Biomarkers are clinical, molecular, or image-based measurable parameters that can characterize an individual's specific biological state, whether normal, pathological, or in response to treatment. A biomarker is considered of clinically valuable if (1) it can be measured repeatedly with accuracy and relatively rapid clinical turnaround, (2) it provides unique, superior information on patient status, and (3) it aids in clinical decision making with high precision. High-quality biomarkers can critically inform clinical diagnosis (eg, high-sensitivity troponin for acute myocardial infarction) and guide therapy (eg, CYP2C19 status for clopidogrel therapy). The ideal biomarkers can further reveal underlying biological processes, inform therapeutic deployment, and pave the way for true personalized precision medicine. In this issue of Circulation, Ngo et al demonstrate the use of a developing proteomics technology to rapidly screen for protein biomarkers in patients with planned and spontaneous myocardial infarcts.

How valuable are new, additional biomarkers in cardiovascular medicine? The current medical literature is replete with publications on biomarkers in cardiovascular medicine. A search of the PubMed database revealed 6421 articles on biomarkers and cardiovascular disease in 2015 alone. Unfortunately, despite the wealth of publications, truly high-value biomarkers that are etiologically specific, reproducibly validated in multiple populations, and informative in decision making, and that can be implemented in clinical care are very few indeed. As highlighted in the most recent guidance on personalized medicine for cardiovascular disease from the Food and Drug Administration, the need for high-value, validated biomarkers to guide treatment development is now more urgent than ever.

Advances in systems biology over the past decade have provided more opportunities for innovative discovery of biomarkers than ever before. These advances range from identifying disease-causing genes using deep genome sequencing, the characterization of mRNA, microRNA, and noncoding RNAs through RNA-Seq, and profiling the expressed proteins and their modified states through deep proteomics. For common conditions such as hypertension, coronary disease, or heart failure, the individual genetic influences are relatively small. Analysis of the expressed genome, integrating environmental imprint and influence, such as RNA or proteins, will likely be more fruitful. By using an integrated approach, one can begin to understand the precise interactions of susceptibility, environmental epigenetic regulation, and functional protein production and turnover, which can be potentially detected as a biomarker, to relate to the individual's phenome and ultimately disease outcome.

Proteomics, however, is perhaps the most challenging of the different systems biology approaches, given the complexity of the human proteome. For each expressed gene, there can be tens to hundreds of expressed protein variations, and, with posttranslational modifications, there is another order of magnitude of complexity. Many important signaling or regulatory proteins have concentrations in the
microgram per milliliter to picogram per milliliter range; yet, in the blood, protein concentrations can differ by over 10 orders of magnitude, with high-abundance proteins such as globulin or coagulation factors in the range of grams per milliliter often masking low-abundance proteins. Following tissue injuries, tissue-resident proteins leak into the bloodstream and become diluted into this heterogeneous mixture over \( \approx 5 \) L of peripheral blood. Discovering these needles in a haystack requires technologies capable of exquisite sensitivity and specificity.

For the past decade, researchers have grappled with how to dig deeper into the low-abundance proteins within a sample. First-generation proteomics platforms detected proteins by their overall charge and molecular weight, and were limited to analyzing \( \approx 200 \) different proteins at a time (Figure). Since 2006, such gel-based methods have gradually been replaced by gel-free tandem mass spectrometry, which can characterize up to \( 10,000 \) proteins from complex samples, a number that approaches the entire proteome by some estimates. This has the advantage of being an unbiased, systems-based approach, but significant challenges persist for biomarker screening. For instance, the stochastic sampling methods of shotgun mass spectrometry create a bias against low-abundance proteins. The throughput of discovery-based experiments is usually low, and resulting data sets are often riddled with missing values, particularly irritating methodological weakness in cohort studies. Targeted mass spectrometry approaches can alleviate these problems by explicitly programming the instrument to pick out target proteins, but often require extensive technical fine tuning, resulting in limited numbers of proteins (up to 100) that can be tested at a time.

Aptamer-based protein arrays represent a third way to characterize the proteome that can circumvent some of these limitations in sensitivity, by capturing target proteins through the use of conjugated DNA aptamers. DNA aptamers are single-stranded deoxyribonucleotides that fold into specific 3-dimensional structures with unique binding surfaces as defined by their sequences, giving them affinity to specific epitopes on target proteins. Unlike protein antibodies, DNA aptamers have the advantage of being able to be quickly synthesized and sequenced in large quantities. This allows iterative selection of aptamer ligands with high target affinity from a large sequence library. SOMAmer (slow off-rate modified aptamers) arrays further expand the versatility of target capture by introducing artificial chemical modifications to nucleotides, eg, by linking deoxyuridine with benzyl side chains or other moieties to enhance hydrophobic interactions and stability of binding.

To profile proteins on aptamer arrays, samples are introduced to bead-immobilized aptamers with specific sequences and target affinity. The protein molecules bound on the beads are tagged by biotinylation. The biotinylated aptamer-protein complexes are then cleaved from the beads together and captured by streptavidin, whereas unbound molecules are washed off. Finally, the aptamers that found their protein targets are released from the captive proteins, then identified by using microarray or next-generation sequencing as surrogates of their target proteins’ abundance. An advantage of this approach is the ability to consistently quantify proteins with very low abundance, including cytokines such as interleukin-6 and tumor necrosis factor that are present at the picogram per milliliter range in plasma, a concentration that is currently difficult to detect with mass spectrometry. Combined with the high throughput of nucleotide sequencing, array analysis delivers a sweet spot that combines high sensitivity for low-abundance proteins and high throughput for large numbers of samples.

Despite its great potential, aptamer arrays have yet to be widely tested in clinically relevant replication studies or in cardiovascular cohorts until recently. Ngo et al present a comprehensive demonstration of its applicability to biomarker discovery in cardiovascular medicine. The authors analyzed the abundance of 1129 proteins in a cohort of 30 individuals undergoing planned myocardial infarction, then validated the data in 23 patients with spontaneous myocardial infarction, finding 79 candidate infarct markers. Known markers were reproduced (eg, troponin I and creatine kinase MB) along with novel candidates (eg, fibroblast growth factor 18 and interleukin-11) that were missed in previous studies. The authors further analyzed the plasma of 899 Framingham Heart Study Offspring participants and found association between the abundance of 156 proteins with Framingham Risk Score in the cohort. The baseline concentration of 1 protein (tissue plasminogen activator) was significantly associated with future cardiovascular disease risk after adjustments for age and sex.

The study by Ngo et al is notable in its inclusion of low-abundance cytokines and cell-surface proteins, which are largely excluded in previous proteomics studies. It will be exciting to see whether subsequent work will find validated markers emerging from these previously inaccessible corners of the plasma proteome. At the same time, several limitations of the study warrant consideration. First, although the authors made an admirable attempt to validate aptamer binding, further investigations are needed to ascertain whether any aptamers may be susceptible to off-target ligands, and whether quantification results may be confounded by alteration of antigens (eg, disease-specific posttranslational modifications). Second, the off-the-shelf aptamer array covers only selected proteins; hence, the status of important cardiovascular proteins not covered by the arrays cannot be examined. Last, although several well-established markers are reproduced (eg, troponin I and creatine kinase MB), the biological significance and clinical utility of the remaining candidates have not been examined. Follow-up studies may triage these candidates by determining their tissue specificity, half-life, relevance to known cardiovascular physiology, and whether they...
may be present in other injuries that would confound diagnostic insight. For any new technology, the ultimate test is cross-validation with the gold standard. At present, the most sensitive and accurate technique for low-abundance proteins remains enzyme-linked immunosorbent assay with specific validated antibodies. Commercially available Food and Drug Administration–approved tests are usually the most robust because of the reagent quality and reproducibility requirement for approval. Any new technology will need to be compared with the currently accepted standards for reliability.

Of timely interest is the recent publication by Ganz et al., using the same aptamer-based technology consisting of 1130 low-abundance proteins in a derivation cohort and validation of the proteomic risk score in a second coronary disease cohort. The investigators evaluated the aptamer-based protein candidates initially in the plasma samples of 938 patients from the Heart and Soul Study of stable coronary disease and identified a 9-protein cluster that best predicted cardiovascular outcome. This cluster was then retested in a separate 971-patient sample from a Norwegian HUNT3 coronary cohort. In comparison with the Framingham secondary-event risk model, the addition of the 9-protein risk cluster improved the C-statistic by 0.09 (95% confidence interval, 0.06–0.12) in the derivation cohort and 0.05 (95% confidence interval, 0.02–0.09) in the validation cohort.

The Ganz study confirmed the ability to analyze low-abundance proteins in archived blood samples from large patient cohorts by using high-throughput proteomics, and provide additional value to the current risk models even in these early stages of technology development. However, going forward, investigators using this technology should endeavor to harmonize the testing strategies, standardize analysis protocols, share quality control approaches, and perform linked statistical analyses. This will permit cross-population comparisons, cross-validation of individual protein candidates, pooling of data to increase statistical power, and eventual opportunities to link with data from genetics and other expressed genome analyses in the same cohorts. Ultimately, it is the identification of the most robust independent predictive markers of outcomes, and the understanding of their biological rationale in disease pathophysiology, that will advance the collective goal of personalized precision medicine in cardiovascular disease.

Before then, much work will still need to be done in terms of technological improvements, biological validation, and insights. The study by Ngo et al is a clear demonstration of how new technologies can have a trans-
formative effect on long-standing research objectives. One can almost certainly look forward to more exciting developments in other similar proteomics technologies in the near future. For example, tailored arrays may be envisioned that specifically target the most intensively studied proteins in the cardiovascular system, whereas off-the-shelf arrays may be applied to identify signatures of other organ injuries such as type 1 diabetes mellitus or early-stage renal dysfunction. The high tissue specificity of cell-surface proteins targeted by the aptamer array can potentially provide valuable information to a variety of diagnostic and prognostic scenarios, and may complement ongoing efforts to identify tissue damage via tissue-specific methylation patterns of circulating DNA and other molecules. If these technologies continue to develop in apace, we can look forward to a bounty of new insights for patient care even from minute amounts of liquid biopsies.

SOURCES OF FUNDING
This work was supported in part by grants from the National Heart, Lung, and Blood Institute, Canadian Institutes of Health Research (CIHR), Heart and Stroke Foundation, and Genome Canada. Dr Gramolini is supported by a Canada Research Chair.

DISCLOSURES
None.

AFFILIATIONS
From Department of Physiology, Ted Rogers Centre for Heart Research, University of Toronto, Canada (A.G., P.P.L.); NIH BD2K Center of Excellence in Biomedical Computing, Department of Physiology, UCLA David Geffen School of Medicine, Los Angeles, CA (E.L.); and University of Ottawa Heart Institute, Departments of Medicine & Cellular Molecular Medicine, University of Ottawa, Canada (P.P.L.).

FOOTNOTE
Circulation is available at http://circ.ahajournals.org.

REFERENCES
Identifying Low-Abundance Biomarkers: Aptamer-Based Proteomics Potentially Enables More Sensitive Detection in Cardiovascular Diseases
Anthony Gramolini, Edward Lau and Peter P. Liu

Circulation. 2016;134:286-289; originally published online July 21, 2016; doi: 10.1161/CIRCULATIONAHA.116.022940
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/134/4/286

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation 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 is online at:
http://circ.ahajournals.org//subscriptions/