Challenges in the phenotypic characterization of patients in genetic studies of coronary artery disease

Albert K Luo, Brian K Jefferson, Mario J Garcia, Geoffrey S Ginsburg and Eric J Topol

J. Med. Genet. published online 20 Dec 2006; doi:10.1136/jmg.2006.045732

Updated information and services can be found at:
http://jmg.bmj.com/cgi/content/abstract/jmg.2006.045732v3

These include:

Rapid responses
You can respond to this article at:
http://jmg.bmj.com/cgi/eletter-submit/jmg.2006.045732v3

Email alerting service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

Online First contains unedited articles in manuscript form that have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Online First articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Online First articles must include the digital object identifier (DOIs) and date of initial publication.

To order reprints of this article go to:
http://www.bmjjournals.com/cgi/reprintform

To subscribe to Journal of Medical Genetics go to:
http://www.bmjjournals.com/subscriptions/
Challenges in the phenotypic characterization of patients in genetic studies of coronary artery disease

Albert K. Luo, BA
Brian K. Jefferson, MD
Mario J. Garcia, MD
Geoffrey S. Ginsburg, MD, PhD
Eric J. Topol, MD

Running title: Genetic Studies of CAD

From the School of Medicine, Case Western Reserve University, Department of Cardiovascular Medicine, Cleveland Clinic, Cardiovascular Institute, Mount Sinai School of Medicine, Institute for Genome Science and Policy, Duke University, and Department of Genetics, Case Western Reserve University

Key Words: Genetics, Coronary Artery Disease, Myocardial Infarction, Atherosclerosis

Address for correspondence: Eric J. Topol MD, Department of Genetics, Case Western Reserve University, BRB 724, 10900 Euclid Avenue, Cleveland, OH 44106-4955, eric.topol@case.edu; 216-368-4943; fax 216-368-0491

Abstract
Coronary artery disease and acute myocardial infarction are complex traits in which there has been recent research to identify the principal genes that engender susceptibility or provide protection. Although there has been exceptional progress in the technology that now allows genotyping of hundreds of thousands of single nucleotide polymorphisms in each individual, there remains a pattern of inconsistency in the studies performed to date, in part owing to the difficulties in defining cases and controls. In this paper, we review the salient issues to facilitate research in this important field.

Introduction
In recent years there has been some meaningful progress in the identification of genes that are associated with susceptibility to the development of coronary artery disease (1-3). But there has also been remarkable lack of replication from one study to the next, and difficulty in identifying genes that underlie impressive linkage peaks (4-12). A primary reason for the lack of consistency in the studies performed to date may be attributable to the variability in phenotypic characterization of the cases and controls, or in linkage studies, differentiation between “affected” and “unaffected” family members. This is especially challenging for the complex trait of coronary artery disease, since it has a wide range of clinical manifestations, from asymptomatic to acute myocardial infarction (MI) and sudden cardiac death. CAD is generally of late-onset, in the sixth or seventh decade of life, and there are multiple anatomical and physiological modalities to assess the condition. In this paper, we review the specific challenges of phenotypic characterization to facilitate progress in this vital and burgeoning field of research.

Background of Inconsistency
Despite an extensive body of research dedicated to the genomic basis of coronary artery disease, there are few genes that have been identified which have been independently replicated. Perhaps the most telling lack of consistency is demonstrated by the aggregate consideration of 8 linkage studies involving genome wide scans of sibling pairs (4-12) (Table 1). Despite the finding of a locus of interest in each of the published reports, there was not one primary locus that was the same between the multiple linkage studies. Of note, the patient inclusion criteria varied considerably, with some studies focused on premature coronary artery disease, which has a higher level of heritability, while others enrolled patients with typical, later onset CAD. As a result, almost a 20 year age gap separates the youngest and oldest cohorts that have been assessed. Although some investigators only analyzed the data for CAD, others used MI as the primary phenotype of interest (6-8, 10, 11). The definitions of CAD and MI were not consistent across the studies. Besides the heterogeneity of phenotype and age, there were important differences in ancestry, sample size, and how the data were analyzed. The issues of heterogeneity of ancestry and population stratification are especially important. Even among inbred mouse models, there are marked difference in atherosclerosis susceptibility across strains. Two multiplex family genome wide scans have identified genes underlying a linkage peak has not yet had independent replication. For ALOX5AP, which was found to be associated with both MI and stroke, there has only been replication
for stroke (13). The other leukotriene pathway gene variant finding, LTA4H, was derived from the same genome-wide scan. Of note, this variant had an ancestry-specific risk of MI; the specific phenotype was late-onset MI with peripheral vascular disease, or stroke or all 3 associated atherosclerotic conditions. Although the study showed replication of this particular phenotype in 3 additional cohorts (Atlanta, Cleveland and Philadelphia) beyond the Icelandic in which it was first identified, there has yet to be independent investigator replication. Of note, the definitions of controls in the 3 American cohorts varied from those without a coronary angiogram, to < 10% narrowing, to <50% stenosis (11). The other linkage peak which has led to identification of a specific gene is the GATA-2 finding (14), which has not yet been independently replicated.

Lack of replication has led to controversy over the finding of deletion mutant in MEF2A, a transcription factor, which was identified in a large pedigree with an autosomal dominant inheritance pattern of MI and CAD (15-19). The MEF2A 21bp deletion in exon 11, the stop codon, co-segregated with the presence or absence of CAD or MI in the family, and was not found in hundreds of control subjects without documented CAD by angiography (15). Furthermore, this deletion was correlated with lack of nuclear translocation and a marked reduction of transcription activity. But Weng et al identified a family with the same MEF2A deletion with a proband who had a transient cerebrovascular attack (16, 17). Two of her siblings were reported to not have CAD or MI, but neither of these individuals had undergone coronary angiography and only one had stress testing. Beyond the 21bp deletion, point mutations in exon 6 and 7 were reported to be associated with an increased risk of MI, further implicating the association of MEF2A with coronary disease, and one of these point mutations Pro297Leu has been independently replicated (18,19). This replication study had coronary angiographic data for the cases, but not controls (19).

High throughput genotyping has suggested several genes that are associated with MI, although not CAD, including THBS-4, which has been independently replicated (20-22) but also many that have not yet been confirmed by other investigators (22). The first genome wide association study identified LTA as a susceptibility gene for MI, and this finding has been refuted by one negative replication study (23, 24). Recently, Shiffman and colleagues have published 2 genome wide association studies, one for MI of typical age onset and another for early-onset, in which SNPs were screened in 3 different cohorts to provide internal replication (25). But none of the 6 gene variants that have been collectively identified to be associated with MI by these 2 studies have been confirmed yet by others. Moreover, the definition for MI varied considerably between all of the association studies that have thus far been published (20, 22-26). Accordingly, the inconsistencies and difficulties in replication may, in part, be related to the lack of uniform phenotypic definitions.

Defining Cases
Unfortunately, atherosclerotic involvement of the coronary arteries is an exceptionally common phenomenon. In a recent study of young individuals, from teenage to 35 years, who had succumbed to motor vehicle accidents and underwent post-mortem, the vast majority already had subclinical atheromatous
coronary plaques (27). Knowing the disease is relatively endemic in the Western world, and expected to become the dominant cause of death and disability worldwide by 2020 (28), it is a challenge to define a threshold which partitions to having or not having coronary artery disease (CAD).

If a patient presents with angina, has a functional test which demonstrates ischemia, and then undergoes a coronary angiogram that shows critical stenosis of \( \geq 70\% \) narrowing of one or more epicardial artery, then an indisputable diagnosis of CAD can be made. But a large proportion of patients who present with CAD do not have symptoms of angina, and only surface because they have had a routine stress test which is abnormal. If intravascular ultrasound is used, the majority of patients who undergo coronary angiography will be found to have at least some or even substantial plaque accumulation in the arterial wall but without any encroachment of the arterial lumen by the corresponding angiogram (29). This occurs because during the early phases of atherosclerotic development, there is extensive arterial remodeling with lipid pool accumulation in the arterial wall while fully preserving normal luminal dimension. By strict anatomic definition, this could be classified as coronary artery disease with even quantification of the plaque volume burden. But this is complicated because it may be occult, and not associated with any symptoms of angina or abnormalities of functional testing for ischemia. Furthermore, the coronary angiogram is an incomplete assessment that only demonstrates the accumulation of plaque which results in narrowing the arterial lumen. A substantial proportion of patients have irregularities of their angiographic borders which correspond to some plaque accumulation. Complicating this matter further, there is often a gray zone—a patient with atypical symptoms, an abnormal stress test, and a 50-60\% sub-critical narrowing. Should this patient be classified as having CAD? And should the patient with “silent” ischemia who has a tight narrowing be considered a case? There are just some of the vagaries in the definition of CAD at any given time of assessment.

The complexity takes another step higher when one considers the multiple longitudinal angiographic studies which demonstrate that in a \( \geq 6 \) month time span, an individual can progress from having a mild (\(< 30\%\)) narrowing to a critical \( >70\% \) stenosis (30). This can result from a plaque fissure, erosion or rupture event and typically occurs superimposed at a site of only minimal luminal encroachment. Thus, it is important to acknowledge that the assessment of a “case” is only relevant to the actual time in which the study was performed, and that this is a dynamic phenomenon. Safeguards are therefore needed to preempt the potential re-categorization of a control to case.

Patients who present for a coronary angiogram represent a bias of ascertainment. It is important to acknowledge that much of the disease is occult with lack of definition by coronary angiography because symptoms may be non-specific or non-existent.

The phenotype of myocardial infarction, one of the manifestations of CAD, is more restrictive and has thus far proven more useful in terms of identifying susceptibility genes. But it, too, leaves us with a skewed group of patients. For patients to have presented with the triad of (1) symptoms of protracted chest pain or associated, unequivocal signs of ischemia, (2) classic electrocardiographic ST segment elevation changes, and (3) enzymatic confirmation of myocardial necrosis, a diagnosis of MI is straightforward. However, this is not at all
A significant minority of patients with MI presents with sudden cardiac death and never reaches the hospital, leading to a survival bias. Such patients are systematically excluded from any genetic study, translating MI genetic investigation to an initial non-fatal presentation phenotype. An even larger proportion of patients have an acute coronary syndrome without ST segment elevation MI, but instead have a normal ECG, or non-specific ECG abnormalities, including ST segment depression, or T wave changes, or both. There are important differences between ST segment elevation MI, which involves occlusive coronary thrombus and more extensive myocardial necrosis, and non-ST segment elevation MI which typically occurs with mural, non-occlusive thrombosis, may result from more extensive collateral flow, and results in less necrosis. The difference in clinical and angiographic profile may be accompanied by differences in genetic predisposition. Thus, the specific type of MI would be helpful to know in any given cohort or subgroup that is analyzed.

The age of onset of MI is an especially important determinant. The genome wide association studies of Shiffman et al showed completely different gene variants were associated with MI by defining the phenotype as a function of age (25,26). The linkage study which focused on the youngest MI cohort appeared to have the most yields in terms of significant loci of interest (8). Thus, specific age and type of MI, and the fact that we are dealing only with initial, non-fatal MI are all important considerations to acknowledge. Finally, it is important to note that MI represents a very small fraction of the individuals who have the CAD phenotype and it is likely that there are specific susceptibility genes that play a role for both phenotypes and only one of these processes. There are vast pathophysiologic differences between the chronic accumulation of arterial plaque as compared to a sudden plaque rupture event with attendant thrombosis. Consideration for a set of criteria for cases is provided in Table 2.

**Defining Controls**

The “ideal” control might be conceived as an individual who has lived to 90 years or more and has undergone a post-mortem which demonstrates completely normal coronary arteries. But this is not so “ideal” for many reasons. First, finding such individuals is nearly impossible because it is so rare for autopsies to be performed, especially in individuals of such advanced age. Second, since cases are typically age 50 to 60 years of age, there is a substantial gap in longevity that may well be confounding in genetic analysis. Third, the use of a pathologic definition as compared to younger individuals who have coronary angiography likely is a mismatch, with the latter modality far less sensitive.

From a practical point of view, it would be ideal to identify controls without having to subject individuals to an invasive assessment using coronary angiography. However, stress testing, even with adjunctive nuclear or echocardiographic imaging, will miss a significant proportion of patients with atherosclerotic coronary involvement. Thus, the only way that one can reliably define anatomic CAD or lack of the disease would involve angiography. Currently the “gold” standard technique is an invasive diagnostic angiogram, requiring cannulation of the left and right coronary ostia and direct arterial injection of contrast dye. This is an expensive procedure which carries some, albeit low, risk of morbidity and
fatality. Even lack of a significant stenosis by an angiogram may not be a suitable definition of a control.
Syndrome X, which is thought to be attributable to small vessel disease, and is more frequent in women who typically present with debilitating chest discomfort, is associated with normal appearing coronary arteries (30,31). But such patients deserve separate genetic studies and would be inappropriate to be classified as controls because of the “normal” appearance of the epicardial arteries by angiography.
One of the most important points of controversy is what should be the threshold for a control insofar as extent of any narrowing or irregularity of the angiographic arterial borders. Quantitative angiography is not generally used, so this assessment relies on a subjective interpretation of the angiogram for what may be interpreted as truly normal or “pristine” appearance of the angiogram in all of the 3 major epicardial and branch vessels. But such an appearance of a truly “normal” angiogram is not particularly common, as compared to finding individuals who have minor irregularities with no frank luminal encroachment that would approximate a 10% narrowing. The actual threshold of what can be accepted as “normal” classification has yet to be resolved.
An important confounding effect in our experience is to find that a large proportion of patients who have a true “normal” appearing angiogram have presented to the cardiac catheterization laboratory for evaluation of valvular heart disease, undergoing the pre-operative standard is to assess the coronaries to exclude occult CAD. But this introduces another potentially obfuscating variable if the controls have valvular heart disease, even though they have “normal” coronary arteries, because they certainly are not without other cardiovascular phenotypic abnormalities. Potential criteria for categorization of controls are provided in Table 2.
Era of Cardiac Computed Tomography
A non-invasive way to anatomically define coronary disease may be to use cardiac computed tomography. Both electron-beam (EBCT) and multi-detector computed tomography (MDCT) have established accuracy and reproducibility for the detection and quantification of coronary calcifications. The burden of coronary calcification is represented in the “calcium score” and correlates well with the extent of calcification measured by histopathology (32). Both the presence and extent of calcification appear to be independent predictors of cardiac events in large population studies (33). However, the absence of coronary calcification does not completely rule out the finding of coronary artery disease. Acute coronary events may occur in younger subjects who do not demonstrate coronary calcifications. Conversely, most elderly individuals who have calcified coronary plaques have never experienced a cardiac event. The prevalence and extent of coronary calcification increase with age, suggesting that there may be a lag period during which coronary atherosclerotic plaques may be active before they develop calcifications (34). For this reason, a “negative calcium score” may be inadequate to reliably exclude coronary artery disease in young or middle age individuals. More recently, MDCT in particular, has been shown to provide high-resolution 3-dimensional images of the coronary vessels, allowing both visualization of the contrast-enhanced lumen and atherosclerotic plaques in the vessel walls,
including those that are non-calcified. In recently published studies, the volume of coronary atherosclerotic plaque determined by CT correlated well with the volume of plaque as determined by IVUS(35) and plaque not obstructing the lumen was detected (36). Moreover, based on specific local X-ray attenuation, CT may estimate the constituents of atherosclerotic plaques (37). Thus, non-calcified plaques that are more lipid-rich appear to have lower Hounsfield units (HU) values than fibrocalcific or calcified plaques. Current 64-row MDCT scanners are now able to detect most plaques greater than 0.5 mm in diameter. Although spatial resolution is limited compared to IVUS, MDCT is non-invasive, less expensive and provides coverage of the entire coronary tree. CT coronary angiography is limited to patients in normal sinus rhythm and like invasive coronary angiography it requires contrast injection, but intravenous rather than intra-arterial. While there is a considerable radiation exposure, this new technique may eventually be well suited for defining controls.

**Affected Versus Unaffected Individuals**

While the accurate categorization of cases and controls represents a critical step for association studies, the parallel challenge in linkage studies is to appropriately identify who is affected with coronary artery disease, and who is not affected. A significant proportion of family members may not be properly classified, because they are either too young to know for certain that CAD or MI would not develop later in life, or that suitable functional or anatomical testing has not been performed. In contrast to cases and controls there is a family history such that the level of risk and concern for an individual is higher. Accordingly, there may be justification to perform exercise testing with nuclear or echocardiographic imaging, and the possibility for justification of MDCT for equivocal functional test results may be enhanced.

**Consideration for Future Studies**

On the basis of the concerns that have been reviewed, we present some considerations that may facilitate replication and progress in this research arena (Table 2).

For cases defined as having the CAD phenotype, angiographic confirmation of > 70% stenosis of a major epicardial artery, with correlation of symptoms and/or a definitively abnormal functional test seems appropriate. For the MI phenotype, “premature” characteristically refers to age < 45 years for men or < 50 years for women for the index event (8). It may be useful to define the type of MI, categorized as ST segment elevation, non-ST elevation, or an acute coronary syndrome without evidence of myocardial necrosis.

The definition of controls is more challenging than cases. Angiography be available, either invasive or by CT angiography, with either “normal”, minor irregularities, or appearance of < 30% narrowing in the absence of visualized atherosclerotic plaques. Ideally, controls would all be “normal” but it is unrealistic when a very careful assessment of the angiogram is performed looking for luminal irregularities. It seems that controls should not have other forms of atherosclerotic disease, such as stroke or transient ischemic attack or peripheral vascular disease. The absence of risk factors may be helpful (Table 2). If controls
are used with valvular heart disease, an adjustment in the analysis will be necessary to control for this important co-variate. One of the important considerations is matching cases and controls. There exists the distinct possibility that controls may, later in life, become cases. Accordingly, controls that are at least 10 to 20 years older than cases may represent a useful assessment to pre-empt this concern. While there is the trade-off of an age mismatch, the accuracy of phenotype may be overriding.

An important classification is the “indeterminate” group. If an individual is less than age 60, it would be difficult to project that CAD or MI would not occur later in life. Similarly, in patients who are older, age > 70, but who have not undergone functional testing, there may be the issue of occult CAD. An age of > 70 years with functional testing, or anatomical definition with coronary angiography or MDCT, may be necessary for accurate phenotyping. Otherwise, in family studies individuals could be regarded as “indeterminate” and for selecting controls younger patients or those without suitable screening may best be avoided for use as controls. Family studies of this disease are further complicated by the frequent phenomena of “phenocopies” because CAD is quite prevalent and may occur independently from genetic susceptibility.

As summarized in Table 2, some of the criteria for cases and controls would make patient accrual more challenging. Quality assurance of categorization would be enhanced with the use of formal adjudication of phenotype by an independent group. With the difficulties inherent in fulfilling these strict criteria, setting up a common resource of cases and controls, across different ancestry groups, for use in validation and discovery studies, may be a particularly worthwhile consideration.

**Conclusion**

Extraordinary progress is being made in defining the genetic underpinning of complex traits. Recent examples include complement factor H for age-related macular degeneration, transcription factor 7 2-like for type II diabetes mellitus and a yet unnamed microsatellite variant for prostate cancer (38-40). One of the advantages in progress for these other disease conditions has been a well accepted phenotype with clearcut, standard definition. Careful consideration to phenotypic definitions will hopefully catapult the research efforts in identifying CAD susceptibility genetic variants in the future.

Acknowledgement: The work was supported by a Specialized Center of Clinically Oriented Research (SCCOR) NIH Grant P 50 HL077107
References


34 Greenland P, LaBree L, Azen SP, Doherty TM and Detrano RC. Coronary artery calcium score combined with Framingham score for risk prediction in asymptomatic individuals. *JAMA*. 2004;291:210-5.


Table 1. Genome wide scans for coronary artery disease and myocardial infarction

<table>
<thead>
<tr>
<th>Ref</th>
<th>Study Population</th>
<th>Families</th>
<th>Age</th>
<th>1°Locus</th>
<th>Analysis Program</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pajukanta 4</td>
<td>Finnish</td>
<td>156</td>
<td>&lt;55</td>
<td>2q21</td>
<td>MAPMAKER</td>
<td>ND</td>
</tr>
<tr>
<td>Francke 5</td>
<td>Mauritian</td>
<td>99</td>
<td>47</td>
<td>16p13</td>
<td>GENEHUNTER</td>
<td>ND</td>
</tr>
<tr>
<td>Broeckel 6</td>
<td>European</td>
<td>513</td>
<td>52</td>
<td>14</td>
<td>SOLAR</td>
<td>ND</td>
</tr>
<tr>
<td>Harrap 7</td>
<td>Australian</td>
<td>61</td>
<td>62</td>
<td>2q36</td>
<td>MAPMAKER</td>
<td>ND</td>
</tr>
<tr>
<td>Wang 8</td>
<td>American</td>
<td>428</td>
<td>44</td>
<td>1p34</td>
<td>SAGE</td>
<td>ND</td>
</tr>
<tr>
<td>Hauser 9</td>
<td>Euro-Amer</td>
<td>438</td>
<td>56</td>
<td>3q13</td>
<td>GENEHUNTER</td>
<td>ND</td>
</tr>
<tr>
<td>Helgadottir 10</td>
<td>Icelandic</td>
<td>296</td>
<td>--</td>
<td>13q12</td>
<td>ALLEGRO</td>
<td>ALOX5AP</td>
</tr>
<tr>
<td>Helgadottir 11</td>
<td>Icelandic</td>
<td>296</td>
<td>--</td>
<td>12q22</td>
<td>ALLEGRO</td>
<td>LTA4H</td>
</tr>
<tr>
<td>Brit Ht Found 12</td>
<td>British</td>
<td>1933</td>
<td>53</td>
<td>2p11</td>
<td>ALLEGRO</td>
<td>ND</td>
</tr>
</tbody>
</table>

Abbreviations: Brit Ht Found-British Heart Foundation, Euro-Amer-European-American, ND-not determined. Modified from Reference 2.
Table 2. Considerations for Phenotype Categorization

Cases for Coronary Artery Disease

- Angiographic stenosis of $\geq$ 70% stenosis in a major epicardial artery
- Family history of coronary artery disease
- No tobacco history
- No diabetes
- Normal LDL and HDL cholesterol, normal C-reactive protein

Cases for Myocardial Infarction

- All of the above plus
- Documented of MI by ECG, enzymes

Controls

- Normal coronary arteries by selective coronary angiography or multidetector CT (no $> 10\%$ narrowing)
- No family history of coronary artery disease
- No history of cerebrovascular or peripheral artery disease
- Age $\gg$ cases by 10-20 years

*Above criteria adjudicated by a panel of independent investigators
The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive license (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in JMG and any other BMJPG products and sublicenses such use and exploit all subsidiary rights, as set out in your license (http://JMG.bmjournals.com/ifora/licence.pdf)