Published in final edited form as:
Circ Cardiovasc Genet. 2011 August 1; 4(4): 465–466. doi:10.1161/CIRCGENETICS.111.960989.

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

Nehal N. Mehta, MD, MSCE
Cardiovascular Institute, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA

Keywords

genomics; genetics; coronary artery disease; genome wide association; meta-analysis

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 10–30%. The authors state 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 endpoints. Therefore, they assembled the Coronary Artery Disease (C4D) Genetics Consortium.

How Was the Hypothesis Tested?

The authors performed a meta-analysis of four large genome-wide association studies of CAD, two of European ancestry (PROCARDIS and HPS) and two of South Asian ancestry (PROMIS and LOLLIPOP) with ~575,000 genotyped SNPs in a discovery dataset comprising 15,420 individuals with CAD (cases) (8,424 Europeans and 6,996 South Asians) and 15,062 controls. 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 (1–5%), which have been excluded from GWAS either due to sample size or because imputation has been required to combine data from different genotyping platforms. Following 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; $P < 1.0 \times 10^{-4}$), the European only meta-analysis (8 SNPs; $P < 3.0 \times 10^{-5}$), the South Asian only meta-analysis (6 SNPS; $P < 3.0 \times 10^{-5}$) and three loci with strong biological plausibility but only suggestive *P* values (4 SNPs). These SNPs were tested in ten replication studies involving a total of 21,408 CAD cases and 19,185 controls largely by *de novo* genotyping. Finally, to understand potential mechanisms and intermediate pathways by

Correspondence to: Nehal N. Mehta, MD, MSCE, 6 Penn Tower, One Convention Center Blvd., Philadelphia, PA 19104, Phone: 215.662.7988, Fax: 215.349.5927, nehal.mehta@uphs.upenn.edu.
From the Early Career Committee of the American Heart Association Functional Genomics and Translational Biology Council

Disclosures
None.

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, as well as analyzed whether an of 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, five SNPs (rs1412444 *LIPA*, rs974819 *PDGFD*, rs4380028 *ADAMTS7-MORF4L1*, rs10953541 7q22, rs2505083 *KIAA1462*) in the newly associated loci achieved the pre-specified threshold for replication ($P < 8.5 \times 10^{-4}$; which is $P < 0.05$ after Bonferroni correction for 59 independent tests), and each also achieved conventional genome-wide significance ($P < 5.0 \times 10^{-8}$), with P 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 previously unidentified loci in either the discovery or replication meta-analyses. In addition to the five newly associated loci, rs9349379, located in an intron of *PHACTR1*, was significantly associated in the replication alone ($P = 9.9 \times 10^{-10}$) 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 pre-determined significance level in both the replication alone ($P = 1.1 \times 10^{-3}$) and in the combined discovery and replication ($P = 2.5 \times 10^{-7}$) 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 they performed 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 specific genes can be implicated at two 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 monocytesin 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 (LD). They found a significant *PDGFD* eQTL for rs974819 in aortic media ($P < 2.3 \times 10^{-7}$), with suggestive

associations in aortic adventitia ($P < 7.7 \times 10^{-4}$) and mammary artery ($P < 7.2 \times 10^{-4}$), and in all three 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 three 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 (1–5%) excluded from prior GWAS either due to 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 which presumably segregated in African populations prior to the major “modern human” diaspora. Hence investigation of populations with diverse genetic ancestries with trans-ethnic differences in LD patterns and haplotypic structures potentially enables discovery of novel signals. In fact, this was elegantly demonstrated by Musunuru et al.[2] 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 two 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 will hopefully be as fruitful as demonstrated by these two studies.

Acknowledgments

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

References

1. Coronary Artery Disease (C4D) Genetics Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. Nat Genet. 2011; 43:339–344. [PubMed: 21378988]

2. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs KV, Li X, Li H, Kuperwasser N, Ruda VM, Pirruccello JP, Muchmore B, Prokunina-Olsson L, Hall JL, Schadt EE, Morales CR, Lund-Katz S, Phillips MC, Wong J, Cantley W, Racie T, Ejebe KG, Orho-Melander M, Melander O, Kotelianski V, Fitzgerald K, Krauss RM, Cowan CA, Kathiresan S, Rader DJ. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. Nature. 2010; 466:714–719. [PubMed: 20686566]