A Genome-Wide Association Study in Europeans and South Asians Identifies 5 New Loci for Coronary Artery Disease

Nehal N. Mehta, MD, MSCE

Study Hypothesis
Recently, genome-wide association studies (GWAS) have identified several common variants associated with increased risk of coronary artery disease (CAD) and myocardial infarction (MI) by 10% to 30%. The authors state that these loci explain only a small proportion of the predicted genetic risk and that all of the current loci for CAD and MI by GWAS have been discovered in European populations. The authors hypothesized that discovery of new susceptibility loci of smaller effect sizes would be aided by conducting much larger studies in addition to an emphasis on early-onset CAD and clearly defined clinical end points. Therefore, they assembled the Coronary Artery Disease (C4D) Genetics Consortium.

How Was the Hypothesis Tested?
The authors performed a meta-analysis of 4 large GWAS of CAD, 2 of European ancestry (PROCARDIS and HPS) and 2 of South Asian ancestry (PROMIS and LOLLIPOP), with ~575,000 genotyped single-nucleotide polymorphisms (SNPs) in a discovery data set comprising 15,420 individuals with CAD (cases) (8,424 Europeans and 6,996 South Asians) and 15,062 control subjects. They observed little evidence for ancestry-specific associations, supporting the use of combined analyses. Furthermore, the authors state that because all individuals were genotyped on the same platform (whole-genome Illumina BeadChips), a meta-analysis of actual genotypes rather than imputed data enabled analysis of low-frequency variants (1% to 5%), which have been excluded from GWAS either because of sample size or because imputation has been required to combine data from different genotyping platforms. After the meta-analysis, they selected 59 SNPs from 50 loci that showed potential new associations from the meta-analysis of the European and South Asian studies (41 SNPs; P < 1.0 × 10⁻⁵), the European-only meta-analysis (8 SNPs; P < 3.0 × 10⁻⁸), the South Asian-only meta-analysis (6 SNPs; P < 3.0 × 10⁻⁸), and 3 loci with strong biological plausibility but only suggestive probability values (4 SNPs). These SNPs were tested in 10 replication studies involving a total of 21,408 CAD cases and 19,185 control subjects, largely by de novo genotyping. Finally, to understand potential mechanisms and intermediate pathways by which novel loci may mediate risk, they investigated associations between the expression levels of all genes within 200 kb of each of the confirmed risk SNPs in tissue samples of aortic media and adventitia, mammary artery, carotid plaque, liver, adipose tissue, transformed lymphoblastoid cell lines, and skin and analyzed whether the novel loci were associated with established CAD risk factors.

Principal Findings
They confirmed the power and representative nature of the discovery-stage studies with data supporting the relevance of 11 known CAD susceptibility loci (9p21, CELSR2-PSRC1-SORT1, PHACTR1, WDR12, SLC5A3-MRPS6-KCNE2, MRAS, LDLR, CXCL12, MIA3, SH2B3, PCSK9) with similar previously reported effect sizes and directionally consistent effects in the European and South Asian populations for all 11 loci. Of the 59 SNPs carried forward from the discovery stage into the replication stage, 5 SNPs (rs1412444 LIPA, rs974819 PDGF, rs4380028 ADAMTS7-MORF4L1, rs10953541 7q22, rs2505083 KIAA1462) in the newly associated loci achieved the prespecified threshold for replication (P < 8.5 × 10⁻⁷, which is P < 0.05 after Bonferroni correction for 59 independent tests), and each also achieved conventional genome-wide significance (P < 5.0 × 10⁻⁸), with probability values ranging from 2.8 × 10⁻¹³ to 3.9 × 10⁻⁸ for the combined discovery and replication meta-analysis. They noted heterogeneity between the European and South Asian effect for rs4380028 in the ADAMTS7-MORF4L1 locus in the discovery meta-analysis, which was not supported by the independent replication, and observed no evidence of ancestry-specific heterogeneity for any of the other previ-
ously unidentified loci in either the discovery or replication meta-analyses. In addition to the 5 newly associated loci, rs9349379, located in an intron of PHACTR1, was significantly associated in the replication alone ($P=9.9 \times 10^{-16}$) and in the combined discovery and replication ($P=8.7 \times 10^{-26}$) meta-analyses, and rs17114046, in an intron of PPAP2B, showed consistent support in the discovery and replication meta-analyses but did not meet the predetermined significance level in both the replication alone ($P=1.1 \times 10^{-5}$) and in the combined discovery and replication ($P=2.5 \times 10^{-5}$) meta-analyses.

To better understand the biology of these novel loci, the authors examined whether novel risk alleles were associated with traditional CAD risk factors and found that none of these loci had any previously reported associations with established CAD risk factors (lipids, blood pressure, glucometabolic traits, or body mass index). Interestingly, rs4380028, in the ADAMTS7-MORF4L1 locus, is ∼200 kb downstream of a robust QTL for cigarettes smoked per day, but a conditional analysis for cigarettes smoked per day showed no attenuation of the CAD risk ($P=0.38$).

Finally, the authors investigated associations between the expression levels of all genes within 200 kb of each of the confirmed risk SNPs in tissue samples of aortic media and adventitia, mammary artery, carotid plaque, liver, adipose tissue, transformed lymphoblastoid cell lines, and skin. Combining regional association plots of each locus and their own or previously published eQTL data, they found that specific genes can be implicated at 2 of the new loci (rs1412444 and rs974819). rs1412444 is in an intron of the LIPA, the lysosomal acid lipase gene, and the risk allele of this SNP has been strongly linked with increased expression level of LIPA mRNA in circulating monocytes in the literature. This was the lead eQTL SNP for LIPA in their data and had suggestive association with increased expression in liver ($P<1.6 \times 10^{-3}$), and loss of function in LIPA has been associated with hypercholesterolemia despite the locus having weak association with LDL levels, implicating alternate pathways of this gene’s effects. Furthermore, rs974819 is near PDGFD, which is 117 kb downstream in an adjacent block of linkage disequilibrium. They found a significant PDGFD eQTL for rs974819 in aortic media ($P<2.3 \times 10^{-5}$), with suggestive associations in aortic adventitia ($P<7.7 \times 10^{-5}$) and mammary artery ($P<7.2 \times 10^{-5}$), and, in all 3 tissues, the risk allele was associated with increased expression. Platelet-derived growth factor D, encoded by PDGFD, is expressed in several cell types in atherosclerotic plaques and is predicted to stimulate atherosclerosis by influencing matrix metalloproteinase activity and monocyte migration in prior work from the literature. The remaining 3 loci (rs4380028, rs10953541, and rs2505083) could not be attributed to a specific gene region by eQTL analysis but were in proximity to genes noted above.

The authors state that this large analysis yielded a substantial increase in the number of confirmed susceptibility loci for CAD without finding any susceptibility variants with material differences in effect size or allele frequency between South Asians and Europeans, noting that current GWAS arrays may not capture all important variants in South Asians. Despite this, the study also demonstrates the importance of the genes associated with CAD beyond the European ancestry groups in which they were first defined.

**Implications**

This study represents how collaborations to increase sample sizes can potentially reveal common susceptibility variants with small effect sizes in complex traits. Additionally, as the authors note, combining directly genotyped data rather than imputed data provides the advantage of enabling analysis of low frequency variants (1% to 5%) excluded from prior GWAS either because of sample size or because imputation was required to combine data from different genotyping platforms. Furthermore, this study demonstrates the importance of large-scale GWAS carried out in populations of different ancestry can be informatively combined in genetic discovery, demonstrating how larger efforts in multi-ethnic groups could identify additional variants that influence CAD risk. There is growing evidence that many common variant signals detected by GWAS are driven by causal alleles shared between populations that presumably segregated in African populations before the major “modern human” diaspora. Hence, investigation of populations with diverse genetic ancestries with transethnic differences in linkage disequilibrium patterns and haplotypic structures potentially enables discovery of novel signals. In fact, this was elegantly demonstrated by Musunuru et al in dissecting the SORT1 locus in dyslipidemia, where comparison of the haplotype structures in the locus in European Caucasians and African Americans found a single SNP that was common to the cholesterol-associated haplotypes in the 2 ethnic groups, with functional studies confirming that the one SNP was the causal variant. Future studies combining large multi-ethnic populations from GWAS of CAD and MI are ongoing and may be as fruitful as demonstrated by these 2 studies.

**Acknowledgments**

The author is a member of the Early Career Committee of the American Heart Association Functional Genomics and Translational Biology Council.

**Disclosures**

None.
A Genome-Wide Association Study in Europeans and South Asians Identifies 5 New Loci for Coronary Artery Disease
Nehal N. Mehta

doi: 10.1161/CIRCGENETICS.111.960989
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2011 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/4/4/465

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/