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

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

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, which 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, salient issues to facilitate research in this important field are reviewed.

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 (CAD). 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. A primary reason for the lack of consistency in the studies performed to date may be the variability in phenotypic characterisation 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 CAD, as it has a wide range of clinical manifestations, from asymptomatic to acute myocardial infarction 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 characterisation 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 CAD, there are few genes that have been identified which have been independently replicated. Perhaps the most telling lack of consistency is shown by the aggregate consideration of eight linkage studies involving genome-wide scans of sibling pairs (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 CAD, 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 analysed the data for CAD, others used myocardial infarction as the primary phenotype of interest. The definitions of CAD and myocardial infarction 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 analysed. The issues of heterogeneity of ancestry and population stratification are especially important. Even among inbred mouse models, there are marked differences in atherosclerosis susceptibility across strains.

Two multiplex family genome-wide scans that have identified genes underlying a linkage peak have not yet had independent replication. For ALOX5AP, which was found to be associated with both myocardial infarction and stroke, there has only been replication for stroke. The other leucotriene pathway gene variant finding, LT4H, was derived from the same genome-wide scan. Of note, this variant had an ancestry-specific risk of myocardial infarction; the specific phenotype was late-onset myocardial infarction with peripheral vascular disease, or stroke or all three associated atherosclerotic conditions. Although the study showed replication of this particular phenotype in 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 three American cohorts varied from those without a coronary angiogram, to <10% narrowing, to <50% stenosis. The other linkage peak that has led to identification of a specific gene is the GATA2 finding, which has not yet been independently replicated.

Lack of replication has led to controversy over the finding of a deletion mutant in MEF2A, a transcription factor, which was identified in a large pedigree with an autosomal dominant inheritance pattern of myocardial infarction and CAD. The MEF2A 21 bp deletion in exon 11, the stop codon, cosegregated with the presence or absence of CAD or myocardial infarction in the family, and was not found in hundreds of controls without documented CAD by angiography. Furthermore, this deletion 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

Abbreviations: CAD, coronary artery disease; MDCT, multidetector computed tomography

See end of article for authors’ affiliations

Correspondence to:
Dr E J Topol, Scripps Genomic Medicine, The Scripps Research Institute, MEM-275, 10550 North Torrey Pines Rd, La Jolla, CA 92037, USA; etopol@scripps.edu

Received 3 August 2006
Revised 30 October 2006
Accepted 12 November 2006
Published Online First
15 December 2006

who had a transient cerebrovascular attack.\(^{16} 17\) Two of her siblings were reported as not having CAD or myocardial infarction, but neither of these individuals had undergone coronary angiography and only one had stress testing. Beyond the 21 bp deletion, point mutations in exon 6 and 7 were reported to be associated with an increased risk of myocardial infarction, 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 for the controls.\(^{20}\)

High throughput genotyping has suggested several genes that are associated with myocardial infarction, although not CAD, including THBS4, which has been independently replicated\(^{20–22}\) but also many that have not yet been confirmed by other investigators.\(^{23}\) The first genome-wide association study identified LTA as a susceptibility gene for myocardial infarction, and this finding has been refuted by one negative replication study.\(^{24}\) Recently, Shiffman et al\(^{25}\) have published two genome-wide association studies, one for myocardial infarction of typical age onset and another for early-onset, in which single nucleotide polymorphisms were screened in three different cohorts of patients to provide internal replication. But none of the six gene variants that have been collectively identified to be associated with myocardial infarction by these two studies have been confirmed yet by others. Moreover, the definition for myocardial infarction 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 that 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 CAD.

If a patient presents with angina, has a functional test that shows ischaemia, and then undergoes a coronary angiogram which shows critical stenosis of \(\geq 70\%\) narrowing of one or more epicardial arteries, 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, most 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 remodelling with lipid pool accumulation in the arterial wall while fully preserving normal luminal dimension. By strict anatomical definition, this could be classified as CAD 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 ischaemia. Furthermore, the coronary angiogram is an incomplete assessment which only shows the accumulation of plaque that results in narrowing of 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 grey zone—a patient with atypical symptoms, an abnormal stress test, and a 50–60% subcritical narrowing. Should this patient be classified as having CAD? And should the patient with “silent” ischaemia who has a tight narrowing be considered a case? These are just some of the vagaries in the definition of CAD at any given time of assessment.

The complexity increases when one considers the multiple longitudinal angiographic studies that show that, in a \(\geq 6\)-month time period, an individual can progress from having slight (<30%) narrowing to 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 recategorisation of a control to a 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 ischaemia, (2) classic electrocardiographic ST segment elevation changes, and (3) enzymatic confirmation of myocardial necrosis, a diagnosis of myocardial infarction is straightforward. However, this is not at all representative. A significant minority of patients with myocardial infarction present with sudden cardiac death and never reach the hospital, leading to a survival bias. As patients are systematically excluded from any genetic study, translating myocardial infarction genetic investigation to an initial

---

**Table 1: Genome-wide scans for coronary artery disease and myocardial infarction**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study population</th>
<th>Families</th>
<th>Age (years)</th>
<th>Primary locus</th>
<th>Analysis program</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papadakis(^a)</td>
<td>Finnish</td>
<td>156</td>
<td>-55</td>
<td>2q21</td>
<td>MAPMAKER</td>
<td>ND</td>
</tr>
<tr>
<td>Francke(^b)</td>
<td>Mauritian</td>
<td>99</td>
<td>47</td>
<td>16p13</td>
<td>GENEHUNTER</td>
<td>ND</td>
</tr>
<tr>
<td>Broeckel(^c)</td>
<td>European</td>
<td>513</td>
<td>52</td>
<td>14</td>
<td>SOLAR</td>
<td>ND</td>
</tr>
<tr>
<td>Horr(^d)</td>
<td>Australian</td>
<td>61</td>
<td>62</td>
<td>2q26</td>
<td>MAPMAKER</td>
<td>ND</td>
</tr>
<tr>
<td>Wang(^e)</td>
<td>American</td>
<td>428</td>
<td>44</td>
<td>1p34</td>
<td>SAGE</td>
<td>ND</td>
</tr>
<tr>
<td>Hauser(^f)</td>
<td>Euro-Amer</td>
<td>438</td>
<td>56</td>
<td>3q13</td>
<td>GENEHUNTER</td>
<td>ND</td>
</tr>
<tr>
<td>Helgadottir(^g)</td>
<td>Icelandic</td>
<td>296</td>
<td>-</td>
<td>13q12</td>
<td>ALLEGRO</td>
<td>ALGKSAP</td>
</tr>
<tr>
<td>Helgadottir(^h)</td>
<td>Icelandic</td>
<td>296</td>
<td>-</td>
<td>12q22</td>
<td>ALLEGRO</td>
<td>LTA4H</td>
</tr>
<tr>
<td>Brit Hf Found(^i)</td>
<td>British</td>
<td>1933</td>
<td>53</td>
<td>2p11</td>
<td>ALLEGRO</td>
<td>ND</td>
</tr>
</tbody>
</table>

[Reference 1]: [www.jmedgenet.com]
non-fatal presentation phenotype. An even larger proportion of patients have an acute coronary syndrome without ST segment elevation myocardial infarction, 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 myocardial infarction, which involves occlusive coronary thrombus and more extensive myocardial necrosis, and non-ST segment elevation myocardial infarction, 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 profiles may be accompanied by differences in genetic predisposition. Thus, it would be helpful to know the specific type of myocardial infarction in any given cohort or subgroup that is analysed.

The age of onset of myocardial infarction is an especially important determinant. The genome-wide association studies of Shiffman et al.\(^a\) showed that completely different gene variants were associated with myocardial infarction by defining the phenotype as a function of age.

The linkage study that focused on the youngest myocardial infarction cohort seemed to have the most yields in terms of significant loci of interest.\(^b\) Thus, specific age and type of myocardial infarction cohort seemed to have the most yields in terms of significant loci of interest.\(^a\) Thus, specific age and type of myocardial infarction, and the fact that we have been dealing only with initial, non-fatal myocardial infarction are all important considerations to acknowledge. Finally, it is important to note that myocardial infarction 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 in both phenotypes and only one of these processes. There are vast pathophysiological differences between the chronic accumulation of arterial plaque as compared with a sudden plaque rupture event with attendant thrombosis. Box 1 provides considerations for a set of criteria.

### Box 1 Considerations for phenotype categorisations (the following criteria were adjudicated by a panel of independent investigators)

- **Cases for coronary artery disease**
  - Angiographic stenosis of ≥70% in a major epicardial artery
  - Family history of coronary artery disease
  - No history of smoking
  - No diabetes
  - Normal low density cholesterol and high density cholesterol, normal C reactive protein.

- **Cases for myocardial infarction**
  - All of the above and
  - Myocardial infarction documented of by ECG, enzymes.

- **Controls**
  - Normal coronary arteries by selective coronary angiography or multidetector CT (<10% narrowing)
  - No family history of coronary artery disease
  - No history of cerebrovascular or peripheral artery disease
  - Age much much greater than cases by 10–20 years.

### DEFINING CONTROLS

The “ideal” control might be conceived as an individual who has lived to ≥90 years and has undergone a postmortem that shows 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, as the cases are typically 50–60 years of age, there is a substantial gap in longevity that may well be confounding in genetic analysis. Third, the use of a pathological definition as compared with younger individuals who have coronary angiography probably 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 large 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 death. 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 common in women who typically present with debilitating chest discomfort, is associated with normal appearing coronary arteries.\(^10\)–\(^10\) 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 a truly normal or “pristine” appearance in all three major epicardial and branch vessels. But the appearance of a truly “normal” angiogram is not particularly common, as compared with 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 catheterisation laboratory for evaluation of valvular heart disease, undergoing the preoperative standard 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. Box 1 shows the potential criteria for categorisation of controls.

### ERA OF CARDIAC COMPUTED TOMOGRAPHY

A non-invasive way to anatomically define coronary disease may be to use cardiac computed tomography. Both electron-beam computed tomography and multidetector 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 seem to be independent predictors of cardiac events in large population studies.13

However, the absence of coronary calcification does not completely rule out the finding of CAD. Acute coronary events may occur in younger subjects who do not show 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 CAD in young or middle aged individuals.

More recently, MDCT in particular, has been shown to provide high-resolution 3-dimensional images of the coronary vessels, allowing visualisation of both 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 intravascular ultrasound,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 values than fibrocalcific or calcified plaques.37

As Box 1 lists, some of the criteria for cases and controls matching cases and controls. There exists the distinct possibility that controls may, later in life, become cases. Accordingly, controls that are at least 10–20 years older than cases may represent a useful assessment to pre-empt this concern. Although 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 <60 years, it would be difficult to project that CAD or myocardial infarction would not occur later in life. Similarly, in patients who are older, >70 years, 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 be best avoided for use as controls. Family studies of this disease are further complicated by the common phenomena of “phenocopies” because CAD is quite prevalent and may occur independently from genetic susceptibility.

As Box 1 lists, some of the criteria for cases and controls would make patient accrual more challenging. Quality assurance of categorisation 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-like 2 for type II diabetes mellitus, and an as yet unnamed microsatellite variant for prostate cancer.38–40

One of the advances for these other disease conditions has been a well-accepted phenotype with clearcut, standard definition. Careful consideration given to phenotypic definitions will hopefully catapult the research efforts in identifying CAD susceptibility genetic variants in the future.
REFERENCES


Challenges in the phenotypic characterisation 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 2007 44: 161-165 originally published online December 8, 2006
doi: 10.1136/jmg.2006.045732

Updated information and services can be found at:
http://jmg.bmj.com/content/44/3/161

These include:

References
This article cites 30 articles, 9 of which you can access for free at:
http://jmg.bmj.com/content/44/3/161#BIBL

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 online article.

Topic Collections
Articles on similar topics can be found in the following collections
Editor's choice (122)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/