Proteomics
Proteomics and Systems Biology
www.proteomics-journal.com

Comparison of Proteomic Assessment Methods in Multiple Cohort Studies

Laura M. Raffield, Hong Dang, Katherine A. Pratte, Sean Jacobson, Lucas A. Gillenwater, Elizabeth Ampleford, Igor Barjaktarevic, Patricia Basta, Clary B. Clish, Alejandro P. Comellas, Elaine Cornell, Jeffrey L. Curtis, Claire Doerschuk, Peter Durda, Claire Emson, Christine M. Freeman, Xiuqing Guo, Annette T. Hastie, Gregory A. Hawkins, Julio Herrera, W. Craig Johnson, Wassim W. Labaki, Yongmei Liu, Brett Masters, Michael Miller, Victor E. Ortega, George Papanicolaou, Stephen Peters, Kent D. Taylor, Stephen S. Rich, Jerome I. Rotter, Paul Auer, Alex P. Reiner, Russell P. Tracy, Debby Ngo, Robert E. Gerszten, Wanda K. O'Neal, Russell P. Bowler, and NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium*

Novel proteomics platforms, such as the aptamer-based SOMAscan platform, can quantify large numbers of proteins efficiently and cost-effectively and are rapidly growing in popularity. However, comparisons to conventional immunoassays remain underexplored, leaving investigators unsure when cross-assay comparisons are appropriate. The correlation of results from immunoassays with relative protein quantification is explored by SOMAscan. For 63 proteins assessed in two chronic obstructive pulmonary disease (COPD) cohorts, subpopulations and intermediate outcome measures in COPD Study (SPIROMICS), and COPDGene, using myriad rules based medicine multiplex immunoassays and SOMAscan, Spearman correlation coefficients range from −0.13 to 0.97, with a median correlation coefficient of ≈0.5 and consistent results across cohorts. A similar range is observed for immunoassays in the population-based Multi-Ethnic Study of Atherosclerosis and for other assays in COPDGene and SPIROMICS. Comparisons of relative quantification from the antibody-based Olink platform and SOMAscan in a small cohort of myocardial infarction patients also show a wide correlation range. Finally, *cis* pQTL data, mass spectrometry aptamer confirmation, and other publicly available data are integrated to assess relationships with observed correlations. Correlation between proteomics assays shows a wide range and should be carefully considered when comparing and meta-analyzing proteomics data across assays and studies.

1. Introduction

Immunoaffinity assays have long been the standard method for assessing protein concentrations in plasma and other tissues. Some multiplexed immunoaffinity methods are available, but only ≈30–50 proteins can be accurately assessed at once, and cross-reactivity may reduce sensitivity as compared to monoplex immunoassays.[1] There is increasing interest in the use of novel large scale proteomics platforms to facilitate biomarker discovery for sub-phenotyping and risk prediction in complex diseases.[2] Oligonucleotide aptamer-based platforms, such as the SOMAscan assays marketed by SomaLogic (Boulder, CO), can be used for quantification of >1300 proteins,[3–6] with >4000 proteins assayed in recent publications.[7] Aptamers are randomly generated nucleotide sequences that bind specific antigens, mimicking the function of antibodies. Aptamer-based assays can detect very low abundance proteins compared to immunoassays,

Dr. L. M. Raffield,
Department of Genetics
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599
E-mail: laura_raffield@unc.edu

Membership in the Trans-Omics for Precision Medicine Consortium is listed in the Supporting Information.

DOI: 10.1002/pmic.201900278

Dr. H. Dang, Dr. C. Doerschuk, Dr. W. K. O'Neal
Marsico Lung Institute
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599

Dr. K. A. Pratte, S. Jacobson, L. A. Gillenwater, Dr. R. P. Bowler
Division of Pulmonary Medicine
Department of Medicine
National Jewish Health (NJH)
Denver, CO 80206

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

but, as a discovery tool, the large SOMAscan platform provides only relative quantification, rather than absolute concentrations. Although the SOMAscan platform has been useful to detect reproducible protein quantitative trait loci (pQTLs) and epidemiological associations, problems with specificity have been reported for a subset of proteins.[8,9] Relatively little information is publicly available on how well the protein levels assessed using the SOMAscan platform correlates with multiplex or individual immunoaffinity assays. In a small sample ($n \leq 42$) comparing nine proteins measured with both aptamer assays and other clinical assays in the Atherosclerosis Risk in Communities Study, the median Spearman correlation coefficient was >0.8 at two separate visits.[5] Comparisons of 20 proteins assessed with both Luminex xMAP immunoassays and SOMAscan in a small number of serum samples demonstrated that nearly half had Pearson correlation coefficients <0.5[10] when compared to SOMAscan relative fluorescence units. Poor correlation across assays, like that reported comparing SOMAscan and Luminex immunoassays, is certainly not unique to aptamer based platforms, as immunoaffinity assays also often have relatively low interassay correlations.[10,11] However, to facilitate large genetic and epidemiological meta-analyses across traits, more publicly available information is necessary regarding which proteins correlate reliably across commercially available platforms. There is also very little information on correlation between SOMAscan and Olink assays; Olink is another popular high throughput platform using proximity extension assay technology to allow for multiplex immunoassays (92 proteins across 96 samples) without the extensive cross-reactivity that would occur using conventional ELISAs at this scale.[12]

Dr. E. Ampleford, Dr. A. T. Hastie, Dr. V. E. Ortega, S. Peters
Department of Internal Medicine
Section for Pulmonary
Critical Care
Allergic and Immunologic Diseases
School of Medicine
Wake Forest University
Winston-Salem, NC 27157

Dr. I. Barjaktarevic
UCLA Division of Pulmonary and Critical Care
David Geffen School of Medicine
University of California Los Angeles
Los Angeles, CA 90095

Dr. P. Basta
Department of Epidemiology
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599

Dr. C. B. Clish
Metabolomics Platform
Broad Institute
MIT and Harvard
Cambridge, MA 02142

Dr. A. P. Comellas
Department of Internal Medicine
University of Iowa
Iowa City, IA 52242

E. Cornell, Dr. P. Durda, Dr. R. P. Tracy
Department of Pathology and Laboratory Medicine
Larner College of Medicine
University of Vermont
Burlington, VT 05405

Dr. J. L. Curtis, Dr. C. M. Freeman, Dr. W. W. Labaki
Division of Pulmonary and Critical Care Medicine
University of Michigan
Ann Arbor, MI 48109

Dr. J. L. Curtis, Dr. C. M. Freeman
VA Ann Arbor Healthcare System
Ann Arbor, MI 48105

Dr. C. Emson
Translational Science and Experimental Medicine
Research and Early Development
Respiratory and Immunology
BioPharmaceuticals R&D
AstraZenenca
Gaithersburg, MD 20878

Dr. X. Guo, Dr. K. D. Taylor, Dr. J. I. Rotter
The Institute for Translational Genomics and Population Sciences
Department of Pediatrics
The Lundquist Institute for Biomedical Innovation
Harbor-UCLA Medical Center
Torrance, CA 90502

Dr. G. A. Hawkins
Department of Biochemistry
Wake Forest School of Medicine
Winston-Salem, NC 27517

Dr. J. Herrera, Dr. B. Masters, Dr. M. Miller
ProterixBio, LLC
Billerica, MA 01821

W. C. Johnson
Department of Biostatistics
University of Washington
Seattle, WA 98195

Dr. Y. Liu
Department of Epidemiology and Prevention
Wake Forest School of Medicine
Winston-Salem, NC 27517

Dr. G. Papanicolaou
Epidemiology Branch
Division of Cardiovascular Sciences
National Heart, Lung and Blood Institute
Bethesda, MD 20892

Prof. S. S. Rich
Center for Public Health Genomics
University of Virginia
Charlottesville, VA 22908

Dr. P. Auer
Zilber School of Public Health
University of Wisconsin Milwaukee
Milwaukee, WI 53201

Dr. A. P. Reiner
Fred Hutchinson Cancer Research Center
Seattle, WA 98109

Dr. A. P. Reiner
Department of Epidemiology
University of Washington
Seattle, WA 98195

Dr. R. P. Tracy
Department of Biochemistry
Larner College of Medicine
University of Vermont
Burlington, VT 05405

Dr. D. Ngo
Division of Pulmonary, Sleep and Critical Care Medicine
Beth Israel Deaconess Medical Center
Boston, MA 02215

Dr. R. E. Gerszten
Division of Cardiovascular Medicine
Beth Israel Deaconess Medical Center
Boston, MA 02215

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

This study explores inter-platform correlations of proteomics assays using data available from four cohorts: Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS), COPDGene, the Multi-Ethnic Study of Atherosclerosis (MESA), and a small cohort of patients undergoing septal ablation for hypertrophic cardiomyopathy (inducing planned myocardial infarction).[3] The first two cohorts (SPIROMICS[13] and COPDGene[14]) include current and former smokers with or without chronic obstructive pulmonary disease (COPD). The MESA cohort is diverse, recruited from the general population, and free of cardiovascular disease at baseline.[15] Each cohort had multiple assays performed on multiple platforms on aliquots taken from the same plasma or serum sample. These platforms included the SOMAscan 1.1 k or 1.3 k array, a custom Myriad Rules Based Medicine (RBM) set of multiplexed Luminex assays,[16,17] Meso Scale Discovery (MSD) assays, Olink panels, and ProterixBio assays, along with many additional individual immunoassays. The different platforms have overlapping proteins, allowing us to explore correlations of target proteins across platforms. Similar assays available in COPDGene and SPIROMICS allowed assessment of many of the same correlations (notably between the SOMAscan 1.3 k array and Myriad RBM assays) in two independent cohorts.

Significance Statement

This paper provides information on the comparability of antibody- and aptamer-based protein measures across multiple cohort studies. As new multi-cohort and multi-platform meta-analyses and replication efforts are initiated using novel proteomics assays, our analysis suggests that investigators must more fully consider differences in protein concentration measurements obtained from different platforms and assess the level of correlation between those platforms. In addition to correlation data, metrics such as mass spectrometry-based aptamer confirmation or the presence of *cis* pQTLs may help infer the specificity of different proteomics platforms when results differ.

2. Experimental Section

2.1. Cohorts

SPIROMICS (ClinicalTrials.gov Identifier: NCT01969344) was a multi-center cohort study that recruited ever-smokers (n = 2,772, ≥20 pack-years, no exacerbation for >30 days) with and without COPD as well as age and gender-matched never-smokers (n = 202). SPIROMICS study design has been previously described.[13] All subjects provided serum using an SST tube (Becton Dickinson) and fresh frozen plasma collected using either an 8.5 mL EDTA collection tube or a P100 tube with K2EDTA, which contained anti-coagulant and proprietary protease inhibitor additives (antiproteases; BD product number 366448).[18] SOMAscan data (1.3 k version) was available in 288 participants, of whom 176 overlap with proteomics data available from a custom 13-panel multiplex assay (114 protein measures, Myriad-RBM, Austin, TX).[16] Additionally, data was available from a Meso Scale Discovery (MSD) assay assessing nine proteins (TNFα; IL-2, -6, -8, -10; interferon γ; CCL11 (eotaxin); CCL26 (eotaxin 3); CCL17)[19] and from unpublished ProterixBio ELISA data for 23 target proteins. ProterixBio immunoassays were developed using commercially sourced antibodies and reagents;[20] detailed methods are included in the Supporting Information.

COPDGene (ClinicalTrials.gov Identifier: NCT00608764) was a multi-center cohort study designed to identify genetic factors associated with COPD. Phase 1 visits were conducted from January 2008 until April 2011, with a subset of participants returning for a second phase approximately five years later. COPDGene recruited 10 263 current and former smokers (≥10 pack-years, no exacerbation >30 days) with and without COPD of non-hispanic white and African American ancestry,[14] age- and sex-matched healthy individuals ($n = 108$) with no history of smoking were enrolled as controls during phase 1, with an additional 347 never-smokers enrolled during phase 2. A subset of subjects in both phase 1 and phase 2 provided fresh frozen plasma collected using an 8.5 mL p100 tube (Becton Dickinson—BD).[16,19,21] SOMAscan data (1.3 k version) was available in 1248 participants at phase 1 and 1086 participants at phase 2. Data from the same multiplex RBM assays as used in SPIROMICS was available in 602 individuals at phase 1, of whom 371 overlapped with SOMAscan.[16] The same participants from phase 2 also had MSD assays available at this timepoint ($n = 500$ overlapping).[19] Additionally, data from monoplex quotient bio research (QBR) assays at phase 1 was available for some proteins, such as C-reactive protein (n = 1096 overlapping).[21]

MESA was a community-based cohort study designed to determine the prevalence, determinants, and progression of subclinical cardiovascular disease (CVD). MESA recruited men and women aged 45–84 free of clinical CVD at baseline from four major race/ethnicity groups from six different locations in the United States (2000–2002).[22] Exam 5 was conducted from 2010–2011. Correlations for target proteins were assessed where individuals were assessed with both conventional ELISA assays and SOMAscan using the same blood sample. SOMAscan data was newly generated for the MESA cohort by the NHLBI Trans-Omics for Precision Medicine Initiative (TOPMed). Biomarker assays were previously reviewed[23] (see Methods section in Supporting Information for references).

Olink and SOMAscan 1.1 k data was available in 48 samples from ten patients undergoing septal ablation for hypertrophic cardiomyopathy, with four timepoints each (baseline, 10 min, 1 h, and 24 h after intervention).[3] Raw data is available in Data Object 1, Supporting Information. From the overlapping Olink and SOMAscan 1.1 k results, a small number of analytes (eukaryotic translation initiation factor 5A-1 (eIF-5A-1), interferon gamma (IFN-γ), interleukin 13 (IL-13), interleukin 1 alpha (IL-1α), interleukin 20 (IL-20), interleukin 22 receptor subunit alpha 1 (IL22RA1), tumor necrosis factor alpha (TNF-α), and thymic stromal lymphopoietin (TSLP)) were excluded due to Olink assay results below the lower limit of detection.

In all cases, the results of samples drawn on the same individual participants were compared during the same study visit. Note that, by design, there was no overlap of study participants between the cohorts. Participants in all included studies provided

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

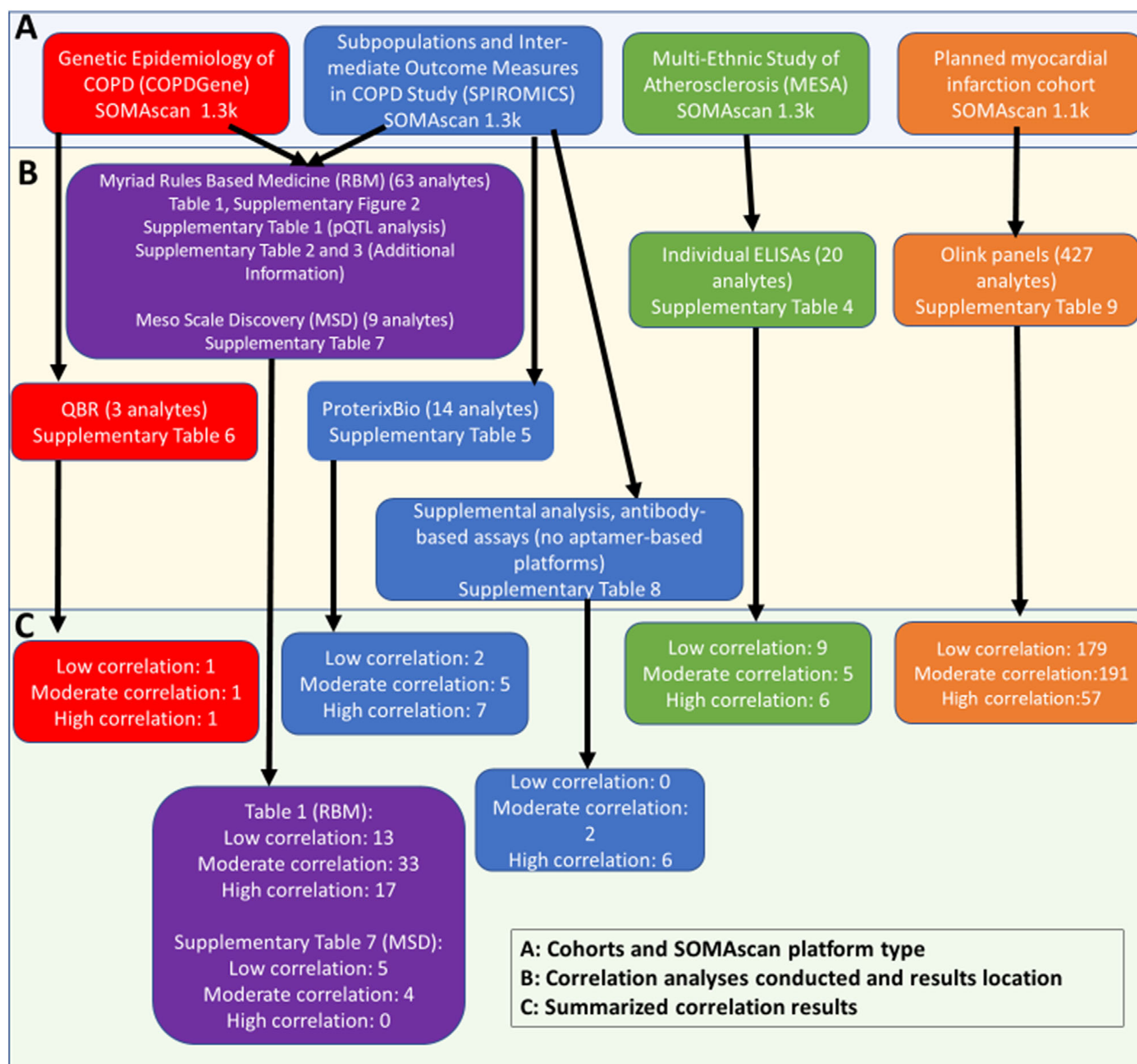

Figure 1. Flowchart of proteomics assays compared, by cohort. We also have listed total numbers of low correlation ($r_s < 0.3$ in all included cohorts), moderate correlation, and high correlation ($r_s > 0.7$ in all included cohorts) proteins in each results table.

written informed consent, and studies were approved by the relevant institutional review boards at all participating centers.

2.2. Analysis Methods

Cohorts and proteomics assays compared are summarized in **Figure 1**. To compare the protein assessment platforms, only proteins for which ≤20% of the assay results were below the lower limit of detection (LLOD) were analyzed. As many assessed proteins had non-normal distributions, a non-parametric method was used, Spearman correlation coefficients, calculated using R 3.5.3 or SAS 9.3. Proteins were matched wherever possible by Uniprot ID, taking into account additional annotation information if needed.

The largest set of overlapping proteins in an adequate sample size for genetic analysis (from Myriad RBM and SOMAscan in COPDGene and SPIROMICS) was also used for *cis* pQTL analysis, using the same set of overlapping samples used for correlation analyses. Genetic association analyses were performed using R with MatrixEQTL version 2.2, biomaRt version 2.38.0, and snpStats version 1.32.0. An additive genetic linear regression model was used, including five genotype principal components to control for population substructure, sex, age, and current smoking status. All protein concentration values were inverse normal transformed for the analysis. Genotype data in COPDGene was from a single version of the

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Illumina HumanOmniExpress Beadchip. SPIROMICS used data from three versions of the Illumina OmniExpress HumanExome BeadChip.[16] Variants with a minor allele frequency <1% were removed from the analysis. The *p*-value threshold for declaring a significant *cis* pQTL was $p < 1 \times 10^{-4}$ within 1 Mb of the gene encoding the protein product.

Information on non-specific aptamer binding, mass spectrometry validation, and pQTLs were compiled and integrated from large previous SOMAscan studies[7,24] and a previous analysis of Myriad RBM data in COPDGene and SPIROMICS[16] in a large sample size.

3. Results

3.1. SOMAscan versus Myriad-RBM Immunoassays

We first compared the results of a previously measured custom 13-panel multiplex assay (Myriad-RBM)[16] with SOMAscan 1.3 k data (**Table 1**, **Figure 2**a). There were overlapping data from 371 subjects in COPDGene and 176 subjects from SPIROMICS. While all data was from blood collection at the same study visit, the SPIROMICS SOMAscan data used P100 fresh frozen plasma, but the SPIROMICS RBM data were generated from EDTA plasma or SST serum.[18] Spearman correlation coefficients for individual proteins had a similar range in both cohorts ($-0.07 \leq r_s \leq 0.97$ in COPDGene, $-0.13 \leq r_s \leq 0.97$ in SPIROMICS); the median correlation coefficient was 0.57 in COPDGene and 0.46 in SPIROMICS (Table 1).

For the 63 proteins assessed, the correspondence between the intra-study correlation coefficients was high between the COPDGene and SPIROMICS studies ($r_s = 0.88$) (Figure S1, Supporting Information). There are a few notable outliers, such as angiopoietin-1, with $r_s = 0.26$ in SPIROMICS and $r_s = 0.75$ in COPDGene and superoxide dismutase [Cu-Zn], with $r_s = 0.21$ in SPIROMICS and $r_s = 0.6$ in COPDGene. We investigated if those disparities might result from use of serum for some RBM assays in SPIROMICS,[16] instead of plasma (as used in COPDGene). However, five of the six proteins with the greatest difference in correlation between cohorts had assays performed using plasma in SPIROMICS, arguing against that explanation (though differences between the p100 plasma used in COPDGene and EDTA plasma used in SPIROMICS for Myriad RBM assays,[16] particularly in terms of protease digestion, may play a role). In total, 17 proteins (27%) had $r_s \geq 0.7$ in both cohorts (high correlation), and 13 proteins (21%) had $r_s < 0.3$ in both cohorts (low correlation) (Figure S1, Supporting Information). We next assessed the presence/absence of *cis* pQTLs with reference to the SOMAscan and Myriad-RBM assays in the same samples used for evaluation of Spearman correlation coefficients (Table 1; Table S1, Supporting Information). The presence of a *cis* pQTL can provide a measure of aptamer validation, suggesting robust binding to the target protein (though not excluding off-target binding). Of the 63 proteins assessed, 31 had evidence of a *cis* pQTL in at least one study for at least one assay. For proteins with *cis* pQTL evidence in at least one study for both assays, the median correlation was 0.505, similar to the overall median. Discordant *cis* pQTL evidence was also evaluated. Five proteins had *cis* pQTL evidence with Myriad-RBM assays in both COPDGene and SPIROMICS and did not have *cis* pQTL evidence from SOMAscan data in either cohort. These five proteins (CCL24, CXCL5, IL2RA, SFTPD, MMP3) had a low median correlation coefficient across the two cohorts ($r_s = 0.1$). ICAM1 had a *cis* pQTL in both cohorts for SOMAscan data only and also has a low correlation coefficient between assays ($r_s \leq 0.13$). Based on this data, presence of concordant *cis* pQTLs does not ensure high correlation across assays, but discordant *cis* pQTLs are generally found only for low correlation assays, with the assay with the *cis* pQTL likely having higher binding specificity for the target protein.

We additionally compiled publicly available data which might be informative for antibody and aptamer binding specificity to evaluate if this information was predictive of protein assay correlations. Of the 63 Myriad RBM target proteins, 42 (67%) have had reported *cis* pQTLs in either the AGES[7] or INTERVAL[24] studies (median $r_s = 0.48$, similar to all tested proteins), with 21 having a reported *cis* pQTL in a previously published COPDGene/SPIROMICS Myriad RBM pQTL study (in a larger sample size than the samples used here with overlapping SOMAscan data)[16] (Table S2, Supporting Information). Proteins with a *cis* pQTL for both assays in these previous studies had a median r_s of 0.40. Details on post-translational modifications and isoforms from UniProtKB and on common coding SNPs (from gnomAD server), which could impact antibody/aptamer binding, on proteins with discordant pQTL evidence or with $r_s < 0.3$ in SPIROMICS and COPDGene are also listed in Table S3, Supporting Information, with high correlation proteins also listed for comparison. Few clear patterns are observed. Many SOMAscan aptamers have also been tested for cross-reactivity against homologous proteins (at least 40% sequence homology).[24] Of the 49 tested, seven have comparable binding with at least one protein (median Spearman correlation coefficient of 0.46 with RBM assays, versus median correlation of 0.56 for all 49 aptamers tested). Finally, Emilsson et al. additionally examined evidence of aptamer binding to target protein using a subset of aptamers by multiple reaction monitoring or data-dependent analysis mass spectrometry from AGES.[7] Proteins with mass spectrometry confirmation were quite clearly enriched in the high correlation group (12/17) versus the low correlation group (0/14, see Table 1, Figure 2a). Confirmation of aptamers by mass spectrometry was associated with higher correlation of SOMAscan with immunoassays, but few clear patterns were seen for other factors, such as known aptamer cross-reactivity with homologous proteins, that might also have been hypothesized to systematically interfere with accurate quantification.

3.2. SOMAscan versus MSD, ProterixBio, and Other Assays

We also assessed Spearman correlation coefficients between SOMAscan data and other antibody-based platforms, including MSD (SPIROMICS/COPDGene), individual ELISAs (COPDGene, MESA), and ProterixBio (SPIROMICS). Median correlation values were similar for immunoassays in MESA ($r_s = 0.41$ for 18 unique proteins at Exam 1 (Table S4, Supporting Informatio)), ProterixBio assays in SPIROMICS (median $r_s = 0.63$ for 14 unique proteins (Table S5, Supporting Information)), and QBR assays in COPDGene (median $r_s = 0.49$ for

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Correlation coefficients for the 63 proteins assessed between the SOMAscan and Myriad RBM platform in both COPDGene and SPIROMICS. Presence or absence of a *cis* pQTL variant ($p < 1 \times 10^{-4}$ within one megabase of the gene encoding the protein product) is also listed. r_s, estimated Spearman's rho. We have separated out low correlation ($r_s < 0.3$ in both cohorts) and high correlation ($r_s \geq 0.7$ in both cohorts) proteins, with all other proteins listed in the moderate correlation category. We also list whether aptamers have *cis* pQTLs found in genetic analyses from either of the larger AGES[7] and INTERVAL[24] studies using SOMAscan data or from previous analysis of the full published sample size in SPIROMICS and COPDGene cohorts with Myriad RBM data.[16] Finally, we list if an aptamer has been validated as having at least some binding to its target by mass spectrometry in AGES,[7] and whether an aptamer has comparable binding in cross-reactivity testing against homologous proteins (at least 40% sequence homology) in ref. [24]. UniProt ID, gene name, and aptamer SomaId are listed in Table S2, Supporting Information.

	SPIROMICS				COPDGene				Evidence from prior publications			
Protein	N	r_s	Myriad pQTL	SOMAscan pQTL	N	r_s	Myriad pQTL	SOMAscan pQTL	Any evidence of SomaScan pQTL	Myriad pQTL	AGES Mass Spec.	INTERVAL Comparable binding to other proteins
Low correlation												
Alpha-1-antitrypsin	175	0.26	yes	yes	371	0.24	yes		yes	yes		not tested
Cadherin-1	176	0.21			371	0.14	yes		yes			no
C—C motif chemokine 13	171	-0.02			364	0.12						no
C—C motif chemokine 24	175	0.1	yes		370	-0.02	yes			yes		no
C—C motif chemokine 8	176	0.2	yes		369	0.18	yes	yes	yes	yes		no
Haptoglobin	175	0.2	yes	yes	367	0.26	yes	yes	yes	yes		yes
Hepatocyte growth factor	175	0			368	0.15		yes	yes			no
Intercellular adhesion molecule 1	174	0.04		yes	369	0.13		yes	yes			not tested
Interleukin-2 receptor subunit alpha	176	0.08	yes		370	0.12	yes			yes		not tested
Platelet endothelial cell adhesion molecule	175	0.04			370	-0.07				yes		not tested
Pulmonary surfactant-associated protein D	175	-0.13	yes		370	0.02	yes			yes		not tested
Tumor necrosis factor receptor superfamily member 11B	176	-0.11			370	-0.03			yes			no
Vascular endothelial growth factor A	175	0.03	yes	yes	364	0.08	yes	yes	yes	yes		no
Moderate correlation												
Alpha-2-macroglobulin	175	0.45	yes		371	0.7					yes	no
Angiopoietin-1	175	0.26			370	0.75			yes		yes	no
Antileukoproteinase	176	0.41			371	0.62			yes		yes	no
Beta-2-microglobulin	176	0.66			371	0.49					yes	no
Brain-derived neurotrophic factor	175	0.48			368	0.82			yes			no
C—C motif chemokine 2	175	0.42			370	0.46						no
C—C motif chemokine 22	175	0.62			370	0.74			yes			yes
C—C motif chemokine 23	175	0.69	yes	yes	371	0.59	yes		yes	yes		yes
C—C motif chemokine 5	175	0.56			371	0.91			yes			no
Coagulation factor VII	176	0.45	yes	yes	369	0.56	yes	yes	yes	yes	yes	no
Complement C3	175	0.35			371	0.45	yes		yes	yes	yes	yes
Creatine kinase M-type:Creatine kinase B-type heterodimer	174	0.46			370	0.44			yes		yes	yes
C—X—C motif chemokine 5	175	0.31	yes		369	0.39	yes		yes	yes	yes	no
C—X—C motif chemokine 9	175	0.66			371	0.56						no
Decorin	174	0.38			370	0.54						no
Fibrinogen	173	0.12			370	0.32					yes	yes
Interleukin-16	175	0.32	yes	yes	370	0.33	yes	yes	yes	yes		not tested
Interleukin-18-binding protein	175	0.69			370	0.57					yes	not tested
Interleukin-8	175	0.52			366	0.55						no
Lactotransferrin	174	0.34			358	0.48	yes		yes	yes		no

(*Continued*)

 © 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1. Continued.

	SPIROMICS				COPDGene				Evidence from prior publications			
Protein	N	r_s	Myriad pQTL	SOMAscan pQTL	N	r_s	Myriad pQTL	SOMAscan pQTL	Any evidence of SomaScan pQTL	Myriad pQTL	AGES Mass Spec.	INTERVAL Comparable binding to other proteins
Mast/stem cell growth factor receptor Kit	175	0.62			371	0.65					yes	no
Matrix metalloproteinase-9	167	0.46			368	0.63		yes	yes		yes	no
Metalloproteinase inhibitor 1	176	0.54			371	0.25						not tested
Metalloproteinase inhibitor 2	176	0.45			371	0.13	yes	yes	yes		yes	no
Myoglobin	175	0.62			371	0.57					yes	no
Neuronal cell adhesion molecule	173	0.38			370	0.4	yes		yes	yes		no
Plasminogen activator inhibitor 1	174	0.69			367	0.89			yes			no
Stromelysin-1	176	0.34	yes		369	0.1	yes		yes	yes		no
Superoxide dismutase [Cu-Zn]	176	0.21			370	0.6						not tested
Thyroxine-binding globulin	175	0.64			371	0.65					yes	no
Transforming growth factor beta-1	176	0.21			370	0.38			yes			no
Tumor necrosis factor receptor superfamily member 6	175	0.44			358	0.61		yes	yes			not tested
von Willebrand factor	175	0.48			365	0.45			yes		yes	not tested
High correlation												
Adiponectin	175	0.94			371	0.91			yes		yes	no
Advanced glycosylation end product-specific receptor, soluble	172	0.7			366	0.7	yes	yes	yes	yes		no
C—C motif chemokine 16	175	0.77	yes	yes	371	0.79	yes	yes	yes	yes	yes	no
C—C motif chemokine 18	175	0.81	yes	yes	371	0.94	yes	yes	yes	yes	yes	no
Chromogranin-A	176	0.92			370	0.89		yes	yes		yes	not tested
C-reactive protein	174	0.97	yes		371	0.96			yes		yes	no
C—X—C motif chemokine 10	175	0.76			371	0.78	yes	yes	yes			no
E-selectin	175	0.84			371	0.87			yes		yes	no
Ferritin	175	0.97			371	0.97					yes	yes
Immunoglobulin A	175	0.88			370	0.85					yes	no
Immunoglobulin M	175	0.72			371	0.73						no
Interleukin-6 receptor subunit alpha	176	0.76	yes	yes	368	0.73	yes	yes	yes	yes	yes	not tested
Serum amyloid P-component	175	0.8			371	0.86			yes		yes	no
Sex hormone-binding globulin	175	0.92	yes		371	0.97			yes	yes	yes	not tested
Tumor necrosis factor receptor superfamily member 1A	176	0.7			370	0.73						no
Tumor necrosis factor receptor superfamily member 1B	176	0.8		yes	371	0.83			yes			no
Vascular cell adhesion protein 1	176	0.78			371	0.72					yes	no

Of the 114 analytes on the RBM platform, 18 do not have a corresponding protein on the SOMAscan panel, and missingness was high (>20% of samples were below the lower limit of detection (LLOD)) in an additional 33 analytes (COPDGene ($n = 30$) and SPIROMICS ($n = 32$)). A few proteins had low missingness in only one study, three in COPDGene (SL000248/Alpha-1-antichymotrypsin, $r_s = 0.33$, $n = 370$, SL001802/IFN-g, $r_s = 0.11$, $n = 339$, SL002621/midkine, $r_s = 0.23$, $n = 369$) and one in SPIROMICS (SL002785/N-terminal pro-BNP, $r_s = 0.94$, $n = 172$).

three unique proteins (Table S6, Supporting Information)). Correlations were lower for the nine unique proteins in the inflammation-focused MSD platform ($r_s = 0.38$ in SPIROMICS, $r_s = 0.28$ in COPDGene) (Table S7, Supporting Information). Comparisons between antibody-based platforms were not the central focus of our analyses, as this topic has been assessed in a number of previous efforts,[11,25] but are presented for SPIROMICS (Table S8, Supporting Information), with a median r_s of 0.76 for Myriad RBM and ProterixBio, higher than for antibody versus SOMAscan comparisons (n = 323 individuals,

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

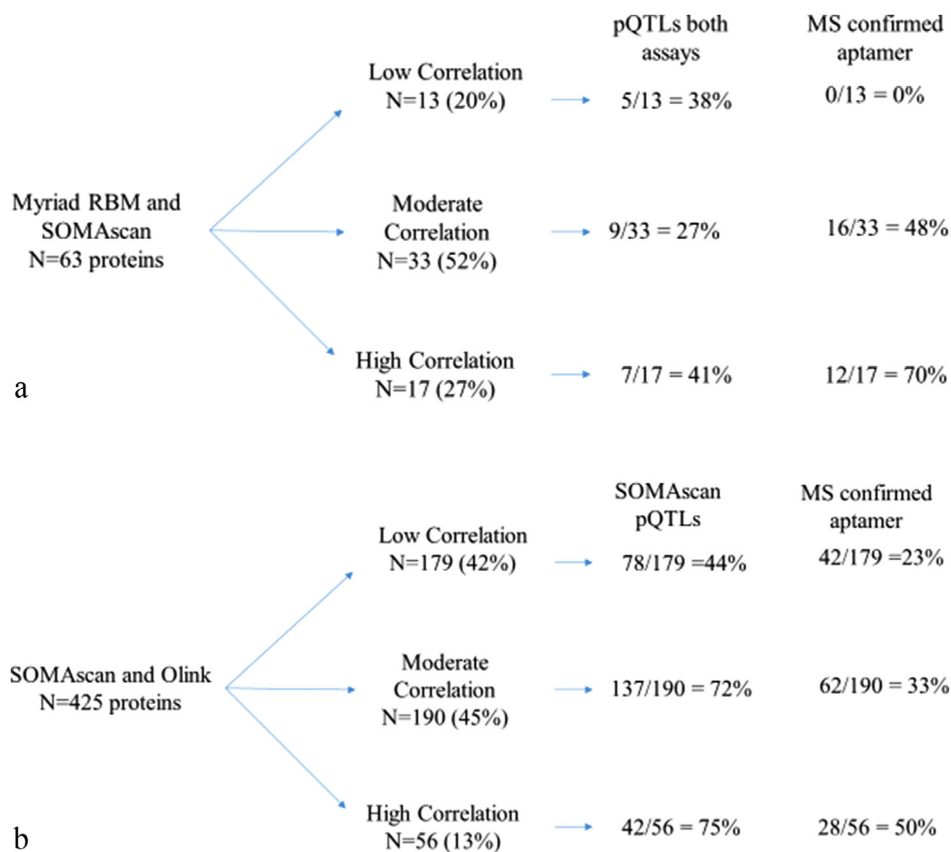

Figure 2. Summary of correlations in datasets with the greatest number of proteins tested. We have separated out low correlation ($r_s < 0.3$ in both cohorts) and high correlation ($r_s \geq 0.7$ in both cohorts) proteins, with all other proteins listed in the moderate correlation category. a) Comparisons between SOMAscan 1.3 k and Myriad RBM assays in SPIROMICS and COPDGene. Concordance of pQTLs (i.e. presence of a *cis* pQTL for both assays) is evaluated for Myriad RBM *cis* pQTLs in any dataset (SPIROMICS samples with SOMAscan, COPDGene samples with SOMAscan, or previously published meta-analysis[16]) versus SOMAscan pQTLs detected in any dataset (including published data from refs. [7, 24]). Mass spectrometry (MS) confirmation of aptamers is from ref. [7]. b) Comparisons between SOMAscan 1.1 k and Olink panels in planned myocardial infarction patients. pQTL data is from previously published SOMAscan meta-analyses.[7,25] Most Olink analytes are not available in currently published genetic analyses.

eight proteins compared, three not assessed due to >20% of samples below the LLOD).

Finally, we assessed correlations between the SOMAscan platform (1.1 k version) and overlapping assays from Olink in a small cohort of myocardial infarction patients (ten patients, up to 48 samples due to multiple time points per patient) (**Table 2**; Table S9, Supporting Information); the median r_s of 0.36 was similar to the comparisons between SOMAscan and conventional immunoassays. Similar enrichment of mass spectrometry confirmed aptamers[7] in the high correlation group was observed, as well as enrichment for proteins with a SOMAscan *cis* pQTL (Figure 2b).

3.3. Summary

Correlation between immunoassays and SOMAscan varied widely by protein. Approximately 27% of the proteins, for example, C-reactive protein (CRP), were well correlated in both studies ($r_s \geq 0.7$) for the RBM platform. Similarly, a CRP immunoassay compared to SOMAscan in MESA ($r_s = 0.96$, $n = 976$) (Table S4, Supporting Information) and a QBR CRP immunoassay to SOMAscan in COPDGene ($r_s = 0.94$, $n = 1096$) (Table S6, Supporting Information) had high correlations, suggesting good consistency across cohorts and platforms for this biomarker. However, other target proteins measured in both COPDGene and SPIROMICS were essentially uncorrelated between SOMAscan and the RBM platform ($r_s < 0.3$, representing 20% of all compared proteins), for example, tumor necrosis factor receptor superfamily member 11B (TNFRSF11B), C—C motif chemokine 24 (CCL24), stromelysin-1 (MMP3), vascular endothelial growth factor A (VEGFA), pulmonary surfactant-associated protein D (SFTPD), C—C motif chemokine 13 (CCL13), and platelet endothelial cell adhesion molecule (PECAM1). This wide range in correlation coefficients is broadly consistent comparing other platforms, such as MSD and Olink, with SOMAscan. For example, for Olink, correlation coefficients range from −0.58 to 0.93 for the 425 tested proteins, with 13% of proteins with $r_s \geq 0.7$ (well correlated) and 42% with an $r_s < 0.3$ (poorly correlated). For proteins present in the reference HUPO plasma dataset, protein abundance in reference plasma samples was not

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 2. Distribution of Spearman correlation coefficients for comparison of Olink analytes which overlap with SOMAscan 1.1 k array in a small cohort of hypertrophic cardiomyopathy patients undergoing septal ablation ($n = 48$). Full results in Table S9, Supporting Information. r_s, estimated Spearman's rho.

	Correlation range	Count	Example proteins
Low correlation	<0.10	110	Brain-derived neurotrophic factor, C—C motif chemokine 24, Cathepsin D, Interleukin-27
	$0.10\leq r_s <0.20$	40	Cadherin-3, Cystatin-C, Ficolin-2, Tissue factor
	$0.20\leq r_s <0.30$	29	CD40 ligand, L-Selectin, Stromelysin-1, Tumor necrosis factor receptor superfamily member 19
Moderate correlation	$0.30\leq r_s <0.40$	55	Angiogenin, Eotaxin, Granzyme B, Wnt inhibitory factor 1
	$0.40\leq r_s <0.50$	40	Endostatin, Resistin, Tissue-type plasminogen activator, von Willebrand factor
	$0.50\leq r_s <0.60$	56	Cathepsin S, Glucagon, Myoglobin, P-Selectin
	$0.60\leq r_s <0.70$	40	C—C motif chemokine 22, Granulysin, Leptin, Vitamin K-dependent protein C
Well-correlated	$0.70\leq r_s <0.8$	31	Angiopoietin-1, E-Selectin, Myeloperoxidase, Platelet-derived growth factor subunit B
	$0.8\leq r_s <0.9$	20	Angiopoietin-2, Interleukin-6, Renin, Tissue factor pathway inhibitor
	$0.9\leq r_s <1.0$	6	C—C motif chemokine 21, Insulin-like growth factor-binding protein 1, Interleukin-1 receptor-like 1, Spondin-1

correlated with reported Spearman correlation coefficients across platforms in our analyses ($r_s = 0.082$), suggesting low and high abundance proteins were equally likely to have poor correlation across platforms (Table S10, Supporting Information). However, we do note that for the 43 proteins included in both our Myriad RBM/SOMAscan comparisons in COPDGene and SPIROMICS (Table 1) and Olink/SOMAscan comparisons in the planned MI cohort (Table 2; Table S9, Supporting Information), there was a strong relationship between the correlation coefficients ($r_s = 0.63$), showing some consistency for SOMAscan comparisons with two different antibody-based platforms (Figure S2, Supporting Information).

4. Discussion

In four cohorts (SPIROMICS, COPDGene, MESA, and a small cohort of planned myocardial infarction patients) comprising both adult participants in good health and those with cardiovascular and smoking-related diseases, we identified a wide range of correlations between SOMAscan aptamer results and data from multiple antibody-based assays. Assay correlations ranged from very high (e.g., CRP) to non-existent (SFTPD, PECAM1). We found that *cis* pQTLs and mass spectrometry confirmation of aptamers are more often observed with well correlated assay pairs than poorly correlated assay pairs, suggesting that these are valuable measures of aptamer/antibody specificity for target proteins. Availability of many of the same assays in COPDGene and SPIROMICS demonstrated the similarity in assay correlations across two cohorts. Epitope availability, cross-reactivity, and negative cooperative binding could all contribute to lack of concordance between methods. Some non-specificity has previously been reported for the SOMAscan platform, for example, the known binding of an aptamer for GDF11 to GDF8[8] (though this has been addressed on the current version of the platform[26]). Specificity problems and lack of correlation between immunoassay platforms have also been reported, however.[10,11] A careful examination of platform concordance is essential for proteomics analyses, and the data presented here should be useful in the design of studies which combine proteomics data across platforms.

Our results support the general feasibility of meta-analysis with immunoassays for some proteins assessed by the novel SOMAscan platform, while highlighting a subset of proteins that may be problematic to compare across platforms. Many analyses have been performed to determine the most likely protein biomarkers for a disease or trait using large-scale proteomics platforms, for example,[27,28,29] but few include replication analyses across many cohorts, with a few recent exceptions (such as ref. [28]), or effectively utilize multiple proteomics platforms. Although pQTLs often have larger effect sizes than are seen in other complex trait genetics analyses,[24,30] increasing power through meta-analysis will also be essential to pQTL discovery, particularly for *trans* variants with more modest effect sizes or for analyses of gene × gene and gene × environment interactions. It is important to utilize alternate protein assessment platforms to validate pQTLs (as done for a subset in ref. [24]) or disease association results. Despite differing scales (as both Olink and SOMAscan provide only relative, not absolute, quantification), meta-analysis with traditional ELISAs or other technologies should be appropriate for well-correlated assays but would be meaningless for essentially uncorrelated assays.

We argue that the presence or absence of *cis* pQTLs and reproducibility of pQTLs across platforms may provide clues to assay specificity. The high reproducibility (74.6% replicated) of previously published non-aptamer identified pQTL associations in the recent AGES study[7] points to reasonable consistency of aptamer and non-aptamer methods for many proteins, while not excluding problematic quantification for a subset of assays. Sun et al. also found that, of 74 tested *cis* pQTLs from SOMAscan, 60 replicated with the Olink platform, again pointing to reasonable platform concordance.[24] Preliminary results released by Olink suggest that 90% of biomarkers on their CVD I-panel (80 of 90 tested proteins) have a genome-wide significant *cis* pQTL in a sample of $n = 22\ 000$, suggesting high specificity for their paired antibody-based system. However, further follow-up of an increased number of Olink assays using both pQTLs and other analysis techniques is needed in future work,[31] as are better powered meta-analyses of SOMAscan assays to definitively confirm the presence or absence of *cis* pQTLs. Our analyses of *cis* pQTLs in SPIROMICS and COPDGene suggest that discordant *cis* pQTLs may be helpful for determining which assay is

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

likely more specific for the target protein for poorly correlated assay pairs. For example, SFTPD has a *cis* pQTL in Myriad RBM based analyses from COPDGene and SPIROMICS (lead pQTL rs2146192),[16] which coincided with previous associations for COPD phenotypes. However, this protein had no reported *cis* pQTLs in SOMAscan pQTL results from COPDGene and SPIROMICS or in two other previous studies.[7,24] One potential reason for this lack of replication is differences in assay specificity and performance, with the antibody for SFTPD for Myriad RBM likely having higher affinity or less off-target binding than the SOMAscan aptamer for this protein, allowing detection of this *cis* pQTL. Conversely, for analytes with a strong *cis* pQTL for SOMAscan data but not for antibody-based assays, there may be an issue with the specificity of antibody-based assays. However, concordant *cis* pQTLs do not ensure that assays will be well correlated. For poorly correlated analyte C—C motif chemokine 8, a *cis* pQTL signal is found both for SOMAscan data and Myriad RBM data, suggesting at least some binding to the target protein for both these assays, but likely with variable specificity or binding properties. Finally, for *cis* pQTL missense variants, such as lead variant rs5498 with SOMAscan measured ICAM1 (Table S1, Supporting Information), it is important to note that differences in antibody or aptamer binding to different isoforms may drive missense variant *cis* pQTLs, as opposed to true differences in protein levels. For example, differences in antibody binding causing spurious pQTLs was found for VDBP (not assessed here) in prior work using the Myriad RBM platform[16] and for other ELISAs in past efforts (such as ref. [32]). Missense variants impacted levels of >32% of proteins assessed in INTERVAL, in some cases likely due to artefactual differences in aptamer binding unrelated to biologically relevant changes in protein abundance.[24] Future studies should also consider common missense variants which may cause assay interference and decrease biomarker correlations in only some ancestry groups, particularly those common only in non-European ancestry populations (and therefore not assessed in existing pQTL studies[7,24]). In cases where missense variants do impact aptamer or antibody binding, stratifying by genotype may improve correlation across assays.[16] While we did not see a systematic pattern for presence/absence of common missense variants among well- and poorly-correlated proteins in our analysis (Table S3, Supporting Information), careful consideration of missense variants which may interfere with appropriate antibody/aptamer binding to target epitopes is necessary. Along with these known concerns for coding *cis* pQTLs, *cis* genetic variants that alter protein oligomerization and posttranslational modifications through their regulatory roles could also impact antibody or aptamer-based quantification without truly influencing protein abundance.

Along with *cis* pQTLs, we also explored other potential explanatory factors for observed correlation coefficients for assessed proteins. As explored in Table S3, Supporting Information, for proteins from SPIROMICS and COPDGene compared across the Myriad RBM and SOMAscan platforms, we did not see systematic differences in post-translational modifications or number of isoforms between highly correlated and poorly correlated (or pQTL discordant) proteins. However, more high correlation proteins versus low correlation proteins had aptamers confirmed by mass spectrometry,[7] for both Myriad RBM and Olink comparisons (Figure 2). Other studies prior to Emilsson et al. also validated SOMAscan aptamers with mass spectrometry, providing important validation of cardiovascular disease related findings for proteins such as epidermal growth factor receptor[3] and adenylosuccinate synthase 1,[33] among others; data from these studies could additionally provide validation for some aptamers of interest. Cumulatively these studies provide convincing evidence of the utility of mass spectrometry-based aptamer confirmation. Cross-reactivity for aptamers or antibodies might also be hypothesized to lead to low correlation between assays. Recent work by Sun et al.[24] extensively tested cross-reactivity of 920 SOMAmers against homologous proteins (at least 40% sequence homology), and found that 126 SOMAmers (14%) showed comparable binding with a homologous protein. However, nearly half of these were binding to alternative forms of the same protein; the likelihood of cross-reactivity also increased with increasing amino acid homology (the median was 70% for those with comparable binding). While we observed few systematic differences in cross-reactivity between low and high correlation assays, cross-reactivity could be a cause for low correlations for a few of the assays examined here. For example, the D-dimer aptamer had comparable binding to fibrinogen and fibrinogen γ chain (but no appreciable binding to fibrinogen β chain) in the analyses by Sun et al., suggesting a cause for the modest observed correlation for D-dimer immunoassays with SOMAscan ($r_s = 0.14$ in MESA).

A central strength of our analysis is concurrent assessment of correlations with an immunoassay platform in four independent cohorts, two of which (COPDGene and SPIROMICS) have very similar proteomics datasets. This helped us obtain a more consistent picture of the true correlations between platforms. Our study has important limitations as well. As is standard for most cohort studies, data sharing is based upon participant informed consent; thus, not all underlying data can be freely shared. All data are available through data access procedures as described in the Associated Data. There are many novel options for proteomics analysis (such as high throughput mass spectrometry pipelines) which are not considered here. We also cannot comment on which assays (SOMAscan, MSD, QBR, etc.) most accurately reflect the true levels of protein biomarkers without comparing them to fully validated immunoassays that have been vetted for specificity, linearity, and possible interference (which for the majority of measured analytes do not exist). In the COPDGene cohort, data for four well-vetted immunoassays in clinical use were available; concordance with SOMAscan data was $\approx$0.6 or higher for three of these (but not for alpha-1-antitrypsin, which was also poorly correlated with RBM values) (Figure S3, Supporting Information). However, in most cases there is no clinical assay available, increasing the utility of data on mass spectrometry confirmation or on *cis* pQTLs. Presence of a *cis* pQTLs for both platforms does not guarantee high correlation (as evidenced by similar median correlation for proteins with/without a *cis* pQTL for Myriad RBM comparisons for example), but presence of a *cis* pQTL suggests an antibody/aptamer has at least some binding with its target protein. Discordant *cis* pQTLs can be helpful in interpreting low correlation between assays, with the assay with a *cis* pQTL likely more accurately assessing abundance of the target protein. We do note as a limitation of our analyses the different genotyping platforms and use of genotyped variants only in comparing pQTL analyses between COPDGene and SPIROMICS, but our purpose was simply to evaluate the

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

presence or absence of *cis* pQTLs. Comparison of serum and plasma protein quantification is also an issue for some assays (notably some of the SPIROMICS Myriad RBM/SOMAscan comparisons in Table 1, as annotated in Table S2, Supporting Information), as there may be differences in protein stability, the formation of aggregates, and other attributes for some proteins between serum and plasma (though previous work[18] from SPIROMICS has shown high correspondence between serum and plasma results for the Myriad RBM platform). Our Olink/SOMAscan comparison dataset is small ($n = 48$) and thus could not be used for *cis* pQTL analysis. It also uses an older (1.1 k) version of the SOMAscan platform (a small subset of the aptamers have been retired in more recent versions).

Our results provide an important starting point for proteomics investigators, but differences between platforms still need to be more fully explored. High-throughput proteomics technologies such as the SOMAscan platform[12] can assess a very large number of proteins in a single assay at relatively low per analyte cost and low sample volume. However, it is evident that some proteins examined do not correlate well with antibody-based assays, including novel Olink proximity extension assays, potentially due to lack of specificity for some aptamers or antibodies or differences in signal to noise ratios. Different assays may have varying benefits and drawbacks. For example, in Olink proximity extension assays, binding of a pair of antibodies to each protein[12] might be hypothesized to increase specificity. However, the real-time PCR quantification may differ in signal to noise ratio compared to other quantification approaches. Additional validation for top results with any novel discovery platform (using mass spectrometry, ELISAs, etc.) is important, especially if there is evidence of lack of correspondence between platforms and a lack of *cis* pQTL evidence for the target protein with the novel assay. The use of relative quantification for SOMAscan and Olink also leads to some uncertainty about which proteins may indeed be below reasonable detection limits in a large portion of individuals, with again orthogonal methods required for some applications. However, we do note that protein abundance in reference data from HUPO was not significantly correlated with Spearman correlation coefficients across platforms, so low protein abundance is unlikely to be the main reason for discrepant quantification. Still, we feel a comprehensive study of many available platforms (including mass spectrometry-based platforms) in a reasonably sized set of overlapping samples ($\approx$100), with comparison to serum and plasma reference standards, would be a precious asset for future research. Complete understanding of the comparability of results from aptamer technology to other protein biomarker or proteomics platforms is still lacking and is vital for interpretation of research findings.

To conclude, more analysis is needed to assess the potential to integrate data across proteomics platforms fully. Olink and SOMAscan do not provide exact quantification and are not designed to correspond perfectly to exact quantification methods. However, sample ranking (as assessed here using Spearman correlation coefficients) should be quite similar if both the novel assay (Olink or SOMAscan) and existing immunoassays perform well. Ideally, all novel assays should be confirmed by multiple orthogonal methods (e.g., both MRM and well-validated immunoassays) in a well-powered sample, as well as using well-characterized reference samples in relevant tissue types (such as serum, plasma, and urine). Investigators should be careful to treat discovery platforms like Olink and SOMAscan appropriately, confirming key results with additional methods, and using comparisons to reference standards to obtain exact protein quantification. However, the correlation information presented here provides a starting point for evaluating the comparability of antibody- and aptamer-based protein measures, similar to previous efforts publicly compiling the coefficients of variation and stability measurements for SOMAmers.[34] As new multi-cohort meta-analyses and replication efforts are initiated, our analysis suggests that investigators must be aware of differences in biomarkers obtained from different platforms and can use metrics such as the presence of *cis* pQTLs or orthogonal assays like mass spectrometry to infer the specificity of different proteomics platforms.

5. Associated Data

Data is available through dbGaP or other secure access mechanisms to approved researchers. Full data is available from MESA (https://www.mesa-nhlbi.org/), SPIROMICS (https://www.spiromics.org/spiromics/), and COPDGene (http://www.copdgene.org/) by approved manuscript proposal and data use agreement at the respective study websites. The Multi-Ethnic Study of Atherosclerosis (MESA) proteomics data is available at dbGaP phs000209; SOMAscan data will soon be posted to phs001416. Genetic Epidemiology of COPD (COPDGene) data is available at dbGaP phs000179. Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS) data (genetic data only) is available at phs001119. Data from the small cohort of patients undergoing septal ablation for hypertrophic cardiomyopathy (inducing planned myocardial infarction) is included as a supplemental data object.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

SPIROMICS

The authors thank the SPIROMICS participants and participating physicians, investigators, and staff for making this research possible. More information about the study and how to access SPIROMICS data is at www.spiromics.org. The authors would like to acknowledge the following current and former investigators of the SPIROMICS sites and reading centers: Neil E. Alexis, Ph.D.; Wayne H. Anderson, Ph.D.; R. Graham Barr, MD, DrPH; Eugene R. Bleecker, MD; Richard C. Boucher, MD; Russell P. Bowler, MD, Ph.D.; Elizabeth E. Carretta, MPH; Stephanie A. Christenson, MD; Alejandro P. Comellas, MD; Christopher B. Cooper, MD, Ph.D.; David J. Couper, Ph.D.; Gerard J. Criner, MD; Ronald G. Crystal, MD; Jeffrey L. Curtis, MD; Claire M. Doerschuk, MD; Mark T. Dransfield, MD; Christine M. Freeman, Ph.D.; MeiLan K. Han, MD, MS; Nadia N. Hansel, MD, MPH; Annette T. Hastie, Ph.D.; Eric A. Hoffman, Ph.D.; Robert J. Kaner, MD; Richard E. Kanner, MD; Eric C. Kleerup, MD; Jerry A. Krishnan, MD, Ph.D.; Lisa M. LaVange, Ph.D.; Stephen C. Lazarus, MD; Fernando J. Martinez, MD, MS; Deborah A. Meyers, Ph.D.; John D. Newell Jr, MD; Elizabeth C. Oelsner, MD, MPH; Wanda K. O'Neal, Ph.D.; Robert Paine, III, MD; Nirupama Putcha, MD, MHS; Stephen I. Rennard,

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

MD; Donald P. Tashkin, MD; Mary Beth Scholand, MD; J. Michael Wells, MD; Robert A. Wise, MD; and Prescott G. Woodruff, MD, MPH. The project officers from the Lung Division of the National Heart, Lung, and Blood Institute were Lisa Postow, Ph.D., and Thomas Croxton, Ph.D., MD. SPIROMICS was supported by contracts from the NIH/NHLBI (HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C, HHSN268200900019C, HHSN268200900020C), which were supplemented by contributions made through the Foundation for the NIH from AstraZeneca, Bellerophon Therapeutics, Boehringer-Ingelheim Pharmaceuticals, Inc., Chiesi Farmaceutici SpA, Forest Research Institute, Inc., GSK, Grifols Therapeutics, Inc., Ikaria, Inc., Nycomed GmbH., Takeda Pharmaceutical Company., Novartis Pharmaceuticals Corporation., Regeneron Pharmaceuticals, Inc., and Sanofi.

MESA

MESA and the MESA SHARe projects were conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA was provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Also supported in part by the National Center for Advancing Translational Sciences, CTSI grant ULTR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.

Whole genome sequencing (WGS) for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the National Heart, Lung and Blood Institute (NHLBI). WGS for "NHLBI TOPMed: Multi-Ethnic Study of Atherosclerosis (MESA)" (phs001416.v1.p1) was performed at the Broad Institute of MIT and Harvard (3U54HG003067-13S1). Centralized read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN268201800002I). Phenotype harmonization, data management, sample-identity QC, and general study coordination were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1; contract HHSN268201800001I).

The authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

COPDGene

The project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the National Heart, Lung, and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung, and Blood Institute or the National Institutes of Health. The COPDGene project was also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens, and Sunovion.

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.

L.M.R. was funded by T32 HL129982. W.W.L. was funded by T32 HL007749. J.L.C. was funded by Merit Review I01 CX000911 from the Department of Veterans Affairs. A.P.R. was funded by R01-HL132947. R.E.G. was supported by R01HL132320 and R01HL133870. C.M.F. was funded by Merit Review I01 CX001553 from the Department of Veterans Affairs.

Conflict of Interest

A.P.C. is a consultant for VIDA Diagnostics. J.L.C. has funding from MedImmune Corporation, Ltd., for a project outside the scope for this work. C.E. is an employee of AstraZeneca and owns stock options.

Keywords

antibody microarrays, biomarkers, multiplexings

Received: August 23, 2019
Revised: April 30, 2020

[1] A. A. Ellington, I. J. Kullo, K. R. Bailey, G. G. Klee, *Clin. Chem.* **2010**, *56*, 186.
[2] J. G. Smith, R. E. Gerszten, *Circulation* **2017**, *135*, 1651.
[3] D. Ngo, S. Sinha, D. Shen, E. W. Kuhn, M. J. Keyes, X. Shi, M. D. Benson, J. F. O'Sullivan, H. Keshishian, L. A. Farrell, M. A. Fifer, R. S. Vasan, M. S. Sabatine, M. G. Larson, S. A. Carr, T. J. Wang, R. E. Gerszten, *Circulation* **2016**, *134*, 270.
[4] L. Gold, D. Ayers, J. Bertino, C. Bock, A. Bock, E. N. Brody, J. Carter, A. B. Dalby, B. E. Eaton, T. Fitzwater, D. Flather, A. Forbes, T. Foreman, C. Fowler, B. Gawande, M. Goss, M. Gunn, S. Gupta, D. Halladay, J. Heil, J. Heilig, B. Hicke, G. Husar, N. Janjic, T. Jarvis, S. Jennings, E. Katilius, T. R. Keeney, N. Kim, T. H. Koch, et al. *PLoS One* **2010**, *5*, e15004.
[5] A. Tin, B. Yu, J. Ma, K. Masushita, N. Daya, R. C. Hoogeveen, C. M. Ballantyne, D. Couper, C. M. Rebholz, M. E. Grams, A. Alonso, T. Mosley, G. Heiss, P. Ganz, E. Selvin, E. Boerwinkle, J. Coresh, *J. Appl. Lab. Med.* **2019**, *4*, 30.
[6] C. H. Kim, S. S. Tworoger, M. J. Stampfer, S. T. Dillon, X. Gu, S. J. Sawyer, A. T. Chan, T. A. Libermann, A. H. Eliassen, *Sci. Rep.* **2018**, *8*, 8382.
[7] V. Emilsson, M. Ilkov, J. R. Lamb, N. Finkel, E. F. Gudmundsson, R. Pitts, H. Hoover, V. Gudmundsdottir, S. R. Horman, T. Aspelund, L. Shu, V. Trifonov, S. Sigurdsson, A. Manolescu, J. Zhu, Ö. Olafsson, J. Jakobsdottir, S. A. Lesley, J. To, J. Zhang, T. B. Harris, L. J. Launer, B. Zhang, G. Eiriksdottir, X. Yang, A. P. Orth, L. L. Jennings, V. Gudnason, *Science* **2018**, *361*, 769.
[8] R. G. Walker, T. Poggioli, L. Katsimpardi, S. M. Buchanan, J. Oh, S. Wattrus, B. Heidecker, Y. W. Fong, L. L. Rubin, P. Ganz, T. B. Thompson, A. J. Wagers, R. T. Lee, *Circ. Res.* **2016**, *118*, 1125.
[9] L. Christiansson, S. Mustjoki, B. Simonsson, U. Olsson-Stromberg, A. S. I. Loskog, S. M. Mangsbo, *EuPA Open Proteomics* **2014**, *3*, 37.
[10] SomaLogic Inc., http://somalogic.com/wp-content/uploads/2017/07/Correlation-of-SOMAmer%C2%AE-reagents-in-the-SOMAscan%C2%AE-assay-and-commercially-available-immunoassays-SS-501-051916-1-1.pdf (accessed: **2019**).
[11] A. K. Chaturvedi, T. J. Kemp, R. M. Pfeiffer, A. Biancotto, M. Williams, S. Munuo, M. P. Purdue, A. W. Hsing, L. Pinto, J. P. McCoy, A. Hildesheim, *Cancer Epidemiol. Biomark. Prev.* **2011**, *20*, 1902.
[12] E. Assarsson, M. Lundberg, G. Holmquist, J. Björkesten, S. Bucht Thorsen, D. Ekman, A. Eriksson, E. R. Dickens, S. Ohlsson, G. Edfeldt, A.-C. Andersson, P. Lindstedt, J. Stenvang, M. Gullberg, S. Fredriksson, *PLoS One* **2014**, *9*, e95192.
[13] D. Couper, L. M. LaVange, M. Han, R. G. Barr, E. Bleecker, E. A. Hoffman, R. Kanner, E. Kleerup, F. J. Martinez, P. G. Woodruff, S. Rennard, SPIROMICS Research Group, *Thorax* **2014**, *69*, 491.
[14] E. A. Regan, J. E. Hokanson, J. R. Murphy, B. Make, D. A. Lynch, T. H. Beaty, D. Curran-Everett, E. K. Silverman, J. D. Crapo, *Int. J. Chron. Obstruct. Pulmon. Dis.* **2010**, *7*, 32.
[15] D. E. Bild, D. E. Bild, R. Detrano, D. Peterson, A. Guerci, K. Liu, E. Shahar, P. Ouyang, S. Jackson, M. F. Saad, *Circulation* **2005**, *111*, 1313.
[16] W. Sun, W. Sun, K. Kechris, S. Jacobson, M. B. Drummond, G. A. Hawkins, J. Yang, T.-h. Chen, P. M. Quibrera, Common Genetic Polymorphisms Influence Blood Biomarker Measurements in COPD

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Anderson, R. G. Barr, P. V. Basta, E. R. Bleecker, T. Beaty, R. Casaburi, P. Castaldi, M. H. Cho, A. Comellas, J. D. Crapo, G. Criner, D. Demeo, S. A. Christenson, D. J. Couper, J. L. Curtis, C. M. Doerschuk, C. M. Freeman, N. A. Gouskova, M. K. Han, N. A. Hanania, N. N. Hansel, et al. *PLoS Genet.* **2016**, *12*, e1006011.

[17] J. D. Keene, S. Jacobson, K. Kechris, G. L. Kinney, M. G. Foreman, C. M. Doerschuk, B. J. Make, J. L. Curtis, S. I. Rennard, R. G. Barr, E. R. Bleecker, R. E. Kanner, E. C. Kleerup, N. N. Hansel, P. G. Woodruff, M. L. K. Han, R. Paine III, F. J. Martinez, R. P. Bowler, W. K. O'Neal, COPDGene and SPIROMICS Investigators, *Am J. Respir. Crit. Care Med.* **2017**, *195*, 473.

[18] W. K. O'Neal, W. Anderson, P. V. Basta, E. E. Carretta, C. M. Doerschuk, R. G. Barr, E. R. Bleecker, S. A. Christenson, J. L. Curtis, M. K. Han, N. N. Hansel, R. E. Kanner, E. C. Kleerup, F. J. Martinez, B. E. Miller, S. P. Peters, S. I. Rennard, M. B. Scholand, R. Tal-Singer, P. G. Woodruff, D. J. Couper, S. M. Davis, SPIROMICS Investigators, *J. Transl. Med.* **2014**, *12*, 9.

[19] E. Bradford, S. Jacobson, J. Varasteh, A. P. Comellas, P. Woodruff, W. O'Neal, D. L. DeMeo, X. Li, V. Kim, M. Cho, P. J. Castaldi, C. Hersh, E. K. Silverman, J. D. Crapo, K. Kechris, R. P. Bowler, *Respir. Res.* **2017**, *18*, 180.

[20] a) J. Crowther, *The ELISA Guidebook, 2nd ed.*, Humana Press, Totowa, NJ **2009**; b) CLSI.org/standards; c) J. W. Lee, V. Devanarayan, Y. C. Barett, R. Weiner, J. Allinson, S. Fountain, S. Keller, I. Weinryb, M. Green, L. Duan, J. A. Rogers, R. Millham, P. J. O'Brien, J. Sailstad, M. Khan, C. Ray, J. A. Wagner, *Pharm. Res.* **2006**, *23*, 312.

[21] B. R. Celli, N. Locantore, J. Yates, R. Tal-Singer, B. E. Miller, P. Bakke, P. Calverley, H. Coxson, C. Crim, L. D. Edwards, D. A. Lomas, A. Duvoix, W. MacNee, S. Rennard, E. Silverman, J. Vestbo, E. Wouters, A. Agustí, ECLIPSE Investigators, *Am J. Respir. Crit. Care Med.* **2012**, *185*, 1065.

[22] D. E. Bild, D. A. Bluemke, G. L. Burke, R. Detrano, A. V. Diez Roux, A. R. Folsom, P. Greenland, D. R. Jacobs Jr., R. Kronmal, K. Liu, J. C. Nelson, D. O'Leary, M. F. Saad, S. Shea, M. Szklo, R. P. Tracy, *Am. J. Epidemiol.* **2002**, *156*, 871.

[23] N. S. Jenny, N. C. Olson, M. A. Allison, D. E. Rifkin, L. B. Daniels, I. H. de Boer, C. L. Wassel, R. P. Tracy, *Glob. Heart* **2016**, *11*, 327.

[24] B. B. Sun, J. C. Maranville, J. E. Peters, D. Stacey, J. R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M. A. Kamat, B. P. Prins, S. K. Wilcox, E. S. Zimmerman, A. Chi, N. Bansal, S. L. Spain, A. M. Wood, N. W. Morrell, J. R. Bradley, N. Janjic, D. J. Roberts, W. H. Ouwehand, J. A. Todd, N. Soranzo, K. Suhre, D. S. Paul, C. S. Fox, R. M. Plenge, et al. *Nature* **2018**, *558*, 73.

[25] Q. Fu, J. Zhu, J. E. van Eyk, *Clin. Chem.* **2010**, *56*, 314.

[26] SomaLogic Inc., http://www.somalogic.com/wp-content/uploads/2017/01/SSM-067-Rev-1-Technical-Note-Characterization-of-SOMAmer-Reagents-Binding-Specificity-in-the-SOMAscan-1.3k-Assay.pdf (accessed: **2019**).

[27] A. Penn-Nicholson, T. Hraha, E. G. Thompson, D. Sterling, S. K. Mbandi, K. M. Wall, M. Fisher, S. Suliman, S. Shankar, W. A. Hanekom, N. Janjic, M. Hatherill, S. H. E. Kaufmann, J. Sutherland, G. Walzl, M. A. De Groote, U. Ochsner, D. E. Zak, T. J. Scriba, ACS and GC6–74 cohort study groups, *PLoS Med.* **2019**, *16*, e1002781.

[28] S. A. Williams, M. Kivimaki, C. Langenberg, A. D. Hingorani, J. P. Casas, C. Bouchard, C. Jonasson, M. A. Sarzynski, M. J. Shipley, L. Alexander, J. Ash, T. Bauer, J. Chadwick, G. Datta, R. K. DeLisle, Y. Hagar, M. Hinterberg, R. Ostroff, S. Weiss, P Ganz, N. J. Wareham, *Nat. Med.* **2019**, *25*(12), 1851.

[29] a) A. Christensson, J. A. Ash, R. K. DeLisle, F. W. Gaspar, R. Ostroff, A. Grubb, V. Lindström, L. Bruun, S. A. Williams, *Proteomics - Clin. Appl.* **2018**, *12*, 1700067; b) W. Chua, Y. Purmah, V. R. Cardoso, G. V. Gkoutos, S. P. Tull, G. Neculau, M. R. Thomas, D. Kotecha, G. Y. H. Lip, P. Kirchhof, L. Fabritz, *Eur. Heart J.* **2019**, *40*, 1268; c) T. Feldreich, C. Nowak, T. Fall, A. C. Carlsson, J.-J. Carrero, J. Ripsweden, A. R. Qureshi, O. Heimburger, P. Barany, P. Stenvinkel, N. Vuilleumier, P. A. Kalra, D. Green, J. Ärnlöv, *J. nephrol.* **2019**, *32*, 111.

[30] K. Suhre, M. Arnold, A. M. Bhagwat, R. J. Cotton, R. Engelke, J. Raffler, H. Sarwath, G. Thareja, A. Wahl, R. K. DeLisle, L. Gold, M. Pezer, G. Lauc, M. A. El-Din Selim, D. O. Mook-Kanamori, E. K. Al-Dous, Y. A. Mohamoud, J. Malek, K. Strauch, H. Grallert, A. Peters, G. Kastenmüller, C. Gieger, J. Graumann, *Nat. Commun.* **2017**, *8*, 14357.

[31] Olink Proteomics, https://www.olink.com/content/uploads/2019/06/Olink-technical-comparisons-and-orthogonal-validation-v1.0.pdf (accessed: **2019**).

[32] a) D. C. Croteau-Chonka, Y. Wu, Y. Li, M. P. Fogarty, L. A. Lange, C. W. Kuzawa, T. W. McDade, J. B. Borja, J. Luo, O. AbdelBaky, T. P. Combs, L. S. Adair, E. M. Lange, K. L. Mohlke, *Hum. Mol. Genet.* **2012**, *21*, 463; b) L. M. Polfus, L. M. Raffield, M. M. Wheeler, R. P. Tracy, L. A. Lange, G. Lettre, A. Miller, A. Correa, R. P. Bowler, J. C. Bis, S. Salimi, N. S. Jenny, N. Pankratz, B. Wang, M. H. Preuss, L. Zhou, A. Moscati, G. N. Nadkarni, R. J. F. Loos, X. Zhong, B. Li, J. M. Johnsen, D. A. Nickerson, A. P. Reiner, P. L. Auer, *Hum. Mol. Genet.* **2019**, *28*(3), 515.

[33] J. Jacob, D. Ngo, N. Finkel, R. Pitts, S. Gleim, M. D. Benson, M. J. Keyes, L. A. Farrell, T. Morgan, L. L. Jennings, R. E. Gerszten, *Circulation* **2018**, *137*, 1270.

[34] J. Candia, F. Cheung, Y. Kotliarov, G. Fantoni, B. Sellers, T. Griesman, J. Huang, S. Stuccio, A. Zingone, B. M. Ryan, J. S. Tsang, A. Biancotto, *Sci. Rep.* **2017**, *7*, 14248.

© 2020 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim