contrasts. Statistical criteria should be stringent, with p values of less than 0.01 and high odds ratio (OR) with CI. It is important that the CI excludes 1 if the OR is less than 2. An OR greater than 2 is desirable. When there is no a priori reason to test pooled genotypes, the appropriate test for three genotypes is a 2-degree of freedom analysis of variance (that is, $n - 1$ degrees of freedom for n genotypes). Similarly, logistic regression should test the genotype effect as an $n - 1$ degree of freedom test. A replication study or sample may use a 1 degree of freedom test when only one genotype is implicated. Informally screening the data for genotype effects and then pooling genotypes causes false-positive associations and does not accurately reflect the prior hypothesis.
- Biology: A plausible biological role for the relevant gene in the pathogenesis of the disease increases the possibility that the association is true. Ideally, the polymorphism should represent a functional change in physiology. It is recognized that such biological connections cannot always be made.
- Replication: Replication of the results in an independent group of subjects is especially convincing and should be accomplished whenever possible.
- Negative results: Journals should be willing to publish high-quality association studies with negative results, especially if a positive study has already been published in the same journal.

Investigators matching their investigations to these guidelines are likely to produce the most successful and valuable studies.

References

1. Aird I, Bentall HH, Mehigan JA, et al. The blood groups in relation to peptic ulceration and carcinoma of colon, rectum, breast, and bronchus. BMJ 1954;1:315–321.
2. Haines JL, Terwedow HA, Burgess K, et al. Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. Hum Mol Genet 1998;7:1229–1234.
3. Mignot E. Genetic and familial aspects of narcolepsy. Neurology 1998;50:S16–S22.
4. Cardon LR, Bell JI. Association study designs for complex diseases. Nat Rev Genet 2001;2:91–99.
5. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 1993;261:921–923.
6. St. George-Hyslop PH. Molecular genetics of Alzheimer disease. Semin Neurol 1999;19:371–383.
7. Schellenberg GD, D'Souza I, Poorkaj P. The genetics of Alzheimer's disease. Curr Psychiatry Rep 2000;2:158–164.
8. Editorial. Freely associating. Nat Genet 1999;22:1–2.
9. Payami H, Lee N, Zareparsi S, et al. Parkinson's disease CYP2D6 polymorphism, and age. Neurology 2001;56:1363–1370.
10. Weinshenker BG, Hebrink DD, Atkinson E, Kantarci OH. Association of a tumor necrosis factor α polymorphism with MS susceptibility. Neurology 2001;57:1341–1342.
11. Desautels A, Turecki G, Montplaisir J, et al. Dopaminergic neurotransmission and restless legs syndrome: a genetic association analysis. Neurology 2001;57:1304–1306.

Genetic association studies: Genes in search of diseases

Thomas D. Bird, Gail P. Jarvik and Nicholas W. Wood

Neurology 2001;57;1153-1154
DOI 10.1212/WNL.57.7.1153