Editorial

Peering Into the Future of CAD Genomics

Daniel I. Chasman

About 10 years ago, the promise of genome-wide association studies (GWASs) for reducing the burden of coronary artery disease (CAD) was in crisis. GWASs had not enabled the genetic prediction of CAD in populations. Nor had they revealed novel pathobiology. Even GWASs including up to 2000 cases of CAD, an ambitious proposition at the time, found only a single common association at chromosomal position 9p21 that surpassed the stringent significance threshold required for common variants on a genome-wide basis (ie, $P<5\times10^{-8}$).[1–4] Although this signal was robust, it only minimally accounted for the family-based heritability and thus risk of CAD,[5] and its biological mechanism of action could not be discerned because it mapped to an extended intergenic region. Subsequent GWAS over the following few years incrementally identified more CAD loci but still failed to account appreciably for CAD heritability.

Article, see p 433

What was wrong? Did family-based heritability overestimate the role of genetics in populations? Did the heritability reside in rare instead of common genetic variants?[6] Was there a need to consider gene-by-environment interactions? Certainly, all of these explanations contribute to a complete picture of CAD genetics. However, since the initial CAD GWASs, it has also been recognized that the heritability of CAD, like other complex traits, is substantially composed of many more variants than had been anticipated, each with a small effect. Such a genetic architecture, consisting of hundreds or even thousands of small genetic influences, can only be dissected with large samples; and the carefully nurtured cohort studies that have been highly effective for conventional CAD epidemiology are typically not large enough. In other words, as resource intensive as the initial CAD GWASs were, their study designs limited statistical power to detect the majority of common genetic variants contributing to CAD.

A sea change in thinking about how to address this challenge is occurring. Led by the example of the UK Biobank, we may soon have access to extremely large samples for genetic analysis that are sufficiently powered to elucidate the genetic architecture of CAD and other complex traits.[7] The UK Biobank leveraged the UK National Health Service to enroll a population-based sample of ≈500 000 consented Britons designed for comprehensive analysis health determinants on a formerly unprecedented scale. It includes whole-genome genotype data in the entire population. Participants were aged 40 to 69 years at enrollment, and although initial analysis will be largely retrospective, the proposed 30 years of follow-up enable prospective studies. Remarkably, the charter of the UK Biobank explicitly encourages public access to these data for approved research projects with a modest administrative fee.

An interim release of data from ≈150 000 UK Biobank participants had already shown promise for analysis of CAD.[8–10] Several papers during the past year considered CAD cases in the UK Biobank defined by *International Classification of Diseases*, Tenth Revision codes, collected by self-report touchscreen questionnaire and a verbal interview in person at assessment centers or from linked medical records. This ascertainment approach yielded ≈5000 to 6000 strictly defined or ≈10 000 more liberally defined cases that included, for example, angina. Combining GWAS of these cases (and suitable controls) with pre-existing meta-analysis of genome-wide scans from a variety of case/control and cohort studies within the CARDIoGRAMplusC4D consortium (which already included 60 801 CAD cases),[6] 3 independent efforts each added ≈13 to 15 novel CAD loci for a grand total of 96 (including loci identified via other strategies[11–13]).

In the current work by van der Harst and Verweij,[14] the authors now avail themselves of the full UK Biobank with up to 34 541 cases of CAD and 261 984 controls and present their new GWAS only ≈6 months after these data were released. In their analytic strategy, the authors first select suggestively significant associations in the UK Biobank for replication in the pre-existing, cohort-based genome-wide meta-analysis from CARDIoGRAMplusC4D, yielding 13 novel loci that satisfy conventional significance standards for replication and also reach genome-wide significance in meta-analysis of all studies. They then perform the reciprocal analysis, identifying suggestively significant candidates in the CARDIoGRAMplusC4D meta-analysis for replication in the UK Biobank and find 21 more loci. However, although reciprocal replication provides a convincing association, it is less statistically powerful than direct meta-analysis of all data. Taking the latter approach, the authors identify 30 additional associations, resulting in a total of 64 novel loci or 161 overall. Considered together, the 161 loci are estimated to explain 15.1% of the heritability of population-based CAD liability attributable to the common genetic variation sampled by the study.

As is customary in GWASs today, the authors further characterize for their novel loci by a series of additional analyses. Using the *International Classification of Diseases*, Tenth Revision–coded UK Biobank event definitions, they

From the Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA.

Correspondence to Daniel I. Chasman, PhD, Division of Preventive Medicine Brigham and Women's Hospital, Boston, MA 02215. E-mail dchasman@research.bwh.harvard.edu

(*Circ Res*. 2018;122:391-393.
DOI: 10.1161/CIRCRESAHA.117.312502.)

***Circulation Research* is available at http://circres.ahajournals.org**
DOI: 10.1161/CIRCRESAHA.117.312502

confirm a previous report showing that 2 loci are selectively associated the acute phase of CAD, namely myocardial infarction, rather than the progressive phase defined by CAD without myocardial infarction. However, selectivity for myocardial infarction was not revealed for any of the novel loci. Many of the novel loci have been previously implicated in known CAD-related phenotypes, such as plasma lipids, inflammatory biomarkers, blood pressure, diabetes mellitus, glomerular filtration rate, and abdominal aortic aneurysm. Still others are related to developmental phenotypes, such as anthropometric measures and age of menarche. At a false discovery rate of 0.05, candidate genes mapping to the novel loci are predominantly active in tissues associated with blood vessels, but there are also expression signals in tissues associated with musculoskeletal system, the digestive system, connective tissue (eg, adipose tissue), and the gastrointestinal tract, including liver. The candidate genes are enriched in 1525 gene sets (ie, genes participating in a shared biological function), derived from semisupervised integration of known biology with tissue-specific patterns of correlated gene expression and described with terms such as complete embryonic lethality during organogenesis, blood vessel development, and anemia among others.

Perhaps, more surprising than the identities of these new biological pathways is their number. Gene-set enrichment analysis on the previous CAD GWAS from CARDIoGRAMplusC4D[6] yielded 457 significant gene sets. This number increased to 889 in analysis with the interim release of 150 000 UK Biobank samples, but leapt to 1525 with the current GWAS representing 13.9% of the 10 968 gene sets that were tested. Because enrichment in gene sets essentially reflects a correlation between the genetic association signal for CAD and coordinated biological functions, the marked increase in signal validates the biological importance of the new findings. Moreover, the increase is consistent with the emerging notion that a great multiplicity of genetic loci, often with apparently disparate biological functions, all conspire to influence the heritability of complex traits, such as CAD.

In spite of these advances, the contribution of the genome-wide significant loci to CAD risk remains modest. Although a genetic risk score aggregating the effects of all loci was statistically significant, its effect would likely not be useful in a clinical setting for predicting incident CAD. Extrapolating from the incremental rate of CAD locus discovery as a function of sample size, it can be estimated that a GWAS incorporating ≈1 million cases may be needed to yield a clinically useful CAD genetic predictor[15] for the general population. Moreover, it is anticipated that genetic predictors will be different for different ancestries. Predictors derived from most ongoing genetic analyses of CAD emphasize European ancestry and would likely not be optimal in populations with non-European, especially African, ancestry, which would require additional GWASs at full scale.

However, the UK Biobank is a great start toward the needed sample sizes, and it will soon be joined by comparable initiatives in the United States and elsewhere. For example, the Million Veterans Project is enrolling veterans on a voluntary basis to collect comprehensive genomic information linked to longitudinal clinical data through the Veteran's Administration healthcare system while the NIH-funded All of Us project is enrolling 1 million individuals to integrate conventional clinical profiles with genomic data toward precision medicine. As these projects mature, the feasibility of involving still larger populations for research may become apparent, for example, through anonymized linking of whole-genome genetics to electronic medical records.

Our understanding of CAD genomics has progressed rapidly in the past 10 years, measured not only by the number of known loci but also by elucidation of CAD genetic architecture and candidate genes for functional follow-up. Peering into the future, the prospects for realizing the full potential of genomics in understanding prevention and treatment of CAD are excellent. We may be in the midst of that future now.

Sources of Funding

D.I. Chasman is supported by the National Institute of Heart, Lung, and Blood Research (R01HL118305), the National Institute of Neurological Disorders and Stroke (R21NS092963, R21NS104398), and the National Eye Institute (R01EY021900).

Disclosures

None.

References

1. McPherson R, Pertsemlidis A, Kavaslar N, Stewart A, Roberts R, Cox DR, Hinds DA, Pennacchio LA, Tybjaerg-Hansen A, Folsom AR, Boerwinkle E, Hobbs HH, Cohen JC. A common allele on chromosome 9 associated with coronary heart disease. *Science*. 2007;316:1488–1491. doi: 10.1126/science.1142447.
2. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316:1491–1493. doi: 10.1126/science.1142842.
3. Samani NJ, Erdmann J, Hall AS, et al; WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *N Engl J Med*. 2007;357:443–453. doi: 10.1056/NEJMoa072366.
4. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661–678. doi: 10.1038/nature05911.
5. Paynter NP, Chasman DI, Buring JE, Shiffman D, Cook NR, Ridker PM. Cardiovascular disease risk prediction with and without knowledge of genetic variation at chromosome 9p21.3. *Ann Intern Med*. 2009;150:65–72.
6. Nikpay M, Goel A, Won HH, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47:1121–1130. doi: 10.1038/ng.3396.
7. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779. doi: 10.1371/journal.pmed.1001779.
8. Verweij N, Eppinga RN, Hagemeijer Y, van der Harst P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Sci Rep*. 2017;7:2761. doi: 10.1038/s41598-017-03062-8.
9. Klarin D, Zhu QM, Emdin CA, et al; CARDIoGRAMplusC4D Consortium. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat Genet*. 2017;49:1392–1397. doi: 10.1038/ng.3914.
10. Nelson CP, Goel A, Butterworth AS, et al; EPIC-CVD Consortium; CARDIoGRAMplusC4D; UK Biobank CardioMetabolic Consortium CHD working group. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet*. 2017;49:1385–1391. doi: 10.1038/ng.3913.
11. Stitziel NO, Stirrups KE, Masca NG, et al; Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Coding variation in

ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N Engl J Med*. 2016;374:1134–1144. doi: 10.1056/NEJMoa1507652.
12. Webb TR, Erdmann J, Stirrups KE, et al; Wellcome Trust Case Control Consortium; MORGAM Investigators; Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. *J Am Coll Cardiol*. 2017;69:823–836. doi: 10.1016/j.jacc.2016.11.056.
13. Howson JMM, Zhao W, Barnes DR, et al; CARDIoGRAMplusC4D; EPIC-CVD. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. *Nat Genet*. 2017;49:1113–1119. doi: 10.1038/ng.3874.
14. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res*. 2018;122:433–443. doi: 10.1161/CIRCRESAHA.117.312086.
15. Chatterjee N, Wheeler B, Sampson J, Hartge P, Chanock SJ, Park JH. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat Genet*. 2013;45:400–5, 405e1. doi: 10.1038/ng.2579.

KEY WORDS: Editorials ■ cohort studies ■ coronary artery disease ■ genetic variation ■ genome-wide association study

Circulation Research
JOURNAL OF THE AMERICAN HEART ASSOCIATION

Peering Into the Future of CAD Genomics

Daniel I. Chasman

Circ Res. 2018;122:391-393
doi: 10.1161/CIRCRESAHA.117.312502
Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0009-7330. Online ISSN: 1524-4571