

VIEWPOINT

## Taking Cardiovascular Genetic Association Studies to the Next Level

Geoffrey S. Ginsburg, MD, PhD,\*† Svati H. Shah, MD, MHS,† Jeanette J. McCarthy, PhD\*‡  
*Durham, North Carolina*

Genetic information is beginning to have a direct impact on patient care and it is important that cardiologists appreciate the value and approaches to associating genetic variation and health outcomes. Genetic associations should be based on compelling genetic and biological hypotheses and should be statistically sound so as to reduce the possibility of “false discovery” in the setting of testing multiple hypotheses. Study designs should clearly define cases and controls and measurement of phenotypes. Finally, findings should be replicated in at least 1 independent cohort. Consideration of these principles should provide insight into disease biology based on genetic findings and encourage their meaningful adoption into clinical practice. (J Am Coll Cardiol 2007;50:930–2) © 2007 by the American College of Cardiology Foundation

Cardiovascular medicine is undergoing a paradigm shift from acute intervention to predictive and preventive care. The latter approach has been afforded largely through the availability of the human genome sequence and technologies that allow us to access its vast and complex information content. The enthusiasm for these new approaches, particularly genetic association studies, to predicting acute coronary syndromes, myocardial infarction, sudden death from arrhythmias, and response to commonly used cardiovascular medications needs to be tempered by a critical evaluation of data supporting their clinical validity. To this end, we believe that it is important for every cardiologist to appreciate the value and approaches to associating genetic variation and outcomes such that the true meaning of these findings can be readily assimilated into the mainstream of cardiovascular patient care.

It has been over 10 years since the *Journal of the American College of Cardiology (JACC)* published its first genetic association studies. At that time there was a flurry of papers on the angiotensin-converting enzyme gene insertion/deletion polymorphism and its association with idiopathic cardiomyopathy, myocardial infarction, left ventricular hypertrophy, or the lack of these (1–4). A prescient accompanying editorial stated, “The increasingly widespread availability and easy applicability of molecular genetic tools provides the research community with a formidable opportunity to, at long last, begin uncovering the heritable

component of complex diseases. In many cases . . . these tools may not be applied with the stringency that an epidemiologic geneticist may deem appropriate. However, as long as we view the data with critical reservation . . . dissemination of such studies is vital to stimulate the field, to usher in new developments, and to generate new hypotheses that eventually will be tested in more robust study designs” (5). Many journals, including *JACC*, continue to strive toward these goals of dissemination of innovation, particularly in genetics and genomics. However, in the case of genetic studies, in our opinion, it is now time to raise the bar and increase the stringency for publication, particularly in terms of study design.

The last decade has seen seminal milestones in genetics: the complete sequencing of the human genome, the completion of HapMap, the development and enhancement of statistical analysis tools, and the banking of deoxyribonucleic acid (DNA) from large populations. Today, with genotyping a mere commodity and many clinical researchers collecting DNA routinely as part of longitudinal cross-sectional studies and clinical trials, one can easily envision a deluge of single gene/single nucleotide polymorphism (SNP) studies to generate new hypotheses of the type envisioned by Lindpaintner and Pfeiffer (5). It is most certainly true that some reported “positive” associations are the result of false discovery, and concomitantly there is a negative publication bias that ensues because the vast majority of “negative” associations are not reported on at all.

The cardiovascular literature is now robust with positive genetic association studies (6), but few have become clinically meaningful. Many have failed to replicate outside the original study, and for others replication has not been attempted. Many reasons exist for failure to replicate, particularly for complex polygenic disorders, including in-

From the \*Duke Institute for Genome Sciences and Policy, Durham, North Carolina; and the †Division of Cardiology, Department of Medicine, and ‡Department of Community and Family Medicine, Duke University Medical Center, Durham, North Carolina.

Manuscript received April 20, 2007; revised manuscript received May 11, 2007, accepted May 23, 2007.

appropriate SNP selection, failure to account for confounding variables, lack of sufficient power in replication cohorts, analyses in inappropriate subgroups, and genetic and phenotypic heterogeneity. Invaluable resources have gone into procuring DNA from patients, SNP detection, genotyping, analysis, and publication, the latter raising false hopes and expectation for a novel tool that will benefit patients and make physicians' diagnostic and therapeutic acumen more precise. Moreover, as a result of some of these errant studies, the "hype" of genetics' potential contribution to medicine now casts a shadow on its real opportunity to change the field. We believe that there is now a need for higher standards and greater uniformity in the design of studies in pursuit of novel genetic susceptibility loci for cardiovascular diseases and that, if they adopt these standards, studies will be more likely to contribute to the field of cardiovascular health care.

Both readers and authors of genetic association studies should look for certain features in the evaluation of these studies and in deciding whether the "take-home message" is one that is of potential clinical import.

**Association studies should be based on a compelling genetic hypothesis.** If the study is not the first to examine a variant or set of variants at a particular locus, all the previous genetic studies of this locus should be summarized as they relate to the specific phenotype under study. If it is the first study of a gene, supporting biological data that makes the gene a good candidate should be outlined. Good candidate genes will have multiple lines of evidence to support a possible role in disease, including, but not limited to, expression in the appropriate tissue, differential expression in experimental models of disease/normal tissue, phenotypes of transgenic or knockout animals, and so forth, or a location consistent with previous linkage or whole-genome association study results. Although the gene is not required to be a positional candidate, the chromosomal location of the gene and whether it is consistent with linkage or whole-genome association hits should be provided. Even in whole-genome association studies that are inherently unbiased, a biological hypothesis for selection of the gene or genes further evaluated on the basis of the study should be clearly formulated.

**Association studies should provide a clear rationale for selection of SNPs for study.** With 1 SNP occurring in as few as every 185 bases and as many as 80 SNPs per gene, it is no longer sufficient to genotype 1 or a few randomly selected SNPs in a gene to draw any concrete conclusions. The exception might be an SNP that has been clearly demonstrated and accepted to be functional, *i.e.*, whose translation product results in a protein with altered function. But even then, we cannot ignore the possibility of allelic heterogeneity at the locus, with different functional alleles arising in distinct founder populations. Studies should aim to unambiguously define the methods and process used for selecting one particular set of SNPs over another. Because haplotypes reduce the need for defining individual SNPs,

ideally a comprehensive survey of haplotype-tagged SNPs that capture common variation and functional SNPs should be included.

If more than 1 SNP per gene is examined, measures of linkage disequilibrium (LD) between the loci should be presented. If only 1 SNP is examined, it would be useful to provide public HapMap data on LD in the region of that variant, when it exists, with the goal of delineating the boundaries of the association. This would allow readers as well as the authors to define the relevant "blocks" of DNA that are being covered by the studied variants.

Details on allele frequency—particularly for the ethnic group(s) under investigation—as well as any information on known (or putative) functionality of SNPs should also be included to develop and allow readers to understand the rationale for SNP selection.

**Association studies should provide a solid biological foundation for the findings.** This notion directly stems from the preceding concept, that functional SNPs be given a significant weight in selection for study. When a genetic finding is directly tied to the biology underlying the disease of interest, more likely than not it will be clinically relevant or lead to clinically relevant findings. Investigations should strive to provide novel functional data pertaining to the gene and associated genes and pathways—or substantiation of these from the literature—and to provide a measure of the effect of the gene, genes, or pathway gene products in the human cohort(s) studied. These data will greatly strengthen the plausibility of the association. For example, if the study claims a positive association of a functional C-reactive protein (CRP) promoter polymorphism with myocardial infarction, the study should also investigate the effect of the variant on the intermediate phenotype of CRP levels.

**There is a need to reduce the possibility of "false discovery" in the setting of testing multiple hypotheses.** In many studies, numerous variants, genetic models, and phenotypes are tested in the study population, and often within substrata of the population. This multiple hypothesis testing increases the opportunity for type I error. Appropriate statistical analyses to correct for multiple independent comparisons (Bonferroni correction, Monte-Carlo simulation, and permutation testing, false discovery rate) should be performed and the results discussed in this context. A design that contemplates validating the findings in a second data set is preferred to reduce the possibility of a false positive finding.

**Findings should be replicated in at least 1 independent cohort.** In the early days of genetic association studies, merely finding an association was significant. Today, with the deluge of associations published and the complexity of the study design, populations, and phenotypes, this is no longer sufficient. Clinical relevance of these findings will

#### Abbreviations and Acronyms

<b>CRP</b>	= C-reactive protein
<b>DNA</b>	= deoxyribonucleic acid
<b>LD</b>	= linkage disequilibrium
<b>SNP</b>	= single nucleotide polymorphism

only be assured if they are robust to replication. This may be one of the most important aspects of genetic associations and establishing their relevance to medicine and broad populations. If the study is a replication study, then the original study or studies that identified the candidates and the populations in which the association was originally described should be clearly described. Furthermore, there should be clear and careful delineation of the phenotype in both the initial and validation cohorts, with discussion regarding potential dissimilarities (and, therefore, potential confounders). If it is an original study of a novel SNP, then evidence of replication of the association in an independent cohort should be provided. Thus true replication obviates the need for power calculations, because the replication shows that the study is sufficiently powered to detect the association.

**When are negative results important?** We recognize the negative publication bias incurred by not publishing studies that show a lack of association. There may be cases of import if such a study makes a significant contribution to the scientific literature. In many cases, such studies might include the testing of undisputable functional SNPs or comprehensive genotyping (tagging) to exclude association with at least common variants of the gene. There also must be a demonstration of adequate power to detect effect sizes that are below the original reported association, consistent with the lower confidence limit of the original effect measured (7). It should also be recognized that it is very difficult to rule out an association with a gene owing to complexities of interaction, allelic heterogeneity, and the effect of rare variants that are not always captured.

**Study designs should be clear on the selection of cases and controls or cohorts, definition and measurement of the phenotype, and acknowledgment of potential biases and confounding.** What are the clinical characteristics or phenotype the study is aiming to associate with underlying genetic variation? Phenotypic definition must be clear for both cases and controls and consistent with standards of practice or clinical guidelines where possible (8). For example, in association studies for coronary artery disease, distinctions need to be made for myocardial infarction versus angiographic coronary disease phenotypes. If standards exist that define a phenotype, event, or outcome, then these

should be used in the definition of the association under study. Confounding by ethnicity, or population stratification, is of particular concern in genetic association studies. At a minimum, the potential for population stratification in the study should be addressed and, ideally, methods should be used to control for such confounding, such as conducting family-based studies or using genomic control or ancestry informative markers.

It is difficult to prescribe with exactitude the ideal genetic study. The goal for both readers and writers of genetic association studies is to ensure that the evidence will instigate investigation into mechanism and biology based on genetic findings and encourage the performance of studies that will enable the meaningful adoption of genetics into clinical practice.

---

**Reprint requests and correspondence:** Dr. Geoffrey S. Ginsburg, Duke Institute for Genome Sciences and Policy, Box 3382, Durham, North Carolina 27710. E-mail: geoffrey.ginsburg@duke.edu.

---

#### REFERENCES

1. Montgomery HE, Keeling PJ, Goldman JH, Humphries SE, Talmud PJ, McKenna WJ. Lack of association between the insertion/deletion polymorphism of the angiotensin-converting enzyme gene and idiopathic dilated cardiomyopathy. *J Am Coll Cardiol* 1995;25:1627–31.
2. Andersson B, Sylven C. The DD genotype of the angiotensin-converting enzyme gene is associated with increased mortality in idiopathic heart failure. *J Am Coll Cardiol* 1996;28:162–7.
3. Samani NJ, O'Toole L, Martin D, et al. Insertion/deletion polymorphism in the angiotensin-converting enzyme gene and risk of and prognosis after myocardial infarction. *J Am Coll Cardiol* 1996;28:338–44.
4. Perticone F, Ceravolo R, Cosco C, et al. Deletion polymorphism of angiotensin-converting enzyme gene and left ventricular hypertrophy in southern Italian patients. *J Am Coll Cardiol* 1997;29:365–9.
5. Lindpaintner K, Pfeffer MA. Molecular genetics crying wolf? The case of the angiotensin-converting enzyme gene and cardiovascular disease. *J Am Coll Cardiol* 1995;25:1632–3.
6. Ginsburg GS, Donahue MP, Newby LK. Prospects for personalized cardiovascular medicine: the impact of genomics. *J Am Coll Cardiol* 2005;46:1615–27.
7. Schafmayer C, Tepel J, Franke A, et al. Investigation of the Lith1 candidate genes ABCB11 and LXRA in human gallstone disease. *Hepatology* 2006;44:650–7.
8. Luo AK, Jefferson BK, Garcia MJ, Ginsburg GS, Topol EJ. Challenges in the phenotypic characterization of patients in genetic studies of coronary artery disease. *J Med Genet* 2007;44:161–5.