Editorial

Eliminating the Improbable
Sherlock Holmes and Standards of Evidence in the Genomic Age

David Herrington, MD, MHS

In 2003, Shearman and colleagues reported that a common variant in the estrogen receptor-alpha gene (ESR1 IVS1 −397 C allele) occurred more frequently in male subjects with myocardial infarction (MI) than in those free of MI. The following year, Schuit et al described an association between the common haplotype containing the alternate (T) allele and MI, but only in women. In this issue of Circulation, Koch et al report no association, in either men or women, between the ESR1 −397 T>C genotype or the common haplotypes containing this allele and MI. What does this collection of heterogeneous results tell us about genetic association studies in general and genetic variation in the estrogen receptor in particular; and how can we use this understanding to improve our ability to understand, treat, and prevent cardiovascular disease?

Frequent failure to replicate initially promising genetic associations is a source of consternation and confusion for scientists and journal editors alike. Recent reviews of the literature indicate that 70% to 95% of reported genetic associations were not confirmed in subsequent studies. It has been argued that false-positive reports of genetic associations distract energy and resources from valid lines of inquiry and erode the credibility of medical research. Some reputable journals now decline to publish genetic associations of complex disorders except in “exceptional circumstances” (http://www.jci.org/misc/jcipoli.pdf), and some scientists have argued that association studies are so unreliable that they should be abandoned altogether.

“The temptation to form premature theories on insufficient evidence is the bane of our profession.”

—Sherlock Holmes, in “The Valley of Fear”

In their simplest form, genetic association studies determine whether the frequency of an allelic variant is more or less common in cases than it is in appropriately selected controls. Like any case-control study, in ensuring that the cases are representative of all cases and controls are representative of all controls, case definitions are precisely defined and uniformly applied; exposure variables, in this case the genetic variants, are measured with the least amount of error possible that is necessary to avoid biased conclusions. Studies also need adequate power to have a high likelihood of detecting a difference, if one really exists. To interpret any collection of genetic association studies with inconsistent results, these details of study design need to be considered first.

“It has long been an axiom of mine that the little things are infinitely the most important.”

—Sherlock Holmes, in “A Case of Identity”

There are, however, additional considerations unique to genetic association studies that also need to be addressed. For instance, allele frequency of many genetic variants differs according to ancestry. Thus, an imbalance in the genetic ancestry of cases versus controls (population stratification) may mask real associations or produce spurious ones. Self-reported race/ethnicity is an inadequate means to capture this genetic variability caused by ancestry, especially in admixed populations. It is unclear to what extent population stratification may have contributed to inconsistent genetic association findings in the past. There are now several analytic tools that make use of panels of unlinked genetic markers to estimate and/or adjust for ancestry in tests of association. These techniques have the added benefit of controlling for unmeasured environmental factors that also differ by genetic ancestry.

Another concern has to do with the difference between tests of association with marker alleles (common) versus tests with true causal variants. The likelihood of a marker allele being associated with a genetically mediated trait depends on its correlation (linkage disequilibrium) with a true causal variant. Before the development of rapid sequencing and genotyping technologies, many association studies used sparse sets of markers (sometimes even single nucleotide polymorphisms [SNPs] within a gene) selected more for their ease of detection than their functional potential or ability to represent the genetic diversity in a gene or region. The value of some of these markers in past association studies is highly questionable. To further complicate matters, linkage disequilibrium is not uniform across populations, making it possible for a marker allele to be associated with a trait in one population but not another. Fortunately, new efforts to document the linkage disequilibrium architecture of the human genome and in multiple populations make it increasingly possible to select sets of informative SNPs that provide comprehensive coverage of an entire gene or region and are

2081
expected to be more reliable indicators of the presence or absence of association with traits of interest.

Perhaps the biggest concern related to the problem of replication is that the staggering number of gene polymorphisms (∼11 million with allele frequency >1%),17 coupled with the newly developed technical ability to detect them in large numbers of subjects, has produced an irresistible opportunity to examine multiple variants with respect to multiple phenotypes, only to report the most promising and seemingly coherent associations.18 In most cases, such statistically significant associations are merely type I errors. Subsequent attempts to replicate the associations either fail or find that the strength of the association is less dramatic than initially reported—the "winners curse."5,19 Unfortunately, conventional adjustments for multiple independent tests (eg, Bonferroni) are too conservative when the tests are performed on correlated phenotypes (eg, hypertension, left ventricular hypertrophy, heart failure) or when the tested genotypes are commonly inherited together (ie, in the same haplotype block), leaving open the door of ambiguity about the real level of significance for any individual finding. Multivariate modeling, cluster analysis, haplotype analysis, and permutation tests in the setting of correlated phenotypes or genotypes can be used to reduce the chance of a type I error, but they still may not completely overcome the problem, especially if all of the comparisons performed are not completely described in the published reports.

Another way to look at this problem, strangely enough, is based on a concept well known to clinicians but frequently forgotten by investigators: The prior probability of a positive test greatly influences the credibility of a test result. Often, investigators errantly assume the P value equals the likelihood that the observed difference is false. (The P value is actually the likelihood that an observed difference occurred because of chance alone when, in fact, there is no real difference.) In fact, for any level of statistical significance (typically P=0.05), the likelihood of a false-positive result is determined by the prior probability of a positive test and, to a lesser extent, the power to detect a difference, if one really exists (Figure). This is just as true for exercise stress tests as it is for tests of association in genetic case-control studies.

The problem with genetic association studies, and especially genome-wide association studies, is that the prior probability of an association being true is generally quite low. Consider, for example, a trait with 5 causal variants. If there are 50 plausible candidate genes and an average of 20 independent variants per gene, there are 1000 possible associations to consider. The prior probability that any of the possible associations are real is only 5/1000=0.005. Even in a study with perfect power, 90% of the observed associations with a P<0.05 will be false positives.20

“I understand, however, from the inquest that there were some objects which you failed to overlook.”
Sherlock Holmes, in “The Adventure of Black Peter”21

One solution is to increase the prior probability of a positive test. In the past, investigators have tried to improve prior probabilities by focusing only on the most promising candidate genes or on the variants most likely to alter the structure or abundance of encoded proteins. Assessing the strength of evidence in support of specific genetic variants in advance is a subjective process, however, and its past success is subject to debate.4 Moreover, focusing only on genetic associations with high prior probability will miss currently unrecognized molecular pathways or novel genetic features that regulate gene expression.

The other solution is to simply study more subjects. This will allow more stringent P values to reduce false positives while retaining power to detect real associations. Several authors have argued that samples sizes of ≥1000 cases and controls and P values in the range of 10⁻³ to 10⁻⁵ and are needed to produce reliable evidence of genetic association in a single study.22,23 Unfortunately, relatively few investigators have access to such large cohorts with uniformly collected phenotypic and confounding variable data.

An alternative strategy is to use ≥2 smaller studies to identify and replicate associations. Under the right conditions, this can produce statistical evidence of an association that is comparable to one large study. Even two studies with comparable associations at the P=0.05 level may not provide compelling evidence of a real association if the prior proba-
bility is sufficiently low. Likewise, failure to replicate an initial observation in a subsequent genetic association study does not guarantee that the initial association was wrong. Studies designed to replicate an initial association need to anticipate that the magnitude of the gene effect is likely to be inflated in the initial report. For this reason, replication studies should be designed to confirm or refute a smaller effect size than that observed in the initial report, ideally with greater than the conventional 80% power. Unless the design of the replication study is identical to that of the initial study, inconsistent results may also occur because of differences in definitions of the trait or populations used for analysis.

Ultimately, the best evidence for genetic association will likely rely on compilation and meta-analysis of data from multiple studies. Interestingly, one of the best examples of a coordinated effort to pool data from multiple smaller cohorts for a meta-analysis of a genetic association involved the same ESR1 variants examined in the article by Koch et al. In this example, the traits of interest were a different set of estrogen-sensitive phenotypes: bone mineral density and fractures. In the analysis of nearly 20,000 subjects, there was strong evidence for a 20% to 50% lower risk for fractures in subjects homozygous for the ESR1 T>G allele or the common haplotype also containing the ESR1 T>C allele. These data provide compelling proof of principle that genetic variation in ESR1 may be associated with at least some clinical phenotypes.

What then can we say about the data from Koch et al with respect to ESR1 variants and risk for MI, and how do these data relate to the earlier findings of Shearman and Schuit and their colleagues? In the study of Koch et al, the selection of controls from subjects referred for angiography with normal coronaries is a potential problem in that these subjects are unlikely to be the same as healthy subjects who are not referred for angiography. However, the distribution of alleles in the control subjects resembles the allele frequency distribution found in other normal populations of European descent, suggesting that no bias is present. The fact that the subjects were recruited from a homogenous population in southern Germany diminishes the likelihood of population stratification, but this was not formally evaluated. The study is clearly the largest of the 3, with >3000 cases compared with only 59 subjects with MI in the report by Sherman et al and 285 in the study by Schuit et al. On the basis of this fact alone, it would appear that the Koch et al study should provide more reliable evidence concerning the presence or absence of an association than either of the other studies. The confidence limits for the association between the −397 T>C variant and MI in the Koch et al study indicate that a positive or negative association as extreme as observed in the other 2 studies is highly unlikely; however, odds ratios as small as 0.9 or as large as 1.2 are still possible. These may seem like small effects, but they are in fact consistent with the common disease/common variant model of gene effects and are in line with the apparent association between the same variants and risk of fracture.

It is important to note that none of the studies provides a comprehensive evaluation of the genetic variation in the ESR1 gene as it relates to risk for MI. The 2 SNPs included in all 3 studies only provide information about common haplotypes for one of the roughly 20 haplotype blocks that exist in the ESR1 gene in white subjects of European descent. In addition, there are no established functional effects of either of the SNPs, although there is suggestive evidence of a potential effect of the −397 T>C on a transcription factor binding site. Thus, more information is needed about other potentially functional SNPs in linkage disequilibrium with the SNPs included in these studies, as well as SNPs in other regions of the ESR1 gene before making claims with confidence about genetic variation in the ESR1 gene and risk for MI.

These studies also provide an opportunity to comment on the publication process for genetic association studies. If we attempt to constrain genetic epidemiology, or at least reported genetic epidemiology to a minimum of false-positive results, then we run the risk of replacing irrational exuberance over genotyping with imprudent pessimism concerning potentially valid genetic associations. With careful attention to study design and use of contemporary methods to address population stratification, linkage disequilibrium, and multiple testing, there is good reason to expect more consistent and informative results from genetic association studies in the future. It is important to acknowledge that most of science is in fact a recursive process of identifying patterns in observable events followed by testing to confirm whether the pattern is reproducible. The medical literature is the necessary forum for this recursive process, even if it means occasionally publishing false-positive or true negative results. The emphasis should be on studies that are performed with methodological rigor and accompanied by a realistic assessment of the strength of evidence for or against a putative association. To that end, the editors of Circulation should be commended for publishing an important negative study that adds to our still-incomplete fund of knowledge about the cardiovascular effects of the estrogen receptor and its genetic variants.

"Any truth is better than indefinite doubt." Sherlock Holmes, in "The Yellow Face"

References

Key Words: Editorials ■ genetics ■ myocardial infarction ■ cardiovascular diseases ■ women
Eliminating the Improbable: Sherlock Holmes and Standards of Evidence in the Genomic Age
David Herrington

Circulation. 2005;112:2081-2084
doi: 10.1161/CIRCULATIONAHA.105.570069

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2005 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/112/14/2081

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org/subscriptions/