Biomarker Discovery
Searching for Quality in Quantity
Deepak K. Gupta, MD; Thomas J. Wang, MD

Cardiovascular disease (CVD) remains a leading cause of morbidity and mortality worldwide, emphasizing the need for both primary and secondary prevention. Defining an individual’s CVD risk informs clinical decisions about preventive and therapeutic strategies. It is well-recognized that CVD is a complex phenotype with multifactorial origins; therefore, it is not surprising that no single model can accurately predict the likelihood that an individual will have a CVD event. For example, the Framingham Risk Score, probably the most validated CVD risk prediction model, classifies a significant proportion of individuals who ultimately have events as low to intermediate risk. Consequently, there is substantial interest in identifying novel risk markers that may enhance prognostication, impact clinical management, or provide insight into the pathophysiologic basis for CVD and CVD-related mortality. Although genetic and imaging markers have received increased attention, circulating molecules remain the focus of most biomarker research, with recent studies focusing on the use of combining multiple circulating biomarkers into a predictive panel.

In this issue of Circulation: Cardiovascular Genetics, Halim et al present the results of their evaluation of multiple circulating proteins to predict the risk of death or myocardial infarction in a high-risk cohort. There have been relatively few multimarker studies in high-risk individuals, for example, those with a high burden of conventional risk factors or existing CVD. In a substudy of the Heart Outcomes Prevention Evaluation (HOPE) trial, the associations between 11 biomarkers and recurrent cardiovascular events or death were evaluated. Only N-terminal pro-brain natriuretic peptide (NT-proBNP) improved risk prediction compared with a model based on clinical predictors alone. In the Heart and Soul Study, the association between 6 biomarkers (NT-proBNP, albuminuria, C-reactive protein, IL-6, cystatin-C, and fibrinogen) and recurrent cardiovascular events was evaluated. When the markers were simultaneously assessed, NT-proBNP, albuminuria, and C-reactive protein remained independently associated with the primary outcome. Inclusion of the 3 biomarkers modestly, but significantly, improved risk prediction beyond clinical factors alone. In the Uppsala Longitudinal Study of Adult Men, 4 biomarkers (NT-proBNP, C-reactive protein, troponin I, and cystatin C) improved CVD risk prediction beyond clinical risk factors. Finally, in a Canadian sample of patients undergoing coronary angiography, the combination of C-reactive protein, IL-6, serum amyloid A, and homocysteine predicted all cause and cardiovascular-related death.

Halim et al used a nested case-control design to examine the association of 53 circulating proteins with the risk of death or myocardial infarction in a subset of patients enrolled in the Measurement to Understand the Reclassification of Disease of Cabarrus and Kannapolis (MURDOCK) Horizon 1 Cardiovascular Disease Study. The authors selected 273 cases and 273 age, race, and sex-matched controls, all of whom underwent invasive coronary angiography at baseline. Rather than using traditional Cox proportional hazard models as in previous studies, the investigators used techniques such as penalized logistic regression with an elastic net and bootstrapping to distill the 53 proteins to a smaller set of 6 proteins (matrix metalloproteinase-3, NT-proBNP, IL-6, intercellular adhesion molecule-1, sCD40L, and IGFBP2) that were associated with death or myocardial infarction. Further adjustment for clinical risk factors led to variable results according to whether all clinical risk factors were forced in or only those that survived simultaneous evaluation with biomarkers. In the model conditioned to include all clinical risk factors, only sCD40L was retained in ≥85% of the bootstrapped samples. In the model with simultaneous evaluation of biomarkers and clinical risk factors, 6 biomarkers (matrix metalloproteinase-3, NT-proBNP, IL-6, intercellular adhesion molecule-1, sCD40L, and IGFBP2) were retained.

The study by Halim et al is a valuable addition to the literature on the evaluation of multiple biomarkers in CVD. The investigators examined a substantially larger number of biomarkers compared with previous studies. These biomarkers were selected primarily based on previously published data indicating associations with death or myocardial infarction among individuals with known CVD or CVD risk factors. They used a well-phenotyped population with uniform assessment of clinical characteristics and outcomes, a single core laboratory to perform the assays of interest, and rigorous statistical methods. The positive associations between NT-proBNP, intercellular adhesion molecule-1, and IL-6 and CVD risk are reassuringly consistent with previous studies.
How should we assess the quality of the biomarkers identified? Biomarkers may enhance clinical prediction, influence clinical decisions, or provide biological insight.10 The findings from Halim et al4 diverge somewhat from previous studies with regard to the prognostic use of measuring multiple biomarkers in high-risk individuals. Although addition of the protein measurements to clinical risk factors raised the C-statistic, a measure of model discrimination, from 0.795 to 0.828, this difference was not statistically significant. Other features of model performance, such as calibration and reclassification, could not be assessed given the case–control design. One potential explanation for the lack of incremental prognostic discrimination in this study is that clinical factors alone performed well, with a C-statistic of nearly 0.80. This is in contrast to the HOPE, Heart & Soul, and Uppsala Longitudinal Study of Adult Men studies, in which clinical risk prediction models alone provided lower C-statistics of 0.65, 0.73, and 0.66, respectively, thereby providing greater room for improvement with the addition of biomarkers.5,2 The explanation for the baseline differences is not entirely clear, although the MURDOCK cohort clearly differed from the others by virtue of including only individuals undergoing clinically indicated coronary angiography. Furthermore, the investigators used a slightly different statistical approach to covariate selection, modeling, and cross-validation compared with previous studies.

What about the clinical effect of measuring multiple biomarkers in individuals at high risk for CVD? High-risk individuals benefit the most from preventive and therapeutic measures; therefore, in clinical practice, these patients are typically managed aggressively. Additional biomarker-guided risk stratification may do little to influence management. Indeed, guidelines do not recommend routine biomarker assessment to guide management decisions in patients already known to be at high risk. This may be one reason why conventional risk algorithms are typically applied to those without previous CVD.11

Biomarkers can also provide insight about pathophysiologic abnormalities that precede overt CVD. As Halim et al4 note, the biomarkers identified in their study reflect distinct pathways, including cardiac stress, inflammation, atherosclerosis, vascular structure and function, and metabolism. The results support the concept that CVD is a complex phenotype involving multiple biological pathways.

One might argue that identifying molecules from different biological pathways is a predictable outcome of multimarker studies, because molecules from the same pathway provide overlapping information and are unlikely to be retained together in a predictive model.11 Using simulation data, we have previously shown that the accumulation of moderately correlated biomarkers leads to minimal change in the C-statistic (Figure). The opposite is true for biomarkers with low correlation to each other; addition of such biomarkers can raise the C-statistic significantly.

Nonetheless, the tendency for models to select orthogonal or uncorrelated biomarkers is more than a statistical phenomenon. It highlights the value of assessing large panels of biomarkers representing diverse biological processes. The result is not only better predictive models but also greater opportunity to acquire new insight about the pathophysiology of complex traits such as CVD. Examining scores of proteins, as in this study, can provide important clues and direct further investigation of specific markers from known pathways (such as sCD40L, IGFBP2, and matrix metalloproteinase-3). The advent of technology that allows simultaneous assessment of much larger panels (hundreds or thousands of protein and metabolite biomarkers) should facilitate future analyses that are less biased by known pathways and could reveal previously unsuspected drivers of CVD and CVD-related mortality.1,11 For instance, mass spectrometry–based metabolomic platforms enable the profiling of several hundred small molecules at a time, and have been used in a growing number of epidemiological studies to predict cardiovascular and metabolic diseases.12,13

Proteomic profiling presents a greater analytic challenge than metabolomics, given the far larger number of circulating peptides and enormous dynamic range. Multiplexed immunoassays, as used in the study by Halim et al,14 permit detection of multiple proteins, but profiling on a more global scale requires other technologies. Mass spectrometry remains a mainstay for proteomic studies, but application of this technique for analysis of human plasma is challenging.14–16 Alternative approaches also exist; for instance, modified DNA aptamers have been used to measure hundreds of proteins at a time.17

The clinical and biological value of biomarkers in CVD remains an active area of debate. The research presented by Halim et al4 extends previous studies by examining an expanded panel of circulating biomarkers, but, as the authors indicate, work remains to be done. Future evaluation of CVD biomarkers will probably involve even larger panels of genetic, protein, and metabolic markers, from easily accessible sources (eg, blood and urine).1,18 Finding quality in this large quantity of biomarkers depends on the identification of biomarkers that add prognostically to existing models or implicate new biological pathways.10

Disclosures
Dr Wang is named as a coinventor on patent applications on metabolic and neurohormonal biomarkers of cardiometabolic disease. He has received grant or assay support from Diasorin, Critical Diagnostics, LabCorp, Siemens, Singulex, and Brahms, and has served on advisory committees with Diasorin, Singulex, and Critical Diagnostics.

Figure. Increment in discrimination from adding hypothetical biomarkers, according to the degree of marker–marker correlation (r). The simulated hazard ratios for the outcome is 1.35 per SD increment in the biomarker. The y axis shows the C-statistic from a model containing traditional risk factors plus a variable number of simulated biomarkers (x axis), each with a fixed association with the outcome. The simulation was performed by Michael Pencina (Reprinted from Wang1 with permission of the publisher. Copyright ©2007, Boston University).
References


Key Words: Editorials ◼ biomarkers ◼ metabolomics ◼ prognosis ◼ proteomics
Biomarker Discovery: Searching for Quality in Quantity
Deepak K. Gupta and Thomas J. Wang

_Circ Cardiovasc Genet_. 2015;8:8-10
doi: 10.1161/CIRCGENETICS.115.000973
_Circulation: Cardiovascular Genetics_ is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 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/8/1/8

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/