Exomes, Proteins, and Cardiovascular Disease
Making Sense of the Signals

Robert W. McGarrah, MD; Svati H. Shah, MD, MHS

The postgenomic era has simultaneously provided huge opportunities and major challenges in the translation of the wealth of genetic information to improve our diagnostic and therapeutic approach to disease. Although many might agree that genome-wide association studies have not met expectations in their ability to capture the full genetic architecture of many medical conditions, there is hope that more sophisticated and comprehensive genetic analyses, such as whole exome and whole genome sequencing, will shed more light on the mechanisms underlying common diseases. A complementary approach is in the integration of molecular profiling platforms (genomics, transcriptomics, proteomics, metabolomics, etc.) to provide a systems-biology perspective. This approach is particularly provocative in common complex diseases such as cardiovascular disease (CVD) where the end disease represents a convergence of diverse biological mechanisms, only some of which may be at play in a given individual, and where the disease process itself varies temporally with fixed genetic variation unable to capture this temporal progression. Such diseases are often defined by an underlying quantitative liability scale. As such, intermediate traits represented by proteins, RNA transcripts, metabolites, etc. may reside closer to the underlying pathogenic gene and thereby provide a stronger genetic signal, in addition to serving as stronger signals for the biological pathways mediating the disease process.

Indeed, many groups have used this functional genomics approach to provide new insights into CVD pathophysiology. Expression quantitative trait loci have been used for years for functional evaluation of genome-wide association studies hits, biomarker discovery, and tissue specificity evaluations in human diseases. More recently, several studies have mapped metabolic traits onto the genome to identify quantitative trait loci associated with CVD phenotypes. These studies have provided valuable understanding about the genetic architecture underlying metabolic variability and have highlighted new biological pathways involved in CVD.

In this issue of Circulation: Cardiovascular Genetics, Solomon et al apply both approaches in a genetic screen, incorporating exome sequencing and mapping of an intermediate quantitative trait, to simultaneously identify potential disease loci, molecular pathways of disease, and novel biological interactions. Specifically, the authors use the Tromsø Study—a prospective, longitudinal population study of men and women in Tromsø, Norway—to identify genetic variation underlying serum concentrations of proteins implicated in CVD (protein quantitative trait loci mapping, pQTL).

Using multiplex sandwich ELISA assays to quantitate 51 CVD-related proteins, the group previously identified 17 proteins that were predictors for incident myocardial infarction in their population. In this study, they examine associations between both rare and common exonic variants in the genes for these 51 proteins, as well as imputed common variants located within 500 kb of the gene, with serum levels of their protein products. They performed these analyses in baseline serum samples of 330 individuals who were part of a nested case–control study of venous thromboembolism in the Tromsø Study. None of the 51 proteins were associated with venous thromboembolism case–control status; therefore, all individuals were combined to determine the effects of genetic variation on baseline protein levels. The genetic data were obtained from whole exome sequencing (n=243) or exome genotyping arrays (n=87), resulting in a total of 158 137 genetic variants (24 915 from direct genotyping and 138 415 from imputation). For cis associations, they tested the variants located in the cis gene loci for associations with their respective protein level. For cis-acting-in-trans associations, they tested all significantly association common variants against each of the other 50 protein levels. For trans associations, they tested all variants in the cis loci for association with each of the 51 protein levels. In addition, they performed in silico analyses using existing genetic databases to validate their findings and to determine the potential clinical significance of their variants.

After correction for multiple comparisons, for cis associations, they identified 8 known (AGT, C3, C3b, CHI3L1, F12, LBP, Lp(a), and MMP3) and 6 novel (a2-AP, ANG, KLKB1, MMP8, Lp(a), and KNG1) common variant pQTLs, as well as 3 known (AGER, Fetuin A, and Lp(a)) and 4 novel (CD40L, MMP8, TAF1, and TIMP4) rare variant pQTLs. They also identified 2 cis-acting-in-trans common variants (F12 and KLKB1), each of which was significantly associated with 3 other proteins (F12: KLKB1, KNG1, N-terminal prohormone of brain natriuretic peptide; KLKB1: KNG1, N-terminal prohormone of brain natriuretic peptide, urokinase-type plasminogen activator

DOI: 10.1161/CIRCGENETICS.116.001549

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

From the Department of Medicine, Duke Molecular Physiology Institute, Durham, NC (R.W.M., S.H.S.); Division of Cardiology, Department of Medicine, Duke University Medical Center, Durham, NC (R.W.M., S.H.S.); and Department of Medicine, Duke Clinical Research Institute, Durham, NC (S.H.S.)

Correspondence to Svati H. Shah, MD, MHS, Duke Molecular Physiology Institute, 300 N. Duke St, Durham, NC 27701. E-mail svati.shah@duke.edu


© 2016 American Heart Association, Inc.

Circ Cardiovasc Genet is available at http://circgenetics.ahajournals.org

DOI: 10.1161/CIRCGENETICS.116.001549
receptor). After permutation adjusting, they did not find additional significant \textit{trans} associations between any of the common or rare variants in the gene regions and the 51 protein levels.

To assess potential functional effects and novelty of the pQTL findings, the authors then cross-referenced their 14 common pQTLs with publicly available expression quantitative trait loci databases. Using this approach, they confirmed that 8 of the pQTLs had been previously identified with an expression quantitative trait loci (either for the \textit{cis} gene or for the nearby genes). Of these 8, 2 were newly identified pQTLs (\textit{a2-AP} and \textit{KLKB1}). Thus, the majority of the pQTL common variants are also associated with gene expression levels, supporting their functional role in regulating protein levels as well. In examining their variants in genome-wide association studies databases, the 8 known common variants, as well as a novel common variant in the kallikrein (\textit{KLKB1}) gene, were associated with a variety of phenotypes; however, they were not associated with venous thromboembolism or myocardial infarction. Finally, to probe the functional implications of the previously unknown relationship between kallikrein and N-terminal prohormone of brain natriuretic peptide that was reflected by the statistical \textit{cis}-acting-in-\textit{trans} association of a \textit{KLKB1} variant with N-terminal prohormone of brain natriuretic peptide levels, they show that kallikrein can cleave proBNP in vitro in a dose-dependent manner.

Integrating large-scale genomic profiling, including whole exome sequencing, with immunoassay-based protein quantification, this is one of the first study to examine the effect of both common and rare variants on circulating protein levels in a human cohort, thus identifying known and new pQTLs and a novel potential molecular interaction. The overall paradigm further supports the proof-of-principle of integrating genetics with other omic platforms, using circulating markers as intermediate traits in genetic studies and to identify molecular pathways and interactions. Although they did not study associations of their identified genetic variants with CVD outcomes in their population, their comprehensive in silico analysis places the results in a genomic context and adds strength to the clinical implications of their pQTLs. Their efforts to translate their findings into biological models through functional validation of statistical signals, albeit cursory, are commendable. Such efforts will be important and necessary to continue to build findings from functional genomics into foundations for new diagnostic and therapeutic tools. It is important to note that despite the use of in silico approaches to validate their findings, the study should be replicated in a separate cohort. Moreover, given the primary goal of the study of identifying pQTLs, the use of a nested venous thromboembolism case-control cohort could bias analyses and introduce unknown deviations in the underlying protein distribution. Medication use such as aspirin, which is known to influence levels of some of these proteins, was not addressed. It is somewhat interesting that no transassociations were identified; however, the small sample size likely resulted in insufficient power for such analyses. Finally, although the proteins that were initially included for analysis in this study are indeed relevant to CVD, many were not able to be measured with the sandwich ELISA platform used here. As such, these proteins only represent a snapshot of the full breadth of the proteome that influences CVD. To increase the applicability of the results, future work should focus on a larger, true population-based cohort that would be sufficiently powered to evaluate exome-wide transassociations with more comprehensive protein data.

The study of Solomon et al\(^5\) provides a glimpse into the use of pQTL to offer new insights into the pathophysiology of CVD and highlights the importance of common and rare variation in the genetic architecture of health and disease. By using proteins as a more proximal marker of disease, this is a significant step in the direction of uncovering the biological relevance of common and rare genetic variation. Future studies will ideally build off the paradigm set by this and others to identify more pQTLs, potential new CVD risk variants and novel molecular disease mechanisms by integrating genomewide, rather than exome-wide, common and rare variation with larger scale protein measurements (including mass spectrometry–based proteomics); and eventually integrating these data with transcriptomics, metabolomics, epigenetics, and exposome (eg, medication, lifestyle, and air quality). Such studies will serve as valuable hypothesis-generating resources for other investigators to drive new discoveries in CVD diagnostics and therapeutics.

Disclosures

None.

References


Key Words: Editorials ■ epidemiology ■ biomarker ■ genotype ■ phenotype
Exomes, Proteins, and Cardiovascular Disease: Making Sense of the Signals

Robert W. McGarrah and Svati H. Shah

_Circ Cardiovasc Genet._ 2016;9:318-319
doi: 10.1161/CIRCGENETICS.116.001549

_Circulation: Cardiovascular Genetics_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 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/9/4/318

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/