Genetic association studies using single nucleotide polymorphisms (SNPs) and insertion/deletion variants are a common feature on the atherosclerosis research landscape. A recent Medline search using the terms “[gene] AND [polymorphism] AND [X],” where X was “atherosclerosis,” “vascular biology,” “thrombosis,” or “lipoprotein,” found >4000 original articles. Furthermore, the yearly number of new reports has been growing exponentially since 1983 (Figure). The allure of SNPs, the release of millions of SNP-based markers from dedicated consortia, and the availability of cost-effective high-throughput detection methods are converging to create a potential explosion of genetic association studies in atherosclerosis. Although the standards and quality seem to be improving, there is nevertheless a risk that SNP-based association analyses will squander academic trust and scientific resources owing to unsatisfactory design and/or analysis.

Like all experimental designs and model systems, genetic association studies in human samples have strengths and limitations.1–3 Their potential strengths include the simplicity of design, ease of noninvasive sampling, reliability and cost-effectiveness of genotyping, uncomplicated statistical analysis, and the potential for clear interpretation and direct relevance to human biology. But many factors collude to undermine confidence in association studies. Often the initial publication of a positive association is followed by reports of non-replication or refutation. There can be good reasons for non-replication, including complexity of mechanisms, multiplicity of causative genes, confounding by gene-environment interactions, and context-dependency of the associations. However, the pattern of non-replication of genetic associations is frequent, familiar, and disconcerting. An index of the tenuous nature of genetic associations in atherosclerosis is that few DNA markers are in routine clinical use, such as in risk stratification protocols, although the application of markers for disease prediction is admittedly distinct from their use in experimental hypothesis testing.

Limiting the potential to publish false-positive (FP) or false-negative (FN) results can be achieved in many ways. For instance, one journal recently expanded its editorial criteria for rapid rejection to include “gene association studies related to complex disorders, including . . . atherosclerotic heart disease” (http://www.jci.org/misc/jcicipol.pdf). This policy can be justified as an effective means to eliminate the likelihood of ever publishing a false, non-replicable result from a genetic association study. However, the bath water could hold an occasional baby. In general, the principle of editorial scrupulousness and placing restrictions on the publication of genetic association studies is valid in the current research environment, but might there be some standards for association studies that would not close the door on their publication, while simultaneously maximizing confidence in their validity? It is hoped that this editorial will stimulate a dialogue among the contributors, editorial staff, and readers of Arteriosclerosis, Thrombosis, and Vascular Biology regarding desirable attributes for genetic association studies in atherosclerosis.

The Basics of Genetic Association Analysis

Using genetic markers as putative atherosclerosis determinants, even risk factors, is intuitively attractive. The classic Hill criteria for a causative relationship between a determinant and an outcome require that an association be temporally correct, specific, strong, and consistent, with clear biological plausibility, with an evident dose-response relationship.4 Genetic markers for atherosclerosis—long regarded as the molecular correlates of family history—can be visualized as fitting such classic criteria. Genetic association studies seem to provide a direct way to probe the relationship between genome and disease. But there is more to association analysis than meets the optimistic eye.

In contrast to genetic linkage studies, which investigate correlations between inheritance of a trait and chromosomal regions within family units such as sibling pairs or multigenerational pedigrees, association studies test for differences in genetic marker frequency between affected cases and unaffected controls. Linkage studies have had many notable successes in identifying the molecular basis of monogenic diseases, but fewer successes with common, more complex phenotypes, such as atherosclerosis. True lasting, replicable successes with association studies are also infrequent, despite the fact that they are more commonly performed because in their simplest form they are not constrained by a requirement for family units.

One common study design for association analysis is: 1) ascertain cases with a trait of interest, such as atherosclerosis or a related phenotype; 2) assemble matched controls without the trait; 3) obtain DNA samples; 4) genotype all subjects for a marker thought to be etiologically important, or with a set of markers covering the genome; and 5) statistically compare allele or genotype frequencies in cases versus controls. Perhaps an even more commonly used alternate procedure for quantitative measures, which would apply to many atherosclerosis-related phenotypes, is to enter the genotype as an independent variable in a multivariate regression analysis that
assesses sources of variation of the continuous trait within a single study sample.

**Weaknesses of Genetic Association Studies**

Case-control genetic association studies suffer from both general problem types that afflict non-genetic case-control studies and from specific problem types that are unique to genetic studies. As with any case-control approach, there are many sources of bias. For instance, unsatisfactory designation of cases and controls is a limitation in this study design. Cases and controls must be representative and must be matched as closely as possible, except for the phenotype of interest, which must be clearly defined, using reliable diagnostic procedures with stated performance attributes.

Etiologic and genetic heterogeneity underlying the affected status of cases or the quantitative trait of interest can obscure the identification of an association, especially in small samples, thus increasing the probability of a FN result. Narrowing the selection criteria for cases, such as defining specific sub-phenotypes using stringent criteria, can reduce the risk of heterogeneity, and can possibly increase the signal-to-noise ratio for a true association. In addition, defining cases and controls should not be influenced by knowledge of individual genotypes. As well, the genotype assay must be uniformly performed using reference standards, and must be blinded to the phenotype. In theory, it is possible to match cases and controls molecularly, by using DNA markers from random genomic locations that have low-to-no priori possibility of being associated with the trait of interest. However, procedures and standards for such molecular matching have not yet been established.

Most criticisms of association studies stem from their potential to generate FP results. High-throughput genotyping and multiple hypothesis testing increase the threat that associations will result from chance alone. For example, if 20 markers are genotyped, but only the two best results are reported without commenting on the actual amount of genotyping, the findings are misleading to the reader. Reports of association studies should thus include a statement of the numbers of markers actually genotyped. Realistic analysis of statistical power should be included, with appropriate assumptions of homogeneity and genetic effect sizes.

Increasing the stringency of the significance level by using an accepted adjustment procedure is a crucial step to control the risk of a FP result. However, this also could reduce statistical power in some association studies that may already be under-powered. Another way to reduce FP results is to increase the previous probability of causation by selecting candidate genes based on evidence of a role in the phenotype from functional, expression, or genetic mapping experiments. Most association studies thus use candidate gene SNPs.

FP associations can also result from inadequate matching of controls and cases, often due to systematic differences. For instance, some population substrata may be more susceptible to the phenotype because of an unmeasured factor, like ethnicity or genetic background, for which the genotype is merely an indirect marker. There are several approaches to reduce population stratification artifacts, such as the use of more restrictive family-based association designs, although some of these approaches may require additional scrutiny to fully define their appropriate use and limitations. In any event, greater awareness of this problem seems to be reducing the hazard of stratification artifacts for recent better-designed studies.

Although true positive or true negative findings should be replicable, there may be valid reasons why this is not always the case. For instance, imagine that four independent investigators conduct separate case-control association studies of the same genetic marker and atherosclerosis phenotype. Assuming a true association and similarity of effect and sample sizes, and a power of 0.85 for each study, the probability that all four studies will detect the association is \( \approx 50\% \). Furthermore, because atherosclerosis and related phenotypes are so complex, the actual assortment of etiologies may vary across samples, making replication less likely between samples.

Publication bias for positive results occurs early in a polymorphism’s history. Also, for some polymorphisms, just as negative associations between a marker/locus and a phenotype were being reported, new positive associations involving the same marker and more distal phenotypes were being published, with a lag period before failure to replicate is demonstrated. The ACE intron 16 insertion/deletion genotype is illustrative: while large meta-analyses were discounting the association with myocardial infarction, smaller case-control studies were reporting new associations with seemingly more remote traits, such as serum triglycerides, exercise tolerance and “leukoaraiosis.”\(^7\)–\(^10\) Replication and persistence of the initial association through systematic, planned meta-analysis and review would appear to be crucial. The rationale to search for new associated phenotypes would need to be justified and withstand critical peer-review for each new study. Also, there might be more convenient ways to publish association studies, with smaller article formats for both initial reports and replication studies.

There are also problems related to the use of SNPs themselves as genetic markers. These include the low power afforded in general by biallelic systems and the inability to examine haplotype phase in unrelated study samples. Most identified SNPs are biologically neutral or have no known function. When a SNP is selected as a marker for a locus,
there may be no functional basis for a relationship with the phenotype. Instead, the SNP may be in linkage disequilibrium with an unmeasured functional variant, which may be weak or variable in large complex populations. Using larger numbers of SNPs might help provide a few functional polymorphisms. If many SNPs at a locus are used, linkage disequilibrium between them should be estimated. Finally, and most importantly, tests of direct association are convincing if a selected SNP has a demonstrated functional consequence.

Is There Any Hope for Genetic Association Studies?

While genetic association studies have significant limitations, they sometimes represent the only practical approach to begin to address a particular biological hypothesis. For new genes or for known genes with an interesting protein product, genetic association analysis may be the fastest initial approach to show a relationship with a biological end point. Because interest in SNPs and association analyses will continue to grow, it is important to recognize their limitations. Some weaknesses of genetic association studies and possible ways to confront them are shown in the Table.

At a bare minimum, a desirable genetic association study should have: 1) justifiable biological rationale; 2) appropriate selection and sampling procedures; 3) rigorous phenotyping and genotyping procedures; 4) large samples; 5) appropriate probability values; and 6) physiologically meaningful evidence supporting a functional role of the polymorphism. Reports that contain an initial study and an independent replication would be particularly desirable, especially if different sampling and/or analytical strategies were used. To reduce the risk that the authors, editorial staff, and readers will be confronted with false results, explicit guidelines for genetic association studies may be required. While few studies will meet all criteria, the confidence in the results will probably be proportional to the number of met criteria. The standards for “desirability” will continue to evolve as insights into complex traits and analytic strategies improve.
Acknowledgments
The author is grateful for laboratory support by grants from the Canadian Institutes for Health Research, the Heart and Stroke Foundation of Ontario, the Canadian Genetic Diseases Network, the Canadian Diabetes Association, and the Blackburn Group. The author is a Career Investigator of the Heart and Stroke Foundation of Ontario and holds a Canada Research Chair (Tier I) in Human Genetics.

References
SNP Judgments and Freedom of Association
Robert A. Hegele

doi: 10.1161/01.ATV.0000026801.56080.14

Arteriosclerosis, Thrombosis, and Vascular Biology is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2002 American Heart Association, Inc. All rights reserved.
Print ISSN: 1079-5642. Online ISSN: 1524-4636

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://atvb.ahajournals.org/content/22/7/1058

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Arteriosclerosis, Thrombosis, and Vascular Biology 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 Arteriosclerosis, Thrombosis, and Vascular Biology is online at:
http://atvb.ahajournals.org//subscriptions/