

GENETICS

Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods

Daniel H. Katz¹, Jeremy M. Robbins¹, Shuliang Deng¹, Usman A. Tahir¹, Alexander G. Bick², Akhil Pampana², Zhi Yu², Debby Ngo¹, Mark D. Benson¹, Zsu-Zsu Chen¹, Daniel E. Cruz¹, Dongxiao Shen¹, Yan Gao³, Claude Bouchard⁴, Mark A. Sarzynski⁵, Adolfo Correa³, Pradeep Natarajan^{2,6,7}, James G. Wilson¹, Robert E. Gerszten^{1,2*}

Copyright © 2022
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

High-throughput proteomic profiling using antibody or aptamer-based affinity reagents is used increasingly in human studies. However, direct analyses to address the relative strengths and weaknesses of these platforms are lacking. We assessed findings from the SomaScan1.3K ($N = 1301$ reagents), the SomaScan5K platform ($N = 4979$ reagents), and the Olink Explore ($N = 1472$ reagents) profiling techniques in 568 adults from the Jackson Heart Study and 219 participants in the HERITAGE Family Study across four performance domains: precision, accuracy, analytic breadth, and phenotypic associations leveraging detailed clinical phenotyping and genetic data. Across these studies, we show evidence supporting more reliable protein target specificity and a higher number of phenotypic associations for the Olink platform, while the Soma platforms benefit from greater measurement precision and analytic breadth across the proteome.

INTRODUCTION

The advent of high-throughput proteomic profiling has greatly enhanced our ability to investigate disease, as proteins are not only mediators of disease but also clinical biomarkers used to diagnose and guide treatment (e.g., B-type natriuretic peptide and troponin) (1–3). New technologies based on affinity reagents for capture and detection of specific proteins are receiving increasing attention in plasma proteomics due to their performance characteristics, cost, and usability. In particular, platforms using paired, nucleotide-labeled antibody probes (Olink) and single-strand DNA aptamer reagents with slow off-rate kinetics (SomaScan) can be automated for efficient multiplexing of thousands of proteins at high sample throughput (4–7). While these platforms have streamlined workflows as compared to liquid chromatography–mass spectrometry (LC-MS)–based methods, this comes at the cost of decreased specificity for molecular characterization (8). Proteomic profiling with these affinity platforms has already been performed in many cohort studies and clinical trials (1, 2, 9–15). As investigators begin to analyze findings from this body of established work and as more studies use these platforms, it is critical to understand the relative strengths and weaknesses of the available technologies. At the same time, independent verification of test performance relative to a gold standard for each of the thousands of proteins on these platforms is limited by cost and time. Alternatively, comparison of available platforms to one another offers an opportunity for high-throughput assessment. Earlier comparisons have been limited by sample size and the number of proteins measured at the time of comparison (16, 17) but did

suggest differences in platform characteristics and reproducibility. More recently, a larger effort to compare these platforms ($N = 485$ overlapping samples) demonstrated very poor correlation between a large number of reagents targeting the same protein (18). While some differences were explained by a variety of platform and protein factors, an assessment of comparative accuracy was absent. Further characterization of these platforms under direct comparison is needed.

The Jackson Heart Study (JHS) and HERITAGE Family Study are ideally suited for platform comparison. In addition to the many clinical traits ascertained in JHS, whole-genome sequencing (WGS) is available on a large subset of participants, permitting assessment of rare or ancestry-specific genetic variants with a variety of effects on circulating proteins. As a cohort of Black adults, there is greater genetic diversity owing to increased African ancestry (19, 20). Our group has previously described the genetic architecture of the circulating plasma proteome from JHS, leveraging the SomaScan1.3K (1301 aptamers, herein “Soma1.3K”) platform for discovery and the Olink Explore (1472 probes, herein “Olink”) platform for validation (21), expanding previous work in this area to a cohort with substantial African ancestry (13, 22–26). While this previous work described genetic associations with plasma protein levels as measured by these platforms, such data provide an opportunity to compare these platforms directly, instead of showcasing discovery and validation of gene-protein relationships. Specifically, matched data with profiling on both platforms can assess proteomic reagent specificity through the identification of variants near the target gene that affect measured levels of the target protein [termed *cis* protein quantitative trait loci (*cis* pQTLs)]. Thus, we profiled a subset of JHS participants ($N = 568$) using both aptamer-based and antibody-based methods and compared their performance with specific attention to precision, accuracy, analytic breadth across the proteome, and phenotypic associations. In the HERITAGE Family Study, we also compare the expanded SomaScan platform with ~5000 proteins (“Soma5K”) with the Olink platform in 219 individuals, leveraging the rigorous clinical phenotyping performed in that study.

¹Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. ²Broad Institute of Harvard and MIT, Cambridge, MA, USA. ³University of Mississippi Medical Center, Jackson, MS, USA. ⁴Human Genomic Laboratory, Pennington Biomedical Research Center, Baton Rouge, LA, USA. ⁵Department of Exercise Science, University of South Carolina, Columbia, SC, USA. ⁶Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. ⁷Department of Medicine, Harvard Medical School, Boston, MA, USA.

*Corresponding author. Email: rgerszte@bidmc.harvard.edu

RESULTS

Given the large amount of published proteomic profiling data on the original Soma1.3K platform, we began by comparing the Olink and Soma1.3K platforms in 568 individuals from JHS. The means \pm SD age of the cohort was 59 ± 12 years, 59% were female, mean body mass index (BMI) was 32 ± 8 kg/m², and mean estimated glomerular filtration rate (eGFR) was 83 ± 19 ml/min/1.73 m² (table S1). As some unique reagents on each platform detect the same protein or protein multimers, Olink profiling included 1472 unique reagents mapping to 1466 unique UniProt identifiers (IDs), while Soma1.3K profiling included 1301 unique reagents mapping to 1297 unique UniProt IDs. Platform reagents were matched on the basis of their target UniProt proteins, revealing 591 overlapping proteins mapping to 602 Soma1.3K aptamers and 597 Olink probes (Fig. 1 and table S2). This merging resulted in 616 unique Soma1.3K-Olink reagent pairs. The platforms were compared, with specific attention to the overlapping proteins, across four domains: precision, accuracy, analytical breadth, and phenotypic association.

Precision: Coefficients of variation

To assess the precision (i.e., reproducibility) of repeated protein measurements, the coefficient of variation (CV) for each reagent was measured using standard pooled plasma samples, which are included on each plate of each platform (different pooled plasma was used for each platform). Each platform measures 88 samples per plate and thus requires multiple plates for all samples to be run: Intra-assay CVs reflect precision within a given plate, while inter-assay CVs reflect precision between plates. As shown in Fig. 2, while most of the protein measurements on both platforms had inter-assay CVs below 20% (81% of the platform for Olink and 99% for Soma1.3K), Soma1.3K CVs were overall lower, whether comparing the full platform or overlapping proteins. Median intra-assay CVs were also lower on Soma1.3K (2%) compared to Olink (10%). As modeled in Fig. 2B, as CV increases, the required sample size to detect a given percent difference in mean protein levels between groups also increases.

Accuracy: Platform correlation and cis pQTLs

To understand the accuracy or specificity of a given protein measurement on either platform without using LC-MS—the gold standard—an “orthogonal” method can provide supportive evidence. For a small number of proteins, we can compare proteomic measurements to an established enzyme-linked immunosorbent assay (ELISA; fig. S1), but these are only available in a very limited number of cases. When

a protein is measured by both platforms, high correlation between the two suggests accuracy. The distribution of Spearman correlations for all overlapping protein targets is shown in Fig. 3A. *K*-means clustering supports three categories of paired reagents by the elbow method: high correlation ($N = 236$ reagent pairs), medium correlation ($N = 173$), and low correlation ($N = 207$).

While high correlation between the distinct assays suggests specificity, for proteins with weaker correlation or those that do not overlap between the platforms, we leveraged WGS data to help further inform each reagent’s specificity. If a protein’s measure by a given platform is associated with genetic variation near the cognate gene (i.e., cis pQTLs), this supports the protein assay’s accuracy. WGS was available in 489 of the JHS participants, and a validated computational pipeline for variant-protein association analysis—matching that previously described for pQTL identification—was used in the present analysis (21). Many of the proteins for which cis pQTLs were identified in the present analysis also have cis pQTLs that were previously identified in the literature in populations with predominantly European ancestry: 425 target proteins have previously known cis pQTLs, although these were not necessarily specific to Olink or Soma1.3K (table S2). Conversely, we identified previously unknown cis pQTLs for 373 protein targets by this method. Cis pQTLs could be identified for 595 of 1472 (40%) of the reagents on the Olink platform and 370 of 1301 (28%) of reagents on the Soma1.3K platform at a $P < 1 \times 10^{-5}$ (table S2). At this threshold, 164 of the reagent pairs show a cis pQTL for both reagents. If a genome-wide significance threshold is used (5×10^{-8}), cis pQTLs are observed for 368 of 1472 (25%) Olink reagents and 206 of 1301 (16%) Soma1.3K reagents. At this threshold, 98 reagent pairs have cis pQTLs for both reagents. Figure 3B shows the presence of pQTLs on either platform for matched reagents, relative to their correlation. While highly correlated reagents were both likely to demonstrate a cis pQTL, proteins with a lower correlation were more likely to show a cis pQTL for the Olink reagent only (Table 1).

Analytical breadth: Protein classifications and PCA

To capture the breadth of known proteomic biology captured by the Olink and Soma1.3K, Fig. 4 shows measured protein distribution for four protein classification systems. Overall, each platform measured similar numbers of proteins in each subcategory. Notable exceptions include expanded coverage of the immunoglobulin receptor superfamily on the Olink platform and more serine/threonine protein kinases on the Soma1.3K platform. Despite targeting fewer proteins than Olink, a nominally larger percentage of Soma1.3K targets had PANTHER annotations (92% versus 87%). Soma1.3K had more representation than Olink among the largest subcategories, whereas Olink proteins were more often classified in low-frequency subcategories (fig. S2).

While Olink and Soma1.3K measure a similar number of proteins from standard protein categories, we sought to understand the variety of captured biology in an unsupervised fashion. Thus, each full platform was decomposed by principal components analysis (PCA). By aligning protein variation along multiple orthogonal axes of variation, PCA captures statistical variety. As seen in Fig. 5A, more than 30% of total variation in the Olink platform is explained in the first two principal components (PCs), compared to approximately 15% of total variation for the Soma1.3K platform. Ultimately, 95% of total variation in Olink is explained in fewer PCs (Fig. 5B). To understand whether certain demographic or clinical factors explain

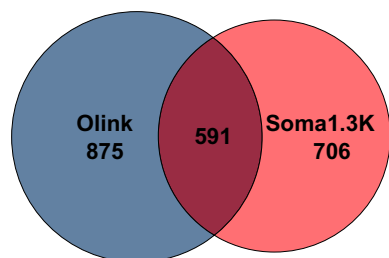


Fig. 1. Unique proteins identified by each platform in analysis of JHS. Venn diagram depicts the overlap between unique UniProt IDs targeted by the Olink Explore and Soma1.3K platforms. Pairing Olink and Soma1.3K reagents based on UniProt target identifies 616 unique reagent pairs.

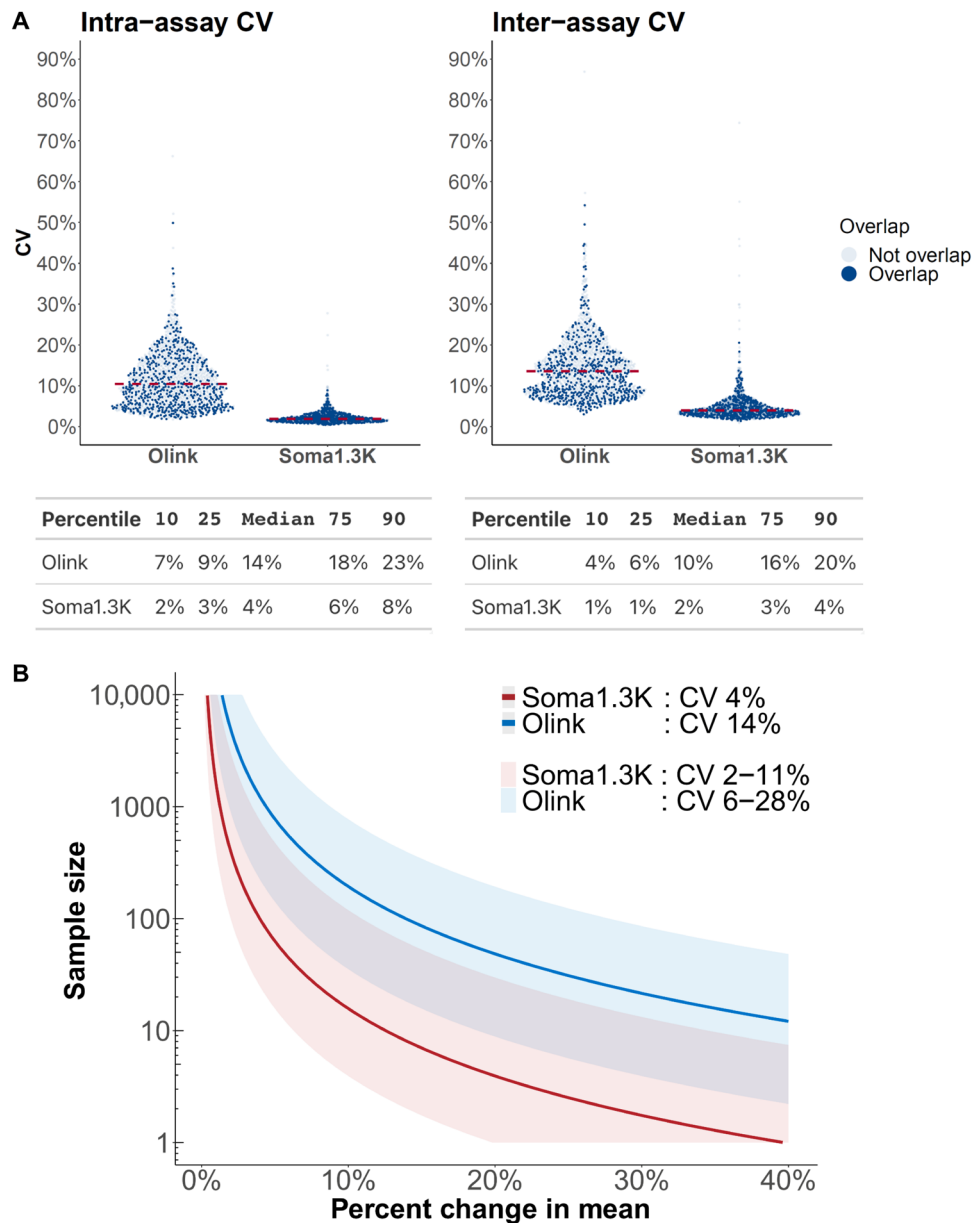


Fig. 2. Intra- and inter-assay CVs. (A) CVs shown are for each reagent on each platform. Intra-assay CVs were calculated using two standard pooled plasma samples included on each plate of a given profiling batch and averaged across all plates. Inter-assay CVs were calculated using 14 pooled plasma samples from seven Olink plates and 10 calibrator samples from five Soma1.3K plates (batch 1 samples only). The CV corresponding to each percentile is shown in the table below the plot. Reagents that have overlapping protein targets are highlighted in darker blue. (B) Shown are a family of curves for each platform showing the relationship between the difference in mean protein level between two groups and the sample size required to detect that difference for a given CV. The mean inter-assay CV for each platform is indicated by the solid line, and the 5th percentile CV and 95th percentile CVs are indicated by the limits of the shaded regions. As CV increases, the required sample size to detect a given percent difference in mean protein levels between groups also increases.

the top two PCs, Fig. 5C shows age, sex, and kidney function as assessed by the eGFR overlaid on the top two PCs from each platform. Across Olink, a gradient of both eGFR and age is apparent across PCs 1 and 2, while only PC 2 from Soma1.3K appears associated with eGFR and age. The relationship between Soma1.3K PC 2 and renal function manifests in the top proteins associated with this PC, which include cystatin C and β 2-microglobulin (fig. S3), two well-established markers of renal function (27). Sex was not associated with PC 1 or 2 on either platform.

Phenotypic association

A principal goal of proteomic profiling is to detect and understand novel disease mediators and biomarkers. The present analysis showed many expected, previously described phenotypic associations on both platforms including cystatin C with eGFR (28), leptin with BMI (29), B-type natriuretic peptide with systolic blood pressure (30), and interleukin-18 receptor 1 with hemoglobin A1c (31) (tables S3 and S4). Figure 6A shows the number of significant associations for eight important clinical traits, at three common significance cutoffs,

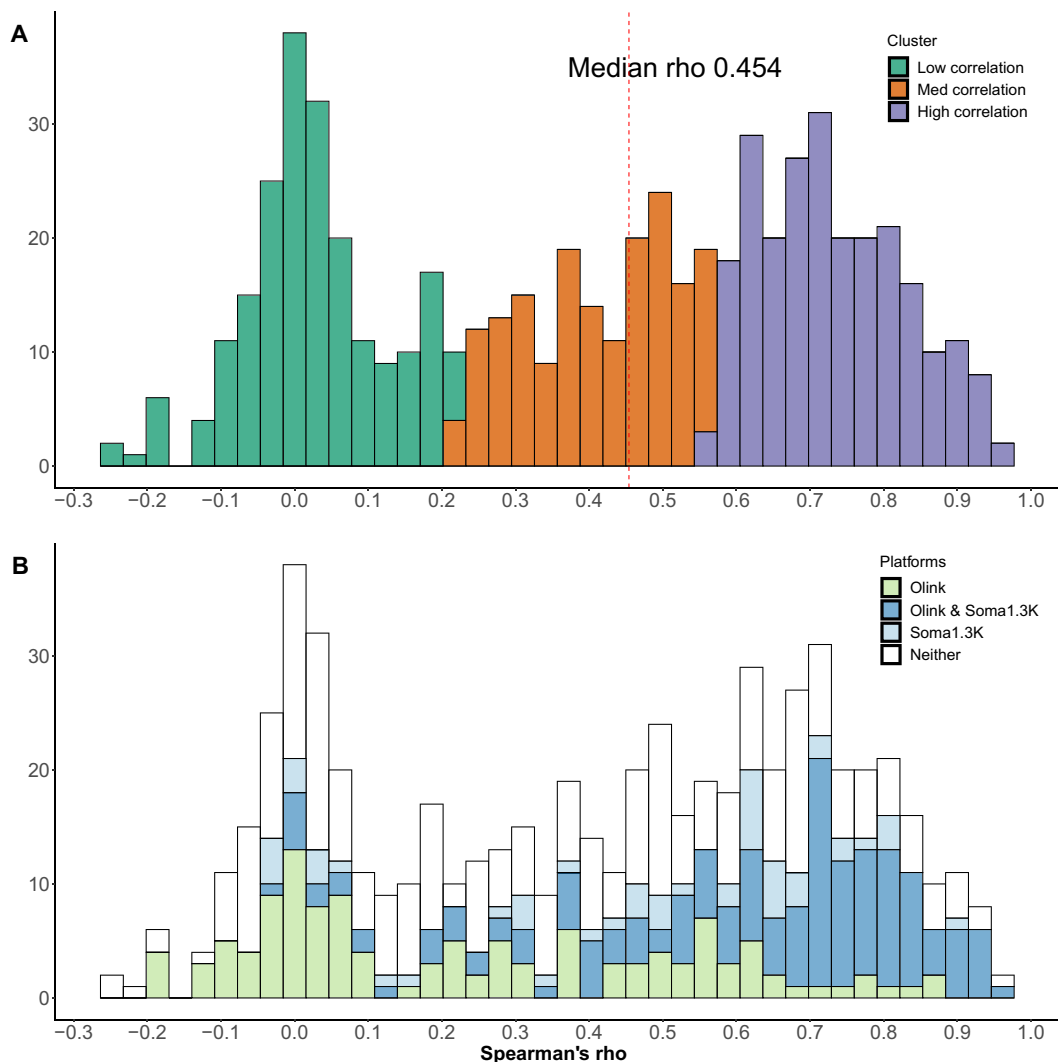


Fig. 3. Spearman correlations between Olink and Soma1.3K reagents, which measure the same protein. (A) *K*-means clustering of correlations into three levels of correlation. (B) Colored bars indicate number of proteins on each platform in that correlation bin that have cis pQTLs, defined as a variant-protein association with $P < 1 \times 10^{-5}$ within 1 Mb of the transcription start site for the cognate gene.

across each full platform. The Olink platform, with a slightly greater number of reagents than the Soma1.3K, demonstrated a higher number of associations with each trait, regardless of the significance threshold used. When only overlapping proteins were considered, Olink maintained more phenotypic associations, particularly among those reagents which correlated poorly with Soma1.3K (Fig. 6B). For example, when considering proteins associated with BMI in the low correlation cluster of 207 reagent pairs, 106 Olink reagents associated with BMI at $P < 0.05$, versus only 44 Soma1.3K reagents.

To examine protein associations that increase the total number of associations but do not expand the total variance explained in the phenotype of interest (a result, possibly, of measuring multiple, highly correlated proteins, sometimes linked in the same biological pathway), we performed separate Lasso regression for each platform and trait (Table 2 and fig. S4). Despite measuring more proteins, the antibody-based platform did not explain more total variance in each phenotype, and overall, the total variance explained by the Olink versus Soma1.3K platform was similar.

Soma5K compared to Olink in HERITAGE

We next extended our analyses from JHS to the HERITAGE Family Study to assess whether our observations of the Soma1.3K would also apply to the Soma5K platform. Thus, we profiled 219 subjects from the HERITAGE Family Study (clinical characteristics in table S5) with the Soma5K and the Olink platforms (Fig. 7). There were 1137 protein targets overlapping both platforms. Despite the smaller sample, similar patterns to the Soma1.3K comparison emerged. The correlation distribution between matched reagents on the Soma5K and Olink platforms was comparable to that of the Soma1.3K comparison, although the median Spearman's correlation was lower (0.35 for the Soma5K and 0.44 for the Soma1.3K), with a similar overall distribution (Fig. 7B). The expanded Soma5K platform provided a larger number of protein assays associated with available clinical traits (Fig. 7C); for instance, there were 1044 associations with BMI and 787 associations with total/high-density lipoprotein (HDL) cholesterol (compared to 439 and 243 on the Soma1.3K platform, respectively). When assessing the clinical trait

Table 1. pQTLs for each platform by correlation cluster.

Cluster	Olink	Soma1.3K	Olink and Soma1.3K	Neither	Total
Low correlation	66	13	19	109	207
Medium correlation	36	15	38	84	173
High correlation	21	26	107	82	236
Total	123	54	164	275	616



Fig. 4. Proteins on each platform by PANTHER protein classification. The number of proteins on each platform in the top 20 subcategories across four classification systems. Soma1.3K is shown in red, Olink is shown in blue. Distribution for all subcategories can be viewed in fig. S2.

associations for overlapping protein targets only, the pattern was similar to the comparison between Olink and Soma1.3K: The antibody-based platform demonstrated a larger number of associations overall, particularly among proteins with weaker correlations ($\rho < 0.3$) between the two platforms (Fig. 7D). For example, when considering proteins associated with BMI in the low correlation cluster of 467 reagent pairs, 174 Olink reagents associated with BMI at $P < 0.05$, versus only 122 Soma1.3K reagents. Lasso regression for traits in HERITAGE again showed that additional protein measurements did not necessarily increase the variance explained in derived models (table S6), although reduced sample size makes these estimates less stable.

Reagent validation with ELISA

As previously noted, high-affinity reagent platforms for protein quantification show substantial gains in efficiency, albeit at a cost to accuracy. Together, the preceding data suggest that when reagents cannot be individually tested against a gold standard, strong correlation to a paired reagent on another platform, the presence of a cis pQTL, or a significant association with a clinical trait can highlight the accuracy or value of a given reagent. To better delineate this, we selected reagents for four protein targets for further testing against a well-validated commercial ELISA. For each protein, one

of the two reagents had either a clinical association or a cis pQTL (or both), while the other did not. CD97 showed in our data a previously unknown association with hemoglobin A1c when measured by the Olink reagent [β (95% confidence interval) = 1.06 (0.81 to 1.31), $P = 2.6 \times 10^{-13}$] as well as a cis pQTL, whereas the Soma1.3K reagent did not. Mesothelin showed an Olink-specific novel association with the atherosclerotic cardiovascular disease (ASCVD) risk score when measured with the Olink reagent [$\beta = 0.017$ (0.006 to 0.027), $P = 0.002$] as well as a cis pQTL for Olink. Heat shock protein, 70 kDa (HSP70) had an association with BMI when measured by the Olink reagent [$\beta = 1.51$ (0.68 to 2.34), $P = 4.0 \times 10^{-4}$]. Last, angiopoietin-like 3 (ANGPTL3) as measured by Olink also showed an association with BMI [$\beta = 2.37$ (1.05 to 3.69), $P = 4.7 \times 10^{-4}$], one previously suggested in the literature (32), as well as an Olink cis pQTL.

When each protein was measured in 60 random samples from either HERITAGE or JHS using the ELISA, the reagent with these clinical and/or genetic associations had a strong positive correlation with the ELISA, while the other did not, suggesting that the reagent with the associations was in fact measuring the protein in question (Fig. 8). Further, in the case of ANGPTL3, where the aptamer reagent was updated from Soma1.3K to Soma5K, we were further able to show that correlations with the ELISA and the Olink reagent also improved (Fig. 8E).

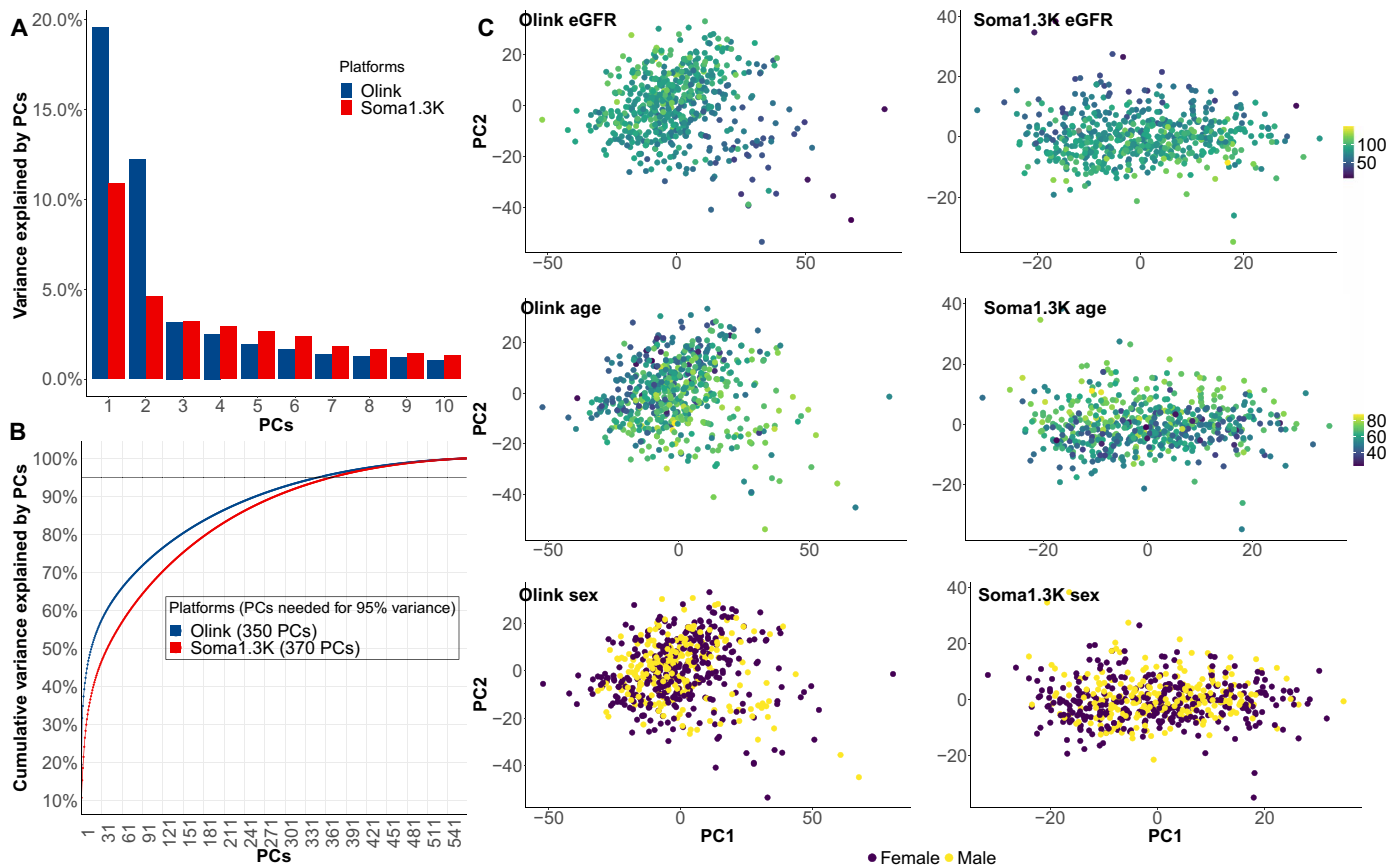


Fig. 5. PCA of each platform. (A) Total platform variance explained by each of the top 10 PCs on each platform. (B) Total cumulative variance explained with 95% variance marked by the black horizontal line. (C) Scatterplot of each participant showing their top 2 PCs and overlaid with eGFR, age, or sex.

DISCUSSION

A deeper understanding of the proteomic profiles of disease is crucial for future research efforts, and cohorts focused on cardiovascular phenotypes are among the largest to leverage these platforms thus far (2, 9, 13, 33). Proteomic insights can improve disease prediction, uncover novel pathways, and identify drug targets (2, 22, 34). Compared to previous proteomic platform comparisons, which were limited by sample size (16), a key strength of our data is that they provide a more comprehensive investigation across multiple domains for two of the most widely used profiling platforms available within two large, well-phenotyped cohorts. Using these data, researchers may be better equipped to interpret existing proteomic data and/or plan future investigations with a clearer understanding of the relative strengths and limitations of each platform.

Our data help illuminate the advantages afforded by each platform. Aptamer-based protein measurements are more consistent over repeated measures, whether examining intra-assay or inter-assay CVs, consistent with previous work (1, 4, 35, 36). The CVs on the Olink platform may be improved if more pooled plasma measurements were used; however, the Soma1.3K platform still outperformed Olink when limited to two pooled plasma measurements. The reason for these differences remains unclear but may in part be related to the exquisitely small sample volume used for Olink, which, in some cases, may itself be considered an advantage, in the case of limited sample availability. Olink antibody reagents are also sometimes

polyclonal, which could affect precision but may also make them more resistant to binding interference. When planning studies with Olink data, greater sample sizes or larger protein effect sizes may be needed to overcome the observed measurement variability, compared to SomaScan.

To draw strong biological conclusions from proteomic analysis, accuracy is paramount. To this end, use of a gold standard is optimal, and we and others have worked to verify a small but important subset of platform reagents using LC-MS (1, 25). In the absence of a cost/time-effective gold standard for these platforms, comparing the platforms to each other and to available genetics data can help inform specificity. The correlations between reagents on each platform revealed a picture of three clusters of proteins. In one cluster, the two measurements are highly correlated, suggesting specificity for both. In another cluster, the mean correlation is near zero, implying that the platforms are not measuring the same target and one or both assays may be inaccurate. Last, there is a distinct middle ground. We hypothesize that many of these medium correlation reagents are measuring the same target protein, likely the correct one, but the reagents on one or both platforms may be affected differentially by interactions with another protein or some posttranslational modification.

Genetic variation provides a useful initial orthogonal tool to support specificity. Genetic variation in or near the gene that codes for a protein can affect plasma protein levels (13, 22–25), such that identifying these cis pQTLs for a reagent can indicate that they are accurately

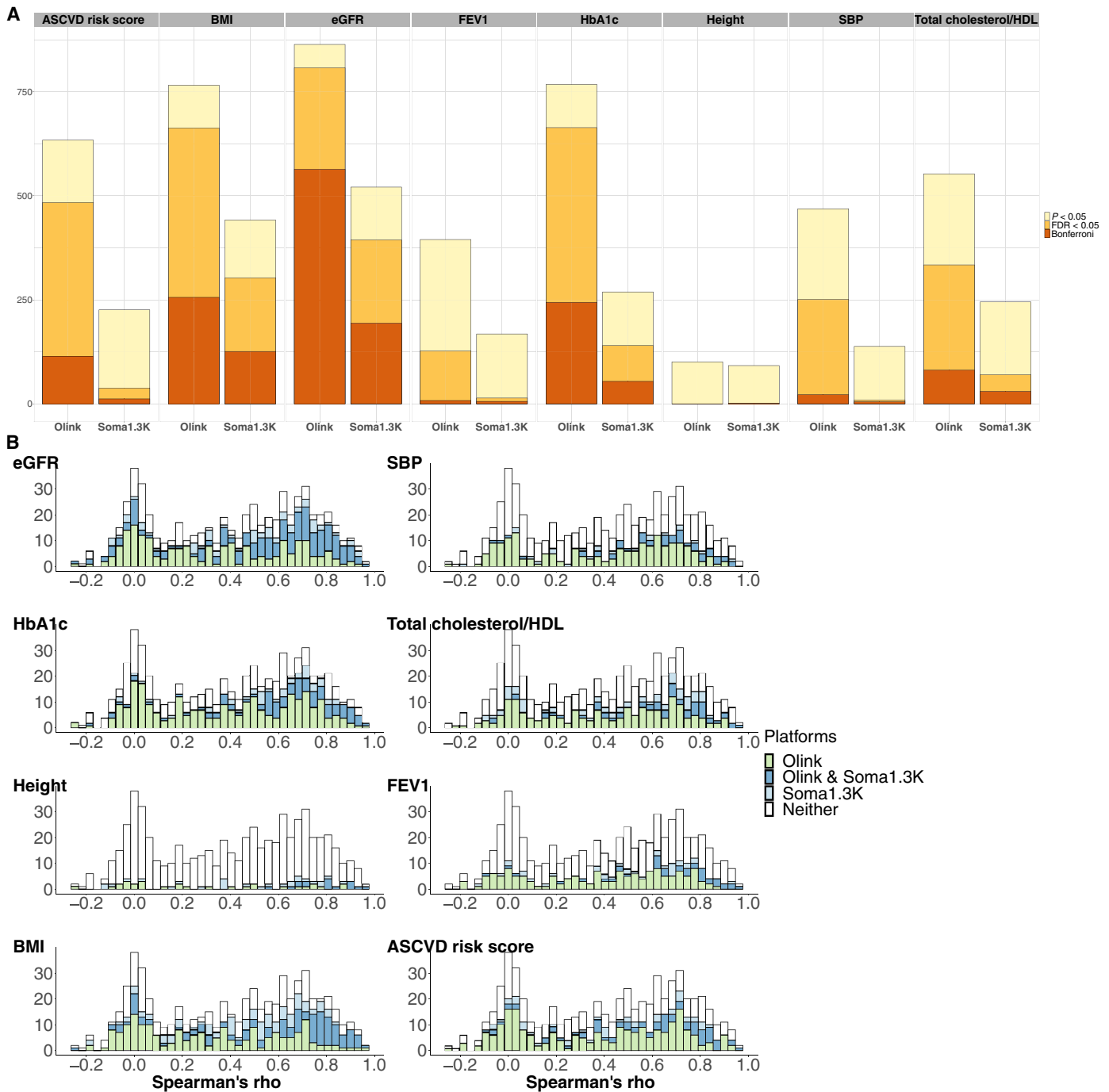


Fig. 6. Phenotypic associations by platform. (A) Number of associations on each platform across eight phenotypes and at three different significance thresholds. (B) Associations at $P < 0.05$ for the same eight phenotypes but limited to overlapping proteins from each platform. The associations are shown on the same distribution of Spearman correlations as seen in Fig. 3. ASCVD, atherosclerotic cardiovascular disease; BMI, body mass index (kg/m^2); eGFR, estimated glomerular filtration rate ($\text{ml}/\text{min}/1.73 \text{ m}^2$); FEV1, Forced expiratory volume in the first second (L); HbA1c, hemoglobin A1c (%); SBP, systolic blood pressure (mmHg), total cholesterol/HDL, total cholesterol divided by high-density lipoprotein cholesterol; FDR, false discovery rate.

targeting the stated protein. In this assessment, Olink held an advantage, as a higher percentage of proteins on that platform had cis pQTLs. Among overlapping proteins at low correlation, the Olink platform showed more pQTLs, suggesting that in a cluster where there is uncertainty about each platform’s specificity, Olink was more likely

binding the specified protein. Furthermore, it is notable that the specificity of antibody-based reagents is more readily confirmed with alternate biochemical methods, as compared to aptamer-based reagents.

Both proteomic platforms are presently expanding: SomaScan has already made available a 7K platform with even greater breadth

Table 2. Lasso regression models in JHS. Lasso regression models are shown for eight phenotypes derived from all proteins available on each platform across 568 samples in JHS. The variance explained by the model and the number of proteins required to achieve that level of variance are shown. FEV1, Forced expiratory volume in the first second. HDL, high-density lipoprotein; ASCVD, atherosclerotic coronary risk score. ASCVD risk score is based on the pooled cohort equation.

Phenotypes	R ² Olink (SD)	R ² Soma1.3K (SD)	No. of proteins Olink	No. of proteins Soma1.3K
Body mass index	0.758 (0.026)	0.718 (0.037)	152	131
Estimated glomerular filtration rate	0.655 (0.058)	0.643 (0.053)	87	144
Hemoglobin A1c	0.613 (0.056)	0.55 (0.075)	137	178
Height	0.544 (0.045)	0.517 (0.038)	143	99
FEV1	0.575 (0.052)	0.579 (0.049)	97	49
Total cholesterol/HDL	0.587 (0.071)	0.596 (0.062)	69	65
Systolic blood pressure	0.217 (0.055)	0.162 (0.057)	51	68
ASCVD risk score	0.567 (0.041)	0.535 (0.034)	81	84

than the Soma5K assessed here. Although Olink and the Soma1.3K measure similar numbers of proteins, our PCA suggests that the aptamer-based platform captures more statistical variation in the proteome. Further, Soma1.3K measures more protein kinases, a particularly important subclass given their utility as drug targets—although the roles of circulating kinases in the plasma are of less clear consequence. Together, our data suggest that the aptamer-based platforms, including the Soma5K, capture a larger statistical and biological breadth of information. Expansions of the Olink platform are also underway and should be compared to the expanded SomaScan7K.

PCA revealed the notable impact of age and eGFR on protein levels, accounting for substantial variation on both platforms, although the effect was smaller for the Soma1.3K platform. This suggests that caution should be taken when extrapolating findings from a cohort with renal dysfunction and highlights the importance of adjusting protein measurements for these variables.

Ultimately, precision, accuracy, and breadth influence each platform's ability to detect meaningful biological associations and insights. Conversely, reagents that capture more noise than variation in protein abundance are unlikely to associate with well-measured clinical variables. The breadth of the Soma5K increases the number of associations detected and, coupled with its larger statistical and biological coverage, may provide a valuable advantage for preliminary or discovery screens. Critically, however, when only overlapping proteins were considered, the Olink platform detected more associations across phenotypes of interest, particularly in the region of low inter-platform correlation, overcoming the slightly lower precision of the platform overall. This pattern was observed

when comparing Olink to either the Soma1.3K or the Soma5K. These data, when considered with correlation and genetic information, suggest that at least across overlapping reagents, Olink's specificity may increase the likelihood of finding reliable phenotypic associations.

It is important to note, however, that more associations do not always translate into more information. Despite measuring fewer protein targets, the earlier Soma1.3K platform was able to explain nearly as much, if not more, total variance in the phenotypes of interest, a feature that could draw benefit in outcome prediction or assessing a wider variety of disease states. Alternatively, multiple, accurately identified proteins tagging the same biological pathway could add weight to a pathway analysis and biological inference. The preferred profile depends on the research goals. Future work assessing these ever expanding and improving platforms will be necessary to ensure their optimal application.

Our ELISA-based experiments support the hypothesis that using genetics or phenotype associations to infer specificity or usability as described above is valid. Across four protein targets, the reagent with the *cis* pQTL or the phenotypic association also showed strong agreement with the ELISA-based assay, while the other reagent did not. In our four examples, this reagent was always the antibody-based reagent, although there is absolutely no reason to suspect an aptamer with the same characteristics would not show the same pattern. Notably, while HSP70 had the stronger phenotypic association for the Olink reagent, which, in turn, was better correlated to the ELISA measurement, there is a weak *cis* pQTL for the Soma1.3K aptamer (table S2). Close investigation showed that this variant, chr6:32604567:G:GA, is actually in the major histocompatibility complex region of chromosome 6. We and others note this to be a region of high linkage disequilibrium, which can interfere with *cis* pQTL identification (21, 22). We suspect this variant is not a true *cis* pQTL for HSP70.

Our work has important limitations. We did not have repeated measurements from participant samples (rather than pooled samples), which would allow for more accurate CVs and the ability to calculate intraclass correlations, another important metric of precision. The number of participants for whom the Soma5K platform overlaps with Olink is limited, and a greater sample size would enhance the veracity of our observations. Determinations of reagent specificity are inferred without direct verification with LC-MS, which remains the “gold standard,” and as reagents are gradually verified by this approach, those results should likely supersede conclusions drawn from the data presented here (25). We are not able to independently determine lower limit of detection or quantification for each reagent given cost and time constraints, although these data are available from each manufacturer. In our validation experiments with ELISAs, each reagent was able to quantify at least as low as the lowest levels as measured by the ELISA.

In summary, our data provide a comprehensive comparison of large-scale plasma proteomic profiling platforms. The antibody-based platform appears to confer a protein-for-protein edge in specificity and phenotypic association, while the aptamer-based approach demonstrates more reproducible measurements and greater breadth of measurement across the proteome. When choosing a platform, other factors, not directly comparable, may also bear consideration, such as required sample volume, scalability, and cost. Both have provided excellent biological insights in multiple investigations and likely will continue to do so, particularly when profiling with LC-MS, despite notable recent improvement (37), cannot presently provide the necessary sample throughput.

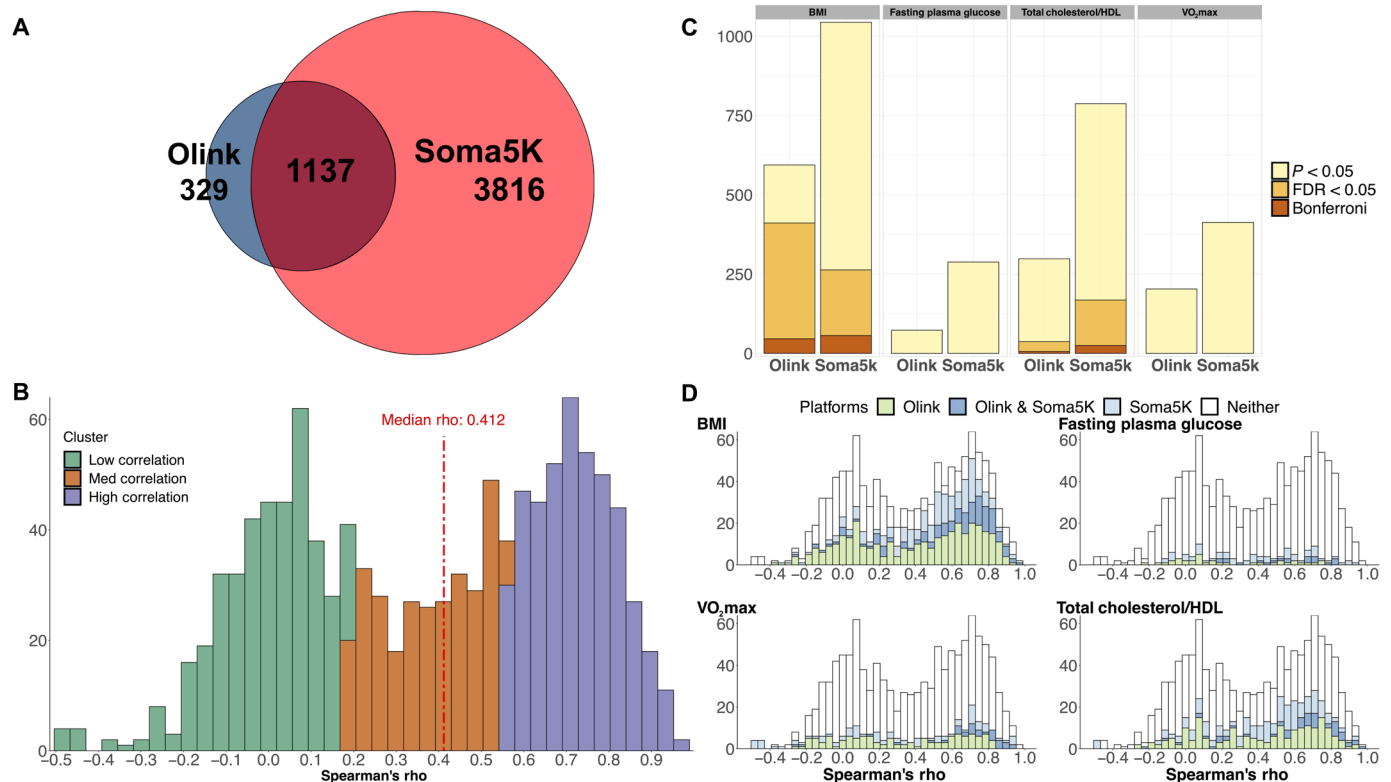


Fig. 7. Comparison between Soma5K and Olink Explore in HERITAGE. Plasma profiling was performed on a random subset of HERITAGE ($N = 219$). (A) Overlap between unique UniProt targets between the two platforms. (B) Spearman correlations between overlapping reagents on Olink and Soma5K. K -means clustering divided the distribution into three clusters. (C) Phenotypic associations between all reagents on each platform and four phenotypes at $P < 0.05$. (D) Associations at $P < 0.05$ for the same four phenotypes but limited to overlapping proteins from each platform. The associations are shown on the same distribution of Spearman correlations as seen in (B).

MATERIALS AND METHODS

Study approval

The JHS study was approved by Jackson State University, Tougaloo College, and the University of Mississippi Medical Center institutional review boards, and all participants provided written informed consent. The human study protocols were approved by the institutional review boards of Beth Israel Deaconess Medical Center, University of Washington, and the four clinical centers of HERITAGE.

Cohorts

The JHS and the HERITAGE Family Study have been described (38, 39). Briefly, the JHS is a community-based longitudinal cohort study begun in 2000 of 5306 self-identified Black individuals from the Jackson, Mississippi metropolitan statistical area (38). Included in the present study are samples collected at visit 1 between 2000 and 2004 from 568 individuals. Clinical traits in JHS have been defined previously (40). Resting blood pressure was measured by recording two measurements in the seated position with a Hawksley random zero sphygmomanometer using one of four cuff sizes selected by measuring arm circumference. Hypertension was defined as use of blood pressure lowering medication or blood pressure $> 140/90$ mmHg. Hypertensive treatment was determined by patient medication inventory or self-report of taking blood pressure medication. Routine laboratory measurements were made at visit 1 using standard venipuncture and laboratory techniques. Glomerular filtration rate

was estimated using the Chronic Kidney Disease Epidemiology Collaboration equation (41). An ASCVD 10-year risk was estimated from the pooled cohort equations (42). Summary statistics are presented as means \pm SD.

HERITAGE enrolled a combination of self-identified white and Black family units, totaling 763 sedentary participants (62% white) between the ages of 17 and 65 years in a 20-week, graded endurance exercise training study across four clinical centers in the United States and Canada in 1992 to 1997 (39). Included in the present study is a subset of 219 individuals demographically representative of the overall HERITAGE cohort with baseline (pretraining), fasted plasma samples. The HERITAGE phenotype measurement protocols have been described (39). Resting blood pressure was measured twice in the fasted state after at least 5 min acclimating to a quiet environment using an appropriately sized automated unit (Colin STBP-780, Colin Medical Instruments, San Antonio, TX) and subsequently averaged. Standard laboratory assessments were performed using 12-hour fasting, morning samples.

SomaScan proteomic profiling

JHS plasma samples were collected at visit 1 in EDTA tubes and then maintained in -70°C freezers (40). Proteomic measurements were performed using Soma1.3K, a single-stranded DNA aptamer-based proteomic platform, which contained 1305 aptamers (43). Nonhuman proteins were excluded from analysis ($N = 4$) for a final count of 1301. Samples were run in two separate batches.

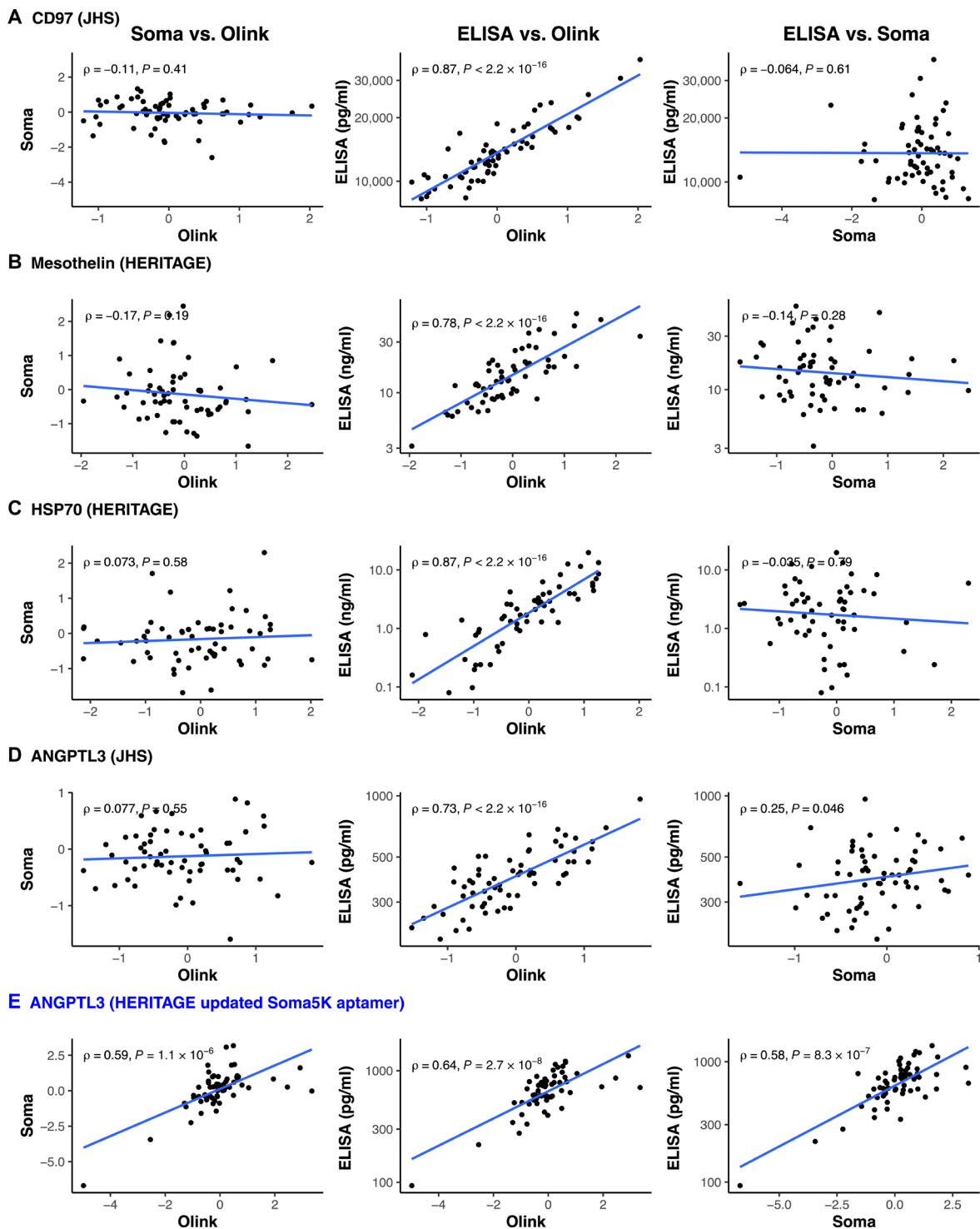


Fig. 8. Correlations between ELISA and both Olink and Soma. In 60 random samples from either JHS or HERITAGE (according to sample availability), protein levels were assayed by ELISA and compared to measurements from each affinity platform. Shown here are the normalized data and Spearman correlations for (A) CD97, (B) mesothelin, (C) HSP70, and (D and E) ANGPTL3. In the case of (A) to (D), aptamers are those featured on the Soma1.3K, while (E) features a new ANGPTL3 aptamer, upgraded on the Soma5K platform. Absolute concentrations by ELISA are shown on a log scale axis, while affinity reagent measurements are log₂-transformed and scaled.

In HERITAGE, plasma samples were collected in EDTA tubes and stored at -80°C , then were diluted in three different concentrations (40, 1, and 0.05%), and profiled using the expanded Soma5K platform (4979 aptamers) in a single batch. Plasma samples had either zero or one freeze-thaw cycle before proteomic profiling.

Assays were performed using SomaScan reagents according to the manufacturer's detailed protocol (43). Briefly, a SomaScan reagent is a single-stranded DNA-based aptamer that is chemically modified to enhance binding to conformational protein epitopes. In addition, the aptamers are fluorophore-tagged to allow detection by standard oligoarray readers. The assay measures proteins directly from plasma using a multistep capture, release, and recapture enrichment process. Plasma proteins first bind to the bead-immobilized aptamers. Aptamer-bound proteins are then biotinylated. Aptamer-protein complexes are next released by a photocleavage process. Biotinylated proteins are then bound to a second set of streptavidin beads. Following a washing step, aptamers are released from the protein targets and collected. The fluorophore-tagged modified nucleotides are quantitated using an oligoarray plate reader providing relative fluorescent unit readout, which is proportional to protein concentration in the sample. The assay was performed on 96-well plates with 85 wells on each plate dedicated to study samples and 11 wells used for quality controls (QCs). QC samples include seven "calibrator" plasma samples from a single "pool" that are used by the manufacturer to assess intra-assay CVs and standardize across experiments; four samples from a distinct "QC" plasma pool are used to assess inter-assay CVs across plates. Sample data were normalized to remove hybridization variation within an oligoarray reader set followed by median normalization across all samples to remove other assay biases within the run and lastly calibrated to remove assay differences between runs. Samples are \log_2 -transformed and scaled to mean of 0 and SD of 1. This is done within batch if samples were run in batches, which was the case in JHS but not HERITAGE. Outlier analysis was performed by PCA (see below); no outliers were identified.

Olink Explore proteomic profiling

Profiling was performed in a single batch in JHS and two batches of $N = 88$ and $N = 121$ in HERITAGE using the Olink Explore panel (Olink Proteomics AB, Uppsala, Sweden) according to the manufacturer's instructions using separate aliquots. The Proximity Extension Assay technology used for the Olink protocol has been described (5), and Olink enables analysis using 2.8 μl of each sample. Briefly, pairs of oligonucleotide-labeled antibody probes bind to their targeted protein, and if the two probes are brought in close proximity, the oligonucleotides hybridize in a pair-wise manner. The addition of a DNA polymerase leads to a proximity-dependent DNA polymerization event, generating a unique double-stranded DNA barcode for each specific antigen. The resulting DNA sequence is subsequently detected and quantified using next-generation sequencing (Illumina NovaSeq). Data are then quality-controlled and normalized using an internal extension control and a plate control, to adjust for intra- and inter-run variation. The final assay readout is presented in Normalized Protein eXpression (NPX) values, which is \log_2 -transformed ratio of sample assay counts to extension control counts; a higher value corresponds to a higher protein expression. Internal controls for incubation, extension, and amplification are included on each plate. Outlier analysis was performed by PCA; two samples were removed across all JHS analyses resulting in the

present $N = 568$ of the present study. No samples were excluded in HERITAGE analyses. All assay validation data (detection limits, intra- and inter-assay precision data, etc.) are available on manufacturer's website (www.olinkexplore.com).

Pairing platform reagents by protein targets

Protein targets are identified here by their UniProt ID (www.uniprot.org), which uniquely identifies a peptide sequence. As proteins are commonly found in multimers, some affinity reagents target multiple UniProt IDs. Conversely, on both platforms, there are instances where multiple reagents target the same protein. Thus, reagents on each platform were paired to one another for direct comparison if they target the same UniProt ID. These reagents are identified as "overlapping."

Comparing CVs

While Soma1.3K plates include several replicates for calculating CVs (see above), for the direct comparison described here, calculations of CV on the Soma1.3K platform were limited to the first two QC samples on each plate to match the number of pooled plasma samples on Olink plates. Intra-assay CVs were calculated using those two standard pooled plasma samples on each plate and then averaged across all plates. Inter-assay CVs were calculated using 14 pooled plasma samples from Olink (two each from seven plates) and 10 pooled plasma samples from Soma1.3K (two each from five plates, batch 1 samples only). The median CV was determined for each platform, as well as the 10, 25, 75, and 90 percentile CVs. Because Olink NPX values are \log_2 -transformed, CVs were calculated using the equation recommended by Olink: $\text{CV} = \sqrt{e^{(\ln(2) \times \sigma_{\text{NexusNPX}})^2} - 1}$ compared to $\text{CV} = \sigma \div \mu$ for Soma1.3K.

Correlation of matched reagents

For reagents on each platform targeting the same UniProt protein, Spearman correlation was calculated using \log_2 -transformed and scaled measurements. *K*-means clustering was used to identify subgroups of correlation. Three clusters were created on the basis of the elbow method.

Genotyping and imputation

WGS in JHS has been previously described (44, 45). Included in the present study are participants included in Freeze 6 of the Trans-Omics for Precision Medicine (TOPMed) project, sequenced at the Northwest Genome Center at University of Washington. Samples underwent $>30\times$ WGS. Genotype calling with vt (46) and QC were performed by the Informatics Resource Center at the University of Michigan (44).

WGS association analysis in JHS

\log -transformed and scaled (to mean = 0 and SD = 1) Soma1.3K measurements were residualized on age, sex, batch, and PCs of ancestry 1 to 10 as determined by GENetic ESTimation and Inference in Structured samples (47). The resulting residuals were then inverse normalized. Olink protein measurements underwent the same normalization but did not require adjustment for batch. The association between these values and genetic variants was tested using linear mixed effects models adjusted for age, sex, the genetic relationship matrix, and PCs 1 to 10 using the fastGWA model implemented in the Genome-wide Complex Trait Analysis (GCTA) software package (version 1.93.2beta/gcta64) (48). Repeat adjustment for covariates was implemented to

reduce type I error and improve statistical power (49). Variants with a minor allele count less than five were excluded from analysis.

Identifying cis pQTLs

Cis pQTLs were defined as variants associated with protein measurement and located within 1 megabase (Mb) of the transcription start site of the cognate gene of the target protein. A *P* value threshold was set at 1×10^{-5} . Given that genome-wide significance accounts for 3 billion bases, adjusting genome-wide significance to the 2-Mb window gives $5 \times 10^{-8} \times (3 \times 10^9 / 2 \times 10^6) = 7.5 \times 10^{-5}$. Thus, 1×10^{-5} is a conservative threshold.

Identifying previously identified cis pQTLs in PhenoScanner

To determine whether pQTLs were previously unknown, we used the PhenoScanner package (version 2) for R (50, 51). For each protein-locus association identified above, we divided the locus into 1 Mb or less segments (maximum permitted by PhenoScanner application programming interface) if needed. The resulting region or regions were then passed to the phenoscanner function in R, with the following arguments: build was set to “38,” *P* value to 1×10^{-5} , catalog to “pQTL,” proxies set to “None” (query date 5 April 2022). To supplement PhenoScanner, we reviewed the literature for additional studies using SomaScan or Olink to identify the genetic architecture of the plasma proteome, and we identified that three are not in the PhenoScanner (2, 13, 25). Results from these studies were considered using the same criteria as above.

Protein annotation using PANTHER

The PANTHER classification system (<http://pantherdb.org/>) was used to annotate each protein, using the complete set of UniProt IDs covered by each platform, and protein counts for each category on each platform are displayed. Categories are arranged by the total number of proteins from either platform.

Principal components analysis

After \log_2 transformation and scaling of measurements were performed on each platform to achieve normal values as above, missing values on the Olink platform (0.2% of all measurements) were imputed by substituting the mean value for that protein. There were no missing values in the Soma1.3K data. PCA was performed on each full platform using the “tidymodels” package in R 4.0.2 (Vienna, Austria). The percent variation explained by each PC and the number of PCs to explain 95% of the total platform variation were determined.

Clinical trait associations

Associations between clinical traits (dependent variable) and \log_2 -transformed and scaled proteins (independent variable) on each platform were determined by linear regression. Models were adjusted for age and sex. Models for Soma1.3K proteins were also adjusted for batch. Lasso models were fit for each trait-platform combination with age, sex, and batch (for Soma1.3K in JHS), and all proteins were entered into the model. The tuning parameter giving the minimum integrated mean squared error was identified by fivefold cross-validation repeated five times.

Validation against ELISA

Four proteins (ANGPTL3, CD97, HSP70, and mesothelin) were selected on the basis of ELISA availability and the criteria that one reagent for the protein had a cis pQTL and/or a phenotypic

association, while the other reagent did not. Sixty random samples from JHS or HERITAGE or both when sample availability allowed were selected, and protein levels were measured by ELISA using commercially available kits. Kits for ANGPTL3 (#EH29RB) and HSP70 (#BMS2087) are from Thermo Fisher Scientific, and kits for CD97 (#ab213763) and mesothelin (#ab216168) are from Abcam. Standard curves were run with protein standards serially diluted in buffers according to the manufacturer’s instruction. Spearman’s correlation rho (ρ) was calculated for ELISA versus Olink, ELISA versus Soma, and Soma versus Olink within the 60 samples. For proteins measured in HERITAGE samples, where aptamer measurements come from the Soma5K platform, the same aptamer is used as on the Soma1.3K platform, with the exception of ANGPTL3 that had an updated aptamer on the Soma5K platform. ANGPTL3 ELISA measurements were thus done in JHS (Soma1.3K) and HERITAGE (Soma5K) for comparison.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abm5164>

[View/request a protocol for this paper from Bio-protocol.](#)

REFERENCES AND NOTES

1. 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, Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation* **134**, 270–285 (2016).
2. M. D. Benson, Q. Yang, D. Ngo, Y. Zhu, D. Shen, L. A. Farrell, S. Sinha, M. J. Keyes, R. S. Vasan, M. G. Larson, J. G. Smith, T. J. Wang, R. E. Gerszten, Genetic architecture of the cardiovascular risk proteome. *Circulation* **137**, 1158–1172 (2018).
3. A. Egerstedt, J. Berntsson, M. L. Smith, O. Gidlöf, R. Nilsson, M. Benson, Q. S. Wells, S. Celik, C. Lejonberg, L. Farrell, S. Sinha, D. Shen, J. Lundgren, G. Rådegran, D. Ngo, G. Engström, Q. Yang, T. J. Wang, R. E. Gerszten, J. G. Smith, Profiling of the plasma proteome across different stages of human heart failure. *Nat. Commun.* **10**, 5830 (2019).
4. 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, Assessment of variability in the SOMAScan assay. *Sci. Rep.* **7**, 14248 (2017).
5. E. Assarsson, M. Lundberg, G. Holmquist, J. Björkstén, S. B. Thorsen, D. Ekman, A. Eriksson, E. Renzel Dickens, S. Ohlsson, G. Edfeldt, A.-C. Andersson, P. Lindstedt, J. Stenvang, M. Gullberg, S. Fredriksson, Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS ONE* **9**, e95192 (2014).
6. M. D. Benson, D. Ngo, P. Ganz, R. E. Gerszten, Emerging affinity reagents for high throughput proteomics: Trust, but verify. *Circulation* **140**, 1610–1612 (2019).
7. J. G. Smith, R. E. Gerszten, Emerging affinity-based proteomic technologies for large-scale plasma profiling in cardiovascular disease. *Circulation* **135**, 1651–1664 (2017).
8. J. E. Van Eyk, K. Sobhani, Precision medicine. *Circulation* **138**, 2172–2174 (2018).
9. D. Ko, M. D. Benson, D. Ngo, Q. Yang, M. G. Larson, T. J. Wang, L. Trinquart, D. D. McManus, S. A. Lubitz, P. T. Ellinor, R. S. Vasan, R. E. Gerszten, E. J. Benjamin, H. Lin, Proteomics profiling and risk of new-onset atrial fibrillation: Framingham Heart Study. *J. Am. Heart Assoc.* **8**, e010976 (2019).
10. 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, Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci. Rep.* **8**, 8382 (2018).
11. L. Adamo, J. Yu, C. Rocha-Resende, A. Javaheri, R. D. Head, D. L. Mann, Proteomic signatures of heart failure in relation to left ventricular ejection fraction. *J. Am. Coll. Cardiol.* **76**, 1982–1994 (2020).
12. J. A. Chirinos, L. Zhao, Y. Jia, C. Frej, L. Adamo, D. Mann, S. V. Shewale, J. S. Millar, D. J. Rader, B. French, J. Brandimarto, K. B. Margulies, J. S. Parks, Z. Wang, D. A. Seiffert, J. Fang, N. Sweitzer, C. Chistoffersen, B. Dahlbäck, B. D. Car, D. A. Gordon, T. P. Cappola, A. Javaheri, Reduced apolipoprotein m and adverse outcomes across the spectrum of human heart failure. *Circulation* **141**, 1463–1476 (2020).
13. L. Folkersen, S. Gustafsson, Q. Wang, D. H. Hansen, Å. K. Hedman, A. Schork, K. Page, D. V. Zhernakova, Y. Wu, J. Peters, N. Eriksson, S. E. Bergen, T. S. Boutin, A. D. Bretherick, S. Enroth, A. Kalnapenkis, J. R. Gådin, B. E. Suur, Y. Chen, L. Matic, J. D. Gale, J. Lee, W. Zhang, A. Quazi, M. Ala-Korpela, S. H. Choi, A. Claringbould, J. Danesh, G. D. Smith,

- F. de Masi, S. Elmståhl, G. Engström, E. Fauman, C. Fernandez, L. Franke, P. W. Franks, V. Giedraitis, C. Haley, A. Hamsten, A. Ingason, Å. Johansson, P. K. Joshi, L. Lind, C. M. Lindgren, S. Lubitz, T. Palmer, E. Macdonald-Dunlop, M. Magnusson, O. Melander, K. Michaelsson, A. P. Morris, R. Mägi, M. W. Nagle, P. M. Nilsson, J. Nilsson, M. Orho-Melander, O. Polasek, B. Prins, E. Pålsson, T. Qi, M. Sjögren, J. Sundström, P. Surendran, U. Vösa, T. Werge, R. Wernersson, H.-J. Westra, J. Yang, A. Zernakova, J. Ärnlöv, J. Fu, J. G. Smith, T. Esko, C. Hayward, U. Gyllensten, M. Landen, A. Siegbahn, J. F. Wilson, L. Wallentin, A. S. Butterworth, M. V. Holmes, E. Ingelsson, A. Mälärstig, Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* **2**, 1135–1148 (2020).
14. J. M. Robbins, B. Peterson, D. Schraner, U. A. Tahir, T. Riemmüller, S. Deng, M. J. Keyes, D. H. Katz, P. M. J. Beltran, J. L. Barber, C. Baumgartner, S. A. Carr, S. Ghosh, C. Shen, L. L. Jennings, R. Ross, M. A. Sarzynski, C. Bouchard, R. E. Gerszten, Human plasma proteomic profiles indicative of cardiorespiratory fitness. *Nat. Metab.* **3**, 786–797 (2021).
15. K. Lifshitz, Y. Ber, C. Shenhar, J. Nillson, A. Peer, E. Rosenbaum, J. Baniel, D. Kedar, O. I. Ben Zadok, D. Margel, Cardiovascular proteomics: A post-hoc analysis from a phase II randomized clinical trial comparing GnRH antagonist vs GnRH agonist among men with advanced prostate cancer. *J. Urol.* **206**, 952–959 (2021).
16. L. M. Raffield, H. Dang, K. A. Pratte, S. Jacobson, L. A. Gillenwater, E. Ampleford, I. Barjaktarevic, P. Basta, C. B. Clish, A. P. Comellas, E. Cornell, J. L. Curtis, C. Doerschuk, P. Durda, C. Emson, C. M. Freeman, X. Guo, A. T. Hastie, G. A. Hawkins, J. Herrera, W. C. Johnson, W. W. Labaki, Y. Liu, B. Masters, M. Miller, V. E. Ortega, G. Papanicolaou, S. Peters, K. D. Taylor, S. S. Rich, J. I. Rotter, P. Auer, A. P. Reiner, R. P. Tracy, D. Ngo, R. E. Gerszten, W. K. O'Neal, R. P. Bowler; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Comparison of proteomic assessment methods in multiple cohort studies. *Proteomics* **20**, e1900278 (2020).
17. S. Y. Lim, J. H. Lee, S. J. Welsh, S. B. Ahn, E. Breen, A. Khan, M. S. Carlino, A. M. Menzies, R. F. Kefford, R. A. Scolyer, G. V. Long, H. Rizos, Evaluation of two high-throughput proteomic technologies for plasma biomarker discovery in immunotherapy-treated melanoma patients. *Biomark. Res.* **5**, 32 (2017).
18. M. Pietzner, E. Wheeler, J. Carrasco-Zanini, N. D. Kerrison, E. Oertner, M. Koprulu, J. Luan, A. D. Hingorani, S. A. Williams, N. J. Wareham, C. Langenberg, Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat. Commun.* **12**, 6822 (2021).
19. 1000 Genomes Project Consortium, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
20. J. M. McClellan, T. Lehner, M.-C. King, Gene discovery for complex traits: Lessons from Africa. *Cell* **171**, 261–264 (2017).
21. D. H. Katz, U. A. Tahir, A. G. Bick, A. Pampana, D. Ngo, M. D. Benson, Z. Yu, J. M. Robbins, Z.-Z. Chen, D. E. Cruz, S. Deng, L. Farrell, S. Sinha, A. A. Schmaier, D. Shen, Y. Gao, M. E. Hall, A. Correa, R. P. Tracy, P. Durda, K. D. Taylor, Y. Liu, W. C. Johnson, X. Guo, J. Yao, Y.-D. Ida Chen, A. W. Manichaikul, D. Jain, C. Bouchard, M. A. Sarzynski, S. S. Rich, J. I. Rotter, T. J. Wang, J. G. Wilson, P. Natarajan, R. E. Gerszten; National Heart, Lung, and Blood Institute TOPMed (Trans-Omics for Precision Medicine) Consortium, Whole genome sequence analysis of the plasma proteome in black adults provides novel insights into cardiovascular disease. *Circulation* **145**, 357–370 (2022).
22. 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, J. Danesh, H. Runz, A. S. Butterworth, Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
23. L. Folkersen, E. Fauman, M. Sabater-Lleal, R. J. Strawbridge, M. Frånberg, B. Sennblad, D. Baldassarre, F. Veglia, S. E. Humphries, R. Rauramaa, U. de Faire, A. J. Smit, P. Giral, S. Kurl, E. Mannarino, S. Enroth, Å. Johansson, S. B. Enroth, S. Gustafsson, L. Lind, C. Lindgren, A. P. Morris, V. Giedraitis, A. Silveira, A. Franco-Cereceda, E. Tremoli; IMPROVE study group, U. Gyllensten, E. Ingelsson, S. Brunak, P. Eriksson, D. Ziemek, A. Hamsten, A. Mälärstig, Mapping of 79 loci for 83 plasma protein biomarkers in cardiovascular disease. *PLoS Genet.* **13**, e1006706 (2017).
24. 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, Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.* **8**, 14357 (2017).
25. 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, Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
26. J. Zhang, D. Dutta, A. Köttgen, A. Tin, P. Schlosser, M. E. Grams, B. Harvey; CKDGen Consortium, B. Yu, E. Boerwinkle, J. Coresh, N. Chatterjee, Plasma proteome analyses in individuals of European and African ancestry identify cis-pQTLs and models for proteome-wide association studies. *Nat. Genet.* **54**, 593–602 (2022).
27. L. A. Inker, H. Tighiouart, J. Coresh, M. C. Foster, A. H. Anderson, G. J. Beck, G. Contreras, T. Greene, A. B. Karger, J. W. Kusek, J. Lash, J. Lewis, J. R. Schelling, S. D. Navaneethan, J. Sondheimer, T. Shafi, A. S. Levey, GFR estimation Using β -trace protein and β 2-microglobulin in CKD. *Am. J. Kidney Dis.* **67**, 40–48 (2016).
28. T. W. Ferguson, P. Komenda, N. Tangri, Cystatin C as a biomarker for estimating glomerular filtration rate. *Curr. Opin. Nephrol. Hypertens.* **24**, 295–300 (2015).
29. C. S. Mantzoros, The role of leptin in human obesity and disease: A review of current evidence. *Ann. Intern. Med.* **130**, 671–680 (1999).
30. F. Macheret, D. Heublein, L. C. Costello-Boerrigter, G. Boerrigter, P. McKie, D. Bellavia, S. Mangiafico, Y. Ikeda, K. Bailey, C. G. Scott, S. Sandberg, H. H. Chen, L. Malatino, M. M. Redfield, R. Rodeheffer, J. Burnett Jr., A. Cicaliotti, Human hypertension is characterized by a lack of activation of the antihypertensive cardiac hormones ANP and BNP. *J. Am. Coll. Cardiol.* **60**, 1558–1565 (2012).
31. Z. Chen, F. Miao, B. H. Braffett, J. M. Lachin, L. Zhang, X. Wu, D. Roshandel, M. Carless, X. A. Li, J. D. Tompkins, J. S. Kaddis, A. D. Riggs, A. D. Paterson, R. Natarajan, DNA methylation mediates development of HbA1c-associated complications in type 1 diabetes. *Nat. Metab.* **2**, 744–762 (2020).
32. M. Aghasizadeh, R. Zare-Feyzabadi, T. Kazemi, A. Avan, G. A. Ferns, H. Esmaily, E. Miri-Moghaddam, M. Ghayour-Mobarhan, A haplotype of the ANGPTL3 gene is associated with CVD risk, diabetes mellitus, hypertension, obesity, metabolic syndrome, and dyslipidemia. *Gene* **782**, 145525 (2021).
33. D. Ngo, D. Wen, Y. Gao, M. J. Keyes, E. R. Drury, D. H. Katz, M. D. Benson, S. Sinha, D. Shen, L. A. Farrell, B. D. Peterson, D. J. Friedman, S. Elmariyah, B. A. Young, J. G. Smith, Q. Yang, R. S. Vasan, M. G. Larson, A. Correa, B. D. Humphreys, T. J. Wang, M. R. Pollak, J. G. Wilson, R. E. Gerszten, E. P. Rhee, Circulating testican-2 is a podocyte-derived marker of kidney health. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 25026–25035 (2020).
34. J. Zheng, V. Haberland, D. Baird, V. Walker, P. C. Haycock, M. R. Hurlé, A. Gutteridge, P. Erola, Y. Liu, S. Luo, J. Robinson, T. G. Richardson, J. R. Staley, B. Elsworth, S. Burgess, B. B. Sun, J. Danesh, H. Runz, J. C. Maranville, H. M. Martin, J. Yarmolinsky, C. Laurin, M. V. Holmes, J. Z. Liu, K. Estrada, R. Santos, L. McCarthy, D. Waterworth, M. R. Nelson, G. D. Smith, A. S. Butterworth, G. Hemani, R. A. Scott, T. R. Gaunt, Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nat. Genet.* **52**, 1122–1131 (2020).
35. 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, Application of large-scale aptamer-based proteomic profiling to planned myocardial infarctions. *Circulation* **137**, 1270–1277 (2018).
36. 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, Reproducibility and variability of protein analyses measured using a multiplexed modified aptamer assay. *J. Appl. Lab. Med.* **4**, 30–39 (2019).
37. P. E. Geyer, N. A. Kulak, G. Pichler, L. M. Holdt, D. Teupser, M. Mann, Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195 (2016).
38. H. A. Taylor, J. G. Wilson, D. W. Jones, D. F. Sarpong, A. Srinivasan, R. J. Garrison, C. Nelson, S. B. Wyatt, Toward resolution of cardiovascular health disparities in African Americans: Design and methods of the Jackson Heart Study. *Ethn. Dis.* **15**, 56–4–17 (2005).
39. C. Bouchard, A. S. Leon, D. C. Rao, J. S. Skinner, J. H. Wilmore, J. Gagnon, The HERITAGE family study. Aims, design, and measurement protocol. *Med. Sci. Sports Exerc.* **27**, 721–729 (1995).
40. M. A. Carpenter, R. Crow, M. Steffes, W. Rock, J. Heilbraun, G. Evans, T. Skelton, R. Jensen, D. Sarpong, Laboratory, reading center, and coordinating center data management methods in the Jackson Heart Study. *Am. J. Med. Sci.* **328**, 131–144 (2004).
41. A. S. Levey, L. A. Stevens, C. H. Schmid, J. L. Zhang, A. F. Castro, H. I. Feldman, J. W. Kusek, P. Eggers, F. Van Lente, T. Greene, J. Coresh; CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration), A new equation to estimate glomerular filtration rate. *Ann. Intern. Med.* **150**, 604–612 (2009).
42. D. M. Lloyd-Jones, L. T. Braun, C. E. Ndumele, S. C. Smith, L. S. Sperling, S. S. Virani, R. S. Blumenthal, Use of risk assessment tools to guide decision-making in the primary prevention of atherosclerotic cardiovascular disease: A special report from the American Heart Association and American College of Cardiology. *J. Am. Coll. Cardiol.* **73**, 3153–3167 (2019).
43. 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, S. Kraemer, L. Kroiss, N. Le, D. Levine, W. Lindsey, B. Lollo, W. Mayfield, M. Mehan, R. Mehler, S. K. Nelson, M. Nelson, D. Nieuwlandt, M. Nikrad, U. Ochsner, R. M. Ostroff, M. Otis, T. Parker, S. Pietrasiewicz,

- D. I. Resnicow, J. Rohloff, G. Sanders, S. Sattin, D. Schneider, B. Singer, M. Stanton, A. Sterkel, A. Stewart, S. Stratford, J. D. Vaught, M. Vrkljan, J. J. Walker, M. Watrobka, S. Waugh, A. Weiss, S. A.-K. Wilcox, A. Wolfson, S. K. Wolk, C. Zhang, D. Zichi, Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLOS ONE* **5**, e15004 (2010).
44. D. Taliun, D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, S. A. G. Taliun, A. Corvelo, S. M. Gogarten, H. M. Kang, A. N. Pitsillides, J. LeFaive, S.-B. Lee, X. Tian, B. L. Browning, S. Das, A.-K. Emde, W. E. Clarke, D. C. Shetty, T. W. Blackwell, A. V. Smith, Q. Wong, X. Liu, M. P. Conomos, D. M. Bobo, F. Aguet, C. Albert, A. Alonso, K. G. Ardlie, D. E. Arking, S. Aslibekyan, P. L. Auer, J. Barnard, R. G. Barr, L. Barwick, L. C. Becker, R. L. Beer, E. J. Benjamin, J. F. Bielak, J. Blangero, M. Boehnke, D. W. Bowden, J. A. Brody, E. G. Burchard, B. E. Cade, J. F. Casella, B. Chalazan, D. I. Chasman, Y.-D. I. Chen, M. H. Cho, S. H. Choi, M. K. Chung, C. B. Clish, A. Correa, J. E. Curran, B. Custer, D. Darbar, M. Daya, M. de Andrade, D. L. DeMeo, S. K. Dutcher, P. T. Ellinor, L. S. Emery, C. Eng, D. Fatkin, T. Fingerlin, L. Forer, M. Fornage, N. Franceschini, C. Fuchsberger, S. M. Fullerton, S. Germer, M. T. Gladwin, D. J. Gottlieb, X. Guo, M. E. Hall, J. He, N. L. Heard-Costa, S. R. Heckbert, M. R. Irvin, J. M. Johnsen, A. D. Johnson, R. Kaplan, S. L. R. Kardia, T. Kelly, S. Kelly, E. E. Kenny, D. P. Kiel, R. Klemmer, B. A. Konkle, C. Kooperberg, A. Kottgen, L. A. Lange, J. Lasky-Su, D. Levy, X. Lin, K.-H. Lin, C. Liu, R. J. F. Loos, L. Garman, R. Gerszten, S. A. Lubitz, K. L. Lunetta, A. C. Y. Mak, A. Manichaikul, A. K. Manning, R. A. Mathias, D. D. McManus, S. T. McGarvey, J. B. Meigs, D. A. Meyers, J. L. Mikulla, M. A. Minear, B. D. Mitchell, S. Mohanty, M. E. Montasser, C. Montgomery, A. C. Morrison, J. M. Murabito, A. Natale, P. Natarajan, S. C. Nelson, K. E. North, J. R. O'Connell, N. D. Palmer, N. Pankratz, G. M. Peloso, P. A. Peyser, J. Pleines, W. S. Post, B. M. Psaty, D. C. Rao, S. Redline, A. P. Reiner, D. Roden, J. I. Rotter, I. Ruczinski, C. Sarnowski, S. Schoenherr, D. A. Schwartz, J.-S. Seo, S. Seshadri, V. A. Sheehan, W. H. Sheu, M. B. Shoemaker, N. L. Smith, J. A. Smith, N. Sotoodehnia, A. M. Stimp, W. Tang, K. D. Taylor, M. Telen, T. A. Thornton, S. A. Tracy, K. J. Van Den Berg, R. S. Vasani, K. A. Viaud-Martinez, S. Vrieze, D. E. Weeks, B. S. Weir, S. T. Weiss, L.-C. Weng, C. J. Willer, Y. Zhang, X. Zhao, D. K. Arnett, A. E. Ashley-Koch, K. C. Barnes, E. Boerwinkle, S. Gabriel, R. Gibbs, K. M. Rice, S. S. Rich, E. K. Silverman, P. Qasba, W. Gan; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, G. M. Peloso, D. A. Nickerson, S. R. Browning, M. C. Zody, S. Zöllner, J. G. Wilson, L. A. Cupples, C. C. Laurie, C. E. Jaquish, R. D. Hernandez, T. D. O'Connor, G. R. Abecasis, Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
45. L. M. Raffield, N. A. Zakai, Q. Duan, C. Laurie, J. D. Smith, M. R. Irvin, M. F. Doyle, R. P. Naik, C. Song, A. W. Manichaikul, Y. Liu, P. Durda, J. I. Rotter, N. S. Jenny, S. S. Rich, J. G. Wilson, A. D. Johnson, A. Correa, Y. Li, D. A. Nickerson, K. Rice, E. M. Lange, M. Cushman, L. A. Lange, A. P. Reiner; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Hematology & Hemostasis TOPMed Working Group, D-Dimer in African Americans: Whole genome sequence analysis and relationship to cardiovascular disease risk in the Jackson Heart Study. *Arterioscler. Thromb. Vasc. Biol.* **37**, 2220–2227 (2017).
46. A. Tan, G. R. Abecasis, H. M. Kang, Unified representation of genetic variants. *Bioinformatics* **31**, 2202–2204 (2015).
47. M. P. Conomos, M. B. Miller, T. A. Thornton, Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet. Epidemiol.* **39**, 276–293 (2015).
48. L. Jiang, Z. Zheng, T. Qi, K. E. Kemper, N. R. Wray, P. M. Visscher, J. Yang, A resource-efficient tool for mixed model association analysis of large-scale data. *Nat. Genet.* **51**, 1749–1755 (2019).
49. T. Sofer, X. Zheng, S. M. Gogarten, C. A. Laurie, K. Grinde, J. R. Shaffer, D. Shungin, J. R. O'Connell, R. A. Durazo-Arviso, L. Raffield, L. Lange, S. Musani, R. S. Vasani, L. A. Cupples, A. P. Reiner; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, C. C. Laurie, K. M. Rice, A fully adjusted two-stage procedure for rank-normalization in genetic association studies. *Genet. Epidemiol.* **43**, 263–275 (2019).
50. M. A. Kamat, J. A. Blackshaw, R. Young, P. Surendran, S. Burgess, J. Danesh, A. S. Butterworth, J. R. Staley, PhenoScanner V2: An expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
51. J. R. Staley, J. Blackshaw, M. A. Kamat, S. Ellis, P. Surendran, B. B. Sun, D. S. Paul, D. Freitag, S. Burgess, J. Danesh, R. Young, A. S. Butterworth, PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics* **32**, 3207–3209 (2016).
- JHS. We thank A. S. Leon, D.C. Rao, J. S. Skinner, and the late J. H. Wilmore for contributions to the planning, data collection, and conduct of the HERITAGE project. Thanks are also expressed to J. Gagnon and T. Rankinen who served as Project Directors of the study. 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. **Funding:** D.H.K. is supported by NHLBI T32 postdoctoral training grant (T32HL007374-40). J.M.R. is supported by NHLBI K23 HL150327-02. U.A.T. is supported by the Ruth L. Kirchstein postdoctoral individual National Research Award (F32HL150992). A.G.B. is supported by NIH DP5-OD029586-01 and is a recipient of a Career Award for Medical Scientists from the Burroughs Wellcome Foundation. D.E.C. is supported by the KL2/Catalyst Medical Research Investigator Training award from Harvard Catalyst (NIH/NCATS Award TR002542). M.D.B. is supported by NHLBI K08HL145095 award. P.N. is supported by R01HL142711 and R01HL127564. M.A.S. is supported by R01HL146462. R.E.G. and M.A.S. are supported by R01NR019628. R.E.G. and J.G.W. are supported by NIH R01 DK081572. R.E.G. is supported by HL132320, HL133870 U24 DK112340, and RF1AG063507. The JHS is supported and conducted in collaboration with Jackson State University (HHSN2682018000131), Tougaloo College (HHSN2682018000141), the Mississippi State Department of Health (HHSN2682018000151/HHSN26800001), and the University of Mississippi Medical Center (HHSN2682018000101, HHSN2682018000111, and HHSN2682018000121) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). Molecular data for the TOPMed program were supported by the National Heart, Lung, and Blood Institute (NHLBI, contract HHSN2682016000341). Genome sequencing for “NHLBI TOPMed: The Jackson Heart Study” (phs000964.v1.p1) was performed at the Northwest Genomics Center (HHSN268201100037C). Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering, was provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN2682018000021). Core support including phenotype harmonization, data management, sample identity QC, and general program coordination was provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN2682018000011). This research was partially funded by National Heart, Lung, and Blood Institute grants HL-45670, HL-47317, HL-47321, HL-47323, and HL-47327, all in support of the HERITAGE Family Study. C.B. is partially funded by the John W. Barton Sr. Chair in Genetics and Nutrition and NIH COBRE grant (NIH P30GM118430-01). M.A.S. is supported by R01HL146462. R.E.G., M.A.S., and J.M.R. are supported by NIH R01 NR019628. **Author contributions:** D.H.K., J.M.R., S.D., U.A.T., and R.E.G. contributed to the original concept of the project, planned these analyses, and formulated the methods. D.H.K., J.M.R., S.D., U.A.T., D.N., M.D.B., Z.-Z.C., D.E.C., D.S., Y.G., A.C., J.G.W., and R.E.G. collected, organized, and contributed to the QC and management of JHS proteomic data, via both the SomaLogic platform and the Olink platform. D.H.K., J.M.R., A.G.B., A.P., Z.Y., U.A.T., D.N., M.D.B., D.E.C., S.D., D.S., C.B., M.A.S., and R.E.G. collected, generated, organized, and contributed to the QC and management of HERITAGE Family Study proteomic and genomic data. D.H.K., D.S., S.D., and R.E.G. conceived and ran ELISA assays and analyzed the resulting data. D.H.K., U.A.T., A.G.B., A.P., Z.Y., P.N., and R.E.G. developed the WGS analysis pipeline and statistical methods across all proteomic data. D.H.K., J.M.R., S.D., and R.E.G. analyzed the data and wrote/revised the article. **Competing interests:** P.N. reports investigator-initiated grants from Amgen, Apple, AstraZeneca, Boston Scientific, and Novartis; personal fees from Apple, AstraZeneca, Blackstone Life Sciences, Foresite Labs, Novartis, and Roche/Genentech; is a cofounder of TenSixteen Bio; is a shareholder of geneXwell and TenSixteen Bio; and spousal employment at Vertex, all unrelated to the present work. The authors declare that they have no other competing interests. **Data and materials availability:** WGS data for TOPMed whole genomes and proteomic data for JHS are available through restricted access via dbGaP. Accession numbers for JHS are phs000964 and phs002256.v1.p1. Summary-level genotype data are available through the BRAVO browser (<https://bravo.sph.umich.edu/>). Full GWAS summary statistics are available for general research use through controlled access at dbGaP accession phs001974: NHLBI TOPMed: Genomic Summary Results for the TOPMed program. HERITAGE Soma5K data are available at <https://motrpc-data.org/related-studies/heritage-proteomics>. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 26 September 2021

Accepted 7 July 2022

Published 19 August 2022

10.1126/sciadv.abm5164

Acknowledgments: We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed. We wish to thank the staff and participants of the