A Gene-Centric Approach to Elucidating Cardiovascular Risk

Ruth McPherson, MD, PhD, FRCPC

Although behavioral precedents including diet and cigarette smoking play an important role, coronary artery disease (CAD) exhibits significant heritability, a part of which can be attributable to genetic variants affecting known biochemical and metabolic risk factors. In the last 2 years, impressive progress has been made in understanding the genetic basis of several of these using both candidate gene and genome-wide association study (GWAS) approaches.

Candidate genes, coding for proteins of known biological significance in a disease process, provide a logical first step in understanding the genetics of common disease states. Populations of affected and unaffected individuals can be studied by genotyping common single-nucleotide polymorphisms (SNPs) within a gene and its regulatory sequences. Although economically attractive, this approach is acknowledged to have a number of limitations. By definition, studies are limited to genes with a known or suspected role in defining a given phenotype and do not provide new insight into biological pathways leading to disease. Furthermore, candidate gene associations often fail to replicate for multiple reasons. Sample sizes have often been inadequate to provide the statistical power required when multiple variants of small effect are tested. The original observation may have been false positive, emphasizing the need for stringent statistical thresholds. Other issues include heterogeneity of causality. For example, genetic variants affecting plasma lipid traits may have significant effects on CAD risk, but these may be less important than or interact with other risk factors including diabetes and smoking. Finally, appropriate measures must be taken to test for population stratification.

Genome-wide association studies (GWASs) are an alternative approach, facilitated by the International HapMap Project and advances in genotyping technology, permitting analysis of a million or more markers across the human genome. GWAS provides an unbiased approach that has led to the identification of novel loci associated with several complex phenotypes including CAD. Notably, many of these were unanticipated based on prior knowledge. GWASs have identified common variants linked to type 2 diabetes in genes affecting islet cell function, putting a renewed emphasis on β cell failure rather than adipocyte hypertrophy in the etiology of this disease. A number of known loci linked to lipoprotein metabolism have been identified by GWASs, supporting the overall validity of this approach. In addition, multiple novel genes or loci have been linked to plasma lipoprotein fractions (Table). One of the most robust associations with low-density lipoprotein cholesterol (LDL-C) and also linked to CAD risk is at the CELSR2-PSRC1-MYBPHL-SORT1 locus on 1p13. Although SORT1 mediates the endocytosis and degradation of lipoprotein lipase, an effect on LDL-C is unanticipated and CELSR2 and PSRC1 have no known effect on lipoprotein metabolism. A common nonsynonymous SNP in MLXIPL, a transcription factor involved in glucose utilization and energy metabolism, has shown a strong and consistent association with triglycerides providing new insights into its role in hepatic triglyceride synthesis. The fatty acid desaturase genes encoding for proteins that function in arachidonate metabolism, FADS1, FADS2, and FADS3 on chromosomes 11q12, have been linked to triglycerides, LDL-C, and high-density lipoprotein cholesterol (HDL-C). GALNT2 at 1q42 coding for a receptor modifying glycosyltransferase has been consistently linked to HDL-C, pointing to a previously unknown pathway in HDL metabolism. Functional studies of these regions may provide new insights into lipid metabolism and reveal potential drug targets for CAD risk reduction.

Beyond the identification of several novel loci, a few key messages have arisen from GWAS for plasma lipoprotein traits. Genetic variants predictive of LDL-C concentrations in genes of known function in apo B lipoprotein metabolism including APOB, LDLR, PCSK9, and HMGR are also predictive of CAD risk to an extent greater than would be anticipated by the apparent effect on LDL-C concentrations supporting the premise that genetic variants affecting LDL-C are a better marker of lifetime exposure to LDL-C as compared with a single measurement later in life. Despite the fact that HDL-C is a more powerful predictor of CAD risk than LDL-C, not all common variants associated with HDL-C or Apo AI have shown a consistent association with CAD. Multiple explanations are possible. The functional relationship of HDL to CAD risk is inherently complex and plasma concentrations of HDL-C are not always a reliable marker of reverse cholesterol transport. Genetic variation at the CETP locus has important reciprocal effects of plasma concentrations of CETP and HDL-C/Apo AI. Although CETP facilitates the generation of atherogenic lipoproteins, it also functions directly in reverse cholesterol transport and available...
data suggest that either high or low levels of CETP may increase atherosclerosis susceptibility. In contrast to CETP, SNPs in hepatic lipase (LIPC) and lipoprotein lipase have been associated with both HDL-C and CAD risk.

It must be acknowledged that despite the recent successes of the GWAS approach, the ability to identify the genetic basis of complex disease remains modest. Despite the fact that heritability (h²) estimates for plasma lipoprotein fractions including triglycerides, LDL-C, and HDL-C ranges from 0.5 to 0.6,14 in large study populations, common variants in both known and novel gene regions have accounted for only 3% to 10% of variability in plasma lipoprotein fractions. Current GWAS technology is limited by reliance on HapMap, based on genetic data from a small number of individuals and providing proxies for the majority of SNPs with a minor allele frequency (MAF) <5%. The Encyclopedia of DNA Elements project indicated that 60% of SNPs had an MAF <0.05. As noted by Gorlov et al,15 there is an inverse relationship between MAF and proportion of nsSNPs predicted to be damaging. Thus, multiple lower frequency variants are likely to have a cumulatively more important phenotypic effect. Indeed, current GWAS platforms fail to capture most of the rare variants present on the nsSNP platform.16 Multiple lower frequency common variants (MAF between 0.5% and 5%) have been shown to contribute importantly to complex phenotypes relevant to CAD. Resequencing of the PCSK9 gene identified low-frequency common variants (MAF, 0.5–1.0) affecting LDL-C concentrations and CAD risk. Even rarer variants with an MAF <0.5% are important determinants of extreme phenotypes. Approximately 15% of subjects with HDL-C below the 10% for age and sex harbor multiple rare mutations in ABCA1, APOA1, or PLTP.18 In addition to lack of rare variant coverage, current GWAS platforms still have significant gaps in coverage. For example, the APOE SNPs, rs7412 and rs429358, defining the common 2, 3, and 4 isoforms are not genotyped on the Affymetrix or Illumina arrays, and the best proxy, rs4420638, lies 3 of APOC1 and is in only partial linkage disequilibrium with the Apo E e4 SNP, rs429358 (r²=0.72).

Thus, considerable interest remains in the use of a candidate gene approach for the evaluation of cardiovascular risk. Keating et al19 have recently described the design and implementation of a cardiovascular gene-centric 50 k SNP array for large-scale genomic association studies. Led by investigators at the Institute of Translational Medicine and Therapeutics (ITMAT), the Broad Institute and the Candidate-gene Association Resource (CARe) supported by the National Heart, Lung and Blood Institute, the ITMAT-Broad-CARe (IBC) array provides an important complement to current GWAS products. In addition to 1349 loci of potential involvement in phenotypes of interest, the array includes dense coverage of 435 genes and regions of high biological plausibility in relationship to cardiovascular dis-

---

**Table. Recent Candidate Gene and GWA Studies of Plasma Lipoprotein Traits**

<table>
<thead>
<tr>
<th>Candidate Gene Studies</th>
<th>Sample</th>
<th>SNPs</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kathiresan et al.5</td>
<td>5 414</td>
<td>11 SNPs in 9 genes</td>
<td>4.1% LDL, 4.4% HDL</td>
</tr>
<tr>
<td>Anand et al.12</td>
<td>8 034</td>
<td>940 SNPs in 69 genes</td>
<td>4.3% Apo B/A1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genome Wide Association Studies</th>
<th>Initial Sample</th>
<th>GWAS Platform(s)</th>
<th>Replication Sample</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willer et al.11</td>
<td>8 816</td>
<td>Illumina HumanHap300, Affymetrix 500K</td>
<td>11 569 (21–196 SNPs)</td>
<td>5%–8% TG, LDL-C, HDL-C</td>
</tr>
<tr>
<td>Kathiresan et al.6</td>
<td>8 816</td>
<td>Illumina HumanHap300, Affymetrix 500K</td>
<td>18 554 (196 SNPs)</td>
<td>4.5% TG, 5.7% LDL-C, 5.2% HDL-C</td>
</tr>
<tr>
<td>Aulchenko et al.4</td>
<td>22 562</td>
<td>Illumina HumanHap300-Duo</td>
<td>4.8% HDL-C, 3.4% LDL-C, 3.0% HDL</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GWAS Supplemented by Custom Content Related to Candidate Loci</th>
<th>Initial Sample</th>
<th>GWAS Platform(s)</th>
<th>Custom Content</th>
<th>Replication Sample</th>
<th>Variance Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chasman et al.13</td>
<td>6 382</td>
<td>Illumina HumanHap 300 Duo</td>
<td>33 923 SNPs iSelect</td>
<td>970 (6 SNPs)</td>
<td>4.3% TG, 7.3% LDL-C, 4.3% HDL-C, 6.4% Apo A1, 12.6% Apo B</td>
</tr>
<tr>
<td>Kathiresan et al.7</td>
<td>19 840</td>
<td>Illumina 317K/550K Affymetrix 5.0</td>
<td>50 000 SNPs IBC17</td>
<td>20 623 (52–66 SNPs)</td>
<td>7.4% TG, 7.7% LDL-C, 9.3% HDL-C</td>
</tr>
</tbody>
</table>

---

*4 Circ Cardiovasc Genet February 2009*
ease or metabolic phenotypes. Tag SNPs were selected to capture variation with MAF >0.02 and r² >0.8 in HapMap populations and Seattle SNPs, and all rsSNPs were included. Notably, more than 20% of the IBC group I SNPs have not been assayed in HapMap and were derived from Seattle SNPs or published reports.

In this issue of Circulation: Cardiovascular Genetics, on a smaller scale but similarly based on previously published data on genetic variants associated with CAD risk factors or myocardial infarction (MI) or biological pathways associated with these risk factors, the INTERHEART investigators have genotyped 1326 SNPs from 103 genes in 8034 MI cases and controls from 5 ethnic groups, including 2306 South Asian subjects, an understudied population at apparently increased risk for CAD. In accordance with other recent attempts to replicate candidate gene findings in CAD, none of these were associated with MI after adjustment for multiple testing. Of 940 SNPs in 69 genes tested for association with MI risk factors, 13 showed significant associations. Each of these has been previously well studied, including SNPs in APOE, LDLR, and APOB associated with Apo B concentrations and 3 tightly linked SNPs in CETP associated with Apo AI levels. By reanalyzing the association of these 13 SNPs with MI risk, the investigators corroborate previous reports on the relationship of SNPs defining the Apo E e2 (rs7412) and e4 (rs429358) isoforms to MI risk and the association of LDLR SNPs (rs6511720 and rs1433099) to both Apo B concentrations and MI risk. In accordance with other recent data, CETP SNPs associated with Apo AI concentrations were not associated with MI risk. Overall, the 13 significant SNPs accounted for less than 5% of variability in apolipoprotein concentrations. Because plasma lipids were not measured and samples were nonfasting, it was not possible to study the relationship of candidate gene SNPs to plasma triglycerides or other lipid traits.

Importantly, more attention must be given to phenotype ascertainment in both GWAS and candidate gene studies. Plasma lipoprotein levels vary over the course of a lifetime, and a single measurement is unlikely to reflect the lifetime exposure to a given lipid risk factor. Many CAD cases have been on long-term statin therapy obscuring the effect of specific polymorphisms on LDL-C. Accurate and precise phenotyping is similarly important for CAD. In many GWAS, CAD cases are compared with age- and sex-matched controls but given the prevalence of subclinical atherosclerosis in middle aged individuals, many 50-year-old controls may harbor a similar disease burden to cases with incident MI. Angiographic ascertainment of disease burden is an optimal solution but is not generally feasible in apparently healthy controls. An alternative and attractive approach is to compare young subjects with extensive multivessel CAD with healthy elderly controls although confounding is possible due to other genetic variants associated with longevity.

Overall, as indicated by Anand et al., a candidate gene approach is valuable and complementary to the discovery nature of GWAS. In 2008, several genome-wide and candidate gene studies provided new information on the genetic basis of variability in plasma lipoprotein concentrations and relationship to cardiovascular disease. It is notable that studies that have supplemented standard arrays with custom content have had the greatest success in explaining variation in individual lipoprotein traits (Table). Further success in deciphering the genetic components of lipoprotein metabolism and CAD risk will require access to large comprehensive primary data sets to dissect potentially important gene–gene and gene–environment interactions and dense resequencing of significant loci that are likely to harbor multiple rare variants of cumulative importance.

Sources of Funding

The Canadian Institutes of Health Research and the Heart and Stroke Foundation of Ontario supported this study.

Disclosures

None.

References


A Gene-Centric Approach to Elucidating Cardiovascular Risk
Ruth McPherson

Circ Cardiovasc Genet. 2009;2:3-6
doi: 10.1161/CIRCGENETICS.109.848986
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 American Heart Association, Inc. All rights reserved.
Print ISSN: 1942-325X. Online ISSN: 1942-3268

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circgenetics.ahajournals.org/content/2/1/3

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Genetics 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: Cardiovascular Genetics is online at:
http://circgenetics.ahajournals.org//subscriptions/