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

Key Words: Editorials ■ aptamers, nucleotide ■ biomarkers ■ oligonucleotides ■ polymorphism, single nucleotide ■ proteomics

© 2018 American Heart Association, Inc.
https://www.ahajournals.org/journal/circ
with disease outcomes were validated by independent measurements of either elevated low-density lipoprotein cholesterol or ultrasound-confirmed atherosclerosis. Direct protein measurements were made for the 7 gene-protein instruments that virtually associated with elevated low-density lipoprotein cholesterol in MCDS and were found to explain a median variation of 3% of these proteins’ abundance. Four of these proteins were isoforms of apolipoprotein E. The other 3 proteins (catalase, interleukin-27, and granulin) did not survive independent validation. When focusing on atherosclerosis, only 2 of the 6 associated virtual proteins could be validated: CLC1B (C-type lectin domain family 1 member B) and PDGFR-B (platelet-derived growth factor receptor beta).

A potential strength of this study is the achievement of high-throughput, scaled data collection and analysis. The authors achieve this in 3 scaling steps. The first is to use the SOMAscan platform to make the initial protein measurements. The second is to impute protein results in a larger cohort using prediction instruments to generate the virtual proteome. This leverages the huge amount of readily available genomic data and sidesteps the requirement to measure and analyze individual samples, with its attendant complexities of adequate sample quality, storage, preparation, and analysis. Finally, the use of EHRs to generate pheno codes allows comparisons across many disease processes and clinical diagnoses, and effectively multiplexes the diseases that can be analyzed.

The first scaling step relies on the specificity, reproducibility, and quantitative accuracy of the SOMAscan aptamer assays. This platform uses a single binder and direct readout that can jeopardize specificity (Table). In an early assessment using chicken plasma as a negative control, 27.6% of SOMAmers (312/1129) generated signals 10-fold greater signal intensity than in human plasma. A few SOMAmers showed nonspecific binding; 7% bind more than 1 protein, and the remaining 7% bound a protein isoform. Amino acid substitutions conferred by SNPs significantly alter the affinity of aptamers for their target proteins in 32% of the panel, meaning that variations in the predicted abundance assigned to these SNPs are possibly artefactual. When validation using an antibody-based platform was performed in the study by Sun and colleagues, 27.6% of aptamers detect proteins in chicken plasma with 10-fold greater signal intensity than in human plasma. 32% of aptamers display altered binding affinity attributable to SNPs. The second scaling step involves the use of EHRs to generate pheno codes allowing comparisons across many disease processes and clinical diagnoses, and effectively multiplexes the diseases that can be analyzed.

Table. Description of Possible Sources of Errors in Quantification of Protein Abundance in the SOMAscan Platform

<table>
<thead>
<tr>
<th>Cause of Interference</th>
<th>Mechanism</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-reactivity</td>
<td>Aptamers can bind with high affinity to nontarget human proteins, or to different isoforms of the same human protein</td>
<td>7% bind to a different protein 7% bind to an isoform^3</td>
</tr>
<tr>
<td>Nonspecificity</td>
<td>Aptamers are not specific to human proteins but bind to proteins from other species</td>
<td>27.6% of aptamers detect proteins in chicken plasma with 10-fold greater signal intensity than in human plasma^4</td>
</tr>
<tr>
<td>Sequence variation</td>
<td>Single amino acid changes in the aptamer binding site can alter the affinity of binding, and thus spuriously alter the measured abundance</td>
<td>32% of aptamers display altered binding affinity attributable to SNPs^4</td>
</tr>
<tr>
<td>Batch effects</td>
<td>Because aptamers do not provide absolute quantification, calibrators are required to ensure a constant measurement across each plate, and normalization strategies are applied</td>
<td>Intraplate coefficient of variability can be as high as 3.6%, similar to the size of some reported genetic effects^5</td>
</tr>
<tr>
<td>Protein complexes</td>
<td>Plasma proteins can form complexes that may alter their tertiary structure and aptamer affinity</td>
<td>To be determined</td>
</tr>
<tr>
<td>DNA/RNA-binding proteins</td>
<td>DNA/RNA-binding proteins^6 and other carrier proteins, such as albumin, may differentially sponge aptamers in plasma and compromise accurate quantitation</td>
<td>To be determined</td>
</tr>
</tbody>
</table>

SNP indicates single nucleotide polymorphism.
leagues,3 35% of SNP-protein predictive instruments were not replicated. The authors of the current study acknowledge this limitation; a high false-negative rate is the price of statistical prudence, but false positives in studies such as these can undermine or misdirect future research. Beyond the current study, Sun and colleagues’ findings raise concerns not only about the validity of correlating SNPs with this method of affinity-based protein measurement, but also for the wider use of a technique that could give spurious results where minor allele frequencies approach 5%.

The final scaling strategy, the use of EHR, effectively allows the investigators to interrogate over a thousand clinical diagnoses, increasing the potential rate of biomarker discovery. Proteomic investigations previously have only been able to interrogate the prespecified target disease. The use of EHR in epidemiological studies is in its infancy, but has already been used, for example, to predict bleeding outcomes during antiplatelet therapy.7 However, EHR diagnosis codes are not collected with the aim of providing robust scientific data. In the United States, where the eMERGE cohort is recruited, EHR codes are used for billing and costing services. In the United Kingdom, the same codes are used by primary care physicians to document medical diagnoses, in part, to meet incentivized treatment areas. In both cases, reporting biases would be expected, but have not yet been adequately studied or quantified to allow claims about their accuracy.8,9 There is currently no consensus on the accuracy of EHRs, but biases will likely include underdiagnosis of rare diseases and overreporting of diseases in incentivized areas of medical practice. Until the landscape is better mapped, studies using phecodes as surrogates for outcomes will remain speculative.

The integrated use of genomic, proteomic, and phenomic data in combination with empirical and statistical methodology could change the way biomarkers are discovered. A similar aptamer-based approach in a prospective analysis of acute cardiovascular events in a cohort with stable coronary artery disease identified a group of 9 proteins, including troponin I and matrix metalloproteinase-12, that modestly increased the predictive power of a clinical risk score.10 These are already well-described biomarkers. The current study faces the same challenge; some of the protein biomarkers that withstand adjustment for the multiple statistical analyses are already well known, whereas others are missing. Apolipoprotein E levels have been previously identified as a biomarker for cardiovascular risk by direct measurement with mass spectrometry, the gold standard for specific protein identification, which avoids reliance on affinity-based techniques.11–13 However, other well-known SNP-regulated plasma proteins were not detected. For example, apolipoprotein B is strongly related to cardiovascular outcomes but apparently missing from the analysis, despite being included in the SOMAscan panel. These missing findings do not invalidate the positive results presented in this study, but highlight how many other disease associations the genetically predicted protein levels based on the aptamer platform might miss. Another surprising observation in this study is that almost a quarter of associations between genetic predictors and clinical phenotypes are attributed to thrombosis. Alternative explanations for this overrepresentation of thrombosis might include that a prothrombotic state or preanalytical variations in plasma preparations impact aptamer binding. Finally, the high number of associations with the ABO locus highlight the potential for single SNPs to confer pleiotropic effects, rendering the individual proteins they control less meaningful.14

In summary, commercial solutions offer a rapid and convenient way of outsourcing protein measurements, and imputation removes the inconvenience of empirical measurement. Proof of principle, however, still requires independent validation of protein levels through orthogonal validation techniques and linkage to cis-acting Mendelian randomization that do not have epistatic effects. In particular, changes in electric charge caused by amino acid substitution may alter the binding properties of the negatively charged aptamers. Thus, the SOMAscan may not be as unbiased as it first may seem. The imputation approach may work well for certain plasma proteins that have a high degree of heritability and can be demonstrated to avoid the potential confounding we outline. These protein quantitative trait loci, made publicly available by the authors, could provide the map for further investigations in highly genotyped populations. Overall, the increasing abstraction from empirical science leaves us with a sense of virtual reality; to quote from Tractatus Logico-Philosophicus by the Austrian philosopher Ludwig Wittgenstein, “Whereof one cannot speak, thereof one must be silent.”15

ARTICLE INFORMATION
Correspondence
Manuel Mayr, MD, PhD, King’s British Heart Foundation Centre, King’s College London, 125 Coldharbour Ln, London SE5 9NU, United Kingdom. Email manuel.mayr@kcl.ac.uk

Affiliations
King’s College London British Heart Foundation Centre, School of Cardiovascular Medicine and Sciences, United Kingdom (A.J., M.M.). Bart’s Heart Centre, St. Bartholomew’s Hospital, London, United Kingdom (A.J.).

Sources of Funding
Dr Joshi is a British Heart Foundation Clinical Research Training Fellow (FS/16/32/32184). Dr Mayr is a British Heart Foundation Chair Holder (CH/16/3/32406) with British Heart Foundation program grant support (RG/16/14/32397).

Disclosures
Dr Mayr is named inventor on patents related to biomarkers in cardiometabolic disease.
REFERENCES


