nature communications

Article

https://doi.org/10.1038/s41467-025-67814-1

Cross-ancestry comparison of aptamer and antibody protein measures

Received: 11 February 2025

Accepted: 9 December 2025

Published online: 22 January 2026

Check for updates

Jayna C. Nicholas[1], Daniel H. Katz[2], Usman A. Tahir[3], Catherine L. Debban[4], Francois Aguet[5], Thomas Blackwell[6], Russell P. Bowler[7], K. Alaine Broadaway[1], Jingsha Chen[8], Clary B. Clish[9], Josef Coresh[10], Elaine Cornell[11], Daniel E. Cruz[3], Rajat Deo[12], Margaret F. Doyle[11], Peter Durda[11], Lynette Ekunwe[13], James S. Floyd[14], Dipender Gill[15], Xiuqing Guo[16], Ron C. Hoogeveen[17], Craig Johnson[18], Leslie A. Lange[19], Yun Li[20], Alisa Manning[21,22,23], James B. Meigs[24,25], Michael Y. Mi[26], Josyf C. Mychaleckyj[4], Nels C. Olson[11], Katherine A. Pratte[27], Brucy M. Psaty[28,29,30], Alexander P. Reiner[31], Peifeng Ruan[32], Magdalena Sevilla-Gonzalez[33,34,35], Amil M. Shah[36], Quan Sun[20,37,38,39], Russell P. Tracy[11], Jia Wen[1], Alexis C. Wood[40], James G. Wilson[41], Kristin L. Young[42], Bing Yu[43], Mary R. Rooney[8,44], Ani Manichaikul[4], Ruth Dubin[45], Karen L. Mohlke[1], Stephen S. Rich[4], Jerome I. Rotter[16], Peter Ganz[46], Robert E. Gerszten[3], Kent D. Taylor[16] & Laura M. Raffield[1] ✉

Measures from affinity-proteomics platforms often correlate poorly, challenging interpretation of protein associations with genetic variants and phenotypes. Here, we examine 2157 proteins measured on both SomaScan 7k and Olink Explore 3072 across 1930 participants with genetic similarity to European, African, East Asian, and Admixed American ancestry references. Inter-platform correlation coefficients for these 2157 proteins follow a bimodal distribution (median r = 0.30). We evaluate protein measure associations with genetic variants, and find approximately 25-30% of the signals on each platform are likely driven by protein-altering variants. We highlight 80 proteins that correlate differently across ancestry groups likely in part due to differing protein-altering variant frequencies by ancestry. Furthermore, adjustment for protein-altering variants with opposite directions of effect by platform improves inter-platform protein measure correlation and results in more concordant genetic and phenotypic associations. Hence, protein-altering variants need to be accounted for across ancestries to facilitate platform-concordant and accurate protein measurement.

Genetic and epidemiological studies of circulating protein abundances yield insights that promise to enhance the discovery of clinical biomarkers and drug targets. Recent advances in multiplexed affinity-based proteomics platforms have made biobank-scale studies of the circulating proteome possible[1–6]. Currently, the leading affinity-based proteomics platforms are SomaScan aptamer assays, which use oligonucleotide affinity probes to quantify 1100–11,000 proteins in a plasma/serum sample[7], and Olink proximity extension assays, which employ conjugated antibody-nucleic acid assay probes to quantify 48–5400 proteins[8]. With efficient pipelines for sample analysis, and

A full list of affiliations appears at the end of the paper. ✉e-mail: laura_raffield@unc.edu

lower intra-assay variability compared to many existing mass-spectrometry (MS) proteomics approaches[9,10], both platforms hold great potential for continued use in biobank-scale studies and future clinical applications.

However, existing comparisons of SomaScan and Olink platforms have shown that some protein measures are not concordant (median inter-platform correlation = 0.38–0.50)[11–16], which challenges the accuracy of protein associations observed, especially as there is no "gold-standard" reference measure for most proteins. Additionally, there is no reported method to reconcile protein measure discordance across platforms, preventing effective cross-platform meta-analysis of many protein measures. Furthermore, existing comparisons of SomaScan and Olink protein measures are limited to populations of predominantly one ancestry; it remains unknown if assay performance varies across diverse ancestries. Recognizing which protein measures commonly agree—or disagree—across platforms and ancestries will be critical for equitable application of proteomics data in research and clinics.

Genetic variants may alter circulating protein abundances through many biological mechanisms, including protein-altering variants (PAVs) such as missense variants, which may influence protein stability, or alterations to transcriptional or translational efficiencies and post-translational modifications (PTMs) that could also impact downstream protein stabilities. Genetic variation may also impact protein structure or PTMs in a manner that alters epitope binding to affinity reagents. Protein quantitative trait loci (pQTL), or associations between circulating protein abundances and genetic variants either within or nearby (*cis*-) to the protein-encoding gene or elsewhere in the genome (*trans*-), are commonly detected in large-scale studies[1–6]. The detection of a *cis*-pQTL for a given protein measure is often used as a metric to confirm on-target affinity-probe binding[12,15,16]; existing studies comparing SomaScan and Olink measures have established that the detection of a *cis*-pQTL is a strong predictor of inter-platform correlation coefficients for a given protein[12,15,16].

Nevertheless, interpreting *cis*- and *trans*-pQTL remains complicated, as PAVs may result in pQTL associations that reflect differential probe binding to alternate isoforms[12,17–19] rather than differential protein abundances. Recent MS-based pQTL studies[10,18] have further suggested that some pQTL may reflect isoform differences (although these studies are limited in number and generally encompass fewer proteins/participants). As these genetic variants may vary in frequency across populations, such variants could drive systematic differences in protein measures[20]. Failure to detect and account for such genetic variation may challenge the interpretation of results obtained across ancestries and introduce a source of bias[21]. This holds important implications for many methods, including Mendelian Randomization (MR), which relies on genetic associations reflecting effects on protein abundances to infer protein genetic liability and drug target efficacy.

Here, we evaluate potential genetic drivers of inter-platform and inter-ancestry variation between SomaScan 7k and Olink Explore 3072 protein measurements from 1930 participants from the Multi-Ethnic Study of Atherosclerosis (MESA). For the 2157 proteins measured on both platforms in the same plasma sample, we assess inter-platform correlation and examine the cross-platform concordance of pQTL and epidemiological associations within and across ancestry groups.

Results

Participant demographics

The SomaScan 7k and Olink Explore 3072 platforms were used to generate protein abundance measures at MESA exam 1 (2000–2002) in EDTA-plasma samples from 1930 participants. The mean age of this cohort was 58.2 years (±8.8), with a mean body mass index (BMI) of 28.1 kg/m^2 and a mean eGFR of 80.7 mL/min/1.73m^2 (Supplementary Data 1). Most participants exhibited >0.5 genetic similarity to African (AFR, $n = 439$), Admixed American (AMR, $n = 231$), East Asian (EAS, $n = 277$), or European (EUR, $n = 939$) 1000 Genomes Project reference populations, based on global ancestry proportion estimates (Supplementary Data 1), enabling stratification by genetic similarity for some analyses.

Proteomics platforms

The SomaScan 7k assay in MESA includes 7289 aptamers, targeting 6407 distinct human UniProt IDs. The Olink Explore 3072 assay includes 2941 antibody complexes, targeting 2923 distinct human UniProt IDs. Additionally, SomaScan offers two versions of data: with adaptive-normalization by maximum likelihood performed on quality control samples only (ANML-QC) or performed on all participant samples (ANML-SMP). Similar numbers of protein principal components (PCs) were required to explain 95% of the variation in protein measures from each platform (940 ANML-SMP SomaScan vs. 880 Olink) (Supplementary Fig. 1), despite the different number of proteins measured.

Here, we compared SomaScan 7k and Olink Explore 3072 protein abundances for the 2157 distinct human proteins (as determined by UniProt identifiers) that were assessed on both platforms in 1889 unrelated MESA participants of diverse genetic ancestry with whole genome sequencing data (Fig. 1). Protein measures from each platform were standardized and inverse normal transformed to facilitate comparison of SomaScan Relative Fluorescence Units (RFU) and Olink Normalized Protein Expression (NPX) values. Although both platforms offer multiple affinity probes for a subset of proteins, we present results corresponding to measures obtained from the SomaScan aptamer and one Olink antibody-conjugate probe pair with the highest inter-platform correlation (unless otherwise stated). Results for all probes are in the Supplementary Data 2, 6, and 19–35 and Supplementary Notes 2, 5, and 6.

Correlation and cross-ancestry differences in SomaScan and Olink measures

To evaluate cross-platform agreement of protein measures, we computed Pearson's correlation of transformed protein measures obtained from probes targeting the same protein. Consistent with prior studies, correlation coefficients comparing SomaScan measures with ANML normalization only on QC samples with Olink protein measures followed a bimodal distribution with a median Pearson's correlation of 0.36 (Quartile 1 = 0.09, Quartile 3 = 0.67), and a median of 0.30 (Quartile 1 = 0.02, Quartile 3 = 0.63) when comparing SomaScan ANML-SMP measures and Olink (Fig. 2A and B; Supplementary Fig. 2 and Supplementary Data 2). Proteins with highly discordant inter-platform correlations included Growth arrest and DNA damage-inducible proteins-interacting protein 1 (G45IP, Pearson's $r = -0.38$) and Caspase-8 (Pearson's $r = -0.30$), while proteins with highly concordant inter-platform correlation included Chitosidrase-1 (Pearson's $r = 0.88$) and insulin-like growth factor-binding protein 1 (Pearson's $r = 0.90$). Several proteins showed near-zero correlations, including interleukin-15 (Pearson's $r = 7.2 \times 10^{-6}$) and angiotensin converting enzyme 2 (Pearson's $r = -5.9 \times 10^{-4}$). Correlation coefficients for each protein generally agreed with those observed in prior, single population studies (Supplementary Fig. 4). As prior studies demonstrate ANML-SMP SomaScan normalization lowers intra-assay coefficients of variation[17,22] and maximizes potential for genetic discovery, and per SomaLogic recommendations, we used ANML-SMP SomaScan values throughout remaining analyses.

Given the multi-ancestry composition of MESA, we had the unprecedented opportunity to explore protein measure agreement within ancestry groups and establish whether there were any proteins with significant differences in platform-agreement across ancestry groups. Overall, platforms followed similar patterns of agreement within each ancestry, evidenced by similar median correlations and distributions of Pearson's correlation coefficients observed within

https://doi.org/10.1038/s41467-025-67814-1

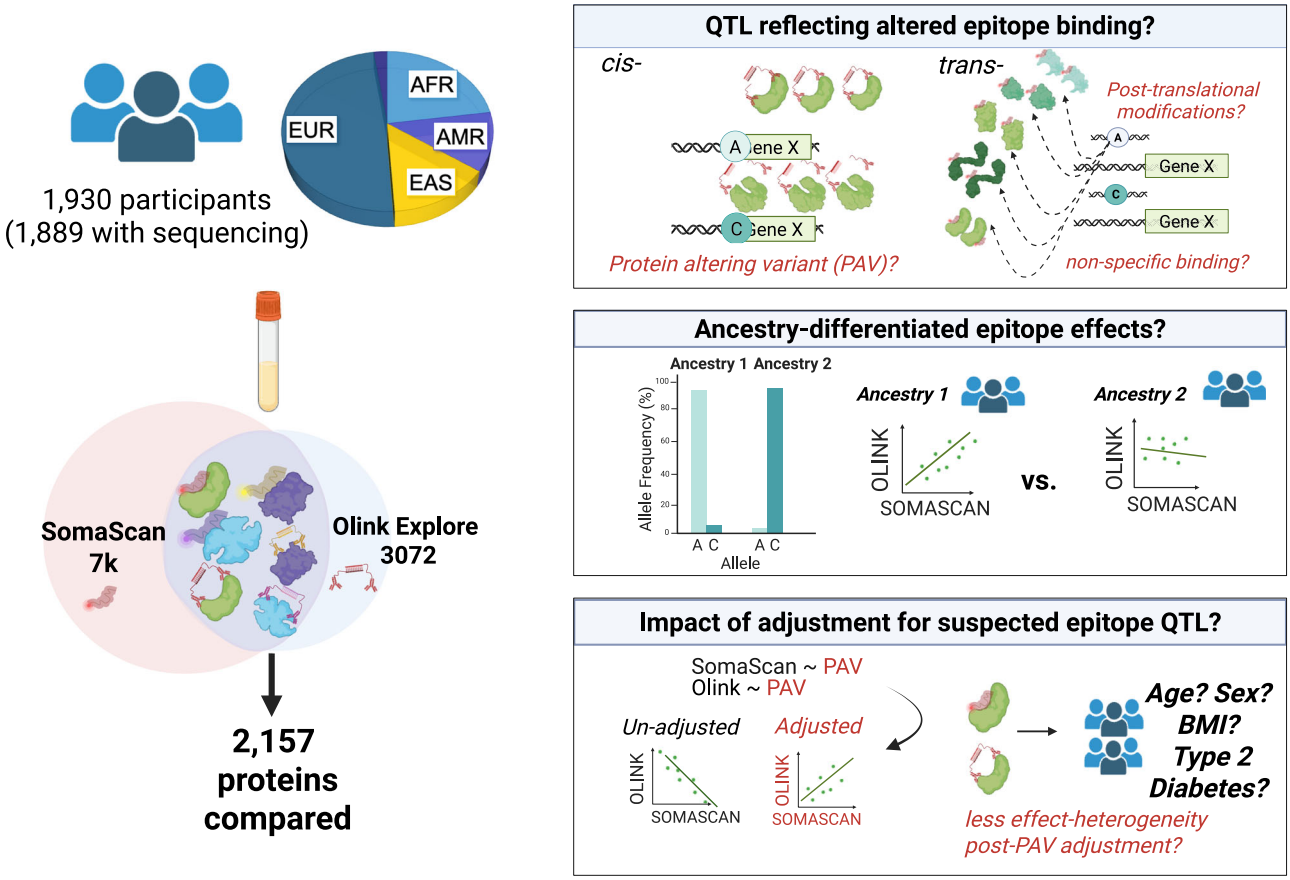


Fig. 1 | Graphical summary of main analyses with (Left) depiction of relative numbers of participants clustering with each ancestry-reference (>0.5 similarity to 1000 G ancestry cluster). A small portion of participants ($n = 44$, 2%) did not cluster with any ancestry reference population. 2157 proteins measured on both SomaScan 7k and Olink Explore 3072 were compared in present analyses. (Right, top panel) Examples illustrating how *cis*- and *trans*- pQTL associations may capture differences in epitope binding rather than differences in protein abundance. (Right, middle panel) Genetic variants which alter epitope binding efficiency and differ in frequency across ancestries may systematically bias affinity-probe measurements in one or more ancestry groups, resulting in systematic differences in measurements and subsequently differences in cross-platform correlation in at least one ancestry. (Right, bottom panel) A central hypothesis of this study is that adjusting an individual's protein measures for variants which likely impact affinity probe binding, rather than abundance, may strengthen accuracy of protein measures on the impacted platform, resulting in stronger inter-platform agreement of protein measures and more concordant downstream epidemiological and genetic associations. Created in BioRender. Nicholas, J. (2025) https://BioRender.com/l96x055.

each ancestry [median Pearson's r AFR = 0.30, AMR = 0.28, EAS = 0.31, EUR = 0.32] (Supplementary Fig. 3 and Supplementary Data 2). While the majority of individual proteins showed similar cross-platform agreement across ancestries, 80 protein targets exhibited significant differences in assay agreement across ancestries (Cochran's Q $P < 1.85 \times 10^{-5}$, Bonferroni adjustment for the total number of probe pairs (2708)) (Tables 1 and 2; Supplementary Data 3), suggesting systematic differences in measurement accuracy across ancestries, on one or both platforms. These differences may be of high relevance to note as research using affinity-based proteomics platforms across diverse populations continues to expand.

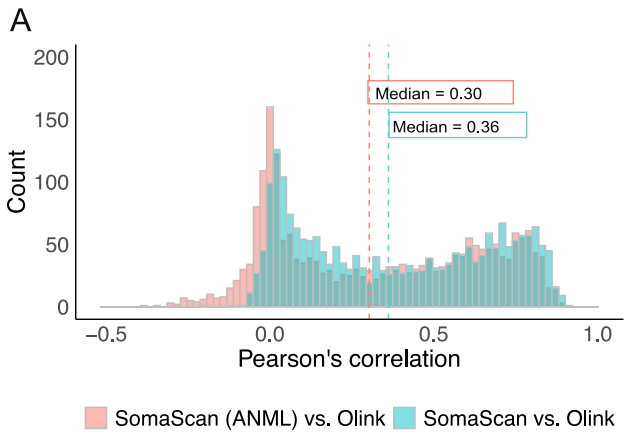
Influence of genetic variants on protein levels

To explore genetic drivers of platform-discordance and cross-ancestry heterogeneity in inter-platform correlation estimates, we mapped associations between genetic variants and SomaScan/Olink protein measures in 1889 MESA participants with whole genome sequencing (WGS) data, excluding first-degree relatives. As *cis*-pQTL are in close proximity to the protein-encoding gene (within 1MB of the transcription start site (TSS)), these associations are most likely to reflect on-target epitope binding, and, in a subset of coding variants, genetic variation that impacts epitope binding (Fig. 1).

Here, we performed ancestry-pooled pQTL mapping, with adjustment for age, sex, site, ten genotype principal components (PCs), and 10 protein PCs on each platform. We show that adjustment for ten protein PCs increases *cis*-pQTL discovery (Supplementary Fig. 5) while minimizing the potential of adjustment for *trans*-pQTL effects (Supplementary Fig. 6, "Methods"). SuSiE[23] fine-mapping of pQTL with individual-level linkage disequilibrium (LD) data was used to define *cis*-pQTL credible sets—which are sets of variants predicted to contain the causal variant and are used here to define genetic signals. Here, we observed significant credible sets (containing at least one variant $p < 5 \times 10^{-8}$) for 840 proteins on SomaScan (Supplementary Data 4) and 1036 proteins on Olink (see "Methods", Table 3, Supplementary Data 5 and Fig. 3A). More proteins showed a *cis*-pQTL on Olink compared to SomaScan (48% vs. 39%, respectively); additionally, we observed nearly twice as many credible sets on Olink compared to SomaScan (although we note this is only among the 2157 proteins measured on both platforms - we did not consider proteins measured on only one platform). 79% and 93% of SomaScan credible sets and 65% and 98% of Olink credible sets reported here were within 1 Mb regions previously reported in deCODE and UK Biobank, respectively (Table 3). A higher proportion of secreted proteins had a significant *cis*-pQTL association compared to intracellular and membrane proteins (Supplementary Fig. 7).

https://doi.org/10.1038/s41467-025-67814-1

While the majority of *cis*-pQTL variants were non-coding on both platforms (Supplementary Fig. 8), we observe several *cis*-pQTL signals driven by protein altering variants (PAVs—most commonly missense variants) on each platform, raising concerns as to whether the signal captures abundances or differences in isoform, as has been shown in other studies[18,24,25]. As PAVs can impact affinity-probe binding, *cis*-pQTL associations driven by PAVs may reflect assay interference due to different affinities for encoded isoforms, particularly when the association is detected on only one platform or detected on both with different effect directions between platforms. Among the 196 and 392 proteins with a *cis*-pQTL association detected with SomaScan or Olink measures only, similar proportions were associated with at least one PAV (25% SomaScan, 30% Olink) (Fig. 3C). We further classified overlapping signal pairs by concordance of lead variant effect-directions (Supplementary Fig. 9 and 10). While most of these signals were led by variants with concordant directions of association across platforms (referred to here as "platform-concordant" signals), we identified 19 proteins with platform-overlapping *cis*-pQTL credible sets for which lead variants had opposite effect-directions ("platform-discordant") (Supplementary Fig. 10). Importantly, the majority of platform-discordant credible sets (15/19) contained a missense variant for the encoded protein (Table 4; Supplementary Fig. 10; Supplementary Data 6 and Supplementary Note 1), suggesting that discordant genetic associations may be driven by platform differences in affinity probe binding. The 18 platform-discordant PAV associations identified in analyses considering all probes per overlapping protein are displayed in Table 4. Encoded changes to protein structure were generally on the surface of the protein; visualizations of the impact of amino acid changes encoded by platform-discordant PAVs on the structure of the protein can be seen in the supplement (Supplementary Note 1 and Supplementary Figs. 36–53).



B

	Minimum	Median	Mean	IQR	Maximum
SomaScan (ANML) vs. Olink	-0.38	0.30	0.33	[0.02, 0.63]	0.90
SomaScan vs. Olink	-0.20	0.36	0.38	[0.09, 0.67]	0.91

Fig. 2 | Inter-platform Pearson's correlation estimates for 2157 proteins measured with both SomaScan 7k aptamer and Olink Explore 3072 antibodies. Results from the comparison of one probe pair per UniProt ID are depicted in this figure. **A** Correlation coefficients between base-normalized SomaScan and Olink Explore measures are plotted in blue. Correlation coefficients obtained by comparison of SomaScan measures with adaptive-normalization by maximum likelihood (ANML) performed on all samples vs. Olink Explore measures plotted in pink. **B** Summary of minimum, median, mean, interquartile range (IQR), and maximum Pearson's *r* values observed in comparisons of SomaScan - measures with ANML performed on all samples vs. Olink (top row) and SomaScan and Olink measures (bottom row). Source data is available in Supplementary Data 2.

Similar to prior studies[12,17], the 644 proteins with a *cis*-pQTL detected on both platforms had higher inter-platform correlation (mean Pearson's $r = 0.58$, SD = 0.20) than proteins with no significant *cis*-pQTL (mean Pearson's $r = 0.21$, SD = 0.30)(Fig. 3C and Supplementary Data 7). Similarly, proteins with Pearson's correlation $r > 0.6$ were most likely to show a platform-concordant *cis*-pQTL signal (Supplementary Fig. 11). Furthermore, the correlation of proteins with a platform-concordant, significant *cis*-pQTL was generally higher (median Pearson's $r = 0.61$, SD = 0.18) than the correlation of proteins with a platform-discordant *cis*-pQTL (median Pearson's $r = 0.40$, SD = 0.20) or a *cis*-pQTL on one platform only (median Pearson's $r = 0.12$, SD = 0.28) (Fig. 3B, C; Supplementary Data 7 and Supplementary Fig. 11). Although we note that some platform-specific *cis*-pQTL were observed even for proteins with highly-correlated measures (Fig. 3C and Supplementary Fig. 11) (which may be due to factors such as differences in intra-assay variability, or differing sensitivities for assays to environmental factors interacting with genetic variation to regulate protein abundance), overall, we propose that *cis*-pQTL presence/absence may be a strong predictor of inter-platform concordance, and that proteins with high inter-platform correlation, for example of Pearson's $r > 0.6$, may hold high potential for meta-analysis of pQTL across SomaScan and Olink platforms.

As PAVs that drive putative epitope effects may differ in frequency between ancestry groups, PAVs may be one driver of inter-ancestry heterogeneity in cross-platform correlation (Fig. 1). Overall, 72 proteins on SomaScan and 84 proteins on Olink were associated with *cis*-

Table 1 | Characterization of *cis*-pQTL status for proteins with heterogeneous inter-platform correlation estimates by ancestry

Category		*N*	cis-pQTL	PAV cis-pQTL	Ancestry-differentiated cis-pQTL	Ancestry-differentiated PAV cis-pQTL	Trans-pQTL	Ancestry-differentiated trans-pQTL	Cis+trans-pQTL	Ancestry-differentiated cis+trans-pQTL
All proteins		2157	1239 (57%)	425 (20%)	563 (26%)	119 (5.5%)	605 (28%)	190 (8.8%)	348 (16%)	58 (2.7%)
Proteins with ancestry-heterogeneous inter-platform correlation estimates	Nominal significance ($p < 0.05$)	611	417 (68%)	150 (25%)	228 (37%)	58 (9.5%)	166 (27%)	54 (8.8%)	115 (19%)	25 (4%)
	Bonferroni significance ($p < 0.05/2708$)	80	69 (86%)	31 (39%)	45 (56%)	13 (16%)	29 (36%)	16 (20%)	24 (30%)	10 (13%)

Counts of proteins with nominal ($p < 0.05$) and Bonferroni significant (Cochran's Q $p < 0.05/2,708$, with correction for 2708 probe pairs tested) differences in inter-platform correlation estimates across ancestries according to the Cochran's Q test of heterogeneity (rows, one-sided test of significance) and proteins which associate with a *cis*-pQTL, a *cis*-pQTL driven by a protein altering variant (PAV), a *cis*-pQTL with ancestry-differentiated effect allele frequencies, a PAV-driven *cis*-pQTL with ancestry-differentiated effect allele frequencies, a *trans*-pQTL, an ancestry-differentiated *trans*-pQTL, a *cis*- and *trans*-pQTL, or an ancestry-differentiated *cis*- and *trans*-pQTL on either platform (columns). pQTL were generated with linear regression under an additive model (two-sided test of significance). Variants with ancestry-differentiated allele frequencies were determined according to the Chi-square test (one-sided test) using allele frequencies derived in each contributing ancestry group (significance defined as a X^2 value in the 75th percentile). Counts displayed reflect one probe per UniProt ID.

https://doi.org/10.1038/s41467-025-67814-1

Table 2 | 20 proteins with the greatest heterogeneity in inter-platform correlation estimates across contributing ancestry groups

Protein summary				Probe association with ancestry-differentiated PAV cis-pQTL		Correlation estimates (Pearson's *r*) per ancestry and cross-ancestry heterogeneity						
UniProt	**Target full name**	**SeqId**	**OlinkID**	**SomaScan**	**Olink**	**ALL**	**AFR**	**AMR**	**EAS**	**EUR**	**Q**	**Q *p* value**
P05154	Plasma serine protease inhibitor	seq.3389.7	OID30763	TRUE	TRUE	−0.27	−0.38	−0.10	0.04	−0.28	39.41	1.42E-08
P12318	Low affinity immunoglobulin gamma Fc region receptor II-a	seq.3309.2	OID20391	FALSE	FALSE	0.30	0.18	0.38	−0.02	0.46	70.13	4.01E-15
Q96HD1	Cysteine-rich with EGF-like domain protein 1	seq.7628.40	OID30619	FALSE	FALSE	0.17	−0.01	0.18	0.25	0.26	25.10	1.47E-05
Q86SJ6	Desmoglein-4	seq.22568.7	OID21325	FALSE	FALSE	0.07	0.02	0.02	0.10	0.31	37.32	3.93E-08
Q9Y3E2	BolA-like protein 1	seq.15370.5	OID30837	FALSE	FALSE	0.28	0.37	0.13	0.49	0.22	31.39	7.03E-07
P17813	Endoglin	seq.4908.6	OID20287	FALSE	FALSE	0.39	0.40	0.38	0.16	0.47	26.78	6.54E-06
Q9H4A9	Dipeptidase 2	seq.8327.26	OID21305	FALSE	FALSE	0.31	0.46	0.34	0.17	0.19	32.83	3.50E-07
Q6UWN8	Serine protease inhibitor Kazal-type 6	seq.5731.1	OID21450	TRUE	FALSE	0.33	0.23	0.37	0.55	0.45	30.38	1.15E-06
Q03405	Urokinase plasminogen activator surface receptor	seq.2652.15	OID20764	TRUE	TRUE	0.44	0.24	0.61	0.64	0.47	55.64	5.02E-12
P41439	Folate receptor gamma	seq.15495.9	OID21485	FALSE	FALSE	0.47	0.61	0.24	0.43	0.42	34.70	1.41E-07
P15289	Arylsulfatase A	seq.3583.54	OID21138	FALSE	FALSE	0.40	0.47	0.49	0.43	0.25	28.56	2.76E-06
P02008	Hemoglobin subunit zeta	seq.6919.3	OID30307	FALSE	FALSE	0.40	0.52	0.29	0.50	0.29	31.88	5.56E-07
P04217	Alpha-1B-glycoprotein	seq.16561.9	OID30771	FALSE	FALSE	0.54	0.45	0.29	0.58	0.61	38.09	2.71E-08
Q13790	Apolipoprotein F	seq.12370.30	OID30701	FALSE	FALSE	0.49	0.29	0.48	0.67	0.55	50.38	6.62E-11
A6NHS7	MANSC domain-containing protein 4	seq.9578.263	OID30141	FALSE	FALSE	0.45	0.31	0.30	0.60	0.48	32.80	3.55E-07
P51858	Hepatoma-derived growth factor	seq.16758.96	OID21455	FALSE	FALSE	0.48	0.65	0.31	0.48	0.45	36.83	4.99E-08
Q9NQ38	Serine protease inhibitor Kazal-type 5	seq.8028.22	OID21148	FALSE	TRUE	0.49	0.32	0.56	0.62	0.51	29.97	1.40E-06
P36959	GMP reductase 1	seq.19254.125	OID20641	TRUE	FALSE	0.42	0.65	0.35	0.60	0.32	71.80	1.75E-15
O14791	Apolipoprotein L1	seq.11510.31	OID30708	TRUE	TRUE	0.54	0.33	0.54	0.58	0.55	27.54	4.54E-06
O15455	Toll-like receptor 3	seq.16918.198	OID20612	FALSE	FALSE	0.58	0.38	0.55	0.63	0.60	29.85	1.48E-06

Displayed here are 20 out of the 80 proteins with Bonferroni-significant heterogeneity in inter-platform Pearson's *r* correlation estimates across the 4 ancestry groups represented in this study. The 20 proteins displayed in this table show the greatest degree of heterogeneity in per-ancestry Pearson's *r* correlations according to the lowest Cochran's Q *p* values (generated in Cochran's Q test of heterogeneity, one-sided test of significance), and the greatest range in Pearson's *r* values across ancestry groups. "*cis*-pQTL summary" columns denote whether the SomaScan protein measure (SeqID) or the Olink protein measure (OlinkID) associates with an ancestry-differentiated protein altering variant (PAV) in *cis*- (i.e., *cis*-pQTL significantly associates with protein measure at traditional genome-wide significance ($p < 5 \times 10^{-8}$) and is driven by a PAV with ancestry-differentiated allele frequencies according to Chi-square test (one-sided test of significance) using allele frequencies derived in each contributing ancestry group, with significance defined as a X^2 value in the 75th percentile). pQTL were generated with linear regression under an additive model (two-sided test of significance). Additional details about these proteins, as well as the remaining 60 proteins with Bonferroni-significant ancestry-heterogeneous inter-platform correlations can be found in Supplementary Data 3.

AFR African, *AMR* Admixed American, *EAS* East Asian, *EUR* European Ancestry.

https://doi.org/10.1038/s41467-025-67814-1

Table 3 | Summary of pQTL signals identified by platform

		SomaScan	Olink
cis-pQTL	Total *N* probes/UniProt IDs assessed[a]	2157	2157
	N UniProt IDs with a significant *cis*-pQTL credible set	840	1036
	Number of credible sets	1476	1779
	Positive directions of effect	659	748
	Negative directions of effect	817	1031
	Containing PAV[b] (for protein-encoding gene)	292	359
	Containing ancestry-differentiated[c] PAV	74	84
	Containing GTEx cis-eQTL for protein-encoding gene	459	587
	deCODE cis-pQTL in region	1156	1163
	UKB cis-pQTL in region	1375	1753
	Containing MS-pQTL[d] lead variant	7	11
	MS-pQTL[d] in region	49	78
	Containing MS potential non-abundance QTL[e] lead variant	3	5
	MS non-abundance QTL[e] in region	35	53
	Containing GWAS catalog reported lead variant	1337	1779
trans-pQTL	N UniProt IDs with a significant trans-pQTL credible set ($p < 1 \times 10^{-11}$)	424	262
	Number of credible sets	529	315
	Positive directions of effect	341	182
	Negative directions of effect	188	133
	Containing PAV	200	65
	Containing ancestry-differentiated variant	133	97
	Maps to pleiotropic region[f]	276	158
	Maps to platform shared pleiotropic region[g]	140	151
	Maps to platform "specific" pleiotropic region[h]	136	7
	deCODE trans-pQTL in region	310	139
	UKB trans-pQTL in region	172	279
	Containing MS-pQTL[d] lead variant	1	1
	MS-pQTL[d] for protein target in region	4	6
	Containing MS- potential non-abundance QTL[e] lead variant	0	0
	MS-potential non-abundance QTL[e] for protein target in region	0	0

Significant *cis*-pQTL and *trans*-pQTL credible set characteristics and overlap with previously identified GTEx version 8 expression quantitative trait loci (eQTL)[75], affinity-based pQTL identified in deCODE ($n = 35{,}559$)[1] or UK BioBank ($n = 54{,}219$)[5], mass-spectrometry based pQTL (MS-pQTL) and non-abundance MS-PAVs from the QMDiab Study ($n = 345$ participants)[10] and/or from the Tarkin Study ($n = 1260$ participants)[18] and GWAS catalog variants[75]. pQTL were generated with linear regression under an additive model (two-sided test of significance). *cis*-pQTL were considered significant at a traditional genome-wide significance *p*-value threshold ($p < 5 \times 10^{-8}$). *trans*-pQTL were considered significant at a traditional genome-wide significance threshold further corrected for the 2708 probe pairs tested ($p < 1 \times 10^{-11}$). Signals (defined by credible sets) identified in the present study were compared to previously reported signals either with a signal level approach, in which we assessed overlap between previously reported lead variants and credible sets identified in the present study, or with a region-based approach, where we assessed whether previously reported lead variants fell within a 1 Megabase (Mb) window of the lead variants for the signals identified in the present study.

[a]Counts correspond to results from one probe per UniProt ID.

[b]*PAV* protein altering variant.

[c]Ancestry-differentiated = EAF X^2 value in 75th percentile. Variants with ancestry-differentiated allele frequencies were determined with the Chi-square test using allele frequencies derived in each contributing ancestry group (significance defined as a X^2 value in the 75th percentile, one-sided test).

[d]Plasma mass-spectrometry (MS) pQTL lead variants were obtained from two previous studies[10,18].

[e]Plasma MS-potential non-abundance QTL lead variants were generated in the QMDiab study[10] and are considered most likely to impact peptide measurements (and potentially, protein measurements), rather than circulating protein abundances.

[f]Pleiotropic regions are defined as signal regions collectively associated with at least 5 proteins on the platform.

[g]Platform-shared pleiotropic regions are defined as platform-overlapping regions associated with at least 5 proteins (but not necessarily the same proteins) on both platforms.

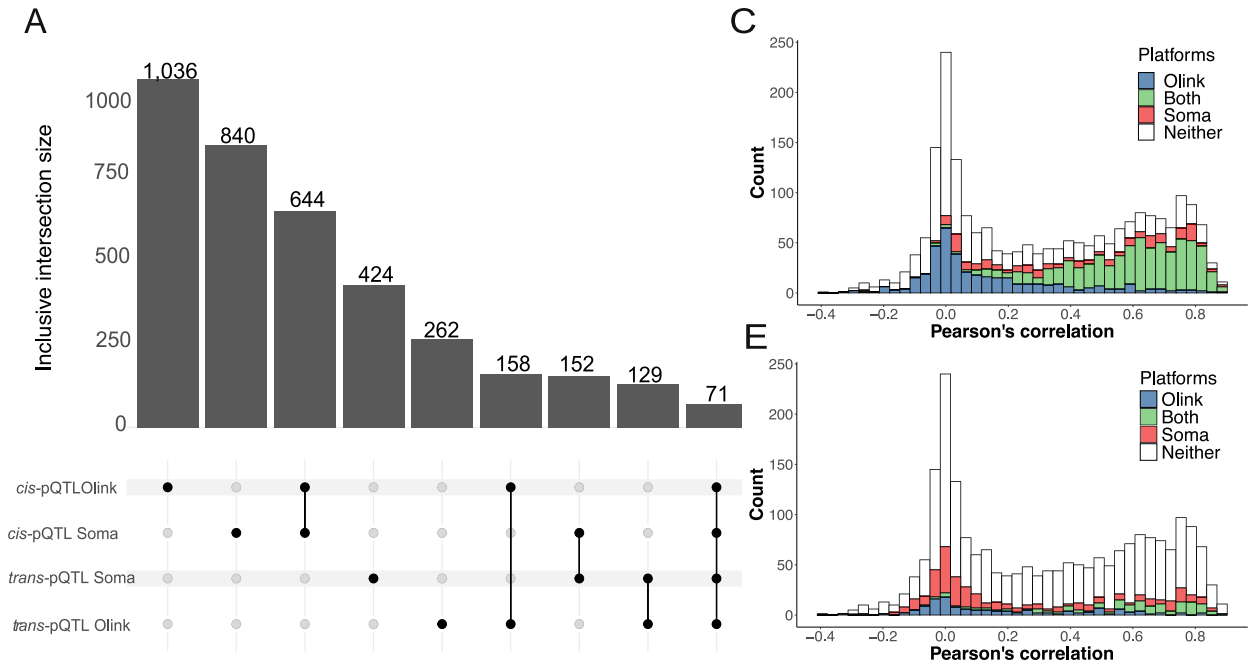
[h]Platform "specific" pleiotropic regions are defined as regions associated with at least 5 proteins on one platform, and 0 or 1 proteins on the other platform.

pQTL containing highly ancestry-differentiated PAVs (Chi-Square >202, the 75th X^2 percentile among variants tested in MESA). Indeed, we observed a higher rate of ancestry-differentiated *cis*-pQTL containing PAVs among the 80 proteins with Bonferroni-significant heterogeneity in inter-platform correlation estimates between ancestries (Cochran's Q $P < 1.8 \times 10^{-5}$, significant at $\alpha = 0.05$, with correction for 2708 probe pairs tested) (16% compared to 5.5% (118/2157) of proteins overall) (Table 1). While proteins with ancestry-differentiated correlation may have multiple contributors to inter-platform variation, including ancestry-differentiated *cis*-pQTL (which may be non-coding) or *trans*-pQTL (which 56% and 20% of proteins with ancestry-heterogeneous correlation were associated with, respectively) (Table 1), confounding of ancestry with environmental and lifestyle factors, etc, we focus on the potential effects of ancestry-differentiated PAVs on protein measures here.

Orthogonal genetic evidence to characterize *cis*-pQTL

Incorporating orthogonal lines of genetic evidence, such as expression quantitative trait loci (eQTL), which are statistically significant genetic variant associations with transcript abundance, may increase confidence that pQTL associations capture true alterations in protein abundance. When a pQTL variant is also associated with a gene's transcript abundance, that may support the variant has a true impact on protein abundance. Consistent with prior reports[17,19], we found that platform-concordant *cis*-pQTL signals were most likely to contain a GTEx v8 eQTL lead variant for the protein-encoding gene in the

https://doi.org/10.1038/s41467-025-67814-1



B

	Median Correlation	N Proteins	N PAV-Associated (%)
Platform-Concordant cis-pQTL	0.61	515	134 (26%)
Platform-Discordant cis-pQTL	0.40	19	16 (84%)
SomaScan-only cis-pQTL	0.41	196	48 (25%)
Olink-only cis-pQTL	0.13	392	119 (30%)

D

	Median Correlation	N Proteins	Pleioptropic (%)
trans-pQTL on both platforms	0.55	129	67 (52%)
SomaScan-only trans-pQTL	0.04	295	178 (60%)
Olink-only trans-pQTL	0.09	133	55 (41%)

Fig. 3 | Comparison of genetic signals detected for 2157 proteins measured on both SomaScan and Olink Explore in 1889 MESA participants. A UpSet plot depicting total counts of proteins with *cis*- and *trans*-pQTL on each platform. Note that, as this is an inclusive UpSet plot, a given protein can be represented in more than one bar. **B** Table depicting the percentage of proteins associated with a platform-concordant, platform-discordant, SomaScan specific, or Olink specific *cis*-pQTL, which were also associated with a protein-altering variant (PAV). **C** Histogram of Pearson's correlation coefficients (*r* values), with bars colored to indicate which proteins in the bin had a genome-wide significant *cis*-pQTL ($p < 5 \times 10^{-8}$) detected on SomaScan, Olink, both, or neither platform. **D** Table depicting the percentage of proteins associated with a pleiotropic *trans*-pQTL (signal significantly ($p < 1 \times 10^{-11}$) associated with ≥5 protein measures on a given platform, not necessarily the same proteins) among the proteins with a *trans*-pQTL on both platforms or one platform only. **E** Histogram of Pearson's correlation coefficients (*r* values), with bars colored according to whether proteins in bin had a significant *trans*-pQTL detected on SomaScan, Olink, both, or neither platform. Source data is available in Supplementary Data 4, 5, 9, 10 (**A**), Supplementary Data 2, 4, 5 (**B**, **C**) and Supplementary Data 2, 9, 10 (**D**, **E**).

credible set, compared to platform-specific or platform-discordant signals. Thirty-six percent of platform-concordant pQTL credible set pairs, for 207 unique proteins, contained a GTEx eQTL with a concordant direction of effect, compared to 18% and 22% of SomaScan and Olink "specific" *cis*-pQTL, respectively (Supplementary Data 7). Among the 15 platform-discordant, PAV-driven *cis*-pQTL, 4 had a reported genetic effect on transcript abundance in one or more tissues; the eQTL direction of effect agreed with the SomaScan pQTL direction of effect for 2 proteins (CTSS, HDGF) and with Olink *cis*-pQTL effect-direction for 2 proteins (HNMT, ACP1). Concordant eQTL evidence may support that pQTL associations detected on one platform only capture true differences in abundance. 9% (13/139) of SomaScan-specific PAV-containing pQTL credible sets also contain a concordant GTEx *cis*-eQTL, compared to 6% (14/208) of Olink-specific PAV credible sets (Supplementary Data 7). The majority of *cis*-pQTL credible sets identified on each platform contained a GWAS catalog lead variant (Table 3; Supplementary Data 4 and 5). While we consider the presence of a directionally-concordant eQTL within a pQTL signal as a helpful indication to suggest that a pQTL captures differences in protein abundance, we note that the *absence* of a shared eQTL is challenging to interpret, as there may be an undetected eQTL due to power, context specificity of eQTL effects, etc., or the variant may alter protein abundance through a mechanism other than expression.

https://doi.org/10.1038/s41467-025-67814-1

Table 4 | Proteins associated with platform-discordant cis-pQTL containing a protein altering variant (PAV) in discordant credible set

Credible set summary						Overall protein measure correlation (Pearson's *r*)		PAV cis-pQTL summary statistics						Allele frequency per ancestry group					
								SomaScan			Olink								
Variant ID	Entrez gene name	SeqID	OlinkID	Amino acid change* (Ref/Alt)	UniProt	Without PAV adjustment	With PAV adjustment	*B*	SE	PVAL	B	SE	PVAL	All	AFR	AMR	EAS	EUR	Ancestry differentiated variant
chr15:78944905:C:G	CTSH**	seq.8465.52	OID20113	C/S	P09668	0.28	0.45	0.42	0.03	2.18E-37	−0.50	0.03	6.47E-49	0.48	0.28	0.49	0.12	0.68	TRUE
chr19:43665370:T:C	PLAUR	seq.2652.15	OID20764	T/A	Q03405	0.43	0.57	−2.10	0.07	2.03E-155	0.61	0.07	2.58E-17	0.03	0.10	0.01	0.00	0.00	TRUE
chr2:138002079:C:T	HNMT	seq.23178.95	OID20975	T/I	P50135	0.32	0.47	0.66	0.05	5.52E-39	−1.05	0.04	4.52E-146	0.07	0.01	0.08	0.03	0.11	FALSE
chr16:12803721:G:T	CPPED1	seq.21108.5	OID21053	A/D	Q9BRF8	−0.02	0.47	−1.20	0.02	1.00E-300	0.66	0.02	1.20E-148	0.33	0.61	0.28	0.37	0.20	TRUE
chr1:150755063:G:A	CTSS	seq.3181.50	OID21056	R/W	P25774	0.41	0.50	0.35	0.03	8.43E-37	−0.37	0.03	1.92E-43	0.35	0.28	0.40	0.35	0.38	FALSE
chr6:41198411:A:G	TREML2	seq.5736.1	OID21120	V/A	Q5T2D2	0.67	0.76	0.57	0.05	3.97E-33	−0.58	0.04	1.50E-54	0.10	0.03	0.10	0.07	0.13	FALSE
chr7:100374211:A:G	PILRA**	seq.10816.150	OID21129	R/G	Q9UKJ1	−0.39	0.26	−1.35	0.02	1.00E-300	0.92	0.02	2.98E-300	0.34	0.14	0.53	0.60	0.32	TRUE
chr1:156743766:G:A	HDGF	seq.16758.96	OID21455	P/L	P51858	0.48	0.68	−0.33	0.02	1.69E-42	1.02	0.02	1.00E-300	0.22	0.11	0.14	0.10	0.32	FALSE
chr4:55113391:C:T	KDR	seq.3651.50	OID21497	V/I	P35968	0.57	0.59	0.26	0.04	5.79E-09	−0.30	0.04	3.25E-12	0.14	0.24	0.09	0.15	0.09	FALSE
chr1:10413143:G:A	PGD	seq.4187.49	OID30344	D/N	P52209	0.49	0.50	−0.21	0.03	4.60E-12	0.74	0.06	2.65E-39	0.04	0.05	0.04	0.06	0.03	FALSE
chr2:277003:A:G	ACP1	seq.3858.5	OID30524	Q/R	P24666	0.15	0.36	1.25	0.01	1.00E-300	−0.17	0.03	2.74E-07	0.29	0.23	0.24	0.26	0.35	FALSE
chr14:20781965:G:A	RNASE6	seq.5646.20	OID30689	R/Q	Q93091	0.35	0.63	1.16	0.02	1.00E-300	−0.24	0.03	1.45E-18	0.23	0.10	0.26	0.41	0.23	FALSE
chr22:36265284:G:A	APOL1	seq.11510.31	OID30708	E/K	O14791	0.54	0.59	0.51	0.03	9.77E-49	−0.22	0.03	1.12E-10	0.29	0.63	0.14	0.18	0.20	TRUE
chr19:44908822:C:T	APOE	seq.2937.10	OID30727	R/C	P02649	0.36	0.40	−0.26	0.05	3.63E-08	0.51	0.06	1.71E-19	0.08	0.10	0.02	0.10	0.08	FALSE
chr8:27599962:C:T	CLU	seq.4542.24	OID30732	D/N	P10909	0.12	0.16	0.66	0.10	5.66E-11	−1.51	0.12	2.61E-32	0.01	0.03	0.00	0.00	0.00	FALSE
chr19:15472052:C:T	PGLYRP2	seq.5601.2	OID30742	R/Q	Q96PD5	0.40	0.61	0.26	0.04	3.78E-12	−1.01	0.03	5.63E-199	0.21	0.17	0.26	0.28	0.19	FALSE
chr11:57614516:G:A	SERPING1**	seq.13710.6	OID30777	V/M	P05155	0.28	0.45	1.12	0.03	1.21E-217	−0.18	0.03	1.98E-08	0.20	0.11	0.14	0.13	0.28	FALSE
chr11:9748015:C:G	SWAP70	seq.13552.7	OID31477	Q/E	Q9UH65	0.40	0.50	−0.52	0.03	8.05E-76	0.22	0.03	3.17E-16	0.41	0.49	0.38	0.57	0.33	FALSE

Protein-altering variants (PAVs) significantly associated (credible-set lead variant $p < 5 \times 10^{-8}$, traditional genome-wide significance threshold) with protein measures in *cis*- on both SomaScan and Olink, with different directions of effect across platforms, for all probe pairs measuring a protein measured on both platforms. 15 PAVs were found in main analyses (considering one probe per platform for each protein), while 3 platform-discordant PAV associations were found only when considering all probes targeting a protein measured on both platforms (marked with **). pQTL summary statistics were generated with linear regression under an additive model (two-sided test of significance). Ancestry-differentiated allele frequencies were determined according to the Chi-square test (one-sided) using allele frequencies derived in each contributing ancestry group (significant difference in allele frequencies across ancestries defined as a X^2 value in or above the 75th percentile).

AFR African, *AMR* Admixed American, *EAS* East Asian, *EUR* European Ancestry.

*Amino acid changes encoded by the PAV, as obtained from Variant Effect Predictor.

**Indicates proteins with a platform-discordant PAV association found only when considering all probes targeting a protein measured on both platforms.

While MS-based pQTL studies are relatively limited in size and number (with some of the broadest studies, used for pQTL comparison here, currently including 1260 Americans and 1980 proteins[18], or 325 participants and ~2800 proteins[10]), comparing affinity-based pQTL to pQTL identified in MS studies may facilitate signal interpretation. A small proportion of affinity-identified pQTL (0.4% SomaScan, 0.6% Olink) harbor a previously reported MS-pQTL lead variant in the credible set (Table 3; Supplementary Data 8, 4, and 5). MS-pQTL studies may also help differentiate PAV-containing pQTL that do not reflect differences in overall protein abundance, but may, for example, reflect differences in specific peptide abundances, altered affinity reagent binding, or technical errors. Platform-discordant signals for 2 proteins, APOL1 and PGLYRP2, had evidence of impacting specific peptides but not overall protein abundance in an MS study[10] (Supplementary Data 4, 5, and 8); many of the proteins and pQTL with platform-discordant signals were not studied in the relatively limited current MS-pQTL literature but will hopefully be interrogated in future studies. Hence, MS pQTL findings may be used to better disentangle true biological abundance signals from signals possibly capturing isoform abundance differences, or other effects.

trans-pQTL identification and pleiotropy

Associations between protein levels and genetic variants that are not in the vicinity of the protein-encoding gene are also commonly detected in large-scale studies and may reflect a host of biological or technical factors. Upon mapping *trans*-pQTL (pQTL >1 Mb of protein-encoding gene TSS) for SomaScan and Olink protein measures, we identified 424 significant *trans*-pQTL credible sets (credible set contains at least one variant with $p < 1 \times 10^{-11}$) for 529 distinct proteins on SomaScan, and 315 for 262 proteins on Olink (Fig. 3A; Table 3; Supplementary Data 9, 10 and 11). While proteins with a *trans*-pQTL on both platforms exhibited higher inter-platform correlation ($N = 129$, mean Pearson's $r = 0.55$, SD = 0.26) (Fig. 3D, E), proteins with a significant *trans*-pQTL on one platform only generally exhibited lower inter-platform correlation estimates ($N = 428$, median Pearson's $r = 0.07$, SD = 0.29) than proteins with no *trans*-pQTL at all ($N = 1600$, median Pearson's $r = 0.34$, SD = 0.32) (Supplementary Data 11).

trans-pQTL associated with several protein measures may, among many other potential mechanisms, reflect non-specific affinity-probe binding or alterations to proteins involved in depositing PTMs that may interfere with affinity-probe binding. We identified 10 pleiotropic *trans*-pQTL regions that were associated with at least 5 proteins on each platform (although not necessarily the same proteins on both platforms), including well-characterized *ABO, FUT2*, and *F12*. Nine *trans*-pQTL regions were pleiotropic for SomaScan protein measures only (significant *trans*-pQTL associations with ≥5 proteins on SomaScan, 0 or 1 on Olink), including *APOE, F5*, and *HPX*. The *FUT6; FUT3* region was identified as a platform-specific pleiotropic region for Olink protein measures in the main analysis, but was also pleiotropic for SomaScan protein measures when considering all probes for each protein target (Supplementary Data 12 and 13). While proteins associated with a pleiotropic region on both platforms had highly correlated measures ($N = 67$, median Pearson's $r = 0.62$, SD = 0.26), the 136 proteins associated with a *trans*-pQTL region pleiotropic for SomaScan only exhibited low inter-platform correlation (median Pearson's $r = 0.02$, SD = 0.21).

Impact of limit of detection on protein measure correlation and pQTL detection

Limit of detection (LOD) refers to the lowest quantity of protein that can reliably be detected by an assay. The LOD for an assay may impact protein quantification precision or pQTL detection, particularly for low-abundance proteins. Here, the majority (2155/2157) of SomaScan proteins had >50% of measures above the company-reported LOD, while 63% (1361/2157) of Olink proteins measured had >50% of measures above the LOD (calculated per company-recommendations) (Supplementary Fig. 12 and Supplementary Data 14). While proteins with >50% measures above LOD were modestly correlated ($N = 1361$, median Pearson's $r = 0.43$, SD = 0.32), proteins with <50% measures above LOD exhibited low inter-platform correlation ($N = 796$, median Pearson's $r = 0.04$, SD = 0.32), consistent with prior reports (Supplementary Data 15 and 16)[26]. 83% and 85% of proteins with a *cis*-pQTL or *trans*-pQTL detected on Olink had at least 50% of measures above the LOD, respectively. This suggests that additional scrutiny of how to best handle protein measures below LOD measures may be necessary for pQTL and other proteomics analyses, especially in cross-platform studies.

Epidemiological protein associations

Proteomics measurements are increasingly used to understand how proteins influence traits and diseases; however, using protein measures that do not necessarily capture overall protein abundances may mask some associations. Here, similar proportions of SomaScan and Olink protein measures were associated with age, sex, BMI, and type 2 diabetes (T2D) case/control status at the Bonferroni significance threshold ($P < 1.8 \times 10^{-5}$, significant at $\alpha = 0.05$, with correction for 2708 probe pairs tested) (Supplementary Data 17 and 18). Across all proteins, effect sizes for each phenotype were modestly correlated (Pearson's r of approximately 0.6) (Supplementary Fig. 32). Among each phenotype, proteins with a Bonferroni significant phenotype association on both platforms were more likely to associate with a *cis*-pQTL on both platforms (Supplementary Data 17). For age, sex, and BMI, we observed a small number of proteins with platform-discordant, Bonferroni significant associations on both platforms (6 to 25, Supplementary Data 17), suggesting that differences in protein measurements may indeed result in differences in epidemiological associations.

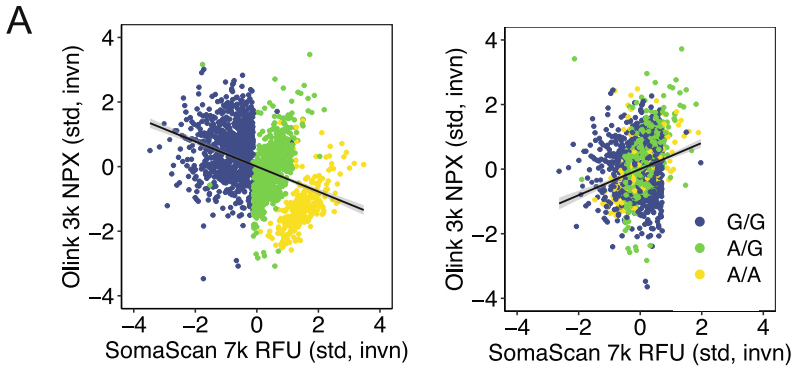
Adjustment for ancestry-differentiated, platform-discordant pQTL improves concordance of genetic and epidemiological associations

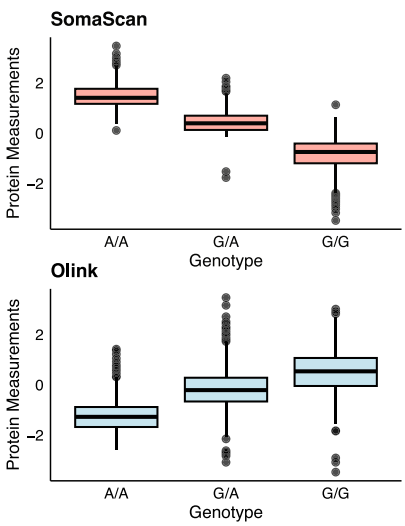
If the platform discordance for a given protein measure is largely driven by genetic variation, we hypothesized that adjustment for PAVs putatively driving inter-platform variation may improve concordance of protein measures and strengthen downstream epidemiological and genetic associations not reflecting assay interference. For the proteins associated with a platform-discordant *cis*-pQTL PAV, we adjusted each platform's measures for copies of the PAV and computed inter-platform correlation estimates with residual measures. Prior to PAV-adjustment, the mean inter-platform correlation estimate for the 15 proteins was 0.38 (SD = 0.18), with estimates ranging from −0.015 to 0.67. Post PAV-adjustment, the mean inter-platform correlation increased to 0.51 (SD = 0.15), with estimates ranging from 0.15 to 0.76 (results for all probes described in Supplementary Note 2 and Supplementary Data 6). Inter-platform correlation improved for all these proteins (mean improvement = 0.14) (Supplementary Fig. 13).

Next, we performed *cis*-pQTL mapping, adjusting for PAV copies as a covariate in mapping (Supplementary Data 19 and 20). We found that a subset of significant *cis*-pQTL associations was strengthened (8 SomaScan, 4 Olink) (Supplementary Data 21 and 22), and 2 *cis*-pQTL associations on SomaScan were significant only when adjusting for the platform-discordant PAV (Supplementary Data 23) (Supplementary Figs. 17, 18 and Supplementary Note 3), demonstrating that adjustment for putative assay interference effects may strengthen the potential to detect true associations with protein abundances, for a subset of proteins. Genomic regions surrounding platform-discordant PAVs prior to and after adjustment for the PAV in pQTL mapping can be viewed in Supplementary Figs. 15–27.

Additionally, we evaluated the impact of PAV adjustment on protein-phenotype associations by regressing adjusted protein measures

https://doi.org/10.1038/s41467-025-67814-1





C

PAV Summary Statistics				EAF				
Platform	B	SE	P-Value	AFR	AMR	EAS	EUR	ALL
Soma	-1.35	0.02	$1x10^{-300}$	0.14	0.53	0.60	0.32	0.35
Olink	0.92	0.02	$3x10^{-306}$					

D

	Correlation Without PAV Adjustment				Correlation With PAV Adjustment			
Protein	AFR	AMR	EAS	EUR	AFR	AMR	EAS	EUR
PILRA	-0.25	-0.35	-0.47	-0.39	0.04	0.37	0.46	0.27

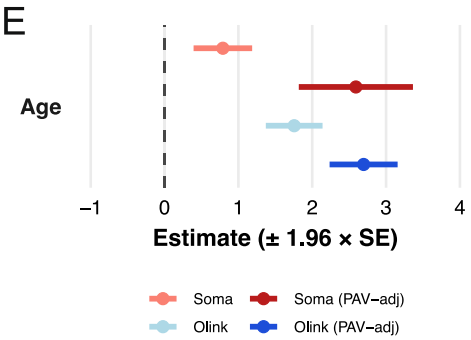


Fig. 4 | Adjustment for PILRA-associated PAV improves inter-platform correlation of measures and strengthens association with age. A Olink vs. SomaScan standardized ("std") and inverse normal transformed normalized ("invn") protein expression (NPX) and relative fluorescence units (RFU), respectively, without adjustment for PAV rs1859788 (chr7:100374211:A:G; p.Arg78Gly) (left) and with adjustment for the PAV (right). Shaded bands indicate the 95% confidence interval for the linear regression fit. **B** SomaScan (pink) and Olink (blue) PILRA measures by genotype in 1889 unrelated MESA participants with protein measures and whole genome sequencing data available. Data are presented as the median protein levels (center line) with the 25th and 75th percentiles as box bounds; whiskers extend to 1.5× the interquartile range, and points beyond represent outliers. **C** pQTL summary statistics (generated with linear regression under an additive model, two-sided test of significance) and effect allele frequencies per MESA ancestry-group. *cis*-pQTL were significant at traditional genome-wide significance threshold ($p < 5 \times 10^{-8}$). This variant is most common among AMR and EAS ancestry groups. **D** Inter-platform protein measure correlation per ancestry without and with PAV adjustment. **E** Forest plot of age association statistics generated from linear regression of age versus protein (two-sided test of significance), using SomaScan and Olink measures without and and with PAV adjustment ("PAV-adj") in 1889 participants. Boxes represent estimated beta coefficients, and horizontal lines show 95% confidence intervals. Associations were significant at $\alpha = 0.05$ correcting for 2708 probe pairs tested ($p < 1.8 \times 10^{-5}$). Source data is available in Supplementary Data 4, 5 (**C**), and 6 (**D**, **E**). Source data for (**A**, **B**) is individual-level data available through dbGap and MESA.

with age, sex, BMI, and T2D case-control status. For proteins with a platform-discordant PAV, we observed an improved correlation of effect sizes observed among protein-phenotype associations when using PAV-adjusted measures as input (Supplementary Data 6 and Supplementary Fig. 33). While adjustment for platform-specific PAVs generally improved protein measure correlation (Supplementary Fig. 14) and the strength and concordance of some phenotypic associations, effects were small if present (Supplementary Figs. 34 and 35). Similarly, adjusting proteins associated with platform-specific pleiotropic *trans*-pQTL loci had little effect on protein measure correlation (Supplementary Note 5, Supplementary Fig. 54).

One protein with platform-discordant measures that exemplifies the impact of PAV adjustment is paired immunoglobulin-like type 2 receptor alpha (PILRA). This protein showed strong platform-discordant measures (Pearson's $r = -0.35$) across the present and prior studies[16,17] (Fig. 4A). A platform-discordant *cis*-pQTL for PILRA is driven by an ancestry-differentiated missense-variant [rs1859788 G > A, PILRA, p.Gly78Arg, SomaScan: $B = -1.34$ (0.02), $p = 1 \times 10^{-300}$, Olink: $B = 0.92$ (0.02), $p = 2.98 \times 10^{-300}$] (Supplementary Figs. 15 and 48) most common among EAS (EAF = 0.60) and AMR (EAF = 0.53) participants, compared to AFR (EAF = 0.14) and EUR (EAF = 0.32) participants (Fig. 4B, C). Upon adjusting for copies of the PAV allele, the correlation of residual PILRA measures improved to 0.27 (Fig. 4A). Inter-platform protein measure correlation improved most among AMR and EAS participants (Fig. 4D and Supplementary Fig. 28). Furthermore, incorporating PAV-adjusted PILRA measures into epidemiological models improved the strength and concordance of downstream PILRA-age associations (Fig. 4E). Finally, adjusting for this PAV in the *cis*-pQTL mapping model unmasks and strengthens a second, independent association with the Olink measure of PILRA, driven by non-coding variant chr7:100480786:C:G (Supplementary Fig. 15). We propose that this alternate non-coding signal may be a true abundance pQTL for the PILRA protein, while the primary, platform-discordant PAV signal may capture platform differences in ability to bind and measure different genetically-encoded PILRA isoforms.

Soluble urokinase plasminogen activator receptor (suPAR) is another protein associated with an ancestry-differentiated, platform-discordant PAV. suPAR measures are modestly correlated across all

participants (Pearson's $r = 0.43$), but there is significant heterogeneity in inter-platform correlation estimates between ancestries (Cochran's Q = 55.6 , Cochran's Q $p = 5.02 \times 10^{-12}$), with lower correlation estimates among individuals clustering with the AFR ancestry reference population (Pearson's $r = 0.24$, Fig. 5A, D, E). Additionally, a platform-discordant *cis*-pQTL driven by a missense variant [rs399145 T > C, PLAUR, p.Thr86Ala SomaScan: $B = -2.09$ (0.070), $p = 2.03 \times 10^{-155}$, Olink: B = 0.61 (0.070), $p = 2.58 \times 10^{-17}$] (Supplementary Figs. 16, 49) common among individuals with African ancestry [EAF in MESA participants AFR = 0.10, AMR = 0.009, EAS = 0.000, EUR = 0.004, EAF $X^2 = 230.7$ (ancestry-differentiated), X^2 $p = 9.49 \times 10^{-50}$] (Fig. 5B) associates with suPAR measures. Upon adjusting for copies of the PAV allele, the correlation of protein measures improved from 0.44 to 0.57, with the largest improvement in inter-platform correlation estimates among individuals of African ancestry (Pearson's $r = 0.59$ post-PAV adjustment) (Fig. 5C). In summary, adjustment for PAVs may improve concordance of protein measures and downstream associations. Additional examples for APOL1 (Supplementary Figs. 20, 29, 30, and 38) and CPPED1 (Supplementary Figs. 18, 31 and 40) proteins are described further in Supplementary Note 4.

Impact of platform-discordant and platform-specific PAV instruments on Mendelian Randomization results

One valuable application of pQTL is to provide genetic instruments for MR studies, which evaluate genetic liability for protein abundances for impact in diseases. When pQTL are applied as genetic instruments in MR, it is generally assumed that they capture the effects of overall protein abundances. However, if a pQTL captures effects of isoform differences instead of overall protein abundances, as may be the case for platform-discordant and platform-specific *cis*-pQTL driven by PAVs, this instrument may bias MR effect-estimates.

The PILRA protein discussed above has been linked to Alzheimer's Disease (AD) in multiple MR studies using Olink pQTL instruments[27–29]; however, different directions of effect for PILRA on AD have been reported across studies using different platform's pQTL instruments. We and others observed a platform-discordant PAV *cis*-pQTL for this protein[16]. To evaluate the potential impacts of this, and other, platform discordant *cis*-pQTL PAV on the concordance of MR results, we performed MR between the 2157 proteins assessed in this study and AD as an outcome[30]. Similar proportions of proteins tested in MR on each platform showed significant ($p < 0.05$) associations with AD. While the majority of the 51 proteins with a nominally significant AD-MR association on both platforms showed concordant directions of MR estimates on each platform, 5 showed opposite directions of effect between analyses with SomaScan versus Olink instruments, including PILRA (Fig. 6). Notably, all 5 of these proteins associated with a platform-discordant or platform-specific PAV on at least one platform.

To further evaluate the impact of potential assay-interference PAVs on MR results, we performed sensitivity analyses excluding first platform-discordant, then platform-discordant and platform-specific PAV instruments from MR analyses. While we note that less proteins could be tested across analyses excluding PAVs (as these were the only instruments available for some proteins), we observed the concordance of significant MR effect estimates increase (from Pearson's r 0.48, 0.51, to 0.63, Fig. 6). Furthermore, we observed 5 proteins in SomaScan MR (CHIT1, GDF-15, HGFAC, HRC, qSOX1), and 5 in Olink MR analyses (OSCAR, LTA, SPINT1, TMPRSS5, TNFRSF13C), which were significant in models excluding platform-discordant and platform-specific PAV instruments but were not significant in the model including these instruments as PAVs. This suggests that including pQTL instruments that do not necessarily capture overall abundance effects, but rather isoform differences, may, in some cases, bias MR results towards the null. Hence, excluding these PAVs as instruments may reveal potentially causal relationships between overall protein abundances and disease outcomes.

Discussion

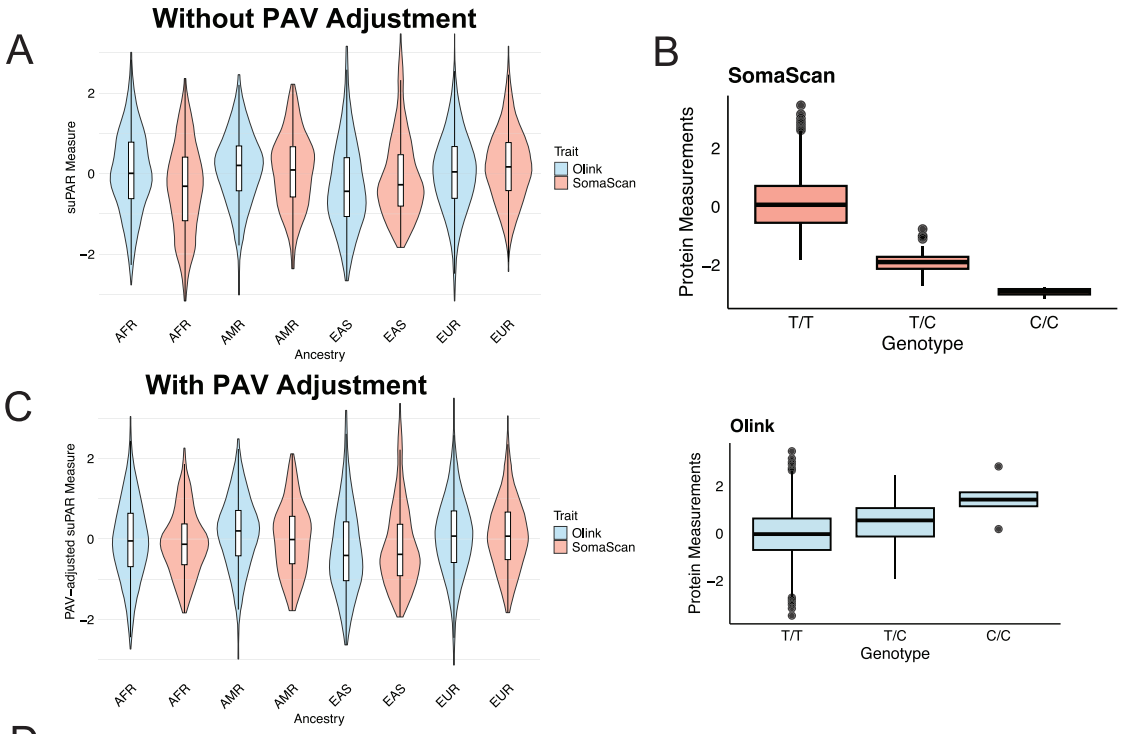
As affinity-based proteomic platforms shape our understanding of genetics and disease, it is imperative to consider what each platform's protein measures capture and account for putative genetic drivers of affinity-based protein measure discordance across diverse populations to increase the utility of data across all participants. This study represents the first comprehensive comparison of affinity-based protein measure agreement and genetic signals in an ancestrally diverse cohort. We highlight proteins with systematic differences in assay measurements by ancestry. We further demonstrate that genetic variants with different frequencies across ancestries may contribute to these measurement differences. Moreover, we demonstrate that adjusting for likely assay-interference variants can, in some cases, unmask previously undetected biology.

Among the 2157 proteins with SomaScan 7k and Olink Explore 3072 measures compared here, we observe a wide range of inter-platform correlations for given protein measures, as in prior studies. One prior study demonstrated that biological (such as presence of a trans-membrane domain, protein length) and technical factors (percentage of outlier or below limit of detection protein measures, etc.) can explain this wide range of correlation coefficients[16]. While relatively uncommon, negatively correlated protein measures are interesting, as these suggest that both platforms measure consistently different, but biologically related values for the same proteins, such as different isoforms. Importantly, we show that, while most proteins did not show systematic differences in platform agreement by ancestry, a subset of 80 proteins had significantly heterogeneous correlation across ancestries, suggesting that one or both platforms may vary in performance across ancestries. Although this study focuses on the contribution of coding genetic variation to cross-ancestry differences, additional factors such as ancestry-differentiated non-coding *cis*-genetic variation or *trans*-pQTL, differential recruitment of participants of varied ancestry by site, or other environmental or clinical factors, may also contribute to ancestry-heterogeneity in correlation estimates.

Here, and across prior studies, *cis*-pQTL presence/absence is thought to provide one predictor of on-target protein measures[12,15,16]. We observed similar proportions of protein measures associated with any *cis*-pQTL on each platform (39% SomaScan, 48% Olink), with more distinct *cis*-pQTL identified for Olink overall (for the set of 2157 proteins measured on both platforms), consistent with single-ancestry genetic comparisons of earlier versions of SomaScan and Olink platforms[12,16,17]. Proteins associated with concordant *cis*-genetic signals exhibited the highest average inter-platform correlation, increasing confidence that the affinity probes for these proteins captured the same entities and may have high potential for cross-platform meta-analysis.

However, *cis*-pQTL driven by coding genetic variation, such as protein altering variants (PAVs) may, in some cases, reflect impacts to the target protein structure that affect protein measurements, particularly for *cis*-pQTL with platform-discordant directions of effect or platform-specific significance. While PAVs may certainly alter a protein's overall abundance through impacts on protein stability and other mechanisms, studies incorporating comparisons to MS protein measures, measures of gene transcript abundance, and/or functional validation have established cases where platform-discordant PAV *cis*-pQTL reflect differences in probe binding to genetically encoded isoforms, including for uromodulin[31] and apolipoprotein-1[25]. While the majority of platform-concordant and platform-specific *cis*-pQTL identified are driven by non-coding genetic variants, we find that nearly all the platform-discordant signals, and ~30% of platform-specific *cis*-pQTL, are driven by PAVs on each platform. It remains challenging to establish the pervasiveness of assay-interference effects; however, we note that several putatively impacted proteins

https://doi.org/10.1038/s41467-025-67814-1



D

PAV Summary Statistics				EAF				
Platform	B	SE	P-Value	AFR	AMR	EAS	EUR	ALL
Soma	-2.09	0.02	$2x10^{-155}$	0.10	0.01	0.00	0.00	0.03
Olink	0.61	0.02	$3x10^{-17}$					

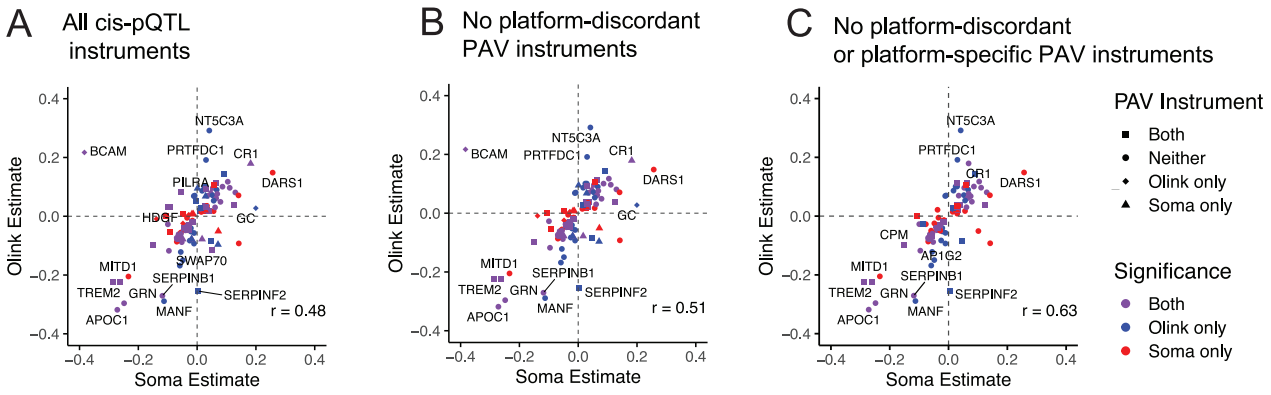
E

	Pearson's r of suPAR Measures				
	AFR	AMR	EAS	EUR	ALL
Without PAV Adjustment	0.24	0.60	0.64	0.48	0.44
With PAV Adjustment	0.59	0.65	0.64	0.51	0.57

Fig. 5 | Adjustment for suPAR-associated, ancestry-differentiated PAV improves correlation of suPAR measures across all participants, with improvement driven by individuals of African ancestry. suPAR measures per platform, per-ancestry, without (**A**) and with (**C**) adjustment for PAV rs399145 (chr19:43665370:T:C; p.Thr86Ala) in 1852 unrelated participants with protein measures, whole genome sequencing data available, and ancestry group assigned (433 African ancestry (AFR), 220 Admixed American (AMR), 267 East Asian (EAS), and 932 European (EUR)). Violin plots show the full distribution of suPAR values by ancestry and platform; embedded box plots indicate the median (center line) and interquartile range (box). **B** SomaScan and Olink suPAR measures by genotype in 1889 participants with protein measures. Data are presented as the median protein levels (center line) with the 25th and 75th percentiles as box bounds; whiskers extend to 1.5× the interquartile range, and points beyond represent outliers. **D** suPAR pQTL summary statistics (generated under an additive model, two-sided test of significance) per platform and effect allele frequencies per ancestry group in MESA. *cis*-pQTL were significant at traditional genome-wide significance threshold ($p < 5 \times 10^{-8}$). Variant is most common in AFR ancestry group. **E** Inter-platform correlation of suPAR measures without and with PAV adjustment in 1889 participants. Values highlighted in red indicate which ancestry groups experienced the largest improvement in correlation following adjustment for PAV (here, AFR). Source data is available in Supplementary Data 4, 5 (**D**) and 6 (**E**). Source data for **A**, **B**, **C** is individual-level data available through dbGap and MESA.

are potential biomarkers or drug targets (for example, suPAR[32,33] and PILRA[27]). As sample sizes and power for pQTL detection and platform protein coverage expand, we predict that more impacted proteins will be identified, underscoring the importance of considering differential binding effects in proteomic analyses. Integration of larger MS, expression, and splicing datasets will further facilitate the identification of assay-interference pQTL and help differentiate, in situations of discordant pQTL effects, what a given platform's measure captures, as affinity reagents may have differential binding to different protein isoforms[18,24,25].

https://doi.org/10.1038/s41467-025-67814-1



D

Category	cis-pQTL Instruments Used							
	All cis-pQTL (1)		No platform-discordant PAV (2)			No platform-discordant or platform-specific PAV (3)		
	Tested	Significant (p_1<0.05), (% of tested)	Tested	Significant (p_2<0.05), (% of tested)	Newly Significant (p_1>0.05, p_2<0.05)	Tested	Significant (p_3<0.05), (% of tested)	Newly Significant (p_1>0.05, p_2>0.05, p_3<0.05)
SomaScan MR	807	77 (10%)	798	98 (11%)	0	738	91 (12%)	5
Olink MR	974	84 (9%)	966	124 (13%)	0	882	116 (13%)	5
SomaScan MR + Olink MR	601	50 (8%)	590	47 (8%)	0	552	44 (8%)	1

Fig. 6 | Comparing SomaScan/Olink effect-estimates from significant ($p < 0.05$) Alzheimer's Disease (AD) Mendelian Randomization (MR) results, from inverse-variance weighted MR analyses (two-sided tests). The genetic instruments tests for these MR tests include **A** all cis-pQTL for each platform (Referred to as Model 1 (p value p_1), **B** all cis-pQTL except for platform-discordant protein-altering variants (PAV) (Referred to as Model 2—p value p_2), **C** all cis-pQTL except for platform-discordant and platform-specific PAVs for each platform (Referred to as Model 3—p value p_3). **D** Tabular summary of significant results from MR models, including SomaScan pQTL instruments only or Olink pQTL instruments only. The bottom row summarizes the results for the subset of genes that were tested in both the SomaScan and Olink MR models. In all models, a nominal p value threshold of 0.05 was used to define significance. Source data is available in Supplementary Data 26 and 27.

Notably, we show that for protein measures associated with platform-discordant PAV *cis*-pQTL, adjustment for the PAV improved protein measure correlation and resulted in improved strength and concordance of some genetic and epidemiological signals, even unmasking new significant associations. When *cis*-pQTL do reflect differential binding to encoded isoforms, we note that studying the impact of individual isoforms can be biologically informative in studies of disease, as established with GDF8/11 for cardiovascular events[34]. However, it may not be the case that all protein isoforms have biologically distinct impacts, and careful interpretation and analysis on a protein-by-protein fashion is needed. Also, platforms often do not report which isoform a given probe targets (or do not provide evidence of specific targeting) and disentangling which isoform each platform measures requires additional analyses incorporating genetic or MS evidence, which may not be feasible in all studies. If the measure of interest is overall protein abundance, rather than the abundance of specific protein isoforms, PAV-adjustment may provide an avenue of accounting for potential isoform differences in affinity-reagent binding. Careful consideration is needed on whether information may be lost regarding differential isoform abundance by adjusting for discordant PAV, counterbalanced by the potential to have better power to detect correlates of overall protein abundance with such adjustment. However, more work needs to be done to integrate MS with affinity-derived pQTLs, as started in very recent analyses[24].

Additionally, we established that proteins with cross-ancestry differences in platform agreement were enriched for association with PAVs with ancestry-differentiated allele frequencies, supporting epitope-binding effects as one driver of cross-ancestry differences in platform agreement. Adjustment for platform-discordant, ancestry-differentiated PAVs generally reduced cross-ancestry heterogeneity in inter-platform correlation estimates. However, the improvement in inter-platform correlation across ancestry groups was not linear, and several proteins with ancestry-heterogeneity were also associated with ancestry-differentiated non-coding *cis*-pQTL or *trans*-pQTL, further suggesting the interplay of environmental or other genetic effects apart from PAVs. Researchers may not always want to adjust for non-PAV factors that explain ancestry differences in these settings. Nonetheless, careful consideration of the potential for assay-interference differentially impacting protein measurements across ancestry groups will be imperative to maximize the utility of proteomics data in future research and clinical settings.

Failure to account for assay-interfering PAVs in studies of circulating proteins may hinder the detection of disease-relevant associations or putative biomarkers. We show that platform-specific and platform-discordant PAVs may bias MR results if used as genetic instruments for circulating protein abundances[35], in some cases resulting in platform-specific and/or platform-discordant MR estimates. Furthermore, we establish that excluding platform-specific PAV instruments from MR models may increase power to detect significant associations for a subset of proteins. This is particularly surprising, as, in theory, these instruments explain less of total trait variance, supporting that the platform-specific PAV may, in some cases, be masking other signals (for example, from non-coding or regulatory variants) that do reflect differences in overall protein abundance. Similarly, if proteomics platforms are employed clinically, failure to account for genetic variation that does not capture overall protein abundance measures may result in systematically different measures among carriers of such genetic variants, thereby creating challenges for downstream applications in disease diagnosis/prognosis.

Interpretation of affinity-based *trans*-pQTL, particularly pleiotropic *trans*-pQTL, is challenging. Existing affinity-based proteomics studies exploring *trans*-pQTL often interpret pleiotropic *trans*-pQTL as biologically relevant signals, which may help confirm appropriate assay performance[36,37]. While pleiotropic regulators certainly exist and are relevant in metabolism and disease risk[38], it is challenging to interpret whether pleiotropic *trans*-pQTL reflect true differences in protein abundances, systematic alterations to PTMs that interact with epitope binding, or other factors that may influence interactions, especially in the absence of cross-platform replication or orthogonal validation with a different assay. For example, while the detection of any platform-shared *trans*-pQTL association for a protein corresponds with high inter-platform correlation, proteins associated with regions pleiotropic on SomaScan only had low inter-platform correlation. The same trend has been observed in the UK BioBank/deCODE SomaScan and Olink platform-comparison, where significant platform-specific pleiotropic *trans*-pQTL were seen at some of the same regions, such as complement factor gene *CFH* and glycoprotein-encoding *VTN* for SomaScan protein measures, and proteins associated with these loci often showed lower inter-platform correlation compared to other proteins[17]. Eldjarn et al. propose that such loci may capture an interaction between SomaScan technical processes and the protein pathways influenced by the *trans*-pQTL. In a similar vein, we observe platform-shared and/or platform-specific pleiotropic *trans*-pQTL regions at loci encoding proteins involved in coagulation processes (*F12*, *F5, CPB2*), blood type (*ABO*), fat-transport proteins *(APOE*), other glycoproteins (*HPX, DAG1*), and proteins with roles in glycosylation (*FUT6, FUT2, HRG, GNPTAB*), rather than for transcription factors (as might be expected if a locus captures a true "master-regulator" effect -- for example, *trans*-eQTL hotspots are often located near transcription factors[37,39]). Such loci may result in widespread changes to, for instance, protein glycosylation, which could induce epitope effects for several proteins based on differential binding to proteins with different PTMs. Similarly, the pleiotropic associations with coagulation factors may impact protein isoforms or indicate susceptibility of one or both platforms to sample preparation variability. Further work, particularly with MS, will be needed to better understand the biological and/or technical effects these pleiotropic *trans*-pQTL capture.

While this study is currently the most diverse platform-comparison, the representation of non-European ancestry groups remains limited. Additional studies with increased numbers of non-European ancestry participants will facilitate the identification of genetic drivers of ancestry heterogeneity. Due to sample size constraints, ancestry clusters are also derived at the continental level in this study; future studies should consider heterogeneity within continents for cross-platform protein correlation and protein measurement impacting pQTL. Future studies may also consider evaluating differences in pQTL effect sizes by local ancestry background to evaluate the extent to which platform agreement is explained by differences in pQTL allele frequency versus effect estimates.

Additionally, this analysis is limited to the 2157 protein targets measured on both platforms. However, SomaScan 7k and Olink Explore 3072 investigated in this study measure up to 7000 and 3072 proteins, respectively. Furthermore, each platform has released a new version, with SomaScan and Olink Explore now measuring up to 11,000 and 5400 proteins. We cannot assume that the same patterns observed in this study hold true in these newer versions, particularly as platforms expand into coverage of lower abundance proteins and more intracellular proteins. These analyses should be repeated as data on large numbers of overlapping samples from ancestrally diverse participants accrue on the most recent SomaScan 11k and Olink 5k platforms. Furthermore, power for pQTL (particularly *trans*-pQTL) discovery is limited by sample size. Additional evaluation of pQTL across increased sample sizes will likely yield additional associations.

Here, and in prior comparisons, the inter-platform correlation of protein measures serves as a proxy for on-target affinity-probe binding. While consistent rank-based protein measures increase confidence in protein measures, it remains possible that neither platform provides accurate protein measures. It is especially challenging to interpret platform-discordant measures, though metrics like *cis*-pQTL presence may provide clues. Platform comparisons incorporating MS protein measures are already proving informative to further depict genetically encoded isoform differences and impacts of splicing on detected protein structures[24,40,41], and will be especially important to expand across ancestrally diverse sets of participants. The release of additional information regarding affinity-probe confirmation and target domains from SomaLogic and Olink would prove useful for research interpretation of affinity probe binding.

Methods

Study sample

The MESA has been approved by the institutional review boards (IRBs) of each field center, the data coordinating center, and the genetic analysis center. The IRB at the Lundquist Institute for Biomedical Innovation also reviewed and approved this project. All participants provided written informed consent, including for genetic study, and participant compensation was approved by local IRBs at all MESA sites. All study activities were conducted in accordance with the criteria set by the Declaration of Helsinki.

MESA is a prospective, population-based cohort study of 6814 men and women free of cardiovascular disease at baseline, aged 45 to 84 years, with recruitment from 6 US clinical centers beginning in 2000[42]. At the baseline exam, 53% of participants self-identified as female, 38% as non-Hispanic White, 28% as African American, 12% as Chinese, and 22% as "Spanish, Hispanic, or Latino" (subjects with both Caribbean and Mexican/Central American ancestry are represented).

Blood sample collection

The MESA exam 1 protocol for blood sample collection and processing was performed as described[42,43]. In brief, extracted blood was added to EDTA and mixed for at least 30 s. Blood-EDTA samples were stored upright on ice for no longer than 30 min following extraction. Centrifugation was performed at $2000 \times g$ for 15 min, or $3000 \times g$ for 10 min. Centrifuged samples were placed on ice; plasma was extracted into 0.5 mL aliquots. Cryovials were frozen and stored in an upright position at −70 °C. Samples were shipped on dry ice. Proteins in plasma samples collected in this manner have been shown to be stable over time[44].

Genetic ancestry assignments via clustering

Continental ancestry of each subject was assigned using Admixture[45–47] in a supervised analysis with reference genotypes from the 1000 Genomes project[48] and the Human Genetic Diversity Panel (HGDP)[49]. Uniformly called genotype data for MESA, 1000 Genomes, and HGDP were obtained from TOPMed whole genome sequence data[50]. MESA subjects were assigned to a 1000 G/HGDP sub-population based on similarity greater than 50% to the reference (with a small number of MESA participants ($n = 44$) unclassified using this threshold).

Proteomics measures

Plasma samples for a subset of 1930 participants from the baseline MESA exam 1 (2000-2002) were assayed on both SomaScan 7k and Olink Explore 3072.

MESA blood samples consist of EDTA plasma, cryopreserved at −80 °C. The modified aptamer binding reagents[51], SomaScan™ v.4.1 assay[52,53], and its performance characteristics[40,54,55] have been described[40,51–55]. The assay uses DNA-based slow off-rate modified aptamers (SOMAmerTM reagents) to quantify the relative abundance of approximately 7000 proteins with high specificity and limits of detection largely comparable to antibody-based assays. Briefly, the SomaScan assay starts with a mixture of ~7596 modified aptamer reagents labeled with a 5′ fluorophore, photocleavable linker, and biotin that are immobilized on streptavidin (SA)-coated beads and incubated with 55 µL of EDTA plasma samples to achieve binding of aptamers to proteins under equilibrium conditions. In the second step, meant to reduce non-specific binding, the samples are diluted, causing non-specifically bound aptamer reagents to dissociate while their rebinding is prevented using a polyanionic competitor. Aptamer reagents are hybridized to complementary sequences on a DNA microarray chip and quantified by fluorescence. Standard normalization procedures were conducted[56]. For the MESA study, the assay started with a total number of 7596 modified aptamers. We excluded 307 aptamers to non-human proteins. Thus, the final number of aptamers was 7289, which collectively measure 6407 UniProt IDs. SomaScan delivers two versions of the data; one with the normalization procedures described, and one with an additional normalization procedure—adaptive normalization by maximum-likelihood applied to all samples (ANML/ANML-SMP), where sample measures are further normalized relative to measures in an external reference of healthy controls[57,58]. The median coefficient of variation for all SomaScan ANML-normalized protein measures in MESA is 2.12 (IQR 1.74, 2.78). This value is based on 1292 samples (646 pairs of split duplicates).

Olink Explore 3072 uses 2941 antibody-nucleic acid assay probes to quantify 2172 human protein targets[59,60]. Two antibody probes are designed per protein target; when both antibodies bind, the nucleic acids are brought in proximity to hybridize. Real-time polymerase chain reaction of the hybridized nucleic acids is used to generate a relative protein abundance measure in NPX units. NPX units are company normalized against a series of internal controls (incubation, extension, detection, inter-plate, and negative controls) designed to account for noise in the immune reaction and subsequent amplification step, between samples and plates[61].

In this analysis, measures from both platforms were obtained in one batch. We included only the subset of MESA participants with measures available from the same plasma sample on both platforms. After excluding 7 participants with samples flagged by SomaScan for possible measurement errors, 1930 participants were included in the study. Furthermore, we also only assessed the 2708 probe pairs corresponding to the 2157 human protein target UniProt IDs measured on both platforms (2687 SomaScan aptamers, 2172 Olink IDs in total). Probes were matched to UniProt IDs using reference files from SomaLogic and Olink, respectively. Both assays output a value for all proteins measured, no protein value imputation was performed in this study. A subset of proteins is measured by more than one affinity probe on each platform. For main analyses, we considered one SomaScan aptamer and one Olink antibody-conjugate probe per UniProt ID, prioritizing the probe pair with the highest inter-platform correlation (unless otherwise stated). Hence, the main analyses consider 2157 aptamers and Olink IDs, for 2157 UniProt IDs. Results for all probes per UniProt ID are reported in the Supplementary Data. RFU and NPX measures obtained from each platform were standardized and rank-based inverse normal transformed prior to conducting analyses.

Protein measure comparison and inter-ancestry heterogeneity of protein measures

Pearson's correlation coefficients were computed for transformed SomaScan measures against Olink measures, as well as for ANML-SomaScan measures against Olink measures, using data from all 1930 contributing participants. We also computed Pearson's correlation coefficients among individuals stratified by ancestry group, excluding the 44 participants who could not be assigned to ancestry groups using the criteria defined. To evaluate whether correlation estimates for a given protein differed significantly across ancestry groups, we computed the Fisher's R-to-Z transformation on per-ancestry Pearson's r estimates and performed the Cochran's Q test of heterogeneity across resulting Z-scores. Analyses were run in R Studio (version 4.3.1 or 4.4.0) unless otherwise noted[62].

Whole genome sequencing

TOPMed WGS was performed according to methods were described previously[50]. Briefly, WGS was conducted at six sequencing centers (mean depth >30X, Illumina HiSeq X Ten instruments). Joint variant discovery and genotype calling were conducted by the TOPMed Informatics Research Center (IRC) across all TOPMed studies using the GotCloud pipeline, resulting in a single genotype call set encompassing all of TOPMed (TOPMed Freeze 10). Variant quality control was also performed centrally by the TOPMed IRC. Sample quality control was performed by the TOPMed Data Coordinating Center (DCC).

pQTL mapping

In this study, 1923 participants had both SomaScan 7k and Olink Explore 3072 protein measures from exam 1 plasma samples, as well as TOPMed whole genome sequencing data available. Direct relatives were identified with a pedigree and excluded entirely from analyses, resulting in 1889 participants for pQTL mapping. Prior to QTL analyses, inverse-normal transformed protein measures from each platform were residualized based on participant age, genotypic sex, MESA recruitment site, protein measure plate number, ten genotype principal components (PCs), and ten protein PCs. Genotypic sex was used throughout this study. Protein PCs were calculated across all inverse-normal transformed protein measures for each platform. Early protein PCs likely capture the impacts of latent variables such as technical processes or platform sample handling, or impacts of kidney function or disease status among participants. Here, we show that adjustment for protein PCs improves power for *cis*-pQTL discovery on both platforms (Supplementary Fig. 5). Notably, our pipeline to assess the impact of adjustment for different numbers of protein PCs in pQTL mapping differed from our main pQTL analysis pipeline, in that we used EasyStrata to define loci from tensorQTL based *cis*-pQTL summary statistics ($p < 5 \times 10^{-8}$, requiring lead variants to be at least 500 kilobases apart) rather than the comprehensive fine-mapping strategy to define signals as performed in our main analysis. This resulted in the identification of fewer *cis*-pQTL at ten protein PCs in this exploratory analysis, compared to our main analysis. However, we believe the results observed—where adjustment for protein PCs improved *cis*-pQTL discovery -- would hold true if using our main analysis pipeline to define independent loci. While adjustment for 10 protein PCs does not

maximize *cis*-pQTL detection, we performed a GWAS of the first 100 protein PCs on each platform to show that several later PCs likely capture impacts of *trans*-pQTL (Supplementary Fig. 6), which we did not want to adjust away for the purposes of this study. Hence, we chose 10 PCs for consistency between platforms and *cis*- and *trans*-pQTL analyses, although future research endeavors may want to consider adjusting for different numbers of protein PCs between *cis*- and *trans*-pQTL analyses to maximize discovery.

Here, we performed QTL mapping with SomaScan measures with ANML normalization applied to all samples, and Olink measures. Gencode version 45 gene coordinates were used to define transcription start sites (TSS)[63]. If gene coordinates could not be found in Gencode, Ref Seq gene coordinates downloaded via the USCS Genome Browser on June 5, 2023 were used. *Cis*- and *trans*- QTL mapping was performed ancestry-pooled to maximize power for discovery of genetic signals[64], as performed in prior studies across TOPMed[39,65] within tensorQTL version 1.0.8[66], using linear models and adjusting for the same set of covariates as listed above. In this study, *cis*-pQTL is defined as pQTL within 1 Mb of the TSS, and *trans*-pQTL is defined as >1 Mb from the TSS. We note that, to increase efficiency, the tensorQTL *trans*-pQTL mapping function only outputs variants associated with the protein measure with a p value $< 1 \times 10^{-8}$; hence, *trans*-pQTL summary statistics are only available for these variants. Fine-mapping of resulting *cis*- and *trans*- signals was performed using the cis.susie and trans.susie functions in tensorQTL, which implement a modified version of SuSiE R package[23] using individual level linkage disequilibrium (LD) references, respectively, to obtain credible sets of variants within each signal for each platform. We use these credible sets to define genetic signals throughout this study. Significant credible sets were defined as *cis*-pQTL credible sets containing at least one variant associated with a protein with $p < 5 \times 10^{-08}$, or *trans*-pQTL credible sets containing at least one variant associated with a protein with $p < 1 \times 10^{-11}$. After observing a small subset of cases for which SuSiE output credible sets containing variants in high LD, we filtered credible sets for the same trait that contain variants in LD ($r^2 > 0.1$) across participants to retain only the credible set with the more significant lead variant.

Identification of ancestry-differentiated variants

To identify variants with differing EAF by ancestry, we obtained EAF of each variant within each contributing ancestry group in MESA using plink version 1.90b3[67] (www.pngu.mgh.harvard.edu/purcell/plink/), specifying the TOPMed reference allele as the reference allele for each ancestry. We then computed Chi-Square test statistics and p-values across EAF values for each variant. Given the large number ($>1 \times 10^7$) of highly significant variants following multiple-testing correction ($p < 4.07 \times 10^{-9}$), we consider a Chi-Square value in or above the 75th percentile across all MESA participants as evidence of ancestry-differentiated allele frequencies.

Cross-platform pQTL signal comparison

Here, we considered pQTL credible sets for a given protein "platform-overlapping" if there was at least one shared variant across the credible sets associated with the same protein measure on each platform. If neither platform's signal contained the lead variant for the other platform, we did not compare the concordance of effects. Genetic signals that did overlap for at least one platform's lead variant were categorized as platform-concordant or platform-discordant based on the direction of association of the lead variant with each platform's protein abundance measure (Supplementary Fig. 9). Platform-concordant signals are significant pQTL signals with concordant directions of association with the given protein on both platforms, whereas platform-discordant signals are significant pQTL signals with differing directions of effect with the given protein across platforms.

Assessment of protein subcellular location

Human Protein Atlas[68,69] was used to ascribe protein subcellular locations as membrane, secreted, or intracellular as previously described[4,17]. Briefly, proteins were considered membrane proteins if annotated by HPA only as membrane proteins, secreted proteins if annotated only as secreted proteins, and intracellular if annotated by HPA as both or neither.

Predicting variant consequences and identifying protein altering variants

Functional annotation and further characterization of variants was performed via WGSA 0.95[70], which integrates variant annotations from Variant Effect Predictor (VEP)[71], Annovar[72], and SNPEff[73]. Here, we used the report-predicted consequences output by VEP[71], annotated relative to Ensembl, in the main analyses. The top predicted consequence for each genetic variant reported by VEP was chosen. Intronic variants were classified according to whether they were intronic for the protein-encoding gene or for a different gene. Variants were annotated as regulatory if they contain variants reported as "upstream" or "downstream" of a gene. Protein-altering variants were defined as VEP-Ensembl annotated missense variants, start-loss, stop-gain, or stop-loss variants. The vast majority of PAVs were missense variants. For *cis*- and *trans*-pQTL main analyses, we report the number of signals reported by VEP to alter the protein-encoding gene.

Affinity pQTL overlap with GTEx *cis*-eQTL

To assess whether pQTL may exhibit shared effects with GTEx eQTL, which are statistically significant associations between a genetic variant and a transcript abundance, we evaluated whether any variants within identified *cis*- and *trans*-pQTL credible sets are reported lead-eQTL (for either primary or secondary signals, based on conditional analysis) in GTEx version 8[74]. When comparing overlap between *cis*-pQTL signals and GTEx eQTL signals, we filtered GTEx eQTL counts by whether the observed eQTL corresponded to the same protein-encoding gene. Notably, the GTEx consortium reports eQTL for several tissues. To label a pQTL direction of effect as concordant with the GTEx-reported eQTL direction of effect, we required the pQTL and GTEx eQTL direction of effect for all tissues in which the eQTL was significant. A small subset of pQTL signals (28 credible sets on SomaScan, 27 credible sets on Olink) contained more than one distinct eQTL signal, with conflicting directions of effect. We excluded these signals from effect-direction comparisons.

Affinity pQTL overlap with MS-pQTL

To assess whether pQTL regions are previously reported MS-pQTL, we evaluated whether any lead SEER MS-pQTL in either the 320 Qatar Metabolomics Study of Diabetes (QMDiab)[10] or the larger the Tarkin Study[10] of 1980 participants were present in our SomaScan or Olink pQTL credible sets, and determined whether these signals were associated with the same protein by comparing associated UniProt IDs. We further assessed the overlap of affinity-based pQTL with MS-PAVs, which are MS-based variants hypothesized by Suhre et al.[10] to reflect impacts on protein quantification by PAV, and not true differences in protein abundance.

MESA pQTL overlap with deCODE and UK BioBank reported pQTL

To assess whether pQTL regions identified in the present study were previously reported in the deCODE-based SomaScan version 4 pQTL study (n = 35,559 Icelanders)[1] or the UK Biobank Olink Explore 3072 pQTL study (n = 54,219 UKB Pharma-Proteomics-Project participants)[5], we evaluated whether the 1MB region surrounding the lead variant contained a previously reported lead pQTL variant for the same protein from these two studies.

MESA pQTL overlap with GWAS catalog variants

To evaluate whether pQTL have been previously associated with phenotypes, we assessed whether pQTL signals (defined by SuSiE credible sets) contained lead variant-trait associations reported in the GWAS Catalog (downloaded August 12, 2024)[75].

Characterizing pleiotropic *trans*-pQTL regions

Several regions contained multiple *trans*-pQTL signals that often associated with one or more of the same proteins within and across platforms. Hence, we took a region-based, rather than a signal-based, approach to define platform-shared pleiotropy vs. "platform-differentiated" pleiotropy. For each platform, we defined *trans*-pQTL regions by finding credible sets with the most significant lead variant and defining a 1 Mb window around these variants. If 1 Mb windows overlapped, we merged the windows into one. We subsequently compared regions identified for each platform: if regions overlapped across platforms, we defined them as platform-shared.

Platform-shared pleiotropic regions were defined as overlapping *trans*-pQTL regions associated with at least 5 proteins on both platforms. Platform-specific pleiotropic regions were defined as regions associated with at least 5 proteins on one platform, and 0 or 1 protein on the other.

Limit of detection

SomaScan global limit of detection (LOD) values were obtained from SomaLogic. Per company recommendations, we compared protein LOD values to ANML-normalized sample protein RFU measures. Olink LOD values were calculated from negative control data, per company recommendations. 186 negative controls samples from the 102 plates used in the present analysis were used for calculation of LOD. Protein measures were filtered to intensity normalized values (excluding amplification, incubation, and extension controls). LOD values were calculated for each protein across all plates by taking the median of the negative control measure + 3 fixed standard deviations. We compared LOD values to non-transformed Olink sample NPX values.

Epidemiological analyses

Proteins were associated with participant age and BMI using linear regression models in R[76] adjusted for plate number and site in age association models, and additionally adjusted for age and sex in BMI models. Proteins were also associated with sex and prevalent T2D using logistic regression, with adjustment for age and site (as well as sex in T2D model).

Evaluation of genetic variant impacts on concordance of protein measures and epidemiological associations

To evaluate the impact of *cis*- protein altering genetic variation on protein abundances, we obtained counts of PAV alleles present for each participant and adjusted protein measures for copies of the effect allele in these variants. Protein residuals from these models were used to evaluate the concordance of adjusted-protein measures.

Furthermore, these protein residuals were used as input in epidemiological analyses and *cis*-pQTL mapping to explore the impact of PAV adjustment on downstream associations. To determine cases where adjustment for PAV resulted in more concordant phenotypic associations, we performed effect-size heterogeneity tests across SomaScan and Olink results obtained using (1) non-PAV adjusted protein measures as model input, then (2) using PAV-adjusted protein measures as input. We subsequently compared proteins with effect-size heterogeneity between models using protein measures without vs. with PAV adjustment.

To evaluate the impact of platform-specific pleiotropic *trans*-pQTL protein altering genetic variation on protein abundances, we obtained counts of lead *trans*-pQTL alleles in credible sets mapping to regions defined as platform-specific pleiotropic and adjusted protein measures for copies of the effect allele in these variants.

Adjustment for protein altering genetic variants in QTL analyses

To determine the impacts of platform-discordant PAV adjustment on *cis*-pQTL, we repeated *cis*-pQTL mapping and fine-mapping analyses for the 19 proteins associated with platform-discordant PAVs using the previously described tensorQTL pipeline, this time adjusting for copies of the respective PAV as a covariate. We then determined whether *cis*-pQTL signal significance increased or decreased following PAV adjustment by overlapping credible sets obtained from models without vs. with PAV adjustment from each platform and comparing summary statistics at the lead variants. To account for a subset of signals that appeared to increase in significance with PAV adjustment due to residual LD with the PAV adjusted for in the model, we removed credible sets from models with PAV-adjustment that contained variants in LD with variants in credible sets obtained from models without adjustment for the PAV from the results. Here, "newly-associated" *cis*-pQTL are defined as *cis*-pQTL credible sets obtained in PAV-adjusted models that do not overlap, and are not in LD ($R^2 > 0.1$) with, variants in credible sets from pre-PAV adjusted models.

Mendelian Randomization

To evaluate the potential impacts of platform differences in *cis*-pQTL on Mendelian Randomization (MR) results, we performed an exploratory MR analysis for Alzheimer's Disease (AD) using both SomaScan 7k and Olink Explore 3072 significant *cis*-pQTL for the 2157 proteins measured on both platforms as genetic instruments. MR analyses were performed in R Studio (version 4.4.0) with the MendelianRandomization package (0.10.0)[77] using the Bellenguez et al. GWAS meta-analysis[30] results as outcome data. We identified SNP instruments for each protein by choosing the most significant *cis*-pQTL variant in each platform's credible set that was also available in the outcome GWAS. Our primary analyses were performed using all *cis*-pQTL as instruments, however, we performed subsequent sensitivity analyses excluding (A) platform-discordant PAV *cis*-pQTL instruments, and (B) platform-discordant and platform-specific PAV *cis*-pQTL instruments. Wald/Inverse-Variance Weighted (IVW) MR tests were performed. For exploratory purposes, a nominal significance threshold of $p < 0.05$ was used to define significant MR results in this study.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The *cis*- and *trans*-pQTL association statistics generated in this study have been deposited in a Zenodo site under accession code 10.5281/zenodo.17642644 (https://zenodo.org/records/17642645). Only *trans*-pQTL results for variants associated with $p < 1 \times 10^{-8}$ are available in summary files. The protein measure correlations, pQTL credible sets, protein and pQTL annotations, and protein-phenotype associations generated in this study are provided in the Supplementary Information. The individual-level proteomics measures, whole genome sequences, and participant covariates are available under restricted access to ensure participant privacy; access can be obtained through dbGaP (phs001416 (https://dbgap.ncbi.nlm.nih.gov/beta/study/phs001416.v4.p1/) and phs000209 (https://dbgap.ncbi.nlm.nih.gov/beta/study/phs000209.v13.p3/)). In addition to the public access repository noted above, researchers with an established collaboration with MESA Investigators may also access the data through the MESA Coordinating Center at the University of Washington through standard MESA policies and procedures. Source data are provided with this paper.

Code availability

The code used to perform the analyses and generate results in this study is publicly available and has been deposited in GitHub "Cross-Ancestry-Comparison-of-Aptamer-and-Antibody-Protein-Measures" at https://github.com/jcnicholas/Cross-Ancestry-Comparison-of-Aptamer-and-Antibody-Protein-Measures. The specific version of the code associated with this publication is archived in Zenodo and is accessible via https://doi.org/10.5281/zenodo.17673143.

References

1. Ferkingstad, E. et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat. Genet.* **53**, 1712–1721 (2021).
2. Folkersen, L. et al. Genomic and drug target evaluation of 90 cardiovascular proteins in 30,931 individuals. *Nat. Metab.* **2**, 1135–1148 (2020).
3. Said, S. et al. Ancestry diversity in the genetic determinants of the human plasma proteome and associated new drug targets. Preprint at *medRxiv*, https://doi.org/10.1101/2023.11.13.23298365 (2023).
4. Sun, B. B. et al. Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
5. Sun, B. B. et al. Plasma proteomic associations with genetics and health in the UK Biobank. *Nature* **622**, 329–338 (2023).
6. Dhindsa, R. S. et al. Rare variant associations with plasma protein levels in the UK Biobank. *Nature* **622**, 339–347 (2023).
7. SomaScan 11K Assay - SomaLogic. https://somalogic.com/somascan-11k-assay/.
8. Home - Olink. https://olink.com/.
9. Xu, F. et al. Genome-wide genotype-serum proteome mapping provides insights into the cross-ancestry differences in cardiometabolic disease susceptibility. *Nat. Commun.* **14**, 896 (2023).
10. Suhre, K. et al. Nanoparticle enrichment mass-spectrometry proteomics identifies protein-altering variants for precise pQTL mapping. *Nat. Commun.* **15**, 989 (2024).
11. Wang, B. et al. Comparative studies of 2168 plasma proteins measured by two affinity-based platforms in 4000 Chinese adults. *Nat. Commun.* **16**, 1869 (2025).
12. Katz, D. H. et al. Proteomic profiling platforms head to head: leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Sci. Adv.* **8**, eabm5164 (2022).
13. Haslam, D. E. et al. Stability and reproducibility of proteomic profiles in epidemiological studies: comparing the Olink and SOMAscan platforms. *Proteomics* **22**, e2100170 (2022).
14. Rooney, M. R. et al. Comparison of proteomic measurements across platforms in the atherosclerosis risk in communities (ARIC) study. *Clin. Chem.* **69**, 68–79 (2023).
15. Raffield, L. M. et al. Comparison of proteomic assessment methods in multiple cohort studies. *Proteomics* **20**, e1900278 (2020).
16. Pietzner, M. et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat. Commun.* **12**, 6822 (2021).
17. Eldjarn, G. H. et al. Large-scale plasma proteomics comparisons through genetics and disease associations. *Nature* **622**, 348–358 (2023).
18. Suhre, K. et al. A genome-wide association study of mass spectrometry proteomics using a nanoparticle enrichment platform. *Nat. Genet.* **57**, 2987–2996 (2025).
19. Pietzner, M. et al. Mapping the proteo-genomic convergence of human diseases. *Science* **374**, eabj1541 (2021).
20. Olson, N. C. et al. Soluble urokinase plasminogen activator receptor: genetic variation and cardiovascular disease risk in black adults. *Circ. Genom. Precis. Med.* **14**, e003421 (2021).
21. Vasbinder, A. et al. Assay-related differences in SuPAR levels: implications for measurement and data interpretation. *J. Nephrol.* https://doi.org/10.1007/s40620-022-01344-7 (2022).
22. Ren, Y. et al. Evaluation of a large-scale aptamer proteomics platform among patients with kidney failure on dialysis. *PLoS ONE* **18**, e0293945 (2023).
23. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**, 1273–1300 (2020).
24. Pietzner, M. et al. Nanoparticle enriched mass spectrometry proteomics in British South Asians identifies novel variant-protein-disease mechanisms. Preprint at *medRxiv*, https://doi.org/10.1101/2025.08.12.25333522 (2025).
25. Wang, Q. S. et al. Platform-dependent effects of genetic variants on plasma APOL1. Preprint at *BioRxiv*, https://doi.org/10.1101/2025.01.30.635763 (2025).
26. Rooney, M. R. et al. Correlations within and between highly multiplexed proteomic assays of human plasma. *Clin. Chem.* **71**, 677–687 (2025).
27. Zhan, H., Cammann, D., Cummings, J. L., Dong, X. & Chen, J. Biomarker identification for Alzheimer's disease through integration of comprehensive Mendelian randomization and proteomics data. *J. Transl. Med.* **23**, 278 (2025).
28. Debette, S. et al. Proteogenomics in cerebrospinal fluid and plasma reveals new biological fingerprint of cerebral small vessel disease. *Res. Sq.* https://doi.org/10.21203/rs.3.rs-4535534/v1. (2024).
29. Zhu, Y., Wang, Y., Cui, Z., Liu, F. & Hu, J. Identification of pleiotropic and specific therapeutic targets for cardio-cerebral diseases: a large-scale proteome-wide mendelian randomization and colocalization study. *PLoS ONE* **19**, e0300500 (2024).
30. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
31. Li, Y. et al. Genome-wide studies reveal factors associated with circulating uromodulin and its relationships to complex diseases. *JCI Insight* **7**, e157035 (2022).
32. Reisinger, A. C. et al. Soluble urokinase plasminogen activator receptor (suPAR) predicts critical illness and kidney failure in patients admitted to the intensive care unit. *Sci. Rep.* **11**, 17476 (2021).
33. Jehn, U. et al. Soluble urokinase-type plasminogen activator receptor (suPAR) is a risk indicator for eGFR loss in kidney transplant recipients. *Sci. Rep.* **11**, 3713 (2021).
34. Walker, R. G. et al. Activated GDF11/8 subforms predict cardiovascular events and mortality in humans. *Nat. Commun.* **16**, 6534 (2025).
35. Lemmelä, S. et al. Integrated analyses of growth differentiation factor-15 concentration and cardiometabolic diseases in humans. *eLife* **11**, e76272 (2022).
36. Wang, Q. S. et al. Statistically and functionally fine-mapped blood eQTLs and pQTLs from 1,405 humans reveal distinct regulation patterns and disease relevance. *Nat. Genet.* **56**, 2054–2067 (2024).
37. Yao, C. et al. Dynamic role of trans regulation of gene expression in relation to complex traits. *Am. J. Hum. Genet.* **100**, 571–580 (2017).
38. Cai, W. et al. Master regulator genes and their impact on major diseases. *PeerJ* **8**, e9952 (2020).
39. Orchard, P. et al. Cross-cohort analysis of expression and splicing quantitative trait loci in TOPMed. Preprint at *medRxiv*, https://doi.org/10.1101/2025.02.19.25322561 (2025).
40. Kirsher, D. Y. et al. Current landscape of plasma proteomics from technical innovations to biological insights and biomarker discovery. *Commun. Chem.* **8**, 279 (2025).

41. Sissala, N. et al. Comparative evaluation of Olink Explore 3072 and mass spectrometry with peptide fractionation for plasma proteomics. *Commun. Chem.* **8**, 327 (2025).
42. Bild, D. E. et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am. J. Epidemiol.* **156**, 871–881 (2002).
43. MESA. Timelines and procedures | The multi-ethnic study of atherosclerosis. https://internal.mesa-nhlbi.org/about/timelines-and-procedures.com.
44. Tin, A. et al. Reproducibility and variability of protein analytes measured using a multiplexed modified aptamer assay. *J. Appl. Lab. Med.* **4**, 30–39 (2019).
45. Olson, J. L., Bild, D. E., Kronmal, R. A. & Burke, G. L. Legacy of MESA. *Glob. Heart* **11**, 269–274 (2016).
46. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
47. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **12**, 246 (2011).
48. Birney, E. & Soranzo, N. Human genomics: the end of the start for population sequencing. *Nature* **526**, 52–53 (2015).
49. Cavalli-Sforza, L. L. The Human Genome Diversity Project: past, present and future. *Nat. Rev. Genet.* **6**, 333–340 (2005).
50. Taliun, D. et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
51. Rohloff, J. C. et al. Nucleic acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. *Mol. Ther. Nucleic Acids* **3**, e201 (2014).
52. Brody, E. et al. Life's simple measures: unlocking the proteome. *J. Mol. Biol.* **422**, 595–606 (2012).
53. Gold, L. et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS ONE* **5**, e15004 (2010).
54. Kim, C. H. et al. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci. Rep.* **8**, 8382 (2018).
55. Williams, S. A. et al. A proteomic surrogate for cardiovascular outcomes that is sensitive to multiple mechanisms of change in risk. *Sci. Transl. Med.* **14**, eabj9625 (2022).
56. Williams, S. A. et al. Plasma protein patterns as comprehensive indicators of health. *Nat. Med.* **25**, 1851–1857 (2019).
57. SomaLogic. SomaScan® v4.0 and v4.1 Data Standardization. Accessed from www.SomaLogic.com 5 Nov 2025.
58. SomaLogic. SomaScan® Assay v4.1 White Pages. Accessed from www.SomaLogic.com 5 Nov 2025.
59. Assarsson, E. et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS ONE* **9**, e95192 (2014).
60. Olink. Olink® Explore 3072 User Manual. Accessed from www.olink.com/knowledge/documents 5 Nov 2025.
61. Olink. Olink Data Normalization and Standardization White Paper. Accessed from www.olink.com/knowledge/documents 5 Nov 2025.
62. Posit team. *RStudio: Integrated Development Environment for R*. (Posit Software, PBC, Boston, MA., 2025).
63. Frankish, A. et al. GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res.* **51**, D942–D949 (2023).
64. Dias, J.-A. et al. Evaluating multi-ancestry genome-wide association methods: Statistical power, population structure, and practical implications. *Am. J. Hum. Genet.* https://doi.org/10.1016/j.ajhg.2025.08.006 (2025).
65. Wang, N. et al. Genetic architecture and analysis practices of circulating metabolites in the NHLBI Trans-Omics for Precision Medicine Program. *Am. J. Hum. Genet.* **112**, 2720–2738 (2025).
66. Taylor-Weiner, A. et al. Scaling computational genomics to millions of individuals with GPUs. *Genome Biol.* **20**, 228 (2019).
67. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
68. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
69. The Human Protein Atlas. https://www.proteinatlas.org/.
70. Liu, X. et al. WGSA: an annotation pipeline for human genome sequencing studies. *J. Med. Genet.* **53**, 111–112 (2016).
71. McLaren, W. et al. The ensembl variant effect predictor. *Genome Biol.* **17**, 122 (2016).
72. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
73. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
74. GTEx Consortium The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
75. Sollis, E. et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).
76. R: The R Project for Statistical Computing. https://www.r-project.org/.
77. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).

Acknowledgements

Jayna Nicholas was supported in part by a grant from the National Institute of General Medical Sciences under award T32GM135128, and by fellowship 1F31HL176194-01A1 from the National Heart, Lung and Blood Institute. This project was also funded in part by U01AG082042, UM1DK078616, R01HL151855, R01HL133870, and R01DK072193. 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.v3.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). Phenotype harmonization, data management, sample-identity QC, and general study coordination, were provided by the TOPMed Data Coordinating Center (3R01HL-120393-02S1), and TOPMed MESA Multi-Omics (HHSN2682015000031/HSN26800004). The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92025D00022, 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92025D00026, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92025D00024, 75N92020D00003, N01-HC-95162, 75N92025D00027, 75N92020D00006, N01-HC-95163, 75N92025D00025, 75N92020D00004, N01-HC-95164, 75N92025D00028, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420, UL1TR001881, and R01HL105756. The authors thank the MESA participants and the MESA investigators and staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. Acquisition of proteomic data with SomaScan was funded by R01HL159081 to Ganz. Analysis at The Lundquist Institute is supported in part by NCATS CTSI grant UL1TR001881, and the NIDDK grant DK063491 to the Southern California

Diabetes Endocrinology Research Center. The MESA projects are conducted and supported by the NHLBI in collaboration with MESA investigators. MESA was supported by contracts 75N92020D00001, HHSN 268201500003I, 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 and N01-HC-95169 from the NHLBI, and by grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Support for additional participants (MESA Family Ancillary Study) was by R01HL071051, R01 HL071205, R01HL071250, R01HL071251, R01HL071258, R01HL071259, by the National Center for Research Resources, Grant UL1RR033176, and NCATS Grant UL1-TR-001881. This study was also supported in part by the NHLBI contracts R01HL151855 and R01HL146860.

Author contributions

The following contributions were determined based on CRediT roles (www.authorservices.wiley.com/author-resources/Journal-Authors/open-access/credit.html). Conceptualization—J.C.N., L.M.R., D.H.K., C.L.D., J.C. (Josef Coresh), D.G., Y.L., M.R.R., A.M. (Ani Manichaikul), K.L.M., J.I.R., P.G., R.E.G., and K.D.T. Data Curation—J.C.N., U.A.T., T.B., D.E.C., R.D., X.G., C.J., M.Y.M., R.D., S.S.R., J.I.R., P.G., and R.E.G. Formal Analysis—J.C.N. Funding Acquisition—S.S.R., J.I.R., P.G., R.E.G., and L.M.R. Methodology—J.C.N., L.M.R., D.H.K., C.L.D., K.A.B, M.R.R., and K.L.M. Project Administration—L.M.R. Resources—E.C., M.F.D., P.D., N.C.O., B.M.P., and R.P.T. Software—F.A. Supervision—L.M.R. and K.L.M. Visualization—J.C.N. Writing—Original Draft Preparation—J.C.N. and L.M.R. Writing—Review & Editing - J.C.N., L.M.R., D.H.K., U.A.T., C.L.D., F.A., T.B., R.P.B., K.A.B, J.C. (Josef Coresh), C.B.C., J.C. (Jingsha Chen), E.C., D.E.C., R.D., M.F.D., P.D., L.E., J.S.F., D.G., X.G., R.C.H., C.J., L.A.L., Y.L., A.M. (Ani Manichaikul), J.B.M., M.Y.M., J.C.M., N.C.O., K.A.P., B.M.P., A.P.R., P.R., M.S., A.M.S., Q.S., R.P.T., J.W., A.C.W., J.G.W., K.L.Y., B.Y., M.R.R., A.M. (Alisa Manning), R.D., K.L.M., S.S.R., J.I.R., P.G., R.E.G., and K.D.T.

Competing interests

L.M.R. is a consultant for the TOPMed Administrative Coordinating Center (through Westat). D.G. is the Chief Executive Officer of Sequoia Genetics, a private company that works with investors, pharma, biotech, and academia by performing research that leverages genetic data to help inform drug discovery and development. D.G. has interests in several biotech companies. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-67814-1.

Correspondence and requests for materials should be addressed to Laura M. Raffield.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

[1]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [2]Division of Cardiovascular Medicine, Division of Computational Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [3]Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [4]Department of Genome Sciences, University of Virginia, Charlottesville, VA, USA. [5]Broad Institute, Cambridge, MA, USA. [6]University of Michigan, Ann Arbor, MI, USA. [7]Genomic Sciences and Systems Biology, Cleveland Clinic Research, Cleveland, OH, USA. [8]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [9]Metabolomics Platform, Broad Institute, Cambridge, MA, USA. [10]Department of Population Health, Institute for Optimal Aging, New York, NY, USA. [11]Department of Pathology and Laboratory Medicine, Larner College of Medicine at the University of Vermont, Burlington, VT, USA. [12]Division of Cardiovascular Medicine, University of Pennsylvania, Philadelphia, PA, USA. [13]University of Mississippi Medical Center, Jackson, MS, USA. [14]School of Medicine, University of Washington, Seattle, WA, USA. [15]Sequoia Genetics, London, UK. [16]Department of Pediatrics, The Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. [17]Division of Cardiovascular Research, Department of Medicine, Baylor College of Medicine, Houston, TX, USA. [18]University of Washington, Seattle, WA, USA. [19]Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. [20]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [21]Broad Institute, Boston, MA, USA. [22]Harvard University, Boston, MA, USA. [23]Massachusetts General Hospital, Boston, MA, USA. [24]Department of Medicine, Broad Institute, Boston, MA, USA. [25]Division of General Internal Medicine, Department of Medicine, Harvard Medical School, Boston, MA, USA. [26]Division of Cardiovascular Medicine, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [27]Department of Biostatistics, National Jewish Health, Denver, CO, USA. [28]Cardiovascular Health Research Unit, University of Washington, Seattle, WA, USA. [29]Department of Medicine, University of Washington, Seattle, WA, USA. [30]Department of Epidemiology, University of Washington, Seattle, WA, USA. [31]Fred Hutchinson Cancer Research Center, University of Washington, Seattle, WA, USA. [32]Department of Health Data Science and Biostatistics, University of Texas Southwestern Medical Center, Dallas, TX, USA. [33]Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA, USA. [34]Department of Medicine, Harvard Medical School, Boston, MA, USA. [35]Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA, USA. [36]Department of Medicine, Division of Cardiology, University of Texas Southwestern Medical Center, Dallas, TX, USA. [37]Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. [38]Center for Computational and Genomic Medicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [39]Department of Biomedical and

Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [40]Department of Pediatrics, USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA. [41]Department of Cardiology, Beth Israel Deaconess Medical Center, Boston, MA, USA. [42]Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [43]Department of Epidemiology and Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA. [44]Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [45]UT Southwestern, Dallas, TX, USA. [46]Division of Cardiology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. ✉e-mail: laura_raffield@unc.edu