

Editorial

Associations of Disease With Genetic Markers: Déjà vu All Over Again

Kenneth K. Kidd

The two articles in this issue [Pato et al., 1993; Crowe, 1993] and a recent review in the Journal of the American Medical Association [Gelernter et al., 1993] focus attention on two major problems in psychiatric genetics, indeed, on two problems for genetic studies of any complex disorder: optimal research design and weeding out of false positive findings. Specifically they address association studies involving a genetic marker and a disease. The problems in such studies are not new ones and were vigorously discussed during the mid 1970s when hundreds of studies compared HLA antigen frequencies in patients and controls for diverse diseases. An editorial at that time [Rosenberg and Kidd, 1977] addressed some of the issues in such studies. However, association studies in the human genetics literature go back much further based on such classical markers as ABO and GC; only a minuscule fraction of them yielded any insight into disease. Experience has led to skepticism and here we go, all over again.

What is meant by “association” when genetic markers and complex disorders are involved? Since the marker is a qualitative variable, concluding that there is an “association” means finding a statistically significant difference in the frequency of the marker when comparing patients with a control sample (of either unaffected individuals or randomly chosen individuals from the population). Two questions must be considered: 1) When is a difference statistically significant and 2) if significant, what does the difference mean? For studies of complex disorders and genetic markers both questions involve complex issues.

Let us consider the second question first. As pointed out by Rosenberg and Kidd [1977], a real association will only exist under three conditions: 1) the genetic marker is the etiologically relevant variable, 2) the sample is a stratified mixture derived from two (or more) populations which have different frequencies of the forms of the marker and different frequencies of the illness, and 3) linkage exists between the marker locus and the locus with etiologically relevant variation *and* alleles at those two sites are non-randomly associated on

chromosomes in the population. Situation 1 is a distinct possibility that can generate hypotheses worth testing in such cases as HLA antigens and diseases with autoimmune aspects [cf. Rosenberg and Kidd, 1976] or the extensive variation in the DRD4 coding sequence [Lichter et al., 1993] and diseases (or psychopharmacologic responses) that might involve this “clozapine receptor.” However, direct etiologic relevance is exceedingly unlikely, a priori, to exist for any DNA polymorphism in non-coding regions of a gene. Situations 2 and 3 are special cases of gametic phase disequilibrium, commonly referred to as linkage disequilibrium. Knowler et al. [1988] show that the highly significant association between type 2 diabetes and a particular allele at the Gm locus that they found in an Amerindian population is explained by levels of admixture with the associated allele being merely a marker of European ancestry. Thus, the associated allele has no etiological relationship to the disease. Situation 3 is the one most researchers cite for such studies, but note the compound nature of the condition—close linkage is not sufficient; linkage disequilibrium must also exist. How likely is linkage disequilibrium? There is no universally applicable answer to that question but there are many examples of very close markers that do *not* show linkage disequilibrium with the alleles causing a disease. PKU is an example in which disequilibrium is undetectable between any one of several RFLPs spanning the PAH locus and the class of PKU-causing alleles [Chakraborty et al., 1987] because, it was subsequently shown, two different mutations were common and they occurred on chromosomes with different marker alleles [cf. Kidd, 1987]. Whenever the disease-causing alleles are a collection of several different mutations, all relatively common, linkage disequilibrium is unlikely to exist for the class. While disequilibrium exists for individual PKU mutations, that was only revealed *post facto* [cf. Kidd, 1987]. One way around this problem is haplotype studies, but they involve considerable effort and then one must consider the point emphasized by Crowe [1993]—the high prior unlikelihood of most “candidate genes” for neuropsychiatric disorders.

Now let us return to the first question. “Statistical significance” is a framework of arbitrary conventions. Considerations relevant to disease-marker association studies include the number of independent tests performed, the prior evidence for (or against) a specific hypothesis, and levels of false positives and false negatives that can be tolerated.

This last point is greatly influenced by the prior support

Received for publication April 5, 1993; revision received April 5, 1993.

Address reprint requests to Dr. Kenneth K. Kidd, Yale University School of Medicine, Department of Genetics, 333 Cedar Street, New Haven, CT 06510.

one has for the hypothesis being tested. Crowe [1993] points out that it is not now possible to make a strong case for any particular gene being the (or even a) major cause of any neuropsychiatric disorder. Tens of thousands of genes appear to be expressed solely in the adult brain [Sutcliffe, 1988]. That estimate does not consider the huge number of genes controlling the development of the brain, the differentiation of the dozens (hundreds?) of types of neurons; essentially none of these development genes has been identified and their number has not been estimated. Bloom [1993] has recently emphasized that schizophrenia and other neuropsychiatric disorders appear more likely to be disorders of brain development than disorders of metabolism/function in a structurally normal brain. To date molecular neurobiologists have identified but a tiny fraction of the genes that could, by malfunctioning, cause structural and/or functional abnormalities in the central nervous system. Under these circumstances the choice of a specific candidate gene is not so much based on knowledge but rather mostly based on our ignorance of the highly complex genetic architecture of the central nervous system. For each "candidate gene" that is proposed, there are likely to be hundreds of other genes that, a priori, are equally likely candidates but have not been identified. Consider, for example, how little we know about "the dopaminergic system." There are actually several different interacting systems if we consider the inhibitory neurons, etc. We know nothing about the genes controlling the differentiation of these neurons and those controlling their interconnections. Even in the mature neurons, we know nothing about the regulation of the genes producing the proteins involved in the synthesis, reuptake, degradation, etc. of neurotransmitters, nothing about the signal transduction from the receptors, etc.

Crowe [1993] points out a major statistical distinction between linkage and association studies as currently carried out—the false positive rates. As he shows, because the prior likelihood is so low, the significance level required to have only 5% false positives among the "significant" association findings is $P < .00001$. What he did not emphasize is that a lod score of +3 was chosen by Morton [1955] to be the significance level precisely to deal with the problems of false positives arising from the high prior odds against any randomly chosen pair of loci being linked. Without a very stringent significance level the false positives would be more common than the true positives. Even at a lod score of +3, the proportion of false positives is 1 in 20—5%! Of course, we now know that an even more stringent criterion is required for linkage studies involving complex disorders and situations in which multiple markers are tested.

The statistical complications raised by having to correct significance levels for multiple tests and having to deal with a very high false positive rate have lead most researchers to accept consistent replication as the best evidence for a true association. That has not been met for the DRD2 and alcoholism studies for several reasons. Gelernter et al. [1993] present the best discussion of this but do not emphasize another related problem with the studies of DRD2 and alcoholism that is typical of many

association studies: definition drift. As various studies attempt to replicate an original finding, each slightly revises the definition of the disorder based on the maximum association found in the new study. Thus, the association of the A1 allele at the TaqIA marker system near DRD2 has gone from one with alcoholism to an association with severe alcoholism, to an association with addictive behaviors. If considered properly, this can be a perfectly acceptable aspect of the evolution of scientific understanding. However, in a strict sense each of these refinements of what diagnosis is associated with a marker is in fact a new *post hoc* hypothesis and not strictly a replication of the original findings. The problem comes from the misperception that each step is in fact a valid replication of the original finding when each would be better considered to be only an exploratory analysis generating a new hypothesis after failing to support the original hypothesis. While appropriate statistical procedures might be devised, depending upon the nature of the modification in the hypothesis, it seems unlikely that statistics could ever adequately deal with such *post hoc* changes.

A statistical problem highlighted by the comparison of the Pato et al. [1993] meta analysis and the Gelernter et al. review [1993] is the distinction between classifying individuals by genotype or by whether or not they are positive for the A1 allele (irrespective of homozygosity or heterozygosity) compared with still a third possibility for analysis, using the allele frequencies in the patient and control populations. Pato et al. [1993] have used the approach that Blum et al. [1990] originally used; classifying individuals by whether or not they had the DRD2**TaqI* A1 allele. In this scheme individuals heterozygous as well as those homozygous for this allele are considered equivalent. The approach used by Gelernter et al. [1993] is to calculate the frequency of the A1 allele. There are different fundamental philosophies underlying the different approaches. Classifying by A1 positive or negative implies that the A1 allele is itself causative, which seems most implausible. Examining the allele frequency in the two groups is more appropriate if linkage disequilibrium is involved since homozygotes for the marker are more likely to have the associated causative allele at the etiologically relevant locus.

And so, we come full circle to the two original questions of what is significant and what does it mean. The way people measure differences and the analyses they use to determine statistical significance often reflect their hypotheses of what is causing the differences. Ultimately, consistent replication of a finding is the surest measure of its being real. The association of a marker allele near the DRD2 locus with alcoholism fails to meet that test since both Pato et al. [1983] and Gelernter et al. [1993] demonstrate significant heterogeneity among studies.

When is it worth publishing a positive association? In linkage studies there is a growing consensus that a lod score of +5 or +6 is the minimum for a complex disorder. What would be comparable for an association study? That depends on the prior odds against any particular genetic marker, i.e., on the strength of the argument for a marker being a "candidate gene." If the variation at

the marker translates into functionally relevant variation, e.g., as at DRD4, then maybe a significance level of $P < .01$ represents sufficiently strong evidence (though clearly far from proof) to warrant publication. If the marker is like the vast majority of DNA-based genetic polymorphisms, no strong prior argument for direct relevance is possible and the logic outlined by Crowe [1993] is applicable. Such studies do not provide any meaningful evidence until a conventional significance level of $P < .0001$ is exceeded. Even if such significance levels are attained and the finding consistently replicated (also unlikely), the problem of interpretation still remains and the most likely explanation may still be sample stratification. Given all these problems, are association studies even worth doing?

REFERENCES

- Bloom FE (1993): Advancing a neurodevelopmental origin for schizophrenia. *Arch Gen Psychiatry* 50:224-227.
- Blum K, Noble EP, Sheridan PJ, Montgomery A, Ritchie T, Jagadeeswaran P, Nogami H, Briggs AH, Cohn JB (1990): Allelic association of human dopamine D₂ receptor gene in alcoholism. *JAMA* 263:2055-2059.
- Chakraborty R, Lidsky AS, Daiger SP, Guttler F, Sullivan S, Dilella AG, SLC Woo (1987): Polymorphic DNA haplotypes at the human phenylalanine hydroxylase locus and their relationship with phenylketonuria. *Hum Genet* 76:40-46.
- Crowe RR (1993): Candidate genes in psychiatry: An epidemiological perspective. *Am J Med Genet (Neuropsych Genet)* 48:74-77.
- Gelernter J, Goldman D, Risch N (1993): The A1 allele at the D2 dopamine receptor gene and alcoholism: A reappraisal. *JAMA* 269:1673-1677.
- Goldman D, Brown GL, Albaugh B, Robin R, Goodson S, Trunzo M, Akhtar L, Lucas-Derse S, Linnoila M, Dean M (1993): DRD2 dopamine receptor genotype, linkage disequilibrium and alcoholism in American Indians and other populations. *Alcoholism: Clin Exp Res* (in press).
- Kidd KK (1987): Phenylketonuria: The population genetics of a disease. *Nature News Views* 327:282.
- Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988): Gm^{3,5,13,14} and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* 43:520-526.
- Lichter JB, Barr CI, Kennedy JL, Van Tol HHM, Kidd KK, Livak KJ (1993): A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. *Hum Molec Genet* (in press).
- Morton NE (1955): Sequential tests for the detection of linkage. *Am J Hum Genet* 7:277-318.
- Pato CN, Macciardi F, Pato MT, Verga M, Kennedy JL (1993): Review of the putative association of DRD2 and alcoholism: A meta-analysis. *Am J Med Genet (Neuropsych Genet)* 48:78-82.
- Rosenberg L, Kidd KK (1977): HLA and disease susceptibility: A primer. *N Engl J Med* 297:1060-1062 (Editorial).
- Spielman, RS, McGinnis RE, Ewens WJ (1993): Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506-516.
- Sutcliffe JG (1988): mRNA in the mammalian central nervous system. *Annu Rev Neurosci* 11:157-198.